Chapter 1

What is Calculus? The Six Pillars

Calculus was invented by Isaac Newton and Gottfried Leibniz in the mid-17th century to solve real problems: To understand the orbits of planets. To understand what happens when an apple falls from a tree. To design lenses and clocks and all sorts of machines. Even to make money betting at dice. Over time, people realized how much calculus had to say about almost *everything* in the world. If it's real and it's changing, we can understand it better with calculus.

But as a 21st century student learning calculus, it's easy to get lost in a maze of formulas, computational techniques, obscure theorems, and contrived problems. You can lose sight of what it's all for.

So what is calculus, anyway? A good working definition is

Calculus is the study of things that change.

- If gas costs \$4/gallon and you need to buy 500 gallons this year, you don't need calculus to figure out your cost.

  $$\$4/\text{gallon} \times 500 \text{ gallons} = \$2000.$$

  That's simple arithmetic. But if the price of gas keeps changing, figuring out your total cost involves adding up the *variable* costs for the next 365 days. That's calculus.

- If you are manufacturing widgets and every widget costs \$20 to make but sells for \$23, it's obvious that you should make more widgets, since you make a profit of \$3 on each widget. But what if making more widgets causes the price to drop, thanks to the law of supply and demand? How many widgets should you make to maximize your profit? That's calculus.

- If your company has a 13% share of the market and your market share increases at exactly 1%/year, it's easy to predict what your market share will be in 5 years. 13%+(1%/year×5 years) = 18%. But what if your market penetration slows down as you gain market share? Understanding how that changes the answer requires calculus.

Fortunately, calculus isn't black magic. Almost everything we do boils down to six simple ideas that we call the **Six Pillars of Calculus**:

(1) **Close is good enough.** In algebra, we look for exact answers to problems. If somebody asks you to solve $x^2-3x+2=0$, you might factor that as $(x-2)(x-1)=0$, or you might apply the quadratic formula, or you might just try different numbers and see that $x = 1$ and $x = 2$ work. What you probably *don't* say is "There must be a solution somewhere between 0.9 and 1.1 ..., make that 0.99 and 1.01 ..., no, make that 0.999 and 1.001." But that's exactly the sort of reasoning we use in calculus! When faced with problems that can't be solved right away, we simplify things and look for an approximate answer. Then we try to improve that answer, and improve it some more, until we get something that is fairly accurate. If the calculations are too grungy to do by hand (and they often are), we program a computer to do our dirty work for us. Eventually, we get an answer that's good enough for whatever real-world task we have in mind, and we stop and congratulate ourselves on a job well done.

Once in a while, we need the exact answer. When that happens, we approach that exact answer as the **limit** of better and better approximations. Almost all of the important formulas of calculus come from this limiting process.

(2) **Track the changes.** We can often tell more about something by looking at the way that it's changing than by asking where it is right now. Figures 1.1 and 1.2 show the stock prices of two companies for the first half of 2019. Coty's did very well, while Nordstrom's did very badly. It doesn't take a genius to see that by midsummer Coty's CEO was dreaming of getting a big bonus, while Nordstrom's was worried about being fired. That is despite the fact that Nordstrom's stock price at the end of June was more than twice Coty's. To understand the success of a company, the recent change in the stock price tells us much more than the actual price.

Graphically, the rate of change of a quantity is closely related to the **slope** of the graph of that quantity. We're going to spend a lot of time studying slopes—both how to find them and what to do with them. In mathematical terms, this is called computing a **derivative**.

(3) **What goes up has to stop before it comes down.** Figure 1.3 shows the profit $P(x)$ that a widget factory makes as a function of the number x of widgets that it makes each month. If we are currently making a widgets/month, then we should think about increasing production. This part of the curve is sloping upward, so more widgets mean more profit. However, if we are currently making c widgets/month, then we should decrease production. This part of the curve is sloping downward, so more widgets mean less profit. The optimal production level is at b widgets/month, where the curve is flat and its slope is 0. By studying

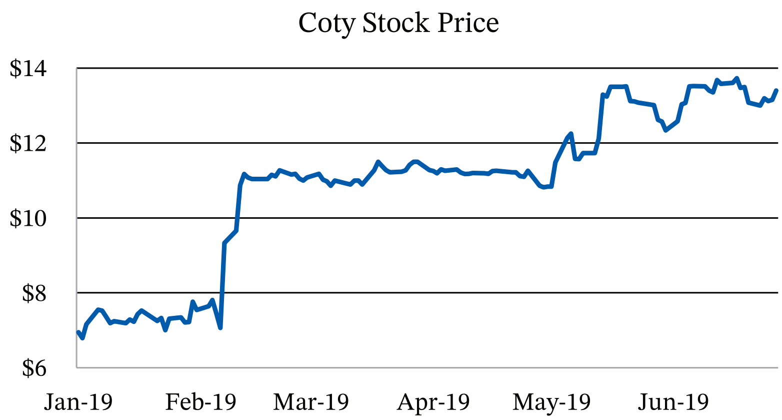


Figure 1.1. Coty's stock went up in the first half of 2019.

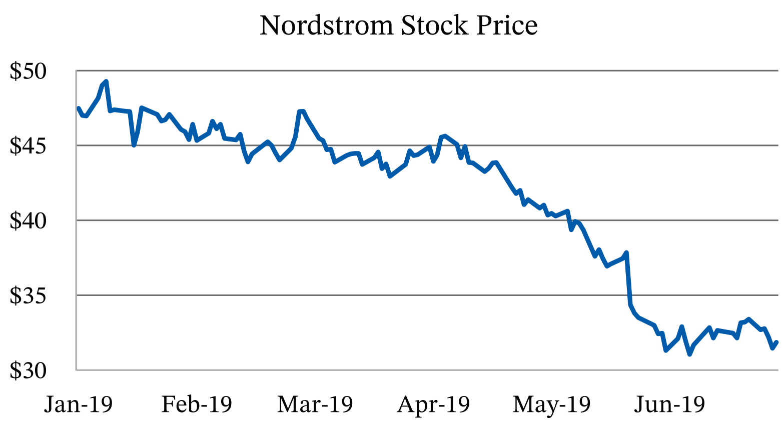


Figure 1.2. Nordstrom's didn't.

the rate of change of the profit curve and figuring out where that rate of change is 0, we can determine the optimal production level.

(4) **The whole is the sum of the parts.** What will the national debt be next year? That's a complicated question, but we can break things down year by year. The national debt next year is the national debt this year plus this year's budget deficit.[1] Similarly, this year's debt is last year's debt plus last year's deficit. Working backward, year by year, we see that the national debt this year is the sum of the national budget deficits every year going back to 1776.

If we plot the budget deficit as a function of time, as in Figure 1.4, the sum of all those values is the same as the area under the curve. We're going to spend a lot of time talking about area, but it isn't because we're obsessed with geometry.

[1] Or at least it would be if the same accounting practices were used for both debt and deficit. See Section 4.1 for the difference in how Social Security is usually counted. In this section, the national debt is adjusted to take that difference into account.

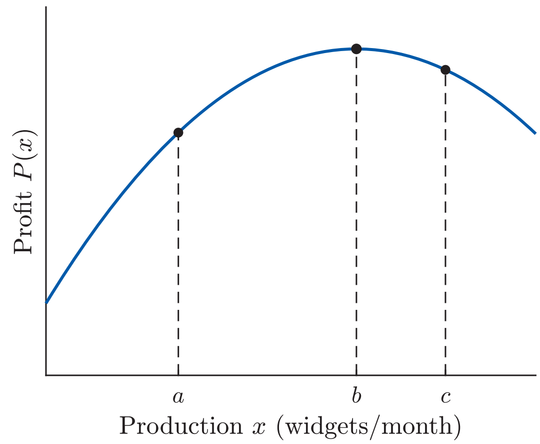


Figure 1.3. Monthly profit as a function of production level

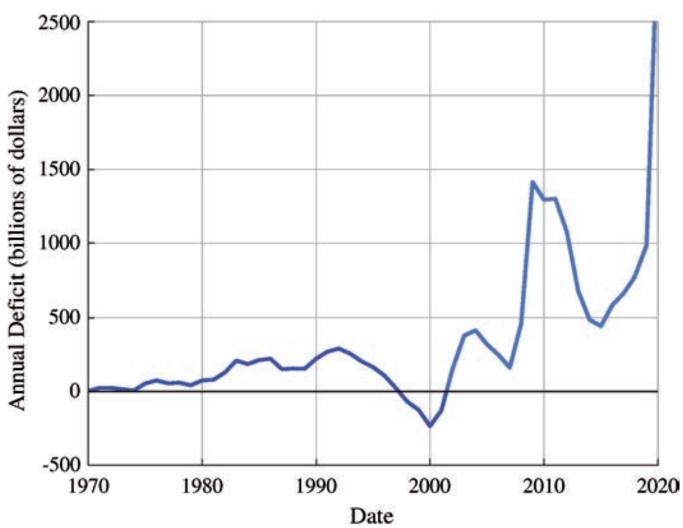


Figure 1.4. US budget deficit by year, 1970–2020

It's because most bulk quantities, like the total power used by the city of Austin in 2007 or the total amount of water carried by the Colorado River in 2011, can be treated exactly like national debt or area. Break the quantity you're studying into little pieces, estimate each piece, and add up the pieces. This process is called **integration**.

Note that this is very different from how integration is often taught. In most calculus classes, students are taught that integration is in some sense the opposite of differentiation. That isn't exactly *wrong*, as we'll see with the 5th Pillar, but it misses the point. Integration is about adding up the pieces, which is why it applies to a host of problems. Anti-derivatives are a great tool for actually *computing* integrals, but they don't explain why such-and-such quantity is represented by such-and-such integral. For that, we need the 4th Pillar.

(5) **One step at a time.** A famous Chinese proverb says that "a journey of a thousand miles begins with a single step". The journey then continues with about two million additional steps. By understanding what happens at each step, we can understand the entire journey.

Instead of asking what the national debt *is*, we can ask how much it *changed* in a short period of time, like a year. Geometrically, the change per year is the **slope** of the debt curve, shown in Figure 1.5. The steeper the curve is, the faster the debt is changing.

In the late 1990s, the deficit was negative (also known as running a surplus), and the debt came down. In the Great Recession of 2009–2012, and again in the Covid pandemic of 2020, the budget deficit was large and the debt shot up. Between 1950 and 1975 (only part of which is shown on the graph) the deficit was close to 0 and the debt was nearly constant.

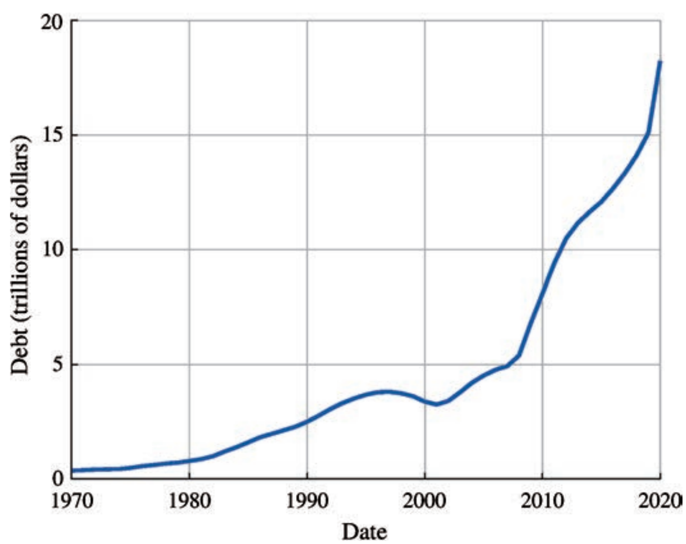


Figure 1.5. US National debt by year 1970–2020

How are Figures 1.4 and 1.5 related? The deficit each year is the same as the change in the debt from that year to the next. This means that the debt is a running total of the deficit and corresponds to the **area** under the deficit curve. The deficit is the rate at which the debt is changing, and corresponds to the **slope** of the debt curve. Mathematically, the debt is the **integral** of the deficit, and the deficit is the **derivative** of the debt. The **Fundamental Theorem of Calculus** (FTC) relates derivatives and integrals in general. With it, we can use what we already know about derivatives to understand integrals. In fact, it is the single most powerful tool we have for evaluating integrals.

The Fundamental Theorem of Calculus comes from thinking about changes one step at a time, but there is more to "one step at a time" than just the Fundamental Theorem. For instance, suppose that we invest $1000 at 6% interest, and we want to know how much money we will have in 30 years. From our initial balance and the interest rate, we can estimate the interest we receive in the first year and compute our bank balance a year from now. From that, we can estimate the interest we receive in the second year and get our bank balance two years from now. Continuing the process, step by step, we can accurately project our bank balance far into the future.

Finally, "one step at a time" is great slogan for how to approach calculus in general. Many problems are way too complicated to be done all at once. By breaking big problems into sequences of smaller problems, and by solving these smaller problems one at a time, we can accomplish wonders.

(6) **One variable at a time.** Many functions involve two or more input variables. The boiling point of water depends on our elevation and on how much salt we put in the water. In the summer, the heat index depends on temperature and humidity. In the winter, the wind chill factor depends on temperature and wind speed. The price of widgets depends on supply and demand. The value of an oil field depends on how much oil it produces and on the price of oil.

To understand functions of two (or more) variables, we always hold everything but one variable fixed and study just that variable. Asking how a change in temperature changes the heat index is a question about a function of just one variable that we already know how to answer. Similarly, we can figure out how changing the humidity affects the heat index. Putting the two answers together, we can understand how changing both temperature and humidity affects the heat index.

It's tempting to say "that's all there is!", but that isn't really true. Over the centuries, lots of really smart people have cooked up lots of really smart ways to solve lots of really hard problems. Now, with the help of computers, we can solve even more problems. In the next ten chapters, we're going to follow in the footsteps of these masters and learn some of their results. Yes, there will be formulas to memorize and techniques to practice and algorithms to implement on computers. And yes, it will take work to really absorb everything.

But hopefully this voyage of discovery won't be a mystery. Every new formula or technique or algorithm will lead straight back to the Six Pillars, so it will be connected to every other formula or technique or algorithm. If you remember the pillars, you'll have a framework for organizing all the little details. Over time, you'll forget many of those details, and that's OK. As long as you remember the pillars and are willing to look up the details as needed, you'll be able to use calculus for your whole life, not just in your classes and in your job, but in understanding the wild and complicated world we live in.

Chapter 2

Predicting the Future: The SIR Model

In this chapter, we will use four of the pillars to get a handle on a real-world problem: product adoption.

- It's impossible to keep track of the behavior of every customer in the world, so we devise a simplified model to describe consumer behavior. *Close is good enough!*
- This involves understanding the *rate* at which consumers start using a product and the rate at which they abandon that product. *Track the changes!*
- Once we understand where things stand and how fast they are changing, we can make realistic predictions about what is likely to happen in the future, as well as what we can do to improve that trajectory. Our projections are only accurate for a short time, so we combine a lot of short-term projections to get a long-term projection. *One step at a time!*
- Even without these long-term projections, we can understand how our market penetration will peak by comparing the factors that increase penetration with those that decrease penetration. *What goes up has to stop before it comes down!*

2.1. A Problem of Market Penetration

Imagine that you work for a company that has just introduced a hot new phone app. People are learning about it by word of mouth, and more and more people are using your product. Of course, that growth can't go on forever. Eventually, people will get tired of your app and will move on to the Next Great Thing. In order to make the most of your product's popularity, you need to forecast usage for the next year or two. You hire a mathematical consultant (your Friendly Author), and together we attack the problem.

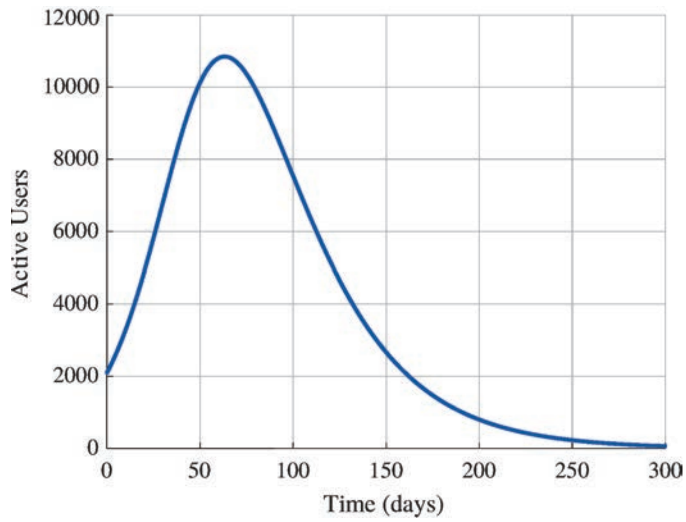


Figure 2.1. Active users as a function of time

The first thing we do is define our quantities. Let t denote time, measured in days, and let $I(t)$ be the number of active users at time t. We call t and I **variables** since they change. The time t is our **input variable**, while I is an **output variable**. The graph of I as a function of t will look something like the plot in Figure 2.1. There is an initial phase where more and more people use the product, a moment of peak usage, a decline in usage, and a long tail. To be successful, we want the rise to be as fast as possible, the peak to be as high as possible, and the decline to be as slow as possible.

Before we try to control I, we need to understand what makes I change over time. That is, we need to make a mathematical **model** for what is going on. The model needs to take into account all the *important* features of what's happening in the world, while being simple enough to be solvable. We are going to ignore a *lot* of details, because *close is good enough!*

Once we have our model, we have to **analyze** it. In this step, we don't care where our equations came from. We just want to solve them. Maybe we can find a formula for the answer. More likely, we can't find a formula, but we can run the model on a computer to generate accurate predictions. Once again, *close is good enough!*

Finally, we need to **interpret** our results. Math can tell us that such-and-such variable will have such-and-such value at such-and-such time, but we need to understand business to say what that means for the success of our company.

In other words, predicting the future is a three step process:

(1) **Model** our system mathematically. Define appropriate variables and write down some equations that describe how these variables change with time. This step requires real-world understanding as well as math.

(2) **Solve** the model to determine what each variable will be at some future time. This step is 100% math, using techniques that we are about to develop.

(3) **Interpret** the results. Take our mathematical results and make real-world sense of them.

We'll tackle these one at a time, which is a lesson in itself. *One step at a time!* We need to break hard problems into bite-sized tasks, and then do each task in turn.

2.2. Building the SIR Model

To understand the adoption of our product, we divide our population of potential customers into three groups, as in Figure 2.2.

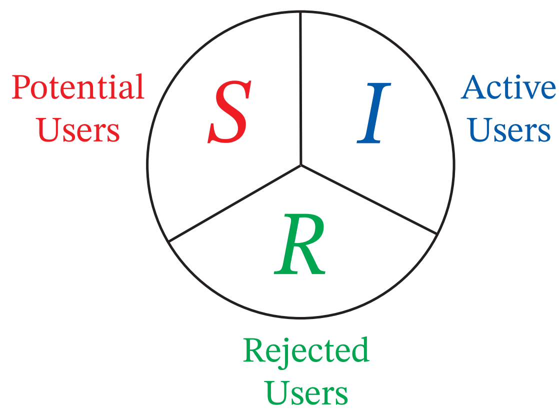


Figure 2.2. Three categories of users

- **Potential** users, or *Potentials*: These are people who might use the product in the future if they hear enough good things about it from their friends.
- **Active** users, or *Actives*: These are people who have already adopted the product. Not only are they using it, but they are spreading the word about it.
- **Rejected** users, or *Rejecteds*: These are people who won't ever use the product. Maybe they just aren't interested. Maybe they used to use the product but got tired of it. Maybe our app doesn't even work on their brand of phone. For whatever reason, they're no longer reachable.

Next, we define our variables. Let $S(t)$ be the number of Potentials at time t, let $I(t)$ be the number of Actives at time t, and let $R(t)$ be the number of Rejecteds at time t. The letters S and I, which obviously don't stand for Potential and Active, are historical.[1]

Obviously, there are differences among the users in each category, but we're going to ignore those differences. This is a model, not a complete description of reality! Instead, we're going to talk about the behavior of the *average* Potential, the *average* Active, and the *average* Rejected. We then ask two questions:

(1) At what rate do Potentials adopt the product and become Actives?

[1] The SIR model was originally developed to study epidemics, as we will see in Section 2.5. In that context, S stands for *Susceptible*, I stands for *Infected*, and R stands for *Recovered* or *Removed*.

(2) At what rate do Actives stop using the product and become Rejecteds?

From these rates, we can figure out the rates S', I', and R' at which the quantities S, I, and R are changing, and from those we can predict the future.

Losing customers: Attrition. Customers don't use an app forever. Sooner or later they get tired of it or switch to a competitor's app. Some apps, like navigation tools and browsers, keep their customers for months or years. Others, like games, can lose their customers after just a couple of weeks.

Suppose that we are analyzing a new game that users keep using for an average of 30 days. Among the Active users, roughly 1/30 of them will grow tired of the game today and will be among the Rejecteds tomorrow. That is,

$$\text{today's change in the Rejected population} = \frac{I(\text{today})}{30}. \tag{2.1}$$

Likewise,

$$\text{tomorrow's change in the Rejected population} = \frac{I(\text{tomorrow})}{30}, \tag{2.2}$$

and in general

$$\text{the change in the Rejected population on day } t = \frac{I(t)}{30}. \tag{2.3}$$

Note the units in this equation. $I(t)$ and $R(t)$ are numbers of *people*, and the number of new R's on any given day is also measured in people. But the *rate* R' at which R is changing is measured in people/day. That is,

$$R'(t) = \frac{I(t)}{30 \text{ days}}.$$

The numerator has units of people, the denominator has units of days, and the ratio has units of people/day.

This is an example of a **rate equation**. A rate equation describes the rate at which something is changing in terms of other data. In this case, it describes R' in terms of I.

More generally, if T is the average length of time that people use a product, then we expect

$$R' = \frac{I}{T}. \tag{2.4}$$

If we define $b = 1/T$, then we can write our rate equation without fractions:

$$R' = bI. \tag{2.5}$$

See Figure 2.3. R' has units of people/day, I has units of people, and b has units of (1/day)s. When the SIR model is used in biology, b is called the **recovery coefficient**. In our business example we'll call it the **attrition coefficient**. If you prefer, you can think of it as the **boredom coefficient**.

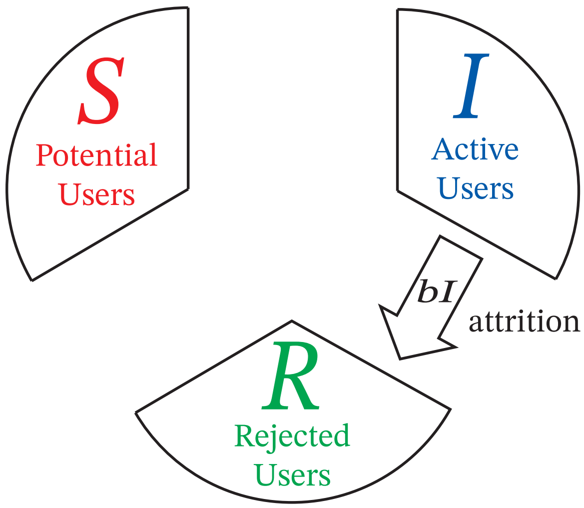


Figure 2.3. Through attrition, active users become rejected users.

The number b is called a **parameter**. It isn't a **variable**, because it doesn't change during the product run. However, it may take on different values for different products. By adjusting the value of b, we can use the same rate equation to model games, navigation aids, music players, you name it.[2]

Gaining customers: Transmission. Losing customers is grim. If our app currently has 2,100,000 users and is only used for an average of 30 days, then we're losing users at a rate of

$$\frac{2{,}100{,}000 \text{ people}}{30 \text{ days}} = \frac{70{,}000 \text{ people}}{\text{day}}. \tag{2.6}$$

That's a lot! If we're going to stay in business, we need to get new customers to replace the old ones.

In our model, we assume that we gain customers by word of mouth. Let's look at this from the perspective of a single Potential user, who we'll call Joe. Joe will hear about our app from a certain fraction p of the Active users each day, for a total of pI contacts. The fraction p is typically very small, but the Active population I can be very large. The more Active users there are, the more times Joe will hear about our product. Every time that Joe hears about our product, there is a probability q that he will be motivated to download it and start using it. That is, there is a probability pqI per day of Joe becoming Active.

We don't actually care about p and q separately, since all that matters is their product. We define $a = pq$, and call a the **transmission coefficient**. Like b, this is a parameter, not a variable. Different products in different communities will have different values of a, but we can use the same reasoning for all of them.

[2]There are some exceptions, such as social media, that follow a different pattern. People don't stop using social media platforms after a certain amount of time. They stop using them when their friends stop using them. That's a much more complicated system to model!

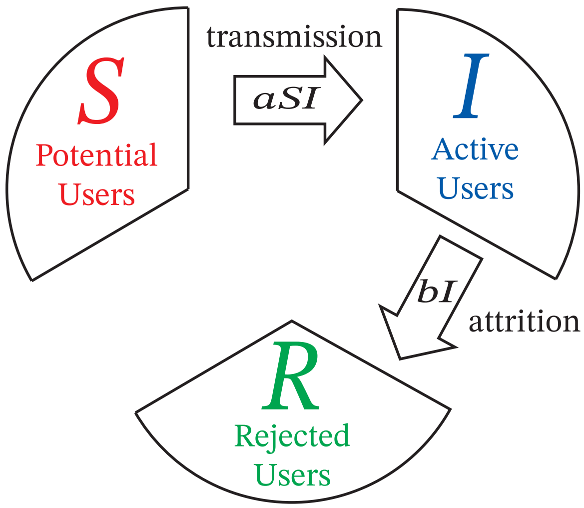


Figure 2.4. Through word of mouth, Potential users become Active users.

Finally, we look at the entire population of Potentials. If there are S Potentials, each of whom has a probability aI (per day) of becoming Active, then each day there will be approximately aSI Potentials who become Actives. That is,

$$S' = -aSI. \tag{2.7}$$

See Figure 2.4. S and I have units of people and S' has units of people/day, so a must have units of 1/(people × days). Also note the minus sign in our equation! The *more* Potentials become Actives, the *fewer* Potentials are left.

What about I? Completing the model. So far we have figured out how the numbers of Rejecteds and Potentials change, but what about the number of Actives? We have

$$\begin{aligned} I' &= +\,(\text{Rate at which people adopt the product}) \\ &\quad -\,(\text{Rate at which they abandon the product}) \\ &= aSI - bI. \end{aligned} \tag{2.8}$$

Another way to see this is that $S + I + R$ is the total number of people out there, and that doesn't change. Since $S' + I' + R' = 0$, we must have $I' = -S' - R'$.

Putting everything together, we have a system of three rate equations. Together, they're called the **SIR model**:

$$\begin{aligned} S'(t) &= -aS(t)I(t), \\ I'(t) &= aS(t)I(t) - bI(t), \\ R'(t) &= bI(t), \end{aligned} \tag{2.9}$$

where

- t is the time. Depending on the setting, you may want to measure t in days, weeks, months, quarters, or years.
- $S(t)$ is the number of Potentials at time t, measured in people.
- $I(t)$ is the number of Actives at time t, also measured in people.
- $R(t)$ is the number of Rejecteds at time t, also measured in people.
- b is the attrition coefficient, which is the reciprocal of the average time T that a user keeps using the product before moving on. The units of b are the reciprocal of whatever units we are using for time. This parameter can vary from product to product.
- a is the transmission coefficient. This can depend both on the product and on the market. If you are marketing the same product in several different places, b will be more or less the same in all markets, but a will typically be bigger in the smaller markets (where each Potential knows a greater fraction of the Actives) and smaller in the bigger markets. The units of a are 1/(time × people).

2.3. Analyzing the Model Numerically

Now that we have our model, let's use it to predict the future. Suppose that we measure t in days, that $a = 0.000002$/person-per-day and $b = \frac{1}{30}$/day, and that we start with $S(0) = 40,000$, $I(0) = 2100$, and $R(0) = 7900$. What will S, I, and R be in two days? In five days? In ten days? For that matter, what were the values yesterday? A week ago? A month ago?

A naive approach is to use the SIR equations to compute S', I', and R' once and for all:

$$\begin{aligned} S' &= -0.000002(40,000)(2100) &&= -168 \text{ people/day}, \\ I' &= 0.000002(40,000)(2100) - 2100/30 &&= 98 \text{ people/day}, \\ R' &= 2100/30 &&= 70 \text{ people/day}. \end{aligned} \tag{2.10}$$

If we have 2100 Actives on day 0, and if I is growing at a rate of $I' = 98$ people/day, then we should expect $2100 + 98 = 2198$ Actives tomorrow. We should expect $2198 + 98 = 2100 + 2(98) = 2296$ Actives the day after tomorrow. After a week, we should expect $2100+7(98) = 2784$ Actives, after a month we should have $2100+30(98) = 5040$ Actives, and after a year we should have $2100 + 365(98) = 37,870$ Actives. In general, after t days we should have

$$I(t) \approx I(0) + I'(0)t = 2100 + 98t. \tag{2.11}$$

This is called a **linear approximation**. Instead of finding the exact equation for $I(t)$, we found the equation of the line that has the right value and the right slope at $t = 0$. We then approximate the value of the true $I(t)$ function by the value of the linear function $2100 + 98t$.

Before we move on, let's recall some basic facts about things that change at a constant rate. If we are driving at a constant speed of 57 miles per hour and pass milepost 253 at 3:00, where will we be at 4:00? At 5:00? At time t?

From 3:00 to time t is $t - 3$ hours. Since rate × time = distance, and since we are going at 57 MPH, we will travel $57(t - 3)$ miles in that time. Adding that to our starting point at milepost 253, we will find ourselves at milepost

$$x(t) = 253 + 57(t - 3) \tag{2.12}$$

at time t. (At least if we're traveling in the direction where the mile markers are increasing. If we're heading in the opposite direction, we will find ourselves at milepost $253 - 57(t - 3)$.)

The same idea works for any quantity that is changing at a constant rate, not just for position. If a quantity Q is changing at rate Q', and if Q starts at a value Q_0 at time t_0, then what will Q be at time t? Since it grows at rate Q' for time $(t - t_0)$, it will increase by $Q' \times (t - t_0)$. Adding that to our starting value of Q gives

$$Q(t) = Q_0 + Q' \times (t - t_0). \tag{2.13}$$

This is the equation of a straight line in **point-slope form**. The linear approximation (2.11) is a special case of this, with I instead of Q, and with starting time $t_0 = 0$. We'll have a lot more to say about equations of lines and linear approximations in Chapter 3.

Returning to the SIR model, we can use the linear approximation (2.11) to study the past as well as the future. According to this approximation, yesterday we had around $2100 - 98 = 2002$ Actives, a week ago we had $2100 - 7(98) = 1414$ Actives, and a month ago we had $2100 - 30(98) = -840$ Actives.

How much do you trust those numbers? You should take them with a grain (or more) of salt. In particular, the estimate $I(-30) \approx -840$ is absurd. You can't have a negative number of Actives!

In making that estimate, something went *seriously* wrong. You should get into the habit of asking whether answers make sense. Reality check! Before reading on, take a minute to think about what went wrong with the method we used to get our negative answer.

The problem with our linear approximation is that it assumed that the rate of change of I is 98 people/day, it always was 98 people/day, and it always will be 98 people/day. In reality, we know that $I' = 98$ people/day *today*, and it's realistic to assume that I' won't change much in the next few days, but in a few weeks, or a few months, it could change by a lot. This means that we can trust our linear approximation when $t = 1$ or $t = 2$ or $t = -1$ or $t = -2$, but we shouldn't trust it when $t = 30$ or 365 or -30. The actual situation is shown in Figure 2.5. The linear approximation tracks $I(t)$ very closely for $-5 < t < 5$, but it does not account for the curvature in the graph of $I(t)$. As t gets more and more negative, the graph of $I(t)$ flattens out and stays positive, while

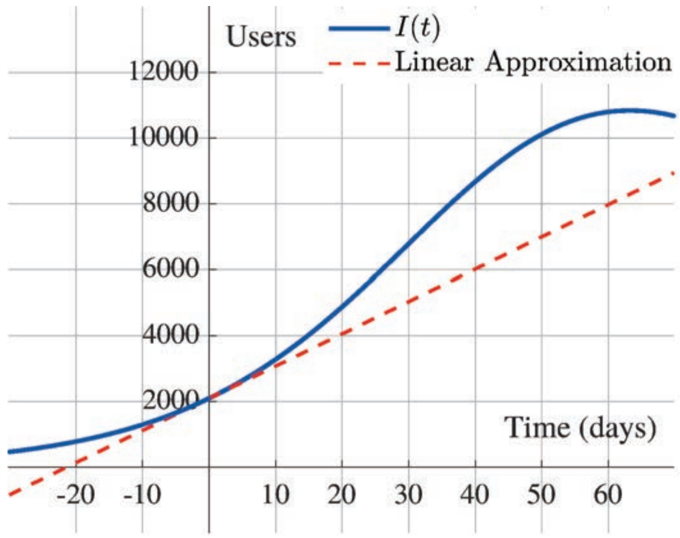


Figure 2.5. Active users over time. The linear approximation is negative when $t < -21.5$, but the actual $I(t)$ isn't.

the linear approximation goes negative. Meanwhile, as t gets more and more positive, $I(t)$ goes up and then down, while the linear approximation just keeps on growing.

To do better than our naive approximation, we need a way of estimating S' and I', and R', not just right now, but in the future and the past. Here's a way to estimate $S(10)$, $I(10)$, and $R(10)$.

First, we need to decide how far we can trust our linear approximation. Since I seems to be changing at about 5% each day, trusting it for a couple of days is reasonable. We're going to pick a time interval of two days, which we'll call Δt days, and predict the future $\Delta t = 2$ days at a time.

(1) Use the initial values $S(0)$, $I(0)$, and $R(0)$ together with the SIR equations to estimate $S'(0)$, $I'(0)$, and $R'(0)$.

(2) Use a linear approximation to estimate $S(2) \approx S(0) + 2S'(0)$, as well as $I(2) \approx I(0) + 2I'(0)$ and $R(2) \approx R(0) + 2R'(0)$.

(3) Use the values of $S(2)$, $I(2)$, and $R(2)$ together with the SIR equations to estimate $S'(2)$, $I'(2)$, and $R'(2)$.

(4) Use a linear approximation to estimate $S(4) \approx S(2) + 2S'(2)$, etc.

(5) Use *those* values and the SIR equations to approximate $S'(4)$, etc.

(6) Lather, rinse, repeat. At each time t, plug the estimated values of $S(t)$, $I(t)$, and $R(t)$ into the SIR equations to get estimated values of $S'(t)$, $I'(t)$, and $R'(t)$. Then use a linear approximation to compute $S(t + \Delta t) \approx S(t) + \Delta t\, S'(t)$, etc. Table 2.1 shows the results for the first ten days.

Likewise, we can go backward in time, using a small time step. We can take $\Delta t = -2$ days and go from $t = 0$ to $t = -2$ to $t = -4$, etc., all the way back to to

Table 2.1. Projecting forward with $\Delta t = 2$

t	$S(t)$	$I(t)$	$R(t)$	$S'(t)$	$I'(t)$	$R'(t)$
0	40,000	2,100	7,900	-168	98	70
2	39,664	2,296	8,040	-182	105.5	76.5
4	39,300	2,507	8,193	-197	113.5	83.5
6	38,906	2,734	8,360	-213	121.5	91.5
8	38,480	2,977	8,543	-229	130	99
10	38,022	3,237	8,741			

Table 2.2. Projecting backward with $\Delta t = -2$

t	$S(t)$	$I(t)$	$R(t)$
0	40,000	2,100	7,900
-6	40,923	1,559	7,518
-12	41,621	1,143	7,236
-18	42,139	832	7,029
-24	42,519	602	6,879
-30	42,796	433	6,771

$t = -30$. The results are shown in Table 2.2, with some times skipped and the values of S', I', and R' omitted to save space. As you can see, $I(-30)$ isn't negative at all.

This method still doesn't give exact answers, but *close is good enough!* If we want to compute $I(10)$ with greater accuracy, we can use the same algorithm with $\Delta t = 1$ instead of $\Delta t = 2$, for an answer of $I(10) = 3261$ instead of 3237. Of course, that requires ten iterations instead of five. If we want even more accuracy, we can take $\Delta t = 0.1$ and do 100 iterations, getting $I(10) = 3282$, or take $\Delta t = 0.01$ and do 1000 iterations, getting $I(10) = 3284$. If we're willing to do the extra work, or if we program a computer to do the extra work for us, we can have as much accuracy as we want. (However, our model is only an approximation of the real world, so even if we can solve our model to great accuracy, that doesn't necessarily mean that we can predict the future with that much accuracy.)

This is the 5th Pillar of Calculus: *One step at a time*. Every big change is made up of many little changes. If we can understand each little change, we can put the pieces together to understand the big change.

2.4. Theoretical Analysis: What Goes Up Has to Stop Before it Comes Down

So far we have used the 1st, 2nd, and 5th Pillars of Calculus. By making approximations and tracking the changes in our variables S, I, and R, and by putting a lot of short-term linear approximations together, we figured out how to obtain good projections of the future, and we were able to use those same projections to understand the past. Now we're going to tackle the question:

What is happening when $I(t)$ reaches its peak?

The key fact is that *the sign of I' tells you whether I is increasing or decreasing.* Whenever $I'(t) > 0$, $I(t)$ must be increasing. Whenever $I'(t) < 0$, I must be decreasing. At the very top of the curve, when $I(t)$ has stopped increasing and hasn't yet started decreasing, $I'(t)$ transitions from positive to negative. At that instant of time, we must have $I'(t) = 0$.

Let's figure out what is happening at that time. The SIR equations tell us that

$$I'(t) = aS(t)I(t) - bI(t) = I(t)\,(aS(t) - b)\,. \tag{2.14}$$

Since $I(t)$ is always positive, the sign of $I'(t)$ is the same as the sign of $aS(t) - b$. As long as $S(t) > b/a$, $I'(t)$ will be positive and $I(t)$ will increase. But eventually we will run out of Potential customers. When $S(t)$ drops below b/a, $I'(t)$ will become negative and $I(t)$ will start to decrease.

The number b/a is called a **threshold**. In our example, it was 16,667. This is **not** the value of $I(t)$ at the peak. Rather, it is the value of $S(t)$ when $I(t)$ hits its peak. In Figure 2.6, $I(t)$ hits its peak value of 10,870 when $t = 63.2$, which is when $S(t)$ passes through $16,667$.

If the threshold b/a is large and there aren't many Potentials to begin with, our product will fizzle from the start. If b/a is small and $S(0)$ is large, we will have a long period of growth before we saturate the market. As a marketer, our considerations for launching a product are the following.

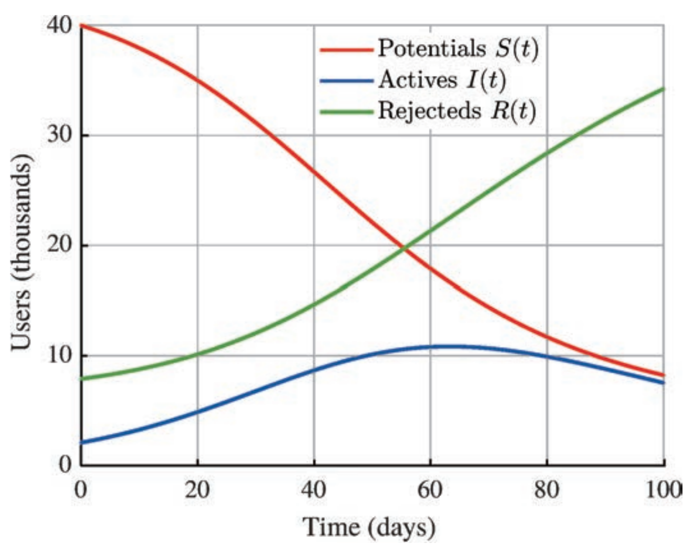


Figure 2.6. Market penetration via the SIR model. Over time, the number of Potentials drops, the number of Rejecteds rises, and the number of Actives rises, reaches a peak, and then falls,

(1) Make b as small as possible. Equivalently, make the usage time $T = 1/b$ as big as possible. This has to do with the quality of our product. The more exciting our product is and the more we provide updates and other continuing benefits to the users, the lower the attrition coefficient will be.

(2) Make a as big as possible. Do everything possible to encourage word-of-mouth communication between Active and Potential users.

(3) Pick a market where $S(0)$ is as big as possible. Even the best products can fizzle if they're directed at the wrong market. A lemonade stand will do a lot better in Texas in August than in Minnesota in February.

2.5. Epidemics

The SIR model originally came from epidemiology, not from business, and was developed to model the spread of disease. In that setting S stands for **Susceptible**, not Potential; I stands for **Infected**, not Active; and R usually stands for **Recovered**, not Rejected.[3] Infected people turn into Recovered people (with immunity against reinfection) when the disease has run its course. If we're talking about the flu, which lasts about one week, then the **recovery coefficient** b is 1/(7 days) and $R' = bI$. The rate at which Susceptible people become Infected is proportional to the number of times that an Infected person sneezes on a Susceptible person, which is proportional to the product $S(t)I(t)$. That is, $S'(t) = -aS(t)I(t)$, where a is still called the **transmission coefficient**.

In other words, the equations for market penetration and for the spread of disease are *exactly the same*! To a mathematician, they're the same problem. The difference is in interpreting the results and deciding what to do. With an epidemic we're trying to keep people from becoming Infected, while in marketing we're trying to get people to become Active.

In particular, if we are trying to contain an outbreak of disease, then we want to the threshold b/a to be as **big** as possible, and we want the number S of Susceptibles to be as **small** as possible. The main tools for doing this are the following.

(1) Vaccination. This transforms Susceptibles (who might get sick) directly into Recovereds. These are people who aren't sick, won't get sick, and can't spread the disease. If we can get the starting value of S to be below the threshold, then we can prevent the epidemic from even starting. This is called **herd immunity**. People often think that it's important to get a shot to protect themselves, and it is. But it's even more important for them to get that shot to protect the rest of us!

(2) Reduce the transmission coefficient a. Usually this is done through voluntary measures like encouraging sick people to stay home, wearing masks, keeping a safe distance from others, and avoiding handshakes. If a lot of people are sick, public health officials may need to take more drastic measures, such as closing schools, canceling large public events, quarantining infected individuals, or issuing shelter-in-place orders.

[3] When dealing with fatal illnesses or with quarantines, R sometimes stands for **Removed**. Either way, we're talking about individuals who can no longer spread the disease.

(3) It would also be good to increase b, which is the same as decreasing the time that people are sick and infectious. Unfortunately, this is difficult. For instance, anti-viral drugs can sometimes reduce the length of the flu by a small amount, but only by a small amount. For a disease like Covid-19 that can be spread by people who don't have symptoms, it takes extensive testing and quarantining to get infectious people out of circulation. Even if somebody hasn't *recovered*, we can try to *remove* them from the general population until they stop being infectious.

The bottom line is that our best defense against epidemics is vaccination, which usually has to be done in advance. Once an epidemic hits, the focus usually turns to improving sanitation and limiting person-to-person contact. Vaccines and face masks provide a lot more bang for the buck than anti-viral drugs.

2.6. Covid-19 and the SIR model

In December 2019, a strange form of pneumonia started affecting residents of Wuhan, China. The virus causing this disease was quickly identified, and the disease was named COVID-19 for "**CO**rona**VI**rus **D**isease 20**19**".[4] By April 2020 Covid-19 had spread around the world and had mutated into a more transmissible form, leading to drastic lockdowns to control the spread. Schools closed or went online, most people were told to stay home except for essential errands like buying food, and everybody was told to practice "social distancing". Despite these measures, Covid-19 continued to spread, and by mid-2022 had killed at least 6 million people worldwide (with some estimates being twice that big) including over one million Americans.

Everybody wanted to know what would happen next, which kept mathematical modelers (including the author) very busy. The details of their models could be complicated, but the main ideas were exactly the same as in the SIR model. In this section we'll go over some of the ways that modelers adapted the SIR model to apply to a world-wide pandemic rather than a localized outbreak.

Scaled models and the replication number R_0. One problem with the SIR model is that the parameters depend on the size of the city where the outbreak is happening. Typically, the larger the population, the smaller a fraction of the population that each person knows or meets and the smaller the transmission coefficient a will be. However, the product of a and the total population tends to be the same in different cities. For this reason, it's useful to write a scaled version of the model.

Let $T = S + I + R$ be the total population. This number is of course constant, or at least approximately constant. We let $s = S/T$, $i = I/T$, and $r = R/T$ be the *fractions* of the population that are susceptible, infected, or recovered, respectively, so that $s+i+r = 1$. Since $s' = S'/T$, $i' = I'/T$, and $r' = R'/T$, we can write rate equations

[4] After a while, people dropped the "19", stopped capitalizing the whole word, and just wrote "Covid" or even "covid".

for these quantities.

$$
\begin{aligned}
s' &= -aSI/T = & -aTsi, \\
i' &= (aSI - bI)/T = & aTsi - bi, \\
r' &= bI/T = & bi.
\end{aligned} \tag{2.15}
$$

Modelers usually use the Greek letters β and γ in this scaled model, with

$$\beta = aT; \qquad \gamma = b. \tag{2.16}$$

This makes the scaled SIR equations:

$$
\begin{aligned}
s' &= -\beta si, \\
i' &= \beta si - \gamma i, \\
r' &= \gamma i.
\end{aligned} \tag{2.17}
$$

Conditions do vary from place to place, with β being larger in cities that are more crowded and where people are less careful, but β varies much less than a and T vary separately. As a result, it's sensible to talk about "the" values of β and γ for each disease.

In the early stages of an epidemic, when s is close to 1, people are getting sick at rate $\beta si \approx \beta i$ and recovering at rate γi. That is, there are β/γ people getting sick for every person who recovers. This ratio is called the **basic replication number**,

$$R_0 = \frac{\beta}{\gamma}, \tag{2.18}$$

and it represents the average number of new people that each sick person infects. If $R_0 > 1$, then the epidemic grows. If $R_0 < 1$, then the epidemic fizzles out.[5] For Covid, R_0 was originally estimated to be between 2 or 3. However, the speed at which it spread through the USA and Europe suggests that R_0 was actually higher. Later variants evolved to be even more infectious, with values of R_0 around or even above 10.

Controlling an epidemic then amounts to getting s and R_0 as low as possible. In the short term, public health measures, such as wearing masks, closing schools, and staying home as much as possible, can reduce β, and so can reduce R_0. Testing helps, too. Sick people don't have to infect others. If they can be identified and quarantined (or **removed**) from the general population, they can be infect**ed** without being infect**ious**.

However, such efforts can't last forever. Sooner or later people want to return to normal, they start behaving as they did prepandemic, and R_0 goes back up. The only long-term solution is reducing the fraction s of susceptibles to below $1/R_0$, either through vaccination (best case) or natural infection (worst case).

The SEIR model. If Alice is sick and coughs on Bob, then Bob might get sick. If Bob then coughs on Carol, then Carol might get sick. However, if Bob coughs on Carol immediately after meeting Alice, then Carol is safe. It takes time for Alice's viruses to grow in Bob's body to the point that he can infect Carol.

[5]It's unfortunate that the basic replication number uses the same letter as the number of Recovered individuals. However, since we're looking at r rather than R and mostly care about s and i, this isn't really a problem.

To take this into account, modelers divide the population into four groups instead of three:

- **Susceptible** people (S, making a fraction $s = S/T$ of the population) are healthy but can become sick if they meet an infectious person.
- **Exposed** people (E, making a fraction $e = E/T$) are incubating the disease, but aren't sick enough to infect anybody else yet.
- **Infectious** people (I, making a fraction $i = I/T$) are a danger to everybody they meet.
- **Removed** people (R, making a fraction $r = R/T$) have recovered, have been quarantined, or have died. In any case, they aren't going to get sick again and they aren't going to infect anybody else.

Susceptible individuals become Exposed at a rate proportional to SI, just as in the SIR model. Exposed individuals become Infectious, and Infectious individuals become Removed, through the passage of time. After rescaling to express everything in terms of fractions of the population, our rate equations become

$$\begin{aligned} s' &= -\beta si, \\ e' &= \beta si - \alpha e, \\ i' &= \alpha e - \gamma i, \\ r' &= \gamma i. \end{aligned} \tag{2.19}$$

The new parameter α describes how long the incubation period is. On average, a person who gets infected is Exposed for time $1/\alpha$ and then is Infectious for time $1/\gamma$ before finally becoming Removed.

SEIR models behave a lot like SIR models, except that the time delay causes epidemics to grow more slowly than in SIR. In particular, we still talk about $R_0 = \beta/\gamma$, and the herd immunity threshold for starting to recover from the epidemic is still when $s = 1/R_0$.

The SIRS model. Immunity doesn't last forever. Over time, you lose the antibodies you had to a virus. Worse, germs can mutate into new forms that your old antibodies don't recognize.[6] Either way, being Removed isn't really permanent.

We take this into account by adding a new parameter δ to our model to indicate the rate at which Removed individuals rejoin the ranks of the Susceptibles.

$$\begin{aligned} s' &= \delta r - \beta si, \\ i' &= \beta si - \gamma i, \\ r' &= \gamma i - \delta r. \end{aligned} \tag{2.20}$$

The average duration of a person's immunity is $1/\delta$. This new model is called the SIRS model, for Susceptible-Infected-Removed-Susceptible. Similarly, we can turn the SEIR model into an SEIRS model.

[6] That's why you need a new flu shot each year. You may still be immune to last year's flu strain but not to this year's.

Asymptomatic transmission and other compartments. One of the most dangerous things about Covid was its ability to be transmitted by people who didn't even know that they were sick, either because they didn't yet have symptoms (**presymptomatic transmission**) or because they never got symptoms (**asymptomatic transmission**). At the beginning of the pandemic, roughly a third of Covid cases were asymptomatic. This fraction increased over time as Covid evolved. While asymptomatic carriers tended to shed fewer viruses than symptomatic or presymptomatic carriers, they tended to go out a lot more. This made disease control very difficult.

To take asymptomatic transmission into account, modelers divided the Infected category (or the Infectious category for the SEIR model) into two groups, I_s and I_a for "symptomatic" and "asymptomatic". They had different rates, β_s and β_a, of infecting others and different recovery rates, γ_s and γ_a. (They might take the same time to actually recover from the disease, but symptomatic individuals get removed from circulation a lot faster than asymptomatic carriers.) The SIR version of the model would then look like

$$\begin{aligned} s' &= -\beta_s s i_s - \beta_a s i_a, \\ i_s' &= \mu(\beta_s s i_s + \beta_a s i_a) - \gamma_s i_s, \\ i_a' &= (1-\mu)(\beta_s s i_s + \beta_a s i_a) - \gamma_a i_a, \\ r' &= \gamma_s i_s + \gamma_a i_a, \end{aligned} \tag{2.21}$$

where μ is the fraction of cases that develop symptoms.

You can also write down versions of the SEIR and SIRS models that take asymptomatic transmission into account. These models are more complicated than the basic SIR model, but they can be solved numerically in exactly the same way that we solved the SIR model. It's too complicated to do by hand, but programming a computer to solve equations (2.21) really isn't any harder than programming one to solve the basic model.

In practice, most models that were used to understand the spread of Covid had many more compartments. Besides sorting people according to whether they were susceptible, exposed, infectious, or removed, there were compartments for being hospitalized, for being in intensive care, and for dying. People were sorted by age and sex, since Covid hit older people and men much harder than it hit younger people and women. To understand patterns of disease spread within a single city, people were also sorted by occupation, race, and zip code, since it turned out that all three factors had a lot to do with who got sick and who didn't. Once vaccines were developed, there were separate compartments for people who were vaccinated or who had recovered from Covid. Some of the models were very complicated! However, they were still just souped-up versions of SIR, with more detail but the same underlying reasoning.

2.7. Chapter Summary

The Main Ideas.

- *Close is good enough.* Most of the time, we don't need an exact answer. A good simple approximation is often more useful than a complicated formula.

- Mathematical models help us to understand real-world problems.
  - Take what we know about our problem.
  - Throw out the unimportant details and keep the main features.
  - Express those features mathematically.
  - Use math to **analyze** the model.
  - **Interpret** the results. Turn numbers and equations into real-world conclusions. Without this last step, the rest is useless.
- *Track the changes.* Many useful models involve **rate equations** that describe how fast something is changing in terms of its current state. Many of the same equations show up in different settings. The SIR equations model market penetration. The same equations also model the spread of disease.
- Once you know how fast something is changing, you can use a **linear approximation** to predict its future or explore its past. This is accurate for a short time but can't be trusted over long time intervals.
- *One step at a time.* If you don't trust a linear approximation to predict what will happen next year, just use it to predict what will happen tomorrow. Then use those results to predict what will happen the day after tomorrow, then the day after that, and so on.
- *What goes up has to stop before it comes down.* You can learn a lot by studying the point when something stops moving; that is, when the rate of change equals 0.

Expectations. You should be able to:

- Model market penetration with a particular set of rate equations (the SIR model).
- Explain how the parameters in the SIR model relate to properties of the product and market being studied.
- Use the SIR equations, together with a linear approximation, to predict future usage rates.
- Iterate this process to generate a table of values for several times.
- Relate the sign of I' to whether your product is gaining or losing market share, and relate the size of I' to how fast this is happening.
- Use the same tools to analyze the trajectory of an epidemic.

2.8. Exercises

Rate of Change

2.1. Suppose that the fees collected by a consulting firm in March and May are \$1.2 million and \$1.25 million, respectively. Let $F(t)$ be the fees collected t months after March. (That is, March is $t = 0$.)

(a) What is the change in fees between March and May? Call your answer ΔF. What are the units for ΔF?

(b) What is the change in time between these two observations? Call your answer Δt. What are the units for Δt?

(c) At what *rate* did the monthly fees change from March to May? Call your answer m_1. What are the units for m_1?
(d) Assuming that fees are a linear function of time, how much can the business expect to collect in August? What value of Δt are you using to find your answer? Use the notation and answers from parts (a), (b), and (c).
(e) Again assuming a linear relationship, how much did the business collect in January of the same year? What value of Δt are you using to find your answer?
(f) Find a linear formula for $F(t)$.

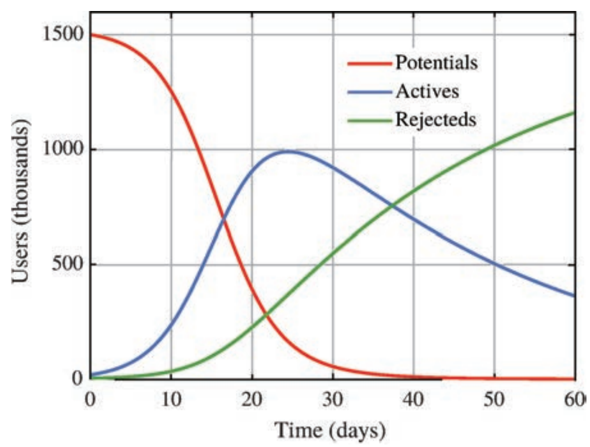
2.2. In Exercise 2.1 you developed a formula for the fees collected t months after March. Suppose that you wish to update your model using additional data. You observe that model you developed in Exercise 2.1 worked well through August, but that the *change* in the fees collected between the months of August and September was \$0.1 million. That is, the firm collected \$0.1 million more in September than in August.
(a) What is the monthly rate of change between August and September? Call your answer m_2.
(b) How does m_2 compare to the value of m_1 you found in Exercise 2.1c? From a business perspective, is this an improvement?
(c) Use m_2 to find an updated linear formula for predicting the fees collected for October through December. (This is similar to the formula you found in Exercise 2.1f.) State the correct domain for this expression.
(d) Use the information given in Exercise 2.1 and the result of Exercise 2.1c to write down a piecewise function that expresses the fees collected each month for the entire year (January through December), letting $t = 0$ represent the month of March.
(e) Use MATLAB to graph the piecewise function you found in part (d). Label your axes and give your graph an appropriate title. Use different colors for the two pieces of your function.
(f) Explain why using the piecewise function to express monthly consulting fees billed for the entire year is better than simply using the linear function you found in Exercise 2.1f for the entire year.

2.3. Explain the term "Rate Equation." How is a rate *equation* different from the rate of *change* m associated with a linear function?

2.4. Suppose we have a rate equation $F' = \cdots$ for a quantity $F(t)$. At a starting time t_0, we have $F(t_0) = F_0$, and the rate equation tells us that $F'(t_0) = R_0$.
(a) Explain, using proper notation, how we can approximate the change in $F(t)$ between times $t = t_0$ and $t = t_0 + a$ using a linear approximation.
(b) Explain, using proper notation, how we can approximate $F(t_0 + a)$.
(c) If $F(t)$ is a linear function, would your answers in parts (a) and (b) still be approximations? Explain.

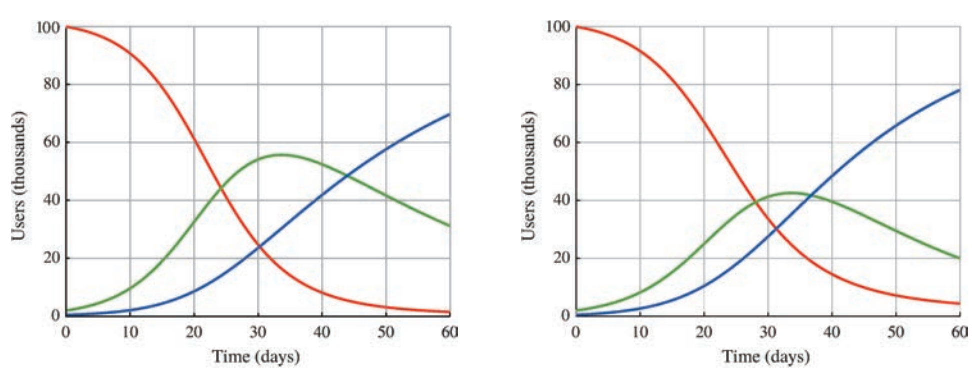
Interpreting SIR Models

2.5. Give approximate answers to these questions about the following graph of SIR model behavior:



(a) When does the number of Actives reach its peak? How many people are Active at that time?
(b) Initially, how many Potentials are there? How many days does it take to cut the Potential population in half?
(c) How many days does it take the Rejected population to reach 500,000? How many people have rejected the product by day 60?
(d) On what day is the size of the Active population increasing most rapidly? When is it decreasing most rapidly? How can you tell?
(e) How many people became Active at some time during the first 30 days? (Note that this is not the same as the number of Active people on day 30!) Explain how you found this information.

2.6. Below are two graphs depicting the marketing of a product according to the usual SIR model. The initial values $S(0)$, $I(0)$, and $R(0)$ and the transmission coefficient a are the same for both graphs. However, the two graphs correspond to different attrition coefficients b.



(a) For each graph, indicate which curve is S, which is I, and which is R.
(b) Which graph corresponds to the *larger* value of b? Explain.

2.7. Consider a marketing campaign with the simplified rate equation $I' = 35{,}000$ persons per day. (This is not the usual SIR model!) Suppose that there are 2,100,000 Actives on December 15.

(a) How many Actives will there be on Christmas Day (December 25)?
(b) When will the Active population reach 3,000,000?
(c) How many Actives were there on December 1?
(d) When did the campaign start? (When were there no Actives?)

2.8. A new product is launched in Smalltown and evolves according to the usual SIR equations

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

It appears that this product keeps customers active for 20 days, and that the initial populations of Potentials, Actives, and Rejecteds was

$$\begin{aligned} S(0) &= 10{,}000, \\ I(0) &= 100, \\ R(0) &= 0. \end{aligned}$$

One day later, it was observed that there were 20 *new* customers.

(a) Use the given information to complete the model. That is, find the transmission and attrition coefficients a and b.
(b) Using $\Delta t = 2$, estimate $S(4)$, $I(4)$, and $R(4)$.
(c) Find the threshold and explain its significance.
(d) Suppose that upon becoming active, customers are given personal customer support and continual product updates that extend the time they stay active to an average of 30 days. What effect will this have on the threshold for S? Do you think the improvement in the threshold is likely to be worth the extra cost of the intervention?

2.9. A company has released a new game app and has created a SIR model to predict the evolution of the market for this game. The SIR equations are

$$\begin{aligned} S' &= -.0000001SI, \\ I' &= .0000001SI - I/30, \\ R' &= I/30. \end{aligned}$$

The initial values at $t = 0$ are thought to be

$$S(0) = 2{,}000{,}000, \qquad I(0) = 21{,}000, \qquad R(0) = 15{,}000.$$

(a) Using the initial rates, estimate $S(1)$, $I(1)$, and $R(1)$.
(b) Using the SIR equations, calculate the rates of change $S'(1)$, $I'(1)$, and $R'(1)$, and then use these values to estimate $S(2)$, $I(2)$, and $R(2)$.
(c) Using the values of $S(2)$, $I(2)$, and $R(2)$ that you computed in part (b), calculate the rates of change $S'(2)$, $I'(2)$, and $R'(2)$. Then estimate $S(3)$, $I(3)$, and $R(3)$.

(d) Go back to the starting time $t = 0$ and to the initial values

$$S(0) = 2{,}000{,}000, \qquad I(0) = 21{,}000, \qquad R(0) = 15{,}000.$$

Recalculate the values of $S(2)$, $I(2)$, and $R(2)$ by using a time step of $\Delta t = 2$. This only requires a single round of calculations, using $S'(0)$, $I'(0)$, and $R'(0)$. How do your answers compare to those computed in part (b)? Which estimates do you think are most accurate, those in part (b) or those in part (d)? Why?

A **product bust** is when a product never catches on, with the number of active users never rising above the initial number. This can happen if the transmission coefficient is too small or if the attrition coefficient is too large.

2.10. In this exercise we consider variations of the situation in Exercise 2.8.

(a) Suppose that the product only keeps customers active for an average of 4 days instead of 20. Will a product bust occur?

(b) Suppose, in addition to the change in b, that $a = 0.00005$. Now will a product bust occur?

(c) In an SIR model where $S(0) = 40{,}000$ and $b = 1/14$, how big does a need to be to avoid a product bust?

(d) In a product bust, are there any new customers? Explain.

2.11. (a) Construct a SIR model for a product that keeps people's interest for an average of 20 days, where an average potential customer meets about 0.01% of the active population each day, and where it takes eight contacts with an Active before a Potential becomes Active.

(b) How many Potentials are needed to avoid a product bust?

2.12. One way to create more interest in a product is to provide rewards for Actives who refer Potentials. By creating an incentive for Actives to reach out to Potentials, a reward program can increase the transmission coefficient.

(a) Suppose, in the setting of Exercise 2.9, that a reward program is put into effect that doubles the chance that a Potential will become Active. What is the new transmission coefficient?

(b) Compute the new threshold if the reward program in part (a) is put into effect, and compare it to the original threshold in Exercise 2.9.

(c) Suppose that, for a different product, $S(0) = 20{,}000$, $b = 1/14$, and $a = 0.000003$. In order to be profitable, this product must have a threshold under 15,000. Will this product be profitable?

(d) Suppose that, for the product in part (c), the company initiates a reward program, hoping to bring the threshold under 10,000. By what factor must the reward program alter the transmission coefficient to achieve this goal?

2.13. Consider two products with the same transmission coefficient a and the same initial conditions. They differ only in the average length of time that someone stays Active. Which product has the higher threshold level: the one with the higher average usage time or the lower average usage time? Explain.

2.14. **Product Trolls**. Imagine the evolution of a product in a population with the following categories of people: Potential users (S), Active users (I), Rejected users (R), and Product Trolls (T). Product Trolls are individuals who continually give negative feedback about the product. Assume the following.

- Trolls never become Actives.
- Once a user becomes Rejected, that user never becomes Potential again.
- Potentials become Active through contact with Actives, as in the SIR model. The rate of transmission is α.
- Actives become Rejected through contact with a Troll (instead of with the passage of time). The rate of contact of the average Active with the average Troll is β.
- Potentials and Rejecteds can become Trolls through the passage of time. This happens with rate constants γ for Potentials and δ for Rejecteds.
- The system is closed. Changes in population only happen when a person moves from one class to another. This also means that the total population $N = S + I + R + T$ is constant.

(a) Draw a diagram describing the interactions between these different classes of people.

(b) Write down a system of rate equations that describes these interactions.

2.15. **Herd Immunity.** Suppose there is a measles outbreak in a small community of 40,000 people, almost all of them Susceptible, with $b = 1/(14 \text{ days})$ and $a = 0.000003$.

(a) What is the threshold in this problem? How many people will get sick before the epidemic reaches its peak?

(b) Now suppose that 10,000 people (out of 40,000 total) are immune to measles, either because they already had the disease or because they were vaccinated. How many will get sick before the epidemic reaches its peak?

(c) If enough people are immunized, the epidemic will fizzle from the start, just like a product bust. How many people would have to be immunized for that to happen?

(d) When enough people are immunized to keep outbreaks from spreading, the population is said to have **herd immunity**. Explain how the number of vaccinations needed for herd immunity is related to the threshold b/a.

(e) If you were a public health official, what would you say to somebody who doesn't want to get vaccinated and who insists that their (not) getting vaccinated is none of your business?

2.16. **Replication Number.** In an epidemic, the ratio of how many people get sick to how many recover is $aSI/(bI) = aS/b$. At the beginning of an epidemic, when S is almost the entire population, this is the same as the **basic replication number** $R_0 = \beta/\gamma$ that we saw in our scaled SIR model. If $R_0 > 1$, then the epidemic will grow, at least at first. If $R_0 < 1$, it will fizzle out.

(a) Suppose that the total population is 40,000. If $R_0 = 1.2$, what is the threshold? How many people will get sick before the epidemic reaches its peak? What if $R_0 = 10$? In general, how are R_0 (the total population) and the threshold related?

(b) In a town with total population T and a disease with basic replication number R_0, how many people would need to become immune to achieve herd immunity?

Food for thought: The number of people needed to reach herd immunity is the same as the number of people who have to get sick *before* the epidemic reaches its peak. However, a substantial number of people can still get sick *after* the peak. For instance, suppose that $R_0 = 2$. The peak will be reached when half the population has gotten sick, but nearly 30% of the population will get sick later. Only about 20% will avoid the disease altogether. This is one reason why it is much better to immunize people before an epidemic than to wait for herd immunity during an epidemic.

2.17. Consider a product evolution that progresses according to the usual SIR model, *except* that Rejecteds become Potentials again after an average of c days. (*"I haven't played that game in a while. Maybe I should play it again."*) After that, they can become Active, just like the other Potentials. Modify the usual SIR equations to take this new feature into account. Your equations should involve three unspecified parameters: a, b, and c.

Exponential Growth. In many cases, the rate of change of a quantity is proportional to the quantity itself. We will see many different examples of that in Chapter 6, but the two simplest are money earning interest and population growth. We will examine these in Exercises 2.18–2.20.

If a bank account containing \$10,000 earns \$150/year in interest, then an account at the same bank containing \$20,000 should earn \$300/year in interest. Either way, the interest is 1.5%/year of the bank balance.

If a city of 100,000 is growing at a rate of 3000 people per year, then we would expect a similar city of 200,000 (or even the same city a few years later) to grow by 6000 people per year. That is, if $P(t)$ is the population at time t, then the **net growth rate** P' is proportional to P:

$$P' = kP,$$

where k is a constant. In this example, $k = .03$/year, or 3%/year.

2.18. In the equation $P' = kP$ for the population of a city, the number k is called the **per capita growth rate**, with the Latin phrase "per capita" literally meaning "per head". Explain why the units for k are 1/year.

2.19. In 2020, the population of Australia was 25 million and was growing at 1.36%/year, while the population of the United States was 330 million and was growing at 0.71%/year. Assume that both countries' populations keep growing at these per capita rates.

(a) Let $A(t)$ and $U(t)$ be the populations of Australia and the United States, respectively. Write down the rate equations that govern the growth of these functions.

(b) What were the net growth rates $A'(2020)$ and $U'(2020)$ in 2020?

(c) In general, does a country with a higher per capita growth rate necessarily have a higher net growth rate?

(d) On average, how many seconds did it take for the population of Australia to increase by 1 in 2020? For the population of the United States? (A year is about 31.6 million seconds.)

Food for Thought: At those per capita growth rates, the populations of Australia and the United States would become equal in 2417, at which time each country would have about 4.3 *billion* people. Exponential growth is great when it comes to money, but it can be scary in other cases.

2.20. A bank account is earning interest at a constant percentage rate. When the account had \$3200, it was growing at a rate of \$79/year.
  (a) Write an equation that links the net growth rate (meaning P', not the interest rate) to the bank balance.
  (b) At a later time, the bank balance is growing at \$100/year. What is the balance at that time?

Scaling the Time Variable. In a rate equation, the quantity Y' has units, namely whatever units we use for Y divided by whatever units we use for time. If Y is people and t is measured in days, then Y' is measured in people/day. In Section 2.6, we generated a scaled model, where S, I, and R were replaced with dimensionless quantities s, i, and r that are proportional to S, I, and R. In Exercises 2.21–2.23, we will see what happens when we similarly replace the time t with a dimensionless variable.

2.21. Suppose that $i' = 0.014$/week. That is, i is increasing at a rate of 0.014 per week.
  (a) How fast is i increasing per *day*? How fast is i increasing per *year*? Which is bigger, the growth per year or the growth per day?
  (b) Let t_d, t_w, and t_y be the number of days, weeks, and years since a fixed starting time. How are the variables related? Which is bigger, t_d or t_y?

To make a model dimensionless, we pick a fixed unit of time T. T might be a second, an hour, a day, a year, or 3723.849 years. We then measure time in units of T. That is, we define a new dimensionless variable $\tau = t/T$ that counts how many T's have elapsed since $t = 0$. If T is small, then τ will be big, while if T is big, then τ is small. A very *large* number of nanoseconds equals a very *small* number of centuries. For variables Y and P that change in time, we let $\dot{Y}$ and $\dot{P}$ denote the changes in Y and P *per change in* τ.

2.22. How is $\dot{Y}$ related to Y'? (*Hint*: If Y' is constant, how much will Y change by the time that $\tau = 1$?)

2.23. Suppose that $P' = rP$. Let $T = r^{-1}$.
  (a) Compute $\dot{P}$ in terms of P.
  (b) Compute T when $r = 1\%$/year, when $r = 3\%$/month, and when $r = 0.06$/minute.

Food for Thought: You should discover that the trajectory of $P' = rP$ looks essentially the same for all of these values of r, only we trace out that trajectory at different speeds. If we can understand the solution to $P' = rP$ for one value of the parameter r (say, for $r = 1$), then we can understand the solution to $P' = rP$ for all values of r.

2.24. Consider the scaled SIR equations (2.17). Let $T = \gamma^{-1}$. Write down a system of equations for $\dot{s}$, $\dot{i}$, and $\dot{r}$.

Food for Thought: The resulting equations still have one dimensionless parameter, namely the basic replication number R_0. The details of a product launch or epidemic depend on a lot of different parameters and initial conditions, but the general shape of the curve only depends on R_0.