

# BB101: INTRODUCTION TO BIOLOGY

## EPIDEMIOLOGY PROJECT REPORT

K.Sai Saran, 2006317, Mechanical Engineering

The region assigned to us is Telangana State, India. The outbreak of the novel corona virus disease (COVID-19) brought considerable turmoil all around the world. We sincerely hope the situation will get better soon but at the same time we are interested about the mechanism of the spread of the disease. It is found that there is a simple mathematical model named SIR model by which predictions can be made which caught our attention and encouraged us to learn about it.

### ABSTRACT:

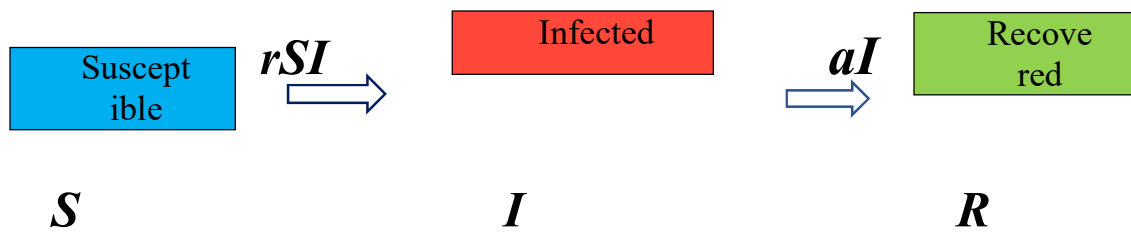
This project aims to understand the basics of epidemiology and appreciate its complexity. The process of fitting the SIR epidemic model to COVID data is addressed. Numerical simulations are carried out. Implications of our analytical and numerical findings are discussed critically.

### INTRODUCTION:

SIR model is a kind of compartmental model describing the dynamics of infectious disease. The model divides the population into compartments. Each compartment is expected to have the same characteristics. SIR represents the three compartments segmented by the model.

- Susceptible
- Infectious
- Recovered

**Susceptible** is a group of people who are vulnerable to exposure with infectious people. They can be patient when the infection happens. The group of **infectious** represents the infected people. They can pass the disease to susceptible people and can be recovered in a specific period. **Recovered** people get immunity so that they are not susceptible to the same illness anymore. But in this project we included deaths also to be recovered because once dead they can't spread the disease anymore neither can they be susceptible again. SIR model is a framework describing how the number of people in each group can change over time. This model is unidirectional.



$$\begin{aligned}\frac{dS}{dt} &= -rSI \\ \frac{dI}{dt} &= rSI - aI \\ \frac{dR}{dt} &= aI\end{aligned}$$

Figure 1: differential equations that determine the SIR model, where  $r$  stand for transmission rate and  $a$  stands for recovery or removal rate

SIR model allows us to describe the number of people in each compartment with the ordinary differential equation. 'r' is a parameter controlling how much the disease can be transmitted through exposure. It is determined by the chance of contact and the probability of disease transmission. 'a' is a parameter expressing how much the disease can be recovered in a specific period. Once the people are healed, they get immunity. There is no chance for them to go back susceptible again. We do not consider the effect of the natural death or birth rate because the model assumes the outstanding period of the disease is much shorter than the lifetime of the human.

*Basic Reproduction number,  $R_0$*  is the average number of people infected from one other person. If it is high, the probability of pandemic is also higher.

## APPLICATION OF SIR MODEL TO COVID DATA:

### Data source:

The data for infectives, recovered and dead is collected from government of india website: <https://www.covid19india.org>

### Guessing parameters:

The first step to follow while applying SIR model to any data set is to guess the parameters 'a' and 'r'. Though this can be done randomly since we had a clear data set we followed the averaging process. We divided our data into chunks of weeks and then using the below equation we calculated  $\rho$  for each week and then obtained the average value of it.

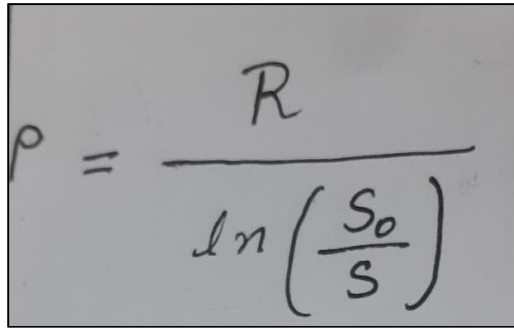

$$\rho = \frac{R}{\ln\left(\frac{S_0}{S}\right)}$$

Figure 2: Formula used for calculating value of  $\rho$

Now using the no. of patients recovered (recovered from infection + dead) in each week and observing the trends in it we calculated  $dR/dt$ . Then using the below relation we calculated 'a' for each week and then averaged it.

Finally using the relation  $\rho = \frac{a}{r}$  we calculated average value of r manually.

Lastly  $\alpha$  and  $\phi$  were calculated using the following equations.

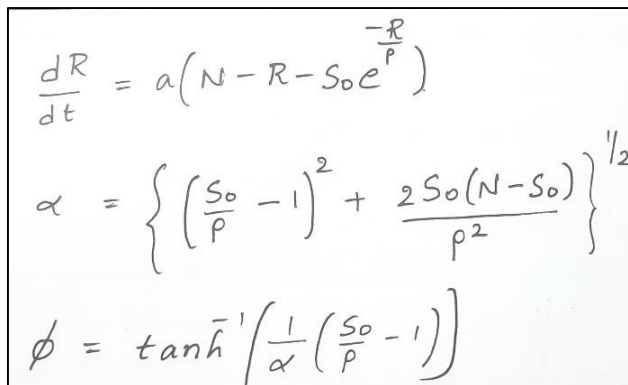

$$\begin{aligned}\frac{dR}{dt} &= a\left(N - R - S_0 e^{\frac{-R}{\rho}}\right) \\ \alpha &= \left\{ \left(\frac{S_0}{\rho} - 1\right)^2 + \frac{2S_0(N - S_0)}{\rho^2} \right\}^{1/2} \\ \phi &= \tanh^{-1}\left(\frac{1}{\alpha}\left(\frac{S_0}{\rho} - 1\right)\right)\end{aligned}$$

Figure 3: Formulae used for calculating parameters.

**Table 1: Values obtained manually(average method):**

| Parameter     | Value           |
|---------------|-----------------|
| $\rho$        | 1398473.66      |
| a             | 0.176019        |
| r             | 1.264e-08       |
| $\alpha$      | 1.8020031503308 |
| $\varnothing$ | 8.64544         |
| Ro            | 2.7923999284    |
| It            | 5.6812 days     |

**Simulation:**

After obtaining the parameters manually we proceeded by writing a code in python that simulates the equations and plots the graph accordingly using these parameters. Few special functions used in code are odeint, integrate, plot.

**Fitting procedure:**

Then using some special functions in python like curve\_fit we tried optimizing errors and obtain exact parameters. We achieved this by writing a code that takes the dataset and average values of parameters (guess values) as input and outputs us the exact values after fitting the curve to the dataset. These are the results that we got after executing that code. The guess value of  $\varnothing$  is changed from 8.64544 to 20.64544 to account for curve fitting.

**Table 2: Values obtained through curve fitting:**

| Parameter     | Value            |
|---------------|------------------|
| $\rho$        | 14246789.6       |
| a             | 0.0242694435     |
| r             | 1.7035026263e-09 |
| $\alpha$      | 1.73746639       |
| $\varnothing$ | 6.58613603       |
| Ro            | 2.7374587605     |
| It            | 41.2041days      |

### Minimizing errors:

After getting the values from the curve\_fit code, we still encountered error in data displayed by our model. In order to minimize errors we tried changing the values of parameters. First by varying the values of 'a' we observed the trends and variations in the graphs and accordingly changed the other parameters till we are left with minimum error possible. These are the final values of parameters that fits very closely to our actual data.

**Table 3: Final Fitted parameters:**

| Parameter | Value                 |
|-----------|-----------------------|
| $\rho$    | 18764243.5040         |
| a         | 0.0707132571          |
| r         | 3.768509945111835e-09 |
| $\alpha$  | 1.79296644            |
| $\phi$    | 6.06904542            |
| Ro        | 2.078420512907        |
| It        | 14.144days            |

### Error Calculation:

Though there are many ways to find the error in a calculation we relied on rms value and obtained the error using the python functions through coding.

#### **When average parameters or guess values are used:**

Error in observed and predicted I(t) values: 3532096.080788228

Error in observed and predicted R(t) values: 23639380.9795963

#### **When parameters after curve fitting are used:**

Error in observed and predicted I(t) values: 1294.5838748974902

Error in observed and predicted R(t) values: 1397.8433604452723

#### **When final parameters after minimizing error are used:**

Error in observed and predicted I(t) values: 393.5777740279937

Error in observed and predicted R(t) values: 394.9774450853022

After multiple trials we finally brought down the error from 3532096.0808 to 393.5777 in case of Infected curve and from 23639380.9796 to 394.9774 in case of Recovered population which seems like promising and satisfying.

## RESULTS:

### Graphs:

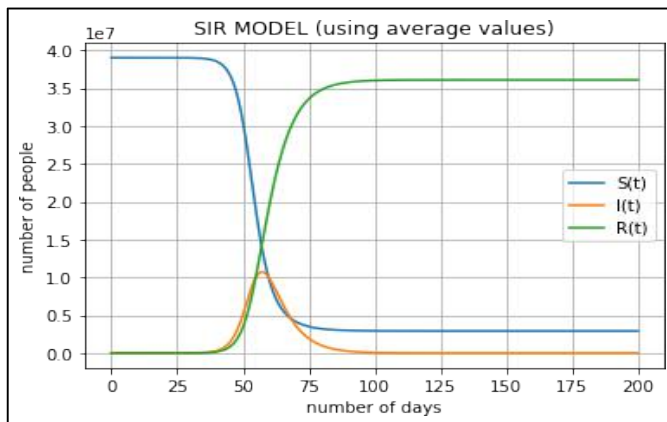


Figure 4 : SIR plot using average values.

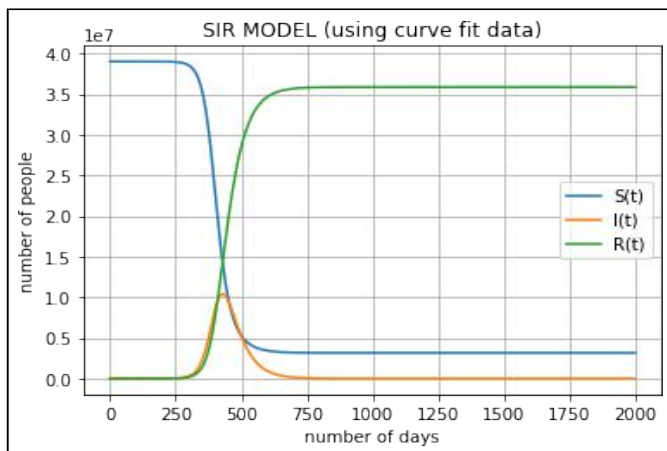


Figure 5 : SIR plot using curve fit values.

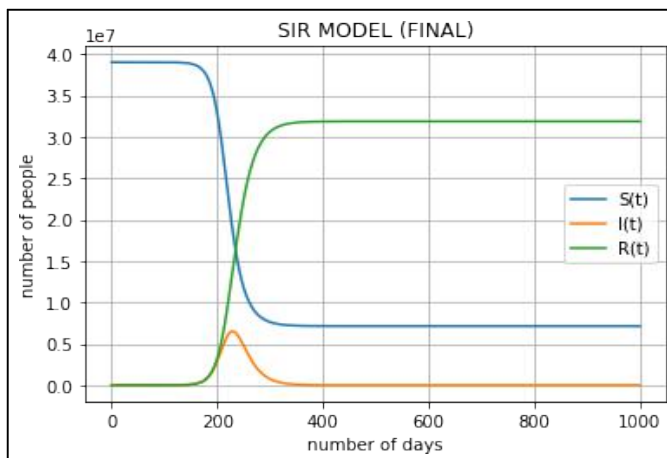


Figure 6: SIR plot using final values.

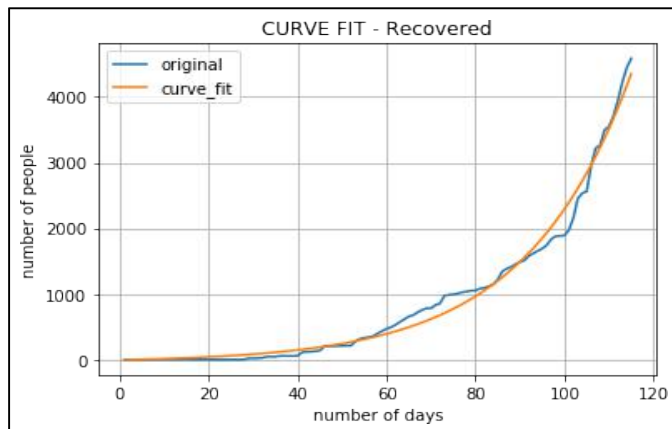


Figure 7 : Graph showing Curve fitting for recovered data.

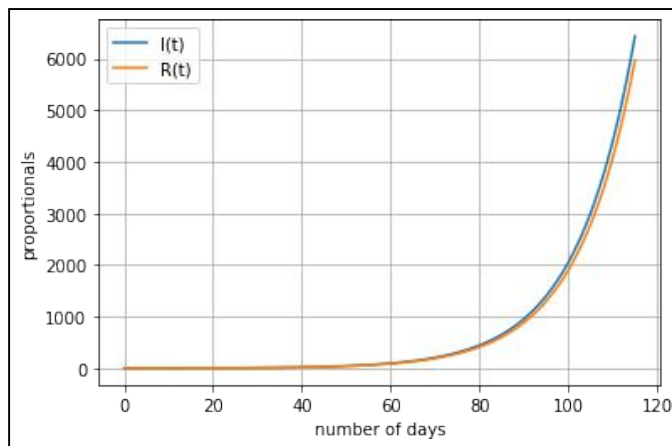


Figure 8: Final plot cut to 115 days

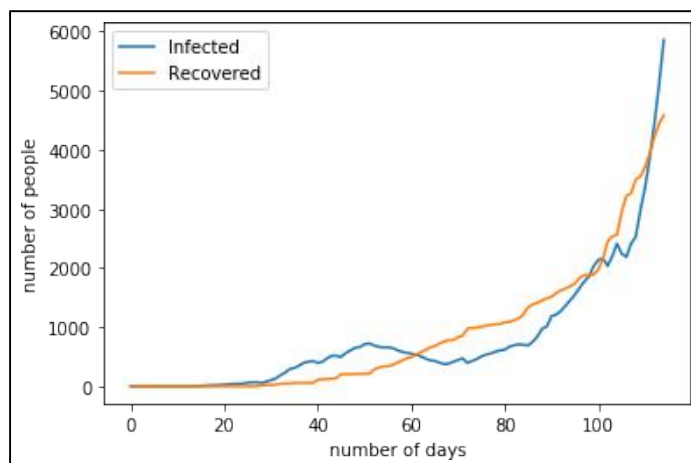


Figure 9: Observed data plot.

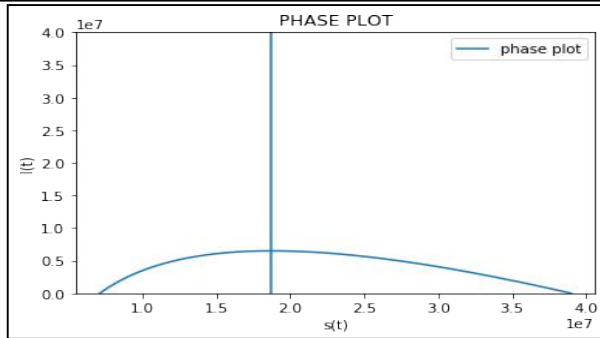


Figure 10: Phase plot for  $S(t)$  and  $I(t)$ . The vertical line is drawn at  $S(t)=\rho$

## Predictions:

After finally simulating the code with final fitted parameters, we succeeded in predicting the following for Telangana state:

Reproduction rate,  $R_0 = 2.078420512907102$

Average infection time,  $1/a = 14.141619846273494$

Maximum number of infected individuals: 6506614.305298221

Time location of maximum number of infected individuals: 229 days

Maximum number of recovered is 31860700.13293012

Duration of epidemic is 610 days.

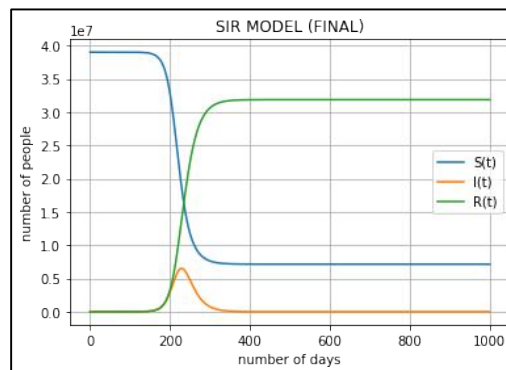


Figure 11 : SIR Model.

## INTERPRETATION:

Experimenting the correctness of the model:

- We tried to fit the data for 105 days (02 march to 15 june) in SIR model by curve fitting, minimising errors and various procedures. And then when this model is simulated for 115 days (02 march to 24 june), the model could correctly predict the number of infectives with minimum error. Whereas a considerable error was observed in recovered data.



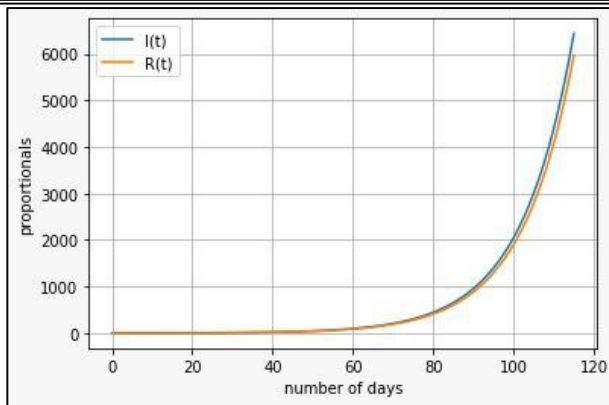


Figure 12: Simulation plot for 115 days.

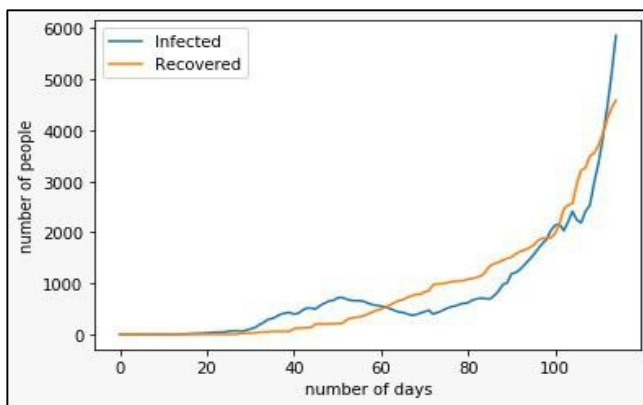


Figure 13: Observed data for 115 days.

## CONCLUSIONS:

1. In the graph simulated by SIR model the number of infected is strictly increasing while in the real life there was a peak between 40 -50 days. The violation of lockdown and social distancing measures by the people can be the reason for that. The same can be applied to other few peaks in the original data.
2. Recently the state government started Home quarantine instead of providing free treatment for the infected people. This affected the rate of recovery especially among the people who couldn't afford for nutritious food because most of the daily wage workers have lost their livelihood due to lockdown. This might be the reason for the deviation in recovered curve.
3. As per the simulated plot the no. of infected should have reached around 6500 by now but due to the spreading awareness among the people about the corona virus and strict lockdown and social distancing measures taken by the state government till now ,the infected population was below 6000(by 24 June) which is encouraging a bit .

4. As predicted by the model, the number of infected individuals keeps increasing drastically from now on i.e after 120 days. In order to control the infection and minimize the peak attained, we need to strictly follow the guidelines from the government and maintain social distancing. Rapid diagnosis of the infected individuals is the need of the hour for the better prevention of outspread of disease.

5. In order to increase the rate of recovery, special measures have to be taken to provide medical support and moral support specially for middle class and below poverty line population.

### **LIMITATIONS:**

1. SIR Model expects the susceptible to be homogenous, well-mixed and accessible to each other. Setting the whole population in the country is not realistic for sure.
2. We didn't consider the effect of natural death or birth rate.
3. Real scenario is not accounted in the model.

### **RESOURCES OR REFERENCES:**

The following are the websites we used as reference to complete the assignment:

- Video lectures by Sreenath Balakrishnan Sir.
- Data collection : <https://www.covid19india.org>
- Reference code: <https://www.youtube.com/watch?v=wEvZmBXgxO0>
- Regression techniques: <https://youtu.be/iOF7f5HCCQE>
- Curve fitting techniques: <https://youtu.be/4vryPwLtjIY>
- More information on SIR model :  
<http://greenteapress.com/modsimpy/ModSimPy3.pdf>