4 Two Variations on the SIR Model

In previous weeks, we described the SIR model as the workhorse model of infectious disease epidemiology. We describe it this way because it can be easily modified to arrive at other useful models. This week, we'll introduce two such models and analyze them.

4.1 The SIS Model

In the SIR model, people who recover from infection enter the R compartment, never to leave. They are therefore often referred to as "removed" because they no longer participate in the dynamics. What if, instead, recovery from infection meant a return to susceptibility? The simplest version of such a model is called the SIS model.

The SIS model assumes that infectious individuals recover at rate γ and simply return to being susceptible. This results in the following equations for the number of susceptible or infected individuals.

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

As before, we can rewrite Eq. (1) using population proportions,

$$\begin{aligned} \dot{s} &= -\beta si + \gamma i \\ \dot{i} &= \beta si - \gamma i \end{aligned} \tag{2}$$

and write a simple flow diagram for the dynamics (Figure 1).

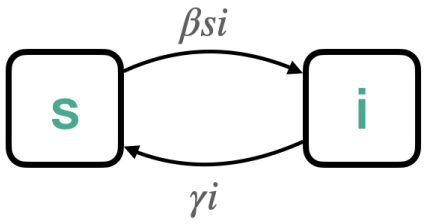

Figure 1: A flow diagram for the SIS model.

One nice observation we can make is that last week's analysis of R_0 can be redone for the SIS model identically. In other words, even in the SIS model, we have $R_0 = \frac{\beta}{\gamma}$. An intuitive way of understanding this is that, in the particular scenario that R_0 asks us to imagine, with a single infection and everyone else susceptible, the two models look identical.

The equations above are redundant because we know that $i = 1 - s$, an observation that enables us to rewrite Eq. (2) as

$$\begin{aligned} \dot{i} &= \beta(1-i)i - \gamma i \\ &= \ldots \\ &= (\beta - \gamma)\, i \left(1 - \frac{i}{1 - \frac{1}{R_0}}\right) , \end{aligned} \tag{3}$$

where the middle steps are not shown.[1]

4.1.1 Equilibria of the SIS model

What are the equilibria of this SIS model? Setting Eq. (3) equal to zero, we get

$$i_{\text{eq}} = 0 \quad \text{and} \quad i_{\text{eq}} = 1 - \tfrac{1}{R_0} \tag{4}$$

Note that $i_{\text{eq}} = 0$ corresponds to a disease-free equilibrium, and this should align with our intuitive understanding of the SIS model: if there are no infections, then everyone must be susceptible, and the system will continue in that disease-free state indefinitely.

What about the other equilibrium? Because this equilibrium depends on the value of R_0, we should interpret it under a few scenarios, using our knowledge of the basic reproduction number's meaning.

Scenario 1 What if R_0 is large? In this scenario, $i_{\text{eq}} > 0$ but is, at most, 1. This would represent a disease-endemic equilibrium. As R_0 grows larger, the disease endemic equilibrium would approach the state where everyone is infected ($i = 1$).

Scenario 2 What if R_0 falls toward 1? Plugging $R_0 = 1$ into the equilibrium equation, this would mean an equilibrium of $i = 0$. In other words, when R_0, both of our system's equilibria are at 0, with no infections.

Scenario 3 What if $R_0 < 1$? In these scenarios, this second equilibrium value is negative, $i < 0$. Because i represents the fraction of the population who is infected, we interpret this equilibrium as mathematically relevant, but epidemiologically irrelevant; it will never be reached!

We can interpret these three scenarios that's telling us that, while there will always be a disease-free equilibrium, the presence of a realistic disease-endemic equilibrium, depends on the value of R_0. Specifically, when $R_0 > 1$ (and we have the potential for epidemic growth), we also see the emergence of a disease-endemic equilibrium which can grow no higher than $i_{\text{eq}} = 1$.

[1] These middle steps are actually just a bit of algebra. Grab a pencil and see if you can find your way to the completed Eq. (3)!

4.1.2 The logistic interpretation

If one squints at Eq. (3), it may look like an old friend,

$$\dot{x} = rx(1 - \frac{x}{K}) \, , \tag{5}$$

also known as the logistic differential equation. The logistic equation appears in population growth models where we typically refer to r as the intrinsic growth rate and K as the carrying capacity. This means that r is the growth rate when the population is small and not yet size limited, and K is the total population that the system can bear, above which the population will collapse down toward K, and below which the population will grow up toward K. This makes K a stable equilibrium of the logistic population growth equation. While $x = 0$ is an equilibrium of the logistic population growth equation, it is unstable because $r > 0$.

By drawing this parallel between our SIS model and the logistic population growth equation, we gain a few things. First, it's just lovely to see old friends where you least expect them, and it reminds us that by learning to be scientists and mathematicians where we can say, mathematically, "*this* looks like *that*" there is much to gain.[2]

Second, we get "for free" that $K = 1 - \frac{1}{R_0}$ is the carrying capacity, and $r = \beta - \gamma$ is the intrinsic growth rate. This means that, as long as $\beta - \gamma > 0$, we will get growth of any non-zero i toward the endemic equilibrium. In other words, as long as the intrinsic growth constant is positive, the system will move toward the endemic equilibrium.

Third, we can see that R_0 shifts the carrying capacity of our system down and up—at most up to a capacity of 1. This means that we can think of R_0 as shifting the population's capacity to carry infections, with higher R_0 moving the capacity higher and lower R_0 shifting the capacity lower.

Finally, the logistic equation typically assumes that $r > 0$, i.e., growth, which for us would correspond to $\beta - \gamma > 0$ which is equivalent to $R_0 > 1$.[3] However, what happens in a logistic growth model when the intrinsic growth rate is negative? Here, we get population collapse toward the equilibrium at zero. The same is true for our SIS model: when $R_0 < 1$, the population of infections people collapses toward zero, i.e., toward the disease free state.

[2] So many systems in our world are alike when you look under the hood to see how they work! Getting to experience the inquiry into life's complexity is one of the joys of mathematical modeling, interdisciplinarity, and science in general. Richard Feynman speaks beautifully about this in this short clip from a 1981 interview. Feynman was a giant in 20th century physics, and he was also problematic in his behavior, viewed through both our 21st century lens and at the time. Nevertheless, this clip strikes me as getting at something lovely about how we get to see the world, and so I share it with a disclaimer.

[3] Do you see why this is so?

4.1.3 A closed form solution

One final advantage to the comparison to the logistic growth model is that we can borrow its closed form solution.[4] The solution to Eq. (5) is

$$x(t) = \frac{K}{1 + \frac{K - x_0}{x_0} e^{-rt}} \tag{6}$$

which means that the solution to Eq. (3) is

$$i(t) = \frac{1 - \frac{1}{R_0}}{1 + \frac{1 - \frac{1}{R_0} - i_0}{i_0} e^{-(\beta - \gamma)t}} , \tag{7}$$

which is, y'know, not the worst.[5]

4.2 SIR model with birth and death

Let's continue our exploration of SIR models with a different twist: birth and death. We will not yet consider mortality associated with the disease, but instead we'll be focusing only on natural births and deaths. This means that we will assume that individuals in our model die at a per-capita rate μ, and so, our typical S, I, and R equations will now include terms of $-\mu S$, $-\mu I$, and $-\mu R$, respectively.[6].

To simplify our analysis, we will assume deaths are balanced by births, which means that, if the total death rate is $\mu S + \mu I + \mu R$, then the total birth rate should be identical. Using $S + I + R = N$, this means that the total birth rate is μN. We will make the further assumption that all individuals are born susceptible.[7] As a consequence, our equations are

$$\begin{aligned} \dot{S} &= -\frac{\beta S I}{N} + \mu N - \mu S \\ \dot{I} &= \frac{\beta S I}{N} - \gamma I - \mu I \\ \dot{R} &= \gamma I - \mu R \, , \end{aligned} \tag{8}$$

where the new terms are colored red. Once more, with a constant N due to the balance of birth and death, we can quickly rewrite the equations in terms of proportions s, i, and r as well, a flow diagram for which is shown in Fig. 2.

[4] Use separation of variables on Eq. (5) with the initial condition that $i(0) = i_0$. If you haven't gone through these steps, it would be a good idea to practice.

[5] This is an understatement. It's actually wonderful to have a closed form solution, and generally rare when it comes to modeling nonlinear systems!

[6] For calibration, note that this also implies that the typical lifetime is μ^{-1}. However, also note that in this simple modeling framework, the distribution of lifetime lengths would be exponential, which is, of course, wildly incorrect for humans!

[7] Can you think of infectious diseases where this assumption is incorrect?

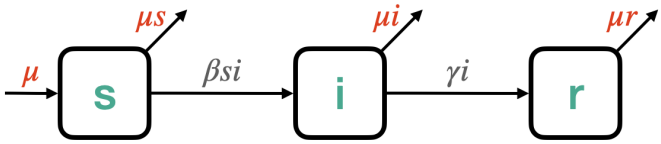

Figure 2: A flow diagram for the SIR model in population normalized units, with per-capita birth rate and per-capita death rate μ .

4.2.1 R_0 for an SIR model with birth and death

Imagine a population with enormous turnover. In fact, imagine cranking up μ so that an infected person is just as likely to leave the I compartment by dying as by recovering. A typical infection would have a 50/50 chance of ending because of recovery vs ending because of death. This, in turn, should affect the typical number of people that can be infected by a single infected person in an otherwise susceptible population. In other words, we should see μ somewhere in our expression for R_0.

Recall that one of our steps in calculating R_0 for the SIR model was to ask: how long does the typical infection last? Previously, the dynamics of I in the absence of new infections was simply $\dot{I} = -\gamma I$, so the typical infection lasted γ^{-1} units of time. Now, however, our equation is $\dot{I} = -(\gamma + \mu)I$, so the typical infection lasts $(\gamma + \mu)^{-1}$ units of time.[8] Consequently,

$$R_0 = \frac{\beta}{\gamma + \mu} \, . \tag{9}$$

4.2.2 Equilibria of the SIR model with birth and death

As with any system of differential equations, we find equilibria by setting all derivatives to zero and learning about the relationships between our variables. Analyzing the equation for $\dot{I}$ implies

$$I\left(\frac{\beta S}{N} - \gamma - \mu\right) = 0 \quad \Longrightarrow \quad I_{\text{eq}} = 0 \text{ or } S_{\text{eq}} = \frac{\gamma + \mu}{\beta} N = \frac{1}{R_0} N \, .$$

The two possibilities $I_{\text{eq}} = 0$ and $S_{\text{eq}} = \frac{1}{R_0} N$ represent different possible equilibria, so we will follow each to its conclusion. First, however, note that analyzing the equation for $\dot{R}$ implies

$$\gamma I - \mu R = 0 \qquad \Longrightarrow \qquad I_{\text{eq}} = \frac{\mu}{\gamma} R_{\text{eq}} \, .$$

This equation means that, at equilibrium, values for I and R must be proportional to each other.

[8] Recall your homework calculations to show why this is so!

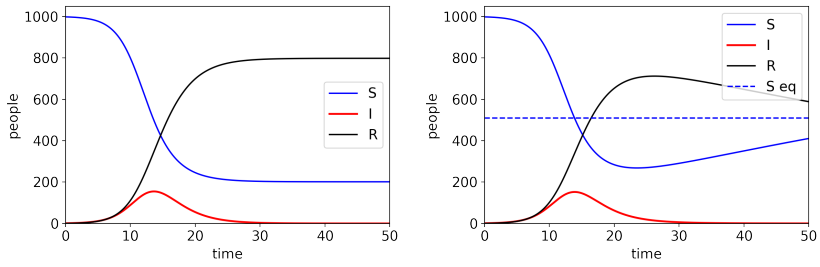

Figure 3: Both figures show SIR model dynamics with $\beta = 1$ and $\gamma = 0.5$. On the left, there is no birth or death ($\mu = 0$), while on the right, $\mu = 0.01$. A reference line shows S_{eq}.

The easier equilibrium is when $I_{\text{eq}} = 0$, the disease-free equilibrium. This, combined with our analysis of the $\dot{R}$ equation, implies that $R_{\text{eq}} = 0$. Further combined with the fact that $N = S + I + R$, we get $S_{\text{eq}} = N$. In other words, if no one is infected, eventually everyone is susceptible. This equilibrium differs from the typical SIR equilibrium because eventually, all recovereds die and are reborn susceptible.

The more interesting equilibrium is if the $\dot{I}$ equation is satisfied by $S_{\text{eq}} = \frac{1}{R_0}N$. This equation produces valid equilibrium values of $S_{\text{eq}} \leq N$ only when $R_0 \geq 1$, decreasing as R_0 increases.[9] Perhaps unsurprisingly, this means that our "more interesting" equilibrium can exist only when $R_0 \geq 1$.

Now, recall the relationship from the $\dot{R}$ equation that $I_{\text{eq}} = \frac{\mu}{\gamma}R_{\text{eq}}$. This means that the $N\left(1 - \frac{1}{R_0}\right)$ people who are *not* susceptible must be divided between R and I, with $\frac{\mu}{\gamma}$ infected people for every 1 recovered person. This means that

$$S_{\text{eq}} = N\frac{1}{R_0}\,, \qquad I_{\text{eq}} = N\left(1 - \frac{1}{R_0}\right)\frac{\mu}{\gamma+\mu}\,, \qquad R_{\text{eq}} = N\left(1 - \frac{1}{R_0}\right)\frac{\mu}{\gamma+\mu}\,. \tag{10}$$

4.2.3 Dynamics of the SIR model with birth and death

We can easily modify our differential equation solver (Forward Euler, for instance) to include our new parameters for birth and death. When we do so, and we pick a scenario with $R_0 > 1$, we can produce plots that help us to see how our dynamics have changed. Figure 3 shows, side by side, a scenario with no birth or death $\mu = 0$ and a scenario with $\mu = 0.01$.

A side-by-side comparison of the dynamics shows that the two appear similar for the first $t = 10$ units of time, but are visibly different by $t = 20$, with the susceptible population slowly increasing toward its equilibrium value (dashed line). However, we should beware an increase in susceptibles, given that our past analyses have shown that outbreaks are possible when s is sufficiently large. In fact, if we integrate forward longer in time, we see fascinating behavior (Figure 4).

[9] Notice that if $R_0 < 1$, the disease will die out on its own, leading us to the other equilibrium where $S_{\text{eq}} = N$.

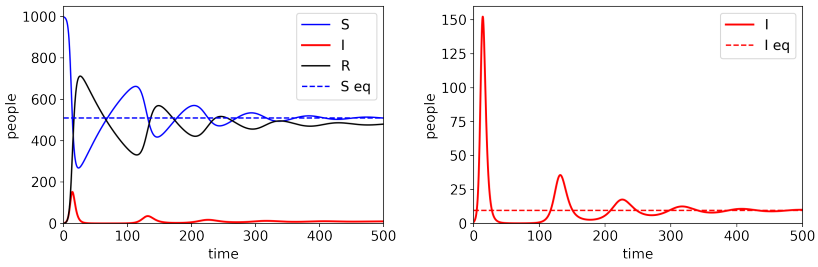

Figure 4: Both figures show SIR model dynamics with $\beta = 1$, $\gamma = 0.5$, and $\mu = 0.01$. Reference lines show S_{eq} (left) and I_{eq} (right). This wider view shows much richer dynamics than might be inferred from Fig. 3 which integrated only until $t = 50$. To gain insights into the dynamics of the I curve, the right-hand panel is rescaled and plots only $I(t)$.

What in the world is going on in Figure 4? Here, we have exactly the right conditions for a disease-endemic equilibrium. In fact, with $\beta = 1$, $\gamma = 0.5$ and $\mu = 0.01$, we should see $I_{\text{eq}} \approx 9.61$, and indeed we do (Fig. 4, right panel). However, we also see fluctuations, such that the infection *almost* dies out entirely while the susceptible population slowly regrows due to births. Eventually, the susceptible population is sufficiently large that an outbreak occurs anew! Eventually, in this scenario, the oscillations are damped in the sense that they eventually settle and the system's nonlinear oscillations around its equilibrium become smaller and smaller in amplitude.

4.2.4 Mathematical analysis of the endemic equilibrium

To analyze this equilibrium, we're going to develop and use a powerful technique from nonlinear dynamics. In our analysis of the typical SIR model, we imagined ourselves to be sitting at the equilibrium, and then considered what would happen if we nudged the system away, just a small amount, from that equilibrium. Now we're going to do the same thing, but instead of applying our thinking to a single equation for i, we're going to have to look at all three equations at once.

First, we'll simplify things by considering the normalized system,

$$\begin{aligned} \dot{s} &= \mu - \beta si - \mu s \\ \dot{i} &= \beta si - \gamma i - \mu i \\ \dot{r} &= \gamma i - \mu r \ , \end{aligned} \tag{11}$$

at the endemic equilibrium $\langle s, i, r \rangle_{\text{eq}} = \left\langle R_0^{-1}, (1 - R_0^{-1}) \frac{\mu}{\mu+\gamma}, (1 - R_0^{-1}) \frac{\gamma}{\mu+\gamma} \right\rangle$. We know that if we plug these values into our differential equations, all the derivatives will be zero. What if, instead, we plug in $\langle s, i, r \rangle_{\text{eq}} + \langle \Delta s, \Delta i, \Delta r \rangle$? Then, we could ask about the behaviors of these little nudges Δs, Δi, and Δr. Do they grow, shrink, or something else? For shorthand, we're going to refer to $\langle s, i, r \rangle_{\text{eq}}$ as $\mathbf{x}_{\text{eq}}$ and $\langle \Delta s, \Delta i, \Delta r \rangle$ as $\Delta \mathbf{x}$.

Let's first look at our equation for $\dot{s}$, and evaluate it at the nudged variables above. We get

$$\dot{s}\left(\mathbf{x}_{\text{eq}} + \Delta\mathbf{x}\right) = \mu - \beta(s_{\text{eq}} + \Delta s)(i_{\text{eq}} + \Delta i) - \mu(s_{\text{eq}} + \Delta s) \; .$$

Multiplying everything out, we get

$$\dot{s}\left(\mathbf{x}_{\text{eq}} + \Delta\mathbf{x}\right) = \mu - \beta s_{\text{eq}} i_{\text{eq}} - \beta \Delta s \, i_{\text{eq}} - \beta s_{\text{eq}} \Delta i - \beta \Delta s \Delta i - \mu s_{\text{eq}} - \mu \Delta s \; .$$

The red terms are special because we know that they sum to zero.[10] The blue term is special because it is the product of two small things, and thus small enough to ignore as we get arbitrarily close to the equilibrium point. And we know that $\dot{s}(\mathbf{x}_{\text{eq}}) = 0$. As a result, we get the approximation

$$\frac{d\Delta s}{dt} \approx -\beta \Delta s \, i_{\text{eq}} - \beta s_{\text{eq}} \Delta i - \mu \Delta s \; .$$

Finally, we're going to collect terms and rewrite the equation as

$$\frac{d\Delta s}{dt} \approx \left(-\beta i_{\text{eq}} - \mu\right) \Delta s - \beta s_{\text{eq}} \Delta i + 0 \, \Delta r$$

Following this same procedure for our equations for i and for r, we get

$$\frac{d\Delta i}{dt} \approx \beta i_{\text{eq}} \Delta s + \left(\beta s_{\text{eq}} - \gamma - \mu\right) \Delta i + 0 \, \Delta r \; ,$$

and

$$\frac{d\Delta r}{dt} \approx 0 \, \Delta s + \gamma \Delta i - \mu \Delta r \; .$$

You have surely noticed that there are some zeros included in the equations above that give a nice ordered structure to these equations. This is not a coincidence, as we can now write the following in matrix-vector form,

$$\frac{d\Delta\mathbf{x}}{dt} \approx \begin{pmatrix} -\beta i_{\text{eq}} - \mu & -\beta s_{\text{eq}} & 0 \\ \beta i_{\text{eq}} & \beta s_{\text{eq}} - \gamma - \mu & 0 \\ 0 & \gamma & -\mu \end{pmatrix} \Delta\mathbf{x} \; . \tag{12}$$

Equation (12) is a linear approximation that tells us how a small nudge near our equilibrium point will behave. Before we analyze it, however, a few remarks.

First, the entries in this matrix are special. The entries in the top row are the partial derivatives of our $\dot{s}$ equation with respect to s, i, and r, respectively. The middle row contains the three partial derivates of the $\dot{i}$ equation. And the bottom row contains the three partial derivatives of the $\dot{r}$ equation. This matrix is called a **Jacobian matrix** and we write it as $J(\mathbf{x})$.

[10] Why is this so?

Second, the partial derivatives (and thus the Jacobian matrix) are evaluated at the equilibrium values of s, i, and r. In other words, the matrix is, in particular, $J(\mathbf{x}_{\text{eq}})$.

Thus, it turns out that the equivalent of linearizing a single-variable nonlinear equation around its equilibrium is to instead write

$$\frac{d\Delta\mathbf{x}}{dt} \approx J(\mathbf{x}_{\text{eq}})\,\Delta\mathbf{x}$$

This provides a nice shortcut to all the steps we showed above. You can simply compute all the partial derivatives, form the Jacobian, and evaluate it at the equilibrium point. Then, the final step: examining its eigenvalues.

Before we do that, we'll make a few simplifications to the entries of the matrix. Note that

$$\begin{aligned}
\beta s_{\text{eq}} &= \frac{\beta}{R_0} = \gamma + \mu \\
\beta i_{\text{eq}} &= \beta\left(1 - \frac{1}{R_0}\right)\frac{\mu}{\gamma+\mu} = R_0\left(1 - \frac{1}{R_0}\right)\mu = (R_0 - 1)\,\mu \\
-\beta i_{\text{eq}} - \mu &= -(R_0 - 1)\,\mu + \mu = -R_0\mu\,.
\end{aligned} \tag{13}$$

Making these substitutions into the Jacobian matrix in Eq. (12), we get

$$J = \begin{pmatrix} -R_0\mu & \gamma+\mu & 0 \\ (R_0-1)\,\mu & 0 & 0 \\ 0 & \gamma & -\mu \end{pmatrix} \tag{14}$$

To determine the behavior of the linear system in Eq. (12), we need only examine the eigenvalues of the Jacobian.[11] Its characteristic equation is

$$(-\mu - \lambda)\big[\,(-R_0\mu - \lambda)\,(-\lambda) - (\gamma+\mu)(R_0-1)\mu\big] = 0$$

which can be manipulated to find

$$(\mu + \lambda)\big[\lambda^2 + R_0\mu\lambda + (\gamma+\mu)(R_0-1)\big] = 0$$

Thus, the three eigenvalues of J are

$$\lambda_1 = -\mu \quad \lambda_{2,3} = -\frac{R_0\mu}{2} \pm \mathbf{i}\sqrt{(R_0-1)\,(\gamma+\mu) - \left(\frac{R_0\mu}{2}\right)^2}\,, \tag{15}$$

where $\mathbf{i}$ is the imaginary number $\sqrt{-1}$. The real parts of all three eigenvalues are negative, meaning that our equilibrium is locally stable, and points near the equilibrium will be attracted toward that equilibrium.

[11] Do you remember how to find the eigenvalues of a matrix? Do you remember how to compute the determinant of a 3×3 matrix using cofactors?

However, we also see a pair of complex eigenvalues, which indicate oscillation. In other words, we expect to see our system move toward the equilibrium, but in an oscillatory way.

One way to see this in action is to plot the trajectories of two of our variables (say, $s(t)$ and $i(t)$), but instead of showing their trajectories vs t, as in previous plots, we'll plot $i(t)$ vs $s(t)$. We call these axes the **phase plane** and our plot is called a **phase portrait**. When we plot these trajectories in the phase plane, it is clear why things that look like oscillations when plotted vs t are also considered spirals (Fig. 5). In this case, we spiral "inward" toward the equilibrium, and this is another way of seeing that the oscillations are damped, and the disease endemic equilibrium is stable.

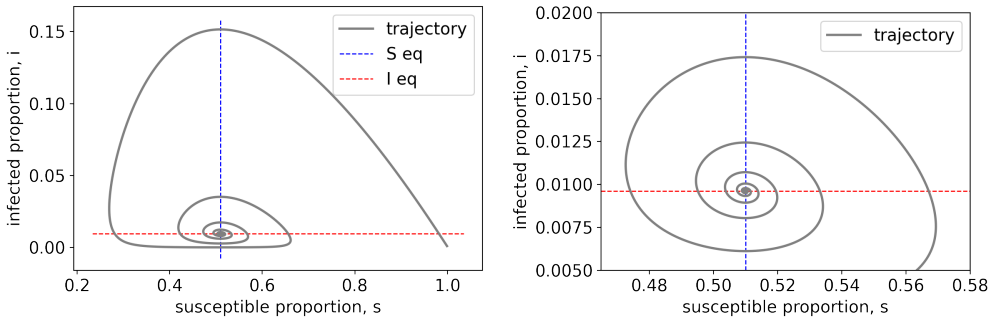

Figure 5: Phase portraits of $i(t)$ vs $s(t)$ with equilibrium lines shown, in both a wide view (left) and zoomed view (right). SIR model with birth and death, parameters: $\beta = 1$, $\gamma = 0.5$, $\mu = 0.01$.

4.3 Remarks and Ethics

The analyses we've presented are both modifications of the SIR model. The SIS model is highly predictable and reduces to a one-dimensional system, the logistic equation. In contrast, the SIR model with birth and death is harder to predict, with a new formulation for R_0 than the ones we've previously seen, and with these oscillations on the way to the endemic equilibrium. Nevertheless, these models are similar in the sense that they both admit an endemic equilibrium, telling us that we should prepare for the possibility that our modeled infectious disease stays in the population forever.

What are the ethical issues we should consider? Most salient are the oscillations we observed in the SIR model with birth and death. The endemic equilibrium has the potential to be misleading about the dynamics because, even in our simulations, the population does not experience the equilibrium in finite time! Instead, they experience wild swings: an epidemic spikes and then subsides to nearly zero, with the spikes repeating until eventually, the dynamics settle toward the equilibrium. Thus, we can see that the equilibrium values here are important, but the dynamics on the way to the equilibrium are just as important. When communicating the results of models to people and especially to those who might make decisions based on the models, we are required as modelers to put effort into explaining these important dynamics.