

Trabajo práctico sobre Series Temporales

Martín Sebastian Armoa

Docente: Horacio Castellini

Universidad Nacional de Rosario

armoa58@gmail.com

1. Introducción

1.1. Series Temporales

Por definición una serie temporal es una recolección de datos, una sucesión de observaciones realizadas a lo largo del tiempo a intervalos regulares, es decir es una sucesión finita o no de variables aleatorias de manera que el tiempo es discreto o numerable.

La característica fundamental de las series temporales es que las observaciones sucesivas no son independientes entre sí, y el análisis debe llevarse a cabo teniendo en cuenta el orden temporal de las observaciones. Los métodos estadísticos basados en la independencia de las observaciones no son válidos para el análisis de series temporales porque las observaciones en un instante de tiempo dependen de los valores de la serie en el pasado. A esto se lo llama procesos con memoria.

En este trabajo se realiza un estudio descriptivo de un DataFrame al cual se le aplican distintas técnicas para sacar conclusiones e intentar reconstruir la dinámica del sistema tales como: -Embedding

-Función de autocorrelación -Transformada de Fourier y densidad espectral -Componentes principales -Método del falso vecino o False Nearest Neighbor

1.2. DataFrame

Primero se realizó una visualización del data frame (1) en RStudio para tener un mejor contexto en el cual empezar a explorar los datos y luego se realizó un resumen estadístico 12.

Figura 1: Visualización de los datos

	Tiempo	Velocidad	R	T
0	0.003022	30.207969	26914.57	9107.59
1	0.004030	29.888208	26914.57	9075.75
2	0.005037	29.569110	26830.73	9043.91
3	0.006045	29.250706	26914.57	9012.07
4	0.007052	28.933022	27249.93	9012.07

El dataframe consta de 5,459 entradas (filas) y 8 columnas de las cuales solo nos interesan las relacionadas al tiempo, la función de referencia (Velocidad) y las variables R y T que son las que refieren al sistema dinámico en cuestión según la información previamente brindada con el conjunto de datos.

Figura 2: Resumen estadístico del dataframe

```
> summary(data1)
```

Tiempo		Velocidad		R		T	
Min.	:0.000	Min.	:2.239e+09	Min.	: 2.00	Min.	:3.516e+06
1st Qu.	:1.000	1st Qu.	:1.954e+13	1st Qu.	:12.00	1st Qu.	:2.073e+12
Median	:2.000	Median	:4.888e+13	Median	:36.00	Median	:5.703e+12
Mean	:2.363	Mean	:1.178e+14	Mean	:36.28	Mean	:1.356e+13
3rd Qu.	:4.000	3rd Qu.	:8.498e+13	3rd Qu.	:61.00	3rd Qu.	:8.878e+12
Max.	:9.000	Max.	:9.994e+14	Max.	:72.00	Max.	:9.999e+13

A continuación se muestran las gráficas de las variables de interés, Tiempo, Velocidad, R y T

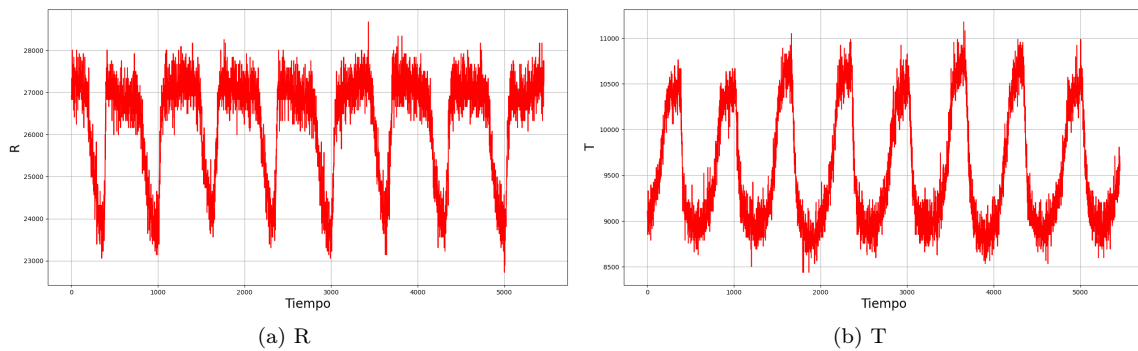


Figura 3: Gráfica de variables R y T

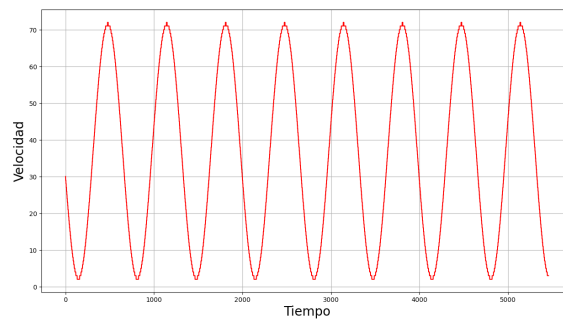


Figura 4: Tiempo vs Velocidad

Estamos interesados en buscar algunas características especiales en nuestros datos para poder hacer predicciones. En concreto, se busca una serie cronológica que sea estacionaria. La estacionariedad es importante porque la mayoría de los métodos estadísticos para realizar análisis y predicciones se basan en la suposición de que las propiedades estadísticas de la serie (media, varianza, correlación, etc.) son constantes en el tiempo.

Se procede a descomponer la serie temporal en sus componentes o fuentes de variación que son consideradas habitualmente:

1. Tendencia
2. Efecto Estacional
3. Componente Aleatoria

Se puede observar que en ambos casos la tendencia es oscilatoria o constante. La componente estacional y aleatoria presentan demasiado ruido como para sacar información.

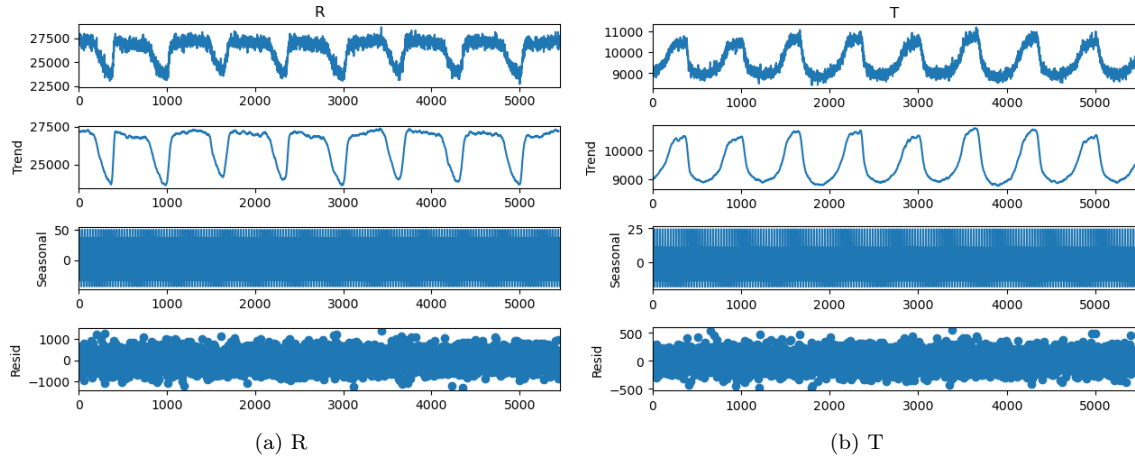


Figura 5: Descomposicion de la serie temporal

1.3. Autocorrelación

La función de autocorrelación Simple (ACF) mide la correlación entre dos variables separadas por k periodos. Mide el grado de asociación lineal que existe entre dos variables del mismo proceso estocástico.

La función de autocorrelación parcial (PACF) mide la correlación entre dos variables separadas por k periodos cuando no se considera la dependencia creada por los retardos intermedios existentes entre ambas. Mide la autocorrelación que existe entre dos variables separadas k períodos descontando los posibles efectos debidos a variables intermedias.

Un valor aproximado de $+1$ implica una fuerte relación entre dos observaciones separadas unidades de tiempo. El signo positivo (+), denota una relación directa entre las dos observaciones, es decir, que cuando el valor de alguna de ellas aumenta, el otro también tenderá a aumentar; o viceversa, si disminuye un valor, el otro tenderá a disminuir también. Un coeficiente de autocorrelación cercano a -1 expresa que la relación entre las observaciones es inversa, el aumento del valor de la variable se asocia a la disminución de la otra. Un coeficiente de 0 indica que los dos valores de una misma serie de tiempo no están relacionados. Las autocorrelaciones proporcionan información importante acerca de la estructura de un conjunto de datos y de sus patrones. Por ejemplo, en un conjunto de datos completamente aleatorios la autocorrelación entre valores sucesivos estará cercana a 0, o será igual a 0, pero los valores de datos de fuerte naturaleza estacional o cíclica estarán sumamente autocorrelacionados.

Figura 6: Autocorrelacion Simple y Parcial de la variable R

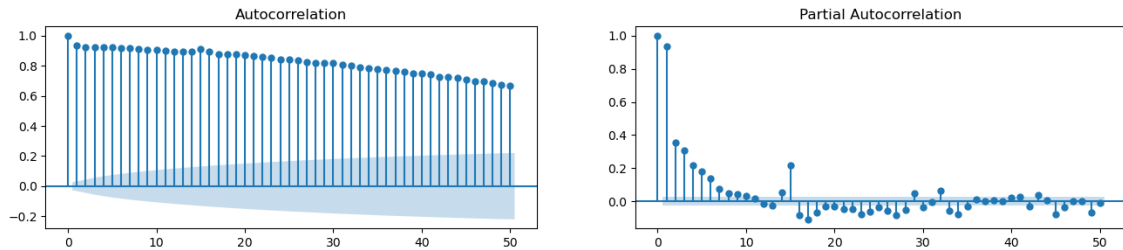
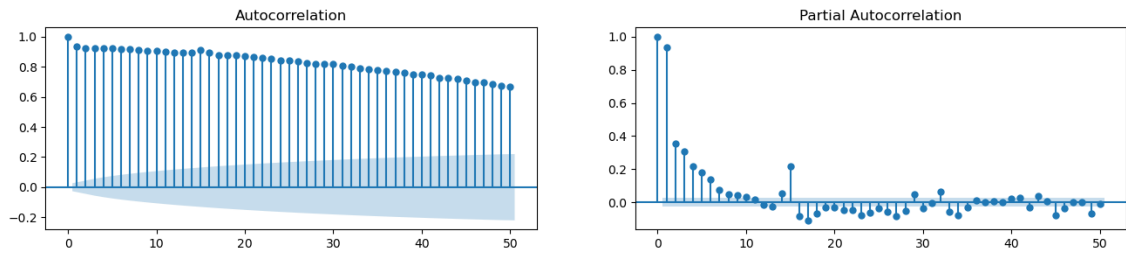


Figura 7: Autocorrelación Simple y Parcial de la variable T



Se puede observar que las series tienen una alta correlación, lo que implica que son de una fuerte naturaleza estacional.

1.4. Suavización de la serie

Se realizó un filtro de la Serie mediante el cálculo de la Media móvil y se obtuvieron las siguientes gráficas.

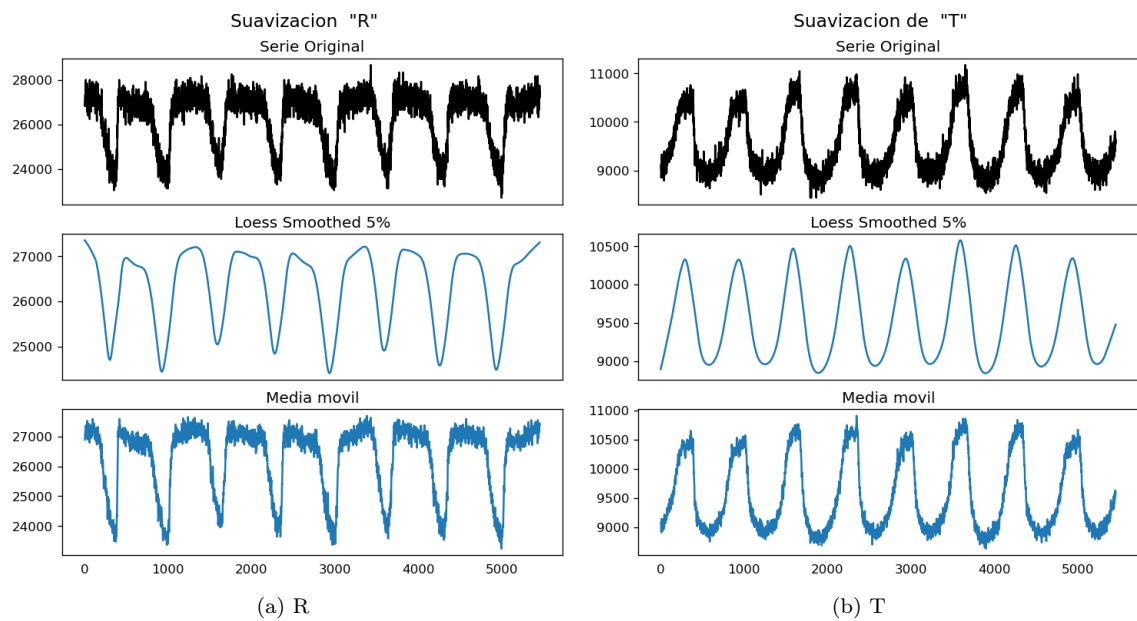
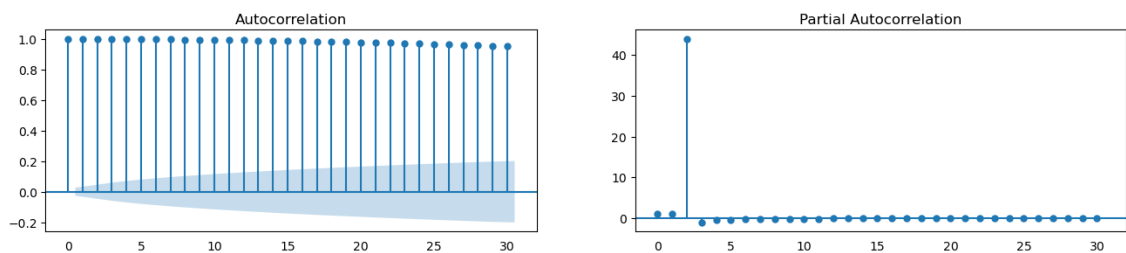


Figura 8

Luego se calcula la correlación nuevamente para "T" se puede ver como esta mejora obteniendo un resultado mucho mas claro.

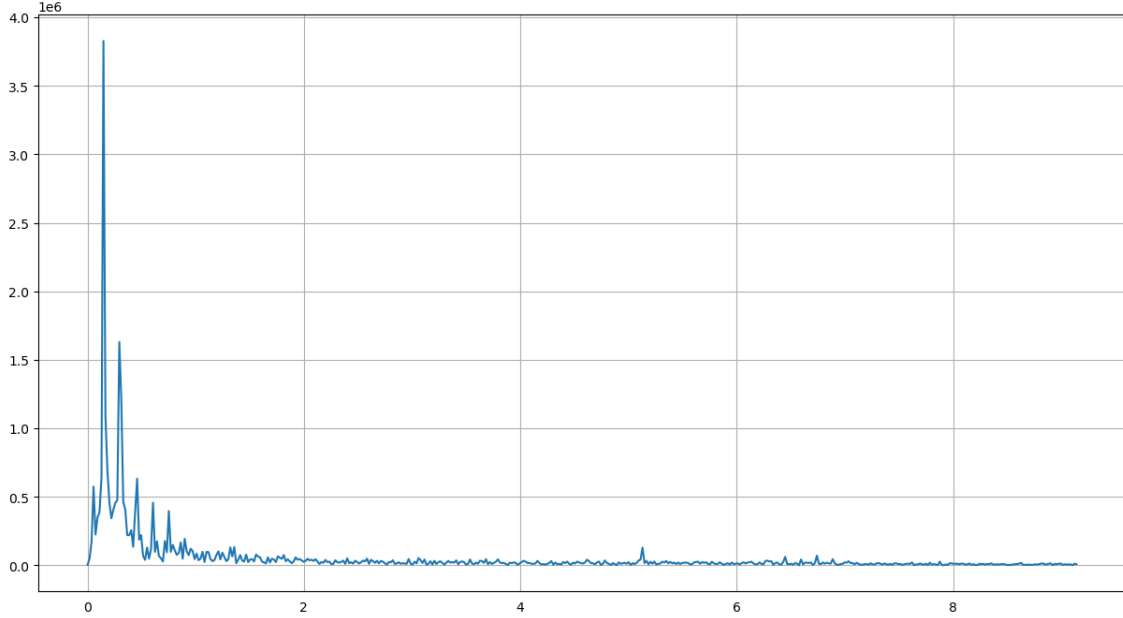
Figura 9: Autocorrelacion de "T" suavizado por media móvil



1.5. Transformada discreta de Fourier

Se realiza una descomposición en componentes principales para maximizar la componente principal obteniendo una nueva serie temporal con menor ruido para luego realizar la transformada de Fourier de la serie para intentar determinar de que sistema se trata obteniendo la siguiente grafica:

Figura 10: Transformada discreta de Fourier



Esto demuestra la existencia de un sistema dinámico no lineal, ya que si el sistema fuera periódico tendríamos un solo valor de frecuencia.

1.6. Reconstrucción del Espacio de fases

Este método se fundamenta en el teorema de inmersión de Takens, que establece que, bajo ciertas condiciones, aunque no sea posible reconstruir la órbita del sistema dinámico en el espacio de fase original, sí será posible obtener una aproximación que resulte topológicamente equivalente, y que por tanto, permitirá extraer toda la información relevante acerca del sistema dinámico subyacente y desconocido que genera la serie temporal. El teorema de inmersión de Takens garantiza que las características dinámicas y topológicas de un atractor se mantienen en ese espacio reconstruido a partir de la serie temporal por el método de los retardos. Esto nos permitirá diferenciar entre un movimiento puramente estocástico frente a la alternativa de caos, ya que las trayectorias caóticas tienen una estructura en el espacio de fases que se puede estudiar a partir del pseudo espacio de fases reconstruido.

A partir de un muestreo dado por:

$$s_0 = x_1(t), s_1 = x_1(t + \tau), s_2 = x_1(t + 2\tau), s_3 = x_1(t + 3\tau), \dots, s_k = x_1(k\tau), \dots$$

tomados a intervalos de longitud τ se construyen los llamados vectores con retardo de dimensión $2d+1$, definidos como: $Y_j = (s_j, s_{j-1}, \dots, s_{j-2d})$ donde habitualmente d corresponde a la dimensión del atractor del sistema.

La elección de τ determina la exactitud del atractor reconstruido. Valores demasiado pequeños producirán el atractor a lo largo de una línea, y valores demasiado grandes tampoco revelarán su estructura de forma correcta.

Para calcular el tiempo de retraso y la dimension adecuada se utilizo una libreria de python llamada Giotto-TDA [1], es una tolbox de machine learning topológico de alto rendimiento en Python construida sobre scikit-learn y se distribuye bajo la licencia GNU AGPLv3 obteniendo el diagrama de fases para la variable R siguiente:

Figura 11: Reconstruccion de diagrama de fases para R''

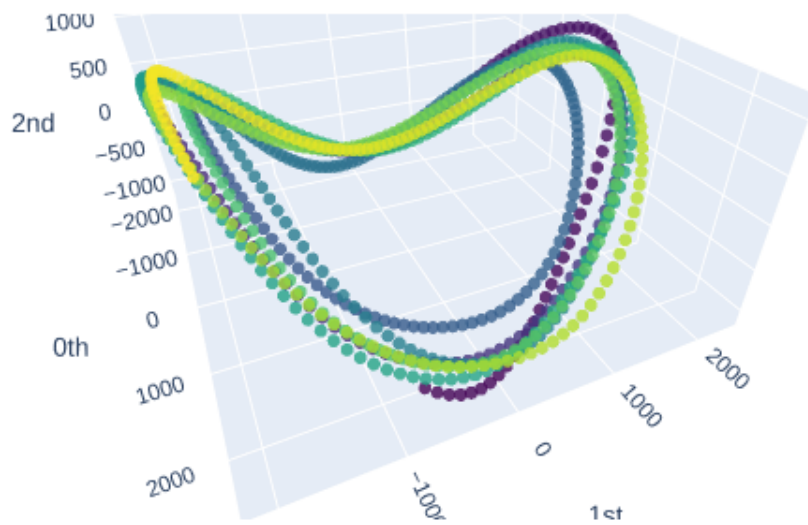
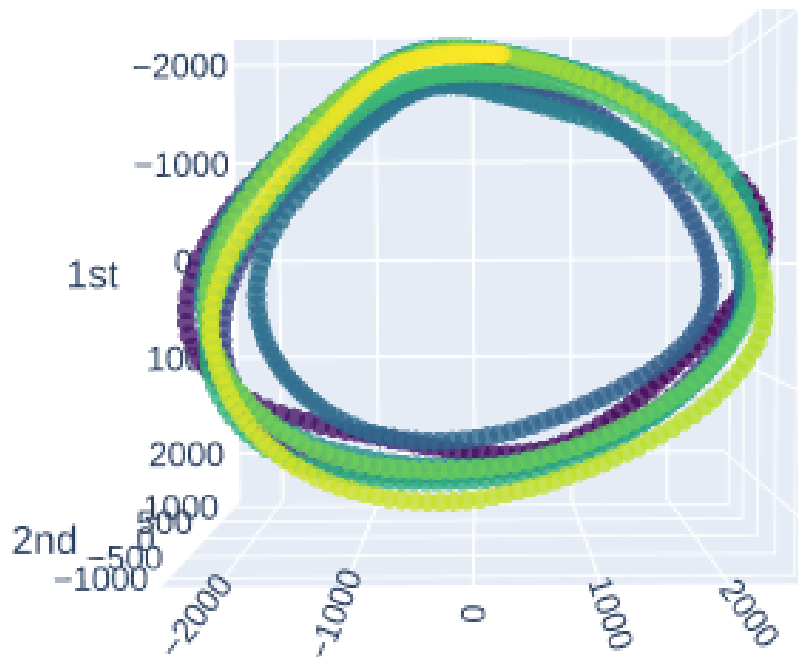


Figura 12: Vista superior del diagrama de fases para R''



Referencias

- [1] Giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration, Tauzin et al, arXiv:2004.02551, 2020.
- [2] Matthew Luckie. CScamper: a scalable, extensible packet prober for active measurement of the internet, 2010.