# Experiment 8

*Author:* Prasad Fidelis D'sa       *Email:* prasadfidelisdsa.191mt032@nitk.edu.in

# 1   Introduction

This week is about modelling real-world data, the growth, stabilisation and saturation of epidemics and studying effects such as virulence and social distancing. We primarily focus on the SARS-CoV-2 virus. All the code for this exercise and the relevant files are included in the `.zip` file submitted along with this report.

# 2   Problems

**1. First Order Model**

**(Part 1: Transfer Function of First Order Model)**

*(Solution)*

**The Basic Reproduction Number $R_o$**

The basic reproduction number $R_o$ is used to measure the transmission potential of a disease. It is the average number of secondary infections produced by a typical case of an infection in a population where everyone is susceptible. For example, if the $R_o$ for a viral infection in a population is 15, then we would expect each new case to produce 15 new secondary cases (assuming everyone around the case was susceptible).

The basic reproductive number is affected by several factors:

1. The rate of contacts in the host population

2. The probability of infection being transmitted during contact

3. The duration of infectiousness

In general, for an epidemic to occur in a susceptible population $R_o > 1$, so the number of cases is increasing. $R_o$ can be a valuable concept, however, the process of defining, calculating, interpreting, and applying $R_o$ is far from straightforward. The simplicity of an $R_o$ value and its corresponding interpretation in relation to infectious disease dynamics masks the complicated

nature of this metric. Although $R_o$ is a biological reality, this value is usually estimated with complex mathematical models developed using various sets of assumptions. Here we present an interpretation of $R_o$ in terms of Digital Signal Processing.

In signal processing terms, the infection mechanism is equivalent to a first-order recursive filter. The assumption we make here is that an initial patient zero appears at day $n = 0$, hence the input is the Kronecker delta function $\delta(n)$ and each infected person spreads the virus over a single day and then recovers. Then the number of newly infected people per day is described by the difference equation

$$y(n) = \delta(n) + R_o y(n-1) \tag{2.1}$$

As $\delta(n) = 0$ for $n > 0$, Equation 2.1 tells us that for $n > 0$, the number of daily new cases $y(n)$ is effectively $R_o$ times the number of new cases of the previous day $y(n-1)$. The transfer function of the system in Equation 2.1, which is also the transfer function of the impulse response of the system is given as

$$H_1(z) = \frac{1}{1 - R_o z^{-1}} \tag{2.2}$$

The system clearly has a zero at $z = 0$ and a pole at $z = R_o$. By definition $R_o$ is real and positive, hence the pole lies on the real axis. This gives us three cases of the Pole-Zero plot corresponding to $R_o < 1$, $R_o = 1$ and $R_o > 1$ as shown in Figure 1. The value of $R_o$ completely decides the behavior of the system, a notion we discuss in Part 2.
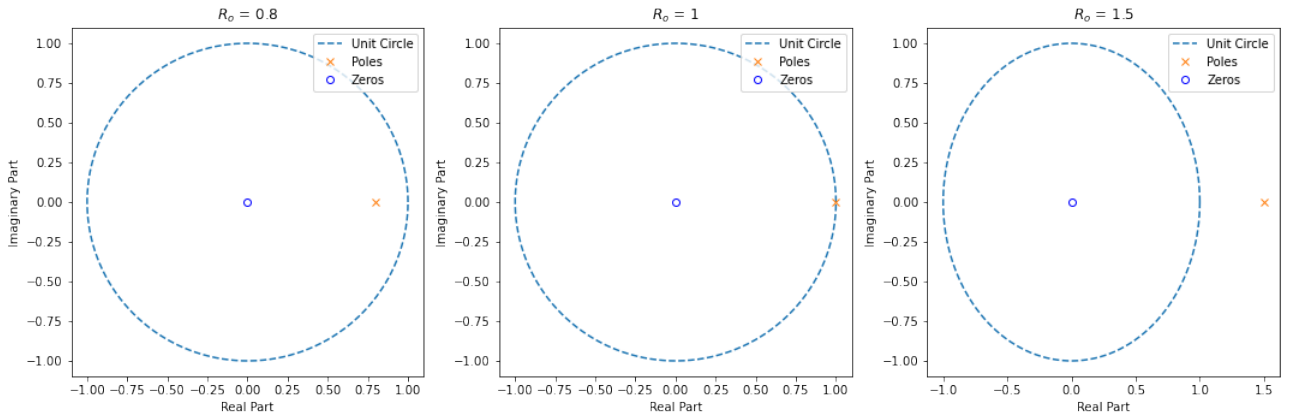


Figure 1: The Pole-Zero plot for different values of $R_o$

**(Part 2: Effect of $R_o$)**
*(Solution)*
Consulting the $z$-transform table we find the inverse $z$-transform of Equation 2.2 as

$$h(n) = R_o^n u(n) \tag{2.3}$$

As our input is $\delta(n)$, $y(n) = h(n)$. Hence for $n \geq 0$, the time domain equation for the number of newly infected people

$$y(n) = R_o^n \tag{2.4}$$

Hence the daily new cases is an exponential function of $R_o$ and completely depends on the value of $R_o$,

- If $R_o < 1$ the daily new cases $y(n) \to 0$ as $n \to \infty$

- If $R_o = 1$, then $y(n) = y(n-1)$ as $n \to \infty$. The number of new cases remain constant

- If $R_o > 1$, then $y(n) \to \infty$ blows up as $n \to \infty$

Revisiting the Pole-Zero plot of Figure 1 and our sense of system stability in the $z$-domain, $R_o < 1$ lies inside the unit circle and hence the impulse response of the system decays to zero as $n \to \infty$ representing a stable system. $R_o = 1$ lies on the unit circle and for a first order system represents a marginally stable system. $R_o > 1$ lies outside the unit circle and represents an unstable system. Hence the impulse response grows without bounds as $n \to \infty$. Figure 2 shows the impulse response of the system for the three cases.
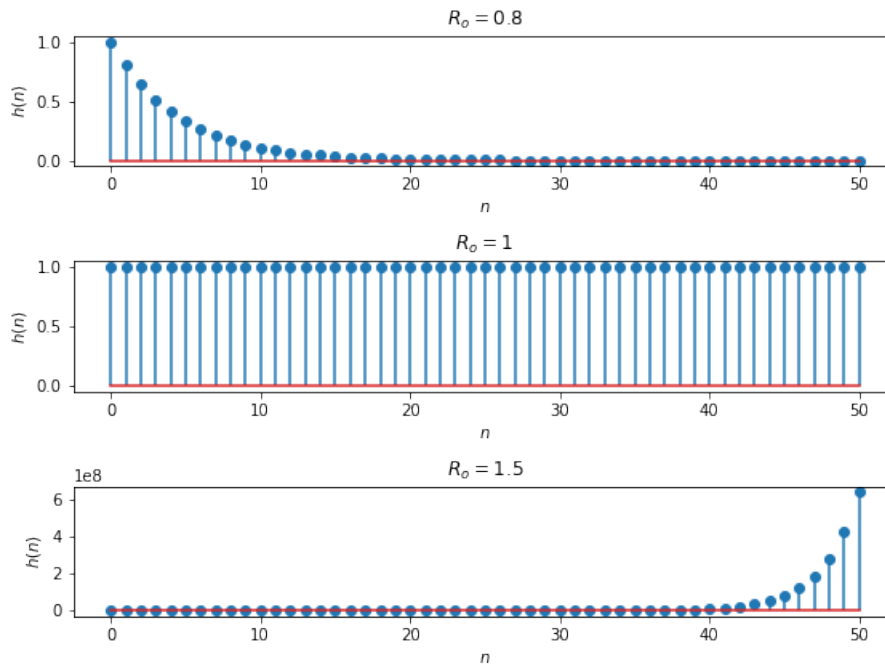


Figure 2: Impulse response for different values of $R_o$

Hence the stability of the system inferred from the Pole-Zero describes the behaviour of the number of new daily infections.

**(Part 3: Determining value of $n$)**
*(Solution)*

Given the value of $R_o$, we can use Equation 2.4 to determine the number of days it takes to reach a particular number of daily new infections. Hence for $R_o = 2.5$, substituting $y(n) = 1000000$ in Equation 2.4 gives $n = 15.0776$. Hence it takes about $n = 16$ days for the daily new infections to reach 1 million.

**(Part 4: Estimating $R_o$)**
*(Solution)*
India had its first case of Covid-19 on 19 March 2020. 182 days later the number of daily new cases were 3809 on 17 September 2020. Substituting these values in Equation 2.4 we get $R_o = 1.0463$. This method is not accurate because it does not consider the clusters in the population and assumes that the members of the population mix homogeneously and transmission and removal rates are constant. A more reliable method will be taking the ratio of the number of new infections generated at time step t, to the total infectiousness of infected individuals at time t or using Bayesian statistical inference assuming a gamma prior distribution for R.

**(Part 5: Total number of infections)**
*(Solution)*
It is useful to know the total number of infections $t(n)$ in the country since the first day along with the daily new cases. This operation can be represented by the recursive difference equation

$$t(n) = y(n) + t(n - 1) \tag{2.5}$$

The transfer function of the impulse response of Equation 2.5 is given as

$$H_t(z) = \frac{1}{1 - z^{-1}} \tag{2.6}$$

The coefficients of Equation 2.6 can be given as input to `sp.lfilter` to obtain the total number of cases for the first $n$ days.
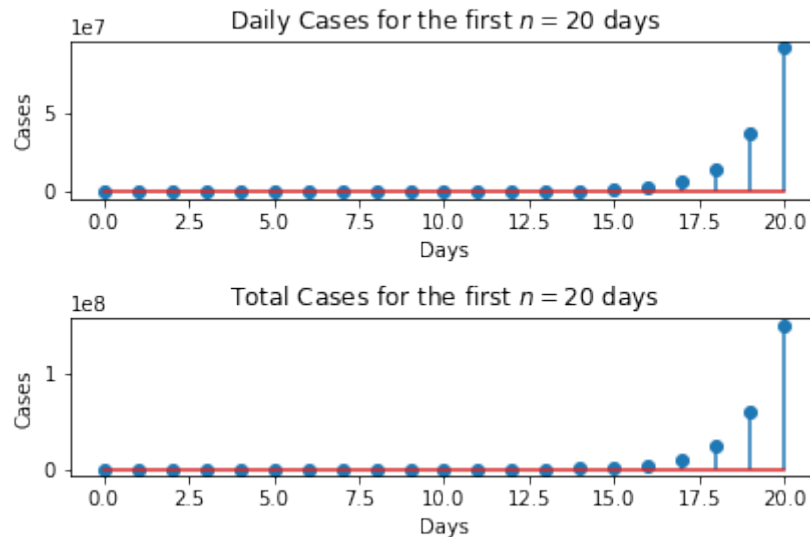
Figure 3: Daily new and cumulative infections for first $n = 20$ days for $R_o = 2.5$

We get $t(20) = 151582449.6288$. Hence the total number of cases for the first $n = 20$ days is approximately 151.6 million.

**2. Increasing the Complexity**

**(Part 1: Accounting for Virulence)**

***(Solution)***

While the First Order model of Equation 2.1 maybe fairly accurate in the initial phase of the pandemic, our assumption that each infected person spreads the virus over a single day and then recovers is unrealistic in practice as the actual infectiousness of a virus is not concentrated in a single day but follows a curve that rises and falls with the duration of the sickness as shown in Figure 4.
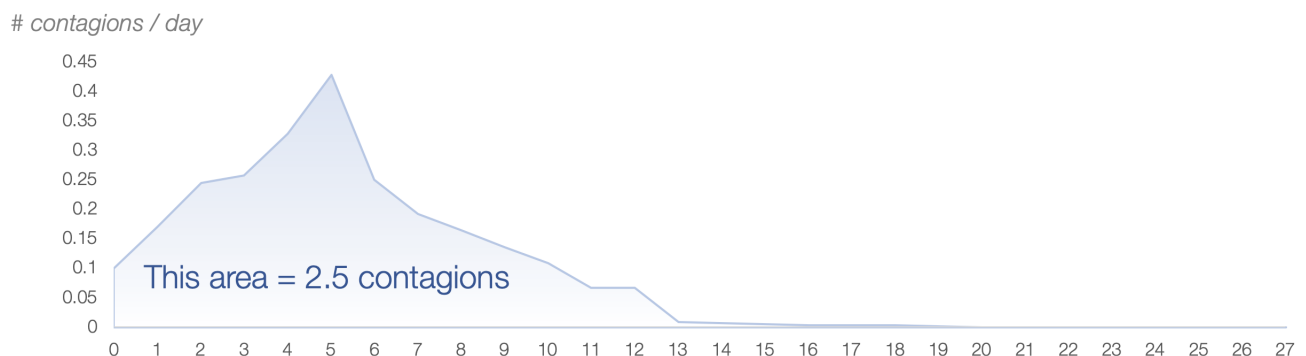


Figure 4: Virulence of SARS-CoV-2

We observe that the virulence decreases from day-7 onwards and becomes negligible after day-

14. This scientifically substantiates the 14-day quarantine rule. From a DSP point of view, we account for this fact by modifying our system to a multi-pole IIR filter with a transfer function

$$H_M(z) = \frac{1}{1 - \sum_{k=1}^{M} a_k z^{-k}} \tag{2.7}$$

Hence our recursive system equation becomes

$$y(n) = \delta(n) + \sum_{k=1}^{M} a_k y(n-k) \tag{2.8}$$

For $M = 12$, the coefficients $(a_k)_{k=1}^{M}$ are the daily values that can be obtained from Figure 4 and are given by $\{.1, .15, .25, .26, .34, .42, .25, .2, .15, .1, .1, .1\}$. Figure 5 shows the Pole-Zero plot of the system. We can observe that the system is still not stable as there is a pole outside the unit circle.
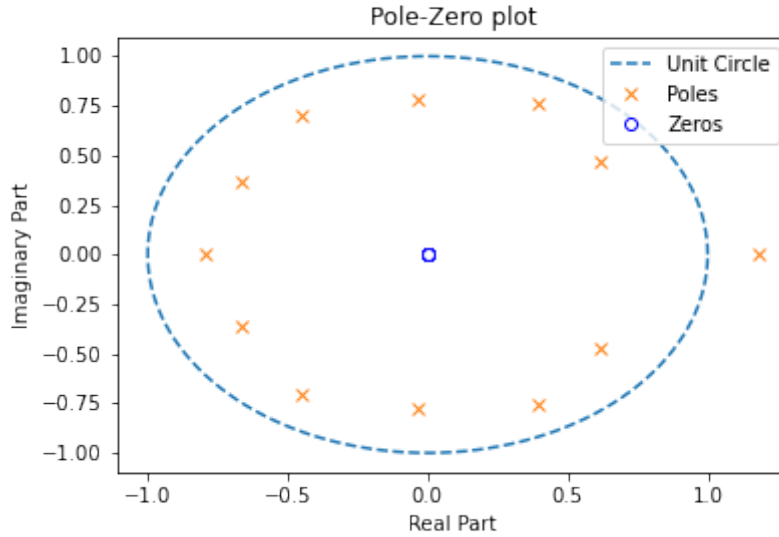


Figure 5: Pole-Zero plot of the Multi-pole IIR filter

Figure 6 shows the new daily and total infections for the first $n = 100$ days by implementing the filter given in Equation 2.7 with the Kronecker delta as the input and integrator filter of Equation 2.6 with the new daily cases as input respectively.
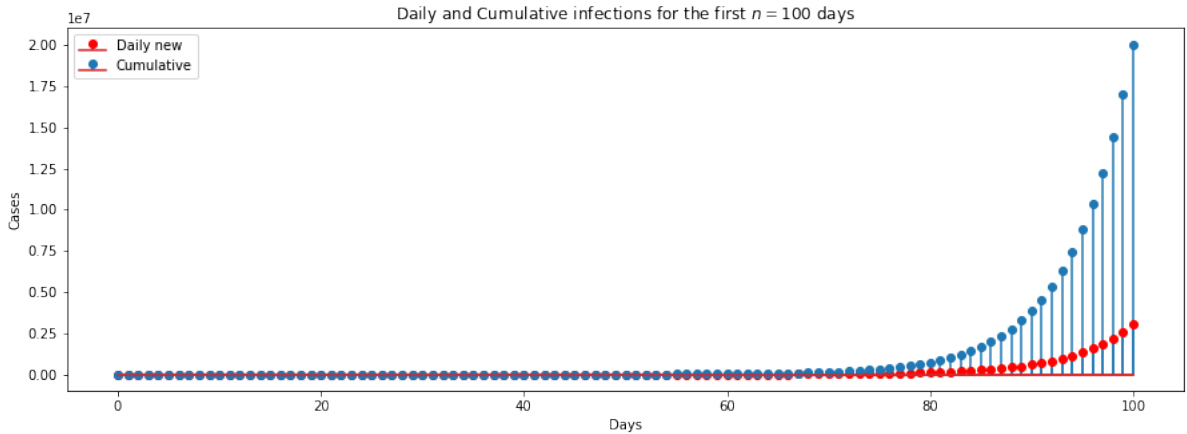
Figure 6: Daily new and Cumulative cases for the multi-pole IIR filter

We get $t(100) = 20044208.412$. Hence the total number of cases for $n = 100$ days is about 20 million.

**(Part 2: Determining $n$ for 1 million daily cases)**
*(Solution)*

In order to determine the value of $n$ for 1 million daily cases here, we loop through each element and break when we reach the first element over 1 million. Hence we get $n = 94$. Therefore it takes $n = 94$ days for the number of new daily cases to reach 1 million in contrast to $n = 16$ days with our first order model. The first order model grows significantly faster than the multi-pole IIR filter. This is because the first order model assumes $R_o = 2.5$ over a single day but if we observe Figure 4 the effective $R_o = 2.5$ is spread over 12 days. Another way of looking at it is in the Pole-Zero plot of Figure 5, the value of the pole that causes instability is lesser than 2.5 and hence the growth of new cases is slower.

**(Part 3: Estimating coefficients)**
*(Solution)*

In order to estimate the coefficients we can take the DFT of the cases and observe the frequency characteristics. This will allow us to set the limits for the IIR filter which can be then designed to get the appropriate coefficients.

**3. Effects of Social Distancing**

**(Part 1: Role of $\rho$)**
*(Solution)*

Social distancing is a one of the measures to curb the spread of a virus among a population. It is based on limiting the number of interactions and hence viral transmissions among individuals. To account for social distancing in our model we will use a parameter $\rho \in [0, 1]$. $\rho$ is a measure

of social distancing or in other words reduction of the interaction of every individual. $\rho = 0$ indicates no social distancing and hence normal interaction among individuals and $\rho = 1$ indicates perfect social distancing where no individual interacts with any other. In essence it is the fraction of interactions that have been reduced. This means that the number of interactions after implementing social distancing will be proportional to $1 - \rho$. The implementation of $\rho$ in our model is predicated on the fact that the coefficients $(a_k)_{k=1}^M$ are dependent on the number of encounters between a sick individual and healthy people. As the new interactions are proportional to $1 - \rho$, we proportionally scale the filter coefficients by $\rho - 1$ to get the transfer function

$$H_M(z, \rho) = \frac{1}{1 - \sum_{k=1}^M (1 - \rho)a_k z^{-k}} \tag{2.9}$$

and our difference equation becomes

$$y(n) = \delta(n) + \sum_{k=1}^M (1 - \rho)a_k y(n - k) \tag{2.10}$$

If we set $\rho = 1$ in Equation 2.10, we get $y(n) = \delta(n)$. This means that for perfect social distancing, the virus will not spread and patient zero will be the only case at $n = 0$. This is true because the patient will not interact and hence infect anybody else irrespective of the virulence of the disease and therefore there will be no other new cases for $n > 0$.

**(Part 2: Infections for different $\rho$ values)**
***(Solution)***
We now implement the filter in Equation 2.9 with the Kronecker delta as the input for first $n = 100$ days for $\rho = 0.25, 0.5, 0.75$. We then use the integrator filter of Equation 2.6 to get the total cases for the first $n = 100$ days. Figure 7 shows the number of daily new cases and cumulative cases different values of $\rho$.
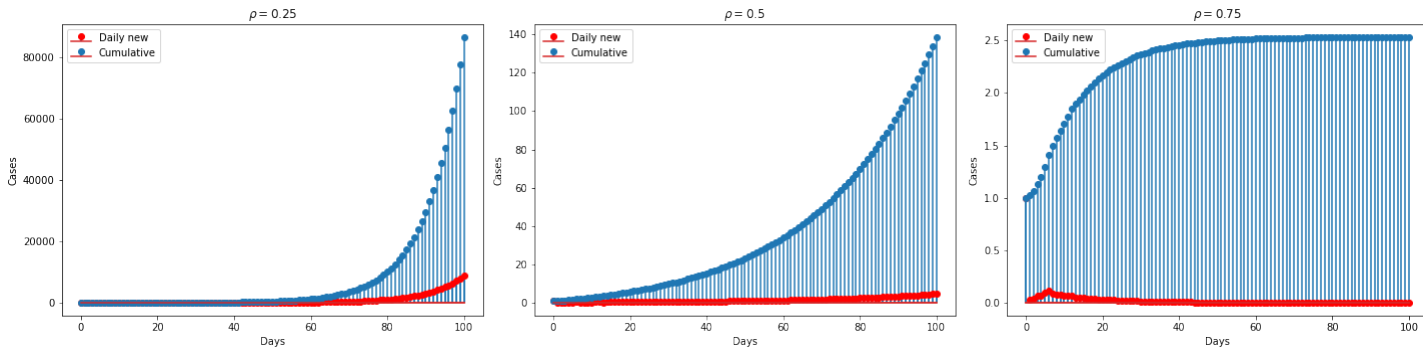


Figure 7: Daily new and Cumulative cases for different values of $\rho$

For $\rho = 0.25$, we get 86548 total cases. For $\rho = 0.5$, we get 139 total cases and for $\rho = 0.75$ we

get about 3 total cases.

**(Part 3: Effectiveness of Social Distancing)**

*(Solution)*

It is clear from Figure 7 that social distancing is very effective in controlling the spread of the virus as it reduces the total number of infections. For $\rho = 0.25$ even though the total cases are unbounded, if we compare it to Figure 6, the number of cases have greatly reduced. Hence social distancing reduces rate of exponential growth of the virus, giving authorities time to arrange health care facilities and other essentials. Same goes for $\rho = 0.5$, however, an interesting thing occurs when $\rho = 0.75$ the number of cases begin to decrease after a certain point and tend to zero. This can be seen in the saturation of total curve indicating that no new cases are being added to the total. We can understand this from the Pole-Zero plot. Whenever we multiply the coefficients with $(1 - \rho)$ it brings the pole on the real axis causing instability nearer to the unit circle and hence closer to stability. For a certain $\rho$ the pole lies on the unit circle and for any value of $\rho$ greater than this, the system will become stable and its impulse response and hence the new daily cases will tend to zero. Thus we can conclude that social distancing is not just effective in reducing the number of daily cases but if implemented well can even bring an end to the pandemic.

## 4. Saturation and Towards Normality

**(Part 1: Logistic Evolution)**

*(Solution)*

All the models discussed so far are reasonable in the initial phase of the pandemic when everyone around the infected individual was susceptible. However no natural system can support a purely exponential growth. In the case of a viral epidemic, as more and more people contract the disease and achieve immunity, the rate of transmission for the infection progressively decreases and the behaviour of new cases deviates from the exponential behaviour. Moreover as the number of cases becomes comparable to population size, if we stick to the exponential model there is a good chance the predicted number of cases goes beyond the population size rendering the exponential model unreliable. Thus it is assumed that the rate of diffusion is inversely proportional to the fraction of healthy people in a population. Then the evolution of the cumulative number of infections since the beginning of the disease can be modeled by a logistic function

$$x(n) = \frac{K}{1 + [K(R_o - 1) - R_o]R_o^{-(n+1)}} - \frac{1}{R_o - 1} \tag{2.11}$$

Where $K$ is the population size. Figure 8 shows the total cases according to Equation 2.11 and our first order model of Equation 2.2 for a population $K = 1$ million and $R_o = 1.15$.
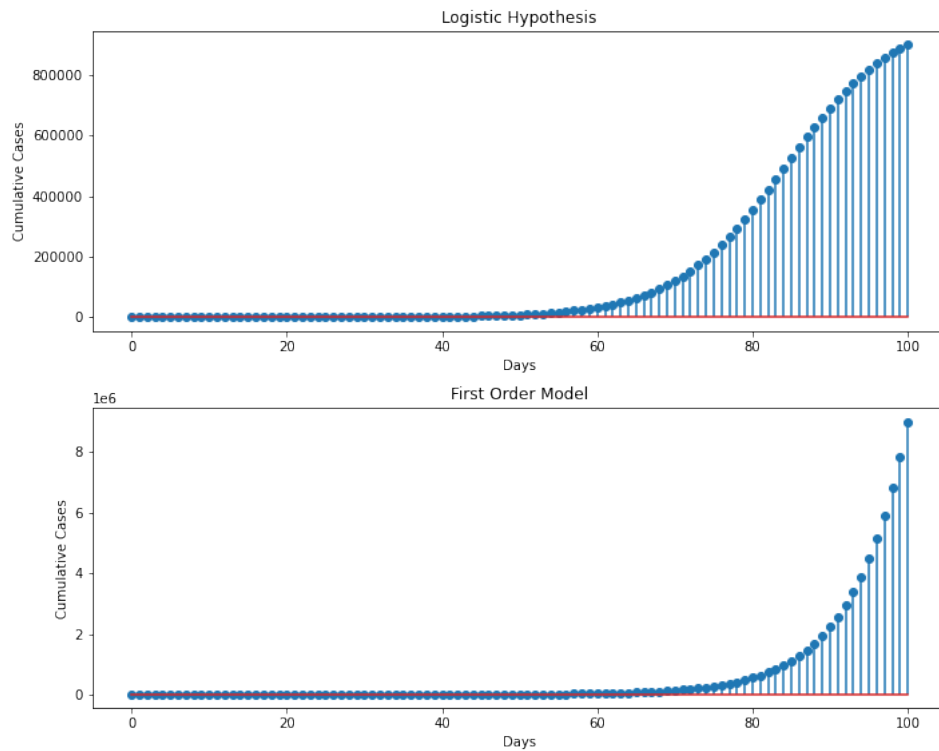
Figure 8: Cumulative cases for population of 1 million and $R_o = 1.15$

The logistic function plateaus at around 800000 while the first order model grows without bound and does not follow the logistic evolution.

The notable thing about the logistic function is that it has a clear inflection point, after which the epidemic starts to level out; this corresponds to the moment in which the implicit reproduction number becomes less than one. It is useful to detect the inflexion point as some of the restrictive measures can be relaxed and people can start to prepare for normalcy gradually.

**(Part 2: Finding the Inflexion Point )**
*(Solution)*
The logistic sigmoid function grows initially at a slow pace, then accelerates to a maximum and then slows down again to saturation. The point at which the growth rate is maximum is the point of inflexion. Hence the inflection point corresponds to the global maximum of the first derivative of the logistic function. Another way to look at it is that the direction of acceleration changes from positive to negative. Thus we can look directly at the zero-crossing of the second derivative.
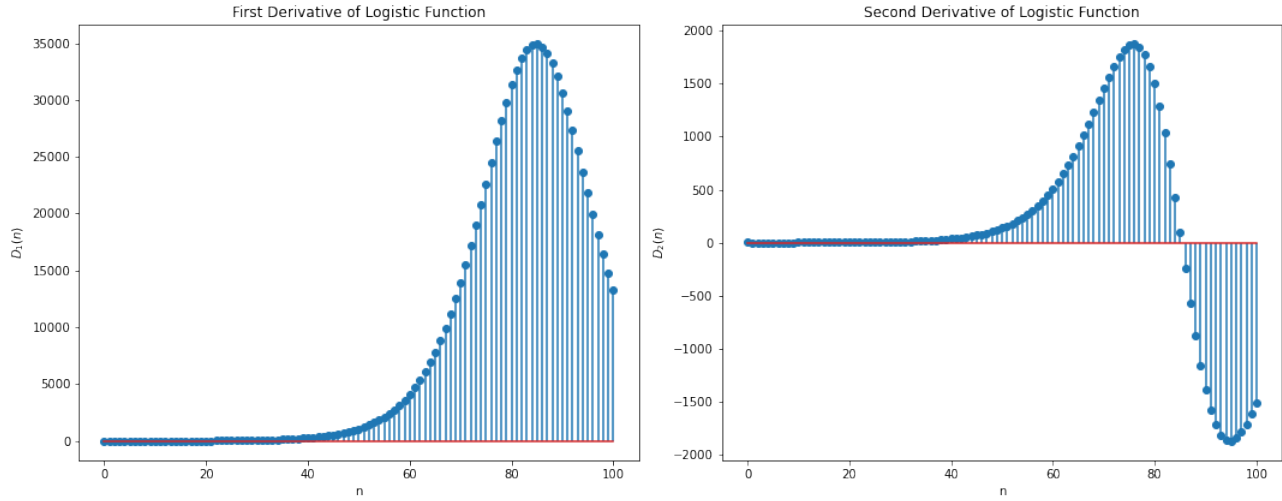
Figure 9: Approximate first and second derivative of logistic function using FIR filters

We can approximate the derivative with a simple two-tap FIR filter of the form

$$D_1(z) = 1 - z^{-1} \tag{2.12}$$

And the second derivative can be approximated with the filter

$$D_2(z) = 1 - 2z^{-1} + z^2 \tag{2.13}$$

Figure 9 shows the first and second derivative of the logistic function obtained from Equation 2.12 and 2.13 respectively by taking the logistic function as input. Utilizing the method described above we get the point of inflexion as $n = 85$, meaning 85 days after the first case, the logistic function reaches its inflexion point.