EPIDEMIOLOGY PROJECT

Course: BB101- Introduction to Biology

Submitted by

Sneha Jayaganthan, I Year, EE Dept (Roll No: 1904135)

Dhasaraiah Gari Sneha I Year, ME Dept (Roll No:1906310)

Indian Institute of Technology Goa

Analysis of COVID-19 Data of UK using SIR MODEL

Abstract

SIR Model has been used to analyse the Covid19 data of North East UK to get an insight in to number of infected, susceptible and recovered people. Runga Kutta technique (4^{th} Order) has been used to obtain the numerical solution of SIR model with the initial conditions assumed for the spread of Covid19. The model parameters β and γ *of* SIR model are estimated by fitting the real Covid19 data of North east UK in order to predict the numbers and calculate the errors such as mean square error and absolute errors. It has been observed that the number of infected cases exhibits an increasing trend at the beginning and then eventually decreases given that individuals recover/decease from the disease. The susceptible fraction of population decreases as the virus is transmitted. An extension of the project on SEIR model has been discussed as second part of the project, which is enclosed along with this report.

1.0 Introduction

An outbreak of "pneumonia of unknown etiology" in Wuhan, Hubei Province, China in early December 2019 has spiralled into an epidemic that is ravaging China and threatening to reach a pandemic state [1]. The causative agent soon proved to be a new betacoronavirus related to the Middle East Respiratory Syndrome virus (MERS-CoV) and the Severe Acute Respiratory Syndrome virus (SARS-CoV). The novel coronavirus SARS-CoV-2 disease has been named "COVID-19" by the World Health Organization (WHO) and on January 30, the COVID-19 outbreak was declared to constitute a Public Health Emergency of International Concern by the WHO Director-General [2]. Despite the lockdown of Wuhan and the suspension of all public transport, flights and trains on January 23, a total of 40,235 confirmed cases, including 6,484 (16.1%) with severe illness, and 909 deaths (2.2%) had been reported in China by the National Health Commission up to February 10, 2020; meanwhile, 319 cases and one death were reported outside of China, in 24 countries [3]. The origin of COVID-19 has not yet been determined although preliminary investigations are suggestive of a zoonotic, possibly of bat, origin [4, 5]. Similarly to SARS-CoV and MERSCoV, the novel virus is transmitted from person to person principally by respiratory droplets, causing such symptoms as fever, cough, and shortness of breath after a period believed to range from 2 to 14 days following infection, according to the Centers for Disease Control and Prevention (CDC) [1, 6, 7]. Preliminary data suggest that older males with comorbidities may be at higher risk for severe illness from

COVID-19 [6, 8, 9]. However, the precise virologic and epidemiologic characteristics, including transmissibility and mortality, of this third zoonotic human coronavirus are still unknown.

Using the serial intervals (SI) of the two other well-known coronavirus diseases, MERS and SARS, as approximations for the true unknown SI, Zhao et al. estimated the mean basic reproduction number (*R*0) of SARS-CoV-2 to range between 2.24 (95% CI: 1.96-2.55) and 3.58 (95% CI: 2.89-4.39) in the early phase of the outbreak [10]. Very similar estimates, 2.2 (95% CI: 1.4- 3.9), were obtained for *R*0 at the early stages of the epidemic by Imai et al. 2.6 (95% CI: 1.5-3.5) [11], as well as by Li et al., who also reported a doubling in size every 7.4 days [1]. Wu et al. estimated the *R*0 at 2.68 (95% CI: 2.47–2.86) with a doubling time every 6.4 days (95% CI: 5.8–7.1) and the epidemic growing exponentially in multiple major Chinese cities with a lag time behind the Wuhan outbreak of about 1–2 weeks [12]. Amidst such an important ongoing public health crisis that also has severe economic repercussions, we reverted to mathematical modelling that can shed light to essential epidemiologic parameters that determine the fate of the epidemic [13]. Here, we present the results of the analysis of Covid19 data of North east UK analysed using SIR model.

2.0 SIR MODEL

SIR model is mode to fit the COVID 19 data of UK collected from the internet. It is one of the popular and simplest compartmental models used for analysing flu infections. The model consists of three compartments such as **S** for the number of **s**usceptible, **I** for the number of **i**nfectious, and **R** for the number **r**ecovered (or immune) individuals. The model can predict rate of infection which is transmitted from human to human and where recovery confers lasting resistance in many flu.

A predominant method of modeling the spread of infectious disease is to categorize individuals in the population as belonging to one of several distinct *compartments*, which represent their health status with respect to the infection. The dynamics of an epidemic can then be analyzed as the rates of transfer between these compartments. One of the most fundamental compartmental models is the SIR model, which forms the basis of much of infectious disease modeling [14]

In the SIR model the population is divided into three compartments, *S* (susceptible), *I* (infected), and *R* (removed). Individuals in the population may exist in any one of these three compartments at a given time.

Susceptible: Susceptible individuals have never been infected, but are susceptible to infection. If they become infected they move into the *Infected* compartment.

Infected: Infected individuals can infect susceptible individuals. After a period of time they move into the *Removed* compartment.

Removed: Removed individuals have either recovered from the infection and are immune to reinfection, or have died.

The rate of transfer from the *Susceptible* population to the *Infected* population is βSI, where β is the per-capita *effective* contact rate (C_e/N). The effective contact rate (C_e) is the number of effective contacts made by a given individual per unit time, where an effective contact is defined as a contact sufficient to lead to infection if it were to occur between a susceptible and an infectious individual. By practicing social distancing we are trying to reduce the value of β. The rate at which *Infected* individuals move into the *Removed* population is I/r, where r is the *recovery delay*. The *recovery delay* represents the length of time an individual remains infectious. The independent variable of the model is the time *t*, and the rates of change of the compartments are expressed as a set of *differential equations* [14]:

$$\frac{dS}{dt} = -\beta SI \quad (1)$$

$$\frac{dI}{dt} = \beta SI - \frac{I}{r} \quad (2)$$

$$\frac{dR}{dt} = \frac{I}{r} \quad (3)$$

The basic reproduction number (R_0) is an indication of the transmissibility of a virus within a particular population. It represents the average number of new infections generated by an infected person in an entirely susceptible population. In this scheme R_0 is given by:

$$R_0 = \beta N r = C_e r \quad (4)$$

where 'N' is the total population:

$$N = S + I + R \quad (5)$$

This simple model predicts behavior similar to that observed in real-world epidemics [15]:

The above equations of SIR model is implemented in python as illustrated in the next section.

2.1 Numerical Simulation of SIR Model

SIR model can be implemented in many ways: from the differential equations governing the system, within a mean field approximation or running the dynamics in a social network (graph). For the sake of simplicity, we've chosen the first option, and ran a numerical method (Runge-Kutta) to solve the differential equations system.

The functions governing the differential equations. are:

```
# Susceptible equation
  def fa(N, a, b, beta):
    fa = -beta*a*b
    return fa

# Infected equation
  def fb(N, a, b, beta, gamma):
    fb = beta*a*b - gamma*b
    return fb

# Recovered/deceased equation
```

```
  def fc(N, b, gamma):
    fc = gamma*b
    return fc
```

In order to solve the differential equations system, we used a 4rth order Runge-Kutta method:

```
# Runge-Kutta method of 4rth order for 3 dimensions (susceptible a,
infected b and recovered r)
def rK4(N, a, b, c, fa, fb, fc, beta, gamma, hs):
    a1 = fa(N, a, b, beta)*hs
    b1 = fb(N, a, b, beta, gamma)*hs
    c1 = fc(N, b, gamma)*hs
    ak = a + a1*0.5
    bk = b + b1*0.5
    ck = c + c1*0.5
    a2 = fa(N, ak, bk, beta)*hs
    b2 = fb(N, ak, bk, beta, gamma)*hs
    c2 = fc(N, bk, gamma)*hs
    ak = a + a2*0.5
    bk = b + b2*0.5
    ck = c + c2*0.5
    a3 = fa(N, ak, bk, beta)*hs
    b3 = fb(N, ak, bk, beta, gamma)*hs
    c3 = fc(N, bk, gamma)*hs
    ak = a + a3
    bk = b + b3
    ck = c + c3
    a4 = fa(N, ak, bk, beta)*hs
    b4 = fb(N, ak, bk, beta, gamma)*hs
    c4 = fc(N, bk, gamma)*hs
    a = a + (a1 + 2*(a2 + a3) + a4)/6
    b = b + (b1 + 2*(b2 + b3) + b4)/6
    c = c + (c1 + 2*(c2 + c3) + c4)/6
    return a, b, c
```

And finally, to obtain the evolution of the disease, we simply define the initial conditions and call the RK4 method:

2.2 Fit SIR parameters to real data

The SIR model is purely theoretical, and we are interested into a real approximation of the COVID-19 expansion in order to extract insights and understand the transmission of the virus. Hence, we need to extract the β and γ parameters for each case if we hope to be able to predict the evolution of the system

3.0 Results and Discussion

The SIR model, which is basically ODE is solved using Runge Kutta (4^{th}) order technique in the present work. With the assumed β value 0.70 and γ value of 0.2, the plot obtained for infected, susceptible, and recovered/deceased people are shown in Fig.1.

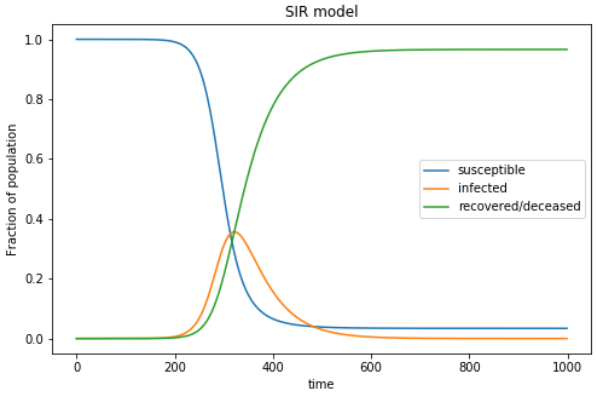

Fig. 1 Fraction of Population (Susceptible, Infected, Recovered/Deceased) versus Time

The following observations are made based on the above figure 1.

- The number of infected cases increases for a certain time period, and then eventually decreases given that individuals recover/decease from the disease
- The susceptible fraction of population decreases as the virus is transmitted, to eventually drop to the absorbent state 0
- The opposite happens for the recovered/deceased case

Fitting of Real Covid Data to SIR Model

The Covid 19 data of North East England is used to fit to SIR Model to analyse the number of infected people and the following plot is obtained (Fig.2). It is observed from the figure that the number of infected people is increasing from the beginning up to 45 days and then follows downward trend. The β value and γ value upon fitting the real covid19 data of north east England was found to be 0.388 and 0.377, respectively.

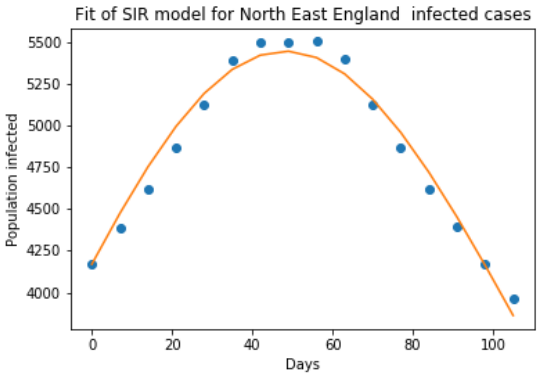

Fig.2: Population infected versus No of days

The Covid-19 data of north east England is used to fit to the SIR Model to get the estimate of recovered/deceased people and plotted in Fig.3. It shows the steady increase of recovered up to 45 days and then attains a constant rate. The β value and γ value upon fitting the real covid19 data of north east England was found to be 0.320 and 0.313, respectively.

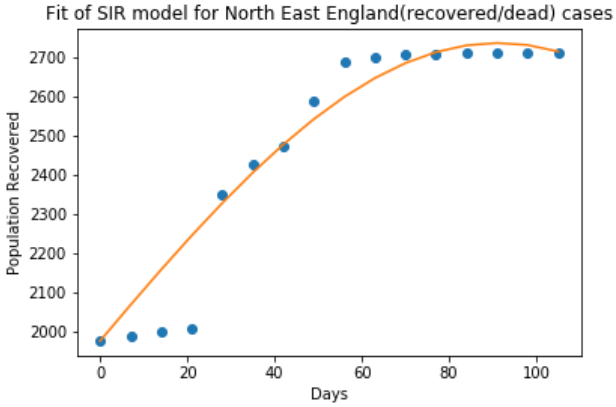

Fig.3 Population Recovered versus No of days

Errors:

The mean square error is defined as the square of the difference between predicted value using SIR model and the actual (real) data of Covid19. The mean square error obtained using SIR model is found to be 6658.38889281338. The absolute mean value obtained through the S IR Model is 74.00606841834758

Limitation of SIR Model:

- SIR model is over simplified model where it assumes that the individual characteristics of immunity, susceptibility, and recovery are same for all members of the population.

- It assumes that the transmission rate remains constant throughout the period of pandemic.
- This model does not differentiate between those who are and who are not in quarantine.
- We assume homogeneity among the whole population, that is, we do not account for some places being initial hot spot and others implementing restrictions earlier and more strenuously.
- There is no account of lock-down, social distancing measures, increasing number of tests carried out to analyse their influence on rate of infections.

4.0 Conclusions

The Covid19 data of North East UK has been used to analyse the number of infected, susceptible and recovered people using SIR model. The numerical solution of SIR model has been obtained using Runge kutta technique (4^{th} Order) with the initial conditions assumed for the spread of Covid19. Subsequently, model parameters β and γ *of* SIR model are estimated by fitting the real Covid19 data of North east UK to the model curve simulated in the present work mimicking the spread of Covid-19. The following conclusions are made in the present work.

- The number of infected cases increases for a certain time period, and then eventually decreases given that individuals recover/deceased.
- The susceptible fraction of population decreases as the virus is transmitted, to eventually drop to the absorbent state 0
- The opposite happens for the recovered/deceased case

References

1. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia; 2020. Available from: https://doi.org/10.1088%2F0951-7715%2F16%2F2% 2F308.
2. Organization WH. WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China; 2020. Available from: https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-ofpneumonia- cases-in-wuhan-china.
3. Organization WH. Novel coronavirus(2019-nCoV). Situation report 21. Geneva, Switzerland: World Health Organization; 2020; 2020. Available from:

https://www.who.int/docs/default-source/ coronaviruse/situation-reports/20200210-sitrep-21-ncov.pdf?sfvrsn=947679ef_2.

4. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet. 2020. https://doi.org/ 10.1016/S0140-6736(20)30251-8.
5. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020. https://doi.org/10.1038/s41586-020-2012-7.
6. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet. 2020. https://doi.org/10.1016/S0140-6736(20)30211-7.
7. Patel A, Jernigan D, nCoV CDC Response Team. Initial Public Health Response and Interim Clinical Guidance for the 2019 Novel Coronavirus Outbreak—United States, December 31, 2019-February 4, 2020. MMWRMorb Mortal Wkly Rep. 2020. https://doi.org/10.15585/mmwr.mm6905e1.
8. Hunag C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020. https://doi.org/10.1016/S0140-6736(20)30183-5
9. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. JAMA. 2020. https://doi.org/10. 1001/jama.2020.1585.
10. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Infect Dis. 2020.
11. Imai N, Cori A, Dorigatti I, et al. Report 3: Transmissibility of 2019-nCoV. Int J Infect Dis. 2019.
12. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet. 2020. https://doi.org/10.1016/S0140-6736(20)30260-9.
13. Siettos CI, Russo L. Mathematical modeling of infectious disease dynamics. Virulence. 2013; 4(4):295–306. https://doi.org/10.4161/viru.24041 PMID: 23552814.
14. Duggan, J. 2016. System Dynamics Modeling with R. Springer
15. Vynnycky, E. and White, R., 2010. An introduction to infectious disease modelling. OUP oxford.