Mathematical Modeling of Epidemics, from influenza to COVID-19

Minimum tech requirements: Youtube, Google spreadsheet, Google Colab (Part 2 only)

Optional tech resources used: Any type of statistic or programming language could be used for solving the SIR model instead of Colab notebooks.

Objectives/Purpose:

1. Students learn the basics of compartmental modeling in an infectious disease context.
2. Students see the connection between discrete and continuous models.
3. Students learn about the relationship between parameters and model behavior; in particular, students learn how the same basic model can be applied to different diseases and/or populations by varying parameters.
4. Students are introduced to reading and making small changes to code for simulations in Part 2.

Possible Courses: Calculus I (after limits or during derivatives), Differential Equations
This module is suitable for Calculus I, to be introduced between limits and derivatives, or as a bridge from calculus topics to the start of a differential equations course.

In a calculus class, I would use the activity from Joanna Wares, "Flattening the curve" at the start of the semester when no knowledge of limits or derivatives is assumed. I would tell the students that we will revisit the ideas when we know more. We would then complete the following activities at an appropriate point mid-semester. Following both activities, I would give students an assignment that calls on them to reflect on the growth in their knowledge of mathematical modeling and tools they have learned to use.

In a differential equations class, I would start the semester with this activity, probably as a precursor to Joanna Wares' "DE Semester Modeling Project that we would work on throughout the remainder of class.

Preview work for students to complete in advance:

Read: "How maths can help us fight infectious diseases" (link https://theconversation.com/how-maths-can-help-us-fight-infectious-diseases-44848)

Watch: Introduction to an infectious disease model, part 1 (11 minutes, link https://youtu.be/XWXqXzAYe4E)

Background
We divide the population into 3 groups:

- **Susceptible** individuals who can catch the disease
- **Infectious** individuals who can spread the disease
- **Removed** individuals who have recovered or otherwise developed immunity to the disease (we will explore vaccination later)

We denote $S(t)$ as the number of individuals who are susceptible at time t, $I(t)$ as the number of infectious individuals at time t, and $R(t)$ as the number of removed individuals at time t.

We make the following assumptions:

- The number of individuals N in the population is large and constant, so for all time *t*

  Total population = **Susceptible** + **Infectious** + **Removed**

  $$N = S(t) + I(t) + R(t)$$

  and there is no birth, death, emigration, or immigration. Individuals move between the subpopulations only. It is helpful to think of this as a closed community, like a college dorm.

- The population is well mixed. This means we are not considering spatial distribution of people and we can picture them as moving around within the space.

- There is no incubation period: once you contract the disease, you are immediately infectious. Is this always true?

1. What are we leaving out in our assumptions? Is it reasonable to assume that all people are equally susceptible to a disease? Explain.

2. What about our assumption regarding an incubation period? Name a disease that DOES have a latency period where there is a lag between contact and infectiousness.

Individuals transition from susceptible to infected at transmission rate β (Greek letter beta). If we look at this at an *individual* interaction level, when sick and healthy people interact, there is a probability β that the healthy person becomes sick.

Since we are modeling at the *population* level, a more precise view is that the rate $\beta = c*x$, where c is the likelihood of contacts between individuals, and x is the likelihood the contact results in infection.

Individuals transition from infectious to removed at rate γ (Greek letter gamma). This rate is inversely proportional to the usual duration of an infection D, $\gamma = \frac{1}{D}$ so diseases with longer duration of infection have smaller recovery rate. This means that people would remain in the infectious population longer.

Equations of change

If $S(t), I(t),$ and $R(t)$ are the number of susceptibles, infectious, and removed individuals at time *t*, then to find the number of individuals in each group at the next time step *t+1* we have:

$S(t+1) - S(t) = -\beta \cdot S(t) \cdot I(t)$, where $S(t+1) - S(t)$ is the change in susceptibles in 1 day

$I(t+1) - I(t) = \beta \cdot S(t) \cdot I(t) - \gamma I(t)$, where $I(t+1) - I(t)$ is the change in infecteds in 1 day

$R(t+1) - R(t) = \gamma I(t)$, where $R(t+1) - R(t)$ is the change in recovereds in 1 day

3. **Notice the signs on the terms**. For example, in the equation that describes the change in the susceptible population, the term $\beta \cdot S(t) \cdot I(t)$ is <u>negative</u>. In the equation that describes the change in the infectious population, the term $\beta \cdot S(t) \cdot I(t)$ is <u>positive.</u> Why?

A schematic of this basic SIR model is shown below:

In a flipped class (like mine), I would assign the previous two pages as preparatory work to be done outside of class as "preview homework" graded for completion. Then we would work through the following portion in class, with students completing the exercises in groups.

Influenza is the disease modeled in this section as a warm-up to COVID-19 in part 2, but these activities can be collapsed into one shorter activity by changing the parameters here to be COVID-19 specific or by only investigating influenza. I think it is worthwhile to both show students different diseases and allow them to "discover" a model of COVID-19 but it is lengthy unless some work is completed out of class.

<u>Activity 1: A discrete math model of influenza</u>

1. Notice that we are making some simplifying assumptions with our SIR model. For example, we are assuming that when you recover from an infection, you are immune to the disease…like chicken pox, for example. Is this true for *any* disease? **Give an example of a disease for which infection does NOT give immunity:**

2. We can rearrange our equations of daily change to find the number of people in each group *S, I,* and *R* at the next time step given our knowledge of the current time step:

$$S(t+1) = S(t) - \beta \cdot S(t) \cdot I(t)$$
$$I(t+1) = I(t) + \beta \cdot S(t) \cdot I(t) - \gamma I(t)$$
$$R(t+1) = R(t) + \gamma I(t)$$

We will use these equations to compute 10 days of a influenza epidemic, with the following assumptions:

Assume that the initial populations (at $t=0$) are

$$S(0) = 760$$
$$I(0) = 3$$
$$R(0) = 0.$$

This is case data from a British boarding school outbreak.

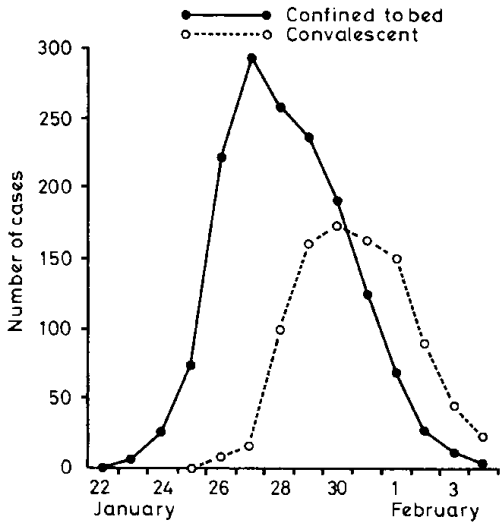
BRITISH MEDICAL JOURNAL 4 MARCH 1978

EPIDEMIOLOGY

Influenza in a boarding school

The following notes are compiled by the Communicable Disease Surveillance Centre (Public Health Laboratory Service) and the Communicable Diseases (Scotland) Unit from reports submitted by microbiological laboratories, community physicians, and environmental health officers.

During January an epidemic of influenza occurred in a boarding school in the north of England. A total of 763 boys between the ages of 10 and 18 were at risk, all except 30 being full boarders; the staff were from the surrounding villages. There were 113 boys between the ages of 10 and 13 in the junior house, while the rest were divided into 10 houses of about 60 boys each.



For influenza, assume the usual duration of infection is $D = 2$ days and in this setting the transmission rate was $\beta = 0.00218$.

a) Find the total number of individuals (N) in the population.

b) Compute the recovery rate $\gamma = \frac{1}{D}$.

c) Use the equations of change to complete a table showing the change in each subpopulation (S, I, R) over the first 10 days of a influenza epidemic.

For example, the susceptible population at time 1 would be:

$$S(1) = S(0) - \beta \cdot S(0) \cdot I(0)$$

Time (t)	0	1	2	3	4	5	6	7	8	9	10
S	760										
I	3										
R	0										

Computing all of this is prone to errors when done by hand, so I recommend that you make a copy of the following spreadsheet and use your results in the sheet to complete the table above and parts e) and f) below: https://tinyurl.com/y9rc4duy

Here I would walk students through the spreadsheet. Then I would break them into groups, trying to make sure that each group has at least 1 student familiar with spreadsheets.

d) Verify that $S + I + R = N$ each time – this will help you know if your model is working correctly!

Use your completed table or the graph generated in the spreadsheet to sketch a very rough graph of S, I, and R versus t (all on the same axis).

Activity 2: Calculus takes us from discrete to continuous

In Activity 1, we used a discrete model in which we are looking at the rate of change in these populations over time steps of one day. However, using calculus, we can take the limit of these discrete rates of change as the time step goes to 0. This is called the ***derivative*** and denoted as:

$$S'(t) = \lim_{\Delta t \to 0} \frac{\Delta S}{\Delta t} = \lim_{\Delta t \to 0} \frac{S(t + \Delta t) - S(t)}{\Delta t}$$

$$I'(t) = \lim_{\Delta t \to 0} \frac{\Delta I}{\Delta t} = \lim_{\Delta t \to 0} \frac{I(t + \Delta t) - I(t)}{\Delta t}$$

$$R'(t) = \lim_{\Delta t \to 0} \frac{\Delta R}{\Delta t} = \lim_{\Delta t \to 0} \frac{R(t + \Delta t) - R(t)}{\Delta t}$$

So, if we let our time intervals get far smaller than one day – infinitely small, our model becomes the system of differential equations below! (Input variable *t* is not shown on the right-hand side for readability).

$$S'(t) = -\beta SI$$
$$I'(t) = \beta SI - \gamma I$$
$$R'(t) = \gamma I$$

1. Based on the formula for *S'(t)* above (the derivative of *S(t)*), do you expect the population of <u>susceptible</u> individuals to **increase or decrease**?

2. Based on the formula for *R'(t)* above, which is the derivative of *R(t)*, do you expect the population of <u>removed</u> individuals to **increase or decrease**?

3. **Notice the connection between the equations and the schematic:** Individuals leave the susceptible group ($-\beta SI$) and join the infectious group ($+\beta SI$). Infectious individuals leave the infectious group ($-\gamma I$) and join the removed group as they recover ($+\gamma I$).

4. Replace parameters β and γ in the general differential equations SIR model above to write the specific SIR model of influenza in the British boarding school:

The following two activities can be omitted or completed in a different order.

Activity 3: Simulating continuous SIR models of influenza and COVID-19

This could be done offline as a group project or in class in a group. This requires that at least 1 student in each group have a Google account to use Colab notebooks to run and edit the Python code for the SIR model.

The solutions to the SIR differential equation system are the population functions S(t), I(t), and R(t) that we would like to see: we want to see how these populations will change over time. We can't, however, solve these equations analytically by hand - we must solve them numerically with a computer. We can use some Python code to solve the differential equations for us.

1. Each group should have at least one student with a Google account who can share a screen, since it is required to use Colab notebooks for Python coding. Go to https://tinyurl.com/yddcj2lg and save a copy of the file to your drive, or designate one or more group members for this.

2. Complete the section titled "A continuous model of an influenza epidemic" – this just involves reading through and gaining some familiarity with the code.

3. *This could be discussed in groups, with someone designated to report for the group.* We will generate a new model specific to COVID-19, first assuming that the population remains the same as in the influenza case: 763 susceptible and 1 infected initially. How do you think the transmission rate β will change for COVID-19? Will it be smaller or larger? How do you think $\gamma = \frac{1}{D}$, where D is the duration of illness, will change? Will it be smaller or larger?

4. Return to the Colab notebook and follow instructions for the next section titled "A continuous model of COVID-19". Choose parameters that reflect your thinking in part 3. Sketch or copy and paste your resulting plot below:

5. *This could be discussed in groups, with someone designated to report for the group.* Reflection questions:

    a. Do you think your results are reasonable?

    b. Where might you look to identify parameters based on data? Did you try to look these up?

    c. Do you think both parameters might be different in different locations?

    d. What are some limitations in the SIR model? For example, do you think all members of the population are equally susceptible? *This could be an opportunity to talk about structured models in a DEs course.*

The activity could wrap up here or could continue offline as a research project to identify parameters through guided research of articles or even data fitting. Alternatively, you can share actual parameters and results that you have found for a particular location with your students for closure.

Activity 4: Data fitting

For a differential equations course, since this module is introductory, you may want to give your students a first sense of parameter estimation that is very low stakes. This activity uses the second tab of the Google spreadsheet that contains data for infecteds over the time course of the British boarding school flu outbreak.

An important part of the mathematical modeling process is fitting the model to real data. If the model can generate predictions that are close to real data collected over time, this gives us some comfort that it is a good predictive model. The connection between the data and the model is in the parameters; ideally the parameters will be able to be tweaked until the model output matches the data reasonably closely, and this process is called data fitting or parameter estimation.

Here we will get a quick sense of what can be a complex process by attempting to match the model values for infected individuals I to real data collected on the number of infected individuals at 14 time points over the course of the outbreak, I_data.

1. Go to the second tab in the Google "Discrete SIR" worksheet, called "Data Fitting": https://tinyurl.com/y9rc4duy.
2. Notice that the graph now shows only the infected population graph, which is in blue for the model output $I(t)$ and red for the data I_data.
3. The goal is to change parameters Beta and Gamma until you get a reasonably good match between the model predictions and the data. This can be evaluated visually and by minimizing the Sum of Squared Errors (SSE) metric on the sheet. The error at each of the 14 time points when data was collected is the difference between the model values and the data values, squared to account for differences in sign. After varying both parameters, write down your best estimates:

Following this activity, groups can share their graphs and parameters, and it may be interesting to compare how different parameter combinations can produce similar results. Discussion could touch on methods of parameter estimation and on the importance of quantifying uncertainty in model predictions and knowing the model limitations given relationships between parameters and output.