MATH 5131

Analysis for SARS Epidemic in Hongkong, China Using SIR Model

Suwen Wu, Wanning Zhou, Qidi Xu, Ning Li, Bowen Ma

Analysis for SARS Epidemic in Hongkong, China Using SIR Model

Introduction

At the end of February 2003, a viral respiratory disease - severe acute respiratory syndrome (SARS) was first identified and reported during an outbreak in China. The US Centers for Disease Control and Prevention (CDC) introduces it as "a viral respiratory illness caused by a coronavirus called SARS-associated coronavirus (SARS-CoV)". In addition, before its globe outbreak was contained in July 2003, it spread to several different countries in Asia, North America, South America, and Europe. And it became the first serious new disease in 21st century that can be easily transmitted. Fortunately, there have been no reported cases of SARS in the world since 2004.

In order to have a better understanding for the characteristics of infectious diseases, such as SARS, as well as predicting the future growth trend and helping inform public health interventions, the mathematical modelling of infectious diseases is needed.

To achieve this task, we analyze the 2003 SARS epidemic within Hongkong, China from March 31st, 2003, to July 11th, 2003, using a published dataset on Kaggle. The reason for the selection of data is that China was under the most severe situation at that time, and data for Hongkong is more reliable, comparing with data for Mainland China and Taiwan. Moreover, the susceptible-infected-removed (SIR) model is used as the mathematical model in this project since SARS is an epidemic disease. The purpose of

this project is to analyze the SARS data and build the corresponding SIR model, to predict the trend of SARS in 2003. In this project, three models have been built, and only one of them performs well.

Model Assumptions

We make a few important assumptions on which to build the model:

- We assume that random differences between individuls can be neglected, so that all individuals have the same probability to contract the disease.

- We ignore the births and immigration in this model, so that no one joins this group.

- We assume all those who recover from the disease are then immune.

- We assume that there is no developed vaccination within this time period.

- We ignore the incubation period for the disease, setting it equal to zero.

The SIR Model

In this project, we apply the SIR model, which is a compartmental model in epidemiology. And it is always applied to the mathematical modelling of infectious diseases.

A mathematical description of SIR model can be represented by the following three differential equations:

$$\begin{cases} \dfrac{dS}{dt} = -\beta SI \\[2mm] \dfrac{dI}{dt} = \beta SI - \gamma I \\[2mm] \dfrac{dR}{dt} = \gamma I \end{cases}$$

The basic reproductive ratio R_0 for the system is $R_0 = \left(\dfrac{\beta}{\gamma}\right) S_0$

β: the transmission rate

γ: the removal rate

S(t): The number of individuals who are vulnerable and are under risk of being infected at time t.

I(t): The number of individuals who have already been infected by SARS and can transmit it to those who are susceptible at time t.

R(t): The number of individuals who have removed from SARS at time t, including those who are recovered and dead.

Analysis of Reported Data

We collect data on SARS from Kaggle and select Hongkong as our research object. The period of SARS data set is from March 31st, 2003, to July 11th, 2003.

1) **Process the data:**

We import the collected data from Kaggle into Python and draw out the following graph. The red curve represents the actual number of people diagnosed and the green curve represents the actual number of people removed.
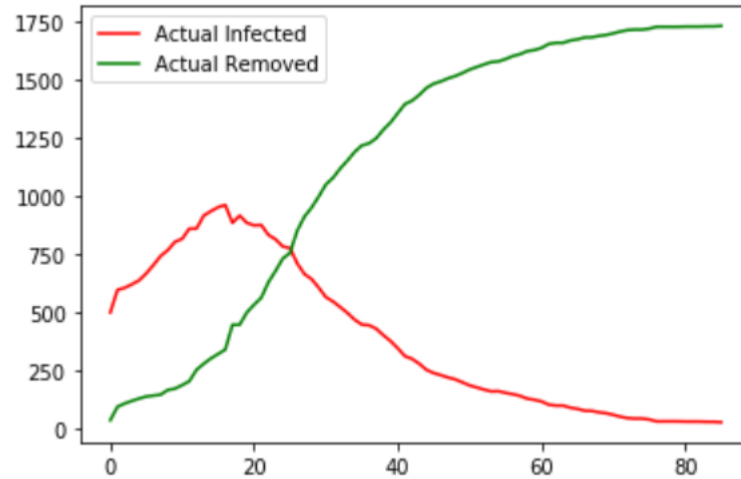
FIGURE 1. The real data of SARS

2) **The differential equations for SIR model for first attempt:**

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta}{N}SI & \text{①} \\[2mm] \dfrac{dI}{dt} = \dfrac{\beta}{N}SI - \gamma I & \text{②} \\[2mm] \dfrac{dR}{dt} = \gamma I & \text{③} \end{cases}$$

There are two major assumptions in the model: in the absence of infection $I + R = 0$ and

$S \approx N$, which means that everyone in Hongkong is susceptible in the absence of infection.

Depending on our initial assumption, we obtain the following equation from ②

$$\frac{dI}{dt} = I(\beta - \gamma)$$

$$\frac{dI}{I} = (\beta - \gamma)dt$$

$$\int \frac{1}{I} dI = \int (\beta - \gamma)\, dt$$

$$lnI = (\beta - \gamma)t + C$$

$$I = e^{(\beta-\gamma)t}K$$

$\because I(0) = I_0$

$\therefore K = I_0$

$$I = I_0 \cdot e^{(\beta-\gamma)t}$$

Determine of transmission rate β and the $m = \beta - \gamma$ value:

$$\frac{dI}{dt} = mI$$

$$I(t) = I_0 e^{mt}$$

$$lnI = mt + lnI_0 \qquad (*)$$

Determine of the removal rate γ:

$$\frac{dR}{dt} = \gamma I$$

When time changes $dt = 1$, we obtain that

$$\gamma = \frac{R(t+1)-R(t)}{I(t)} \qquad (**)$$

3) **Determine the parameters in the model:**

In order to calculate the parameters, we try two different ways. The first way is using daily

new cases as $I(t)$, and the second one is using accumulated cases as $I(t)$.

We tried two different assumptions, which are summarized as follows:

Version 1: $I(t)$: Daily new cases without accumulation

First, we use the Least Square method to compute the value of m and $\ln(I_0)$ based on

(*) equation. We obtain that $m = -0.0473$, $\ln(I_0) = 7.3923$.

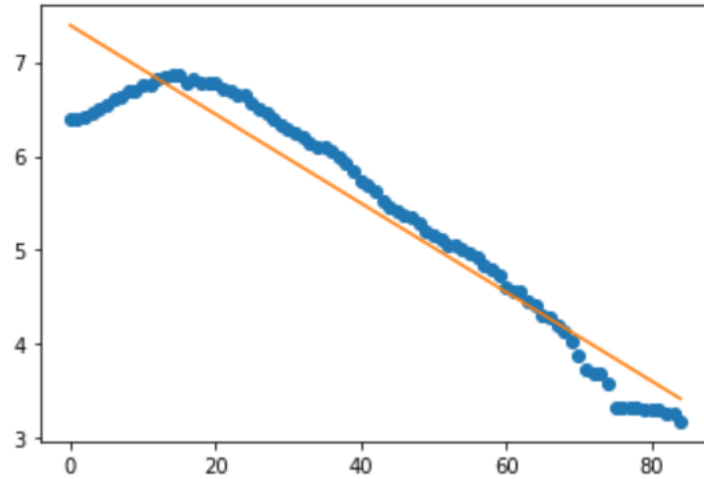Then plug them into equation $y = m \cdot x + \ln(I_0)$, and draw the graph.

FIGURE 2. The fitting image of first try

It can be seen from the graph that the result can fit the line well.

Then we use Least Square method to compute the value of γ based on (**) equation, and

we obtain that $\gamma = \frac{R(t+1)-R(t)}{I(t)} = 0.0464$.

Based on the value of m and γ, we calculate $\beta = m + \gamma = -0.0473 + 0.0464 =$

-0.0009.

Based on the value of β and γ, we calculate $R_0 = \frac{\beta}{\gamma} = -0.01939655172413804$.

Based on the value of γ, we calculate $T = \frac{1}{\gamma} = 21.551724137931036$.

However, we find out that $R_0 < 1$, which means SARS is no longer contagious. It does

not match the fact.

Version 2: $I(t)$: Daily accumulated cases

Similarly, first we use the Least Square method to compute the value of m and $\ln(I_0)$

based on (**) equation. We obtain that $m = 0.0080, \ln(I_0) = 6.9804$.

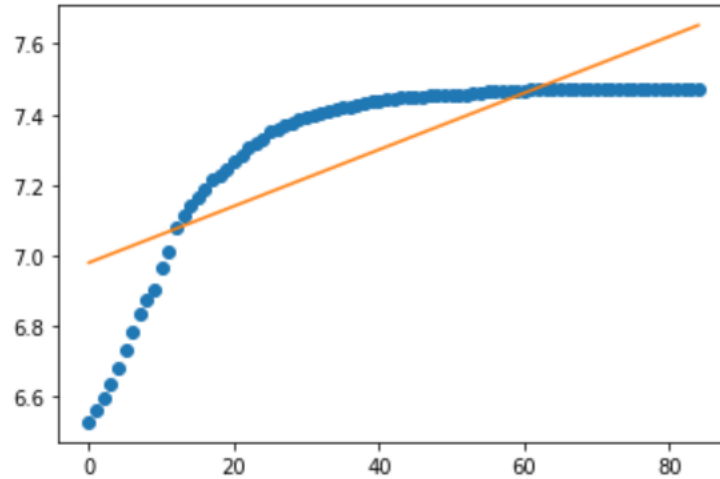Then plug them into equation $y = m \cdot x + \ln(I_0)$, and draw the graph.

FIGURE 3. The fitting image of second try

It can be seen from the graph that the result cannot fit the line well.

Then we use Least Square method to compute the value of γ based on (**) equation, and

we obtain that $\gamma = \frac{R(t+1)-R(t)}{I(t)} = 0.0116$.

Based on the value of m and γ, we calculate $\beta = m + \gamma = -0.0473 + 0.0464 = 0.0196$.

Based on the value of β and γ, we calculate $R_0 = \frac{\beta}{\gamma} = 1.6896551724137931$.

Based on the value of γ, we calculate $T = \frac{1}{\gamma} = 86.20689655172414$.

However, we find out that T is too large, which is impossible.

Thus, we can come up with that our initial assumption $S \approx N$ is wrong, and the correct

assumption is that $S \ll N$, which means under some control measures, it is impossible

for the whole population of Hongkong to be infected.

4) **The differential equations for SIR model for second attempt:**

$$\begin{cases} \dfrac{dS}{dt} = -\beta SI \\ \dfrac{dI}{dt} = \beta SI - \gamma I \\ \dfrac{dR}{dt} = \gamma I \end{cases}$$

We Assume that $S \ll N$

$\because N = S_0 + I_0 + R_0$

$\therefore S = N - I - R = S_0 + I_0 + R_0 - I - R$

$$\frac{dI/dt}{dS/dt} = \frac{\beta SI - \gamma I}{-\beta SI} = \frac{\gamma}{\beta S} - 1 = \frac{\rho}{S} - 1 \quad (\rho = \frac{\gamma}{\beta})$$

$$\frac{dI}{dS} = \frac{\rho}{S} - 1$$

$$dI = (\frac{\rho}{S} - 1)dS$$

$$\int dI = \int (\frac{\rho}{S} - 1)dS$$

$$I = \rho \ln S - S + C$$

$\because$ initial value $I(0) = I_0,\ S(0) = S_0$

$$C = I_0 - \rho \ln S + S_0$$

$$I = I_0 + S_0 - S + \frac{\gamma}{\beta} \ln \left(\frac{S}{S_0}\right) \qquad (***)$$

5) **Determine the parameters in the model:**

Step 1. Calculate γ

The estimated removed days T can be evaluated by Figure 4, in which the number of infected cases at first day equals to the number of removed cases at T days. And we can see that T is about 21.

We can calculate γ using the value of T: $\gamma = \frac{1}{T} = \frac{1}{21}$.
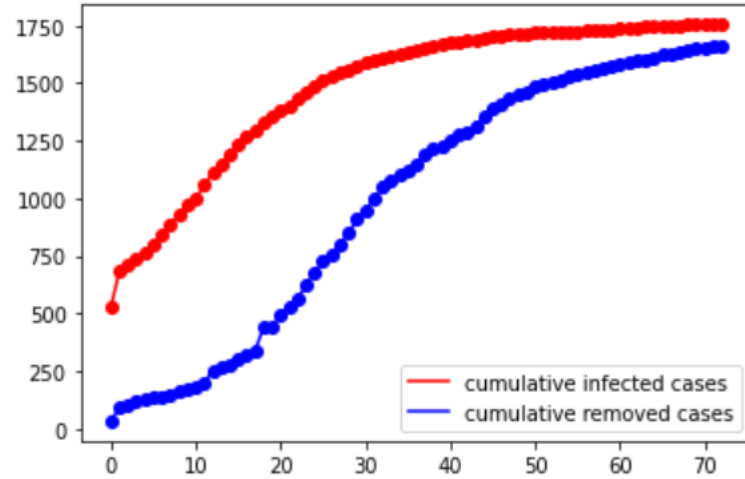
FIGURE 4. The image of cumulative infected cases and removed cases

Step 2. Use Least Square method to calculate the best S_0 and β

We try different S_0 between 800 and 1801, increment by 1, and different β between

$6 * 10^{-5}$ and $(1.8 * 10^{-4} + 1 * 10^{-7})$, increment by $1 * 10^{-7}$. We plug these different

pairs of S_0 and β into the (***) equations.

We use Least Square method to compute the value of S_0 and β based on equation

(***), where $I_0 = 497, R_0 = 33$. And we obtain the best $S_0 = 1271$ and the best

$\beta = 9.420000000000083 * 10^{-5}$.

Step 3. Simulation of the SIR model

By simulating the SIR model using the parameters that we calculated, we can draw out the
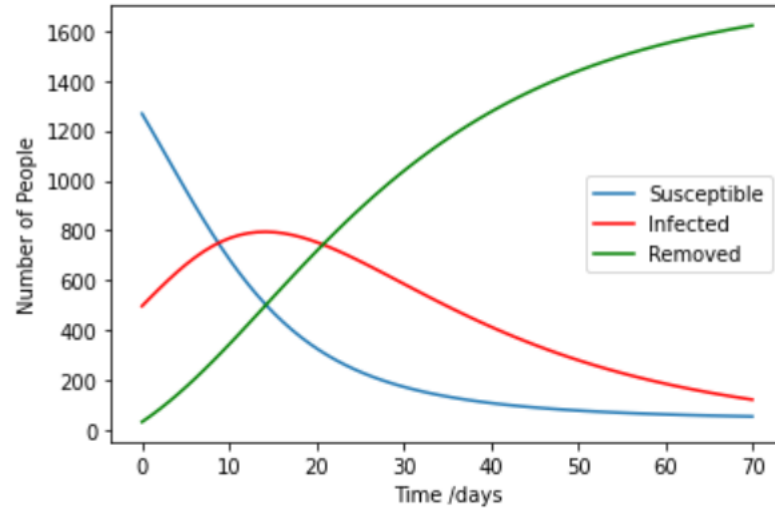
graph Figure 5.

FIGURE 5. The image of prediction through SIR model

Conclusion

1) Review

First, we select Hong Kong data as our research sample, and the time period is from March 31st, 2003, to July 11th, 2003 is the peak time of SARS spread. Thus, the SIR model we established can more accurately predict the subsequent virus transmission rate.

Secondly, we initially assume that $S \approx N$, which means all the people in Hongkong will be infected, but the results of the model do not match the actual situation. Then we assume that $S \ll N$, which means only some people in Hongkong will be infected. By calculating the $\beta = 9.420000000000083 * 10^{-5}$, $\gamma = \frac{1}{21}$ and $S_0 = 1271$, we can compute that $R_0 = \frac{\beta}{\gamma} S_0 = \frac{9.420000000000083 * 10^{-5}}{\frac{1}{21}} * 1271 \approx 2.514$, it means that a SARS infected person may infect 2.514 people. While the basic reproduction number of SARS is around 2.5, it is close to the one that we computed.

As the result, there are some slightly different between the observed data and expected data. Although the trend of infected population and the trend of removed population are similar in the two graphs, we can notice that the peak of infected population can't reach the actual maximum value. So, there may be some errors in our model.

2) Limitation

Our modified SIR model has many limitations.

First, since SARS was an infectious disease in 2003, the statistical method at that time was not as precise as it is now, there may be a deviation in the collection of data from the beginning.

Second, since the data used in this project is between March 31st, 2003, to July 11th, 2003, the time period is limited, which may lead to some errors in the establishment of the final model.

Reference

Dan Feng, et al. "The SARS Epidemic in Mainland China: Bringing Together All Epidemiological Data." Tropical Medicine & International Health, vol. 14, Blackwell, 2009, pp. 4–13.

El Zowalaty, Mohamed E., and Josef D. Järhult. "From SARS to COVID-19: A Previously Unknown SARS- Related Coronavirus (SARS-CoV-2) of Pandemic Potential Infecting Humans – Call for a One Health Approach." One Health, vol. 9, Elsevier B.V, 2020, pp. 100124–100124, doi:10.1016/j.onehlt.2020.100124. Accessed Nov 27th, 2021.

Harko, Tiberiu, et al. "Exact Analytical Solutions of the Susceptible-Infected-Recovered (SIR) Epidemic Model and of the SIR Model with Equal Death and Birth Rates." Applied Mathematics and Computation, vol. 236, Elsevier Inc, 2014, pp. 184–94, doi:10.1016/j.amc.2014.03.030. Accessed Nov 27th, 2021.

Lounis, Mohamed, and Dilip Kumar Bagal. "Estimation of SIR Model's Parameters of COVID-19 in Algeria." Bulletin of the National Research Centre, vol. 44, no. 1, Springer Berlin Heidelberg, 2020, pp. 180–180, doi:10.1186/s42269-020-00434-5. Accessed Nov 27th, 2021.

"SARS." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 6 Dec. 2017, https://www.cdc.gov/sars/index.html. Accessed Dec 9th, 2021.

"SARS 2003 Outbreak Dataset." Kaggle, 24 May 2020, https://www.kaggle.com/imdevskp/sars-outbreak-2003-complete-dataset. Accessed Nov 9th, 2021.