The SIR Models, their Applications, and Approximations of their Rates

Christopher Cano

May 6, 2020

Project Advisor: Dr. Stacey McAdams

Abstract

The SIR (Susceptible-Infected-Recovered) models are used to help predict the spread of diseases. The goals of this paper are: elaborating on the methods of approximating the recovery rate, infection rate, and loss of immunity rate, comparing the SIR models with these approximation methods to real world data, and determining the most accurate combination of the approximation methods for each SIR model. There are several SIR models such as the Kermack-McKendrick, SIRS, and SI models that are designed for specific diseases. Understanding the parameters of these models will assist us in maximizing their accuracy. For example, there is no explicit formula for the any of the rates within the models. Therefore, those rates must be approximated. Using these models to represent real world situations will explain why each disease needs to be represented by a specific model. Understanding the content and the rate approximations of each model can help to determine the level of accuracy the model will have in predicting the spread of the disease.

Keywords: SIR, Kermack-McKendrick, SIRS, SI, Susceptible, Infected, Recovered

1 Introduction

The SIR (Susceptible-Infected-Recovered) models are intricate models that were designed to help predict how diseases will spread within a community based upon several parameters. Understanding where and why these different parameters are used within the models will help to understand what disease specific models are used for. For example, we must first understand the origin of the entire model itself: "The SIR (susceptible-infected-removed)

model, developed by Ronald Ross William Hamer, and others in the early twentieth century, consists of a system of three coupled non-linear ordinary differential equations, which does not possess an explicit formula solution." [10] It is important to understand that the SIR model is simply a class of models that predict the spread of diseases. The SIR models consists of the variables S, I, and R with each being a function of time. S, I, and R represent the number of people susceptible to disease, infected by the disease, and recovered from the disease, respectively.

2 Background Literature and Related Studies

2.1 The SIR Models

The SIR models are each applicable to specific situations. Throughout this paper we will be examining three different types of SIR models: the Kermack-McKendrick model, the SIRS model, and the SI model. Before we continue into the models, it must be understood that the SIR models that we will examine in this paper have several assumptions about the population. The first assumption is that the population is a fixed population. Therefore, it is assumed that no one is born or dies during the duration of the model. The second assumption is that each disease has an incubation period of zero. This means that as soon as a person is infected with the disease, they have the chance to spread it to others. The third assumption is that the time of being infected is equal to the length of the disease. This assumption is self-explanatory as it simply says that having the disease means that person is in the infected category. The last assumption is that of a homogeneous population. Therefore, each individual is assumed to have the same health conditions, age, social status, and so forth. [11] Now that there is a general understanding of the SIR models, we can begin to look at each model individually.

2.2 Kermack-McKendrick Model

One of the most common of the SIR models is the Kermack-McKendrick Model:

$$\frac{dS}{dt} = -\beta SI \tag{1}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

In (1)-(3) we have: β (the infection rate), γ (the recovery rate), $S(t)$ (the percentage of the population that is susceptible to the disease), $I(t)$ (the percentage of the population that is infected with the disease), and $R(t)$ (the percentage of the population that recovered and developed an immunity to the disease). In all SIR models, β is equal to the probability of an infected person giving the disease to a susceptible person during contact and γ is approximately $1/D$ with D being the average duration of the disease. [6] According to Wolfram MathWorld, "The Kermack-McKendrick model is an SIR model for the number of people infected with a contagious illness in a closed population over time. It was proposed to explain the rapid rise and fall in the number of infected patients observed in epidemics such as the plague (London 1665-1666, Bombay 1906) and cholera (London 1865)." [11]. Diseases represented by this model are diseases that people obtain a lifetime immunity. A good example of a disease that would fit in this model is measles. By combining the rate of infection and the rate of recovery, we obtain a property of the model known as the epidemiological threshold:

$$R_0 = \frac{\beta S(t)}{\gamma} \tag{4}$$

The epidemiological threshold predicts whether or not a disease will become an epidemic. When $R_0 < 1$, each infected person will at most infect one other person before dying or recovering. However, if $R_0 > 1$, each infected person will infect 2 or more people causing an exponential rise in the number of people infected. [11] Keep in mind that the values in our variables are between 0 and 1. This is required for the SIR model due to the fact that S, I, and R each resemble a percentage of the total population. Thus, we have $S(t) + I(t) + R(t) = 1$. The infection and recovery rates are being applied to a specific percentage of the population at a given time. This is the same reasoning on why β , $\gamma < 1$. Further within this paper, we will see how β and γ are determined. Determining the epidemiological threshold is one of the key components in a decision such as if quarantine zones need to be created. Note that this decision would also rely on the lethality of the disease, which lies outside the realm of the SIR models. It is also important to consider the possibility of becoming reinfected with the disease. This concept will lead us into the SIRS (Susceptible, Infected, Recovered, Susceptible) model.

2.3 The SIRS Model

The SIRS model is extremely similar to the Kermack-McKendrick model with the addition of ξ, the rate at which recovered individuals return back to the susceptible category. This is important aspect that needs to be taken into consideration when choosing which model to use "We assume that recovery implies permanent or temporary acquired immunity; in the latter case, there is a return of the removed individuals into the susceptibles compartment." [3] Below we can see the set up of the SIRS model.

$$\frac{dS}{dt} = -\beta SI + \xi R \tag{5}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{6}$$

$$\frac{dR}{dt} = \gamma I - \xi R \tag{7}$$

An example of a disease that would require the SIRS model is influenza. The SIRS model is the most complicated of the SIR models as it contains three rates to consider. There is also the SI model which contains only the infection rate.

2.4 The SI Model

The simplest of the SIR models is the SI model. This model contains only the susceptible and the infected ordinary differential equations. Thus, there is no possibility of moving to the recovered category.

$$\frac{dS}{dt} = -\beta SI \tag{8}$$

$$\frac{dI}{dt} = \beta SI \tag{9}$$

One disease that would fit in this model would be HIV since there is currently no cure. Since there is no recovered category this implies, "Once a susceptible entity becomes infected, he or she gets added into the infected set, thereby increasing the size of the infected set and ultimately decreasing the size of the susceptible set of individuals." [5] With this in mind we obtain the equations, $S(t) = 1 - I(t)$ and $I(t) = 1 - S(t)$. In other words, (8) and (9) reflect each other across the line $y = 0.5$. Even though this is the most simple of the models, the approximation of the infection rate still plays a vital role in the model's overall accuracy. Therefore, we must determine how to accurately approximate this rate along with the rates in the other SIR models.

3 Methods

3.1 Approximating Rates in the SIR Models

One of the main difficulties of the SIR models is approximating their rates. The recovery rate, infection rate, and the loss of immunity rate all need to be approximated. We must accurately determine these rates in order for these models to be useful.

3.2 The Recovery Rate and Infection Rate

The rates for the SIR models do not have a standard approximation method. Therefore, we must determine an appropriate approach to approximate these rates. The method used to determine the recovery and infection rates involves Microsoft Excel. First, there is no explicit solution to the SIR models; therefore, the ordinary differential equations are solved separately and not as a system. Second, the solutions of the ordinary differential equations in the models are entered into Excel and programmed to call on one another. Next, the observed data of a disease is placed into Excel. Then, we use Excel to minimize the actual error between the observed values (the observed data) and the expected values (the Infected category).

$$Actual\ Error = (observed - expected)^2 \tag{10}$$

Then, we obtained a Chi Squared (χ^2) statistic by comparing the observed data and the Infected category. It was developed by Karl Pearson and is also referred to as a goodness-to-fit test. [8]

$$\chi^2 = \frac{(observed - expected)^2}{expected} \tag{11}$$

This fact will help us understand why this test is being used. Lastly, we use Excel to determine the P-value based on our Chi Squared statistic. Our P-value is the probability of our null hypothesis not being rejected. [1] The P-value is compared to the α value which is used to determine our confidence percentage of rejecting the null hypothesis ($100 - \alpha$ = confidence percentage). The null hypothesis of a Chi Squared test is that there is no significant difference between our observed and expected values. If the P-value is less than the α value, we reject the null hypothesis with a certain percentage of confidence. However, since our actual errors are minimized, we obtain extremely low Chi Squared statistics for our models. Therefore, in both of our graphs that use real data, we obtain P-values of 1 (i.e. 100 percent probability of our null hypothesis not being rejected). Thus, no matter the value of α, it will never be larger than our P-value. Hence, we have a zero percent confidence to reject our null hypothesis. Ultimately, our P-values reinforce our confidence in the accuracy of our models. Since the rates are optimized, there is no need for a regression line. We can see this method implemented in the data section of this paper.

3.3 The Loss of Immunity Rate

Since there is not much data on diseases that fit in the SIRS model, we will simply review a method of approximating the loss of immunity rate.

$$D(t) = (1 - (1 - \beta)^{\rho(k)I(t)}) \tag{12}$$

In this model we have: $D(t)$ (the loss of immunity rate with respect to time), β (the infection rate), ρ (population density), k (average number of people withing transmitting distance), and $I(t)$ (the percent of population infected at a given time). [12] The SIRS model will use this method along with the method described in the previous section for approximating all three of its rates. It is vital to keep in mind that every disease has different properties. Therefore, the properties of their vaccines would also be different, "The prevention of infection correlates with the induction of specific antibodies. Although loss of antibody after vaccination may render vaccines again susceptible to some infections." [7] The time it takes for vaccinations to become obsolete is directly related to the loss of immunity rate in a vaccinated population. However, the approximation methods are only useful if they are accurate. Thus, the models and their approximation rates must be compared to real world data.

4 Models and Results

4.1 Approximation Methods vs. Real Data

This section contains numerical results generated in MATLAB. These graphs have real world data and the respective SIR model with approximated infection and recovery rates. We will be using the approximation method described in the Methods section. This involves Excel minimizing the actual error in order to fit the Infected category to the observed data.

4.2 The Kermack McKendrick and SI Models

The first graph we will look at is HIV in South Africa. Since HIV currently does not have a cure, we will use the SI model.

The SI Model and HIV

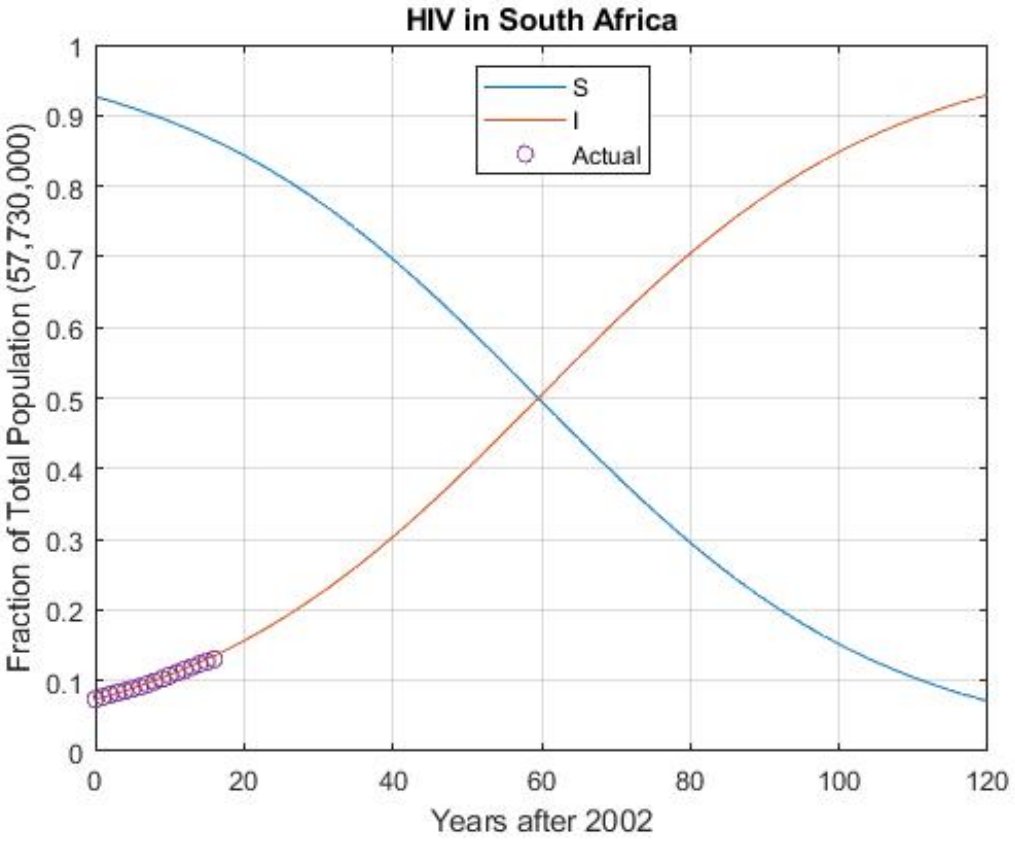

Figure 1: This graph contains data of HIV in South Africa from 2002 to 2018.

In this graph, we have an infection rate of 0.04259. Our total population is $57,730,000$. The initial number of susceptible and infected is $53,480,000$ and $4,250,000$ respectively. [2] Our actual error is $8.4615 * 10^{-6}$ with a χ^2 statistic of $8.76271 * 10^{-5}$.

The next graph we will see is COVID-19 in Rome. It is important to note the two following additional assumptions. The first assumption is people that recover from this disease obtain a lifetime immunity. This was assumed since we are currently unsure if people can become reinfected with this disease. [4] Therefore, we will use the Kermack-McKendrick model. The second assumption is that Rome has Italy's number of infected. This was assumed because the percentage of infected people was very small when using Italy's actual population. Small percentages of infected cause the models to be inaccurate. Therefore, we used Rome's population with Italy's number of infected to increase the percentage of the population that is infected.

The Kermack-McKendrick Model and COVID-19

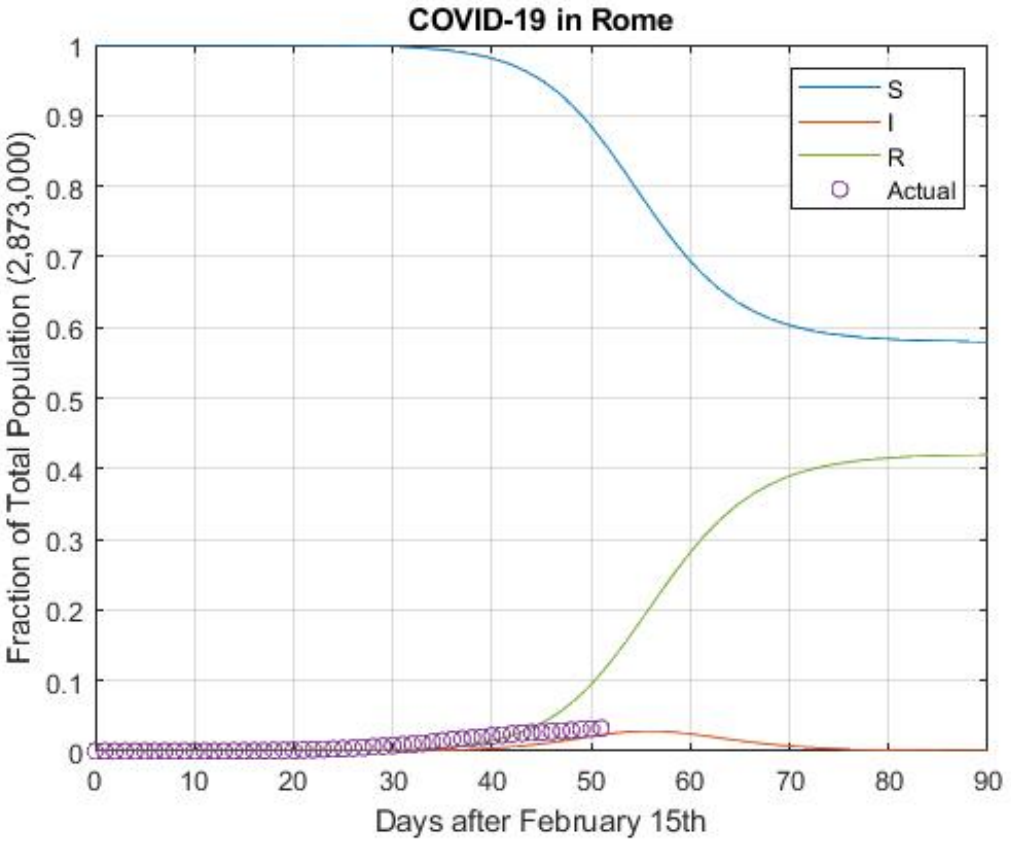

Figure 2: This graph contains data of COVID-19 in Italy from February 15th to April 6th. Note: the population of Rome is used instead of the population of Italy. This is why the graph is titled "COVID-19 in Rome".

In this graph, we have an infection rate of 0.91776 and a recovery rate of 0.70681. Our total population is $2,873,000$. The initial number of susceptible, infected, and recovered is $2,872,997$, 3, and 0 respectively. [9] Our actual error is 0.00399 with a χ^2 statistic of 3.03809.

5 Discussion

In our first graph from the Data and Models section, we have data of HIV in South Africa from 2002 to 2018. The SI model allows us to determine the future spread of the disease based on the observed data. Based on the model, 60 years after 2002, 50% of South Africa's population will have HIV. Then 120 years after 2002, around 93% of the population will be infected with HIV. Note that the SI model has no recovery rate and therefore cannot be evaluated with the epidemiological threshold.

Our second graph from the Data and Models section contains data on COVID-19 in Rome from February 15th to April 6th. Keep in mind that the number of infected is Italy's actual numbers and not Rome's. The assumption of Rome having Italy's active case count was done in order to make up for the inaccuracy of the model due to small percentages. With this in mind, this graph ultimately represents the hypothetical future spread of COVID-19 in Rome if it had Italy's active case count. According to our graph, the epidemiological threshold would be approximately 1.02343 on day 55 and 0.98002 on day 56. Thus, the disease would no longer be considered an epidemic after 55 days. Notice that this also reflects when the number of infected is at its maximum. Then 90 days after February 15th the number of infected would be approximately zero. At this time, around 42% of the population would have recovered from the disease with approximately 58% of the population never becoming infected.

6 Conclusion

The SIR model is a class of models that predict how a communicable disease will spread through a community based on several parameters. Each SIR model is designed for a specific type of disease. The SI, Kermack-McKendrick, and SIRS model is designed for diseases with no recovery, diseases with lifetime immunity after recovery, and diseases that have a chance to reinfect after the victim recovers, respectively. We can use our rate approximation method in combination with observed data to determine how a disease will continue to spread. This aspect is useful for health agencies to determine how fast a disease will spread and what precautions should be taken.

6.1 Goals of the Paper

Recall that the goals of the paper are: elaborating on the methods of approximating the recovery rate, infection rate, and loss of immunity rate, comparing the SIR models with these approximation methods to real world data, and determining the most accurate combination of the approximation methods for each SIR model. The only rate that we did not approximate was the loss of immunity rate. This was due to the focus on the Kermack-McKendrick and the SI models. However, we were not able to determine the most accurate combination of approximation methods since we only used one. This was not the original plan. We changed the plan to using only the approximation method of the paper since it proved to be substantially more accurate than the other methods. All of the other goals listed above were accomplished.

6.2 The Continuation of Research

Here are some potential areas to continue our research. The first area is to determine if there is another approximation method for the rates that makes these models accurate even at small percentages of infected. Implementing this method into another software such as MATLAB may prove to be a good course of action. MATLAB could potentially work more accurately with ordinary differential equations than Excel can. The second area is to test this paper's rate approximation method on the other SIR models that were not discussed in this paper. Some of these models are the SIS model, the SEIR model, and the SEIRS model. This method could potentially work on the rates of those models as well as it did for the SI and Kermack-McKendrick models. Continuing the search of data for the SIRS model is also another research area that could be explored. If suitable data for this model is found, then the approximation method of the loss of immunity rate in Section 3.3 could also be evaluated. These are the areas of research that we recommend.

References

[1] Stephanie Glen. P-value in statistical hypothesis tests: What is it?, Sep 2018.

[2] ISIbalo House, Koch Street, et al. Mid-year population estimates, 2017.

[3] Andrei Korobeinikov and Graeme C Wake. Lyapunov functions and global stability for sir, sirs, and sis epidemiological models. *Applied Mathematics Letters*, 15(8):955–960, 2002.

[4] Hillary Leung. What we know about coronavirus immunity and reinfection, Apr 2020.

[5] Jyoti Sunil More and Chelpa Lingam. A si model for social media influencer maximization. *Applied Computing and Informatics*, 15(2):102–108, 2019.

[6] Sinead E Morris. shinysir: Interactive plotting for infectious disease models.

[7] Raúl Peralta, Cruz Vargas-De-León, and Pedro Miramontes. Global stability results in a svir epidemic model with immunity loss rate depending on the vaccine-age. In *Abstract and Applied Analysis*, volume 2015. Hindawi, 2015.

[8] Rakesh Rana, Richa Singhal, et al. Chi-square test and its application in hypothesis testing. *Journal of the Practice of Cardiovascular Sciences*, 1(1):69, 2015.

[9] Umberto Rosini. Italy, Feb 2020.

[10] Howard Howie Weiss. The sir model and the foundations of public health. *Materials matematics*, pages 0001–17, 2013.

[11] Eric W Weisstein. Kermack-mckendrick model, Sep 2018.

[12] Haifeng Zhao, Jie Jiang, Rongyu Xu, and Yang Ye. Sirs model of passengers' panic propagation under self-organization circumstance in the subway emergency. *Mathematical Problems in Engineering*, 2014, 2014.