MATH 221 - Differential Equations

Project 2: Mathematical Epidemiology 101: The SIR Model and COVID-19

Fall 2020 **Subhadip Chowdhury** **Sep 25-29**

On December 31, 2019, the Chinese city of Wuhan reported an outbreak of a novel coronavirus (**COVID-19**) that has since killed over **979,000** people. As of Sep 23, 2020, over **32,000,000** infections* - spanning **215** countries [1] - 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?

Epidemic spread can be modeled by a system of differential equations. Such models include the **SIR** model and its variations. The 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. Note that it does not represent the *infected* people.
- **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.

We assume that at any given moment, a person must be in exactly one compartment. However, 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.

§B. Derivation of the Model

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 for example, a 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.

*__Interesting (macabre?) Note:__ When I wrote this project last semester, the data was as follows: ''...has since killed over **170,000** people. As of April 20, 2020, over **2,480,000** infections - spanning **210** countries...''

In our closed population of **N** individuals, say that **S** are susceptible, **I** are 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. **1 point**

Explain why, at each time t, we have

$$s(t) + i(t) + r(t) = 1. \tag{0}$$

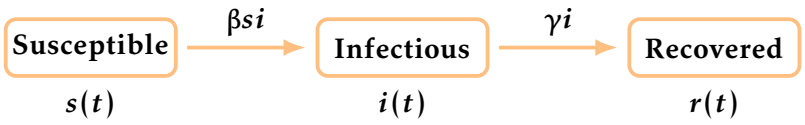
End of Question 1

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

$$\frac{ds}{dt} = -\beta si \tag{1}$$

$$\frac{di}{dt} = \beta si - \gamma i \tag{2}$$

$$\frac{dr}{dt} = \gamma i \tag{3}$$



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 four things:

- the number of individuals currently susceptible,
- the number of individuals currently infectious,
- the amount of contact between susceptibles and infectious, and
- the transmissibility 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. Let's assume that on average, each infectious individual generates $\beta s(t)$ new infectious individuals per day. The constant β depends on the last two factors.

■ 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 term in the differential equation follows from assumptions II and 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 (1)

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

End of Question 2

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 (3) for $r(t)$,

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

follows from the last assumption.

■ Question 4. The Infectious Equation, 1 point

Explain how we can use equations (1) and (3) together to conclude equation (2)

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

Which assumption about the model did you use to get this?

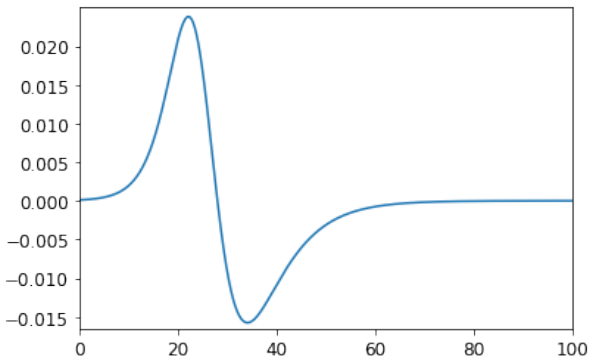
§C. Determining Outcomes: Graphical Representations

A common goal for modeling is to understand likely outcomes in the short term and the long term. These outcomes may be visualized via solutions to the differential equations, or via the differential equations directly. Before moving forward with further calculation, let's try to make some initial observations directly from the set of equations (1)-(3).

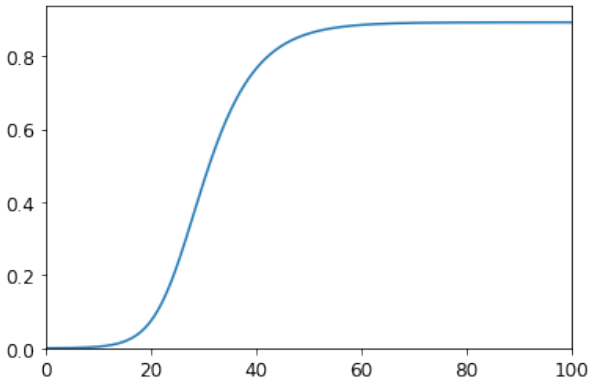
■ Question 5.

3 points

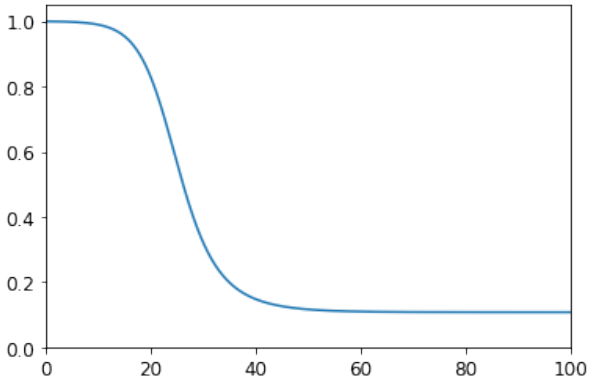
Below are six graphs from the SIR model for a theoretical outbreak of COVID in a population of **3000** people. At the start of the outbreak, there is one Infectious person, and everyone else is Susceptible. There is one graph for each of the following: $s(t)$, $i(t)$, $r(t)$, ds/dt, di/dt, and dr/dt. Which is which? How do you know? Consider shapes of graphs, values on the vertical axis, and other information you believe to be relevant. In particular, two graphs look nearly identical. One is a derivative graph (ds/dt, di/dt, or dr/dt) and one is a solution graph ($s(t)$, $i(t)$, or $r(t)$). Reason through which graph is which. Explain your logic. **You are not allowed to use any technology for this part.**



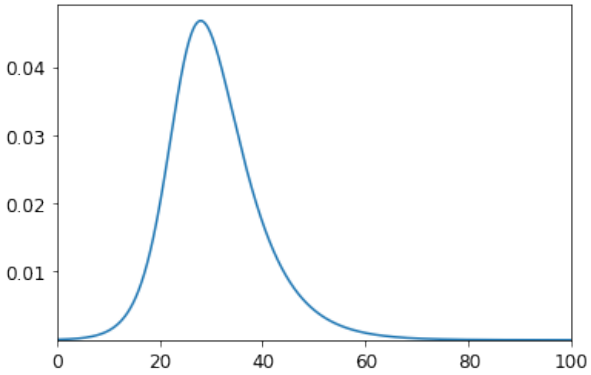
(a)



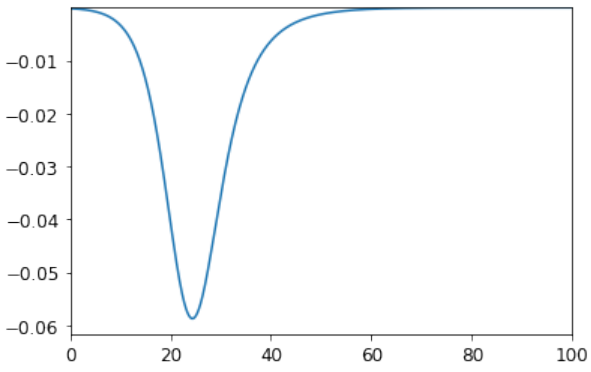
(b)



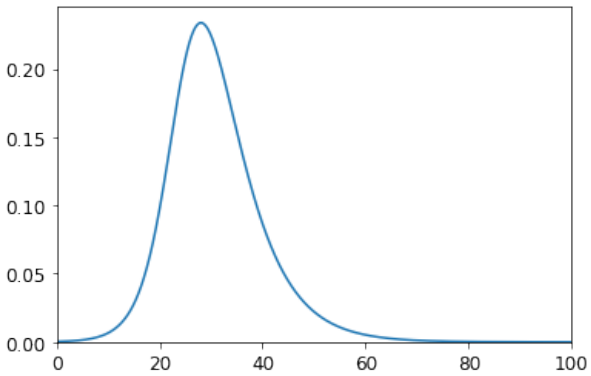
(c)



(d)



(e)



(f)

Figure 1

§D. Basic Reproduction Number R_0

Definition D.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.

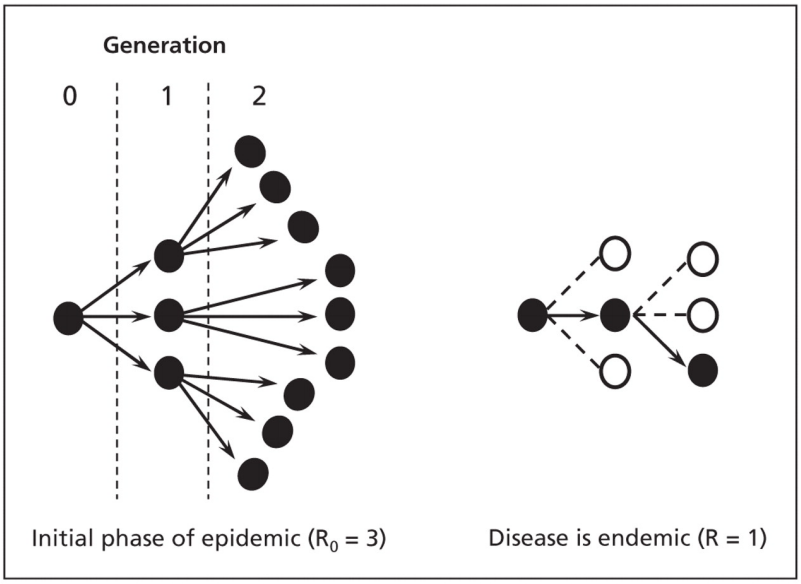


Figure 2: $\mathbf{R_0 > 1}$ is epidemic, $\mathbf{R_0 = 1}$ is endemic

$\mathbf{R_0}$ 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

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

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

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 doesn't require us to know β beforehand.

§E. Qualitative/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 $r(t)$ (using equation 0). 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. We only need to look at the range $[0,1]\times[0,1]$.[†]

■ Question 6. (2+2+2) points

For the two-dimensional system

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

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

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

■ Question 7. 4 points

Find a differential equation for $\frac{di}{ds}$ from equations (2) and (1). [**Hint:** $\frac{di}{ds} = \left(\frac{di}{dt}\right)\Big/\left(\frac{ds}{dt}\right)$.]

Solve it using separation of variables method and show that $i(s)$ has the general formula

$$i(s) = -s + \frac{\ln s}{R_0} + c \tag{4}$$

where c is some arbitrary constant. This is the equation of the general solution curve in the phase plane!

End of Question 7

Since the initial conditions are $(s(0), i(0)) \approx (1,0)$, we can use equation (4) to write

$$0 \approx -1 + \frac{\ln 1}{R_0} + c \implies c \approx 1$$

Now take limit as $t \to \infty$ on both sides of equation (4) and use the fact that $\lim_{t\to\infty} i = 0$ (why?) to get

$$0 = \lim_{t\to\infty}\left(-s + \frac{\ln s}{R_0} + 1\right) \implies R_0 = \lim_{t\to\infty}\left(\frac{\ln s}{s-1}\right) \tag{5}$$

Thus we can find the numerical value of R_0 by collecting data about $s_\infty = \lim_{t\to\infty} s(t)$ in real-life. For countries who have stabilised their COVID transmissions, this number can be found easily and we can consequently find approximate value of R_0. Unfortunately, in the USA we are not yet in the $t \to \infty$ part of the curve[2]; and so the R_0 value gets frequently updated.

■ Question 8. 2+1 points

Use $\beta = 0.5, \gamma = 0.2$ in your PPLANE phase portrait. Draw a solution curve that starts approximately near $(1,0)$.

Attach a picture of the phase portrait with the solution curve.

[†] Strictly speaking, we should only look at the region $s + i \le 1$ since $s + i + r = 1$.

Use the picture to estimate the value of s_∞. Use this value of s_∞ in the limit in equation (5) to find $\mathbf{R_0}$. Your answer should be approximately equal to $\frac{\beta}{\gamma}$.

End of Question 8

One of the most fascinating observation we can make about s_∞ from the phase portrait is that it's not equal to $\mathbf{0}$. Indeed, there is always a fraction of the population who never get infected! This is one of the fundamental insights of mathematical theory of epidemics.

■ Question 9. 2 points

Assume $i(0) > 0$ and $\mathbf{R_0} > 1$.[‡] Show that $s_\infty = \lim_{t\to\infty} s(t)$ is strictly larger than $\mathbf{0}$.

[**Hint:** First note that s_∞ has to be between $\mathbf{0}$ and $\mathbf{1}$ (why?), so it's positive. Why can't s_∞ be equal to zero?]

§F. Numerical/Qualitative Approach

You will need to save and attach your Python Output pictures for different questions of this section. So make sure to provide meaningful labels and titles in the pictures.

You can download and use the `ODE_2D_System.ipynb` file from Moodle as a reference. You will need to convert the code to be used for a 3D stystem. Alternately, you can use `PPLANE`.

Here is a neat trick: Copy all the pictures onto a single page of Word document and convert it into pdf!

■ Question 10. 6 points

Write a program that draws the graphs of $s(t), i(t),$ and $r(t)$ vs. t for $0 \le t \le 100$, either together or separately. You should use different color for each curve. Use the following values for the parameters and the initial conditions:

- $\beta = 0.5$ and $\gamma = 0.2$
- $s(0) = 2999/3000$ and $i(0) = 1/3000$, i.e. we are assuming that is one in a 3000 person is infectious at time $t = 0$.
- $r(0) = 0$

Attach the picture(s). They should look like some of the curves from question 5.

■ Question 11. (2+2+1 points)

Use your plots to answer the following questions.

(a) What are the long term ($t = 100$) approximate values of $i(t)$ and $s(t)$? How does the long term value of $s(t)$ compare to your answer from question (8)?

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

End of Question 11

For the rest of this section, we are going to focus our experimentation on the infectious-fraction, $i(t)$, since that function tells us about the progress of the epidemic. We are going to vary the parameters β and γ and find its effect on the solution curve $i(t)$ vs. t.

[‡] If $\mathbf{R_0} < 1$, on average, an infectious person infects less than one person. I.e. the disease is expected to stop spreading.

■ Question 12. (2+1) points

First let's experiment with changes in β when γ is fixed at $\mathbf{0.2}$.

(a) Plot the graphs[§] of $i(t)$ with β values $\mathbf{0.5, 0.7, 0.9, \ldots, 1.5}$. Describe how changing β affects the graph of $i(t)$.

(b) Explain briefly why the changes you see are reasonable from your intuitive understanding of how β affects the epidemic model.

(c) **(BONUS 2 points)** Modify the program to plot all the graphs for consecutive values of β in a single picture and attach it. All the required code can be found in the Jupyter notebooks for 2D Systems and for one-parameter family of ODEs.

■ Question 13. (2+1+2) points

Now let's experiment with changes in γ when β is fixed to $\mathbf{0.5}$.

(a) Plot the graphs[¶] of $i(t)$ with γ values $\mathbf{0.2, 0.3, 0.4, 0.5, 0.6}$. Describe how these changes affect the graph of $i(t)$.

(b) Explain the changes you see in terms of your intuitive understanding of how γ affects the model.

(c) 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? What happens to $\frac{di}{dt}$ and $i(t)$ when γ is bigger than this particular value? Use equation (2) to justify your answer.

(d) **(BONUS 2 points)** Modify the program to plot all the graphs for consecutive values of γ in a single picture and attach it.

§G. Flattening the curve

■ Question 14. 4 points

In our planetary response to COVID-19, we have come up with many different ways to reduce the $\mathbf{R_0}$: 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 to flatten the curve?

End of Question 14

So what happens to our model when we change β? While public health responses can reduce the value of β, a single social event on a college campus weekend may cause an increase in the value of β. Changes in β lead to changes in how many people become sick, and changes in the model provide a fascinating moment for building insight into the relationship between a differential equation and its solution.

As a first adjustment, consider $\beta = \beta(t)$ to be piecewise constant

$$\beta(t) = \begin{cases} \mathbf{0.5} & \text{for } \mathbf{0} \le t \le \mathbf{20}, \\ \mathbf{0.2} & \text{for } t \ge \mathbf{20} \end{cases}$$

This may represent suddenly changed health policies once authorities realize an outbreak has begun.

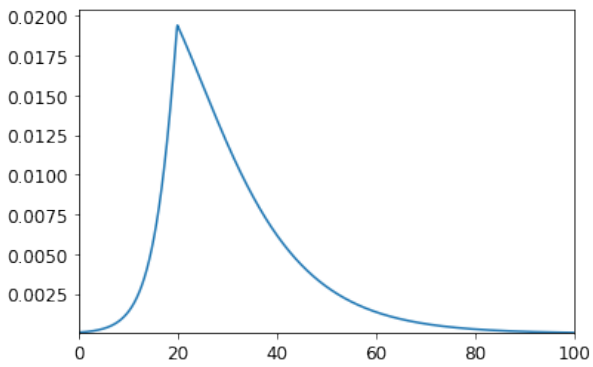
[§] You don't need to attach the pictures unless you want the bonus point from (c).

[¶] You don't need to attach the pictures unless you want the bonus point from (d).

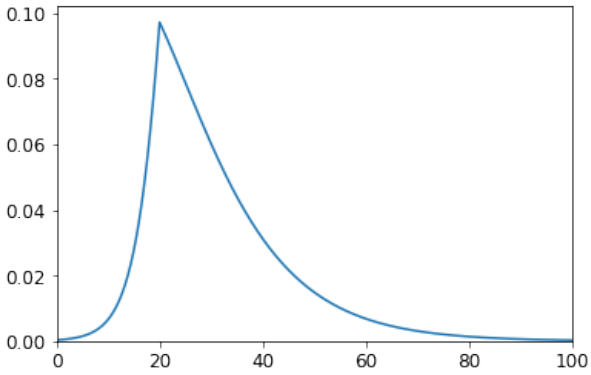
Question 15.

3 points

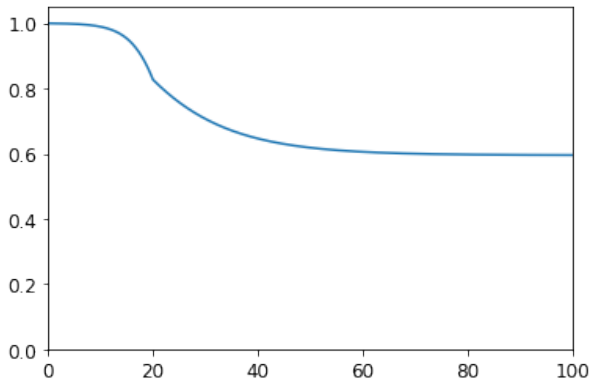
Determine which graph is which below. The choices are $s(t)$, $i(t)$, $r(t)$, ds/dt, di/dt, and dr/dt.



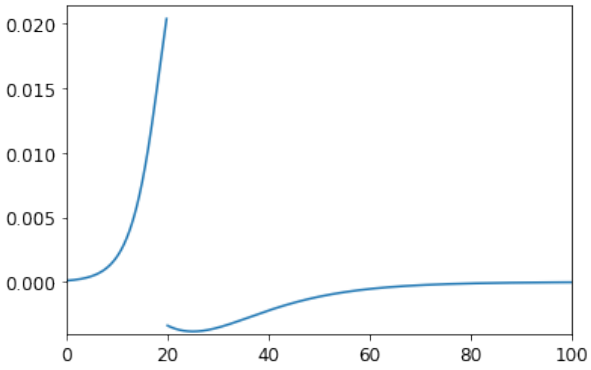
(a)



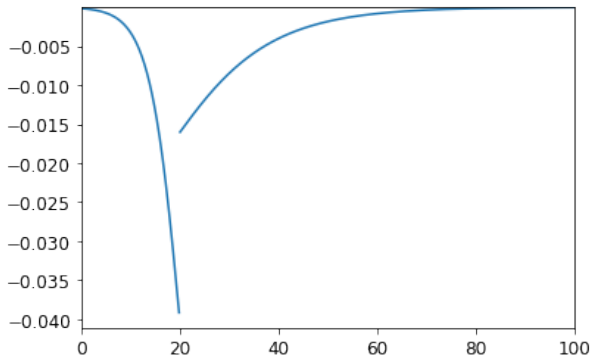
(b)



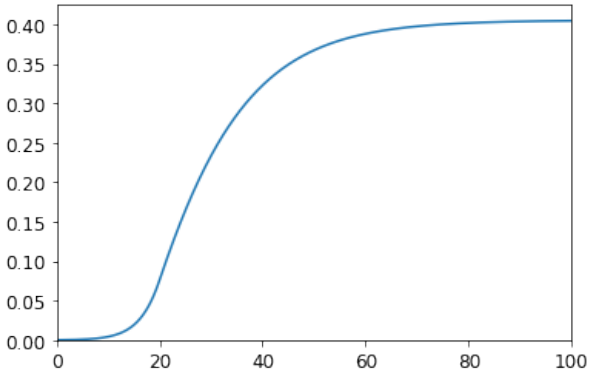
(c)



(d)



(e)



(f)

Figure 3

Question 16.

Optional, Bonus 6 points

Finally, try a more free-form approach.

- What would happen if $\beta(t)$ were constantly decreasing, that is, if $\beta(t)$ were a straight line with negative slope? (Be sure $\beta(t)$ remains nonnegative throughout the time of your outbreak.)
- What would happen if $\beta(t)$ were periodic? Can you construct a periodic function for $\beta(t)$, with a

period of seven days, to represent weekly variation in infectivity? Again, keep $\beta(t) > 0$.

- What other $\beta(t)$ functions could you consider?

In all variations, think through what the graphs of $s(t)$, $i(t)$, $r(t)$, ds/dt, di/dt, and dr/dt should look like. Reason through the answers first, and explain your reasoning carefully. Then, use Python to test your claims by creating the graphs. Ask me if you are unsure how to code variable β in Python.

§G. References

[1] https://www.worldometers.info/coronavirus/

[2] https://covid19.healthdata.org/united-states-of-america

[3] Meredith Greer (2018), "6-007-S-FunctionsAndDerivativesInSIRModels," https://www.simiode.org/resources/4884.

[4] David Smith and Lang Moore, "The SIR Model for Spread of Disease - The Differential Equation Model", https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model

§H. COVID-19 (Optional)

Figure (4) gives an idea of how COVID-19 compares to other infectious diseases. Note that it is extremely hard to estimate $\mathbf{R_0}$ accurately, the picture below only provides a possible upper and lower bound for the value.

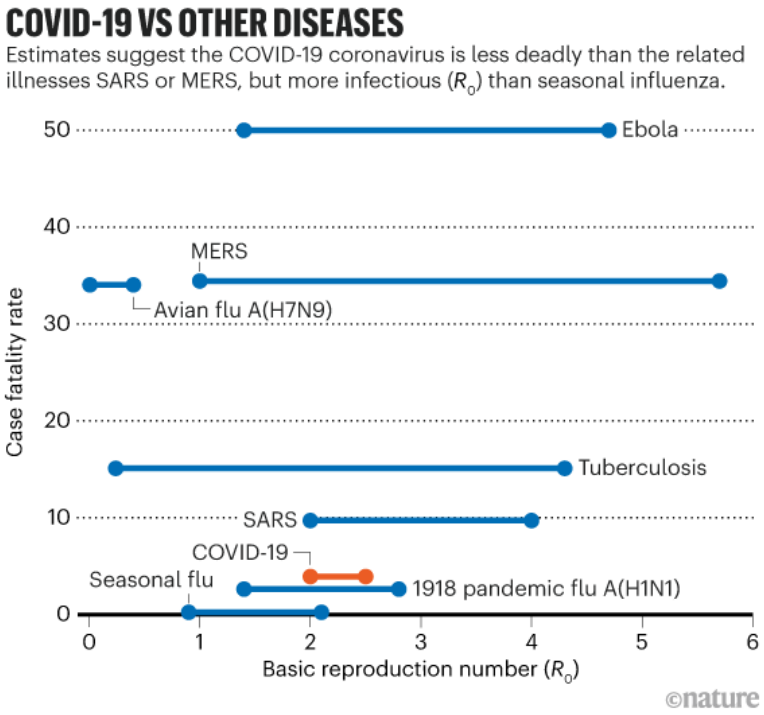


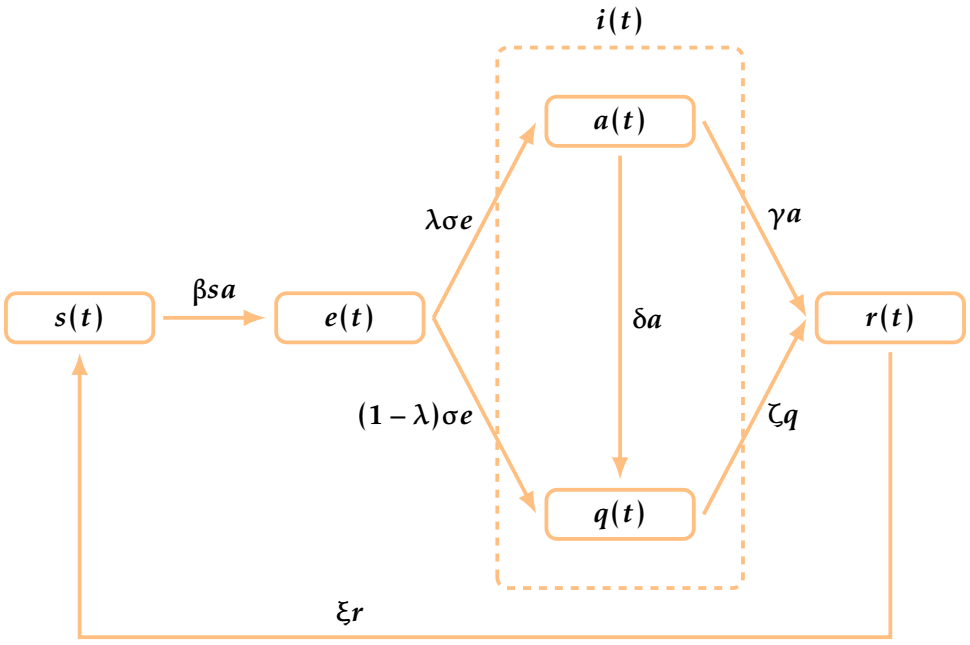
Figure 4: Source: Nature 579, 482-483 (2020)

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 infected 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.

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 below. 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:



- 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 17. **Optional, Bonus 4 points**

Write down the system of ODEs corresponding to this model.

Question 18. **Optional, Bonus 6 points**

Use the Python 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 ζ. Use them to demonstrate the effectiveness of quarantine.