3 Simulation and Analysis of the SIR Model

While the SIR model introduced in the preceding lectures is interesting, two joys of working with a mathematical model are (i) simulation from the model, and (ii) mathematical analysis of the model. Simulation involves choosing a set of conditions at which the model should start, running the clock forward, and then examining what happened. This can be valuable for intuition, but may generalize poorly: what if the initial conditions or parameters change? One would have to simulate anew. Mathematical analysis involves working directly with the equations to find general principles.

3.1 Simulation from the SIR model

One easy way to try to understand the behavior of a model is to simulate from it. The ordinary differential equations (ODEs) from the previous lecture, phrase in terms of the number of people in each compartment, give us a recipe to do conduct such a simulation.

$$\begin{aligned}\dot{S} &= -\frac{\beta SI}{N} \\ \dot{I} &= \frac{\beta SI}{N} - \gamma I \\ \dot{R} &= \gamma I\end{aligned} \tag{1}$$

We'll start with a classic and simple method to solve, called Euler's Method.

3.1.1 Euler's Method

The idea behind Euler's Method goes back to what you learned in Calc I,

$$\frac{dy}{dt} \approx \frac{\Delta y}{\Delta t} \, ,$$

provided that Δt is small. Rearranging this equation, we get

$$\Delta y \approx \frac{dy}{dt} \Delta t \, ,$$

which we should read as *the change in y is approximately equal to y's slope $\times$ the change in t*. This means that to determine the change in y from its existing value, I should (i) compute its current rate of change, and (ii) multiply by a small increment in t. As a result, we can start from an initial value of y, which we'll refer to as y_0, and incrementally step forward along the solution of a differential equation $dy/dt = f(t, y)$.

$$\begin{aligned}y_{n+1} &= y_n + \Delta y = y_n + \Delta t \cdot f(t_n, y_n) \\ t_{n+1} &= t_n + \Delta t \, ,\end{aligned} \tag{2}$$

producing a sequence of y and t values, indexed by n, which approximate the solution of our differential equation. Because we use the current value of the slope $f(t_n, y_n)$ to look forward to the next point y_{n+1}, we call this method **Forward Euler**.[1] In class, we'll implement this approach to solve the SIR model, and you'll be able to return to this thinking to solve other models introduced in future lectures.

3.2 Learning from Simulations

Let's examine the output of simulations of the SIR model. In Figure 1, we've plotted the solutions to the

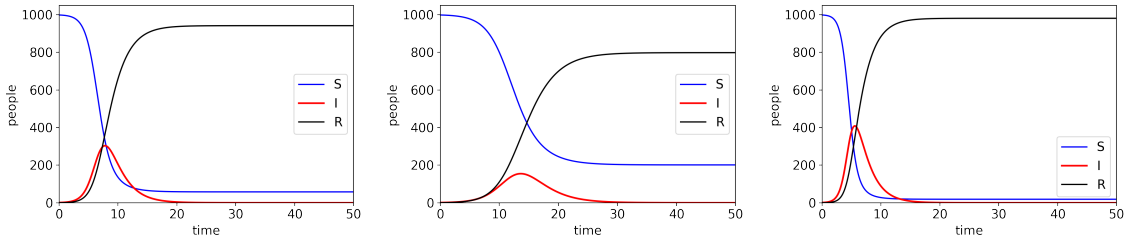

Figure 1: Three simulations from the SIR model, varying β, with $I_0 = 1$ initial infection, and $S_0 = 999$ initially susceptible, and no one initially recovered. See text for discussion.

SIR model starting from a single infection in an otherwise susceptible population. The values of γ are the same in each figure, but the values of β have been changed. Using what you know about the meaning of β, which plot corresponds to the highest value of β, and which the lowest?

Now consider another set of three simulations in Figure 2, but this time with a fixed value of β and varying γ. Using what you know about the meaning of γ, which plot corresponds to the highest value of γ, and which the lowest?

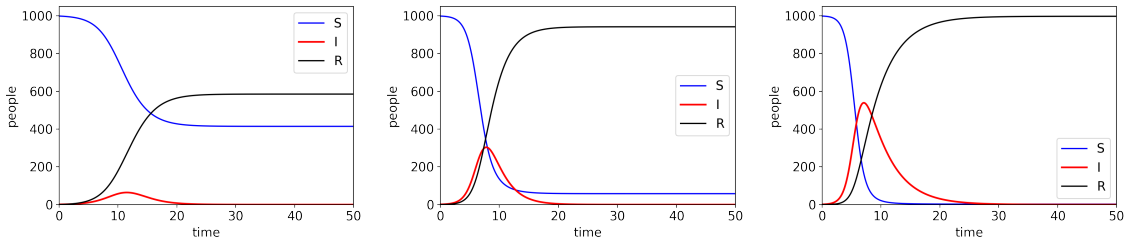

Figure 2: Three simulations from the SIR model, varying γ, with $I_0 = 1$ initial infection, and $S_0 = 999$ initially susceptible, and no one initially recovered. See text for discussion.

[1] There's also a Backward Euler method, also called Implicit Euler. Such methods are taught in numerical analysis courses!

These simulations are helpful for developing a sense of intuition. When we raise β, the red curve,[2] grows more quickly, peaks earlier, and peaks higher. Similarly, when we raise β, the total number of people recovered at the end of the simulations increases. Therefore the order of the panels in Fig. 1 is medium, low, high.

Our intuition about γ should be the opposite. Raising γ means increasing the rate at which people recover. That means that they should spend less time infected, because they recover more quickly. Spending less time infected means creating fewer infections, and therefore a higher γ should mean fewer infections, a lower peak, and fewer total people recovered. Therefore the order of the panels in Fig. 2 is high, medium, low.

Finally, notice that there are two aspects unifying all of these plots. First, the epidemic curve rises, peaks, and falls, meaning that there was an epidemic at all. This is important, and we'll return to it in the math. Second, notice that eventually, the system stabilizes, and in every case we see eventually that $I \to 0$, while $S \to S_\infty$ and $R \to R_\infty$ at some non-zero values. We'll also return to this important point mathematically.

3.3 Analyzing the SIR Model

Simulations are useful, but one downside is that we cannot quickly learn general patterns from simulations. For example, where does the system stabilize, and where *can* it stabilize? Will there be an epidemic? When will the epidemic peak? Is the system stable to the introduction of new infections from "outside" the modeled population? How big will the epidemic be? All of this, and more, with a little math...

3.3.1 Equilibria

A good first place to start with a model is to ask about its equilibria. An **equilibrium** is a set of values of the variables (not the parameters) such that nothing is changing. Mathematically, the equilibrium (singular) or equilibria (plural) can be found by setting all derivatives to zero. When we do that for our equations, what do we find?

We'll strategically start from the equation for R, which tells us that

$$\dot{R} = \gamma I = 0$$

and implies that $I_{\text{eq}} = 0$. Note that there is no other possibility here, as long as $\gamma > 0$. We can interpret this condition as saying that, to be at equilibrium, there must be zero infecteds!

What about the other equations?

$$\dot{S} = -\frac{\beta S I}{N} = 0$$

[2] The red curve, tracking the number of infected people over time, is sometimes called the epidemic curve because it tracks the unfolding of the epidemic.

is already satisfied by our previous discovery that $I_{\text{eq}} = 0$. This means that this equation tells us nothing new. In other words, the previously established fact that $I = 0$ means that $\dot{S} = 0$, i.e. that S is not changing over time. Similarly,

$$\dot{I} = \frac{\beta SI}{N} - \gamma I = 0$$

is also already satisfied by $I = 0$. In other words, when $I = 0$, then $\dot{I} = 0$ too, and thus I is not changing over time.

At this point, we have used up all three of our equations, which means we have established (i) that $I_{\text{eq}} = 0$ means that $\dot{S} = \dot{I} = \dot{R} = 0$, and (ii) that there are no additional constraints on variables S_{eq} or R_{eq}. However, S and R can't be just anything (at equilibrium, and in general). Specifically, we know that $S_{\text{eq}} + R_{\text{eq}} = N$, because the population size is N. This means that $R_{\text{eq}} = N - S_{\text{eq}}$. As a result, our equilibrium is:

$$(S, I, R) = (S_{\text{eq}}, 0, N - S_{\text{eq}}) \,, \tag{3}$$

for any values of S_{eq} between 0 and N.[3]

What can we say about this equilibrium? Most importantly, we call this a **disease-free equilibrium** because it is exactly that: there are 0 people in compartment I, and the system is at equilibrium. Many systems we will study have a disease-free equilibrium.[4] In contrast, other systems, in which there is an equilibrium with $I_{\text{eq}} > 0$, will be described as having a **disease-endemic equilibrium**. More to come on this topic, but for now we will simply observe that the only equilibrium of the basic SIR model is disease free.[5]

3.3.2 Will there be an epidemic?

A fundamental question we can ask is: *will there be an epidemic?* This question is important because if the answer is "no" then we need only think about small chains of transmission, but not large outbreaks. If the answer is "yes" then we should prepare for an epidemic that infects a substantial portion of the population.

First, let's develop some intuition. When we think about an outbreak, what we're saying is: the number of infected people grows to a large number. To infect a large number, the number of infecteds must grow, and this means that $\dot{I} > 0$. Indeed, if the number of infected people never grows, you can't have an outbreak!

[3] You could just as easily say that the equilibrium is $(N - R_{\text{eq}}, 0, R_{\text{eq}})$. Do you see why?

[4] Suppose you are talking to your 7 year old niece, and she asks for an intuitive explanation, with no equations, about why the disease-free state is an equilibrium of this system. In a 1-3 sentences, can you explain to her?

[5] When writing a compound adjective such as "disease free", the rule is that you throw in the hyphen when the modifier comes before the thing it is modifying, as in "disease-free equilibrium," but you omit the hyphen when the modifier comes afterward, as in "the equilibrium was disease free."

To learn more, we can set $\dot{I} > 0$, and see what it implies[6] about our other variables and parameters,

$$\dot{I} = \frac{\beta S I}{N} - \gamma I > 0$$
$$\left(\frac{\beta S}{N} - \gamma\right) I > 0 \tag{4}$$
$$\implies$$
$$\frac{\beta S}{N} - \gamma > 0 \tag{5}$$
$$\implies$$
$$\frac{S}{N} > \frac{\gamma}{\beta}\,. \tag{6}$$

which we might also simply write as $s > \frac{\gamma}{\beta}$, using the proportion s. By the way, it is important that you be able to follow these steps so that you can explain *why* Step (4) implies Step (5), and so on.

3.3.3 When will the epidemic peak?

This conclusion gives us a handy mathematical rule for the SIR model: when $s > \frac{\gamma}{\beta}$, the epidemic has the potential to grow, and when $s \leq \frac{\gamma}{\beta}$, it does not.

What about when $s = \frac{\gamma}{\beta}$? This condition corresponds means that $\dot{I} = 0$, yet it is possible that $I \neq 0$. This means that this value of s tells us about the proportion of the population that is susceptible when the number of infected people achieves a local maximum, i.e.

$$s_{\text{peak}} = \frac{\gamma}{\beta}\,. \tag{7}$$

3.3.4 What is the basic reproductive number?

The **basic reproductive number**[7] is defined as the expected number of secondary infections caused by a single primary infection in a totally susceptible population, with no interventions or countermeasures in place. This number is written as R_0, and pronounced *R-naught*.[8]

Before discussing R_0 in the context of our SIR model, let's think about what R_0 means. When $R_0 > 1$, it means that, on average, in a susceptible population, each infection leads to *more than one* additional infection. This would mean that the number of infected individuals has grown. In some cases, the number

[6] Recall that the $\implies$ symbol can be read as "implies."
[7] *alt:* basic reproduction number
[8] Confusingly, this R_0 has nothing to do with the R for the Recovered compartment.

of infected individuals can grow quite a bit! For example, measles has an R_0 of 12-18[9] as compared with influenza whose R_0 is around 1.3.[10]

One way of calculating R_0 for the SIR model is to directly calculate, via a thought experiment, the number of secondary infections per single infection in an entirely susceptible population. First, remember that the rate of new infections per unit time is given by the incidence $\beta SI/N$. Second, we let $S = (N-1)$ and $I = 1$, arriving at an incidence of $\beta(N-1)/N$ secondary infections per time, which we simplify to simply β on the assumption that N is very large. Third, to convert from secondary infections per time to secondary infections per single infection, we multiply by the amount of time an infected person remains infected on average, $1/\gamma$.[11] Thus, we arrive at a classic result of the SIR model,

$$R_0 = \beta/\gamma \, . \tag{8}$$

Note that with this equation in hand, we can also rewrite Eq. (7) as $s_{\text{peak}} = \frac{1}{R_0}$.

3.3.5 What is the effective reproductive number?

The basic reproductive number R_0 is defined using a hypothetical totally susceptible population, with no interventions or countermeasures in place. What if there *are* countermeasures or there *are* some folks who are no longer susceptible? In these scenarios, the **effective reproductive number** R_e quantifies the expected number of secondary infections caused by a single primary infection.

Following the derivation above, but assuming that we can no longer assume that everyone is susceptible (i.e., no longer setting $S = N-1$), we get

$$R_e = \frac{S}{N} \cdot \frac{\beta}{\gamma} = sR_0 \, . \tag{9}$$

This tells us that by depleting the fraction of the population that is susceptible s, the reproductive number decreases. Reassuringly, if we set $s = 1$, we recover a reproductive number of R_0.

The depletion of susceptibles is just one way that we can affect R_e. We could also do so by other countermeasures such as vaccination, masking, social distancing, and pharmaceuticals. Intuitively, an epidemic will start to die out when $R_e < 1$ because each infection is replaced by less than 1 subsequent infection.[12]

This view of R_e is valuable because it provides us with a target for interventions: find a way to push R_e to below 1, and the epidemic will start to dwindle. However, during an actual epidemic we can never directly

[9] Why might there be a range of estimates for R_0?

[10] Wikipedia includes a simple table of various R_0 values with citations to original estimates.

[11] Don't worry if you're unsure where this comes from, as it will be a homework question!

[12] In fact, just in the case of the depletion of susceptibles, if we evaluate R_e when $s = s_{\text{peak}}$, we get $R_e = 1$, indicating that the peak is the point at which the next generation of infections is as large as the previous, where we have just tipped from growth $R_e > 1$ to decline $R_e < 1$.

calculate R_e, and besides, it varies over time anyway because, in addition to any variation in countermeasures, the number of susceptibles is typically always changing. For this reason, we define the **real-time reproductive number** R_t as simply the reproductive number at time t. During the course of a flu season, for instance, R_t will exceed 1 until the peak and then dip below 1 as the season subsides.

3.3.6 Is the disease free equilibrium stable?

A fundamental question that we can ask of an equilibrium is: *is the equilibrium locally stable?* This question means asking whether, if one nudges the system away from the equilibrium just a little bit, the system will return back to the equilibrium (stable) or will move away from it (unstable). Put differently, does the perturbation away from equilibrium grow or shrink?

Mathematically, what we'll do is take the equilibrium value of $I_{\text{eq}} = 0$ and imagine that we perturb it just a little bit, i.e. $I = I_{\text{eq}} + \varepsilon$, for $\varepsilon > 0$ small.[13] In other words, what if $I = \varepsilon$? Will the perturbation ε grow?

First, note that, because $I = \varepsilon$, we can take a derivative of both sides to get $\dot{I} = \dot{\varepsilon}$. Plugging into our original equation for $\dot{I}$, we get

$$\dot{\varepsilon} = \beta \frac{S}{N} \varepsilon - \gamma \varepsilon$$
$$\dot{\varepsilon} = (\beta s - \gamma)\, \varepsilon$$

This looks like the classic exponential growth equation

$$\dot{y} = ky \,, \tag{10}$$

except instead of variable y and growth rate k, we have variable ϵ and growth rate $\beta s - \gamma$. Thus, in the first moments after a perturbation, we get exponential growth (or decay) at rate $\beta s - \gamma$ when $\beta s - \gamma$ is positive (or negative). The disease-free equilibrium is unstable when a perturbation grows, and thus, we conclude that the disease free equilibrium is unstable when $\beta s - \gamma > 0$.

Where have we seen this before? Rearranging, we see that actually, the disease-free equilibrium is unstable when $s > \frac{1}{R_0}$. Similarly, it is stable whenever $s < \frac{1}{R_0}$.

Something exciting has occurred! We have just shown that there is a special susceptible population *fraction* such that, whenever there are *more* than that fraction susceptible, then a few incoming infections can destabilize the disease-free equilibrium and cause an epidemic. And, whenever there are *fewer* than that fraction susceptible, then incoming infections cannot destabilize the disease free equilibrium.

This value, $s = \frac{1}{R_0}$ is called the **herd immunity threshold**. It tells us that if, at some point, the susceptible fraction is smaller than this critical value, an epidemic cannot occur. This means that any process that

[13] This is a common move in math, where you nudge or perturb away from a particular value by a small amount. In fact it's so common that we almost always use the same letter, ε (epsilon). We assume that ε is positive and arbitrarily small

depletes the population of susceptibles may, eventually, lead to a dynamics that drives infections out of the population naturally. This concept will return later when we study vaccination as a particular countermeasure which decreases the susceptible population.

As a final note, observe that the critical value of s that defines herd immunity is identical to s_{peak}, the value of s at which $\dot{I} = 0$. Here, we get a second interpretation of the herd immunity threshold: it is the susceptible fraction at which the growth rate of infections hits 0, and after which it is negative.

3.3.7 How big will the epidemic be?

How many people will have been infected by the end of an epidemic? How many will remain susceptible? We are, in essence, asking to calculate

$$\lim_{t\to\infty} S(t) = S_\infty \qquad \text{and} \qquad \lim_{t\to\infty} R(t) = R_\infty \ . \tag{11}$$

Note that these limits must exist because both S and R are bounded between 0 and N, and S can never increase while R can never decrease.[14] It is also possible to show that $\lim_{t\to\infty} I(t) = 0$, i.e. that the epidemic will eventually die out.[15] Thus, eventually, everyone is either susceptible or recovered.

To approach this calculation, consider the equations for $\dot{i}$ and $\dot{s}$, and divide them,

$$\frac{\dot{i}}{\dot{s}} = \frac{\beta si - \gamma i}{-\beta si} = -1 + \frac{\gamma}{\beta s} \ .$$

Then, multiply both sides by $\dot{s}$ to get

$$\dot{i} = -\dot{s} + \frac{\gamma}{\beta}\frac{\dot{s}}{s} \ .$$

Integrate both sides from $t = 0$ to $t = \infty$, and use $i(0) = \epsilon \to 0$ and $s(0) = 1 - \epsilon \to 1$ to get[16]

$$0 = -(s_\infty - 1) + \frac{1}{R_0}\left(\ln s_\infty - \ln 1\right) \ ,$$

which can be manipulated to get[17]

$$s_\infty = e^{-R_0(1-s_\infty)} \qquad \text{and} \qquad r_\infty = 1 - e^{-R_0 r_\infty} \ . \tag{12}$$

Both of these equations are transcendental, meaning that we cannot explicitly solve for one variable in terms of the other.

[14] The limit of something that is bounded and monotonic must exist! Consult your favorite real analysis text.
[15] This is rather tricky to show. Come to office hours!
[16] Good practice! Can you work this step out?
[17] Also good practice! See if you can use the equation above to show the next one.