Doubling Times for Infection Growth

Siva Athreya [*] Abhiti Mishra [†]

July 7, 2020

Abstract

In this note we use doubling times to understand the trajectories of COVID-19 infection time line for the States and Union Territories of India. We state a precise definition of doubling time for continuous non-decreasing functions. Using linear interpolation we present a procedure for calculating doubling times for the discrete data. The procedure applies to discrete data that is specified at uniform intervals of time or otherwise.

1 Introduction

One of the earliest works in the study of epidemics was by Bernoulli [4], who attempted to study smallpox and the effectiveness of inoculation against the disease. Modern day mathematical models are based on the classical paper by Kermack and McKendrick [9]. Their work had a major influence on developments in modelling epidemics and the model they proposed came to be known as the SIR model. They divided the entire population into three categories- S- susceptible, I- infected and R- recovered and prescribed dynamics governed by a system of differential equations. This model was quite useful in explaining the trends observed in the number of infected patients in epidemics such as the plague (London 1665-1666, Bombay 1906) and cholera (London 1865). We refer the reader to [5], for a survey of other stochastic as well as deterministic models in the epidemiology literature.

In any epidemic, a basic question is: How fast is the infection spreading? One way to answer this question is to look at how long ago the infection count was half of the current count. This quantity is known as *doubling time*. It is an estimate of the current rate of spread of infection. For example, consider the West African Ebola virus that resulted in immense loss of life in the countries of Liberia, Sierra-Leone and Guinea from 2013 – 2016 (see [12] for further details).

As seen in Figure 1, the infection growth appears to be very slow in the beginning. It starts growing quickly around June 2014. Toward the beginning of December 2014, the infection growth starts to slow down and does not grow in late 2015. Hence, from June 2014 to December 2014, the rate of growth seems high, suggesting that the infection count is doubling quickly. One can infer that the doubling time is low in this period. Towards the end of the epidemic the rate of growth reduces and the doubling time becomes larger (see Section 2.3.2 for precise calculation of the doubling time). Doubling time can also be used to calculate the transmission potential of a disease. In [8], the author studies the outbreak of avian influenze subtype H5N1 in Romania in 2006 and uses the doubling time to estimate R_0, or basic reproductive number. A similar study has been carried out for the SARS epidemic of 2003 in [7].

---

[*] 8th Mile Mysore Road, Indian Statistical Institute, Bangalore 560059, India. Email: athreya@isibang.ac.in

[†] 8th Mile Mysore Road, Indian Statistical Institute, Bangalore 560059, India. Email: bmat1803@isibang.ac.in

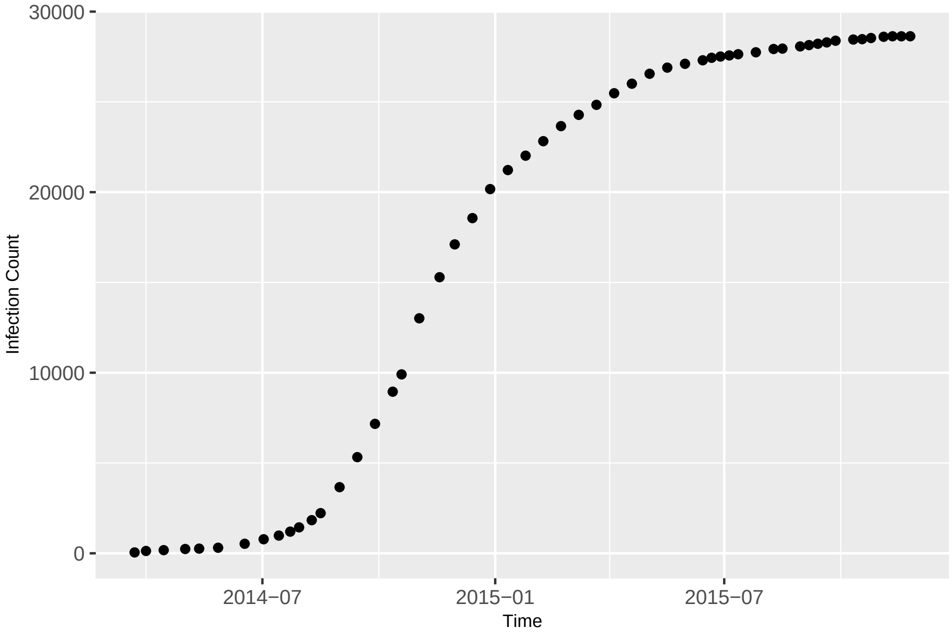


Figure 1: Ebola Cases in Liberia, Sierra-Leone, and Guinea during April, 2014 to November, 2015.

In this note, we shall first define the concept of doubling time for non-decreasing continuous functions (Section 2). Using this defintion we calculate doubling times for various functions that are used as mathematical models for epidemics (see Section 2.1, Section 2.2). We then describe a method to calculate doubling times for discrete-time data. The method is based on first constructing a linear interpolation and then using the definition of doubling time for continuous functions to calculate an approximation for the doubling time of the data set (see Section 2.3). We conclude the article by applying this method to find doubling time for the COVID-19 infection time line for all of India as well as some states of India (see Section 3).

On our accompanying portal COVID−19 India-Timeline an understanding across States and Union Territories, we have calculated Doubling times across States and Union Territories in India. We were inspired in part by Deepayan Sarkar's website https://deepayan.github.io/covid-19/doubling which looks at the doubling times of the COVID-19 pandemic in various countries.

Remark 1. A word of caution before we begin. The data we use is taken from website of the Union Ministry of Health and Family Welfare, Government of India. The data contains daily counts of Infected, Recovered and Deceased across the states of India. The number of infected patients is the number of positive test results in each state on that day. So we are equating the number of those tested positive to the number of infected individuals. This may be an error, because every individual in the population has not been tested. Thus for any inference or conclusion on infection growth we must take into consideration the different policy/rates of testing, population density, quarantine measures, and biological aspects of this epidemic.

Acknowledgements: We would like to thank Nitya Gadhiwala for useful discussions as well as Jaya Mishra and Abhiram Kumar for providing detailed feedback on an earlier version of the note.

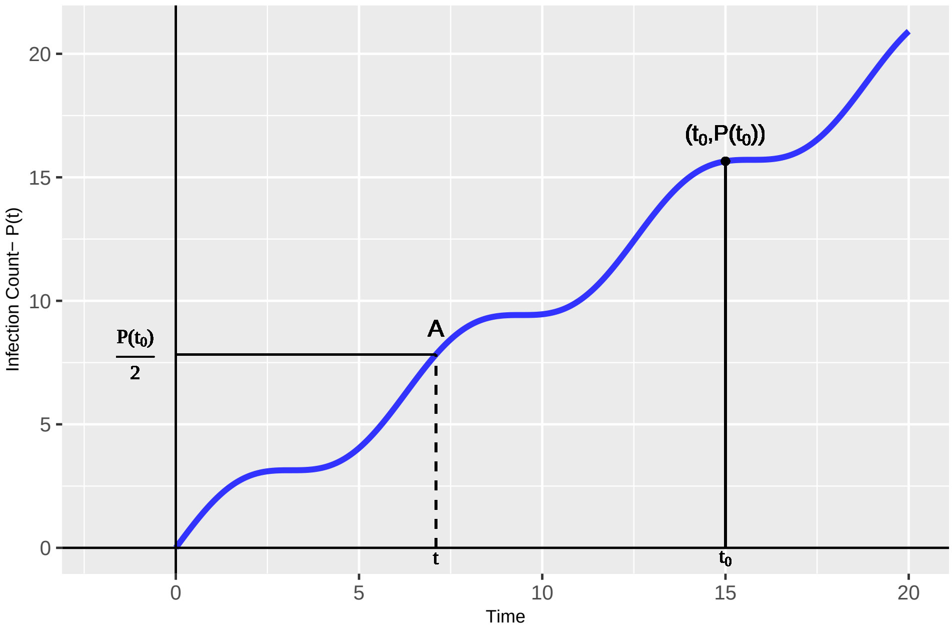


Figure 2: Finding Doubling Time from Infection Timeline

2 Mathematical Preliminaries

In this section we will explain the basic mathematical preliminaries. Section 2.1 and Section 2.2 are intended for the reader who is not familiar with Doubling times, Exponential functions and Logistic functions. We discuss how to apply the methods to discrete data that are available at uniform intervals of time in Section 2.3.1 and in Section 2.3.2 we consider data that are not necessarily provided at uniform intervals of time.

One can understand the notion of doubling time graphically as follows. Consider the graph in Figure 2. Suppose the function $P(t)$ represents the infection count and we want to find the doubling time at a given time t_0. Then we look at the point $(t_0, P(t_0))$ and we want to find the time at which the infection count was half. So, we draw a horizontal line at $y = P(t_0)/2$ and look at the point, say A, at which it intersects the curve. Then the x-coordinate of the point A is the time t at which the count was half. Thus doubling time at t_0 is given by $t_0 - t$.

We now provide a precise definition for doubling time. We will say a function $P : [0, T] \to \mathbb{R}$ is said to be increasing if $P(s) \leq P(t)$ for all $s \leq t$.

Definition 2.1. *Let* $P : [0, T] \to \mathbb{R}$ *be a continuous and increasing function. Then let*

$$\mathcal{D} = \{s \in [0, T] : P(s) \geq 2P(0)\} \tag{1}$$

For any $t_0 \in D$ *let* $t \in [0, t_0]$ *be such that*

$$P(t) = \frac{P(t_0)}{2} \qquad \text{and} \qquad P(s) < \frac{P(t_0)}{2} \text{ for } s < t.$$

Then the doubling time at t_0*, denoted by* $D(t_0)$ *is given by*

$$D(t_0) = t_0 - t.$$

We make a couple of remarks on the definition. The Intermediate Value Theorem will ensure that for any $t_0 \in D$, there will be at least one $t \in [0, t_0)$ such that $P(t)$ is one-half of $P(t_0)$. If there are multiple

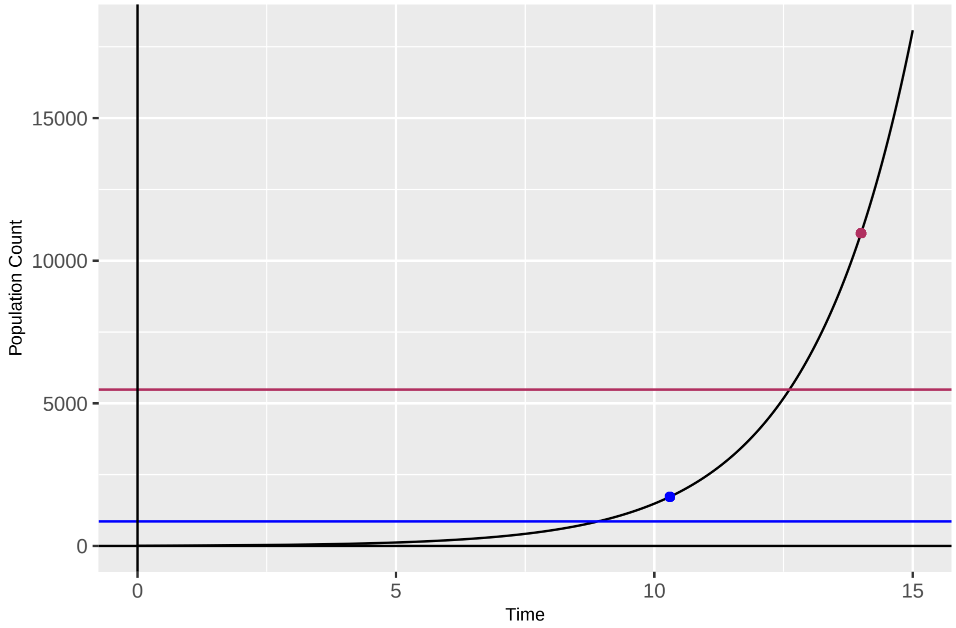


Figure 3: We pick two points on the exponential function, represented by the blue and maroon dots. We have drawn horizontal lines which represent half of the population count at these points. The blue line represents half of the population count at the blue point and the maroon line does so for the maroon point.

$t \in [0, t_0)$ that satisfy $P(t) = \frac{P(t_0)}{2}$, then the condition $P(s) < \frac{P(t_0)}{2}$ for all $s < t$ ensures that we take the smallest such t. So the doubling time for all $t_0 \in D$ is well defined. Secondly, if the function $P : [0, T] \to \mathbb{R}$ is strictly increasing (i.e. $s < t$ implies $P(s) < P(t)$) and continuous then it is not hard to see that there is a unique $t \in [0, t_0)$ such that $P(t) = \frac{1}{2}P(t_0)$. One can also observe that if we scale the function P by a constant factor then this does not change the values of doubling time.

We will now work with two specific class of functions, namely Exponential function and Logistic function. Using the above definition we shall calculate the doubling times for the same.

2.1 Doubling Time for an Exponential Function

We noticed for the Ebloa epidemic (see Figure 1) that during the period June 2014 to December 2014 the infection count grew rapidly. This phase is referred to as the Exponential phase growth of the infections. This feature of the trajectory is typical for any infection growth during the initial stages of the epidemic and can be modelled by the exponential function.

Let $P : [0, 15] \to \mathbb{R}$ (representing the count of infection over a 15 day period) be given by the exponential function. That is,

$$P(t) = P_0 e^{rt}, \qquad \text{for } t \in [0, 15] \tag{2}$$

Putting $t = 0$ in the above equation gives $P(0) = P_0$, hence P_0 is the initial population count. In Figure 3, we have plotted the exponential function in which the initial population P_0 is 10 and the rate r is 0.5. Using Definition 2.1 let us calculate the doubling time for a given instant $t_0 \in \mathcal{D}$ (see 1). The exponential function is a strictly increasing function. Hence there is a unique time $t \in [0, t_0)$ at which

$$P_0 e^{rt} = \frac{1}{2} \times P_0 e^{rt_0}.$$

By cancelling P_0 from both sides we obtain that

$$e^{rt} = \frac{1}{2} \times e^{rt_0}.$$

Taking logarithms on both sides and doing elementary algebra we obtain

$$r(t_0 - t) = \log 2.$$

Therefore,

$$D(t_0) = \frac{\log 2}{r}. \tag{3}$$

We note that $D(t_0)$ is independent of the time t_0 and only depends on the rate of growth r. This is a feature of the exponential function. Thus, when the infection count is given by the exponential function, the doubling time remains constant over time and is inversely proportional to the rate of spread r. In this setting, we can infer that a higher doubling time indicates that the rate of growth is small and it takes a disease longer to spread. Similarly, a lower doubling time indicates a faster rate of growth.

2.2 Doubling Time for Logistic Function

In the previous section we had assumed that infection growth $P(t)$ was given by an exponential funciton. There the infection keeps increasing indefinitely with time. As we saw in the Ebola epidemic (see Figure 1), this behaviour is not seen and the infection growth slows down in late 2015. This is modelled by the logistic function.

Suppose there is a K which represents the highest value the infection count can rise up to. Under this hypothesis, when the population count $P(t)$ is such that $P(t) << K$, then we assume the resources are almost freely available and unimpeded growth of infection occurs. However, as the infection count increases, there is competition for the resources, which slows down the growth rate.

This idea was captured by the Belgian mathematician Pierre Verhulst (1838), who is credited with developing the logisitic model. He suggested that instead of the exponential curve, where the growth rate only depends on $P(t)$, the growth rate for infections in an epidemic model with resource constraints is given by the logistic function

$$P(t) = \frac{KP_0}{P_0(1 - e^{-rt}) + Ke^{-rt}}. \tag{4}$$

If we set $t = 0$ in (4), it gives us $P(0) = P_0$, which represents the initial count of infections. In Figure 4, we plot a logisitic function with $P_0 = 10$, $K = 10000$ and $r = 0.2$.

By defintion, the inflection point is achieved when

$$\frac{d^2P}{dt^2}(t) = 0,$$

with P given by (4). It is an easy exercise to verify that the inflection point is achieved at time

$$t_{\text{inf}} = \frac{1}{r} \ln\left(\frac{K - P_0}{P_0}\right) \tag{5}$$

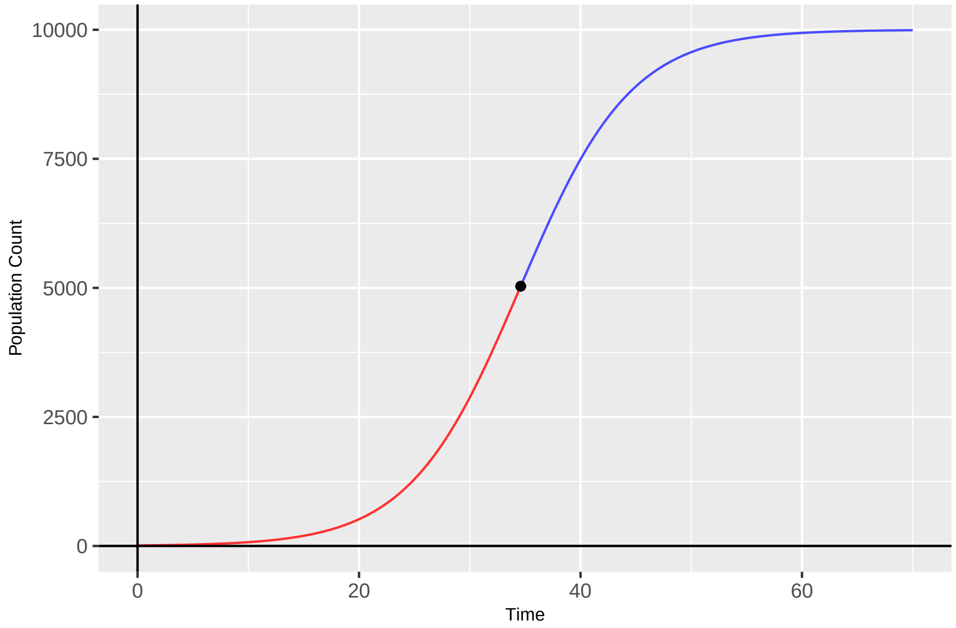


Figure 4: The red part of the curve is known as the *exponential phase* where the function increases sharply. The black dot in the middle represents the point in time where the growth begins to decrease and is called the *inflection point* of the curve. The blue part of the curve is where the growth rate decreases due to increased competition for resources. The curve has an asymptotic value of K.

Now suppose we want to calculate the doubling time at a given time t_0. Suppose that t is the time at which the count was half. Then we observe that the population at time t is

$$P(t) = \frac{1}{2}P(t_0).$$

This implies that

$$\frac{KP_0}{P_0(1-e^{-rt})+Ke^{-rt}} = \frac{1}{2}\frac{KP_0}{P_0(1-e^{-rt_0})+Ke^{-rt_0}}.$$

Cross-multiplying and simplifying in the above, we get

$$P_0 - 2P_0e^{-rt_0} + P_0e^{-rt} + 2Ke^{-rt_0} - Ke^{-rt} = 0.$$

Rearranging the terms we obtain

$$e^{-rt} = \frac{P_0}{K-P_0} + 2e^{-rt_0}.$$

Taking logarithm and dividing by r we get

$$t = -\frac{1}{r}\log\left(\frac{P_0}{K-P_0} + 2e^{-rt_0}\right)$$

As discussed in the previous section the doubling time at time t_0 is given by

$$D(t_0) = t_0 - t = t_0 + \frac{1}{r}\log\left(\frac{P_0}{K-P_0} + 2e^{-rt_0}\right) = \frac{1}{r}\log\left(\frac{P_0e^{rt_0}}{K-P_0} + 2\right) \tag{6}$$

From (6), the doubling time depends on the time at which it is computed, unlike the exponential function when it was constant (see (3)). In Figure 5, we plot the doubling time versus the time for the same logistic model as above ($P_0 = 10, K = 10000, r = 0.2$).

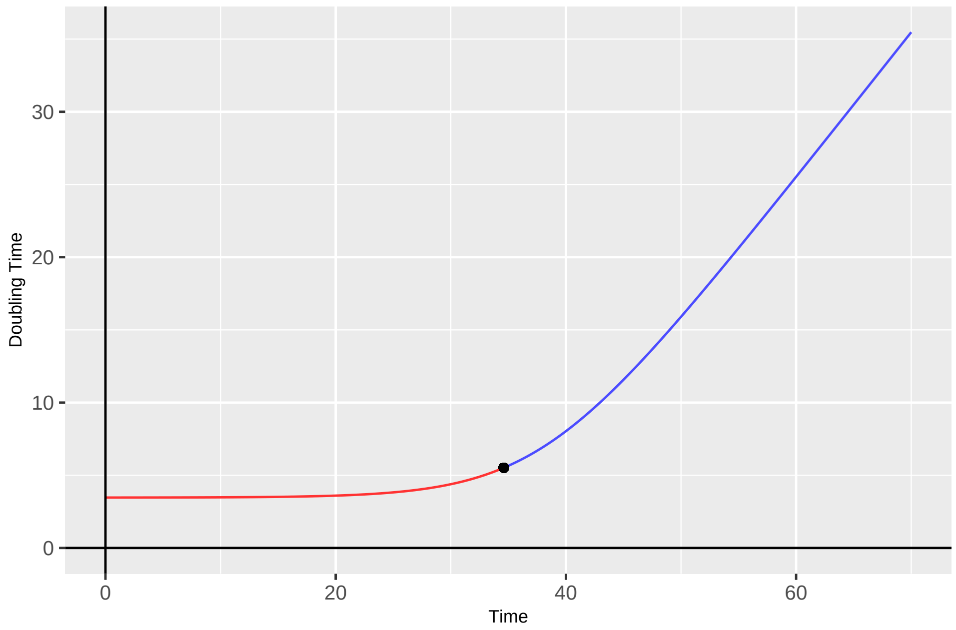


Figure 5: Doubling Time for Logistic Model

An interesting question is whether we can calculate the inflection point, t_{inf} given by (5), when only the doubling times are known. First it is standard to verify that

$$\frac{dD}{dt}(t) = 1 - \frac{2(K - P_0)}{2(K - P_0) + P_0 e^{rt}} > 0$$

for all $t > 0$ and that at t_{inf},

$$D(t_{\text{inf}}) = \frac{1}{r}\log(\frac{K - P_0}{P_0}) + \frac{1}{r}\log(\frac{3P_0}{K - P_0}) = \frac{\log 3}{r} \tag{7}$$

Thus if the rate r is known, then the time at which the inflection point occurs can be determined.

2.3 Estimating Doubling Time for Discrete Data

As stated earlier one of our main goals is to calculate doubling time for the COVID-19 infection time line. The infection count is reported every day. Suppose the daily counts read as $4, 10, 16, 19, \ldots$. If we try to calculate doubling time for Day 2, then we find that we have no data point at which the count was $\frac{10}{2} = 5$. Similarly, for Day 3 there is no data point at which the count was 8. Thus we will need to come with up an approximation for the doubling time.

To do this, we will fit a piecewise linear function that passes through the points given by empirical data. Then use the doubling time definition for functions (recall Definition 2.1) as an approximation for doubling time of the data set. Before describing it in detail we will work it out with an example. Consider the exponential curve that we worked with in Section 2.1. Let the initial population $P_0 = 10$ and the rate $r = 0.5$. Now suppose we are *only* given the evaluations of this function at $k = 0, 1, 2, \ldots, 20$ and asked to find the doubling time at $t_0 = 3$.

We will fit a linear interpolating function that passes through the data points $(k, P(k)) = (k, 10e^{0.5k})$ for $k = 0, 1, 2, \ldots, 20$. Such a function will need to pass through the given set of points and between adjacent data points should be a line segment. One can give a precise definition for the linear interpolating function $(\tilde{P})$ by

$$\begin{aligned}\tilde{P}(t) &= P(k) + (P(k+1) - P(k))(t-k) \\ &= 10e^{0.5k} + (10e^{0.5(k+1)} - 10e^{0.5k})(t-k) \\ &= 10e^{0.5k}[k+1-ke^{0.5}] + 10t(e^{0.5(k+1)} - e^{0.5k}) \end{aligned} \tag{8}$$

when $t \in [k, k+1)$, for $k = 0, 1, 2 \ldots$. We provide a plot of the function in Figure 6.

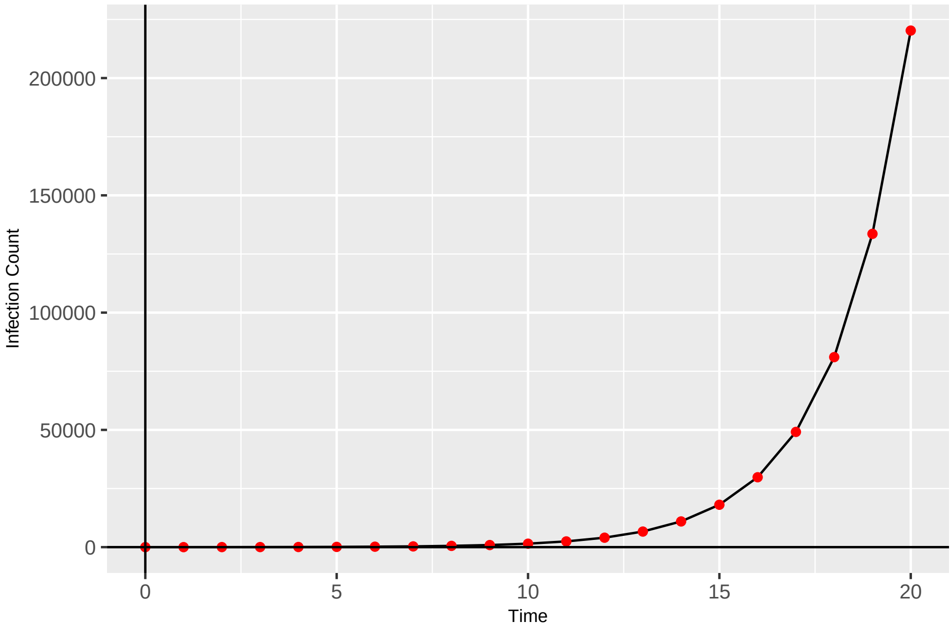


Figure 6: Linear Interpolation of $\{(k, P(k)) : k = 0, 1, 2, \ldots\}$

Note that because the given data set comes from a strictly increasing function, by definition the linear interpolating function $\tilde{P}$ is also strictly increasing and continuous. We proceed as in Definition 2.1 to find doubling time at $t_0 = 3$ for $\tilde{P}$. Note $\tilde{P}(3) = 10e^{1.5} > 5 = \frac{\tilde{P}(0)}{2}$. To find the unique $0 < t < t_0$, we set

$$\tilde{P}(t) = \frac{1}{2}\tilde{P}(t_0).$$

Now, $\tilde{P}(3) = P(3) = 10e^{1.5}$ and $\tilde{P}(2) = P(2) = 10e$. As $2 < e$ and $e^{0.5} < 2$, we have $\tilde{P}(1) < \frac{1}{2}\tilde{P}(3) < \tilde{P}(2)$. This would imply that $t \in (1, 2)$. Therefore we have

$$\tilde{P}(1) + (\tilde{P}(2) - \tilde{P}(1))(t-1) = \tilde{P}(t) = \frac{1}{2}\tilde{P}(3).$$

Elementary algebra yields that

$$t = 1 + \frac{\tilde{P}(3) - 2\tilde{P}(1)}{2(\tilde{P}(2) - \tilde{P}(1))}$$

Therefore the doubling time at $t_0 = 3$ is given by

$$\tilde{D}(3) = 3 - t = 2 - \frac{\tilde{P}(3) - 2\tilde{P}(1)}{2(\tilde{P}(2) - \tilde{P}(1))} = 2 - \frac{e^{1.5} - 2e^{0.5}}{2(e - e^{0.5})} = 1.446.$$

We will now assign this the doubling time at 3 for the data points

$$\{(k, P(k) : k = 0, 1, 2, \ldots, 20\}$$

to be equal to $\tilde{D}(3) = 1.446$. We are now ready to provide a prescription for finding doubling times of a given discrete data set.

2.3.1 Method for Calculating Doubling Time for Discrete Data: Uniform

In this subsection we will decribe the method to calculate doubling time for discrete data that is given at uniform intervals of time (daily, weekly or for allx consecutive unit of time). Suppose that we are given a discrete data set

$$\mathbf{P} = \{(k, P(k)) : k = 0, 1, 2, \ldots T\},$$

where for each time k, $P(k)$ denotes the infection count at time k and T is last time for which the data is available. Let

$$\mathcal{D}_{\text{disc}} = \{k : P(k) \geq 2P(0)\}.$$

To calculate doubling time for the data set $\mathbf{P}$ at $t_0 \in \mathcal{D}_{\text{disc}}$ we shall first find the smallest k_0 with $k_0 < t_0$ such that

$$P(k_0) \leq \frac{P(t_0)}{2} \leq P(k_0 + 1)$$

Then we construct the linear interpolating function for $\mathbf{P}$, namely $\tilde{P} : [0, T] \to \mathbb{R}$ given by

$$\tilde{P}(t) = P(k) + (P(k+1) - P(k))(t - k) \tag{9}$$

when $t \in [k, k+1)$, for $k = 0, 1, 2 \ldots, T - 1$ and then using Definition 2.1 we have to find $t \in [k_0, k_0 + 1)$ for which

$$P(k_0) + (P(k_0 + 1) - P(k_0))(t - k_0) = \frac{P(t_0)}{2} \tag{10}$$

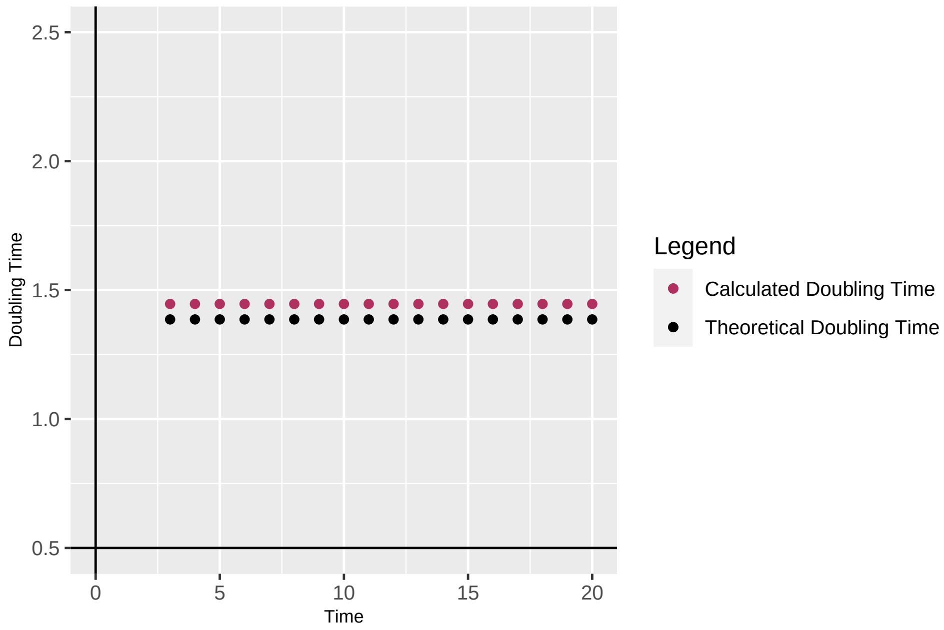


Figure 7: The plot contains the calculated and theoretical values of doubling time for Exponential function with $r = 0.5, P_0 = 10$. The calculated doubling time is constant and has value 1.4464, while the theoretical doubling time value is 1.386. One can check that as one takes more discrete points on the exponential curve, the calculated doubling time values become closer to the theoretical value.

Solving for t gives us

$$t = \frac{1}{2}\left(\frac{P(t_0) - 2P(k_0)}{P(k_0+1) - P(k_0)}\right) + k_0 \tag{11}$$

Thus, doubling time at t_0 is given by

$$D(t_0) = t_0 - k_0 - \frac{1}{2}\left(\frac{P(t_0) - 2P(k_0)}{P(k_0+1) - P(k_0)}\right) \tag{12}$$

We now consider two examples of data sets that arise from functions (Exponential from Section 2.1 and respectively Logistic from Section 2.2) at discrete time points $t = 0, 1, 2, \ldots, 20$ and plot doubling times for them using only the latter information. We also compare them with theoretical doubling times calculated earlier (see Figure 7 and Figure 8).

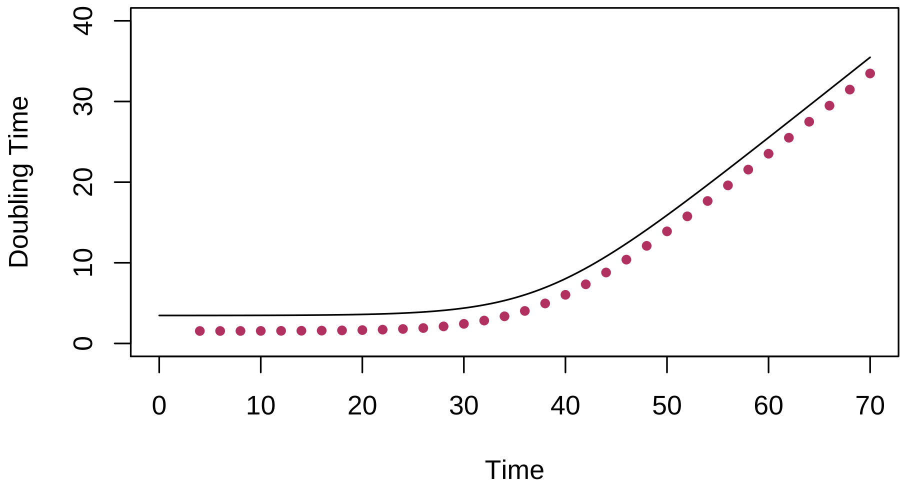


Figure 8: The plot contains the calculated and theoretical values of doubling time for Logistic function with $K = 10,000, r = 0.2, P_0 = 10$. The black curve is the theoretical curve for doubling times. The maroon dots are the doubling times calculated by taking 35 discrete data points (whose x-co-ordinates are equally spaced from 2 to 70. The difference in theoretical and calculated values is around 2 for all data points.

One could try to formulate a theory w.r.t the above approximation. That is, provide precise error bounds between doubling times calculated by discretising the function and the theoretical doubling times calculated by using Definition 2.1 for a given class of functions.

2.3.2 Method for Calculating Doubling Time for Discrete Data: Non-Uniform

We now look at data sets for which infection count data is given at non-uniform intervals of time. In Figure 1, we had plotted the infection time line for the Ebola Cases in Liberia, Sierra-Leone, and Guinea from April, 2014 to November, 2015. One observes that the infection count is not reported daily in the Data set. Below is snap shot of the data from July 2014 to September 2014.

Date	Total Cases
...	...
23/07/14	1201
30/07/14	1437
09/08/14	1835
16/08/14	2225
31/08/14	3664
14/09/14	5327
...	...

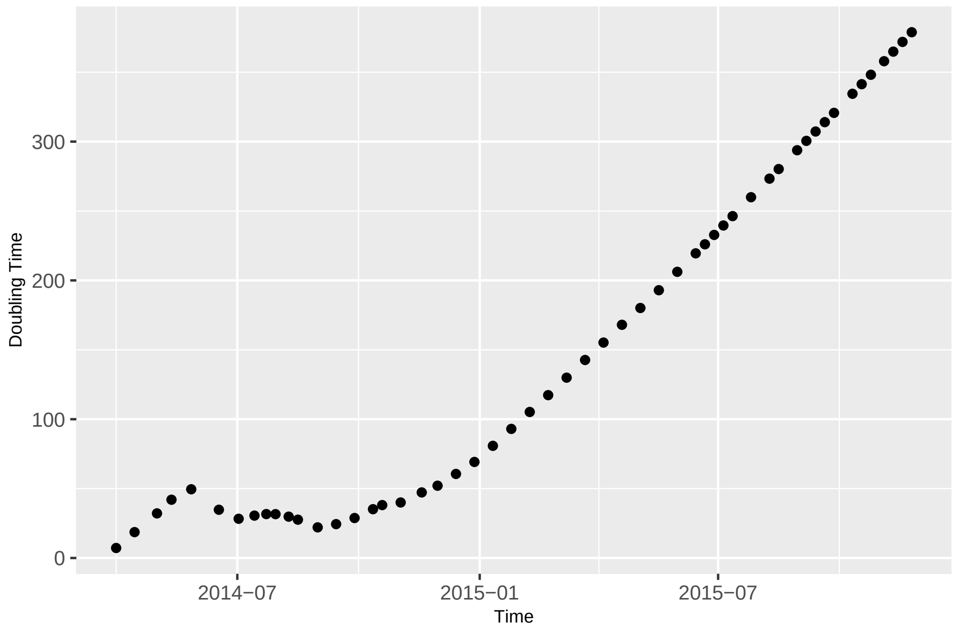


Figure 9: Above is plot of doubling time for Ebola Cases in Liberia, Sierra-Leone, and Guinea during April, 2014 to November, 2015. The data is sourced from [12]

For such a discrete data set in which cases are not reported at uniform intervals of time, the procedure described earlier to compute doubling times will apply but the formula will change. Consider the data set given by

$$\mathbf{P}_x = \{(x(k), P(k)) : k = 0, 1, 2, \ldots, T\} \tag{13}$$

with $x(k) \in [0, \infty)$ and $x(k) < x(k+1)$ for each $k = 0, 1, 2, \ldots$. Thus the sequence x can be used to specify the non-uniformity of the time intervals for the data set. For example, if we were working with the snap shot above as the data set then we could specify $x(0) = 0, x(1) = 7, x(2) = 17, x(3) = 24, x(4) = 39, x(5) = 53$.

Let $\mathbf{P}_x \equiv P$. We will now construct a linear interpolating function for P, namely $\tilde{P} : [0, x(T)] \to \mathbb{R}$, given by

$$\tilde{P}(t) = P(k) + \frac{P(k+1) - P(k)}{x(k+1) - x(k)}(t - x(k)) \tag{14}$$

for $t \in [x(k), x(k+1))$ for $k = 0, 1, 2, \ldots, T-1$. Now, for any $t_0 \in \mathcal{D}_{\text{disc}}$, we first find the smallest $k_0 < t_0$ such that

$$P(k_0) \leq \frac{P(t_0)}{2} \leq P(k_0 + 1)$$

Using Definiton 2.1 and Equation 14, we have to find $t \in [x(k_0), x(k_0 + 1))$ for which

$$P(k_0) + \frac{P(k_0 + 1) - P(k_0)}{x(k_0 + 1) - x(k_0)}(t - x(k_0)) = \frac{\tilde{P}(t_0)}{2} \tag{15}$$

Solving for t gives us

$$t = \frac{1}{2}\frac{\tilde{P}(t_0) - 2P(k_0)}{P(k_0 + 1) - P(k_0)}(x(k_0 + 1) - x(k_0)) + x(k_0)$$

Thus, doubling time at t_0 is given by

$$D(t_0) = t_0 - x(k_0) - \frac{1}{2}\frac{P(t_0) - 2P(k_0)}{P(k_0 + 1) - P(k_0)}(x(k_0 + 1) - x(k_0)) \tag{16}$$

We implement this method to calculate doubling time for the Ebola epidemic of $2013 - 16$ for which discrete data is given at non-uniform intervals of time for the years $2014 - 15$.

2.4 Application to the SIR Model

In this section we will study an epidemic model proposed by Kermack and McKendrick in 1927. Here the entire population can be categorized into three states- Susceptible, which refers to healthy individuals at a risk of contracting the disease; Infected, which refers to the people infected with the disease; and Recovered, which refers to people who have already contracted and then recovered from the disease. At any time instant, an individual of the population falls in one and only one of the above three categories. It is assumed that recovery confers complete immunity from the disease.

We also assume the total population to be fixed at N for all time $t \geq 0$. For each $t \geq 0$, we denote the number of susceptible at time t by $S(t)$, infected at time t by $I(t)$ and recovered at time t by $R(t)$. Let susceptible fraction, infected fraction and recovered fraction be denoted by

$$s(t) = \frac{S(t)}{N}, i(t) = \frac{I(t)}{N}, \text{ and } r(t) = \frac{R(t)}{N}.$$

We will assume no migration or birth of new individuals in the population. Hence, $S(t)$ decreases when a susceptible individual gets infected. The decrease depends on $S(t)$, the number of individuals already infected $I(t)$, and the amount of contact between susceptible and infected individuals. We shall assume that each infected individual has a fixed number β of contacts per day with other individuals. If we assume sufficient mixing of the population, the fraction of these contacts that are with susceptibles is $s(t)$. Thus per day, on average, each infected individual infects $\beta s(t)$ individuals.

$$\frac{dS}{dt}(t) = -\beta s I$$

Using the definition of s we have,

$$\frac{dS}{dt}(t) = -\frac{\beta}{N} S I \tag{17}$$

We also assume that a fixed fraction ν of the infected group will recover during any given day.

$$\frac{dr}{dt}(t) = \nu i$$

or equivalently,

$$\frac{dR}{dt}(t) = \nu I \tag{18}$$

As $S(t) + I(t) + R(t) = N$ we have

$$\frac{dS}{dt}(t) + \frac{dI}{dt}(t) + \frac{dR}{dt}(t) = 0.$$

From (17) and (18) we obtain

$$\frac{dI}{dt}(t) = \frac{\beta}{N} S(t) I(t) - \nu I(t) \tag{19}$$

The coupled system of ordinary differential equations (17), (18), and (19) detail the dynamics of $S, I,$ and R model. It is not possible to solve explicitly for $S(t), I(t)$ and $R(t)$. An approximate solution can be found in [9]. Otherwise one would need to use numerical methods to solve the system. We use the *deSolve* package in *R* software to solve for S, I, R. The detailed methodology used in the package is given in [10].

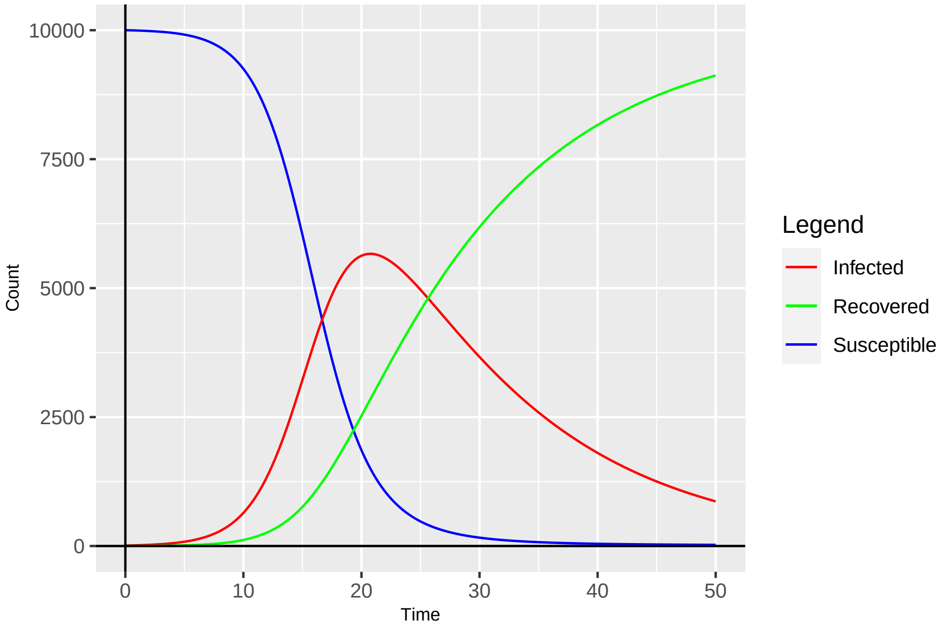


Figure 10: We have plotted the Infected, Susceptible and Recovered Timeline for the SIR Model 1:- $S(0) = 10000, I(0) = 10, R(0) = 1, \beta = 0.50055$ and time duration is for 50 days.

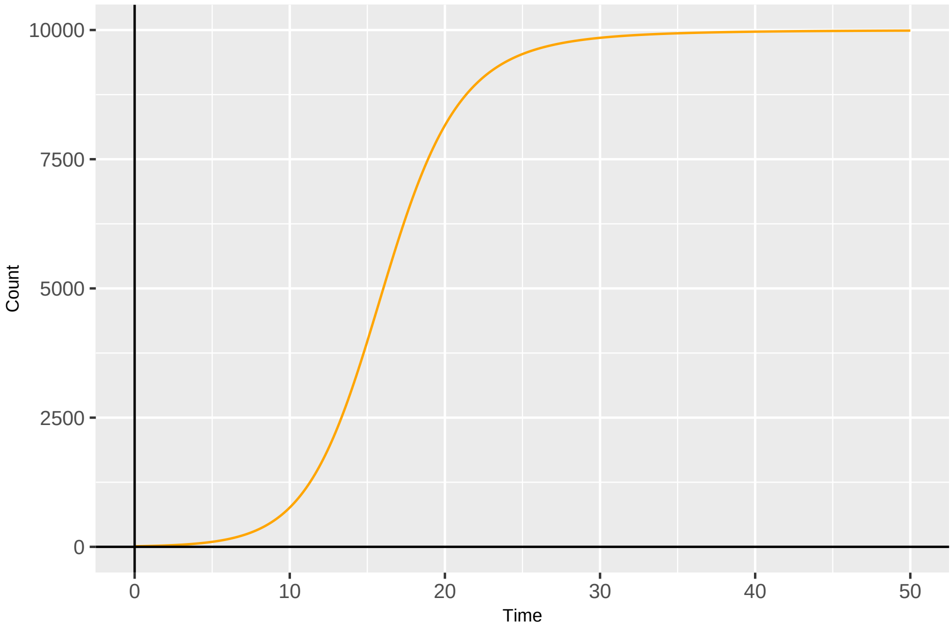
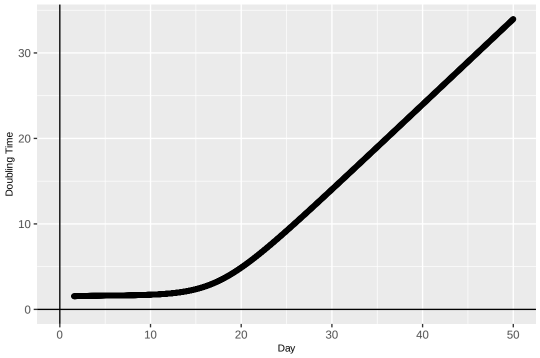


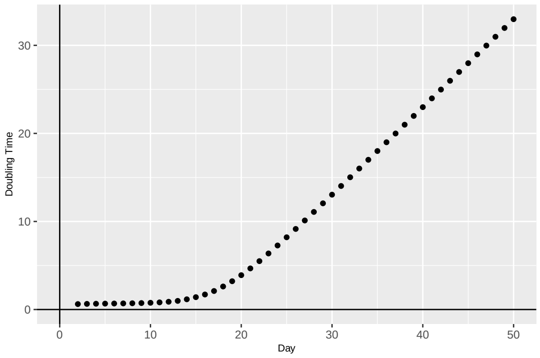
Figure 11: The above plot is of the Total infections timeline for SIR Model 1. That is a plot of $I + R$ vs time. Note that the plot which looks very similar to a logistic curve (See Figure 4).

Now, we use the method decribed earlier for estimating doubling time for discrete data and use the discrete points of the two SIR models to find doubling time. We have calculated doubling time for SIR Model 1 in two ways, using two different time scales. The respective plots are given in the figure below.

Given the doubling time values for an SIR curve, one can estimate the values of β and ν. However,



(a) We assume that 5000 data points are given to us, corresponding to tim $t = 0, 0.01, 0.02, \ldots, 49.99, 50$.



(b) We assume 51 data points are given to us, corresponding to time $t = 0, 1, 2, \ldots, 49, 50$.

the details are beyond the scope of this note. In the Figure 10, we note that the blue line, representing the susceptible people, becomes quite close to zero towards the end, which means that almost the entire population contracted the disease. However, this is not true for all SIR models, as seen below.

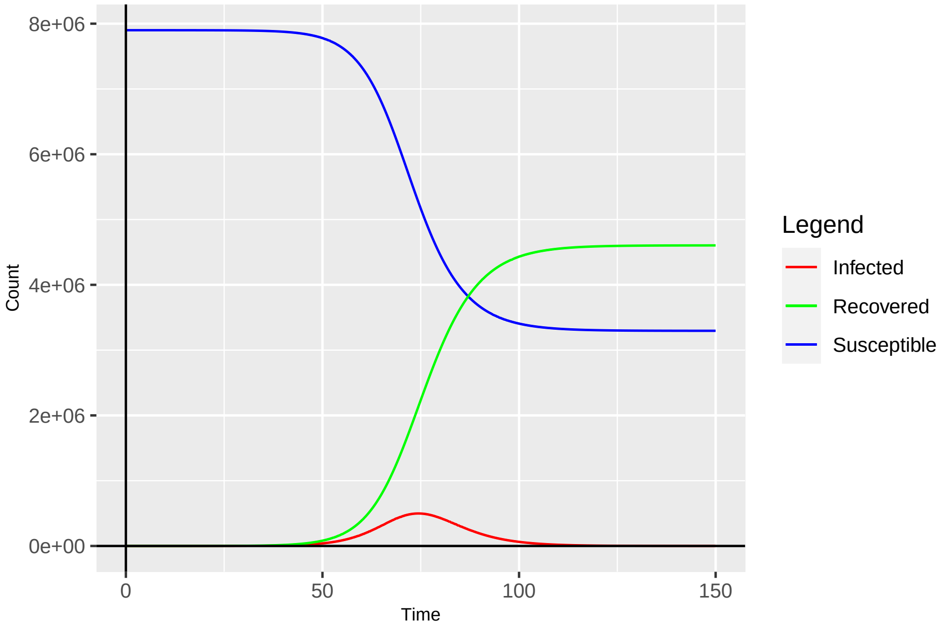


Figure 12: We have plotted the Infected, Susceptible and Recovered Timeline for the SIR Model 2- $S(0) = 7900000, I(0) = 10, R(0) = 0, \beta = 0.5$ and time duration is for 150 days

In Figure 12, due to a very small infection rate β and a relatively higher recovery rate of $\nu = \frac{1}{3}$, we see that only almost half of the population is infected and the peak of the $I(t)$ curve is quite small. When we plot the total infected people, ie, $I(t) + R(t)$ vs time, we will obtain a plot similar to Figure 11. The doubling time graphs will look similar as well.

3 Application to States and Union Territories of India

We now look at the COVID-19 infection timeline of India and some states in India. The data has been taken from the official website of the Union Ministry of Health and Family Welfare, of the Government of India, where the number of positive cases has been given on a daily basis from 10th March onwards, for all of India as well as the states and union territories with at least one infection. The doubling time graphs in this section look quite different from the ones plotted above for various mathematical functions, due to intervention measures adopted to prevent the spread of infection. Intervention measures include social distancing, imposition of lockdown and isolation of infected patients. In India, lockdown was imposed in stages, with restrictions varying in each stage. The first intervention measure was a 14-hour curfew on 22 March from 7 a.m. to 9 p.m., known as the Janata Curfew. Subsequently, lockdown was imposed in four phases, which lasted for-

- Phase 1 -25 March to 14 April
- Phase 2 -15 April to 3 May
- Phase 3 -4 May to 17 May
- Phase 4 -18 May to 31 May

All the doubling time graphs have points color-coded according to the various lockdown phases. This helps us assess the changes in doubling time due to intervention measures and the relaxation in them. We have used the linear interpolation technique for discrete data as described in section 2.1. Moreover, graphs have been plotted from the first date on which the infection count surpassed 50 and the number of infections was at least twice as that on 10th March. Hence, different graphs have different starting dates.

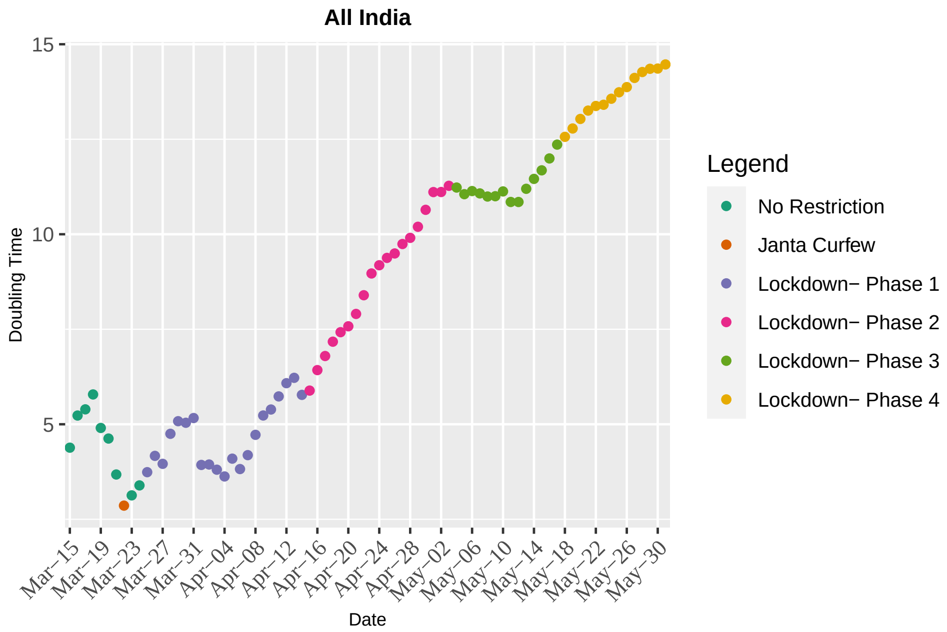


Figure 13: The doubling time is clustered around 5 during Lockdown Phase 1. In Lockdown Phase 2, the doubling time steadily increased from 5.88 on 15th April to 11.27 on 3rd May. The doubling time remained almost constant (around 11) till May 12 and then increased to 12.35 on May 17. Hence Lockdown Phase 3 shows a small increase in doubling time. The doubling time for India continued to increase in Lockdown Phase 4 and it reached 14.46 on May 31.

The clustered doubling times in Lockdown Phase 1 was expected because the effects of intervention measures are typically visible two to three weeks after they are imposed. The increase in Lockdown Phase 2 suggests a succesful lockdown in which the spread of infection was curbed to some extent. The increase in doubling time in Lockdown Phase 4 was an interesting phenomenon. It is in contrast with the doubling time of many states which dropped in Lockdown Phase 4 due to inter-state travel allowance. This suggests that travel relaxations resulted in a mere re-assignment of COVID-19 cases in the country.

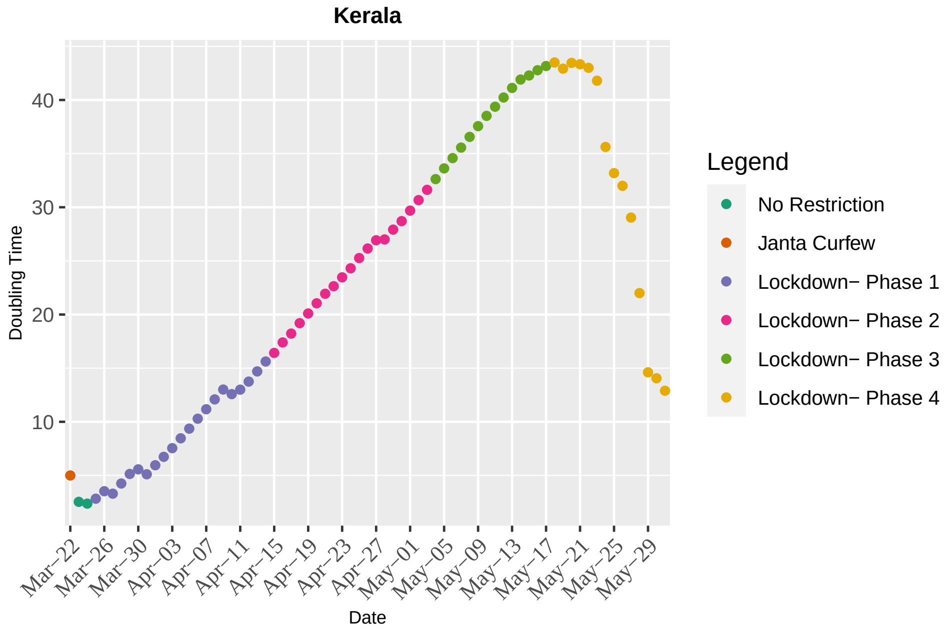


Figure 14: Many Indian states have witnessed a sudden spike in the number of cases in the fourth phase of the lockdown due to the influx of infected individuals from other states. In Kerala, doubling times roughly increased 23rd March onwards. However, in Lockdown Phase 4, the doubling time fell from 41.8 (23rd May) to 14 (30th May).

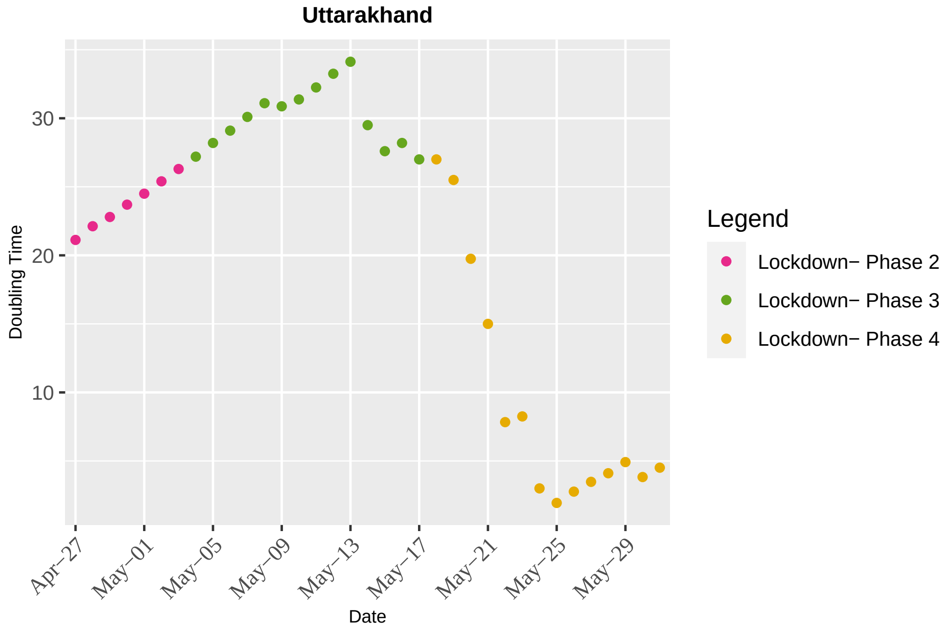


Figure 15: Another state which showed a drastic drop in doubling time in Lockdown Phase 4 is Uttarakahnd. On 13th May, the doubling time was around 34. It started decreasing after that and reached an all-time low of 1.93 on 25th May.

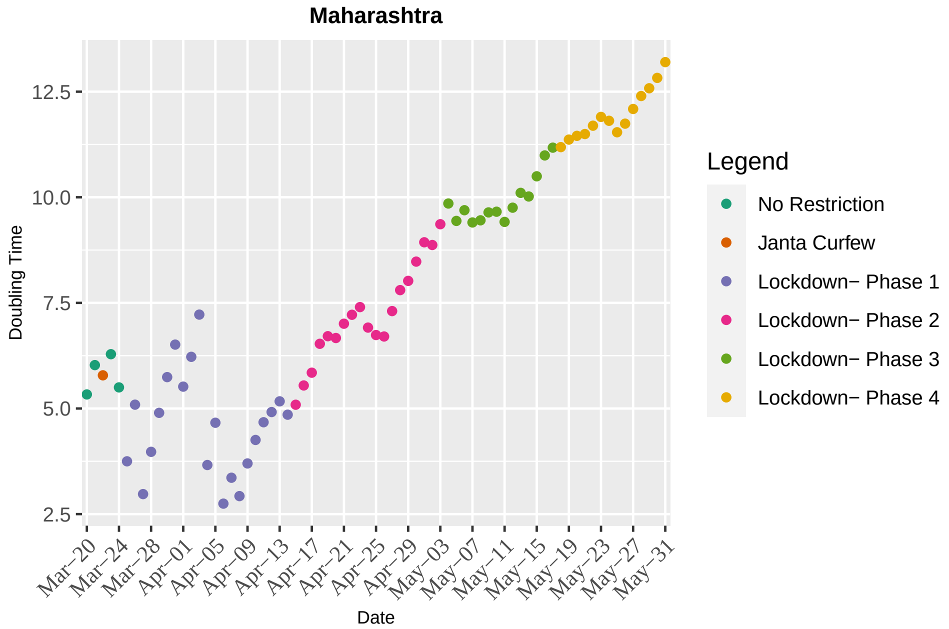


Figure 16: The doubling times for Maharashtra were scattered in the no restrictions, Janta Curfew and Lockdown Phase 1 periods. Since then they have been growing steadily, which suggests that the lockdown has been effective in the state. Even in Lockdown Phase 4, where many states witnessed a sudden spike in cases, Maharashtra showed an increase in doubling time. This suggests that Maharashtra does not have many cases imported from other states.

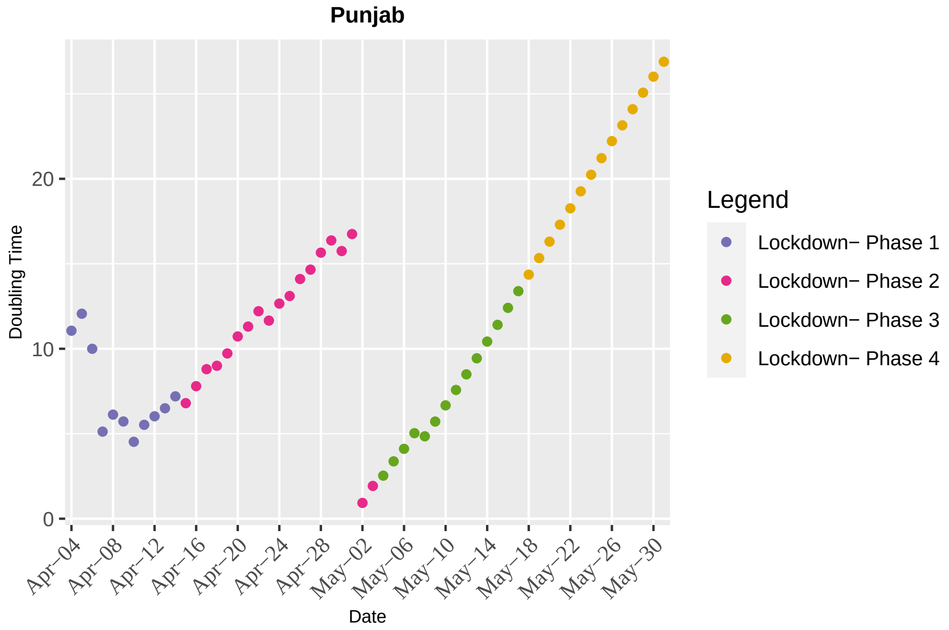


Figure 17: Punjab showed a sudden drop in doubling time towards the end of Lockdown Phase 2. On 1st May, the doubling time was around 16.5, but it dipped to 0.93 the next day. This suggests a tremendous increase in cases in a very short duration of time. However, the doubling times have been increasing since then.

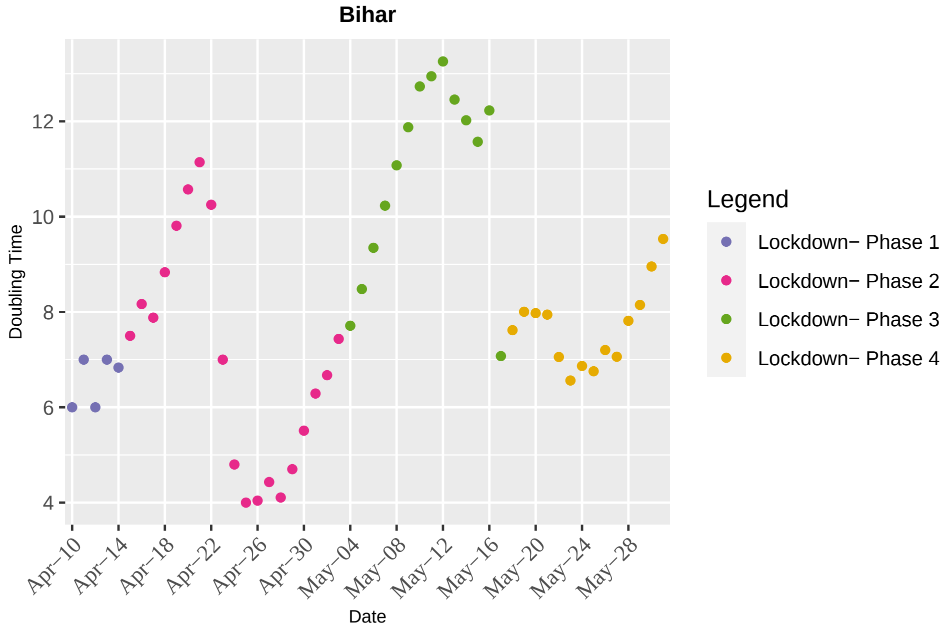


Figure 18: The state of Bihar shows no steady trend in doubling time in any intervention period. In Lockdown Phase 2, the doubling time roughly increased till 21st April, then decreased till 25th April and then roughly increased till 3rd May. In Lockdown Phase 3, the doubling time increased in the beginning, but decreased after 12th May. In Lockdown Phase 4 too, the doubling times are scattered around the horizontal line of $D(t) = 8$.

References

[1] Deepayan Sarkar, Doubling times of COVID-19 cases. https://deepayan.github.io/covid-19/doubling

[2] Siva Athreya, Nitya Gadhiwala, and Abhiti Mishra, COVID−19 India-Timeline an understanding across States and Union Territories., *Ongoing Study at http://www.isibang.ac.in/~athreya/incovid19*, (2020).

[3] John-Hopkins COVID-19 Dashboard- https://coronavirus.jhu.edu/map.html

[4] Bernoulli, D. (1760) Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir, Mém. Math. Phys. Acad. Roy. Sci., Paris (1760) 1–45.

[5] Britton, Tom. Stochastic epidemic models: a survey. *Mathematical biosciences*, 2010, 225.1: 24-35.

[6] Greenwood, Priscilla E and Gordillo, Luis F, Stochastic epidemic modeling, in Mathematical and statistical estimation approaches in epidemiology, Springer (2009) 31–52.

[7] Galvani AP, Lei X, Jewell NP. Severe acute respiratory syndrome: temporal stability and geographic variation in case-fatality rates and doubling times. Emerg Infect Dis. 2003;9(8):991-994. doi:10.3201/eid0908.030334

[8] Ward MP, Maftei D, Apostu C, Suru A. Estimation of the basic reproductive number (R0) for epidemic, highly pathogenic avian influenza subtype H5N1 spread. Epidemiology and Infection. 2009;137(2):219-226. doi:10.1017/S0950268808000885

[9] Kermack, W. O. and McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond. A*, **115**, (1927) 700-721.

[10] Soetaert K., Petzoldt T., and Setzer R.W. Solving Differential Equations in R: Package deSolve. *Journal of Statistical Software*, 33(9), (2010), 1

[11] Union Ministry of Health and Family Welfare, https://www.mohfw.gov.in/

[12] West African Ebola virus epidemic timeline of reported cases and deaths, *Wikipedia*, https://en.wikipedia.org/wiki/West_African_Ebola_virus_epidemic_timeline_of_reported_cases_and_deaths