# SIR Modelling of an Epidemic Outbreak

Joshua Tegegne, Imperial College

Introduction

An SIR Model assumes that at all time, a population can be divided into three groups: susceptible, infected and recovered.

If a susceptible person S comes contact with an infected person I, they may become infected with rate γ. An infected person I can become a recovered person R with rate β (4).

$$S + I \xrightarrow{\beta} 2I, \qquad I \xrightarrow{\gamma} R$$
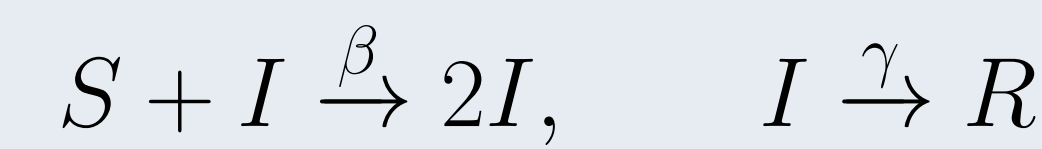
We can then obtain a system of 3 coupled differential equations.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI \qquad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta SI - \gamma I \qquad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I$$

These equations cannot be solved directly so we will use numerical integration.

Motivation

We will use Python to accurately fit the SIR Model to coronavirus outbreak data for Italy for the 81 day period from the 1st March to 20th May. We seek to find the parameters that will minimise the residual sum of squares (RSS). In Italy, the lockdown began on the 9th (Dave Lawler) which will cause the parameters of the model to change - in particular the infection rate as human interaction is significantly reduced.

Method 1

We first want to work out $\theta_1 = (\hat{\beta}_1, \hat{\gamma}_1)$, the vector of parameters that minimises RSS_1, the residual sum of squares **before the lockdown**. Likewise, $\theta_2 = (\hat{\beta}_2, \hat{\gamma}_2)$ minimises RSS_2, which is from the 10th March to 20th May.

Let $\mathbf{I} = (I_1, I_2, \ldots, I_{81})$ and $\mathbf{R} = (R_1, R_2, \ldots, R_{81})$, where I_k is reported proportion of infected people and recovered people respectively at the start of the kth day and vice versa. We let the proportion of susceptible people $\mathbf{S} = 1 - (\mathbf{I} + \mathbf{R})$ We denote $(i_1, i_2, \ldots, i_{81})$ as the proportion of infected people according to the model we choose. We will do this by using a Graphical Monte Carlo simulation.

- Pick an interval in which the parameters lie in, in this case [0,0.5]
- Suppose $\beta, \gamma \sim U[0, 0.5]$
- Take n observations of each variable, for each observation let each pair be the parameters of the model and calculate the respective RSS

- Determine the smallest RSS between those already calculated after n iterations, noting down which parameters this occurs for, and as n increases, by the law of large numbers, this should converge to the true minimum RSS.

$$RSS_1 = \sum_{k=1}^{9} (I_k - i_k)^2 + (R_k - r_k)^2$$

$$RSS_2 = \sum_{k=10}^{81} (I_k - i_k)^2 + (R_k - r_k)^2$$

After 10,000 iterations it is apparent that RSS_1 converges, we find that its minimum value is 2.288×10^{-10}, where $\theta_1 = (0.159, 0.016)$ (3 d.p). Note that we can approximate an interval in which the required parameters belong in by experimenting with different parameters and models on other countries (Wan Nor Arifin, pg 12) With $\theta_2 = (0.101, 0.056)$, we find that the RSS_2 is minimised at 7.39×10^{-5}

We now plot our model and compare it to our reported data;
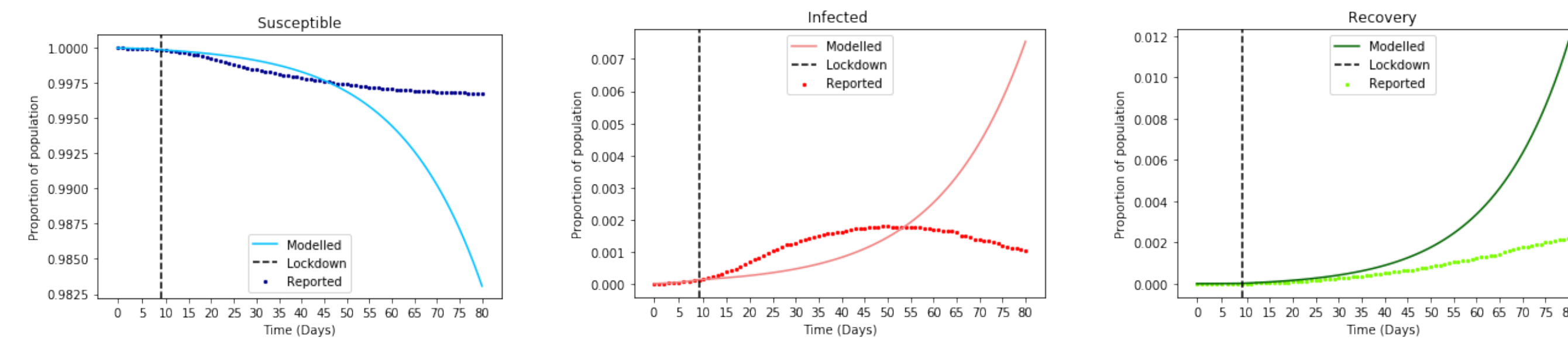
(a) Reported susceptible vs modelled susceptible

(b) Reported infected vs modelled infected

(c) Reported recovered vs modelled recovered

We can see that this method is quite inaccurate as the residuals seem to initially be small, and increase significantly as time passes- we expect the residuals to be normally distributed but they clearly are not. Additionally, the infections in our model reach their peak much later than the reported data does.

Method 2

We will make a few adaptations to our previous model. Instead of using proportions of people infected, we will simply use the actual number of people infected and so on. We want to find N, the effective population size, which we initially assumed to be 60.3 million but it is actually far less than this, most people in Italy will never come into contact with the COVID-19. Other studies that fit COVID-19 data to an SIR model found that the best fit occured when N was taken to be 128 times smaller than the country's population (Wan Nor Arifin, pg 7) Furthermore, we will include the susceptible residuals in the RSS calculations as we aim to have normally distributed residuals.

Note that as we are not using proportions, this slightly changes our differential equations; $\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta SI}{N} \qquad \frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta SI}{N} - \gamma I \qquad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I$
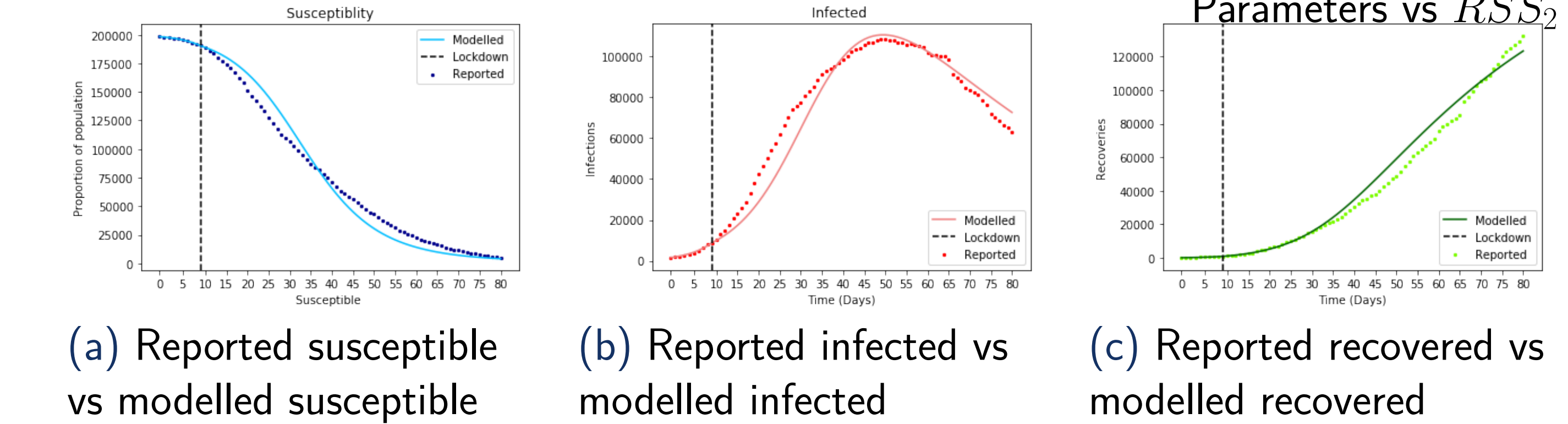
Figure 1 is a plot of N against the minimum RSS_2 for the given N, each of which is calculated Monte Carlo Method again-note I disregard RSS_1 as it clearly does not contribute as much to the overall inaccuracy.

We then find that $\theta_1 = (0.1617, 0.0157), \theta_2 = (0.116, 0.0184)$. The 3D surface plots in Figure 2 and 3 confirm this; here I iterated over all 4 decimal place numbers in an interval containing the components of θ_1 or θ_2, finding where the minimum RSS occurs (at the red dots) and looking at the behaviour near this point.

Figure: 1: N vs RSS_2

Figure: 2: Parameters vs RSS_1

Figure: 3: Parameters vs RSS_2

(a) Reported susceptible vs modelled susceptible

(b) Reported infected vs modelled infected

(c) Reported recovered vs modelled recovered

The model now fits the reported data much better than before, and as expected the recovery rate does not change much when the lockdown is implemented. β is the average number of interactions that transmit the coronavirus so we obviously expect this to decrease once the lockdown is implemented. This is exemplified by Figure 4 which was made by plotting a histogram of the residuals where the integral is 1 so that the frequency became the probability. The best fitting distribution was a $N(-3.33, 7334)$ which is close to a $N(0, \epsilon)$ as desired.
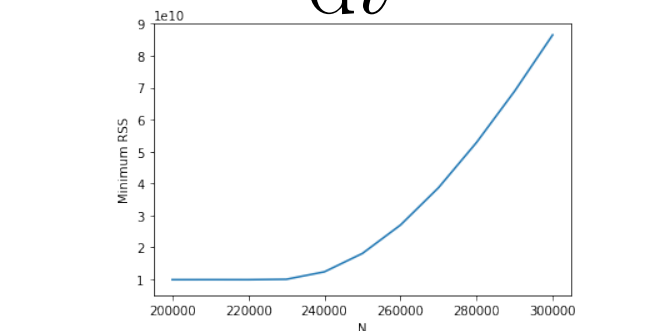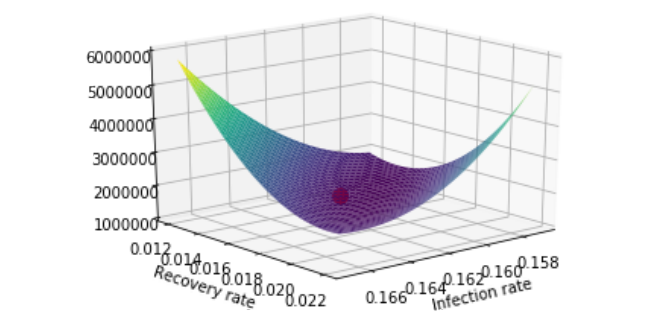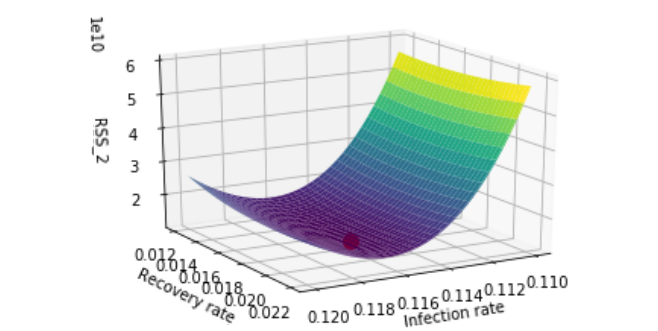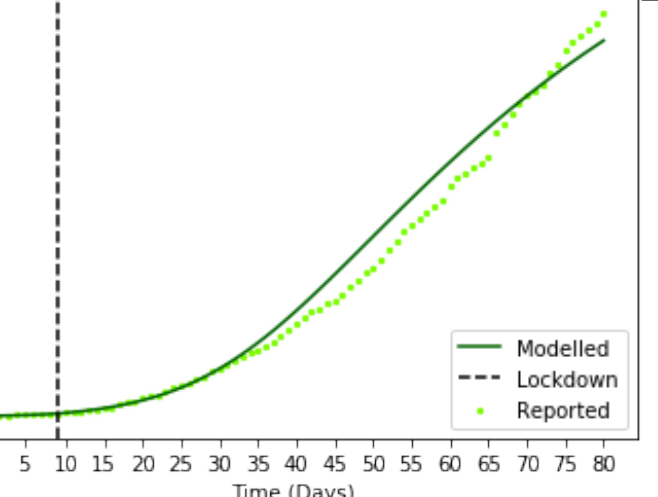

Figure: 4:Residuals fitted to Normal

Sources

[1] CSSE at Johns Hopkins University, *Novel Coronavirus (COVID-19) Cases Data*. Available from:`https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases` [Accessed 4th June 2020]

[2] Dave Lawler. *How Italy's coronavirus crisis became the world's deadliest*. Available from `https://www.axios.com/italy-coronavirus-timeline-lockdown-deaths-cases-2adb0fc7-6ab5-4b` html [Accessed 7th June 2020]

[3] Wan Nor Arifin et al *A Susceptible-Infected-Removed (SIR) model of COVID-19 epidemic trend in Malaysia*. Faculty of Engineering, University Technology Malaysia. Report number: . , 2020

[4] Philip Thomas *M1R[Lecture.]* Imperial College London, 01/06/2020