April 9, 2020

Detemination of the basic reproduction number $\mathscr{R}_0$ using epidemic models.

Notes on modeling the spread of COVID-19

Rohith Krishna

PGDM in Reseach and Business Analytics
Madras School of Economics

mailto: pgdm19rohith@mse.ac.in

Contents

1 Introduction

1.1 Statistics of epidemics

The mathematics of epidemiology is essential for understanding the spread and thus the impact of diseases and their toll on the population. It is essential for policy responses and public health professionals. In the period 1918-'19, the *Spanish flu* is estimated to have caused 50 million deaths worldwide. Annually, the *seasonal flu* is the cause of death for about half a million people worldwide. In the 1300s alone, the *Bubonic Plague* resulted in 400 million deaths and decimated a third of European population. In the year 2011 alone, *tuberculosis* accounted for 1.4 million deaths globally, followed by *AIDS* with 1.2 million and *malaria* leaving 0.6 million dead. Together they alone account for 9000 deaths a day!

Diseases that are endemic to a population have real consequences on life span. In many countries endemic diseases may prove to be huge barriers to development. Epidemics however are much shorter outbreaks on a temporal scale. They infect a substantial portion of the population before dying down. Epidemics also have a high chance of recurrence in small outbreaks over the course of several years. The key difference between the epidemics and endemics are also seen in the way they are modeled. Since epidemic outbreaks last only in a short duration of time, one can assume that population changes and immigration/demographic effects are negligible. This results in the model having a constant number of susceptibles, hence pushing the number of infectives down to zero, eventually. In endemics however, these assumptions are relaxed. In endemics, recovered individuals might still lose immunity or could serve as carriers with no symptoms. All these make endemic diseases, persistent in a population.

Pertinent questions:

1. How long will the disease last in the population? (epidemic or endemic?)
2. How severe will be the spread of an epidemic? How many people could get infected on a single day? (and might require treatment)
3. What is the time duration between being exposed and getting infected and being infected and becoming infective?
4. Would the individuals requiring treatment exceed the capacity of the existing healthcare system?
5. How quickly can one utilize existing resources under epidemic conditions to expand the capacity of the healthcare system?
6. How quickly can research yield a vaccine? At what rate can the vaccine be produced and deployed in a large scale?
7. To what extent is quarantine of infected individuals effective? To what extent is isolation necessary in a network of individuals?
8. What is the probability of recurring outbreaks of the epidemic?

1.2 Historical Note

The mathematical study of epidemics is first seen in the work of Daniel Bernoulli in 1766. He estimated the impact of *inoculation* (a process of vaccination) against fighting smallpox. In his time a variant of inoculation called *variolation* was introduced. In this the susceptible individual was subject to a mild strain of the the microbe that was supposed to offer lifelong immunity against the disease but also came with a small risk of infection and disease. This was at a time when the idea of vaccination was still dealt with scepticism. It became imminent on him to test the benefit of variolation. In order to do so he estimated the increase in life expectancy if smallpox was eliminated (as a cause of death), using methods of variolation. His analysis involved the **study of competing risks associated with vaccination - risk of providing immunity or causing infection/death.** In fact this work is more popular among the *actuaries* than among mathematicians. A modern review, of the original paper written in French, is available at [6].

1.3 Compartmental Models

The period between 1900 and 1935, saw several public health physicians introducing pursuing work in compartmental models for the epidemics. Ronald Ross worked on the dynamics of transmission of malaria from its vectors to humans and was awarded the Nobel Prize in 1902. Kendrick and Kermack first introduced the idea of compartmental models in epidemiology in their seminal works.

Compartmental models divided the population into smaller compartments based on certain defining factors. The population could then be divided into three compartments which are: 1. ***susceptibles*** - people who can get infected if they come into contact with the disease, 2. ***infected*** - people who are infected and can serve as carriers for the spread, and 3. ***recovered*** - people who recover from the disease or die. This is called the SIR model. Some assumptions such as neglecting demographic changes, meaning fixed population, may be made in the baseline model. One could also assume that infected patients make full recovery, i.e, gain full immunity against reinfection, or they die and are removed anyway. We might also assume that the size of the compartments are large enough, that mixing of members within the population is homogeneous, although this might not be the case in reality.

An important term associated with epidemiological models is the ***basic reproduction number***, denoted by $\mathscr{R}_0$. It is defined as the expected number of infections produced by a *typical* infected individual in a wholly susceptible population over the full course of the disease outbreak. The $\mathscr{R}_0$ of 1 is a significant number and is infact a threshold line. If the threshold $\mathscr{R}_0 < 1$, then usually, the epidemic dies down, however, if $\mathscr{R}_0 > 1$ then it might usually be the case that the infection spreads on to become an epidemic.

$$\mathscr{R}_0 - \begin{cases} < 1, & \text{infection dies down.} \\ > 1, & \text{becomes epidemic.} \end{cases}$$

If, however, we assume that there is a constant inflow of susceptible individuals in the population,[1] then this could lead to the system attaining disease-free equilibrium if $\mathscr{R}_0 < 1$ or could lead to an endemic equilibrium $\mathscr{R}_0 > 1$, (where the disease is always present in the population).

$$\mathscr{R}_0 - \begin{cases} < 1, & \text{disease-free equilibrium.} \\ > 1, & \text{endemic equilibrium.} \end{cases}$$

1.4 Stochastic Models

During the initial period of the disease outbreak, when the number of infected individuals is still small, it is apt to consider the transmission of the infection as a stochastic event (a random variable in probability), which depends on the patterns in the population. Some of the early work in stochastic models was done by Galton and Watson. Particularly, the beginning of an outbreak can be described by a ***stochastic branching process***.

A key component of stochastic modeling involves significant use of graph theory. One assumes that the infection spread can be perceived to be on a network of contacts between individuals. The vertices of the graph represent people, the members of the population, while the edges represent the contacts between individuals. One approach is to start with *patient zero* who is infected and causes the infection to pass to several nodes through edges that they are in contact with. Another treatment would use the infected edge as a primer, which can infect only a single individual.

The idea of basic reproduction number can also be extended to stochastic models. If $\mathscr{R}_0 < 1$, then the probability that the infection dies down is 1. Else if $\mathscr{R}_0 > 1$, then there is a nonzero probability that the infection persists and leads to an epidemic. This also imbibes another case with $\mathscr{R}_0 > 1$, where there is a nonzero probability that the infection increases initially only to die down much later producing a minor outbreak and not cause a major epidemic.

$$\mathscr{R}_0 - \begin{cases} < 1, & \mathrm{P}(\text{infection dies down}) = 0 \\ > 1, & \text{either} \begin{cases} \mathrm{P}(\text{major epidemic persists}) > 0 \\ \mathrm{P}(\text{minor outbreak and dies down}) > 0 \end{cases} \end{cases}$$

1.5 Models and equilibrium

In order to come up with a model that reflects reality as close as possible, one could combine the two kinds of models discussed above. In the initial outbreak, one could use the branching process in stochastic models, such that in the long run, it breaks down a deterministic compartmental model.

[1] Through population growth. Either directly from the mother - *vertical transmission*, or the newborn coming in contact with an infected individual.

$$\text{Compartmental models} \begin{cases} \text{Simple} & \Longrightarrow \text{identify qualitiative behaviour for understanding the spread. (Analytic solution possible)} \\ \text{Detailed} & \Longrightarrow \text{making quantitative predictions for policy response to the epidemic. (Numerical solution only possible.)} \end{cases}$$

Further, in detailed compartmental models, it is possible to divide the population based on specific behaviour of its agents - ***agent-based models***. The threshold $\mathscr{R}_0$ in useful in predicting the equilibrium state achieved by the system. The following scenario is usually observed.

$$\mathscr{R}_0 - \begin{cases} < 1 & \Longrightarrow \text{disease-free equilibrium; asymptotically stable.} \\ > 1 & \Longrightarrow \text{unique endemic equilibrium exists and is asymptotically stable.} \end{cases}$$

Further, a unique endemic equilibrium could exist which is unstable. In this can one can observe asymptotically stable periodic orbits around the endemic equilibrium. Extending the model to involve heterogeneity in mixing is an advanced step in these models. The population subgroups in this case have different activity levels.

2 The Classical Model

The *classical model* expounded here is the ***deteministic compartmental model***. The population is divided into smaller compartments based on defining features of that compartment. Key assumptions about the nature of the compartments and the rate of transfer of individuals from one compartment to the other are made. A simple model is the SIR model, where $S(t)$ denotes the number individuals susceptible to infection at time t, $I(t)$ denotes the number of infected individuals at time t, and $R(t)$ is the number of individuals removed from the population. This removal can be done either through isolation of infected individuals, or through developing immunity either by vaccination or by full recovery against reinfection, or could be due to death. Although each of these are distinct in a medical stand-point and a growth/fall in their numbers require different policy responses, from a modeling point of view, these are all identical states in the system.

In the SIS model, the assumption of full recovery is relaxed and the recovered individual becomes susceptible again. In the SIRS model, individuals acquire temporal immunity - immunity that is a function of time, and this leads to complicated dynamics. In the SEIR and SEIS models, involves exposed period where time period between exposure and infection and that between infection and being infective are factored. The MSEIRS model incorporates all the above and an additional compartment M which contains infants with passive immunity to the disease.

The defining feature of this model is that **disease transmission is deterministic**, meaning, the population behaviour (and the spread of the disease) is completely dependent on its past states and the rules describing the model. As is the case with any deterministic system, the governing rule/law of nature and the initial states ought to be known to make any prediction about a future state. Further it is also assumed that **the number of members in a compartment is a differentiable function in time**, for the time rate of change of individuals entering and exiting a compartment is measured, with time flowing continuously.

2.1 Kermack-McKendrick Epidemic Model

This is the SIR model mentioned above. It comprises of 3 compartments - Susceptibles S, Infected I and Removed R. We assume mass action incidence in the spread of covid-19, which means that in a population of size N on an average, an individual makes βN contacts per unit time, sufficient to transmit the disease. We also assume that these contacts made are effective in transfer of the disease. The probability that an infective makes contact with the susceptible is S/N, and given the contact rate is βN and we have I infected individuals at time t, we can say that the rate of new infections per unit time is $(\beta N)(S/N)I = \beta SI$.

The number of individuals in each class changes with time and are hence functions of t. Since the total size of the population is N, the sum of the sizes of these three classes are:

$$N = S(t) + I(t) + R(t) \tag{2.1}$$

In order to formulate the model, we make several assumptions that simplify the reality of the situation. These assumptions are:

1. The rate of new infections is given by mass action incidence.
2. The population remains a constant and demographic effects have no role in this model.
3. Infected individuals are also infectious and they transmit the infection at a constant rate proportional to the size of the already existing size.

When a susceptible individual comes in contact with an infectious individual, the susceptible individual becomes infected with a certain probability and moves from the susceptible class to the infected class. The susceptible population decreases in a unit of time by all individuals who become infected in that time. At the same time, the class of infected increases by the same number of newly infected individuals. The number of individuals who become infected per unit of time is called incidence, and the rate of change of the susceptible class is given by:

$$S'(t) = -\text{incidence} \tag{2.2}$$

We earlier calculated the mass action incidence to be the rate of new infections per unit time , which was $(\beta N)(S/N)I = \beta SI$. Therefore:

$$S'(t) = -\beta SI \tag{2.3}$$

The susceptible individuals who become infected move to the class I. Those individuals who recover or die leave the infected class at constant probability per unit time α, and is called the ***recovery rate***. Hence the number of infected individuals who recover/die per unit time and are removed from the population is αI. Thus the change in the infected compartment is given by:

$$I'(t) = \beta SI - \alpha I \tag{2.4}$$

The recovered individuals leave the infected class and move to the recovered class or removed class vary as:

$$R'(t) = \alpha I \tag{2.5}$$

Thus our model is given by:

$$\begin{aligned} S'(t) &= -\beta SI \\ I'(t) &= \beta SI - \alpha I \\ R'(t) &= \alpha I \end{aligned} \tag{2.6}$$

Consider the initial conditions of the model are: $S = S_0$ and $I = I_0$. Further since the number recovered is initially zero, $R - R_0$. Thus $S + I + R = S_0 + I_0$. From out assumption we know that the population remains a constant. Therefore:

$$\frac{d}{dt}(S + I + R) = 0 \tag{2.7}$$

The spread of the disease is imminent if the $I' > 0$. Further since the number of individuals getting infected increases, the compartment size of S keeps decreasing. Therefore, $S' < 0$. That is $S < S_0$ always in this model. Plugging this into the second equation above,

$$\begin{aligned} I' &< \beta S_0 I - \alpha I \\ I' &< I(\beta S_0 - \alpha) \end{aligned} \tag{2.8}$$

I' is negative when $\beta S_0 > \alpha$. We also define contact ratio by β/α. The basic reproduction number $\mathscr{R}_0$ defined previously can be expressed as:

$$\mathscr{R}_0 = \frac{\beta S_0}{\alpha} \tag{2.9}$$

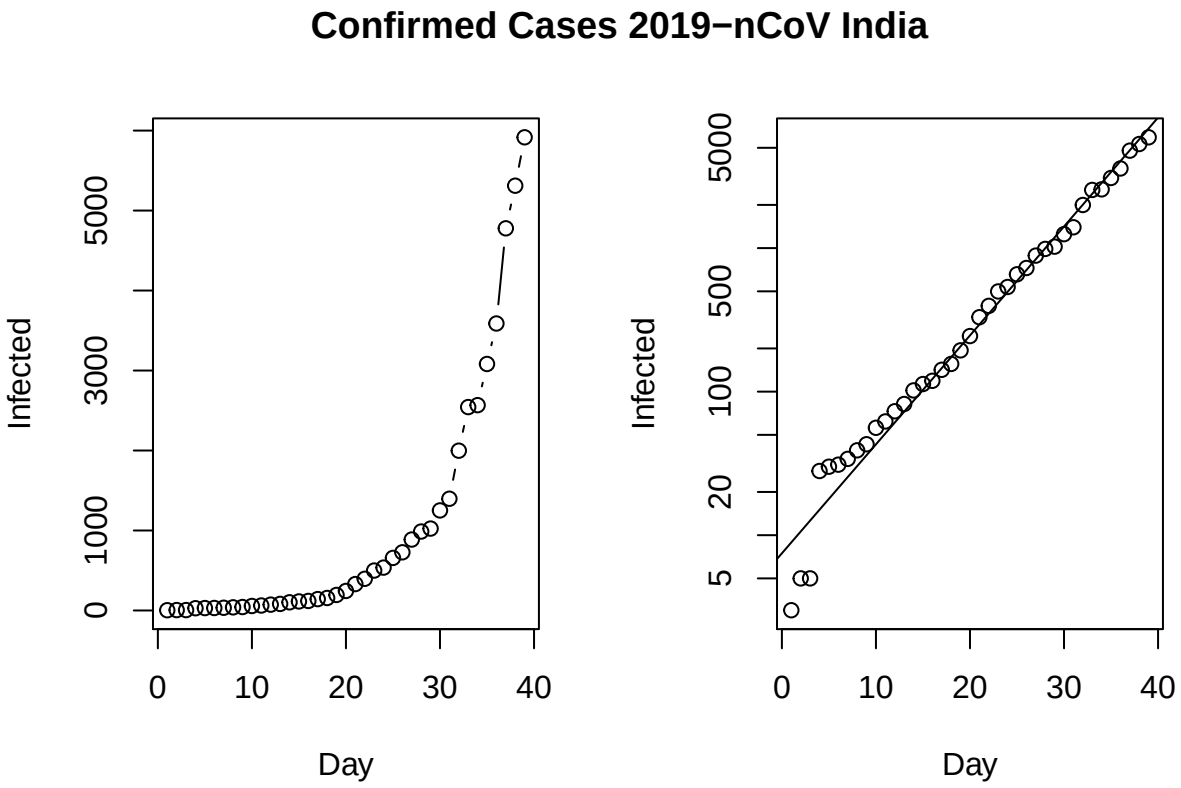
Given that we know $\beta S_0 > \alpha$, we can conclude that the disease transforms into an epidemic if $\mathscr{R}_0 > 1$. We shall now study the case for data from India and determine the basic reproduction number. From its value, we can predict the intensity of the spread.

Calculating $\mathscr{R}_0$ for the spread in India

An important term associated with epidemiological models is the ***basic reproduction number***, denoted by $\mathscr{R}_0$. It is defined as the expected number of infections produced by a *typical* infected individual in a wholly susceptible population over the full course of the disease outbreak. In order to predict the expected number of infections in future time, we start with the existing data. Consider the case for India. The data below represents the number of people infected for a period of 40 days from 28 February 2020 to 8 April 2020. The Indian population in this duration is assumed to be a constant $N = 1387297452$. Further we also perform a linear regression of the *log* of number of infected individuals over time. This functional form regression and the corresponding line-fit is also plotted below.

```
Infected <- c(3,5,5,28,30,31,34,39,43,56,62,73,82,102,113,119,142,156
              ,194,244,330,396,499,536,657,727,887,987,1024,1251,1397
              ,1998,2543,2567,3082,3588,4778,5311,5916)
Day <- 1:(length(Infected))
N <- 1387297452 #Indian Population

old <- par(mfrow = c(1, 2))
plot(Day, Infected, type ="b")
plot(Day, Infected, log = "y")
abline(lm(log10(Infected) ~ Day))
title("Confirmed Cases 2019-nCoV India", outer = TRUE, line = -2)
```



The plot on the left is the number of cases infected as a function of time. One can see the clear

exponential increase in the number of cases of infection. The plot on the right is the log linear plot, with the vertical axis taken on a logarithmic scale.

The SIR model is now defined in R, as shown below. The three set of differential equations representing the rate of change in Susceptible, Infected and Removed compartments.

```
SIR <- function(time, state, parameters) {
  par <- as.list(c(state, parameters))
  with(par, {
    dS <- -beta/N * I * S
    dI <- beta/N * I * S - alpha * I
    dR <- alpha * I
    list(c(dS, dI, dR))
    })
}
```

Now, we fit the model to the data we use: the `ode` function in `deSolve` package to solve the differential equations and the `optim` function to optimize. That is, to minimize the sum of the squared differences between the number of infected I at time t, $I(t)$ and the corresponding number of predicted cases by our model $\hat{I}(t)$

$$RSS(\beta, \alpha) = \sum_t (I(t) - \hat{I}(t))^2$$

Use the `deSolve` package we can solve the given system of ODEs. The residual sum of squares RSS is defined as a function below.

```
library(deSolve)
init <- c(S = N-Infected[1], I = Infected[1], R = 0)
RSS <- function(parameters) {
  names(parameters) <- c("beta", "alpha")
  out <- ode(y = init, times = Day, func = SIR, parms = parameters)
  fit <- out[ , 3]
  sum((Infected - fit)^2)
}
```

Following this, the model is optimized using:

```
Opt <- optim(c(0.5, 0.5), RSS, method = "L-BFGS-B", lower = c(0, 0),
             upper = c(1, 1))
Opt$message
```

```
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
```

The above message states that convergence of the solution has been achieved. The optimal values of β and α are determined now.

```
Opt_par <- setNames(Opt$par, c("beta", "alpha"))
Opt_par
```

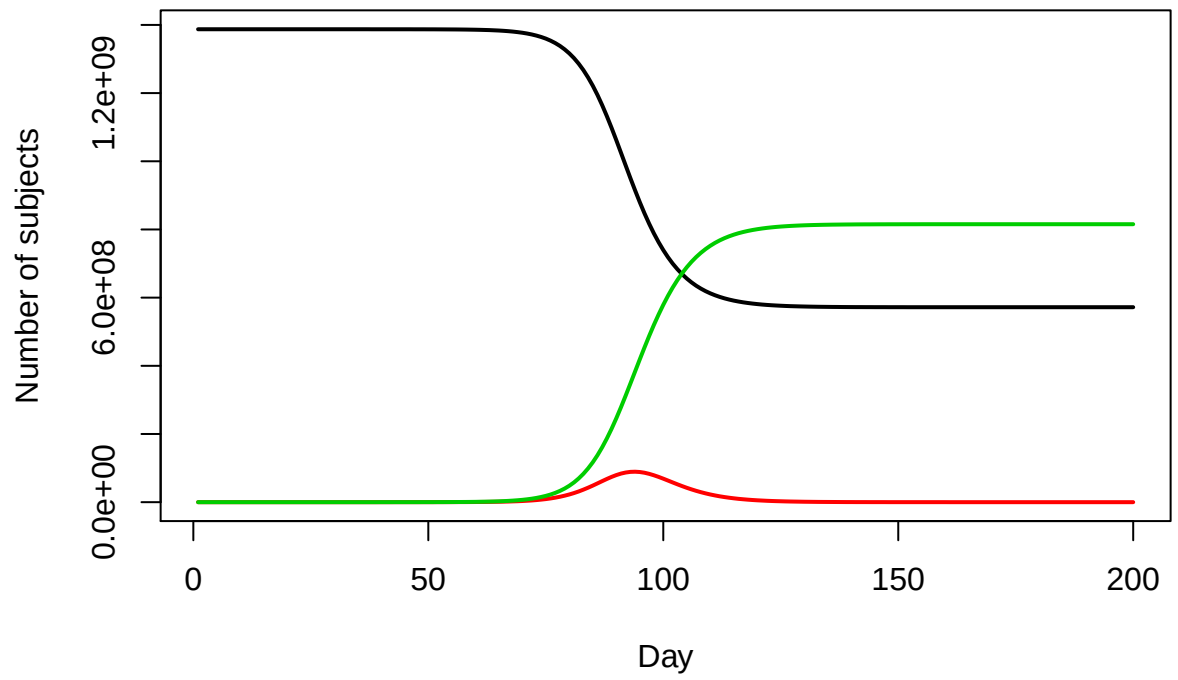
```
##      beta     alpha
## 0.6012267 0.3987733
```

Now, we forcast the spread of the covid outbreak for 200 days from the start date. Thus the time t is set from 1 to 200.

```
t <- 1:200 # time in days
fit <- data.frame(ode(y = init, times = t, func = SIR,
                      parms =Opt_par))
col <- 1:3 # colour
```

The SIR model prediction is now plotted and the trends in the three compartments can be visualized.

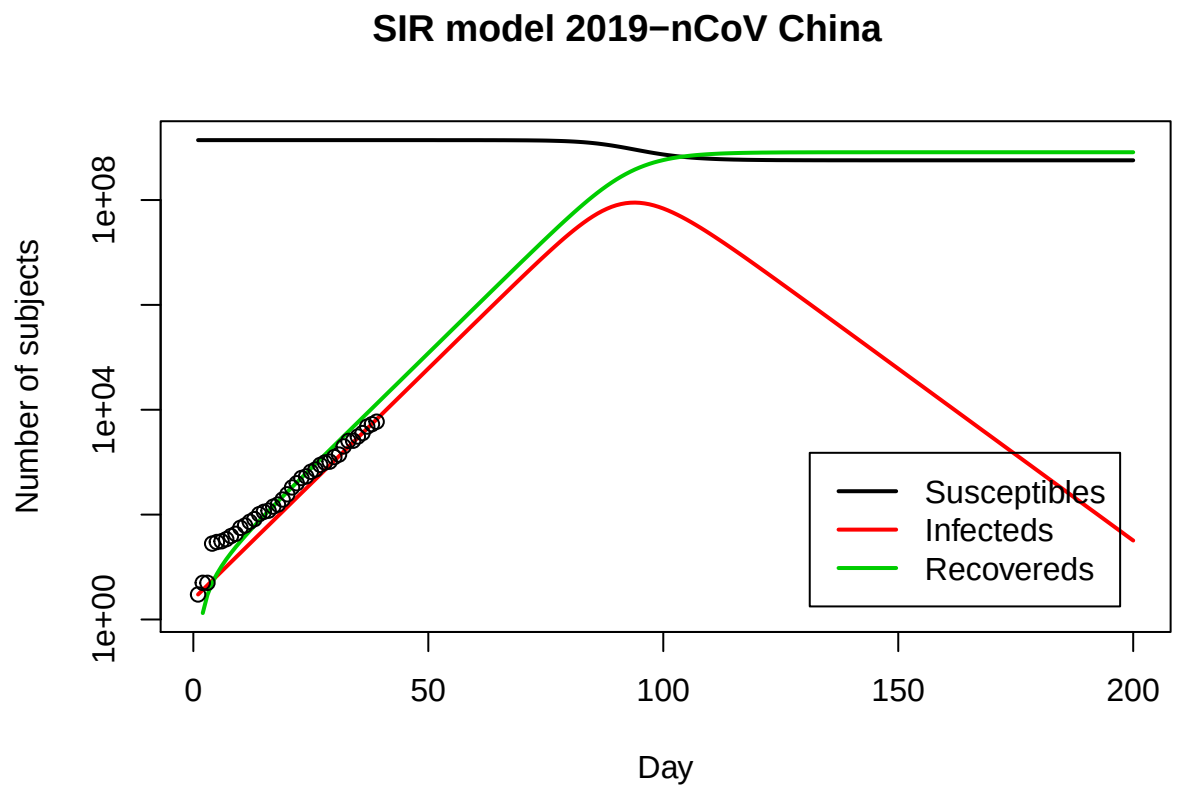
```
matplot(fit$time, fit[ , 2:4], type = "l", xlab = "Day",
        ylab = "Number of subjects", lwd = 2, lty = 1, col = col)
```




```
matplot(fit$time, fit[ , 2:4], type = "l", xlab = "Day",
        ylab = "Number of subjects", lwd = 2, lty = 1, col = col,
        log = "y")
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log = log): 1 y value <= 0 omitted
## from logarithmic plot
```

```r
points(Day, Infected)
legend("bottomright", c("Susceptibles", "Infecteds", "Recovereds"),
       lty = 1, lwd = 2, col = col, inset = 0.05)
title("SIR model 2019-nCoV China", outer = TRUE, line = -2)
```



Calculation of R_0

We can calculate R_0 as shown below. Further, the peak of the pandemic and the maximum number of deaths at a 2% fatality rate is also calculated.

```r
par(old)
R0 <- setNames(Opt_par["beta"] / Opt_par["alpha"], "R0")
R0
```

```
##      R0
## 1.50769
```

```r
fit[fit$I == max(fit$I), "I", drop = FALSE] # height of pandemic
```

```
##           I
## 94 89339796
```

```r
max(fit$I) * 0.02 # max deaths with supposed 2% fatality rate
```

```
## [1] 1786796
```

Inferences.

- According to this model we observe that the number of infected indivuals would tend to grow in a sigmoid shape as characterized by the logistic equation.
- The peak of the spread is expected to be reached in 94 days, which is on June 2, 2020. This model however does not take policy measures into account. Therefore methods of quarantine, isolation and social distancing have not been included. This model, predicts the worst case scenario. The maximum number of infected individuals is estimated at 89,339,796.
- The $\mathscr{R}_0$ for COVID-19 and its spread in India is estimated as 1.5, which is significantly lesser than the number predicted by the data in China and other Western countries.
- Since $\mathscr{R}_0 > 1$ the chances are that the disease breaks out to an unstable epidemic equilibrium or reaches an asymptotic endemic equilibrium.
- Further study on the stability of the equilibrium would throw light into control of the disease. Stastistically significance of the results obtained has not been tested and is scope for further work. But the given data clearly predicts that the disease is here to stay and it's complete eradication would involve gargantuan efforts by policy-makers and public-health officials.

Declaration

I hereby declare that the contents of this report are primarily reading notes sourced from the books [1–5] and lecture notes [7]. The application of the models learnt in these sources to the spread of the COVID-19 pandemic is the author's original work. The data for the same has been sourced from John Hopkins University CSSE. This data can be visualized at https://coronavirus.jhu.edu/ or can be downloaded from github at https://github.com/datasets/covid-19. The data used in the study here, is for the duration from January 22, 2020 to April 4, 2020 spanning 74 days of the spread of the disease. The JHEP article style adopted for this document is edited upon the file sourced at https://jhep.sissa.it/jhep/help/JHEP_TeXclass.jsp.

References

[1] Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). Mathematical Models in Epidemiology. Springer New York. https://doi.org/10.1007/978-1-4939-9828-9

[2] Martcheva, M. (2015). An introduction to mathematical epidemiology (Vol. 61). New York: Springer. https://doi.org/10.1007/978-1-4899-7612-3

[3] Li, M. Y. (2018). An introduction to mathematical modeling of infectious diseases (Vol. 2). Cham: Springer. https://doi.org/10.1007/978-3-319-72122-4

[4] Liu, X., & Stechlinski, P. (2017). Infectious Disease Modeling. Springer. https://doi.org/10.1007/978-3-319-53208-0

[5] Chowell, G., & Hyman, J. M. (Eds.). (2016). Mathematical and statistical modeling for emerging and re-emerging infectious diseases. Springer. https://doi.org/10.1007/978-3-319-40413-4

[6] Dietz, K., & Heesterbeek, J. A. P. (2002). Daniel Bernoulli's epidemiological model revisited. Mathematical biosciences, 180(1-2), 1-21. https://dspace.library.uu.nl/bitstream/handle/1874/8122/heesterbeek_02_daniel_bernoulli's.pdf?sequence=1

[7] Jones, J., (2007). Lecture notes on the Basic Reproduction Number, Stanford University. https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf

Blog entries

[8] ETH Zurich Covid-19 Dashboard. https://bsse.ethz.ch/cevo/research/sars-cov-2.html

[9] Center for Mathematical Modeling of Infectious Diseases. https://cmmid.github.io/topics/covid19/

[10] Tim Churches, Covid-19 Epidemiology in R. RViews. https://rviews.rstudio.com/2020/03/05/covid-19-epidemiology-with-r/, https://timchurches.github.io/blog/posts/2020-02-18-analysing-covid-19-2019-ncov-outbreak-data-with-r-part-1/

[11] Epidemiology: How contagious is Novel Coronavirus (2019-nCoV)? https://blog.ephorie.de/epidemiology-how-contagious-is-novel-coronavirus-2019-ncov