Modelling the epidemic and disaster known as Covid 19

Abstract

We study the outbreak of the disease known as Covid 19. We derive a model which takes into account 4 groups within the population including the fatality rate and the rate of an epidemic. We derive a method to determine the rate of the epidemic in order to provide an estimate on the fatalities of the disaster. We then use data given to us by the governments of a few countries to calibrate this model.

1 Introduction

The epidemic event caused by the disease known as Covid 19 has been called a one in a century occurrence. Quick and efficient measures had to been taken in order to prevent the epidemic from evolving due to its high infection rate. To be able to counteract the fast-spreading disease mathematicians and epidemiologists were tasked with deriving models for the epidemic in order to be able to predict how well the solutions will work. In this project we aim to model the epidemic, before any major actions were taken to prevent its evolution.

2 Simple SIR Model

We begin with the simplest epidemiological model. We can split the population into 4 groups: ***susceptible***, S_i: people who have not been infected yet on day i; **infectious**, I_i: total number of people who are ill on day i; ***recovered***, R_i, and ***fatalities***, F_i: those who recovered or those who haven't.

The way in which the disease is spread begins with people who are infected interact with other people. The probability that a person can become infected is dependant on the number of people they come into contact with and the conditions of the way they interact. Furthermore, the number of people who get infected on a particular day is proportional to the number of susceptible persons and the probability that someone is infectious. This means that as the number of people who are susceptible decreases, the number of people exposed to the disease increases.

This is represented mathematically below:

$$\begin{aligned} \Delta I_i &= K S_i I_i / N \\ S_{i+1} &= S_i - \Delta I_i \\ I_{i+1} &= I_i + \Delta I_i, \end{aligned} \tag{1}$$

where ΔI_i represents the number of people who become infected on day i.

Let d be the days of infection. After some amount of d the infected people will either recover or die, so we can write:

$$\begin{aligned} R_{i+1} &= R_i + (1 - K_f)\Delta I_{i-d} \\ F_{i+1} &= F_i + K_f \Delta I_{i-d} \\ I_{i+1} &= I_i - \Delta I_{i-d}, \end{aligned} \tag{2}$$

where K_f is the fatality rate.

Combining (1) and (2), we get:

$$\begin{aligned} S_{i+1} &= S_i - \Delta I_i \\ I_{i+1} &= I_i + \Delta I_i - \Delta I_{i-d} \\ R_{i+1} &= R_i + (1 - K_f)\Delta I_{i-d} \\ F_{i+1} &= F_i + K_f \Delta I_{i-d}. \end{aligned} \tag{3}$$

Q1

To ensure our model is accurate at this point, we must perform a test to verify that the total population $S_i + I_i + R_i + F_i$ does not vary over time as we know that all individuals have to be accounted for all the time. This is done below:

$$\begin{aligned} S_{i+1} + I_{i+1} &= S_i - \Delta I_i + I_i + \Delta I_i - \Delta I_{i-d} \\ S_{i+1} + I_{i+1} + R_{i+1} &= S_i + I_i - \Delta I_{i-d} + R_i + \Delta I_{i-d} - K_f \Delta I_{i-d} \\ S_{i+1} + I_{i+1} + R_{i+1} + F_{i+1} &= S_i + I_i + R_i + F_i - K_f \Delta I_{i-d} + K_f \Delta I_{i-d} \\ S_{i+1} + I_{i+1} + R_{i+1} + F_{i+1} &= S_i + I_i + R_i + F_i, \end{aligned} \tag{4}$$

which clearly verifies that as i varies the total does not change.

In order to start solving this model we have to look at where $i = 0$ to initialise the iteration. If we assume the total population for the UK is 66 million then S_0 would be the total population as nobody is infected yet. To start the epidemic one person will become infected, so we can let $I_0 = 1$. Obviously then $R_0 = F_0 = 0$.

Epidemiologists give the parameter R to the average number of people who get contaminated by a single infected person. So the relation between K (from (1)) and R is given by:

$$K = \frac{R}{d} \tag{5}$$

The iteration can be performed in Python by using the equations from (3) to evolve the population every i.

Q2

Using Python to iterate this model with different values for R, we can see the results in (1).

Here we can see that increasing the value of R in the model obviously increases the number of infected people per day and so the number of fatalities and recoveries increases with it. However, as we increase the value of R to be more than 1 we can clearly see that there is a sharp increase then decrease in the infected. This is because the more people that get infected by one person, the more people who will recover/die faster.

Clearly, this model is very simple so it does not come as a surprise that it does not take into account some key information. For example the ease spreading infection to other can evolve with time. In addition to this people do not recover or die after an exact set of days as it also varies between each person. In order to deal with this, Epidemiologists have found probability distributions which characterises both of these.

3 Covid 19 SIR Model

In this more general model of infection there are 4 stages: A person who is not yet infected is described as *susceptible*. If this person is then exposed to an infected person,

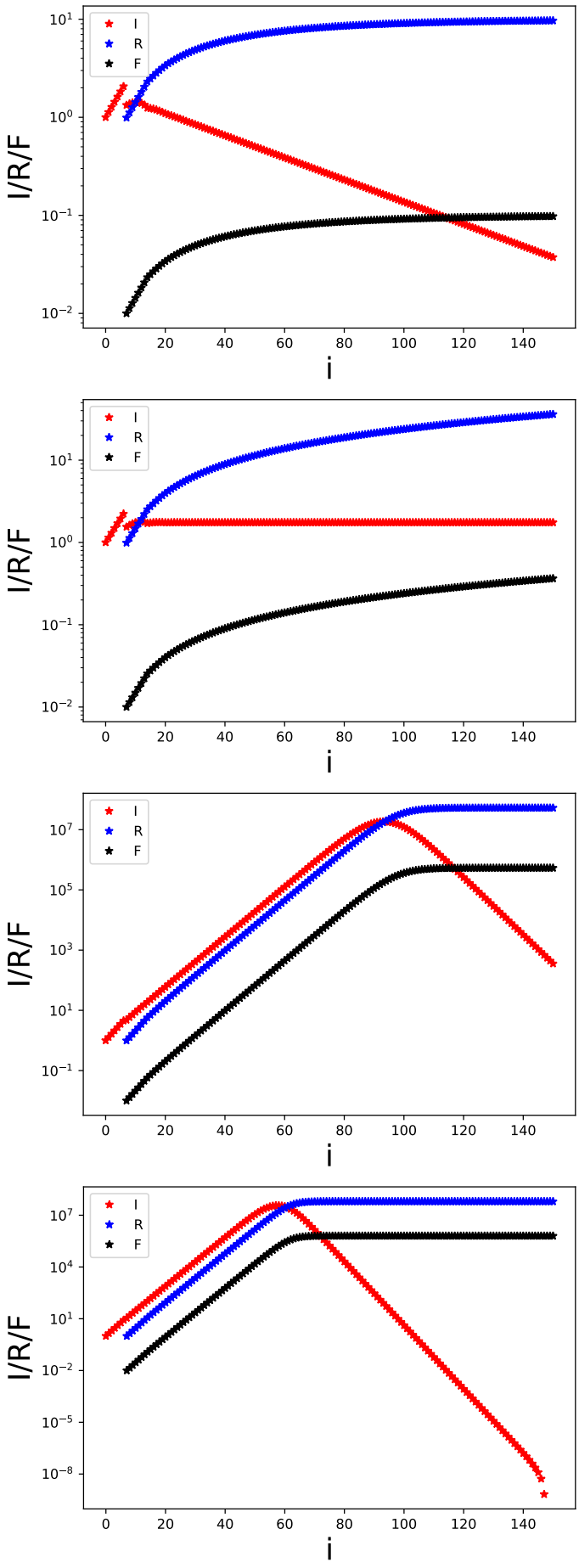

Figure 1: Infected vs Recovered vs Fatalities: a) R=0.9 b) R=1 c) R = 2 d) R=3

they are described as *exposed* as they can contract the disease. After a few days of being exposed they will contract it and become *infectious* so that they can now contaminate others. Finally they either recover or not (*recovered* or *fatalities*).

However, the problem with Covid 19 is that people can be asymptomatic (ie not show any symptoms) even if they are exposed or infected. The only easy way to tell if someone is infected is to test them, however this will only work a few days after the infection has happened so it could have spread to others in that period before knowing they were infectious. In addition to this people who have been infected stop being infectious before they recover meaning there may need to be an extra state to cover these people. As we do not really know when people become infectious once infected, we will consider the time when people become infected.

Now the probability that a person is infected on day d is proportional to the probability to meet someone who was infected on day $d-1$ and the probability, P_inf that a person is infection after 1 day. This is then added to the same thing but on day $d-2$. Mathematically this is:

$$\Delta I_i = RS_i \sum_{d=1}^{n_i} \frac{\Delta I_{i-d}}{N} P_{inf}(d), \tag{6}$$

where n_i is the number of days that $P_{inf}(d) = 0$ if $d > n_i$. In [3] it was shown that the distribution of recovery and death were similar [4], so we can then use the same distribution for them and call it P_r. This gives us,

$$\begin{aligned} \Delta R_i &= (1-K_f) \sum_{d=1}^{n_r} \Delta I_{i-d} P_r(d), \\ \Delta F_i &= K_f \sum_{d=1}^{n_r} \Delta I_{i-d} P_r(d). \end{aligned} \tag{7}$$

Here n_r is the number of days such that $P_r(d) = 0$ if $d > n_r$. We can insert (6) and (7) into (3) to get a more realistic model:

$$\begin{aligned} S_{i+1} &= S_i - \Delta I_i \\ I_{i+1} &= I_i + \Delta I_i - \Delta R_i - \Delta F_i \\ R_{i+1} &= R_i + \Delta R_i \\ F_{i+1} &= F_i + \Delta F_i. \end{aligned} \tag{8}$$

Now we would like to find expressions for the probabilities P_{inf} and P_R. In [1] the we were told that:

- The probability distribution of infecting others during the days after the infection is $P_{inf} = Gamma(6.5, 0.62)$.
- The probability distribution of incubation (showing first symptoms), is $P_{inc} = Gamma(5.1, 0.86)$ [2] [1].
- The probability distribution of recovering after *incubation* is $P_R = Gamma(18.8, 0.45)$

Gamma distribution is a lump-like distribution, where the first argument is the average value and the second describes how spread out it is.

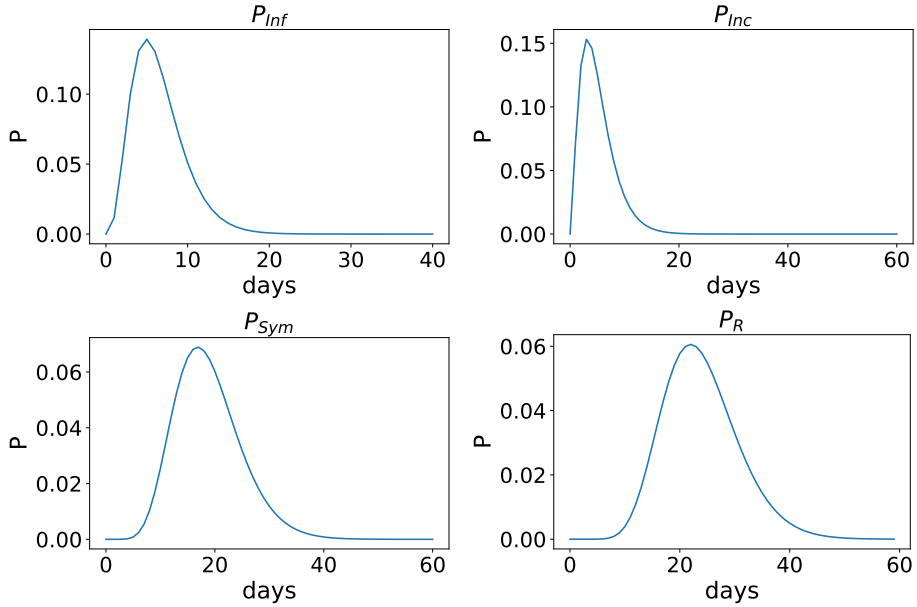

Figure 2: Probability distributions: a) Infectiousness after infection b) Showing symptoms after infection c) Recovery from symptoms d) Recovery from infection after incubation

The probability distribution of recovering after *infection* is given by:

$$P_r(d) = \sum_{n=0}^{d} P_{inc}(n) P_R(d-n) \tag{9}$$

This sum can be read as: the probability that an individual recovers 3 days after being infected, is the probability that they incubated for 3 days times the probability that they immediately recovered after the incubation, plus the probability that they incubated for 2 days and recovered 1 day later and so on.

The distributions for these probabilities are shown in (2).

Q3

If we assume that $P_{inc}(d)$ and $P_R(d)$ are 0 outside the domain $d \in [0,N]$ for some N we can show that $P_r(d)$ is 0 for $d < 0$ as according to (9) the sum cannot happen as it cannot sum to -1. It also uses the Gamma distribution which is 0 when an argument is less than or equal to 0.

In order to show that this is probability distribution, If we take $d = 2N$ then from (9) we know that the only non-zero value for the sum would be for n=N as we know that when n is not in [0,N], P_{inc} and $P_R = 0$. So $P_r(2N) = P_{inc}(N)P_R(d-N)$ is non-zero. We know this because between 0 and $N-1$ $P_R(d-n) = 0$ and it is larger than N and this is the same between $N+1$ and $2N$ as P_{inc} is 0. So this leaves $n = N$. Furthermore, if we take $d = 2N+1$ then it is 0 everywhere making the largest is $2N$.

While the infection rate R parameter is used by many epidemiologists, there is another parameter we can use that is easy to determine from the data. It is the time needed for the number of cases to double. If we say that the number of fatalities is in the of the form $F = Aexp(t)$ then the doubling time τ is defined as the value for which:

$$Aexp(\lambda(t+\tau)) = 2Aexp(\lambda t). \tag{10}$$

Cancelling the common terms leaves us with $exp(\lambda\tau) = 2$ and so

$$\tau = \frac{log(2)}{\lambda}. \tag{11}$$

We this parameter, we are now able to solve the model equations and compare them with real-life data. Now our issue lies in choosing the correct data sets. For every country in the world, the number of cases recorded and the number of fatalities have been widely published. However, these have many problems with their use as for example some countries may have provided tests in bulk and others may not have so the data will be skewed as a result of this.

We are going to look at the United Kingdom and Brazil because fewer measures were taken at the onset of the epidemic in Brazil so it fits our model fairly well.

At the beginning of the epidemic, the number of infected people is much smaller than the total population and so we can assume $S = N$ is constant. Furthermore the recoveries and fatalities will also be very small allowing us to reduce our equation (8) to:

$$I_{i+1} = I_i + \Delta I_i \tag{12}$$

If we say that the infection lasts exactly d days and ignore the recoveries we can get:

$$I_{i+1} = I_i + R(I_{i-d} - Ii - d - 1), \tag{13}$$

as the number of people who become infected on day i is the number of people who became infected on day $i - d$ times R. Then, the number of people who became infected on day $i - d$ is the total number of people infected on day $i - d$ less the number of people infected the day before.

Q4

In order to estimate values for R for the countries we try to derive an expression for R as a function of λ and d

$$\begin{aligned}
I_i &= Aexp(\lambda i) \\
I_{i+1} &= I_i + R(I_{i-d} - I_{i-d-1}) \\
R &= \frac{I_{i+1} - I_i}{I_{i-d} - I_{i-d-1}} \\
&= \frac{e^{\lambda(i+1)} - e^{\lambda}}{e^{\lambda(i-d)} - e^{\lambda(i-d-1)}} \\
&= \frac{e^{\lambda i}(e^{\lambda} - 1)}{e^{\lambda i}(-e^{-d\lambda}) - e^{\lambda(-d-1)})} \\
&= \frac{e^{\lambda} - 1}{(-e^{-d\lambda})(1 - e^{-\lambda})} \\
&= \frac{e^{d\lambda}(e^{\lambda} - 1)}{\frac{e^{\lambda}-1}{e^{\lambda}}} \\
&= e^{d\lambda}e^{\lambda} \\
R &= e^{\lambda(d+1)}
\end{aligned} \tag{14}$$

Using Python to estimate λ for both countries, we get $\lambda = 0.2944175$ for the United Kingdom and $\lambda = 0.0792947$ for Brazil and then taking $d = 6.5$ to estimate the values R using our rearranged equation. We get $R = 9.099$ for the U.K. and $R = 1.813$ for Brazil. This clearly shows that the U.K. is more densely populated since for each infected individual they spread to more people.

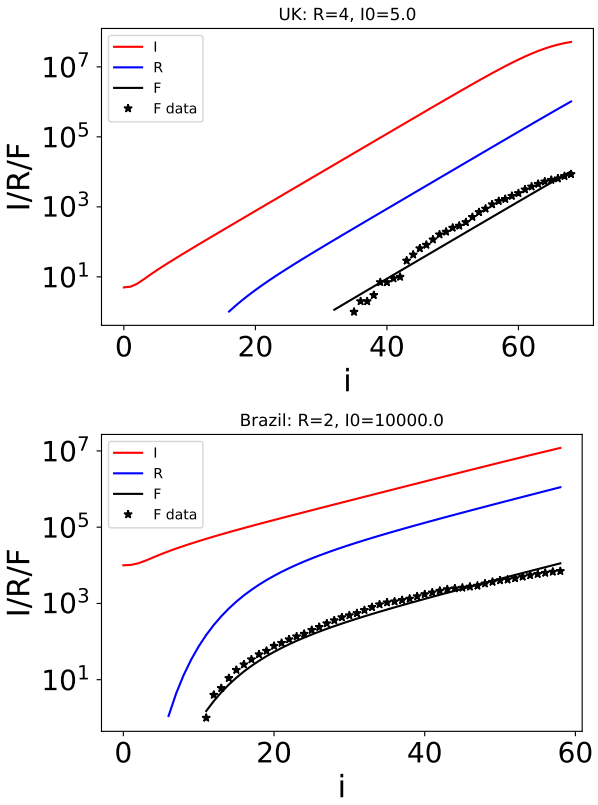

Figure 3: I/R/F for the UK vs Brazil

4 Solving the SIR Covid 19 model

Using a Python program we can solve the model equation 8 to find the parameter R and $I(0)$ of the epidemic in the United Kingdom and Brazil. We find that for Brazil $R = 2$ and $I(0) = 10000$. For the United Kingdom we get $R = 4$ and $I(0) = 5$. As shown in (3) we were able to match the data to the curve by using these values, starting from our initial values estimated earlier from (14). Q5 Q6

The difference between the profile fatalities for the United Kingdom and Brazil is very noticeable as in the logarithmic plot of both we can see in (3) that there is straight line for UK but for Brazil it drops off. This is because in 2020 Brazil stop releasing data of their Covid 19 toll [5] and so there were no fatalities recorded.

5 Conclusions

In conclusion, we have been able to construct a model in order to estimate R values for different countries, where the R values dictate the number of individuals that, on average, an infected person contaminates. We started with a simple that split the population into 4 groups, however, this model missed key information, especially to do with time. For example people do not recover or die after a fixed amount of days. Our final model used probability distributions that characterised the probability of infecting someone else as a function of time since infection. We were then able to use data from the 2 countries of the United Kingdom and Brazil in a Python program in order to make a logarithmic plot, which would help us find these R values. In the end the data from Brazil may have been skewed due to events that would not have been able to be predicted.

References

[1] Flaxman *et. al*, *Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries* https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/

[2] Stephen A. Lauer *et. al*, *The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application* Ann. Intern. Med. doi:10.7326/M20-0504

[3] Sung-mok Jung *et .al*, *Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases* J. Clin. Med. 2020, 9, 523; doi:10.3390/jcm9020523

[4] Fei Zhou *et .al*, *Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study* doi: 10.1016/S0140-6736(20)30566-3

[5] Ana Mano *Brazil takes down COVID-19 data, hiding soaring death toll.*