Mathematical Theory of Infectious Disease Epidemics

Final Project in Numerical Methods
American University of Armenia

Grigor Hovhannisyan Aram Abrahamyan
Meruzhan Khachatryan

May 5, 2024

Contents

List of Figures

Abstract

Infectious diseases are a prevalent issue among everyone. The knowledge of its likely path and development can substantially help the government control the disease and hospitals correctly manage resources to accommodate the high number of patients. This is the aim of the SIR model, which this paper aims to explain and provide examples of. The SIR model has many modifications and extensions, some of which we will explore. We will also showcase and explain the analytic solution to the standard SIR model, but unfortunately, many of the modifications do not have an analytical solution; hence, we will approximate them numerically. The main methods to do so will be the Euler and the backward Euler methods, both of which have an error, which we will derive and showcase on graphs. All of the coding is done in Python with examples and explanations, as well as graphs and the explanation of the given graphs. Furthermore, we gathered and plotted graphs from real-world data concerning some famous diseases. We will compare them to the SIR model and determine how accurate it gets over time. Finally, we will draw conclusions and discuss how much of an impact the model can have on the world if used correctly.

1 Introduction

Infectious disease epidemics cause significant challenges to public health, necessitating mathematical models for understanding and managing their spread. The practical use of epidemic models must rely heavily on the realism put into the models. This means that a reasonable model can include only some possible effects but rather incorporate the mechanisms in the most straightforward possible fashion to maintain significant components that influence disease propagation. Before epidemic models are used to predict real phenomena, great care should be taken. However, even simple models should, and often do, pose important questions about the underlying mechanisms of infection spread and possible means of control of the disease or epidemic. There are classical papers by W. Kermack and A. McKendrick [1] that have greatly influenced the development of mathematical models for disease spread and are still in many epidemic situations. These first papers laid a foundation for modeling infections that confer complete immunity after recovery (or, in case of lethal diseases - death). The population is taken to be constant - no births or deaths other than from the disease are possible - consistent with the course of an epidemic being short compared with an individual's lifetime. Suppose a group of infected individuals is introduced into a large population. In that case, a fundamental problem is to describe the spread of the infection within the population as a function of time. Over time, the epidemic may come to an end. One of the most critical questions in epidemiology is to ascertain whether this occurs only when all the initially susceptible individuals have contracted the disease or if some interplay of infectivity, recovery, and mortality factors may result in an epidemic "die out" with many susceptibles still present in the unaffected population. In their first paper, Kermack and McKendrick start with the assumption that all members of the community are initially equally susceptible to the disease and that complete immunity is conferred after the infection. The population is divided into three distinct classes: the susceptibles, S, - healthy individuals who can catch the disease; the infected, I, - those who have the disease and can transmit it; and the removed, R, - individuals who have had the disease and are now immune to the infection (or removed from further propagation of the disease by some other means). Schematically, the individual goes through consecutive states $S \to I \to R$. Such models are often called the SIR models.

2 Mathematical Formulation of the SIR Model

2.1 Types of Variables

The first step in mathematical modeling process is defining independent and dependent variables. The independent variable is time t, measured in days. Afterward, two related sets of dependent variables are being defined. The first set of dependent variables counts people in each of the groups, each as a function of time:

$$S = S(t) \text{ is the number of } \textit{susceptible} \text{ individuals,}$$
$$I = I(t) \text{ is the number of } \textit{infected} \text{ individuals, and}$$
$$R = R(t) \text{ is the number of } \textit{recovered} \text{ individuals.}$$

The second set of dependent variables represents the *fraction* of the total population in each of the three categories. So, if N is the total population we have:

$$s(t) = \frac{S(t)}{N} \text{ the susceptible fraction of the population,}$$
$$i(t) = \frac{I(t)}{N} \text{ the infected fraction of the population, and}$$
$$r(t) = \frac{R(t)}{N} \text{ the recovered fraction of the population.}$$

Using fractions instead of population counts simplifies calculations. As the two sets of dependent variables are proportional to each other, using either set will give us the same information about the progress of the epidemic.

2.2 Important Assumptions

The next step is making some important assumptions on the rates of changes of our dependent variables. Natural deaths, births, immigration and other similar factors are being ignored, hence no one is being *added* to the susceptible group. The only way an individual *leaves* the susceptible group is by becoming infected. The rate of change of the number of susceptible individuals, $S(t)$, over time depends on the existing number of susceptibles, the number of individuals currently infected, and the level of interaction between susceptibles and infected individuals. Specifically, we assume that each infected individual initiates a fixed number β of contacts per day that can potentially transmit the disease. However, not all of these contacts are made with susceptible individuals. Assuming a uniformly mixed population, the proportion of

these contacts involving susceptibles is denoted by $s(t)$. Therefore, on average, each infected individual gives rise to $\beta \cdot s(t)$ new daily infections. This formulation simplifies complex scenarios, such as instances where a single susceptible individual encounters multiple infected individuals within a single day, particularly where the susceptible population dramatically outnumbers the infected population.

A fixed fraction γ of the infected group will be recovered during any given day. For instance, assuming an average infection duration of three days, approximately one-third of the presently infected population typically transitions to the recovered state daily. It is important to clarify that when referring to "infected," the term specifically denotes "infectious," indicating individuals capable of transmitting the disease to susceptible individuals. Conversely, individuals classified as "recovered" may still experience discomfort, and there remains a risk of complications such as pneumonia-related fatalities at a later stage. These assumptions inform the derivatives of our dependent variables.

$$\frac{\partial s}{\partial t} = -\beta s(t) i(t) \tag{1}$$

$$\frac{\partial r}{\partial t} = \gamma i(t) \tag{2}$$

As the sum of *susceptable, infected and recovered* people gives the whole population, it means that

$$\frac{\partial s}{\partial t} + \frac{\partial i}{\partial t} + \frac{\partial r}{\partial t} = 0 \tag{3}$$

To get the differential equation for *infecteds* it is enough to plug equations (1) and (2) into equation (3). The result will be:

$$\frac{\partial i}{\partial t} = \beta s(t) i(t) - \gamma i(t) \tag{4}$$

Combining all three equations will give the following system of differential equations:

$$\begin{cases} \frac{\partial s}{\partial t} = -\beta s(t) i(t) \\ \frac{\partial i}{\partial t} = \beta s(t) i(t) - \gamma i(t) \\ \frac{\partial r}{\partial t} = \gamma i(t) \end{cases}$$

3 Analytical Solutions

In this chapter we will analytically solve the standard SIR model with the help of the paper by Tiberiu Harko et al. [2]. Our goal at first is to get a second degree differential equation that is equivalent to the given system of differential equation, with changed variables for a better readability.

$$\begin{cases} x' = -\beta xy & \text{(1a)} \\ y' = \beta xy - \gamma y & \text{(1b)} \\ z' = \gamma z & \text{(1c)} \\ x + y + z = N & \text{(1d)} \\ x_0 = N_1, y_0 = N_2, z_0 = N_3 & \text{(1e)} \end{cases}$$

We begin by differentiating (1a) with respect to time t, and substituting y with $y = \frac{-x'}{\beta x}$ from the same equation

$$\begin{aligned} x'' = -\beta(x'y + xy') = -\beta\left(x'\left(\frac{-x'}{\beta x}\right) + xy'\right) &\implies \frac{-x''}{\beta} = \frac{-(x')^2}{\beta x} + xy' \\ &\implies xy' = \frac{-x''}{\beta} + \frac{-(x')^2}{\beta x} \\ &\implies y' = \frac{-1}{\beta}\left(\frac{-x''}{x} - \left(\frac{x'}{x}\right)^2\right) \end{aligned} \tag{2}$$

Next we insert $y = \frac{-x'}{\beta x}$ from (1a) into (1b) and equate it to (2)

$$\begin{aligned} -x' + \frac{\gamma x'}{\beta x} = \frac{-1}{\beta}\left(\frac{-x''}{x} - \left(\frac{x'}{x}\right)^2\right) &\implies \beta x' - \frac{\gamma x'}{x} = \frac{x''}{x} - \left(\frac{x'}{x}\right)^2 \\ &\implies \frac{x''}{x} - \left(\frac{x'}{x}\right)^2 + \frac{\gamma x'}{x} - \beta x' = 0 \end{aligned} \tag{3}$$

We also have that from (1a) $y = \frac{-x'}{\beta x}$ and from (1c) $y = \frac{z'}{\gamma}$. Equating them we get

$$\frac{-x'}{\beta x} = \frac{z'}{\gamma} \implies z' = -\frac{\gamma}{\beta}\left(\frac{x'}{x}\right) \tag{4}$$

Next we integrate (4)

$$
\begin{aligned}
c_1 + z = -\frac{\gamma}{\beta}\int \frac{x'}{x}\,dt \implies c_1 + z &= -\frac{\gamma}{\beta}\left(\ln(x) + c_2\right) \\
\implies \ln(x) &= -\frac{\beta c_1}{\gamma} - \frac{\beta z}{\gamma} - c_2 \\
\implies x &= e^{-\frac{\beta z}{\gamma}} \cdot e^{-\frac{\beta c_1}{\gamma} - c_2} \\
\implies x &= x_0 e^{-\frac{\beta z}{\gamma}} \qquad \text{Where } x_0 \text{ is an integration constant}
\end{aligned} \tag{5}
$$

From (5) we get the relation

$$
x' = -\frac{x_0 \beta}{\gamma} z' e^{-\frac{\beta z}{\gamma}} \tag{6}
$$

Now if we differentiate (4) and preform some manipulation we get

$$
z'' = -\frac{\gamma}{\beta}\left(\frac{x''}{x} - \left(\frac{x'}{x}\right)^2\right) \tag{7}
$$

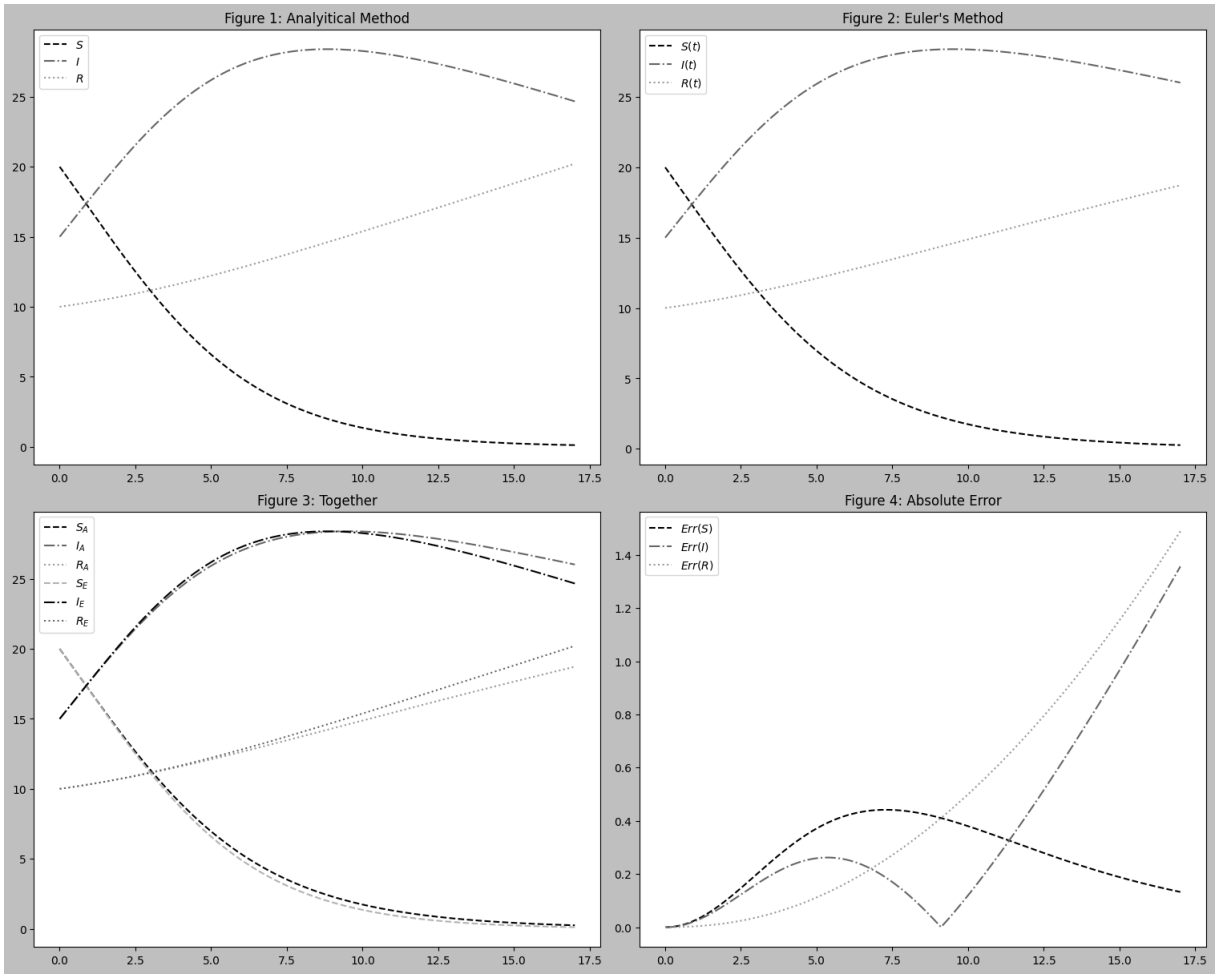
Finally if we combine (6), (7) and (4) into (3) we get

$$
z'' = x_0 \beta z' e^{-\frac{\beta z}{\gamma}} - \gamma z' \tag{8}
$$

Which is equivalent to the system of equation (1).

From this point we can solve the ODE by transforming (8) into a Bernoulli type differential equation and solving it by the given formula. Since this is out of scope for numerical methods class, the full solution can be found in appendix (B).

Figure 1: Graphical Examples of the Analytical Solution



Plotted Using Values $N_1 = 20, \quad N_2 = 15, \quad N_3 = 10, \quad \beta = 0.01, \quad \gamma = 0.02$

4 Numerical Methods for Solving the SIR Model

4.1 Euler's Method for the SIR Model

One of the most known ways of solving an ODE system, is Euler's method. This method is solving the IVP.

$$\begin{cases} y'(x) = f(x, y(x)), x \in [a, b] \\ y'(x_0) = y_0 \end{cases} \tag{1}$$

Here we are taking a mesh of x in the necessary range and calculating the $y(x)$ only at points of x_K After that, using the initial y_0, we approximate the next y, by using the formula

$$y_{k+1} = y_k + hf(x_k, y_k) \tag{2}$$

where $h = \frac{b-a}{n}$ is the stepsize of the x.

In our case, we are going to solve 3 such IVP's;

$$\begin{cases} \frac{\partial s}{\partial t} = -\beta s(t) i(t) \\ s(t_0) = s_0 \end{cases}$$

$$\begin{cases} \frac{\partial i}{\partial t} = \beta s(t) i(t) - \gamma i(t) \\ i(t_0) = i_0 \end{cases}$$

$$\begin{cases} \frac{\partial r}{\partial t} = \gamma i(t) \\ r(t_0) = r_0 \end{cases}$$

We are going to need the initial t_0 value, as well as the number of susceptible, infected, and recovered population at that t_0. After obtaining the sufficient pieces of information, we can use the Euler's formula and obtain

$$\begin{cases} s_{k+1} = s_k - \beta \cdot h \cdot s_k \cdot i_k \\ i_{k+1} = i_k + (\beta \cdot s_k \cdot i_k - \gamma \cdot i_k)h \\ r_{k+1} = r_k + \gamma \cdot h \cdot i_k \end{cases}$$

Thus we are going to be able to determine the needed values of $s(t)$, $i(t)$ and $r(t)$ at the necessary t_k date by doing the necessary amount of iterations.

4.2 Backward Euler's Method for the SIR Model

Just like forward Euler's Method described previously, backward Euler's method also aims to solve the IVP.

$$\begin{cases} y'(x) = f(x, y(x)), x \in [a, b] \\ y'(x_0) = y_0 \end{cases} \tag{1}$$

But this time, the Method uses a different approach, namely

$$y'(x) \approx \frac{y_n - y_{n+1}}{h} \tag{2}$$

By plugging the functions that interest us, we will get the system

$$\begin{cases} \frac{s_{t+1} - s_t}{h} = -\beta \cdot s_{t+1} \cdot i_{t+1} \\ \frac{i_{t+1} - i_t}{h} = \beta \cdot s_{t+1} \cdot i_{t+1} - \gamma i_{t+1} \\ \frac{r_{t+1} - r_t}{h} = \gamma i_{t+1} \end{cases} \tag{3}$$

From here, we have $\beta,\ \gamma,\ n,\ h,\ s_t,\ i_t,\ r_t$. Hence we are left with 3 equations and 3 unknowns. By using direct or numerical approaches, we can find the $s_{t+1}, i_{t+1}, r_{t+1}$ by solving this SLE

$$\begin{cases} s_{t+1} = s_t - h \cdot \beta s_{t+1} i_{t+1} \\ i_{t+1} = i_t + h \cdot (\beta s_{t+1} i_{t+1} - \gamma i_{t+1}) \\ r_{t+1} = r_t + h \cdot \gamma i_{t+1} \end{cases} \tag{4}$$

4.3 Approximation of the Recovery Rate

In order to make the SIR model work, we need to have methods of approximating the parameters β - the probability per unit of time that an infectious quarter will infect a noninfected quarter, and γ - the probability per unit of time that an infectious quarter will recover from the disease.

To calculate the γ, we start by using the formula (1), and supposing that I is a constant

$$\frac{\partial r}{\partial t} = \gamma I_0 \tag{1}$$

By integrating (1) we will obtain

$$r(t) = \gamma t I_0 \tag{2}$$

At t = T days, such as $R(T) = I_0$, we obtain

$$r = \gamma T I_0 \tag{3}$$

By getting the derivative of r with respect of t, we get

$$\frac{\partial r}{\partial t} = \gamma I(t) = \frac{r(t + \delta t) - r(t))}{\delta t} \tag{4}$$

We pick δt as 1, and get the formula for γ

$$\gamma \approx \frac{r(t+1) - r(t)}{I(t)} \tag{5}$$

5 Simulation Results on Real-World Data

In this chapter we will try to predict the epidemic trend based on the Covid-19 Data.

5.1 Simulation Results

Figure 2: USA Data

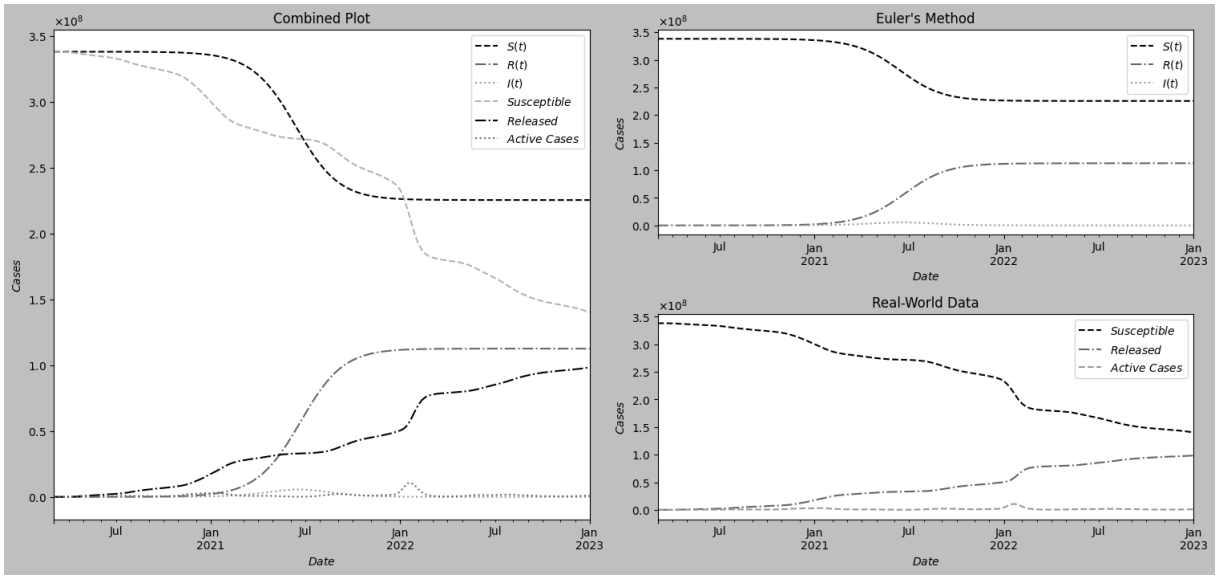
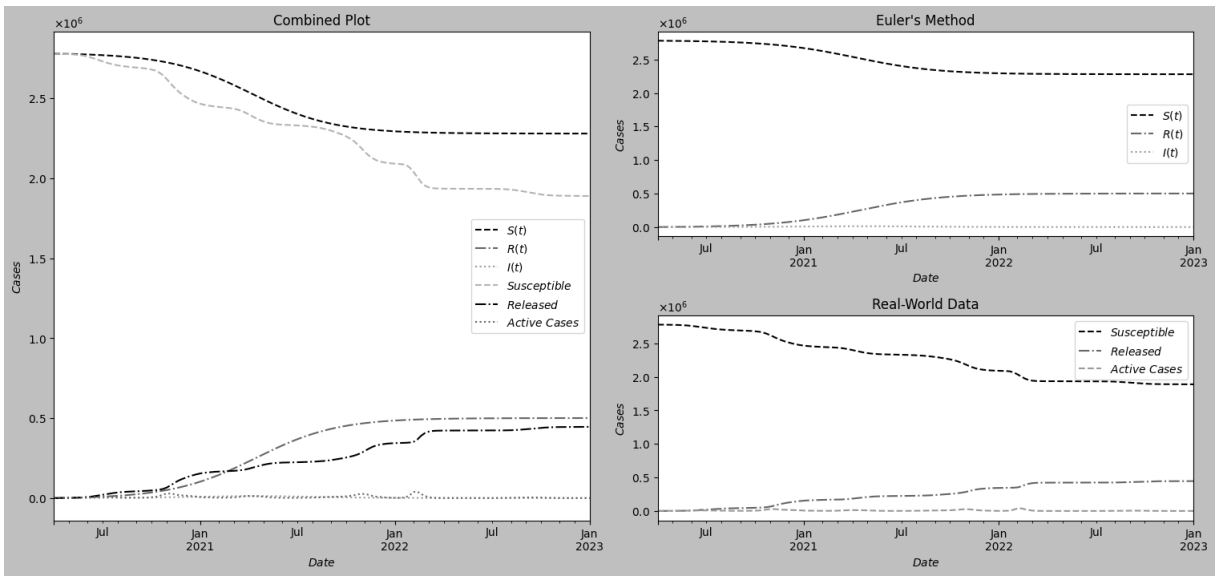


Figure 3: ARM Data



5.2 Calculation & RealWorld Use

In both figures is depicted the prediction we obtained by using half the data of the Covid cases in their respective countries. The difficulty of predictions stems from trying to correctly identify the optimal values for γ (Recovery Rate) and β (Infection Rate). Firstly we need to define what we mean by "optimal" values. In this case optimal was defined in a way such that the function

$$\sum_{t=0}^{n}(y(t) - y_t)^2$$

is minimal.

Here $y(t)$ is the real world data of susceptible, infected or recovered in time t. And y_t is the point in the same time t that we obtained by approximating the data with the SIR model. Hence we want values of γ and β such that

$$\sum_{t=0}^{n}(S(t) - S_t)^2 + \sum_{t=0}^{n}(I(t) - I_t)^2 + \sum_{t=0}^{n}(R(t) - R_t)^2$$

is minimal.

Calculating this function for γ and β is a very difficult task and even the approximations we provided are not optimal, so we went with the brute force approach.
We took half the data and calculated the points for the Euler's approximation with different values for γ and β. Lastly we calculated the error for each combination of γ and β and took the ones corresponding to the min of the error. What we obtained were not the most optimal values but they were still close.

In Figure 2 and 3 we can see that the SIR model can to some extent predict the progression of the disease, but since the real world data is complex and involves a lot of other factors, it is not ideal. Those factors can include the lockdown, social distancing, vaccines and a lot more. Hence fully predicting the progression of the diseases will not be possible, but we can still infer some useful information, such as when approximately the diseases will plateau and what development can we expect from it in general.

6 Conclusion

In conclusion, after thorough research and analysis of the SIR model, we were able to apply it to predict real-world data. Even though we demonstrated this with the simple SIR model, the same idea can be extended to more complex models that can predict epidemics more accurately. There are also better results available that use AI to make the model even more accurate.

Through solving the SIR system of ordinary differential equations analytically and then, numerically, we were able to compare the possible outcome given the parameters and initial values, and check if the SIR model provides a dependable source to rely on for epidemic clinical research. According to the graphs we obtained by comparing Euler's method, Backwards Euler's method, and the Analytical solution given the necessary initial data, we obtained almost identical results and applied one of the methods to the real data. to check whether the error between the real COVID-19 data and our prediction is acceptable.

The tests came in as a success, and even though there are still errors present between the approximation and the real data, we can confidently say that using the model with one of the given methods in the early stage of epidemics will provide us with insights into how big of a threat the disease is and how concerned the researchers should be, hence getting a more precise insight on the regulations of the precautionary behavior for the whole population. Our analysis also proves that the SIR model and its extensions are relevant in informing public health policies and interventions.

This analysis serves as proof of concept, meaning that more research into the SIR model will allow for better results. Even with the simple SIR model and using half the data from the COVID-19 epidemic, we were partially successful in predicting the future behavior of the disease.

A Appendix: Extensions and Modifications

The basic SIR model can be expanded and modified in various ways to incorporate factors such as death and birth rates, vaccination rates, reinfection rates, and other significant variables, leading to more accurate calculations and predictions. One famous modification was made by Yamamoto [4]. Let death rate be denoted by α and death rate be denoted by μ. The birth rate contributes to an increase in the size of the susceptible population, while the death rate leads to a decrease in its size. The extent of this change can be calculated as:

$$\frac{\partial s}{\partial t} = \alpha - \mu s(t) - \beta s(t) i(t)$$

Yamamoto made another modification of the SIR model. In this modification he took into consideration the inoculation rate of vaccine. Let ε be the rate at which individuals are vaccinated within a given population. Incorporating this parameter leads to an increase in the population within the recovered group, thus resulting in the following modifications to the model equations:

$$\frac{\partial s}{\partial t} = \alpha(1-\varepsilon) - \mu s(t) i(t)$$

$$\frac{\partial i}{\partial t} = \beta s(t) i(t) - (\gamma + \mu)(t)$$

$$\frac{\partial r}{\partial t} = \alpha\varepsilon + \gamma i(t) - \mu r(t)$$

Satoru Yamamoto examined the combined impact of both vaccination and the rate of reinfection as well. A small portion of the population might get reinfected at a reinfaction rate σ. The equation will look in the following way after making the modification:

$$\frac{\partial s}{\partial t} = \alpha(1-\varepsilon) - \mu s(t) i(t)$$

$$\frac{\partial i}{\partial t} = \beta[s(t) + \sigma r(t)] i(t) - (\gamma + \mu)(t)$$

$$\frac{\partial r}{\partial t} = \alpha\varepsilon + \gamma i(t) - \mu r(t) - \beta\sigma r(t) i(t)$$

There is an extension of the SIR model which gives more realistic result. The model is called SEIR, where “E” stands for “exposed”. The intensity of the virus depends

on the exposure of an individual to the symptomatic asymptomatic carrier of the virus. According to Peter Turchin [3] there is a need to add another group of exposed individuals to have a more effective model. That group stage would be between susceptible and infected stages. It will represent a group of individuals who have encountered the infection but are currently not infected. Let $e(t)$ be the fraction of the population which is exposed to infection but is not infected. As it was obtained above, for the standard SIR model the rate of change for the susceptible group is:

$$\frac{\partial s}{\partial t} = -\beta s(t) i(t)$$

Let δ be the coefficient which shows the possibility of an exposed person contracting the infection. Then the rate of change of the exposed group can be calculated as:

$$\frac{\partial e}{\partial t} = \beta s(t) i(t) - \delta e(t)$$

Since there is a possibility for exposed individuals to contract the infection, the influence of *susceptibles* on the rate of infected persons is positive. Additionally, considering the possibility for infected individuals to recover, the impact of *recovery* on the rate of infected persons is negative. Thus:

$$\frac{\partial i}{\partial t} = \delta e(t) - \gamma i(t)$$

Similarly, the rate of change of the recovered group is:

$$\frac{\partial r}{\partial t} = \gamma i(t)$$

Combining all the equations above will ge the following system of differential equations:

$$\begin{cases} \frac{\partial s}{\partial t} = -\beta s(t) i(t) \\ \frac{\partial e}{\partial t} = \beta s(t) i(t) - \delta e(t) \\ \frac{\partial i}{\partial t} = \delta e(t) - \gamma i(t) \\ \frac{\partial r}{\partial t} = \gamma i(t) \end{cases}$$

B Appendix: Continuation of Analytical Solutions

Recall that by transforming the ODE we obtained the form

$$z'' = x_0 \beta z' e^{-\frac{\beta z}{\gamma}} - \gamma z' \tag{1}$$

From here we need to introduce a new function

$$u = e^{-\frac{\beta}{\gamma} z}, \quad u_0 = e^{-\frac{\beta}{\gamma} N_3} \tag{2}$$

Also note that we got the initial value by substituting z with $z(0)$.
Next we substitute (2) into (1) and by doing a lot of simplification we get

$$u(u)'' - (u')^2 + (\gamma - x_0 \beta u) u(u)' = 0 \tag{3}$$

Now we introduce another function $\phi = t'_u$, and by substituting it back we get

$$\phi'_u + \phi \frac{1}{u} = (\gamma - x_0 \beta u) \phi^2 \tag{4}$$

which is a Bernoulli type differential equation with the solution given as

$$\phi = \frac{1}{u(C_1 - \gamma \ln u + x_0 \beta u)} \tag{5}$$

where C_1 is an integration constant. Since $\phi = t'_u$ by integrating (5) we get an equation for time

$$t = \int_{u_0}^{u} \frac{du}{u(C_1 - \gamma \ln u + x_0 \beta u)} \tag{6}$$

hence with all of this we can get the parametric equation for the simple SIR model with u taken as parameter

$$\begin{aligned} x &= x_0 u, \\ y &= \frac{\gamma}{\beta} \ln u - x_0 u - \frac{C_1}{\beta}, \\ z &= -\frac{\gamma}{\beta} \ln u \end{aligned}$$

and by adding them we can obtain than $C_1 = -\beta N$.

References

[1] W. Kermack and A. McKendrick. "Mathematical Theory of Epidemics". In: (1927). URL: https://doi.org/10.1098/rspa.1927.0118.

[2] M. K. Mak Tiberiu Harko Francisco S. N. Lobo. "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates". In: (2014). URL: https://doi.org/10.48550/arXiv.1403.2160.

[3] Peter Turchin. "Comples Population Dynamics". In: (2003). URL: https://polydora.github.io/General-ecology/Literature//Turchin%202003%20Complex%20Population%20Dynamics_%20A%20Theoretical_Empirical%20Synthesis.pdf.

[4] Satoru Yamamoto. "Introduction to Mathematical Modeling and Computation". In: (2018). URL: https://www.researchgate.net/publication/295290766_Introduction_to_Mathematical_Modeling_and_Computation.