

Ana García Reyes

*Análisis y modelización de series temporales. Estudio de la demanda de los recambios en vehículos de transporte público.*

Analysis and modeling of time series. Study of the demand for spare parts in public transport vehicles.

Trabajo Fin de Grado  
Grado en Matemáticas  
La Laguna, Septiembre de 2022

DIRIGIDO POR  
*Carlos Pérez González*  
*Ginés León Rodríguez*

*Carlos Pérez González*  
*Departamento de Estadística e*  
*Investigación Operativa*  
*Universidad de La Laguna*  
*38200 La Laguna, Tenerife*

*Ginés León Rodríguez*  
*Departamento de Data Science y*  
*Big Data*  
*Transportes Interurbanos de*  
*Tenerife S.A.*  
*38111 Santa Cruz de Tenerife,*  
*Tenerife*

---

## Agradecimientos

En primer lugar, quisiera darle las gracias a mi familia, a mis padres y a mi hermana, por haberme dado la oportunidad de formarme, por ser los grandes pilares en mi vida, por los consejos, valores y principios que me han inculcado.

A mi tutor académico, Carlos Pérez González, por compartir sus conocimientos de manera profesional e invaluable, por su dedicación y perseverancia. Y a mi tutor de prácticas externas, Ginés León Rodríguez, que me ha permitido aplicar muchos de los conocimientos adquiridos durante la carrera a problemas de la vida real.

A mi pareja, por su paciencia, comprensión, cariño y apoyo, las cuales me han servido para dar el último empujón que me faltaba.

Y a mis amigos, que siempre me han prestado un gran apoyo moral y humano, necesarios en los momentos más difíciles.

Ana García Reyes  
La Laguna, 8 de septiembre de 2022



---

## Resumen · Abstract

### *Resumen*

---

*Saber cómo evolucionarán las observaciones en un futuro próximo ha sido siempre objeto de estudio por los científicos. Cómo evolucionará el precio del alquiler en los próximos meses, el número de habitantes en cierto país por año o la temperatura media mensual, son algunos ejemplos donde se utilizan las series temporales.*

*En este trabajo se estudiarán diferentes modelos de series temporales. En particular, nos centraremos en el Modelo de regresión de Poisson, así como el Modelo Oculito de Markov. Además, se realizará un estudio con datos reales proporcionados por la empresa TITSA, en el que se ajustarán varios modelos para predecir la cantidad de cierto tipo de piezas que se pedirán en los talleres de la empresa.*

**Palabras clave:** *Series temporales – Modelo de regresión de Poisson – Modelo Oculito de Markov*

### *Abstract*

---

*Knowing about the evolution of data from short-term observations has drawn the attention of many scientists. How the rental price will evolve over a number of months, the number of inhabitants in a certain country per year or the average monthly temperature, are some examples where time series are used.*

*In this project, different time series models will be studied. Particularly, we will be focus on the Poisson Regression Model, as well as the Hidden Markov Model. Furthermore, a study will be performed with real data provided by the company TITSA, in which several models will be adjusted to predict the quantity of certain types of pieces that will be ordered in the company's workshops.*

**Keywords:** *Time series – Poisson Regression Model – Hidden Markov Model*



---

# Contenido

|   |     |
|---|-----|
| <b>Agradecimientos</b> .....  | III |
| <b>Resumen/Abstract</b> .....   | V   |
| <b>Introducción</b> .....   | XI  |
| <b>1. Introducción y modelización de series temporales</b> .....      | 1   |
| 1.1. Conceptos básicos y ejemplo .....                                | 1   |
| 1.2. Modelos y ejemplos de series temporales .....                    | 4   |
| 1.2.1. Modelo Naïve, SNaïve .....                                     | 4   |
| 1.2.2. Descomposición estacional .....                                | 5   |
| 1.2.3. Suavizado exponencial .....                                    | 8   |
| 1.2.4. Modelos ARIMA y SARIMA .....                                   | 8   |
| 1.2.5. Modelos lineales dinámicos .....                               | 12  |
| 1.2.6. Modelo TBATS .....   | 13  |
| 1.2.7. Modelo Prophet .....   | 14  |
| 1.2.8. Modelo NNETAR .....  | 16  |
| 1.2.9. Modelo LSTM .....  | 18  |
| <b>2. Modelos para datos de conteo: regresión de Poisson</b> .....    | 21  |
| 2.1. Introducción .....   | 21  |
| 2.2. Distribución de Poisson .....                                    | 21  |
| 2.3. Modelo de regresión de Poisson .....                             | 22  |
| 2.3.1. Entrenamiento del modelo de regresión .....                    | 24  |
| <b>3. Modelos de transición de estados: modelo oculto de Markov</b> . | 27  |
| 3.1. Conceptos básicos sobre las cadenas de Markov .....              | 27  |
| 3.2. Modelo Oculto de Markov .....                                    | 29  |
| 3.3. Modelo Poisson - oculto de Markov (Poisson-HMM) .....            | 30  |
| 3.3.1. Proceso de Poisson .....                                       | 31  |
| 3.3.2. Unión del modelo oculto de Markov .....                        | 31  |

|  |           |
|--|-----------|
| 3.3.3. Entrenamiento y estimación .....                                | 33        |
| <b>4. Predicción con datos reales: piezas de taller de TITSA .....</b> | <b>35</b> |
| 4.1. Introducción al problema .....                                    | 35        |
| 4.2. Preparación de los datos .....                                    | 35        |
| 4.3. Entrenamiento y validación de modelos .....                       | 36        |
| 4.4. Comparación de las predicciones .....                             | 41        |
| 4.4.1. MSE y RMSE .....  | 41        |
| 4.4.2. MAE .....   | 42        |
| <b>Conclusiones .....</b>  | <b>44</b> |
| <b>Bibliografía .....</b>  | <b>47</b> |
| <b>Poster .....</b>  | <b>49</b> |

*“Prediction is truly very difficult, especially if it’s about the unknown future.”*  
— Nils Bohr, Premio Nobel de Física.



---

## Introducción

El pronóstico de series temporales, en términos estadísticos, es el proceso de analizar los datos de series temporales utilizando estadísticas y modelos para hacer predicciones y tomar decisiones estratégicas. El análisis, el modelado y la previsión de series temporales tienen mucha esencia en los ejemplos prácticos de la vida real y en una variedad de campos de investigación, incluidos los negocios, la economía, la medicina, la astronomía, la ingeniería, la política y muchos más.

Los desarrollos teóricos en el análisis de series temporales comenzaron con los procesos estocásticos. La primera aplicación real de modelos autorregresivos a los datos se remonta al trabajo de G. U. Yule y J. Walker en las décadas de 1920 y 1930.

Durante este tiempo, se introdujo el promedio móvil para eliminar las oscilaciones periódicas en la serie temporal, por ejemplo, las oscilaciones debidas a la estacionalidad. Herman Wold introdujo modelos ARMA para series estacionarias, pero no pudo derivar una función de probabilidad para permitir la estimación de máxima verosimilitud (MLE) de los parámetros.

En 1970 se publicó el libro *“Time Series Analysis: Forecasting and Control”* de G. E. P. Box y G. M. Jenkins, que contenía el procedimiento completo de modelado para series individuales: especificación, estimación, diagnóstico y pronóstico.

En esta memoria se hablará sobre algunos de los modelos más usados en la predicción de series temporales, así como la aplicación de alguno de ellos a datos reales.

En el Capítulo 1 se introducirán algunos de los conceptos necesarios para entender las series temporales. Además, se describirán de manera general algunos de los modelos más usados en las series temporales, como pueden ser el modelo Naïve y SNaïve (1.2.1), la descomposición estacional (1.2.2), el modelo ARIMA (1.2.4), ...

A continuación, en el Capítulo 2, se hablará sobre el Modelo de Regresión de Poisson, modelo más avanzado, útil para modelar variables de respuesta (va-

lores  $Y$ ) que son recuentos. Este modelo selecciona qué variables explicativas tienen un efecto estadísticamente significativo en la variable de respuesta. Es un modelo apropiado para describir eventos raros, ya que estos tienden a seguir una distribución de Poisson.

Seguidamente, en el Capítulo 3 se define el Modelo Oculto de Markov (HMM), en el cuál los parámetros observados se utilizan para identificar los parámetros ocultos. Estos parámetros se utilizan luego para un análisis posterior. El HMM es un tipo de cadena de Markov. Su estado no se puede observar directamente, pero se puede identificar observando la serie de vectores.

Y, por último, en el Capítulo 4 aplicaremos algunos de los modelos introducidos a un caso real de análisis de datos de demanda de piezas en talleres de la empresa TITSA, en particular para predecir el número de *filtros de aire* que se instalan en los vehículos de transportes público cuando entran en el taller.

## Introducción y modelización de series temporales

En este capítulo veremos una introducción a los distintos conceptos que serán necesarios para el desarrollo de la memoria. Además, se dará una descripción general de diferentes modelos de predicción de series temporales.

### 1.1. Conceptos básicos y ejemplo

Las siguientes definiciones serán de gran utilidad para entender algunos conceptos que se introducirán en capítulos y secciones posteriores.

**Definición 1.1.** Una serie temporal es una secuencia de observaciones de valores reales  $y_t$ , cada una de ellas tomada en un tiempo específico  $t$ .

**Definición 1.2.** Un proceso estocástico es una secuencia de variables aleatorias, ordenadas y equidistantes cronológicamente,  $\{y_t : t = 1, 2, \dots\}$ , y sirve como modelo para una serie temporal observada.

**Definición 1.3.** Sea  $\{y_t\}$  un proceso estocástico  
La función media de  $\{y_t\}$  viene dada por

$$\mu_y(t) = E(y_t), \quad t = 1, 2, \dots \quad (1.1)$$

La función covarianza de  $\{y_t\}$  es

$$\gamma_y(r, s) = \text{Cov}(y_r, y_s) = E[(y_r - \mu_r)(y_s - \mu_s)] \quad \forall r, s \in \mathbb{Z} \quad (1.2)$$

**Definición 1.4.** Las medias móviles se usan para estimar el ciclo de tendencia. Una media móvil de orden  $m$  se puede escribir como:

$$T_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j} \quad (1.3)$$

donde  $m = 2k + 1$ . Es decir, es el promedio de los valores de la serie temporal dentro de  $k$  períodos de tiempo  $t$ .

**Definición 1.5.** Para algún punto específico  $r$ , la observación  $x_{r-i}$  ( $i$  períodos hacia atrás) se llama el  $i$ -ésimo retraso de  $x_r$ .

**Definición 1.6.** Un proceso estocástico  $\{y_t\}$  es estacionario cuando

1.  $\mu_y(t)$  es independiente de  $t$ .
2.  $\gamma_y(t+h, t)$  es independiente de  $t$  para cada  $h$ ,  $h \in \mathbb{Z}$

**Definición 1.7.** Los pasos aleatorios (Random Walks)  $\{S_t, t = 0, 1, 2, \dots\}$  se obtiene sumando acumulativamente variables aleatorias independientes e idénticamente distribuidas.

**Definición 1.8.** Sea  $\{y_t\}$  una serie temporal estacionaria.

La función de autocovarianza de  $\{y_t\}$  con retraso  $h$  es

$$\gamma_y(h) = \text{Cov}(y_{t+h}, y_t) \quad (1.4)$$

La función de autocorrelación de  $\{y_t\}$  con retraso  $h$  es

$$\rho_y(h) = \frac{\rho_y(h)}{\rho_y(0)} = \text{Cor}(y_{t+h}, y_t) \quad (1.5)$$

**Definición 1.9.** A la secuencia  $\{y_t\}$  de variables aleatorias, independientes e idénticamente distribuidas, cada una con media 0 y varianza  $\sigma^2$  se le conoce como ruido blanco.

**Definición 1.10.** Un proceso estocástico es de tendencia estacionaria si se puede expresar como

$$y_i = f(i) + \epsilon_i \quad (1.6)$$

donde  $f(i)$  es cualquier función  $f: \mathbb{R} \rightarrow \mathbb{R}$  y  $\epsilon_i$  es un proceso estocástico estacionario con media 0.

**Definición 1.11.** En los modelos de regresión lineal se dice que hay heterocedasticidad cuando la varianza de los errores no es igual en todas las observaciones realizadas.

$$\text{Var}(\epsilon_i | X) = \sigma_i^2 = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{pmatrix} \quad (1.7)$$

La varianza del error en la observación  $i$  condicionado a  $X$  (variable explicativa) es igual a la varianza de la misma observación.

**Definición 1.12.** Una función sigmoide es una función real de variable real diferenciable, de la forma:

$$P(t) = \frac{1}{1 + e^{-t}} \quad (1.8)$$

Tiene dos asíntotas horizontales:

$$\lim_{t \rightarrow -\infty} \left( \frac{1}{1+e^{-t}} \right) = 0$$

$$\lim_{t \rightarrow +\infty} \left( \frac{1}{1+e^{-t}} \right) = 1$$

Con una primera derivada no negativa:

$$P'(t) = \frac{dP}{dt} = \frac{e^{-t}}{(1+e^{-t})^2}$$

Esta función es útil en redes neuronales artificiales (1.2.9).

**Definición 1.13.** Una *distribución binomial* es una distribución de probabilidad discreta que describe el número de éxitos al realizar  $n$  experimentos independientes entre sí, acerca de una variable aleatoria. La fórmula para calcular la distribución normal es

$$P_x = \binom{n}{x} p^x q^{n-x} \quad (1.9)$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad (1.10)$$

donde:

$n$  = número de ensayos/experimentos

$x$  = número de éxitos

$p$  = probabilidad de éxito

$q$  = probabilidad de fracaso ( $1 - p$ )

*Ejemplo 1.14.* A modo de ejemplo de serie temporal, en la Figura 1.1 se muestran las ventas mensuales (en kilolitros) de vino tinto australiano desde enero de 1980 hasta octubre de 1991. Se puede decir que las ventas tienen una tendencia ascendente y un patrón estacional con un máximo en julio y un mínimo en enero.

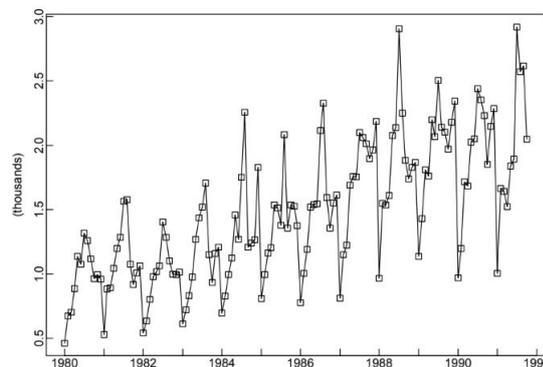


Figura 1.1. Ventas de vino tinto australiano 1980-1991

## 1.2. Modelos y ejemplos de series temporales

Veremos una breve descripción sobre los siguientes modelos:

- Naïve, SNaïve
- Descomposición estacional (Seasonal decomposition)
- Suavizado exponencial (Exponential smoothing)
- ARIMA, SARIMA
- Modelos lineales dinámicos (Dynamic linear models)
- TBATS
- Prophet
- NNTAR
- LSTM

Estos modelos permiten obtener estimaciones y/o predicciones de  $y_t$  que denotaremos como  $\hat{y}_t$ .

Cada uno de los modelos nombrados anteriormente, se utilizará para predecir la evolución del índice de producción industrial que cuantifica los equipos eléctricos fabricados en la Eurozona.

### 1.2.1. Modelo Naïve, SNaïve

El modelo **Naïve** es un modelo en el que se utilizan cantidades mínimas de esfuerzo y manipulación de datos para preparar un pronóstico. En este modelo se le da importancia únicamente a la última observación, ignorando el resto. Por lo tanto, el modelo será de la forma

$$\hat{y}_{t+h|t} = y_t \quad (1.11)$$

La notación  $\hat{y}_{t+h|t}$  es una forma abreviada de la estimación  $\hat{y}_{t+h}$  basado en los datos  $y_1, y_2, \dots, y_t$ .

Este tipo de pronósticos asume que el modelo estocástico (Definición 1.2) que genera la serie temporal es el resultado de hacer sucesivos pasos aleatorios (Definición 1.7).

Una extensión del modelo es el método **SNaïve**, el cuál es útil para datos altamente estacionales. En este caso, se establece que el pronóstico para una temporada, debe ser igual al último valor observado en la misma temporada. El modelo será de la forma

$$\hat{y}_{t+h|t} = y_{t+h-T} \quad (1.12)$$

donde T es el período de estacionalidad.

Estos modelos se utilizan a menudo como modelos de referencia para modelos más complejos.

Las Figuras 1.2 y 1.3 muestran la comparativa de los dos modelos anteriores para las predicciones de la fabricación de equipos eléctricos en el año 2007.

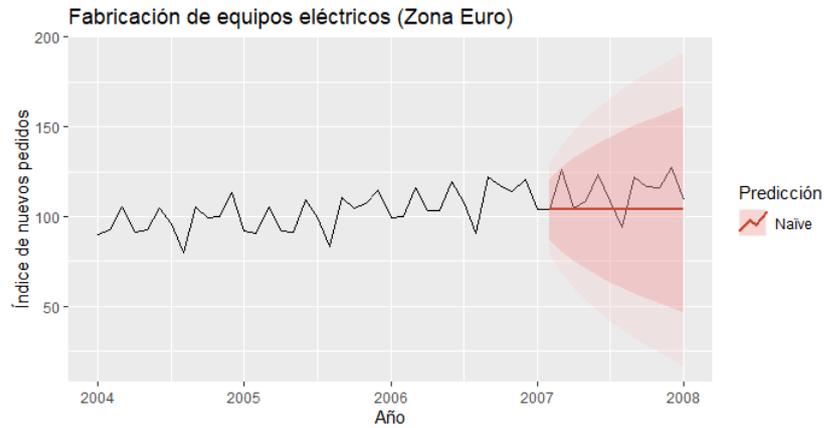


Figura 1.2. Predicción año 2007 usando el modelo Naïve

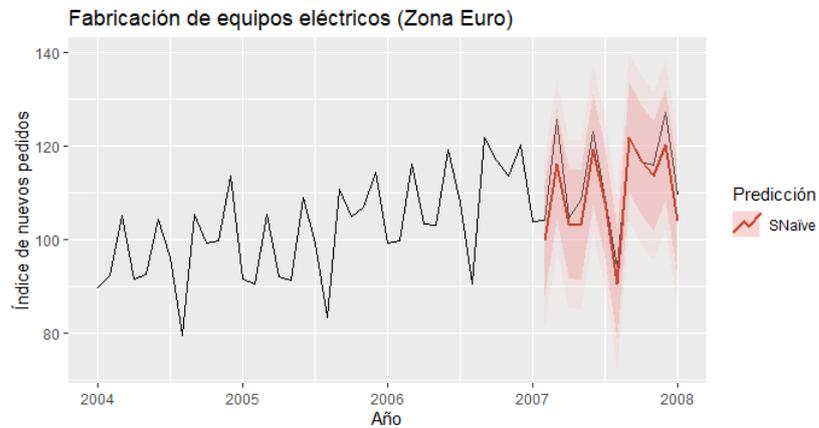


Figura 1.3. Predicción año 2007 usando el modelo SNaïve

### 1.2.2. Descomposición estacional

La **descomposición estacional** se utiliza en series temporales para describir la tendencia y los factores estacionales de las mismas. Cuando los datos presentan cierta estacionalidad, puede ser útil descomponer la serie temporal original en varias componentes, cada una de las cuales representa una de las categorías subyacentes de patrones.

Uno de los principales objetivos de una descomposición es estimar los efectos estacionales que se pueden utilizar para crear y representar valores ajustados estacionalmente.

Las dos estructuras siguientes se consideran modelos básicos de descomposición estacional:

1. Aditivo:

$$\hat{y}_t = S_t + T_t + R_t \quad (1.13)$$

Este modelo es útil cuando la variación estacional es relativamente constante en el tiempo

2. Multipliativo:

$$\hat{y}_t = S_t \times T_t \times R_t \quad (1.14)$$

Mientras que el modelo multiplicativo es útil cuando la variación estacional aumenta con el tiempo.

Las componentes son:

- $S_t$  es la componente estacional, es decir, representa la periodicidad de los datos.
- $T_t$  es la componente del ciclo de tendencia. Se puede definir como el comportamiento o movimiento a largo plazo de los datos.
- $R_t$  es la componente restante. Recoge las alteraciones de la serie que no tiene una pauta periódica ni tendencia reconocible.

Existen varias técnicas para estimar cada descomposición. Los primeros métodos clásicos obtenían las componentes de forma básica, a esto se le llamó *descomposición clásica* y consiste en:

1. Estimar la tendencia  $T_t$  a través de una media móvil (Definición 1.4).
2. Calcular  $S_t$  como la serie sin tendencia promedio  $\hat{y}_t - T_t$  para cada temporada.
3. Calcular la serie restante como  $R_t = \hat{y}_t - T_t - S_t$ .

La descomposición más versátil y robusta es la descomposición STL (*Seasonal and Trend decomposition using Loess*) (Figura 1.4). Esta descomposición se maneja bien con la estacionalidad, la cuál puede cambiar con el tiempo, y el ratio de cambio puede ser controlado por el usuario, así como el suavizado del ciclo de tendencia.

En la Figura 1.5 se muestra el ajuste estacional de la serie temporal correspondiente a la fabricación de equipos eléctricos.

Y la predicción obtenida para el año 2007 utilizando la descomposición estacional se puede observar en la Figura 1.6.

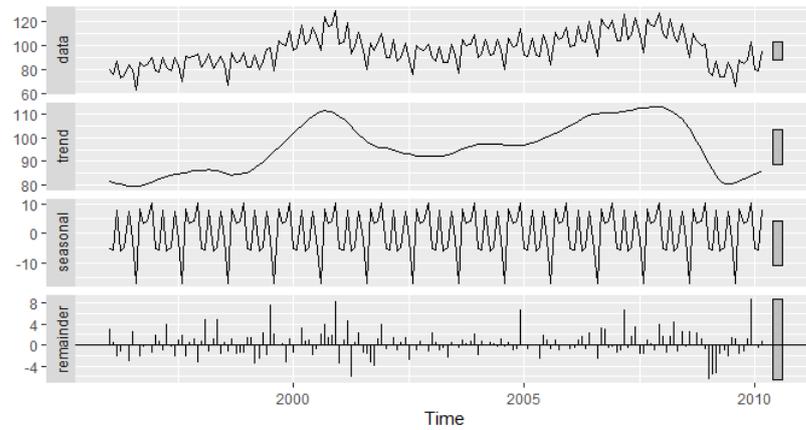


Figura 1.4. Descomposición STL en los datos del índice de producción industrial

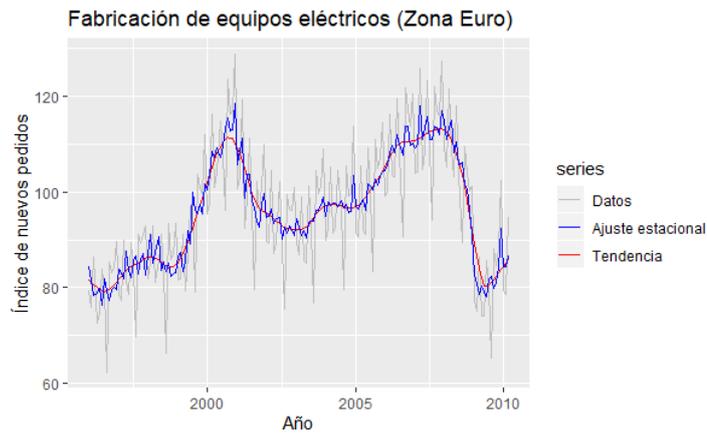


Figura 1.5. Serie temporal del índice de producción industrial desestacionalizado

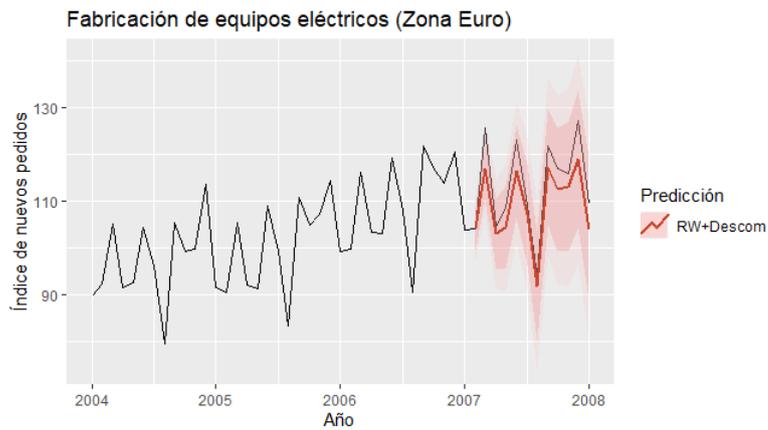


Figura 1.6. Predicción para el año 2007 usando la descomposición estacional

### 1.2.3. Suavizado exponencial

El **suavizado exponencial** es uno de los métodos clásicos de pronósticos más exitosos. Su máxima eficacia se obtiene cuando los datos siguen una tendencia gradual y presentan un comportamiento estacional, es decir, cuando los valores sigan un patrón cíclico repetido a lo largo de un número determinado de períodos. Sus pronósticos vienen dados por

$$\hat{y}_{t+h|t} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \dots + \alpha(1-\alpha)^{t-1} y_1 \quad (1.15)$$

donde  $0 < \alpha < 1$  es el parámetro de suavizado. La predicción para el tiempo  $t+1$  es un promedio ponderado de todas las observaciones en la serie  $y_1, y_2, \dots, y_t$ . La velocidad a la que disminuyen los pesos está controlada por el parámetro  $\alpha$ .

Todos los datos influyen en el pronóstico, pero los pesos de las observaciones disminuyen exponencialmente a medida que retrocedemos en el tiempo (Figura 1.7).

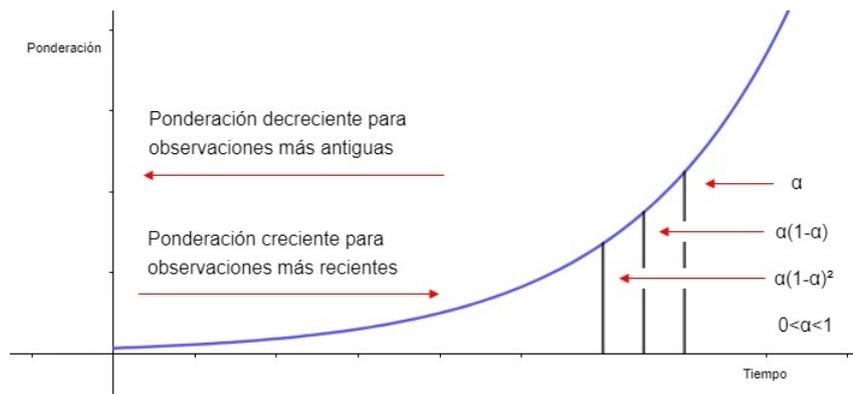


Figura 1.7. Gráfico ponderación - tiempo

Las ventajas más relevantes de los métodos de suavizado son:

- No se necesitan muchos datos históricos.
- Tiene una mayor precisión que otros modelos al utilizar técnicas de modelado exponencial.
- Es un método muy flexible.

Las Figuras 1.8 y 1.9 muestran las predicciones obtenidas para el año 2007 usando el suavizado exponencial para ajustar la serie temporal original y la ajustada estacionalmente.

### 1.2.4. Modelos ARIMA y SARIMA

Estos dos modelos tienen base en los modelos ARMA. En un modelo autorregresivo (**A**uto**R**egressive) los pronósticos corresponden a una combinación

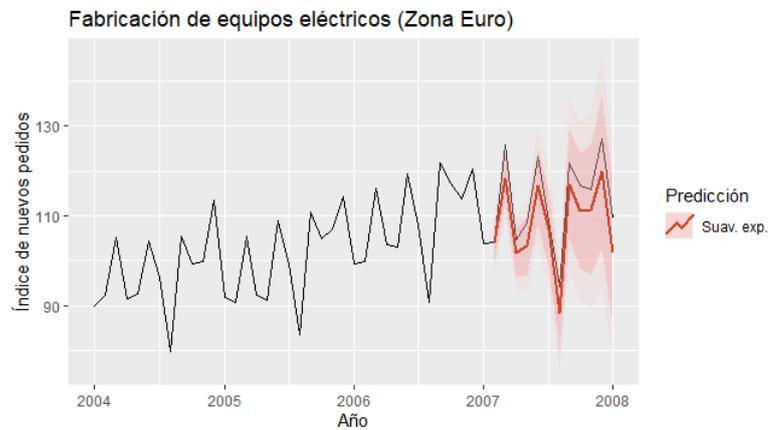


Figura 1.8. Predicción año 2007 serie original usando suavizado exponencial

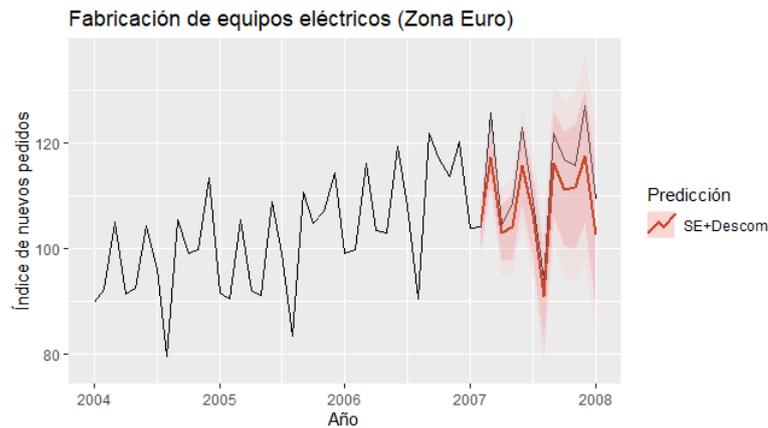


Figura 1.9. Predicción año 2007 serie ajustada usando suavizado exponencial

lineal de valores pasados de la variable. Mientras que en un modelo de promedio móvil (**Moving Average**), los pronósticos corresponden a una combinación lineal de errores de pronósticos anteriores.

Una generalización de los modelos ARMA, que incorporan series temporales no estacionarias (Definición 1.6) utilizando la diferenciación, son los modelos **ARIMA** (**AutoRegressive Integrated Moving Average**). Estos modelos intentan describir la autocorrelación (Definición 1.8) de los datos de la serie y, al igual que el suavizado exponencial, los modelos ARIMA son los más utilizados en el pronóstico de series temporales.

Cuando una serie no estacionaria se quiere transformar en estacionaria, se recurre a la diferenciación. Para hablar de la diferenciación es importante hablar del operador de retardo y el operador de diferencia:

**Definición 1.15.** Se define el operador de retardo  $B$ , también llamado  $L$  (*Lag operator*), como un operador lineal que aplicado a una función temporal proporcional a esta misma función pero retardada un período.

$$By_t = y_{t-1} \quad \forall t > 1$$

**Definición 1.16.** El operador de diferencia  $\Delta$  calcula la diferencia entre observaciones consecutivas  $\Delta = 1 - B$ . La diferencia  $d$ -ésima viene dada por el operador  $\Delta^d = (1 - B)^d$

*Ejemplo 1.17.* La serie diferenciada de primer orden es

$$\Delta y_t = y_t - y_{t-1} = (1 - B)y_t$$

La serie diferenciada de segundo orden es

$$(y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

o lo que es igual,

$$\Delta^2 y_t = (1 - B)^2 y_t = (1 - 2B - B^2)y_t = y_t - 2y_{t-1} + y_{t-2}$$

La diferenciación puede ayudar a estabilizar la media de una serie temporal y, por lo tanto, reducir o eliminar la tendencia y la estacionalidad.

Cada una de las componentes del modelo se especifica como un parámetro. Se utiliza la notación  $ARIMA(p, d, q)$  con  $p, d, q \in \mathbb{Z}$ . Los parámetros del modelo ARIMA se definen de la siguiente manera:

- $p$ : El número de observaciones pasadas incluidas en el modelo, también llamado orden de retardo.
- $d$ : El grado de la primera diferenciación involucrada, también llamado grado de diferencia.
- $q$ : El tamaño del promedio móvil, también denominado orden del promedio móvil.

Por lo tanto, el modelo se puede descomponer en dos partes:

1. La componente basada en la autorregresión  $AR(p)$ :

$$\phi(y_t) = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (1.16)$$

donde  $\epsilon_t$  es el error,  $\phi_i$  los parámetros y  $c$  una constante.

2. La componente basada en las medias móviles  $MA(q)$ :

$$\theta(y_t) = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (1.17)$$

donde  $\epsilon_t$  es el error,  $\theta_i$  los parámetros y  $c$  una constante.

Si se combina la autorregresión, las medias móviles y la diferenciación, obtenemos el modelo:

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\epsilon_t \quad (1.18)$$

El modelo **SARIMA** (**S**easonal **ARIMA**) se forma al incluir términos estacionales adicionales en los modelos ARIMA. Este se escribe como

$$ARIMA(p, d, q) \times (P, D, Q)_m$$

donde  $m$  es el período de estacionalidad.

La parte estacional del modelo consta de términos que son similares a los componentes no estacionales del modelo, pero implican desplazamientos hacia atrás del período estacional.

En los modelos SARIMA, el operador de diferencia estacional de período  $m$  viene dado por

$$\Delta_m = 1 - B^m$$

De este modo

$$\Delta_m y_t = y_t - y_{t-m}$$

Por lo tanto, el modelo será:

$$\phi(B)\Phi(B^m)(1 - B)^d(1 - B^m)^D y_t = c + \theta(B)\Theta(B^m)\epsilon_t \quad (1.19)$$

Las Figuras 1.10 y 1.11 muestran los pronósticos obtenidos para el año 2007 utilizando el modelo SARIMA y el ARIMA sobre la serie temporal desestacionalizada.

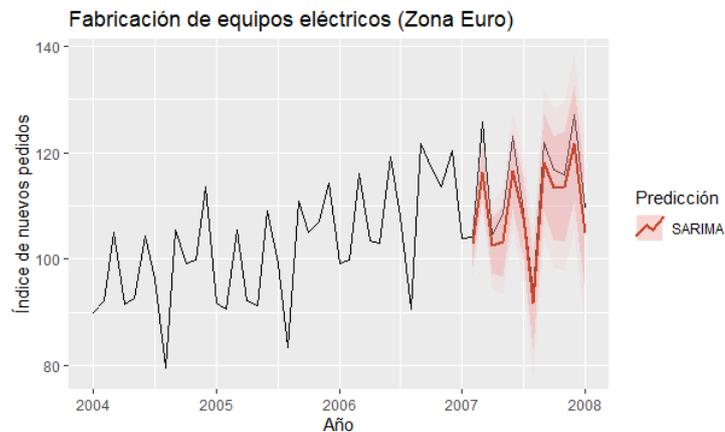


Figura 1.10. Predicción año 2007 usando el modelo SARIMA

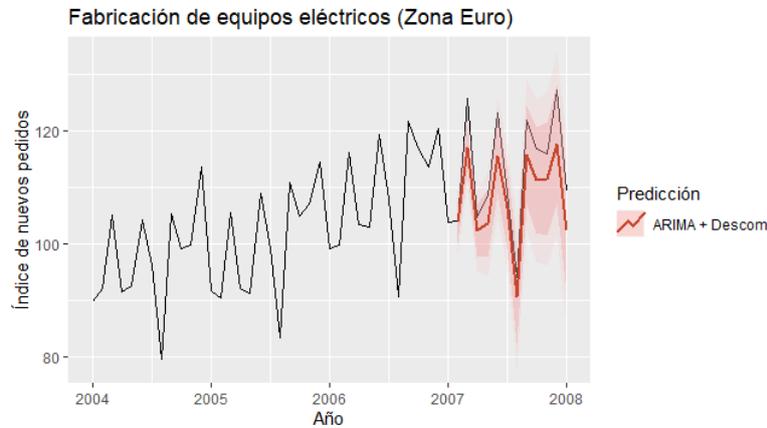


Figura 1.11. Predicción año 2007 usando el modelo ARIMA

### 1.2.5. Modelos lineales dinámicos

Los **modelos lineales dinámicos** representan otra clase de modelos para el pronóstico de series temporales. La idea es que en cada tiempo  $t$  estos modelos correspondan a un modelo lineal, donde los coeficientes de regresión cambian con el tiempo. El modelo lineal dinámico general se puede escribir con la ayuda de la ecuación de observación y la ecuación de estado, como

$$\hat{y}_t = \begin{cases} y_t = F_t \theta_t + \epsilon_t & \text{(ecuación de observación)} \\ \theta_t = G_t \theta_{t-1} + \omega_t & \text{(ecuación de evolución)} \end{cases} \quad (1.20)$$

donde:

- $y_t$  es la observación en el tiempo  $t$ . Ese valor puede ser un escalar o un vector.
- $\theta_t = (\theta_{1,t}, \dots, \theta_{p,t})$  es el vector de parámetros en el tiempo  $t$  y de dimensión  $p \times 1$ .
- $F_t$  es el vector fila (dimensión  $1 \times p$ ) de covariables en tiempo  $t$ .
- $G_t$  es una matriz de dimensión  $p \times p$ , conocida como la *matriz de evolución o transición*.
- Por lo general,  $F_t$  y  $G_t$  están definidas.
- $\epsilon_t$  es el error de observación en el tiempo  $t$  y  $\omega_t$  es el error de evolución (vector  $p \times 1$ ).
- $\epsilon_t$  es independiente de  $\epsilon_s$ ,  $\omega_t$  es independiente de  $\omega_s$  para  $t \neq s$ . Son dos secuencias independientes entre ellas.

A continuación, se muestra la Figura 1.12 con los pronósticos para el año 2007 usando el modelo lineal dinámico para ajustar estacionalmente la serie temporal. Dado que el modelo tiene altos costos computacionales, este se mantuvo extremadamente simple, por lo que los pronósticos son deficientes.

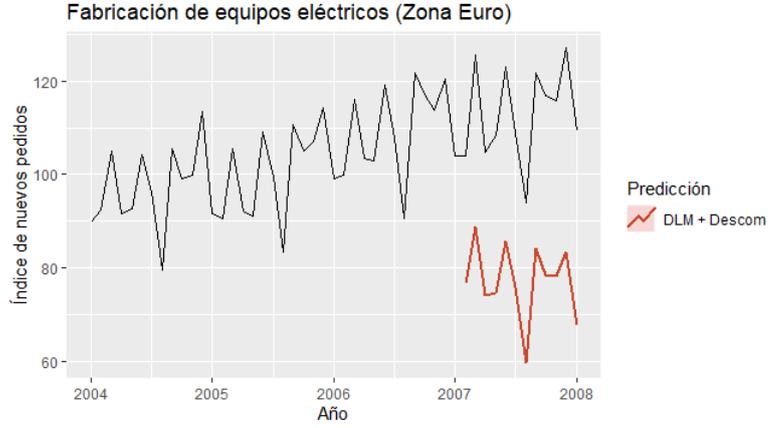


Figura 1.12. Predicción año 2007 usando el modelo lineal dinámico

### 1.2.6. Modelo TBATS

El modelo **TBATS** es un método de pronóstico de series temporales con patrones estacionales complejos. El nombre es un acrónimo de Trigonométrico, transformada de Box-Cox, errores ARMA, componentes de Tendencia y Estacionalidad (**T**rigonometric, **B**ox-Cox transform, **A**RMA errors, **T**rend and **S**easonal components).

Muchas series temporales exhiben patrones estacionales complejos y múltiples. Los modelos más populares (por ejemplo, ARIMA y Suavizado exponencial) solo pueden tener en cuenta una estacionalidad.

Pero el modelo TBATS tiene la capacidad de lidiar con estacionalidades complejas (por ejemplo, estacionalidad no entera, anidada o de grandes períodos). Al modelar cada estacionalidad con una representación trigonométrica basada en series de Fourier, se crean pronósticos detallados a largo plazo.

El modelo es el siguiente:

$$\hat{y}_t^{(\omega)} = \begin{cases} \frac{y_t^\omega - 1}{\omega} & \omega \neq 0 \\ \log y_t & \omega = 0 \end{cases}$$

$$\hat{y}_t^{(\omega)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t$$

en el cual las expresiones  $\ell_t$ ,  $b_t$  y  $d_t$  vienen dadas por:

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = \phi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t$$

donde:

- $\hat{y}_t^{(\omega)}$  es la serie temporal en el momento  $t$ .

- $m_i$  denotan los períodos estacionales.
- $\ell_t$  es el nivel local en el periodo  $t$ .
- $b_t$  es la tendencia a corto plazo en el período  $t$ .
- $s_t^{(i)}$  representa la  $i$ -ésimo componente estacional en el tiempo  $t$ .
- $d_t$  denota un proceso  $ARMA(p, q)$ .
- $\epsilon_t$  es un proceso de ruido blanco gaussiano.

La Figura 1.13 representa los pronósticos para el año 2007 del ejemplo, usando el modelo TBATS.

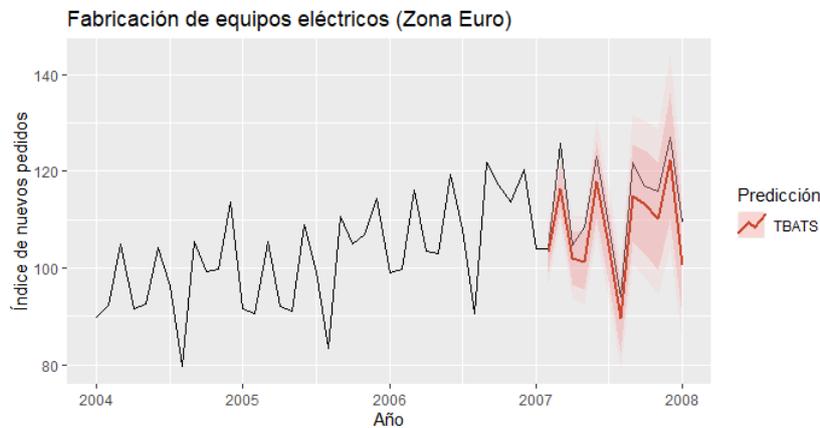


Figura 1.13. Predicción año 2007 usando el modelo TBATS

### 1.2.7. Modelo Prophet

El modelo **Prophet** es otro modelo de predicción que permite tratar con múltiples estacionalidades. Este modelo asume que la serie temporal se puede descomponer de la siguiente manera

$$\hat{y}_t = g(t) + s(t) + h(t) + \epsilon(t) \quad (1.21)$$

donde:

- $g(t)$  describe una tendencia lineal por tramos (o “*término de crecimiento*”). Los nodos para esta tendencia se seleccionan de forma automática si no se especifican. O bien, se puede usar una función logística para establecer un límite superior en la tendencia.
- $s(t)$  describe los patrones estacionales. Estas componentes consisten en términos de Fourier de los períodos relevantes.
- $h(t)$  captura los efectos de los días festivos. Estos efectos se agregan como variables ficticias simples.

- $\epsilon(t)$  es el error de ruido blanco (Definición 1.9).

El ajuste del modelo se enmarca como un ajuste de curvas, por lo que no tiene en cuenta, explícitamente, la estructura de dependencia temporal de los datos. Esto permite tener observaciones espaciadas irregularmente.

Hay dos modelos de tendencia: un modelo de crecimiento saturado y un modelo lineal por partes.

### MODELO DE CRECIMIENTO SATURADO

Para la previsión del crecimiento, la componente central del proceso de generación de datos es un modelo que muestra cómo ha crecido la población y cómo se espera que siga creciendo. Este modelo es similar al crecimiento de la población en ecosistemas naturales, donde hay un crecimiento no lineal que se satura a una capacidad de carga:

$$\frac{dN}{dT} = r_{max} \frac{(K-N)}{K} N$$

donde  $r_{max}$  es la tasa máxima de crecimiento per cápita,  $N$  es el tamaño poblacional y  $K$  la capacidad de carga.

Este modelo de crecimiento saturado generalmente se crea utilizando el modelo de crecimiento logístico, que en su forma básica es

$$g(t) = C / \{1 + e^{-k(t-m)}\}$$

con  $C$  la capacidad de carga,  $k$  la tasa de crecimiento y  $m$  un parámetro de compensación.

### MODELO LINEAL POR PARTES

Hay dos aspectos importantes que no se contemplan en el modelo anterior.

Primero, la capacidad de carga no es constante. Se reemplaza entonces el valor fijo  $C$  por una capacidad variable  $C(t)$ . Y, en segundo lugar, la tasa de crecimiento no es constante. Los nuevos productos pueden alterar la tasa de crecimiento, por lo que el modelo debe ser capaz de incorporar una tasa variable para adaptarse a los datos históricos.

Supongamos que hay  $S$  puntos de cambio en los tiempos  $s_j, j = 1, \dots, S$ . La tasa de crecimiento en cualquier momento  $t$  es

$$k + \sum_{j:t>s_j} \delta_j$$

con  $k$  tasa de crecimientos base y  $\delta_j$  es el cambio de tasa que ocurre en el tiempo  $s_j$ .

Se puede representar más claramente definiendo

$$a_j(t) = \begin{cases} 1 & \text{si } t \geq s_j \\ 0 & \text{en otro caso} \end{cases}$$

Cuando se ajusta la tasa  $k$ , el parámetro de compensación  $m$  también debe ajustarse para conectar los puntos finales del segmento. El ajuste correcto en el punto de cambio  $j$  se calcula como

$$\gamma_j = \left( s_j - m - \sum_{l < j} \gamma_l \right) \left( 1 - \frac{k + \sum_{l < j} \delta_l}{k + \sum_{l \leq j} \delta_l} \right)$$

Por tanto, el modelo de crecimiento logístico por partes es entonces

$$g(t) = C(t) / \{1 + e^{-(k+a(t)^T \delta)(t-(m+a(t)^T \gamma))}\}$$

A continuación, se muestra la Figura 1.14 con el pronóstico para el año 2007, del ejemplo base, usando el modelo Prophet para ajustar la serie temporal.

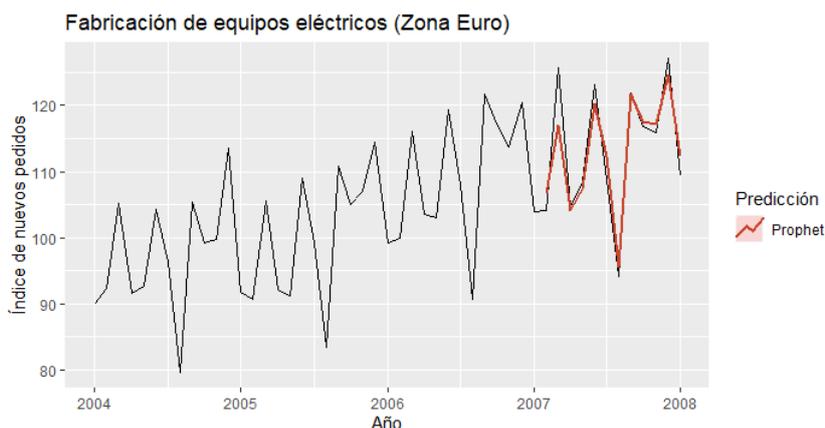


Figura 1.14. Predicción año 2007 usando el modelo Prophet

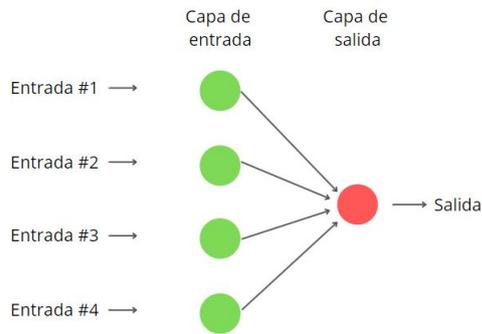
### 1.2.8. Modelo NNETAR

El modelo **NNETAR** es una red neuronal totalmente conectada. El acrónimo significa Autorregresión de Red Neuronal (**N**eural **NET**work **A**uto**R**egression).

Se puede pensar en una red neuronal como una red de “neuronas” que se organizan en capas. Los predictores (o entradas) forman la capa inferior y los pronósticos (o salidas) forman la capa superior. También puede haber capas intermedias que contengan “neuronas ocultas”.

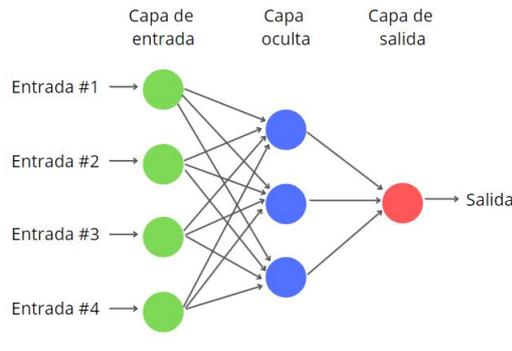
Las redes más simples no contienen capas ocultas y son equivalentes a las regresiones lineales.

La Figura 1.15 muestra una red neuronal con cuatro predictores. Los coeficientes adjuntos a estos predictores se denominan “pesos”. Los pronósticos se obtienen mediante una combinación lineal de las entradas. Los pesos se seleccionan en el marco de la red neuronal utilizando un “algoritmo de aprendizaje”.



**Figura 1.15.** Red Neuronal simple

Una vez que se le añaden neuronas ocultas, la red neuronal se vuelve no lineal (Figura 1.16).



**Figura 1.16.** Red Neuronal con tres neuronas ocultas

Esto se conoce como una “*red de múltiples capas*”, donde cada capa de nodos recibe entradas de las capas anteriores. Las salidas de los nodos en una capa son entradas para la siguiente.

Con datos de series temporales, los valores retrasados (Definición 1.5) se pueden usar como entrada para una red neuronal, tal como usamos valores retrasados en un modelo de autorregresión lineal.

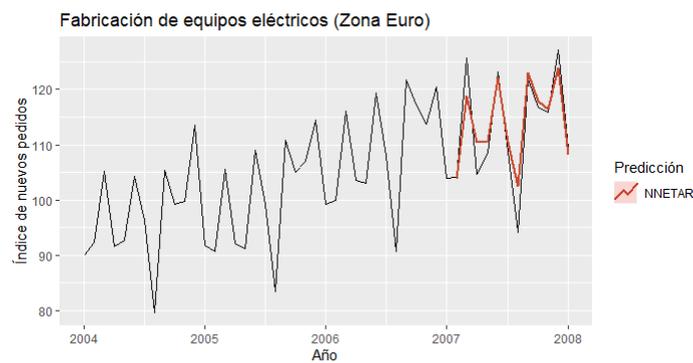
El modelo NNETAR toma como entrada los últimos elementos de la secuencia hasta el tiempo  $t$  y genera el valor pronosticado en el tiempo  $t + 1$ . Para realizar los pronósticos de varios pasos, la red se aplica de forma iterativa.

Cuando haya estacionalidad en los datos, la entrada puede incluir también la serie temporal desfasada estacionalmente.

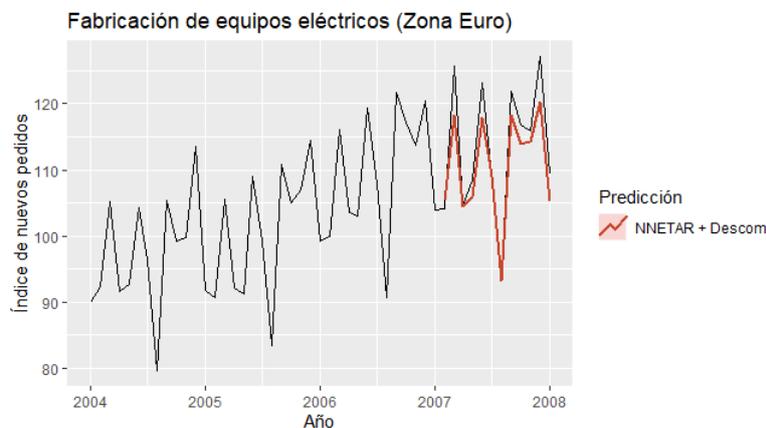
Si solo se consideran las redes de retroalimentación con una capa oculta, y usamos la notación  $NNAR(p, k)$  para indicar que hay  $p$  entradas retrasadas ( $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-p}$ ) y  $k$  nodos en la capa oculta. Un modelo  $NNAR(p, 0)$  es

equivalente al modelo  $ARIMA(p, 0, 0)$ , pero sin las restricciones en los parámetros para asegurar la estacionalidad.

Las predicciones obtenidas para el año 2007 de nuestro ejemplo, usando el modelo NNETAR con entrada desfasada estacionalmente y un modelo NNETAR en la serie temporal ajustada estacionalmente serían: (Figura 1.17 y 1.18)



**Figura 1.17.** Predicción año 2007 usando el modelo NNETAR



**Figura 1.18.** Predicción año 2007 usando el modelo NNETAR + Descomposición

### 1.2.9. Modelo LSTM

Los modelos **LSTM** se pueden usar para pronosticar series temporales (así como otras redes neuronales recurrentes). LSTM es un acrónimo de Memorias a Largo y Corto Plazo (**L**ong-**S**hort **T**erm **M**emories).

Estos modelos son un tipo especial de red neuronal recurrente que es capaz de aprender dependencia a largo plazo en los datos.

La red LSTM está compuesta de varios bloques de memoria llamados celdas. Una celda LSTM tiene cinco componentes vitales que le permiten utilizar datos a corto y largo plazo: el estado de la celda, el estado oculto, la puerta de entrada, la puerta de olvido y la puerta de salida.

- **Capa de puerta de olvido:** la decisión de qué información va a pasar desde el estado de la celda la toma esta capa. Da un número entre 0 y 1, para cada número en el estado de la celda, usando la función sigmoide (Definición 1.12). Mientras que 1 muestra “*deja pasar la entrada*”, 0 significa “*no deja pasar la entrada*”.
- **Capa de puerta de entrada:** gestiona el proceso de adición de información al estado de la celda. En primer lugar, regula qué valores deben agregarse al estado de la celda mediante el uso de una función sigmoide. Luego, crea un vector que incluye todos los valores posibles que se pueden agregar al estado de la celda usando la función  $\tanh$ , que genera valores de  $-1$  a  $+1$ . Multiplica el valor del filtro por el vector creado y así transfiere esta información útil al estado de la celda a través de la operación suma.
- **Capa de puerta de salida:** la red selecciona información útil del estado actual de la celda y muestra cómo la salida se realiza a través de la puerta de salida.

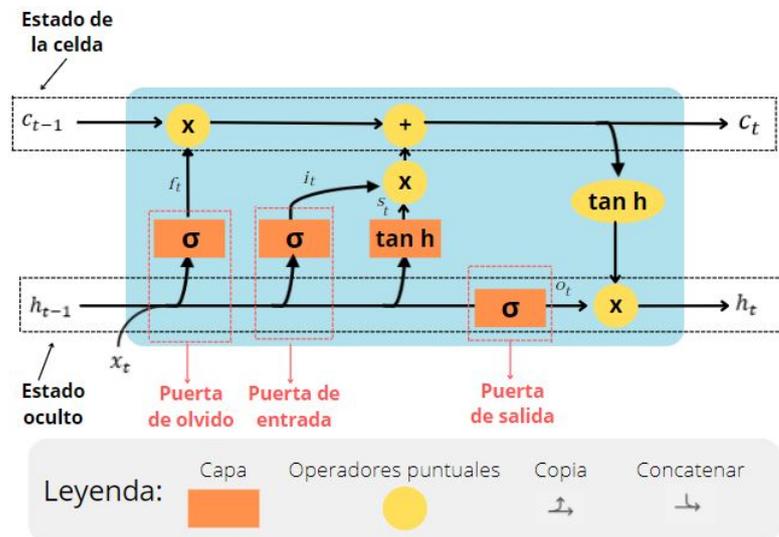


Figura 1.19. Modelo LSTM

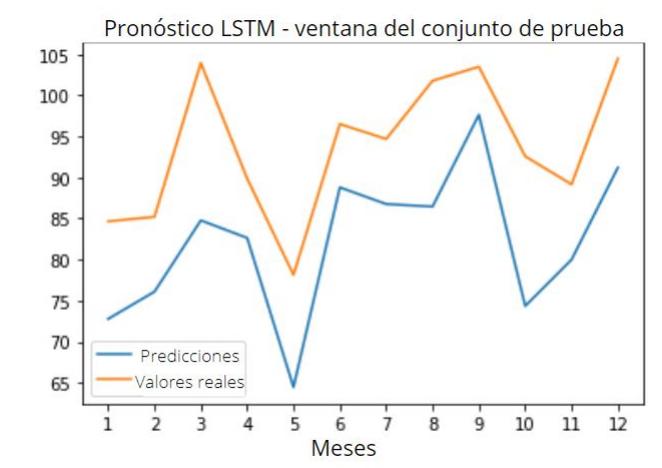
En la Figura 1.19 se muestra un diseño del modelo con cuatro capas de red neuronal en rectángulos naranjas y cinco operadores puntuales en círculos amarillos. Las variables que se muestran son:

- $\vec{x}_t$ : vector de entrada a la unidad LSTM.

- $f_t$ : vector de activación de la puerta de olvido.
- $i_t$ : vector de activación de la puerta de entrada/actualización.
- $s_t$ : vector de activación de la entrada a la celda.
- $o_t$ : vector de activación de la puerta de salida.
- $h_t$ : vector de estado oculto, también conocido como vector de salida de la unidad LSTM.
- $c_t$ : vector de estado de la celda.
- $\tanh$ : función de activación.
- $\sigma$ : función sigmoide de activación.

En términos generales, los LSTM son modelos complejos y rara vez se utilizan para predecir una sola serie temporal, ya que requieren una gran cantidad de datos para su estimación. Sin embargo, se usan comúnmente cuando se necesitan predicciones para una gran cantidad de series temporales.

La Figura 1.20 muestra las predicciones para el primer año en el conjunto de prueba obtenido al ajustar un modelo LSTM en la serie temporal ajustada estacionalmente.



**Figura 1.20.** Predicción para el primer año usando el modelo LSTM

## Modelos para datos de conteo: regresión de Poisson

En el capítulo 1 se introdujeron varios modelos de predicción de series temporales. En este capítulo nuestro principal objetivo será hablar sobre el modelo de regresión de Poisson, útil para modelar datos de series temporales que consisten en recuentos de números enteros. Se estudiará la estructura del modelo, su estimación y cómo construir, y entrenar el modelo.

### 2.1. Introducción

El modelo de regresión de Poisson y el modelo de regresión binomial negativa son dos técnicas populares para desarrollar modelos de regresión para datos de conteo.

Hay que tener en cuenta que los conjuntos de datos de recuento presentan las siguientes características:

- **Datos de números enteros:** los datos consisten en número enteros no negativos  $(0, 1, \dots, \infty)$
- **Distribución sesgada:** los datos pueden contener una gran cantidad de puntos para solo unos pocos valores, lo que hace que la distribución de frecuencia sea bastante sesgada.
- **Escasez:** los datos pueden reflejar la ocurrencia de un evento raro lo que hace que los datos sean escasos.
- **Tasa de ocurrencia:** se puede suponer que existe una cierta tasa de ocurrencia de eventos  $\lambda$  que impulsa la generación de dichos datos. La tasa de eventos puede variar con el tiempo.

### 2.2. Distribución de Poisson

La distribución de Poisson es una distribución de probabilidad discreta que expresa la probabilidad de que ocurra un número dado de eventos en un intervalo fijo de tiempo si estos eventos ocurren con una tasa media constante

conocida e independientemente del tiempo transcurrido desde el último evento. La distribución de Poisson es una forma límite de la distribución binomial en la que el número de observaciones es muy grande.

**Definición 2.1.** Sea  $x$  una variable aleatoria discreta, se dice que  $x$  sigue una distribución de Poisson con parámetro  $\lambda$  si su función de probabilidad viene dada por (Figura 2.1):

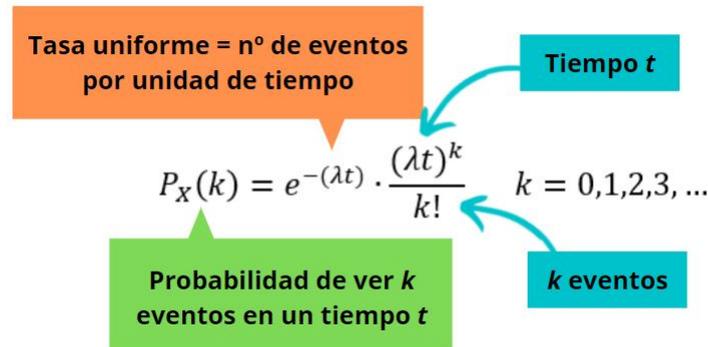


Figura 2.1. Distribución de Poisson

donde  $k = n^{\circ}$  veces que ocurre un evento y  $\lambda =$  parámetro positivo que representa el  $n^{\circ}$  de veces que se espera que ocurra un evento en un período determinado.

Si  $x \sim \text{Poisson}(\lambda)$  entonces la variable aleatoria  $x$  satisface algunas propiedades:

- Los eventos son independientes.
- No pueden ocurrir dos eventos al mismo tiempo.
- $E(x) = \text{Var}(x) = \lambda$ .
- A medida que  $\lambda$  crece, la distribución de Poisson es aproximadamente una distribución normal por el *Teorema Central del Límite*

**Teorema 2.2 (Teorema Central del Límite).** Sea  $x_1, \dots, x_n$  una muestra suficientemente grande de variables aleatorias independientes e idénticamente distribuidas ( $n > 30$ ), con  $E(x) = \mu$  y  $\text{Var}(x) = \sigma^2$ , ambos finitos. Entonces:

$$\bar{x} := \frac{\sum_{i=1}^n x_i}{n} \approx N\left(\mu, \frac{\sigma^2}{n}\right) \quad (2.1)$$

### 2.3. Modelo de regresión de Poisson

En los modelos de regresión de Poisson, se puede dar la situación en la que  $\lambda$  (Sección 2.2) puede cambiar de una observación a la siguiente. En este caso, se

asume que el valor de  $\lambda$  está influenciado por un vector de variables explicativas, también conocidas como *predictores*, *variables de regresión* o *regresores*. A esta matriz de variables de regresión la llamaremos  $\mathbf{X}$ .

El trabajo del modelo de regresión de Poisson es ajustar los recuentos observados  $y_t$  a la matriz de regresión  $\mathbf{X}$  a través de una función que expresa el vector de tasas  $\boldsymbol{\lambda}$  en función de los coeficientes de regresión  $\boldsymbol{\beta}$  y la matriz de regresión  $\mathbf{X}$  (Figura 2.2).

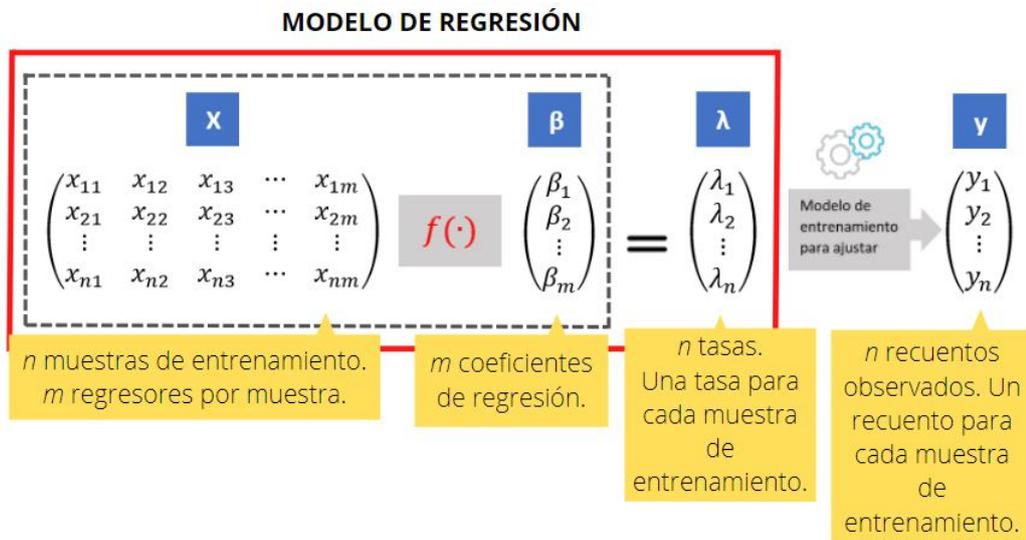


Figura 2.2. Estructura del modelo de regresión de Poisson

La función de enlace,  $f(\cdot)$ , es la función exponencial (Figura 2.3). Dicha función mantiene  $\boldsymbol{\lambda}$  no negativo incluso cuando los regresores  $\mathbf{X}$  o los coeficientes de regresión  $\boldsymbol{\beta}$  tienen valores negativos.



Figura 2.3. Función de enlace exponencial del modelo de regresión de Poisson

Para la  $i$ -ésima observación,  $\mathbf{x}_i$ , correspondiente a la  $i$ -ésima fila de la matriz  $\mathbf{X}$ , la probabilidad de observar el conteo  $y_i$  sigue una distribución de Poisson (Figura 2.4):

$$P(y_i|x_i) = \frac{e^{-\lambda_i} \cdot \lambda_i^{y_i}}{y_i!}$$

Figura 2.4. Probabilidad de observar el conteo  $y_i$  dado  $x_i$

Una vez que el modelo está completamente entrenado (Sección 2.3.1), es decir, se conocen los coeficientes de regresión  $\beta$ , el modelo está listo para hacer predicciones. Para predecir el recuento de eventos  $y_p$  correspondiente a una fila de entrada de regresores  $\mathbf{x}_p$  que se ha observado, se utiliza la fórmula de la Figura 2.5:

$$y_p = \lambda_p = e^{x_p \beta}$$

Figura 2.5. Predicción para el modelo de regresión de Poisson

### 2.3.1. Entrenamiento del modelo de regresión

El entrenamiento de un modelo de regresión de Poisson implica encontrar los valores de los coeficientes de regresión  $\beta$  que harían que el vector de recuentos observados  $y_t$  sea más probable.

La técnica para identificar los coeficientes  $\beta$  se denomina **Estimación de Máxima Verosimilitud (MLE)** (Maximum Likelihood Estimation). Esta técnica consiste en:

1. **Calcular la probabilidad conjunta de ocurrencia:** La probabilidad conjunta de ocurrencia de  $y_1, y_2, \dots, y_n$  puede expresarse como una simple multiplicación de las probabilidades individuales. (Ecuación 2.2)

$$L(\beta) = P(\mathbf{y}|\mathbf{X}) = P(y_1|\mathbf{x}_1) \cdot P(y_2|\mathbf{x}_2) \cdot \dots \cdot P(y_n|\mathbf{x}_n) = \prod_{i=1}^n P(y_i|\mathbf{x}_i) \quad (2.2)$$

2. **Calcular la función logarítmica de verosimilitud:** Esta ecuación se obtiene tomando el logaritmo natural de ambos lados de la Ecuación 2.2, después de sustituir  $\lambda_i$  por  $e^{x_i\beta}$ . La solución de la ecuación 2.3 produce el mismo valor óptimo de  $\beta$ .

$$\ln L(\beta) = \sum_{i=1}^n (y_i x_i \beta - e^{x_i \beta} - \ln y_i!) \quad (2.3)$$

3. **Derivar la función logarítmica de verosimilitud:** Diferenciamos con respecto a  $\beta$  y la igualamos a 0. Esta operación nos da la Ecuación 2.4.

$$\sum_{i=1}^n (y_i - e^{x_i \beta}) x_i = 0 \quad (2.4)$$

En resumen, los coeficiente de regresión  $\beta$  es la solución de la ecuación que se obtiene al diferenciar la ecuación de probabilidad conjunta con respecto a  $\beta$  y ajustando esta ecuación diferencial a 0. Esto producirá la estimación de máxima verosimilitud (MLE) para  $\beta$ .



## Modelos de transición de estados: modelo oculto de Markov

En este capítulo se hablará sobre el modelo oculto de Markov, una herramienta estadística idónea para modelar sistemas dinámicos, y que es utilizada para el reconocimiento de patrones y problemas de clasificación.

### 3.1. Conceptos básicos sobre las cadenas de Markov

**Definición 3.1.** Se dice que una secuencia de variables aleatorias discreta  $\{C_t : t \in \mathbb{N}\}$  es una cadena de Markov (MC) si  $\forall t \in \mathbb{N}$  satisface la propiedad de Markov

$$P(C_{t+1}|C_t, \dots, C_1) = P(C_{t+1}|C_t) = P(C_{t+1}|\mathbf{C}^{(t)}) \quad (3.1)$$

En otras palabras, describe el comportamiento de un sistema que salta de un estado a otro con una cierta probabilidad, y esta probabilidad de transición al siguiente estado depende solo del estado en el que se encuentra el sistema actualmente, es decir, no depende de los estados en los que se encontraba el sistema anteriormente.

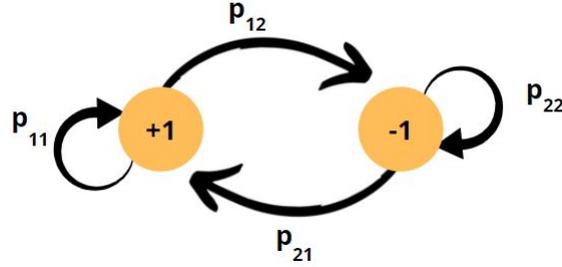
Formalmente, una cadena de Markov es una tupla de 3,  $(\mathbf{Q}, \mathbf{A}, \boldsymbol{\pi})$ , donde las componentes son:

- **Q:** es el conjunto de  $n$  estados.
- **A:** es la matriz de probabilidad de transición (Definición 3.3), donde cada  $a_{ij}$  representa la probabilidad de pasar del estado  $i$  al estado  $j$ .
- **$\boldsymbol{\pi}$ :** es una distribución de probabilidad inicial sobre los estados.  $\boldsymbol{\pi}$  es la probabilidad de que la cadena de Markov comience en el estado  $i$ .  $\sum_{i=1}^n \pi_i = 1$

*Ejemplo 3.2.* La cadena de Markov que se muestra en la Figura 3.1 tiene dos estados o *regímenes*: +1 y -1. Hay cuatro tipo de transiciones de estado posible entre los dos estados:

- Estado +1 a estado +1: Esta transición ocurre con probabilidad  $p_{11}$ .
- Estado +1 a estado -1: Ocurre con probabilidad  $p_{12}$ .

- Estado -1 a estado +1: Ocurre con probabilidad  $p_{21}$ .
- Estado -1 a estado -1: Ocurre con probabilidad  $p_{22}$ .



**Figura 3.1.** Diagrama de transición de estado del proceso de Markov

**Definición 3.3.** En el tiempo  $t$ , la probabilidad de que  $C_t$  tome algún valor  $j$  dado que en  $C_{t-1}$  había tomado un valor  $i$ , viene dada por la probabilidad condicional, llamada probabilidad de transición 3.2:

$$p_{ij} = P(C_t = j | C_{t-1} = i) \quad (3.2)$$

Para una cadena de Markov sobre  $n$ -estados  $(1, 2, 3, \dots, n)$  la matriz de probabilidad de transición es 3.3:

$$\Gamma = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} \quad (3.3)$$

Esta matriz tiene las siguientes propiedades importantes:

- Cada elemento  $p_{ij}$  es una probabilidad, por lo tanto, toma valores  $0 \leq p_{ij} \leq 1$ .
- Cada fila debe sumar 1, es decir,  $p_i = \sum_{j=1}^n p_{ij} = 1$

**Definición 3.4.** Las probabilidades incondicionales  $P(C_t = j)$  de una cadena de Markov se denota por el vector fila

$$u(t) = (P(C_t = 1), \dots, P(C_t = m)), \quad t \in \mathbb{N} \quad (3.4)$$

Si llamamos  $u(1)$  a la distribución inicial de la cadena de Markov. Para deducir la distribución en el tiempo  $t + 1$ :

$$u(t + 1) = u(t)\Gamma = u(1)\Gamma^t \quad (3.5)$$

| Día $t$  | Día $t + 1$ |         |
|----------|-------------|---------|
|          | Lluvioso    | Soleado |
| Lluvioso | 0.9         | 0.1     |
| Soleado  | 0.6         | 0.4     |

**Tabla 3.1.** Tabla de probabilidad del clima

*Ejemplo 3.5.* Imagine que la secuencia de días lluviosos y soleados es tal que el clima de cada día depende sólo del día anterior, y las probabilidades de transición están dadas por la tabla 3.1.

Es decir, si hoy llueve, la probabilidad de que mañana llueva es 0.9. En cambio, si hoy hace sol la probabilidad de que mañana llueva es de 0.6. El tiempo es entonces una cadena de Markov homogénea de dos estados, con matriz de probabilidad de transición:

$$\Gamma = \begin{pmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \end{pmatrix}$$

Ahora suponga que hoy es un día soleado ( $t = 1$ ). Esto significa que la distribución del clima de hoy es

$$u(1) = (P(C_1 = 1), P(C_1 = 2)) = (0, 1)$$

La distribución del tiempo de mañana, pasado mañana, y así sucesivamente, se puede calcular multiplicando repetidamente  $u(1)\Gamma$ , las matrices de probabilidad de transición serán:

$$\begin{aligned} u(2) &= (P(C_2 = 1), P(C_2 = 2)) = u(1)\Gamma = (0.6, 0.4) \\ u(3) &= (P(C_3 = 1), P(C_3 = 2)) = u(2)\Gamma = (0.78, 0.22) \\ &\vdots \\ u(n) &= u(1)\Gamma^{n-1} \end{aligned}$$

## 3.2. Modelo Oculto de Markov

**Definición 3.6.** *Un modelo oculto de Markov (HMM)  $\{Y_t : t \in \mathbb{N}\}$  es un tipo particular de mezcla dependiente, con  $Y_t$  variable aleatoria en el tiempo  $t$ . Las variables  $\mathbf{Y}^{(t)}$  y  $\mathbf{C}^{(t)}$  representan las secuencias de tiempo 1 al tiempo  $t$ . Se puede resumir el modelo más simple por:*

$$P(C_t | \mathbf{C}^{(t-1)}) = P(C_t | C_{t-1}), \quad t = 2, 3, \dots \quad (3.6)$$

$$P(Y_t | \mathbf{Y}^{(t-1)}, \mathbf{C}^{(t)}) = P(Y_t | C_t), \quad t \in \mathbb{N} \quad (3.7)$$

*El modelo se divide en dos partes:*

1. **Proceso de parámetros no observados:**  $\{C_t : t = 1, 2, \dots\}$  que satisface la propiedad de Markov.

2. **Proceso dependiente del estado:**  $\{Y_t : t = 1, 2, \dots\}$  cuando  $C_t$  es conocido, la distribución de  $Y_t$  depende solo del estado actual de  $C_t$  y no es sus estados previos. Esta estructura está representada por la figura 3.2.

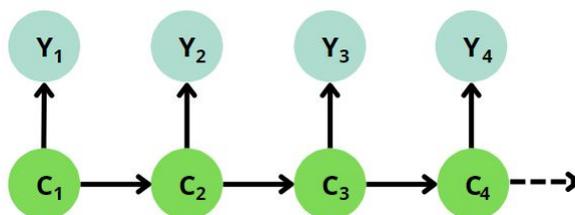


Figura 3.2. Gráfico dirigido del HMM básico.

Un HMM es una tupla compuesta por 5 componentes,  $(Q, A, \pi, V, B)$ :

- **Q:** es el conjunto de  $n$  estados.
- **A:** es la matriz de probabilidad de transición (Definición 3.3), donde cada  $a_{ij}$  representa la probabilidad de pasar del estado  $i$  al estado  $j$ .
- **$\pi$ :** es una distribución de probabilidad inicial sobre los estados.  $\pi$  es la probabilidad de que la cadena de Markov comience en el estado  $i$ .  $\sum_{i=1}^n \pi_i = 1$
- **V:** es el conjunto de posibles observaciones.
- **B:** es una secuencia de probabilidades de observación, también llamadas *probabilidades de emisión*, cada una de las cuales expresa la probabilidad de que se genere una observación de un estado  $i$ .

La razón por la que se les llama modelos “ocultos” de Markov se debe a que las observaciones son causadas por algún estado “oculto” o un estado que no se observa directamente.

### 3.3. Modelo Poisson - oculto de Markov (Poisson-HMM)

El modelo Poisson-HMM utiliza una combinación de dos procesos aleatorios. Por una parte utiliza el modelo de Markov de  $n$  estados discreto que está “oculto”, en el sentido de que, en cada paso de la serie temporal no se sabe con certeza en qué estado del modelo Markov se encuentra el proceso. Y, por otro lado, se estima el efecto de la posible existencia de cada estado en el valor medio que predice el modelo de Poisson. Esto hace que el proceso de Poisson haga significativa una variable aleatoria cuyo valor esperado es la probabilidad de que el modelo de Markov subyacente se encuentre en un estado particular.

Primero se elabora el proceso de Poisson ‘visible’ y luego se mostrará como el proceso de Markov se ‘mezcla’ con el proceso de Poisson.

### 3.3.1. Proceso de Poisson

Suponemos que el vector  $\vec{y}$ , de recuentos observados, obedece a un proceso de Poisson. Y, por tanto,  $y_t$  para  $t = 1, 2, \dots, n$  son  $n$  variables aleatorias, independientes e idénticamente distribuidas, con distribución de Poisson, cada una con una media  $\lambda_t$ , como se muestra en la Figura 3.3.

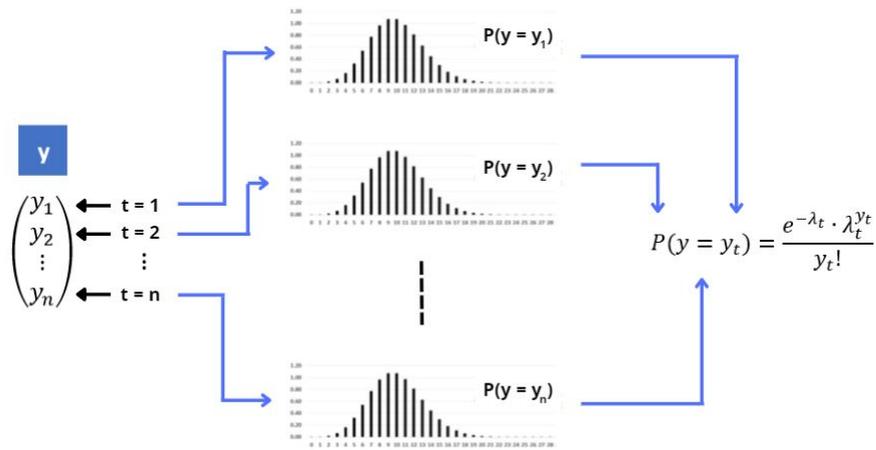


Figura 3.3. Serie temporal  $y$  que obedece al proceso de Poisson

Siguiendo los pasos del Capítulo 2, pero utilizando la matriz de variables de regresión (Figura 3.4) y el vector de coeficientes de regresión (Figura 3.5), obtenemos la ecuación 3.8.

$$y_t = \hat{\lambda}_t + \epsilon_t \tag{3.8}$$

donde:

- $\hat{\lambda}_t = e^{\mathbf{x}_t \cdot \hat{\boldsymbol{\beta}}}$  es el vector de valores predicho, donde  $\mathbf{x}_t \cdot \hat{\boldsymbol{\beta}} = \hat{\beta}_0 1 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \dots + \hat{\beta}_m x_{mt}$ .
- $\epsilon_t$  es el error residual. Suponemos, además, que es una variable aleatoria homocedástica con distribución  $N(0, \sigma^2)$ .

### 3.3.2. Unión del modelo oculto de Markov

Se considera un proceso de Markov de  $k$  estados que está en algún estado  $j \in 1, 2, \dots, k$  en el tiempo  $t$ , aunque no se sabe en cuál. Solo asumimos que influye en el modelo de proceso de Poisson de la forma:

$$y_t = \hat{\lambda}_{tj} + \epsilon_t \text{ cuando } C_t = j \tag{3.9}$$

donde  $\hat{\lambda}_{tj} = e^{\mathbf{x}_t \cdot \hat{\boldsymbol{\beta}}_j}$ .

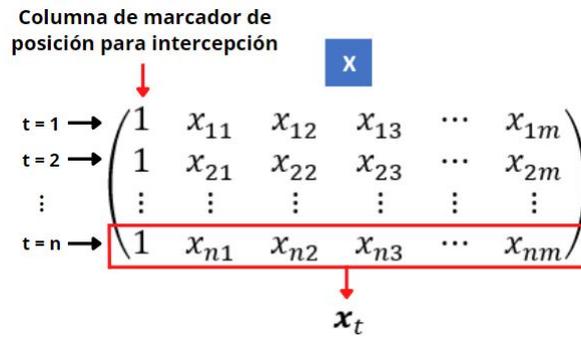


Figura 3.4. Matriz de variables de regresión

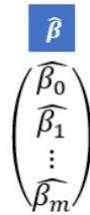


Figura 3.5. Vector de coeficientes de regresión

Obsérvese que la medida ajustada  $\hat{\lambda}_{tj}$  y el vector de coeficientes de regresión  $\hat{\beta}^{(r)}$  depende del estado  $r$ -ésimo del proceso de Markov.

Si el modelo de Markov opera sobre  $k$  estados, los coeficientes de regresión toman la forma (Figura 3.6):



Figura 3.6. Matriz de coeficientes de regresión

Además, su matriz de transición será 3.10:

$$\Gamma = \begin{bmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{bmatrix} \tag{3.10}$$

y la distribución de probabilidad de estado en el paso de tiempo  $t$  es 3.11:

$$u(t) = (P(C_t = 1), P(C_t = 2), P(C_t = 3), \dots, P(C_t = k)) \tag{3.11}$$

Para no generar  $k$  predicciones para  $y_t$  en cada paso de tiempo, se unen estas  $k$  predicciones en una sola predicción  $\hat{\lambda}_t = [\hat{\lambda}_{t1}, \hat{\lambda}_{t2}, \dots, \hat{\lambda}_{tk}]$ . Esta predicción es

una variable aleatoria que tiene  $k$  valores posibles, y cada valor está asociado a una probabilidad incondicional,  $u(t, j)$ .

Por lo tanto, la media predicha del modelo Poisson-HMM es:

$$\hat{\lambda}_t = E(\hat{\lambda}_t) = \sum_{j=1}^k \hat{\lambda}_{tj} u(t, j) \quad (3.12)$$

### 3.3.3. Entrenamiento y estimación

El entrenamiento del modelo de Poisson - oculto de Markov implica estimar la matriz de coeficientes  $\hat{\beta}^{(s)}$  y la matriz de probabilidades de transición de Markov  $\Gamma$ . Esta estimación se realiza mediante el proceso MLE, maximizando el producto:

$$L(\hat{\beta}^{(s)}; \Gamma | \mathbf{y}) = \prod_{t=1}^n P(\mathbf{y} = y_t) \quad (3.13)$$

donde:

- $L(\hat{\beta}^{(s)}; \Gamma | \mathbf{y})$  es la probabilidad de observar  $y$ .
  - $n$  es el tamaño del conjunto de datos.
  - $P(\mathbf{y} = y_t)$  es la función de masa de probabilidad del proceso de Poisson.
- Donde:

$$P(\mathbf{y} = y_t | \hat{\lambda}_{tj}; s_t = j) = \frac{e^{-\hat{\lambda}_{tj}} \hat{\lambda}_{tj}^{y_t}}{y_t!} \quad (3.14)$$

Como antes, para cada paso de tiempo, nos gustaría colapsar estas  $k$  probabilidades en una sola probabilidad. Para hacerlo, utilizamos la Ley de Probabilidad Total.

**Definición 3.7.** *La Ley de Probabilidad Total es un teorema que establece, en su caso discreto, si  $\{B_n : n = 1, 2, \dots\}$  es una partición finita o contablemente infinita de un espacio muestral y cada evento  $B_n$  es medible, entonces para cualquier evento  $A$  del mismo espacio de probabilidad:*

$$P(A) = \sum_n P(A \cap B_n) = \sum_n P(A|B_n)P(B_n) \quad (3.15)$$

Usando la Ecuación 3.15, calcularemos la probabilidad predicha por el modelo Poisson-HMM de observar  $y_t$  en el tiempo  $t$  de la siguiente manera:

$$P(\mathbf{y} = y_t | \hat{\lambda}_t) = \sum_{j=1}^k \left[ \left( \frac{e^{-\hat{\lambda}_{tj}} \hat{\lambda}_{tj}^{y_t}}{y_t!} \right) P(s_t = j) \right] \quad (3.16)$$

Por tanto, volviendo a la Ecuación 3.13, muchas veces es más rápido maximizar el logaritmo natural del producto, ya que tiene la ventaja de convertir productos en sumas y estas últimas son más fáciles de trabajar:

$$\ell(\hat{\beta}^{(s)}; \Gamma | \mathbf{y}) = \sum_{t=1}^n \ln(P(\mathbf{y} = y_t)) \quad (3.17)$$

Para resolver la ecuación 3.17 se tiene que:

1. Calcular la derivada parcial de la ecuación con respecto a cada probabilidad de transición  $p_{ij}$  en  $\Gamma$ , y cada coeficiente específico en la matriz de coeficientes  $\hat{\beta}^{(s)}$  donde el estado de Markov  $r \in [1, 2, \dots, k]$ . Igualarla a cero.
2. Resolver el sistema resultante de  $(k^2 + (m+1)k)$  ecuaciones correspondientes a  $k^2$  probabilidades de transición de Markov,  $(m+1)k$  coeficientes en  $\hat{\beta}^{(s)}$  usando algún algoritmo de optimización como Newton-Raphson.

## Predicción con datos reales: piezas de taller de TITSA

En este capítulo aplicaremos algunos de los modelos que se han visto en el capítulo 1 para estudiar la demanda de los recambios de filtro de aire en vehículos de transporte público de TITSA.

**TITSA:** Transportes Interurbanos de Tenerife S.A. es una sociedad anónima propiedad del Cabildo Insular de Tenerife, cuya actividad principal es la prestación de servicios de transporte público colectivo de pasajeros en guagua en la isla.

### 4.1. Introducción al problema

El análisis de series temporales ayuda a determinar un buen modelo que pronostique métricas comerciales como *stock*, precio de mercado, ventas, facturación, etc. Esto permite a la empresa comprender patrones oportunos en los datos y analizar las tendencias.

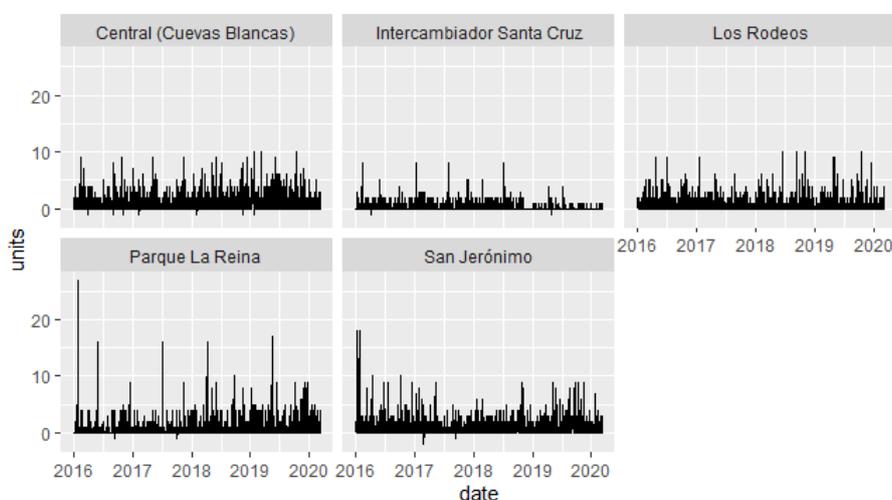
Desde la empresa TITSA, se plantea el problema de predecir el número de filtros de aire que fallarán en las guaguas para tener recambios suficientes en los talleres de la Central (Cuevas Blancas), Intercambiador de Santa Cruz, Los Rodeos, Parque de la Reina y San Jerónimo. Para ello utilizaremos el entorno de desarrollo, dedicado a la computación estadística y gráficos, RStudio, y librerías como *sknifedatar*, *tidymodels*, *timetk*, ...

### 4.2. Preparación de los datos

Desde el departamento de Data Science y Big Data de TITSA, se nos proporciona un Dataset en el cuál se muestran distintas operaciones que se han realizado en los talleres de la empresa, como pueden ser reparaciones, pedidos, consumo de combustible, etc. Este Dataset consta de 1.529.704 entradas, entre los días 1 de enero de 2016 y 10 de marzo de 2020.

Para este estudio, se trabajará solamente con los datos relacionados con el artículo "*FILTRO DE AIRE*". Por lo tanto, filtraremos el Dataset original y crearemos uno nuevo solamente con los datos relacionados con estas piezas de taller. Este nuevo Dataset consta de 7.655 entradas.

En la Figura 4.1 podemos observar el consumo de filtros de aire que se ha realizado, diariamente, en cada uno de los talleres entre los días 1 de enero de 2016 y 10 de marzo de 2020.



**Figura 4.1.** Filtros de aire pedidos diariamente en cada taller entre el 01/01/2016 y 10/03/2020

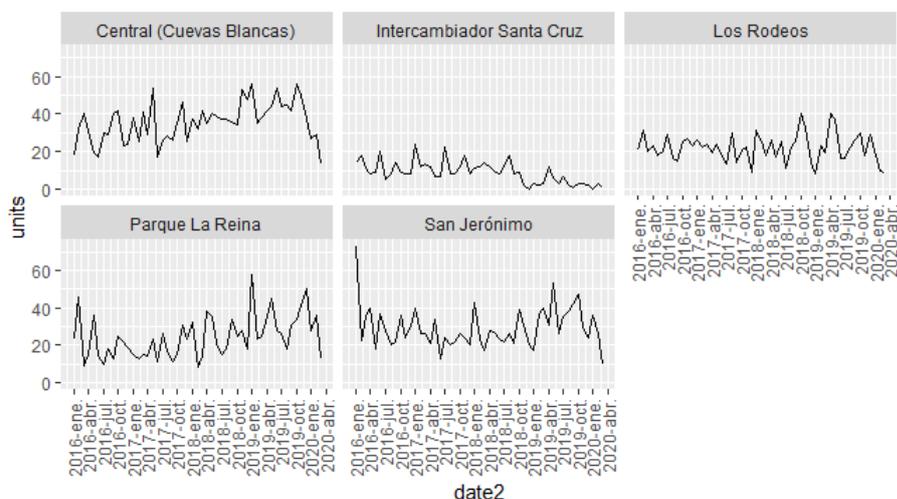
Como algunos de los modelos que se ha introducido en los capítulos anteriores incluyen una componente cíclico o estacional, se ha decidido juntar los datos de forma mensual. (Figura 4.2)

Después de analizar los datos finales obtenidos, se decidió hacer el estudio solamente para uno de los talleres de la empresa, la Central (Cuevas Blancas). De esta forma, podemos aplicar los modelos en una base de datos no tan extensa.

En resumen, se ha pasado de tener un Dataset de 1.529.704 entradas, que contenía todas las operaciones diarias que se realizaron en los cinco talleres de TITSA, a obtener un Dataset con 51 entradas solamente con los datos mensuales de "*FILTRO DE AIRE*" del taller Central (Cuevas Blancas).

### 4.3. Entrenamiento y validación de modelos

Antes de comenzar a utilizar los modelos de predicción, introduciremos algunos términos que serán útiles para analizar el conjunto de datos.



**Figura 4.2.** Filtros de aire pedidos mensualmente en cada taller entre el 01/01/2016 y 10/03/2020

**Definición 4.1.** *El conjunto de entrenamiento (training set) es el conjunto de datos que se utiliza para entrenar y hacer que el modelo aprenda las características y patrones ocultos en los datos. El conjunto de entrenamiento debe tener un conjunto diversificado de entradas para que el modelo se entrene en todos los escenarios y pueda predecir cualquier muestra de datos invisible que pueda aparecer en el futuro.*

**Definición 4.2.** *El conjunto de validación (validation set) es un conjunto de datos, separado del conjunto de entrenamiento, que se utiliza para validar el rendimiento de nuestro modelo durante el entrenamiento.*

Para realizar nuestra predicción se tomará como conjunto de entrenamiento los datos mensuales desde el 01/01/2016 hasta el 01/09/2019. Y el conjunto de validación (Figura 4.3) serán los últimos 6 meses, desde el 01/10/2019 hasta el 01/03/2020.

|    | store                    | period     | indvar |
|----|--------------------------|------------|--------|
| 46 | Central (Cuevas Blancas) | 2019-10-01 | 56     |
| 47 | Central (Cuevas Blancas) | 2019-11-01 | 50     |
| 48 | Central (Cuevas Blancas) | 2019-12-01 | 37     |
| 49 | Central (Cuevas Blancas) | 2020-01-01 | 27     |
| 50 | Central (Cuevas Blancas) | 2020-02-01 | 29     |
| 51 | Central (Cuevas Blancas) | 2020-03-01 | 13     |

**Figura 4.3.** Validation set

Por tanto, gráficamente los conjuntos de entrenamiento y validación quedarán de la siguiente manera (Figura 4.4):

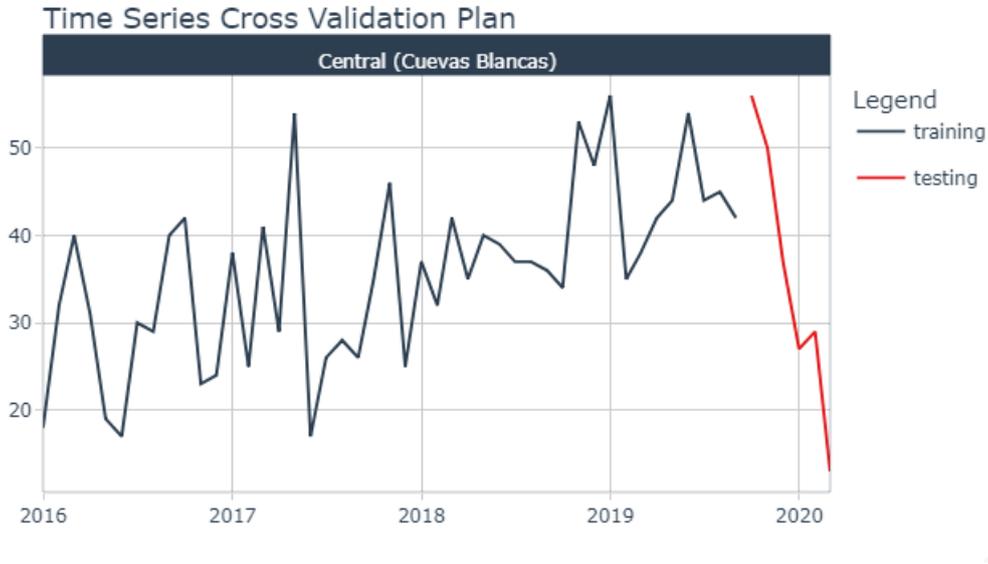


Figura 4.4. Training and validation set

A continuación, se procederá a ajustar algunos de los modelos que se han introducido en el capítulo 1, y así observar qué modelo ajusta mejor la serie temporal creada con el conjunto de datos de los *filtros de aire*.

### MODELO ARIMA

Una vez entrenado el modelo ARIMA, este pronostica los valores que se muestran en rojo y calibra el ajuste del modelo sobre el conjunto de validación (Figura 4.5).

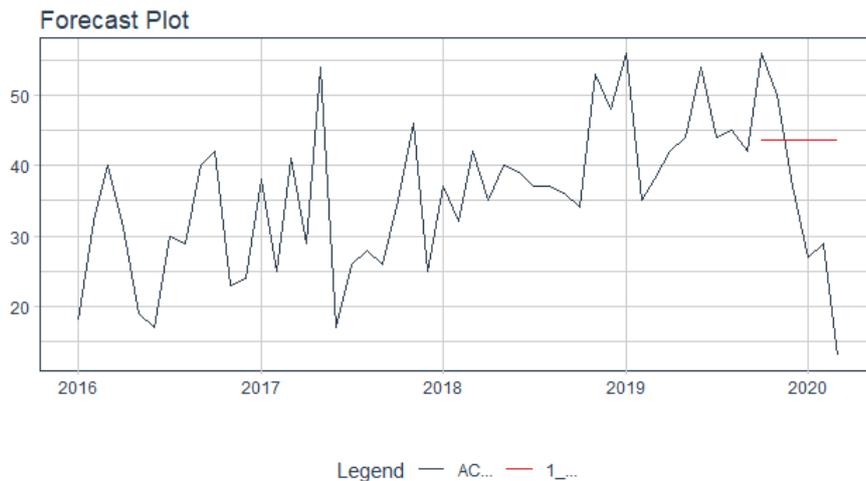
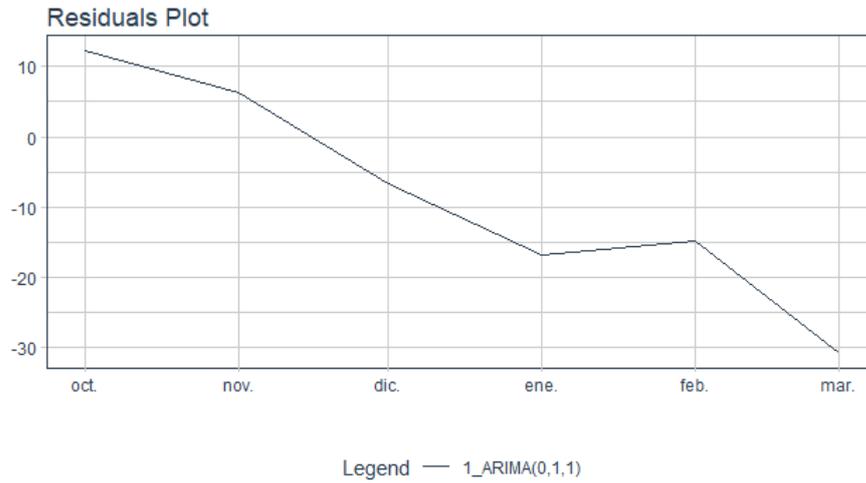


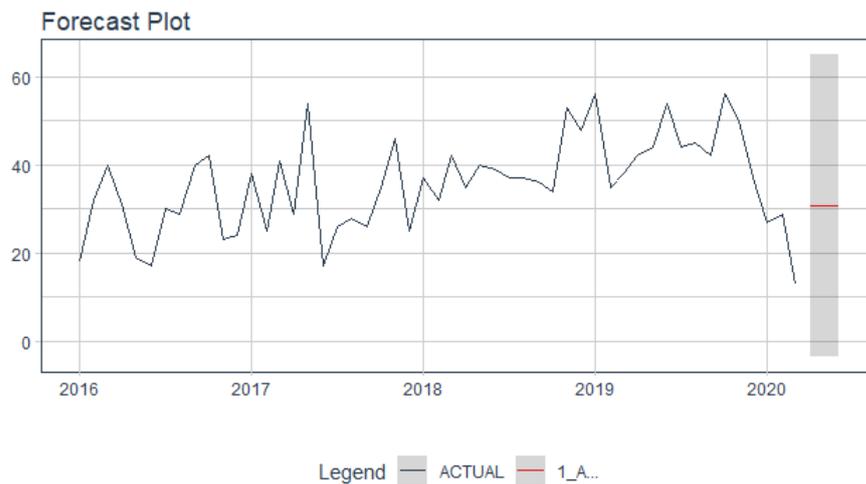
Figura 4.5. Valores observados y pronosticados con el modelo ARIMA

Se obtiene la gráfica de valores residuales, observados frente a predichos, que se muestra en la Figura 4.6.



**Figura 4.6.** Valores residuales del pronóstico observado con el modelo ARIMA

Después de haber entrenado y validado el modelo, este crea una predicción para los meses de abril, mayo y junio de 2020 representado en la Figura 4.7.



**Figura 4.7.** Predicción para los meses posteriores a marzo de 2020 con el modelo ARIMA

### SUAVIZADO EXPONENCIAL

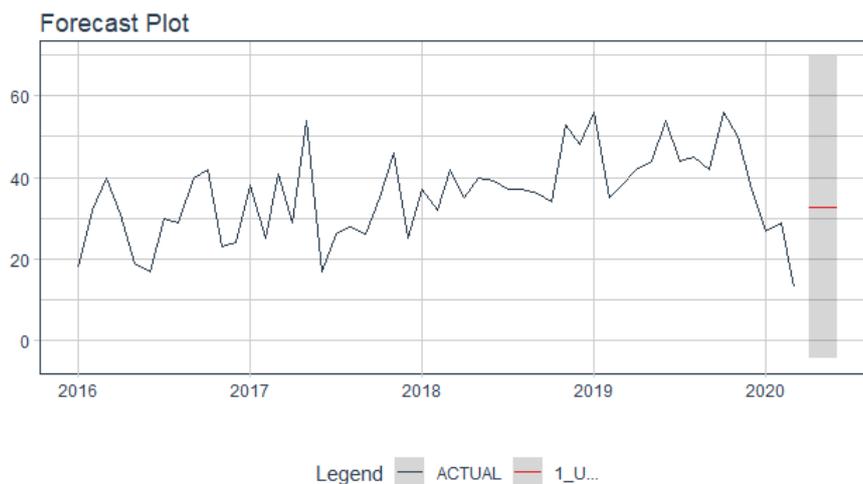
El segundo modelo que se ha utilizado ha sido el modelo del suavizado exponencial, este crea una predicción para los tres meses posteriores a marzo de 2020 que se muestran en la Figura 4.8.



**Figura 4.8.** Predicción para los meses posteriores a marzo de 2020 con el suavizado exponencial

### MODELO TBATS

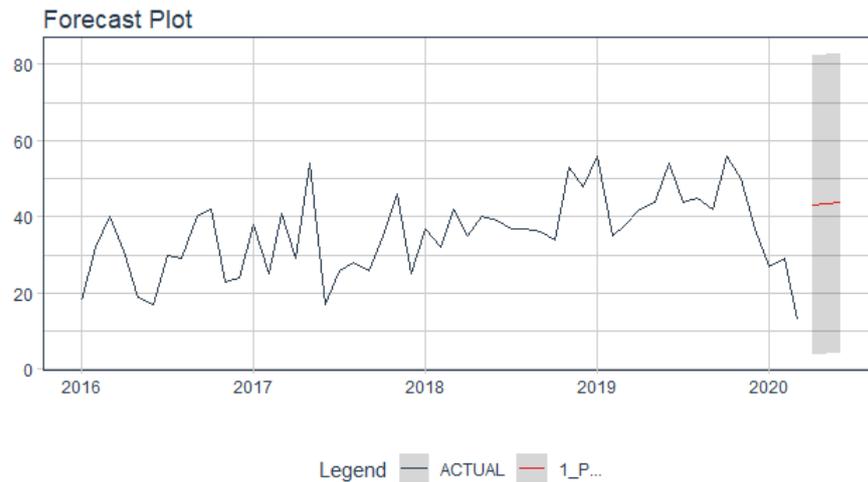
El tercer modelo aplicado a los datos iniciales ha sido el modelo TBATS, este ha realizado la predicción que se muestra en la Figura 4.9.



**Figura 4.9.** Predicción para los meses posteriores a marzo de 2020 con el modelo TBATS

### MODELO PROPHET

Y, por último, aplicamos el modelo Prophet, la predicción que realiza es la mostrada en la Figura 4.10.



**Figura 4.10.** Predicción para los meses posteriores a marzo de 2020 con el modelo Prophet

## 4.4. Comparación de las predicciones

Para finalizar nuestro estudio, debemos comprobar la bondad de los modelos de predicción utilizando diferentes medidas de error.

### 4.4.1. MSE y RMSE

El **error cuadrático medio (MSE)** y el **error de la raíz cuadrada media (RMSE)** son medidas de error que otorgan mayor peso a los errores grandes que a los pequeños. Estas medidas están basadas en el siguiente concepto de error  $e_i$  (donde  $y_i$  representa el  $i$ -ésimo valor real de una serie de tiempo y  $\hat{y}_i$  es el valor que se pronosticó, por la misma posición en la serie, por el modelo):

$$e_i = y_i - \hat{y}_i \quad (4.1)$$

Dado que los errores positivos y negativos tienden a cancelarse entre sí, tomamos los cuadrados de estas diferencias y sacamos el promedio de todos estos cuadrados. Como resultado, obtenemos el error cuadrático medio (calculado sobre  $N$  diferencias entre los valores reales y predichos):

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N} \quad (4.2)$$

Una de las limitaciones del MSE es que la unidad de medida del error es el cuadrado de la unidad de medida de los datos, por ello se toma la raíz del MSE y luego obtenemos el error de la raíz cuadrada media:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \tag{4.3}$$

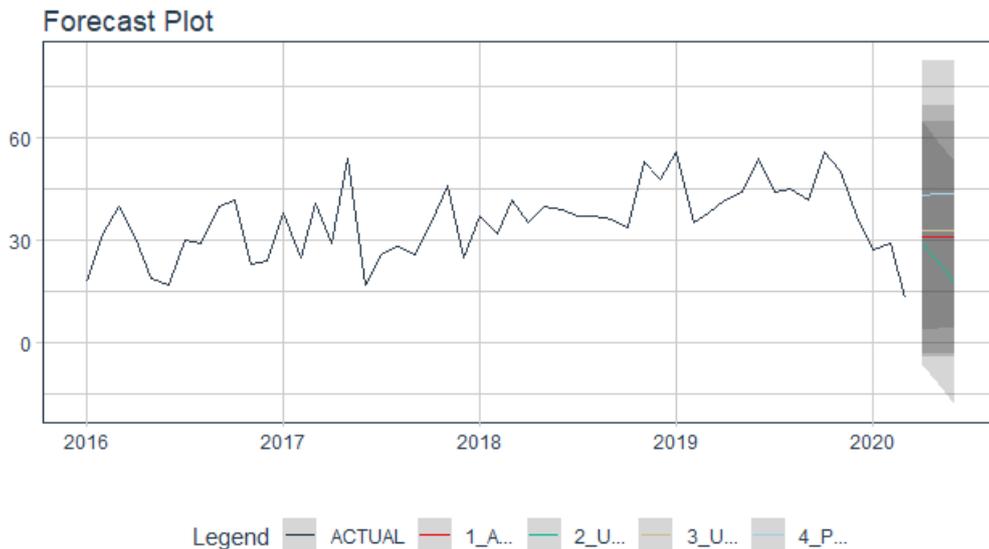
#### 4.4.2. MAE

Si, por el contrario, se le da el mismo peso a los errores grandes y pequeños, podemos usar el promedio de los valores absolutos de los errores y obtener el **error absoluto medio (MAE)**:

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \tag{4.4}$$

.....

Con la ayuda de la herramienta R realizamos una comparación de las predicciones que crean los cuatro modelos que hemos aplicado para de esta forma decidir que modelo ajusta mejor la serie temporal y utilizarlo para posteriores predicciones.



**Figura 4.11.** Predicción para los meses posteriores a marzo de 2020 de los cuatro modelos (ARIMA, Suavizado exponencial, TBATS y Prophet)

Analizando los resultado obtenidos en la Figura 4.12, podemos observar que el modelo ARIMA(0,1,1) tiene el MAE y RMSE más pequeño, de 14.6 y 16.7

```

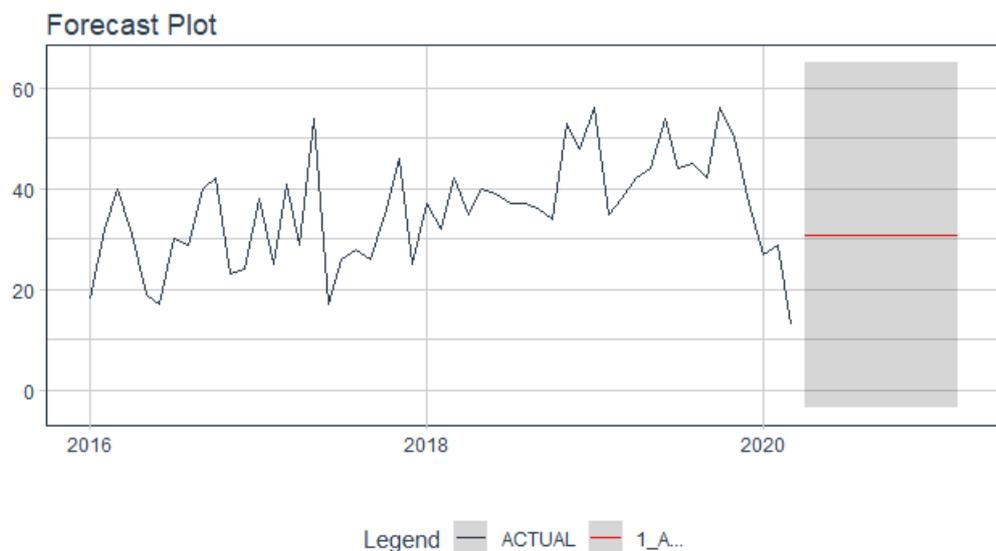
> calibration_table %>% modeltime_accuracy()
# A tibble: 4 × 9
  .model_id .model_desc .type mae mape mase smape rmse rsq
  <int> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1 ARIMA(0,1,1) Test 14.6 67.0 1.55 41.8 16.7 NA
2 2 ETS(M,AD,N) Test 15.0 70.0 1.59 42.4 17.4 0.932
3 3 BATS(0.873, {0,0}, 1, -) Test 15.6 72.5 1.66 43.7 18.1 0.945
4 4 PROPHET Test 16.3 77.1 1.73 44.9 19.2 0.943

```

**Figura 4.12.** Errores de los cuatro modelos (ARIMA, Suavizado exponencial, TBATS y Prophet)

respectivamente. Este sería el mejor modelo que ajusta nuestra serie temporal. También el modelo de suavizado exponencial (ETS) tiene valores pequeños de estos indicadores.

Para finalizar, si utilizamos el modelo ARIMA para predecir la cantidad de *filtros de aire* que necesitará la empresa TITSA para el siguiente año será lo mostrado en la Figura 4.13.



**Figura 4.13.** Predicción para el año 2020 de los filtros de aire en el taller Central (Cuevas Blancas)



---

## Conclusiones

En esta memoria se ha hecho un recorrido por los diferentes modelos de predicción de series temporales.

En el Capítulo 1 se describieron brevemente varios modelos de predicción de series temporales, todos ellos acompañados por un ejemplo práctico en el que se predice el número de nuevo pedidos de equipos eléctrico en la Zona Euro.

Estos modelos tienen sus ventajas y desventajas. Algunos de ellos requieren de una cantidad mínima de esfuerzo, pero sus predicciones son menos fiables. Y otros, en cambio, realizan una predicción más exacta cuando los datos tienen patrones estacionales. Por lo tanto, no existe un modelo perfecto para todas las series temporales, se debe elegir el modelo que aprenda las características y patrones ocultos en los datos, para de esta forma obtener una predicción más elaborada.

Un estudio complementario a esta memoria podría ser desarrollar de manera más extensa los modelos presentados en la Sección 1.2. Entre otros, podemos extraer más información del modelo ARIMA (Subsección 1.2.4) de [8].

En los Capítulos 2 y 3 se estudiaron modelos estadísticos más complejos. El modelo de regresión de Poisson, utilizado para modelar series temporales basadas en el recuento de número enteros, y donde el parámetro  $\lambda$ , de la distribución de Poisson, está influenciado por las variables de regresión. Y el modelo oculto de Markov, en el cual se asume que el sistema a modelar es una cadena de Markov de parámetros desconocidos, y cuyo objetivo es determinar los parámetros desconocidos de dicha cadena a partir de los parámetros observables. Para un estudio más detallado de este modelo es recomendable ver [10].

Además, se formuló la combinación entre el modelo de Poisson visible y el Modelo Oculto de Markov, creando así el modelo Poisson-HMM. Sin embargo, el entrenamiento y estimación de este modelo es mucho más complicado. Se dejó redactado el proceso pero de manera general, por lo que se deja como trabajo futuro desarrollar este proceso.

Por último, en el Capítulo 4, se realizó un estudio de la serie temporal cedida por TITSA, en la cuál se aplicaron varios de los modelos presentados en la memoria para la predicción de los filtros de aire necesarios en el taller Central (Cuevas Blancas). Se ha utilizado el lenguaje de programación estadístico R para manipular los datos y realizar el análisis del conjunto de datos. Como resultado se obtuvo que el mejor modelo de predicción para esta serie temporal era el modelo ARIMA, aunque si observamos la Figura 4.5 observamos que los datos predichos no se ajustan al conjunto de validación.

---

## Bibliografía

- [1] Burba, Davide. *An overview of time series forecasting models* [en línea]. 2019. [Fecha de consulta: 12-01-2022]. Disponible en: <https://towardsdatascience.com/an-overview-of-time-series-forecasting-models-a2fa7a358fcb>.
- [2] Mañas Mañas, Andrés. *Notas sobre pronóstico del flujo de tráfico en la ciudad de Madrid* [en línea]. 2019. [Fecha de consulta: 04-02-2022]. Disponible en: <https://bookdown.org/amanas/traficomadrid/m%C3%A9todos-param%C3%A9tricos.html>.
- [3] Bollerslev, Tim. *Generalized autoregressive conditional heteroskedasticity* [en línea]. 1986. University of California at San Diego, La Jolla, CA 92093, Estados Unidos. Institute of Economics, University of Aarhus, Dinamarca. [Fecha de consulta: 10-02-2022]. Disponible en: [http://public.econ.duke.edu/~boller/Published\\_Papers/joe\\_86.pdf](http://public.econ.duke.edu/~boller/Published_Papers/joe_86.pdf).
- [4] Petris G., Petrone S. y Campagnoli P. *Dynamic Linear Models with R* [en línea]. 2009. [Fecha de consulta: 15-02-2022]. Disponible en: [http://civil.colorado.edu/~balajir/CVEN6833/lectures/Dynamic\\_Linear\\_Models\\_with\\_R.pdf](http://civil.colorado.edu/~balajir/CVEN6833/lectures/Dynamic_Linear_Models_with_R.pdf).
- [5] Taylor S.J. y Letham B. *Forecasting at scale* [en línea]. 2017. [Fecha de consulta: 27-02-2022]. Disponible en: <https://peerj.com/preprints/3190/>.
- [6] Hyndman R.J. y Athanasopoulos G. *Forecasting: Principles and Practice* [en línea]. 2018. Monash University, Australia. [Fecha de consulta: 12-03-2022]. Disponible en: <https://otexts.com/fpp2/nnetar.html>.
- [7] Tayyip Saka, A. *Guide to Long Short Term Memory (LSTM) models in Stock Prediction* [en línea]. 2021. [Fecha de consulta: 20-03-2022]. Disponible en: <https://abdullahsaka.medium.com/long-short-term-memory-lstm-model-in-stock-prediction-d0e1c93717b3>.
- [8] Brockwell P.J. y Davis R.A. *Introduction to time series and forecasting*. Reading: Springer Texts in Statistics, 1996.
- [9] Cameron A.C. y Trivedi P.K. *Regression Analysis of Count Data*. 2nd edition. Reading: Cambridge, 2013.

- [10] Zucchini W. y MacDonald I.L. *Hidden Markov Models for Time Series. An Introduction Using R*. Reading: Chapman & Hall/CRC, 2009.
- [11] R Core Team. *R: A language and environment for statistical computing*. 2019. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- [12] Gráficos y figuras de elaboración propia.

# Analysis and modeling of time series. Study of the demand for spare parts in public transport vehicles.

Ana García Reyes

Facultad de Ciencias • Sección de Matemáticas  
Universidad de La Laguna

alu0101138282@ull.edu.es

## Abstract

Knowing about the evolution of data from short-term observations has drawn the attention of many scientists. In this poster, different time series models will be shown. Particularly, the Poisson Regression Model and the Hidden Markov Model. Furthermore, the results of a study carried out with real data provided by TITSA will be shown.

## 1. Introduction and time series modeling

TIME SERIES MODELS are used to forecast events based on verified historical data. Some of the most used models in time series are:

- Naïve, SNaïve
- Seasonal decomposition
- Exponential smoothing
- ARIMA, SARIMA
- Dynamic linear models
- TBATS
- Prophet
- NNTAR
- LSTM

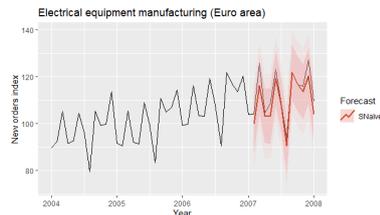


Figure 1: Prediction of the manufacture of electrical equipment with the SNaïve model.

## 2. Poisson Regression Model and Hidden Markov Model

If we go into more complex statistical models we find the Poisson Regression Model and the Hidden Markov Model.

### 1. POISSON REGRESSION MODEL

This model is useful for modeling time series data consisting of integer counts.

**Definition 1.** The **Poisson distribution** gives the probability of a given number of events happening in a fixed interval of time.

$$P_x(k) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}$$

**Definition 2.** The job of the **Poisson regression model** is to fit the observed counts  $y_t$  to the regression matrix  $X$  through a function that expresses the rate vector  $\lambda$  in terms of the regression coefficients  $\beta$  and the regression matrix  $X$ .

$$y_t = \lambda = e^{X\beta}$$

### 2. HIDDEN MARKOV MODEL

This model is suitable for modeling dynamic systems, and it is used for pattern recognition and classification problems.

**Definition 3.** A sequence of discrete random variables  $\{C_t : t \in \mathbb{N}\}$  is said to be a **Markov chain** if  $\forall t \in \mathbb{N}$  it satisfies the Markov property:

$$P(C_{t+1}|C_t, \dots, C_1) = P(C_{t+1}|C_t) = P(C_{t+1}|\mathbf{C}^{(t)})$$

TRABAJO FIN DE GRADO, Convocatoria de septiembre, 2022

**Definition 4.** A **hidden Markov model**  $\{Y_t : t \in \mathbb{N}\}$  is a particular kind of dependent mixture. With  $\mathbf{Y}^{(t)}$  and  $\mathbf{C}^{(t)}$  representing the time series past from time 1 to time  $t$ . The model can be summarize by:

$$P(C_t|\mathbf{C}^{(t-1)}) = P(C_t|C_{t-1}), \quad t = 2, 3, \dots$$

$$P(Y_t|\mathbf{Y}^{(t-1)}, \mathbf{C}^{(t)}) = P(Y_t|C_t), \quad t \in \mathbb{N}$$

## 3. Forecasting with real data

THE problem of predicting when the air filters will fail in the TITSA's buses is analyzed. This problem must be solved in order to have enough stock in the workshops.



Figure 2: Training and validation set.

The predictions created by the ARIMA, Exponential Smoothing, TBATS and Prophet models are:



Figure 3: Forecasting of the four models (ARIMA, Exponential Smoothing, TBATS and Prophet)

If we compare the goodness of the models, we can say that the best model that fits the series is the ARIMA model.

## References

- [1] An overview of time series forecasting models [en línea]. [Consulta: 12-01-2022] Disponible en: <https://towardsdatascience.com/an-overview-of-time-series-forecasting-models-a2fa7a358fcb>.
- [2] Cameron A.C. y Trivedi P.K. *Regression Analysis of Count Data*. 2nd edition. Reading: Cambridge, 2013.
- [3] Zucchini W. y MacDonald I.L. *Hidden Markov Models for Time Series. An Introduction Using R*. Reading: Chapman & Hall/CRC, 2009.