Math 2208: Ordinary Differential Equations

Project 4: The SIR model and COVID-19

Spring 2020 | **Subhadip Chowdhury** | **Due: April 29**

On December 31, 2019, the Chinese city of Wuhan reported an outbreak of a novel coronavirus (**COVID-19**) that has since killed over **170,000** people. As of April 20, 2020, over **2,480,000** infections - spanning **210** countries - have been confirmed by the World Health Organization (WHO). In this project, we try to understand the infectious disease models, exploring how the WHO and other groups are characterizing and forecasting the COVID-19 pandemic.

§A. What is the SIR model?

SIR model is a **compartmental disease model** describing the dynamics of infectious diseases. The letters in **SIR** represent the three compartments of the total population:

- **Susceptible** - Susceptible individuals have no immunity to the disease (immunity can come from prior exposure, vaccination, or a mutation that confers resistance). Susceptible individuals can move into the Infectious compartment through contact with an infectious person.
- **Infectious** - The group of infectious represents people who can pass the disease to susceptible people and can recover after a specific period.
- **Removed** - People who recover from the disease get immunity so that they are not susceptible to the same illness anymore. Many SIR-based models assume that a recovered person remains immune (which is often appropriate if immunity is long-lasting, e.g., chicken pox or the disease is being modeled over a relatively short time period). As a matter of convenience, we include the group of people who do not recover but die in the Removed group -- since they too can no longer contract the disease.

Because people can move between compartments, the number of people in each compartment changes over time. The SIR model captures population changes in each compartment with a system of ordinary differential equations (ODEs) to model the progression of a disease.

As the first step in the modeling process, we identify the independent and dependent variables. The independent variable is time t, measured in days. We are going to make the following simplifying assumptions:

Assumption I.

Assume that the total population size $\mathbf{N}(t)$ is a constant. This is reasonable if the city is on lockdown. We also do not consider the effect of the natural death or birth rate because the model assumes the outstanding period of the disease to be much shorter than the average lifetime of a human.

In our closed population of **N** individuals, say that **S** are susceptible, **I** infectious, and **R** are recovered. Let

$$s = \frac{\mathbf{S}}{\mathbf{N}}, \quad i = \frac{\mathbf{I}}{\mathbf{N}}, \quad r = \frac{\mathbf{R}}{\mathbf{N}}$$

denote the fraction in each compartment.

■ Question 1. (0+1 points)

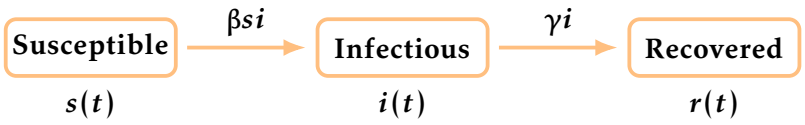
(a) Under the assumptions we have made, how do you expect $s(t)$ to vary with time? What about $r(t)$ and $i(t)$? Sketch on a piece of paper what you think the graph of each of these functions looks like. Do not spend too much time on it, we will draw the graphs using a computer later on.

(b) Explain why, at each time t, we have

$$s(t) + i(t) + r(t) = 1.$$

The complete SIR model is given by the following three dimensional system of ODEs:

$$\begin{aligned} \frac{ds}{dt} &= -\beta si \\ \frac{di}{dt} &= \beta si - \gamma i \\ \frac{dr}{dt} &= \gamma i \end{aligned}$$



Below we will explain how to derive each of the three equations.

Assumption II.

We assume that the population is well-mixed. This means any infectious individual has a constant probability of contacting any susceptible individual. This is often the most problematic assumption, but is easily relaxed in more complex models by taking averages.

Assumption III.

We assume that the time-rate of change of $\mathbf{S}(t)$, the number of susceptibles, depends on the number currently susceptible, the number of individuals currently infectious, the amount of contact between susceptibles and infectious, and the infectivity of the disease. In particular, suppose that each infectious individual has a fixed number of contacts per day and each contact has a fixed probability to transmit the disease. Not all these contacts are with susceptible individuals. On average, each infectious individual generates $\beta s(t)$ new infectious individuals per day.

■ Question 2. (The Susceptible Equation, (1+1+1) points)

The rate of change of $\mathbf{S}$ over time is given by $\frac{d\mathbf{S}}{dt} = -\beta s(t)\mathbf{I}(t)$. Explain carefully how each component of the differential equation follows from assumption III.

(a) Why is the factor of $\mathbf{I}(t)$ present?

(b) Where did the negative sign come from?

(c) Explain how this leads to the equation

$$\frac{ds}{dt} = -\beta si$$

Assumption IV.

Infectious individuals are assumed to recover with a constant probability at any time, which translates into a constant per capita recovery rate that we denote with γ. For example, if the average duration of infection is three days, then, on average, one-third of the currently infectious population recovers each day.

Question 3. (The Recovered Equation, 1 point)

Explain how the corresponding differential equation for $r(t)$,

$$\frac{dr}{dt} = \gamma i$$

follows from the last assumption.

Question 4. (The Infectious Equation, 1 point)

Explain why

$$\frac{di}{dt} = \beta si - \gamma i$$

What assumption about the model does this reflect?

§B. Qualitative/Numerical Approach

Download the Octave file `solve_sir.m`. The file draws the graphs of $s(t), i(t),$ and $r(t)$ vs. t with $\beta = 0.5$ and $\gamma = 0.2$.

Question 5. (2+2+1 points)

(a) Assume $i(0) = 10^{-6}$ and $s(0) = 1 - i(0)$, that is one in a **1000000** person is infectious (these are the default values in the Octave file). What can you say about the long term values of $i(t)$ and $s(t)$? What percent of population never contract the disease?

(b) What is the maximum value of $i(t)$? Find the approximate value of t when it happens.

(c) Attach a screenshot of the picture.

Solution. (a) As $t \to \infty, s \to 0.1$ and $i \to 0$. About **10%** of the population never contract the disease.

(b) The maximum value of $i(t)$ is approximately **0.23** with $t = 47.5$.

■

We are going to focus our experimentation on the infectious-fraction, $i(t)$, since that function tells us about the progress of the epidemic.

- Comment out the commands in your `Octave` file that correspond to plotting $s(t)$ and $r(t)$.
- Change the legend to `'i'` only.
- Comment the `clf;` command.
- Uncomment the lines 66 and 67.

We are interested in finding the effects of the parameters β and γ on the solution curve $i(t)$ vs. t.

■ Question 6. (2+2+1) points

First let's experiment with changes in β when γ is fixed at **0.2**.

(a) Plot the graph of $i(t)$ with β values $0.5, 0.7, 0.9, \ldots, 1.5$. Make sure to overlay the consecutive graphs on a single figure. Describe how these changes affect the graph of $i(t)$.

(b) Attach a screenshot of the picture.

(c) Explain briefly why the changes you see are reasonable from your intuitive understanding of the epidemic model.

Solution. (a)

(b)

(c) Increase in β means the the transmissibility and contact rate increases, which increases the infectivity of the disease. That means the rate at which people will be infected from a infectious individual in given amount of time is **faster**, so the peak in $i(t)$ happens earlier and is steeper.

■

■ Question 7. (2+2+1+2)

Now let's experiment with changes in γ when β is fixed to **0.5**. Change the code in line 67 so that it prints γ instead of β.

(a) Plot the graph of $i(t)$ with γ values $0.2, 0.3, 0.4, 0.5, 0.6$. Make sure to overlay the consecutive graphs. Describe how these changes affect the graph of $i(t)$.

(b) Attach a screenshot of the picture.

(c) Explain the changes you see in terms of your intuitive understanding of the model.

(d) There is a change in the behavior of the $i(t)$ graph for a certain value of γ in the given range. What is the change, and where does it occur? Can you find the exact value analytically?

[Hint: How does the sign of $\frac{di}{dt}$ change?]

Solution. (a)

(b)

(c)

(d) The change happens when $\gamma = 0.5$. Note that at $t = 0$, the rate $\frac{di}{dt}$ changes from positive to negative when

$$(\beta s(0) - \gamma) i(0) = 0 \implies \beta = \gamma$$

since $s(0) \approx 1$. When that happens the graph of $i(t)$ decreases from the very beginning, starting with a very small initial value. So it looks almost flat.

■

§C. Basic Reproduction Number R_0

Definition 3.1

The **basic reproduction number**, $\mathbf{R_0}$, also known as the **contact number**, is defined as the expected number of secondary cases produced by a single infection in a completely susceptible population.

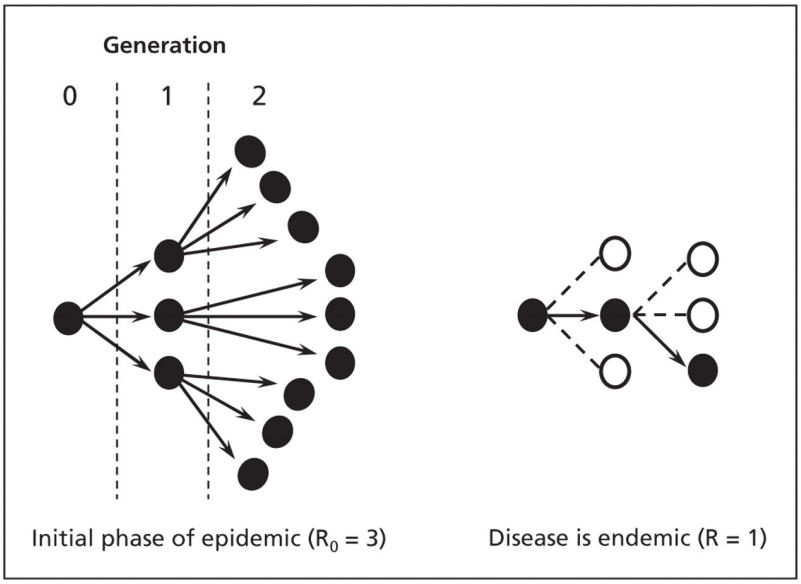


Figure 1: $\mathbf{R_0 > 1}$ is epidemic, $\mathbf{R_0 = 1}$ is endemic

$\mathbf{R_0}$ is a combined characteristic of the population and of the disease, it measures the relative contagiousness of a disease. It is important to understand that $\mathbf{R_0}$ is a dimensionless unitless number and not a rate. We can define

$$\mathbf{R_0} = \underbrace{\underbrace{\left(\frac{\text{infection}}{\text{contact}}\right)}_{\text{transmissibility of disease}} \times \underbrace{\left(\frac{\text{contact}}{\text{time}}\right)}_{\text{average rate of contact}}}_{\beta} \times \underbrace{\underbrace{\left(\frac{\text{time}}{\text{infection}}\right)}_{\text{duration of infectiousness}}}_{\frac{1}{\gamma}}$$

$$= \beta \times \frac{1}{\gamma}$$

We have used assumption IV here: γ is roughly equal to the reciprocal of the number of days an individual is sick enough to infect others. Although γ is directly observable from patients, there is no direct way to observe β. Fortunately, there is an indirect way of calculating $\mathbf{R_0}$ that in turn gives us β.

§D. Analytical Approach

Note that the first two ODEs in the SIR system can be treated as a **2D** system by themselves, since they do not invlove a r variable. Then once we find $i(t)$ and $s(t)$, we can use that to find $r(t)$ (how?). For a **2D** system, we can use PPLANE to draw the (s, i) phase portrait with s as the horizontal axis and i as the vertical axis. You only need to look at the range $[0,1] \times [0,1]$ (Strictly speaking, you can only look at the region $s + i \leq 1$ since $s + i + r = 1$).

■ Question 8. (2+2+2 points)

(a) Find the equations of the nullclines and the equilibrium points.

(b) Explain why for an epidemic to occur (i.e. $i(t)$ increases from its initial value) we must have $s(0) > \dfrac{1}{R_0}$. This number $\frac{1}{R_0}$ is consequently called the **threshold value** of the model.

(c) The point $(1, 0)$ is called the Disease Free Equilibrium (DFE). Linearize the system at the DFE and explain why $i(t)$ looks almost exponential for small values of t.

Solution. (a) The s–nullclines are $s = 0$ and $i = 0$. The i–nullclines are $i = 0$ and $s = \dfrac{\gamma}{\beta}$.

(b) Because if $s(0) \leq \frac{1}{R_0} = \frac{\gamma}{\beta}$, then $\frac{di}{dt} \leq 0$, which means people recover faster than being infected. Hence, the epidemic does not occur.

(c) The Jacobian is given by $\begin{bmatrix} -\beta i & -\beta s \\ \beta i & \beta s - \gamma \end{bmatrix}\Big|_{(1,0)} = \begin{bmatrix} 0 & -\beta \\ 0 & \beta - \gamma \end{bmatrix}$. So the linearized ODEs (which is an approximation near $t = 0$), become

$$\frac{ds}{dt} = -\beta s, \qquad \frac{di}{dt} = (\beta - \gamma) i$$

So in particular, the second ODE solves to $i = e^{(\beta-\gamma)t}$, an exponential function.

■

■ Question 9. (2+2+2 points)

(a) Find $\frac{di}{ds}$ and check that $i(s)$ has the general formula

$$i(s) = -s + \frac{\ln s}{R_0} + c$$

where c is some arbitrary constant.

(b) Since the initial conditions are $(s(0), i(0)) \approx (1, 0)$, we can assume $c = 1$ (why?). Use this to show that

$$R_0 = \lim_{t \to \infty} \left(\frac{\ln s}{s - 1} \right)$$

[Hint: Take limit as $t \to \infty$ on both sides of the $i(s)$ equation.]

(c) Find the numerical value of $\lim\limits_{t \to \infty} s(t)$ from PPLANE for the solution curve with same initial values as in the Octave code. Use this to calculate above limit and consequently R_0. Your answer should be approximately equal to $\frac{\beta}{\gamma} = \frac{5}{2}$.

Solution. (a) $\dfrac{di}{ds} = \dfrac{di/dt}{ds/dt} = \dfrac{\beta s i - \gamma i}{-\beta s i} = -1 + \dfrac{\gamma}{\beta s}$. This is a separable differential equation that solves to the given equation.

(b) Substitute $s = 1, i = 0$ in part (a) to get $c = 1$. Then note that as $t \to \infty$, we get $i \to 0$. Take limit on both sides of above equation, then rearrange and isolate R_0.

(c) The approximate $\lim\limits_{t \to \infty} s(t)$ is **0.10811.** Hence we have

$$R_0 = \lim_{t \to \infty} \left(\frac{\ln 0.10811}{0.10811 - 1} \right) \approx 2.49$$

■

■ **Question 10. (Optional, 4 points)**

Assume $i(0) > 0$. If $R_0 < 1$, on average, an infectious person infects less than one person. I.e. the disease is expected to stop spreading. If $R_0 > 1$, show that $\lim_{t\to\infty} s(t) = L$ exists and is strictly larger than 0.

The conclusion we can draw from this analysis is that, in general, a fraction of the population will always escape infection. This is one of the fundamental insights of mathematical theory of epidemics.

§E. Flattening the curve

■ **Question 11. (4+2 points)**

In our planetary response to COVID-19, we have come up with many different ways to reduce the R_0: via **social distancing**, via **quarantining the infectious**, or via **providing better treatment and healthcare**. Explain which of these correspond to changing β and which ones correspond to γ? Describe what the phrase ''flatten the curve'' means! Which curve are we talking about here? Why is it important?

Solution. The first two reduce β. The third one increases γ. To 'flatten the curve' means to flatten the $i(t)$ curve, specifically lowering the peak number of infected people at a time ($\max i(t)$) and delay the time of at which the peak happens. It is important because otherwise our healthcare system gets overwhelmed and some infectious people can't get treatment when they need it.

■

§F. COVID-19 (Optional)

Figure (2) gives an idea of how COVID-19 compares to other infectious diseases. Note that it is extremely hard to estimate R_0 accurately, the picture below only provides a possible upper and lower bound for the value.

The SIR model is fairly simplistic in nature and needs modification for complex diseases such as COVID-19. One of the more commonly used is the SEIR model, the E stands for **(Exposed)**. It takes into account the fact that some diseases (e.g. COVID-19) have a latent (or incubation) period, during which individuals have been infectious but are not yet infectious themselves (i.e. cannot infect others). Another approach is to use a SIQR model, where the Q stands for **(Quarantined)**. There are also models where β varies over time (you can find such an example at this link).

We are going to consider a SEAQRS model. The 'A' stands for **(Asymptomatic)**.

- When someone first contracts the disease they go from S to E.
- Once the incubation period is over, people move from E to I at a rate σ approximately equal to the reciprocal of the duration of incubation.
- We are breaking I in to two parts: people who show their symptoms go to Q, people who don't move to A. People can move from A to Q but not conversely.
- Both I and Q lead to R.
- Some part of R individuals return to S status due to loss of immunity.

I have provided a graphical representation above. There are more underlying assumptions as before that I am not going to list here. You are encouraged to think about what some of those could be to make this model more accurate. I have purposefully made the model a bit over-complicated to give you an idea of what a general model looks like. Some questions that you could consider are:

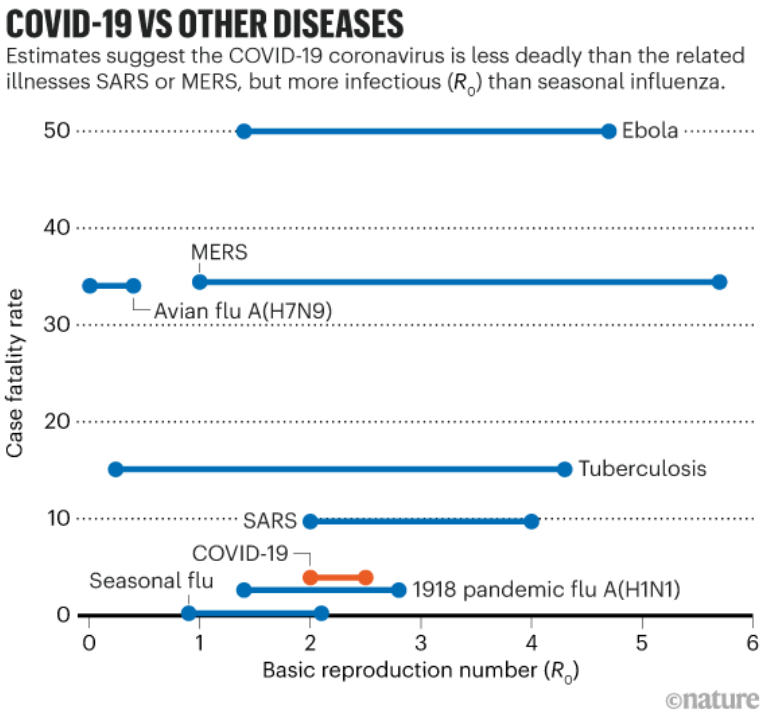
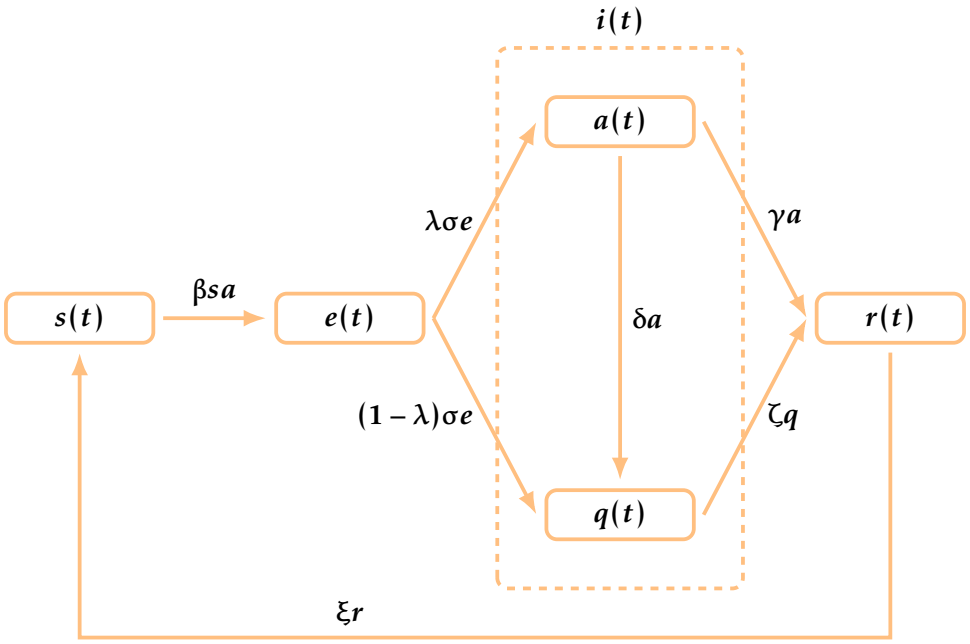


Figure 2: Source: Nature 579, 482-483 (2020)



- what does λ represent?
- why does the arrow between s and e says βsa?
- why is γ and ζ different? Note that ζ is usually bigger than γ.

Here are some stats for COVID-19. The average incubation period has been approximated to have a median of **5.1** days. So $\sigma \approx \frac{1}{5}$. According to a NYT article, $\lambda \approx \frac{1}{4}$. The majority of individuals that contract COVID-19 resolve symptoms within two weeks, so we can take $\zeta \approx \frac{1}{14}$.

■ Question 12. (Optional, 4 points)

Write down the system of ODEs corresponding to this model.

■ Question 13. (Optional, 10 points)

Use the Octave file to draw the $i(t)$ vs. t graph, where $i = a + q$. Discuss how the peak of i changes with respect to β and α. Use them to demonstrate the effectiveness of quarantine.