

Lecture 1

In what is almost certainly an egregious over-simplification, I claim that calculus is essentially a study of the relationship between quantities and their rates. The traditional example is one of speed and distance: how might we use an equation for distance over time to infer an object's instantaneous speed? What is a sensible interpretation for instantaneous speed? While perhaps the most natural, determining speed from position is far from the only example of relating a quantity (distance) to a rate. A similar process is necessary to model the spread of a disease throughout a population. This is the problem that will serve as our lens through which we shall view the concepts of calculus.

1 Modeling Disease

How can we understand the behavior of a disease in a population as an example of the relationship between rates and quantities? What is a practical way to model disease? What factors should we consider?

Naturally, these questions depend on the particular attributes of the disease in question. Our model will change depending on whether the disease is lethal, whether recovered individuals gain permanent immunity, how the virus itself evolves over time, etc. For the sake of this example and the following discussion, we shall make some simplifying assumptions. However, it is worth considering how the model we develop here might be adapted to address other, possibly more complex, circumstances.

In epidemiology, it is common to divide a population into the following three categories:

- **Susceptible:** those people who have not yet been infected and gained immunity;
- **Infected:** those people who are currently suffering from the illness in question;
- **Recovered:** those people who have recovered from the illness and are no longer contagious.

This is known as the SIR model. We shall make the following assumptions about our disease.

Let's suppose that this disease is not lethal, and that those people who have recovered are immune to further infection (the textbook uses the example of measles). In this way, people move from susceptible to infected to recovered in the following manner:

$$S \rightarrow I \rightarrow R.$$

Note that the assumption about immunity prevents people from ever traveling backward along these arrows. Additionally, the assumption that no individuals will die guarantees that the entire population is made up of people in these three categories. If we let S be the number of susceptible people, I be the number of infected people, and R be the number of recovered people, then

$$\text{population} = S + I + R.$$

Therefore, modeling the behavior of this disease is equivalent to determining the values of S , I , and R at all points in time. Having a model that achieves this goal will allow us to roughly (depending on the accuracy of our model) predict the point in time when infection will be highest, and when the disease will have run its course. We shall see that the values of S , I , and R depend on each other.

2 Finding Equations for S , I , and R

Let's develop equations for the values of S , I , and R . As we observed earlier, it would be nice for these equations to provide the values of S , I , and R at any point in time. In mathematical terms, we say that S , I , and R are *functions of time*, and write

$$S(t), I(t), R(t).$$

In this particular example, we will deviate from this convention, and refer to the quantities of S , I , and R at time t as S_t , I_t , and R_t , respectively. This is a purely notational choice, and is due to the fact that recursive equations (which these will be) are represented in this way.

2.1 R_t

We shall begin with the equation for the number of recovered individuals in our population at time t . Because people go in one direction from susceptible to recovered, we expect R_t to increase with time. Since a person can only recover if they have been infected, we would like R_t to be a function of I_{t-1} . Supposing (we shall make a number of assumptions that may not perfectly capture the dynamics of an infection in the real world) that it takes 14 days to recover once infected, we should expect the number of newly recovered individuals to be $1/14$ of the infected population, or

$$\frac{1}{14}I_{t-1}.$$

(It is worth noting that this is perhaps a less accurate assumption than we might hope. Why?) Of course, we must add this to the number of previously recovered individuals, which is conveniently given by R_{t-1} . In total, we then have the following equation:

$$R_t = R_{t-1} + \frac{1}{14}I_{t-1}.$$

Note that $1/14$ was chosen arbitrarily, and that different diseases will have different values. Selecting the value that best represents the real world is part of the art of tuning a mathematical model. Generally, this quantity is referred to as the *recovery coefficient*, and we shall represent it with the letter b . This gives us the following general equation:

$$R_t = R_{t-1} + bI_{t-1}.$$

2.2 S_t

We now consider the equation for the number of susceptible individuals in our population at time t . Based on the dynamics of our disease, we expect S_t to decrease as t increases. (This is a consequence of our assumption that individuals move in one direction from susceptible to infected to recovered.) Susceptible people become infected when exposed to the infected people. Supposing that each infected individual contacts 0.04% of the susceptible population on a given day, the number of susceptible people that are exposed to infection on a given day is:

$$I_{t-1}(0.0004S_{t-1}).$$

However, not every susceptible person who is exposed will become infected. The percentage of exposed individuals who actually become infected is contingent upon the contagiousness of the disease. We might suppose that our disease is particularly virulent, and give this a value of 50%. This means that the number of infected people at time t is:

$$0.5I_{t-1}(0.0004S_{t-1}) = 0.0002S_{t-1}I_{t-1}.$$

For notational simplicity, we shall combine the percentages representing contact rate and contagiousness into a new value, which we shall call the *transmission coefficient*, and represent with the letter a . Since newly infected people are subtracted out of the group of susceptible people, we have the following equation for S_t :

$$S_t = S_{t-1} - aS_{t-1}I_{t-1}.$$

2.3 I_t

We have already done the hard work. Since the population is partitioned into the three groups S , I , and R , the total number of infected individuals is precisely those who were previously infected (I_{t-1}), plus those newly infected ($aS_{t-1}I_{t-1}$), minus those recovered (bI_{t-1}). This gives us the following equation:

$$I_t = I_{t-1} + aS_{t-1}I_{t-1} - bI_{t-1}.$$

In summary, we have the following three equations for our populations of susceptible, infected, and recovered individuals at time t :

$$\begin{aligned} S_t &= S_{t-1} - aS_{t-1}I_{t-1} \\ I_t &= I_{t-1} + aS_{t-1}I_{t-1} - bI_{t-1} \\ R_t &= R_{t-1} + bI_{t-1}. \end{aligned}$$

If we have done our job well, these equations should capture the behavior of an infection in our population over time. However, in their current form, they are a little bit hard to parse. Rather than representing the size of S , I , and R at time t , let us simply keep track of the amount that each of these quantities changes at each time-step. Observe that the change in S , I , and R is given by $S_t - S_{t-1}$, $I_t - I_{t-1}$, and $R_t - R_{t-1}$, respectively. Applying this to the above equations, we obtain:

$$\begin{aligned} S_t - S_{t-1} &= -aS_{t-1}I_{t-1} \\ I_t - I_{t-1} &= aS_{t-1}I_{t-1} - bI_{t-1} \\ R_t - R_{t-1} &= bI_{t-1}. \end{aligned}$$

In mathematical notation, it is common to represent the change in some value X by the character Δ . In this way, the expression ΔX is taken to mean “the change in X ”. Let us use ΔS , ΔI , and ΔR to refer to the changes in S , I , and R after one time-step t . Therefore, we might re-write the above equations as:

$$\begin{aligned} \Delta S &= -aSI \\ \Delta I &= aSI - bI \\ \Delta R &= bI. \end{aligned}$$

We have gotten rid of the subscripts because the idea of moving one time-step from S_{t-1} to S_t is baked into the notation ΔS . We are still saying the same thing: the change in the population of susceptible people after one time-step is a function of the number of currently infected and currently susceptible people.

Let us make one final alteration. Rather than describing the magnitude of change in S , I , and R (i.e. ΔS , ΔI , and ΔR), let us consider the *rate of change* of these quantities. What does this mean? Just as speed is the rate of change of distance (i.e. distance per time), we would like to consider the rate at which the sizes of S , I , and R change per unit time. In mathematical terms, this rate is simply

$$\frac{\Delta S}{\Delta t},$$

or “the change in S divided by the change in time.” We shall use the notation S' (and I' , R') to refer to this rate of change of S (and I , R). Fortunately, in our case $\Delta S = S_t - S_{t-1}$ and $\Delta t = t - (t - 1) = 1$, so $S' = \Delta S$ (and $I' = \Delta I$, $R' = \Delta R$). Accounting for these observations, we have:

$$\begin{aligned} S' &= -aSI \\ I' &= aSI - bI \\ R' &= bI. \end{aligned}$$

These are the equations we shall use to model the dynamics of our infection.

3 Using Our Model

Let’s put this model to use. Right now, our equations tell us how the populations of susceptible, infected, and recovered people change day-to-day, subject to a recovery coefficient b , and a transmission coefficient a . To actually go about modeling a real-world infection, we will need to replace these coefficients with actual numbers, and we will need some additional information about the status of our population (i.e. how many people are susceptible, infected, and recovered). In the context of an actual pandemic, our choices for these values would be based on real-world observations; however, for simplicity, we will just pick our own.

Let’s suppose that we are modeling a measles outbreak in a population of 50,000 people. We shall use our original recovery coefficient of $b = 1/14$. Our transmission coefficient will be $a = 0.00001$, which is roughly in the range of numbers used in epidemic studies (according to the textbook that this lesson is based on). Say that 2,100 people are currently infected, and 2,500 people are already recovered. Since our population has 50,000 people in total, this means that the number of susceptible individuals is $50,000 - 2,100 - 2,500 = 45,400$. We can summarize our model as follows:

$$S' = -0.00001SI \tag{1}$$

$$I' = 0.00001SI - \frac{1}{14}I \tag{2}$$

$$R' = \frac{1}{14}I \tag{3}$$

$$S_0 = 45400, \quad I_0 = 2100, \quad R_0 = 2500. \tag{4}$$

Here, the values S_0 , I_0 , and R_0 represent the initial values of our three population groups (i.e. their values at time $t = 0$). (Note that time $t = 0$ is not necessarily the time of the first infection; we simply use it to represent the first instance of the problem for our model. With this notation, it is possible to determine the number of infections at $t = -1, -2, -3, \dots$) Because our model depends on these initial values, it is called an *initial value problem*.

3.1 $t = 1$

How might we use our initial value problem to determine the behavior of this measles infection over time? Let's begin by finding S_1 , I_1 , and R_1 .

First, we shall determine S' , I' , and R' . This is as simple as plugging the initial conditions S_0 , I_0 , and R_0 into the equations on lines 1, 2, and 3. Therefore:

$$\begin{aligned} S' &= -0.00001S_0I_0 \\ &= -0.00001(45400)(2100) \\ &= -953.4 \end{aligned}$$

and

$$\begin{aligned} I' &= 0.00001S_0I_0 - \frac{1}{14}I_0 \\ &= 0.00001(45400)(2100) - \frac{1}{14}(2100) \\ &= 803.4 \end{aligned}$$

and

$$\begin{aligned} R' &= \frac{1}{14}I_0 \\ &= \frac{1}{14}(2100) \\ &= 150. \end{aligned}$$

This means that from $t = 0$ to $t = 1$, the sizes of S , I , and R are changing at rates of -953.4 , 803.4 , and 150 people per day, respectively. We can now determine S_1 , I_1 , and R_1 by simply adding these values to S_0 , I_0 , and R_0 . Doing so:

$$\begin{aligned} S_1 &= S_0 + S' = 45400 - 953.4 = 44446.6 \\ I_1 &= I_0 + I' = 2100 + 803.4 = 2903.4 \\ R_1 &= R_0 + R' = 2500 + 150 = 2650. \end{aligned}$$

(Note that our notation here is a bit lazy. We have claimed that $S_1 = S_0 + S'$. In reality, $S_1 = S_0 + S'\Delta t$, and $\Delta t = 1$. This is an important distinction; it may not always be the case that the rate S' is calculated over an interval of size 1.)

We can do a sanity check to make sure that everything is working properly: since our measles epidemic is non-lethal, our total population at $t = 1$ should still be 50,000 people. Indeed:

$$44446.6 + 2903.4 + 2650 = 50000.$$