LAGRANGE COLLEGE

Citations Journal of Undergraduate Research
2021, Volume 18

Modeling COVID-19 Data using an SIR Model

Alyssa G. Lord

Faculty Mentor: Stacey L. Ernstberger, PhD
Mathematics

Abstract

When analyzing the spread of infectious diseases, a common model that is used is a differential equations system called the SIR model, representing the population divided into three categories: susceptible, infected, and removed. There are many variations and applications of this model, and in this research we apply the basic model to COVID-19 data from Japan and Brazil.

In February 2020, Coronavirus disease (COVID-19) became a concern worldwide. Various institutes began collecting data regarding the number of infections, recoveries, and deaths throughout the world [3, 7]. We use the gathered data from two countries, Japan and Brazil, and model the effects of the epidemic in the population using an ordinary differential equations system called the SIR Model. This model represents the susceptible (healthy, uninfected) population, the infected population, and the removed (no longer infected) population [1].

The SIR Model

The SIR Model represents the population using a single system of three different functions: the function of the susceptible group, the infected group, and the removed group. In our application, the removed group consists of both recovered individuals as well as the deaths caused by COVID-19. While forming this model, the following assumptions will be made:

- once recovered, individuals will no longer be susceptible
- deaths will only be caused by the infection
- there will be no births introduced to the population
- each individual has the same the chance of susceptibility, infection, and removal.

This model will be represented by a system of ordinary differential equations (ODE) [4]. The system of ODEs that represent the changes across the population is given by

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta S(t)I(t) \\ \frac{dI(t)}{dt} &= \beta S(t)I(t) - \gamma I(t) \\ \frac{dR(t)}{dt} &= \gamma I(t). \end{aligned} \tag{1}$$

Here, β represents the infection parameter that causes the susceptible population to decrease, when the population of the infected group is considered. When the population of the susceptible group is considered, β then causes the infected population to increase. When the removed population is considered, the removal parameter, γ, is introduced, causing a portion of the infected population to move to the removed population. From these parameters, the basic reproduction ratio, $R_0 = \frac{\beta N}{\gamma}$, can be used to find the expected number of new infections from a single infection. Typically when $R_0 > 1$, the infection will spread in a population. An epidemic with a large R_0, will be hard to control [2]. At any given point in time the sum of each populations will be equivalent to the total population, since each individual will either be susceptible, infected, or removed.

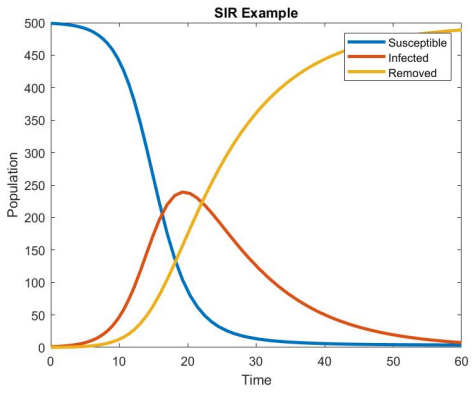

Figure 1: An example of the basic SIR model.

Although the SIR model is a standard differential equation system used to model epidemics within a population, it is interesting to note that the SIR model can also be used to ana-

lyze the use of social networking sites and even the applause of a crowd [6].

Implementing the Model

Now that we have established the basic SIR model, we will create MATLAB code to implement this model so that we can apply the system to our gathered data [5]. Prior to applying the data, we will discuss the basic implementation of the MATLAB code that we use to represent the ODE system.

The SIR Model

We first construct a function to represent the SIR model. This includes creating the basic model and passing in the parameters, β and γ, as well as the three populations at a given time. It will return the rate of change of each population, to be used in an ODE solver.

```
function dydt = odefun(~,y,p)
% input: population: y (in three parts)
%        parameter: p = [beta, gamma]
% output: rates of change of populations
        s = y(1); i = y(2); r = y(3);
        dydt = [-p(1)*s*i; p(1)*s*i - p(2)*i;
                p(2)*i];
end
```

Modeling the ODE

We use an ODE solver, ode45, to approximate the solutions. This uses the populations and the derivatives at each time step to approximate the populations at the next time step. It repeats this process through the duration of the time interval. Then we compute the least squares error between our approximate solution and our gathered data at each of our time steps [4]:

$$J = \sum_{i=1}^{n} y_{model_i} - y_{data_i} \tag{2}$$

Our goal is to find the parameters which form the solution set that best minimizes the least squares error (2). We create a function to form a solution set and return the given error as follows:

```
function err = odefit(act_t, act_y, p)
% input: time, raw data, parameters
% output: least squares error

    IC = [1 0.00015 0];
    % initial conditions:
    % scaled to represent 1 infected (i),
    % the rest are susceptible (s)

    [~,y] = ode45(@(t,y)odefun(t,y,p),act_t,IC);
    % approximates the solution at many points
    % over our time interval

    err = sum((y(:) - act_y).^2);
    % computes the least squares error
end
```

Parameter Estimation

Now, since there are functions representing the ODEs and the error from the ODE solver, we use the MATLAB tool `fminsearch` to approximate the infection and recovery parameters. This tool uses our error function to find the best fitting parameters by minimizing the least squares error. We start with an initial guess for the parameters and we form an approximated solution and a corresponding error. From there, the function tests different parameters and compares the resulting least squares error from each of those corresponding solution sets in order to determine the optimal parameters. If the initial parameter estimations are not reasonable, `fminsearch` will typically not be able to find a close approximation.

Application to COVID-19 Data

Now that we have established the basic SIR model and have implemented it in MATLAB, we attempt to model the COVID-19 data of different countries.

Data Collection

We collected the data for daily infections, deaths, and recoveries from the Johns Hopkins Data Center [3]. The text file of the data was stored as a CSV file to be loaded into a new MATLAB document. The susceptible data was found by taking the total population and subtracting the infected population for each day. The infected data on the site was cumulative and did not include the removed group, so the infected group per day was found by taking the data given for infections and subtracting the sum of deaths and recoveries. In this model we are combining deaths and recoveries into one removed group, and thus the sum of deaths and recoveries were added together to form one removed group. Because the population in most countries was so large in comparison to the actual infected population, the reported population was scaled down, while keeping the same infection and recovered data.

Japan COVID-19 Data

We chose to study the reported COVID-19 data for the island nation of Japan. Japan has a population of about 126.3 million people, but for the purposes of applying an SIR model to our data set, we will instead use a modified total population of one twentieth of the true population. This is because only 3.08 million were ultimate infected, with a maximum daily infected population reaching 46,551. Thus, the susceptible population was substantially larger than the infected and removed population.

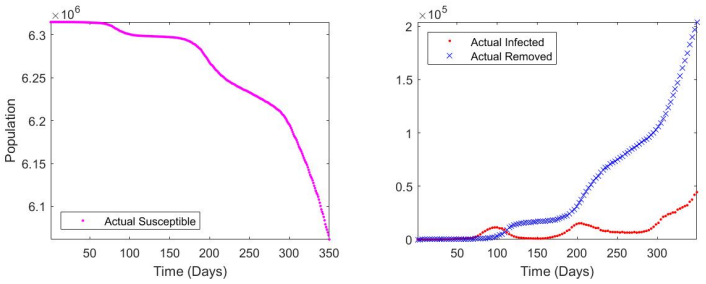

Figure 2: Susceptible, infected, and recovered populations in Japan.

To better visualize the data, the susceptible group was plotted separately from the other two groups as seen in Figure 2. The data and the approximation from our model can be seen together in Figure 3. We can see that as the infected group increases, the susceptible group decreases at nearly the same rate, until the removed population grows. Since the data has multiple spikes throughout the time interval, there are clearly components of the data that could not be modeled as well as others. These are likely contributed to data collection inconsistencies. as well as COVID-19 hotspots, formed from large gatherings o the population. For the Japan model, the infection parameter (β is 0.0800, and the removed parameter (γ) is 0.0681.

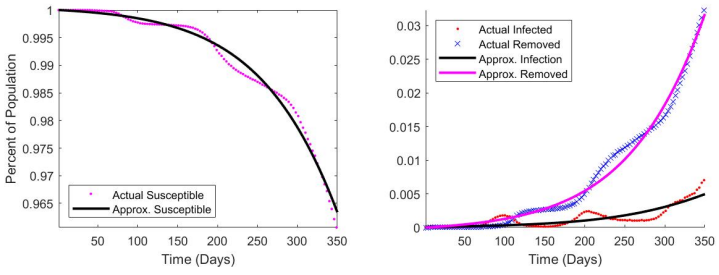

Figure 3: The COVID-19 data and models for Japan.

Brazil COVID-19 Data

We also chose to model Brazil COVID-19 data because it has many differences from Japan. Brazil has many bordering countries and has a larger population, 212.6 million people. We rescaled the magnitude of the susceptible population for the purposes of modeling the data. We can see the data and the SIR model approximation in Figure 4. We initially use only the first

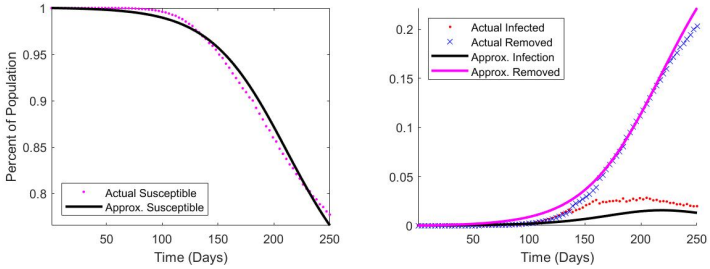

Figure 4: The COVID-19 data and models for Brazil (first 250 days).

250 days of gathered data from Brazil, and obtain an estimate representing the SIR population. We can see that Figure 4 resembles the initial basic model in Figure 1. For this model, the infection parameter (β) is 0.1433, and the removed parameter (γ) is 0.1075.

The reason that we originally only used the first 250 days of data is because there were some errors in the gathered data around day 280 in the model. If we used the optimized parameters from the 250 day model, seen in Figure 4, to attempt to model the entire data set (437 days), we would obtain ill-fitting results as seen in Figure 5. We expected this to be a poor

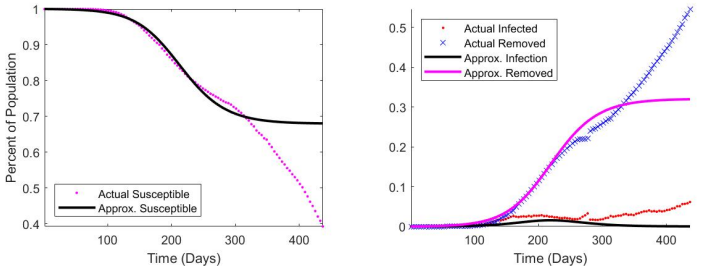

Figure 5: 437 days of COVID-19 data modeled with the 250 day solution set for Brazil.

fit because we are experiencing an ongoing pandemic and by only using a basic SIR model with the first 250 days to model the complete data set, we would certainly not be able to predict the behavior of the population system. To obtain a better fit we would have to make many modifications and use the full amount

of obtained data available. We included this poor fit as an example of why it is hard to use current data to predict the future of population dynamics in an epidemic.

We ultimately also used the full 437 days of Brazil COVID-19 data that were available to us at the time of the research, and in Figure 6 we see the resulting model. In this model,

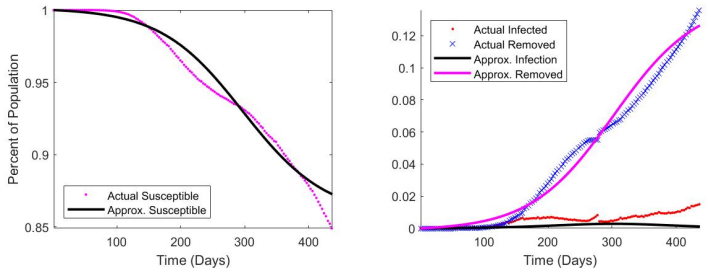

Figure 6: The COVID-19 data and models for Brazil (437 days).

the infection parameter (β) is 0.0756, and the removed parameter (γ) is 0.0423. It is clear that susceptible, infected, and removed population are better modeled in Figure 4 than in Figure 6. This is largely due to inconsistencies in data collection that occurred near day 200. It seems likely that the data was not gathered for several days and then reported all at once, causing a large irregularity which the model was unable to properly handle. Further, at the time of the data collection, Brazil was in the peak of a large second wave of infection, and the basic SIR model is only capable of handling one major infection wave. This shows us that while the SIR model is capable of handling some basic population modeling, it would need substantial modifications to be able to handle the dynamics that we have seen in the gathered COVID-19 data,

Conclusions

The basic SIR model can be used to make a rough approximation to the pandemic population data, and can reasonably represent the overall population dynamics if the data is relatively smooth and only has one primary epidemic wave. In order to truly capture the dynamics of COVID-19, we would have to make many modifications. We could lift the assumption that once the individual has recovered, they are no longer susceptible and we would need to introduce terms to handle the data spikes. We would also have to consider the percentage of the populations that have received the COVID-19 vaccine. Another way to better model the data might be to add parameters such as the demographics of the population, including, age, blood type, or chronic illnesses. This study could also be used to compare the severity of COVID-19 to other diseases, and ultimately we could use the comparison to create action plans and have a better understanding of the new disease.

References

[1] Bonhoeffer, Sebastian and Schafroth, Stefan. “SIR models of epidemics,” BioSym. World Wide Web. 2021.

[2] Fine P, Eames K, Heymann DL (April 2011). “Herd immunity”: a rough guide”. Clinical Infectious Diseases. 52 (7): 911–6.

[3] Johns Hopkins University COVID-19 Data Repository. Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. World Wide Web. 2021.

[4] Kaw, Autar and Kalu, E. E. Numerical Methods with Applications. https://nm.mathforcollege.com, 2009.

[5] Moler, Cleve, “Solving Ordinary Differential Equations in MATLAB, 6: ODE45,” https://www.mathworks.com, 2020.

[6] Rodrigues, Helena Sofia, “Application of SIR epidemiological model: new trends.” International Journal of Applied Mathematics and Informatics, Volume 10, 2016.

[7] Worldometer, “COVID-19 Coronavirus Pandemic,” https://www.worldometers.info/coronavirus, 2021.