

# Inteligencia Artificial en Ciencia de Datos.

## Proyecto 2:

Erick Muñiz Morales  
github.com/ErickMM98  
N.C. 315230372

Julio, 2021

## Contents

<b>1</b>	<b>Introducción</b>	<b>1</b>
<b>2</b>	<b>Análisis descriptivo: Descripción de la base.</b>	<b>2</b>
<b>3</b>	<b>Presentación del modelo: Modelo (SIR)</b>	<b>3</b>
3.0.1	Algunas consideraciones del modelo y consecuencias del mismo. . . . .	4
<b>4</b>	<b>Ajuste.</b>	<b>5</b>
<b>5</b>	<b>Conclusiones.</b>	<b>6</b>

## 1 Introducción

Comprender y modelar la difusión de un ente, tangible o intangible, en un sistema es un problema de índole complejo. Los sistemas dinámicos no lineales son de las primeras aproximaciones a esta problemática. Si bien dichos modelos introducen una variable temporal con varios estados relacionados, tiene una sutil problemática al omitir la importancia de la estructura topológica de los individuos del sistema así como su importancia e interacción. En resumen, se tiene un modelo simple que supone homogeneidad, que generalmente no es cierto, en beneficio de resultados. Dado que estas ecuaciones abstraen la relación de intercambio de individuos en los estados de interés, se puede extrapolar dicha moción para el caso de intercambio de ideas y opiniones.

Este proyecto tiene como objetivo utilizar la base de datos de la dinámica de la dispersión del *SARS-COVID-2* en China. La base de datos es la presentada en las indicaciones del proyecto. La base de datos cuenta con registros desde el 20 de enero del 2020 y hasta el 1 de marzo de 2020.

En este proyecto, se usará el modelo  $SIR$  usual con algunas modificaciones en las condiciones del mismo para mejorar el ajuste en los primeros días. Todas las implemtención del código se encuentran en el siguiente repositorio de GitHub.

## 2 Análisis descriptivo: Descripción de la base.

La base de datos es el conteo del crecimiento de los casos positivos y fallecimientos derivados por el virus  $SARS-COV-2$ .

El objetivo de este proyecto es ajustar la dinámica de la difusión del virus entre la población a través de un modelo  $SIR$ . En este caso, se enfocará en lo recabado por China.

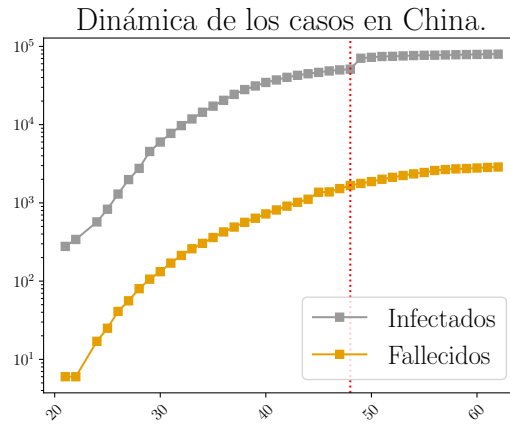


Figure 1: Número de casos y fallecimientos por  $SARS - COV - 2$ . La línea vertical roja muestra el último registro de los datos antes de un ajuste de cuentas de los mismos.

De esta base, sólo tenemos *dos* de los tres estados necesarios para el modelo  $SIR$ ; por sus siglas, nos hace falta el estado  $S$ . Además, en ese caso simplista, podemos considerar el caso  $R$  como aquellas personas que fallecieron. En la figura 1, podemos ver la dinámica que tuvieron. Por un lado, se tienen que en ambos casos se tiene un crecimiento sin aparente final; principalmente, antes de los primeros 30 días, los infectados se incrementaban mucho más que después de esa fecha. Por otro lado, notemos que los datos parecen ser no tan nobles como esperamos. Antes del día 50, tenemos una discontinuidad en la base. Además, de esto no reporta un crecimiento significativo. Por lo cual, para este análisis sólo se considera la información que esté antes de este evento (ver línea vertical roja 1).

Por otro lado, tenemos la problemática presente que es la ausencia de un estado de susceptibilidad ( $S$ ). Para ello, hagamos tengamos un supuesto fuerte en la dinámica el cual será que no exista crecimiento en la población; esto es que

la suma de cada uno de los estados sea igual a una constante. Así, si  $N$  es esta constante y  $I_i$  y  $R_i$  son los infectados y recuperados al tiempo  $i$  respectivamente, entonces

$$S_i = N - I_i - R_i$$

Al dividir entre  $N$ , tenemos

$$\frac{S_i}{N} = 1 - \frac{I_i}{N} - \frac{R_i}{N}$$

Lo cual, además de mencionar que podemos obtener el estado  $S$  a partir de los estados  $R$  e  $I$ , nos es conveniente considerar el porcentaje de la población. Esto es ver qué porcentaje de  $N$  tiene cada estado y cómo cambian en el tiempo. Definamos a  $N = 56 \times 10^3$ .

Considerando lo anterior, debemos ajustar el sistema a los datos en la figura 2.

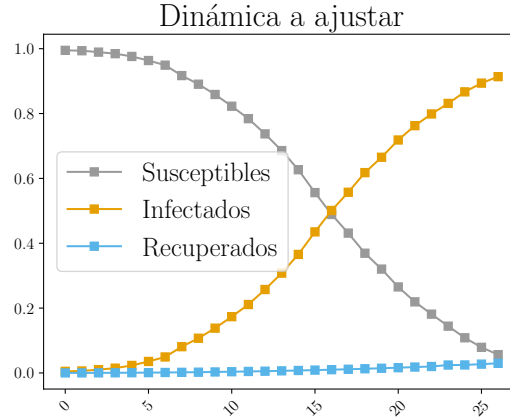


Figure 2: Datos para realizar el ajuste de la dinámica.

De este último gráfico, claramente tenemos una dinámica típica del modelo *SIR*. *A priori*, se tiene que la tasa de contagio es bastante alta; pues, notemos como los infectados tienen un aumento considerable en apenas 15 días. Aunque, por el otro lado, tenemos que la recuperación es bastante lenta.

### 3 Presentación del modelo: Modelo (SIR)

Aquí proponemos el uso de un modelo SIR modificado como anteriormente, basado en el modelo SIR clásico. Cabe recordar que uno de los principales supuestos del modelo SIR clásico es la existencia de una mezcla homogénea de las poblaciones infectadas  $I$  y susceptibles  $S$  y que la población total  $N$  es

constante en el tiempo. Además, en el modelo  $SIR$ , la población susceptible  $S$  disminuye monótonamente hacia cero. Sin embargo, en términos generales, estos supuestos no son válidos en el caso de la propagación del virus  $SARS-COV-2$ , ya que surgen nuevos epicentros en todo el mundo en diferentes momentos y con frecuencias.

Así, el modelo  $SIR$  está dado por el sistema de tres ecuaciones diferenciales ordinarias. En consecuencia, este modelo está diseñado para eliminar muchas de las complejidades asociadas con la evolución en tiempo real de la propagación del virus, de una manera que es útil tanto cuantitativa como cualitativamente, al menos para los primeros días de la evolución del sistema.

Formalmente, este es un sistema dinámico que viene dado por tres EDOs acopladas que describen la evolución temporal de las tres poblaciones, siguientes:

- Los individuos susceptibles,  $S$ : Son aquellos individuos que no están infectados y que pueden infectarse. Un individuo susceptible puede infectarse o seguir siendo susceptible.
- Individuos infectados,  $I$ : Son aquellos individuos que ya han sido infectados por el virus y pueden transmitirlo a los individuos susceptibles. Un individuo infectado puede permanecer infectado, y puede ser retirado de la población siempre y cuando este fallezca.
- Individuos retirados,  $R$ : Son aquellos individuos que han fallecido.

Por lo que, nuestro modelo resultante es

$$\dot{S} = -\beta SI \quad (1)$$

$$\dot{I} = \beta SI - \gamma I \quad (2)$$

$$\dot{R} = \gamma I \quad (3)$$

donde  $1 = S + I + R$ ,  $\beta, \gamma \in \mathbf{R}$ . Notemos los estados son proporciones.

### 3.0.1 Algunas consideraciones del modelo y consecuencias del mismo.

Como se mencionó, un supuesto fuerte es negar la reproducción o crecimiento de la población en cuestión. Entonces, dado que tenemos  $N = S + I + R = 1$ , tenemos que el cambio es

$$0 = \dot{N} = \dot{S} + \dot{I} + \dot{R}$$

Por lo que

$$\dot{S} = -\dot{I} - \dot{R}$$

Esto es, que podemos ver este sistema de 3 variables (S,I,R) a sólo dos. Esto nos da la necesidad vista en la descripción de los datos.

Sabemos que la evolución o crecimiento de los infectados existirá si

$$\dot{I} = \beta SI - \gamma I > 0$$

Equivalentemente,

$$R_0 = \frac{\beta}{\gamma} S > 1$$

Por lo que, el aumento en los casos estará dado si  $R_0 > 1$ ; y se *evitaría* la pandemia si  $R_0 \leq 1$ . Notemos que el valor  $R_0$  depende del valor  $S$  en ese momento. Entonces, si es que tenemos una cantidad no necesaria tal que  $R_0 > 1$ ; entonces la infección de todo el sistema no sólo depende de  $\beta$  y  $\gamma$ , sino de la población susceptible.

## 4 Ajuste.

Para el ajuste, hagamos uso de la paquetería *PDEParams.py* y sobre los mostrados en la figura 2. De esto, tenemos un ajuste bueno al considerar que de 3 variables esperadas a estimar sólo tenemos datos de dos variables.

$\gamma$	$\beta$
0.00924	0.340121

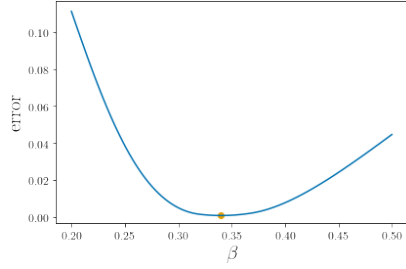
Table 1: Mejores estimadores para  $\gamma$  y  $\beta$

	$\gamma$	$\beta$
count	100.000000	100.000000
mean	0.009240	0.340122
std	0.000005	0.000012
min	0.009210	0.340054
25%	0.009240	0.340121
50%	0.009240	0.340121
75%	0.009240	0.340121
max	0.009267	0.340183

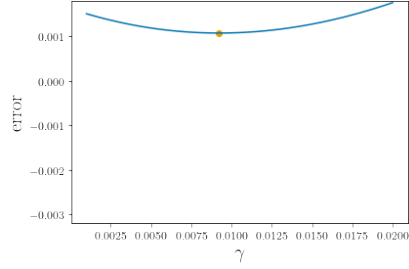
Table 2: Estadísticos para  $\gamma$  y  $\beta$

Así, del ajuste podemos ver que la significancia e identificabilidad de los mismos es buena. Principalmente, viendo de la desviación estándar es bastante pequeña.

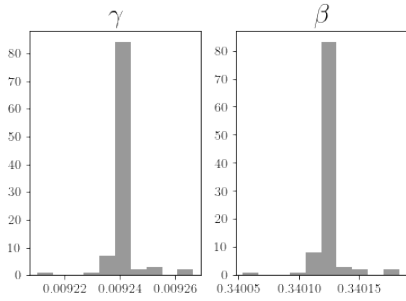
Por parte de sus perfiles obtenidos en 3b y 3a, tenemos una buena identificabilidad vista en el resumen anterior. Si bien, el obtenido para  $\gamma$  pudiera ser malo dando indicio a algo falso, basta con recordar que estamos estimando 3 variables a partir de 2. Por lo cual, a pesar de tener una variable faltante, se tiene un buen resultado. Por otro lado, tenemos que el perfil  $\beta$  es bastante bueno pues es una parábola aunque se esté abriendo por la parte derecha.



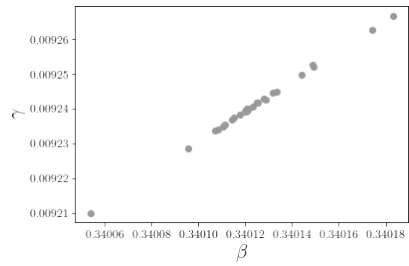
(a) Perfil de  $\beta$



(b) Perfil de  $\gamma$



(c) Distribuciones individuales.



(d) Diagrama de dispersión.

Esta problemática se puede ver en el diagrama de dispersión en la figura 3d ya que se tiene una correlación positiva. Sin embargo, notemos que las distribuciones individuales tienen una media bastante clara y una varianza cercana a cero (ver figura 3c).

Si bien tenemos esas problemáticas, notemos que se tiene un ajuste razonablemente bueno en el gráfico 4. De este podemos ver un error cuadrático medio (*MDE*) de apenas milésimas. Además, cabe recordar que nuestra variable  $S$  era completamente desconocida y se pudo inferir a través del modelo.

Además, cabe recalcar que el comportamiento cualitativo de la epidemia es bastante certero. Notemos que de las curvas  $S$  e  $I$ , antes del tiempo  $t = 15$ , tienen la misma concavidad que en los datos originales. Este mismo hecho ocurre para el comportamiento posterior a  $t = 15$ . Donde queda un tratado especial de los datos es en la variable  $R$  ya que del modelo parece crecer más rápido que en la base original. Esto, seguramente, se pudiera ser al asumir que los recuperados eran únicamente las personas que fallecen después de ser infectados.

## 5 Conclusiones.

Si bien se tienen problemas en una aproximación cuantitativa, la aproximación cualitativa se tiene un buen desempeño pues tenemos las mismas concavidades a través del tiempo.

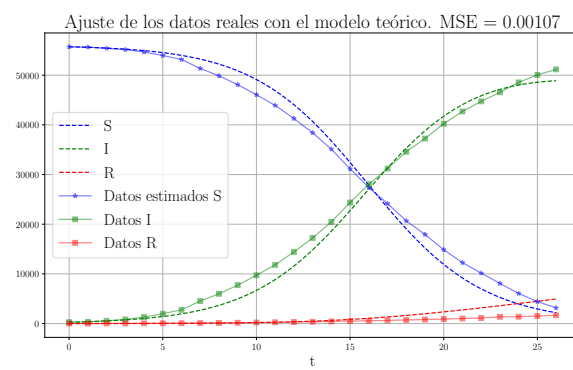


Figure 4: Ajuste final sobre los datos.