

SIR Model Explained

Samuel Atkins

December 2020

1 The Model

There are three variables that the SIR model depends on, S , I , and R . $S(t)$ is the number of susceptible individuals, $I(t)$ is the number of infected individuals, and $R(t)$ is the number of recovered individuals. We define the following variables to make our future calculations more simple:

$$\begin{aligned}s(t) &= \frac{S(t)}{N} \\ i(t) &= \frac{I(t)}{N} \\ r(t) &= \frac{R(t)}{N}\end{aligned}$$

2 Model Assumptions

1. The total number of individuals in the population, N , is constant
2. Individuals do not immigrate to or from the population in question
3. There are no births
4. There is no loss of immunity (we could assume a loss of immunity and derive different equations)

Since the total number of individuals in the population, N , is constant $s(t) + i(t) + r(t) = 1$ because $S(t) + I(t) + R(t) = N$. Given our assumptions, we are effectively assuming that the only way an individual leaves the susceptible group is by becoming infected. Thus, the number of susceptible individuals is always decreasing.

3 The Differential Equations

The rate at which the number of susceptible individuals decreases is dependent on how many infected individuals the susceptible group comes into contact with. If an infected individual comes into contact with b susceptible individuals per day, then, on average, each infected individual generates $b \cdot s(t)$ new infected individuals per day. Furthermore, a fixed fraction of the infected group, k , recovers every day. Using these assumptions, we arrive at the following differential Equations:

$$\begin{aligned}\frac{dS}{dt} &= -bs(t)I(t) \implies \frac{ds}{dt} = \frac{-bs(t)I(t)}{N} \\ \frac{ds}{dt} &= -bs(t)i(t)\end{aligned}$$

The $i(t)$ factor is present because given our assumption, if there is just one infected person then the number of susceptible people decreases by a factor of $bs(t)$. Therefore, if there are 2 infected people, then the number of susceptible people decreases by a factor of $2bs(t)$. For $I(t)$ infected people, the number of susceptible people decreases by a factor of $bs(t)I(t)$.

$$\frac{dR}{dt} = kI(t) \implies \frac{dr}{dt} = ki(t)$$

This follows from the fact that k infected individuals recover every day. Now, since the number of infected people increases as a function of $bs(t)I(t)$ and $kI(t)$, we have the following:

$$\frac{dI}{dt} = -kI(t) + bs(t)I(t) \implies \frac{dI}{dt} = -\frac{dR}{dt} - \frac{dS}{dt} \implies \frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0$$

SIR Equations Summary

$$\begin{aligned}\frac{ds}{dt} &= -bs(t)i(t) \\ \frac{di}{dt} &= -\frac{dr}{dt} - \frac{ds}{dt} \\ \frac{dr}{dt} &= ki(t)\end{aligned}$$

4 Initial Conditions

To solve this set of differential equations, we need to supply some initial conditions. We assume that a tiny fraction of our population is infected and the rest of our population is healthy. Thus, for a population of size 5,000,000, we have the following:

$$S(0) = 4,999,990$$

$$I(0) = 10$$

$$R(0) = 0$$

$$s(0) = 0.999998 \approx 1$$

$$i(0) = 2 \times 10^{-6}$$

$$r(0) = 0$$

5 Euler's Method

We do not numerically solve the SIR equations. Instead, we use Euler's method. For a single time dependent variable, x , Euler's method is as follows:

$$x_i = x_{i-1} + \frac{dx(t-1)}{dt} \cdot \Delta t$$

Since we have three variables in the context of the SIR model, we have three Euler formulas:

$$s_i = s_{i-1} + \left. \frac{ds}{dt} \right|_{i-1} \cdot \Delta t$$

$$i_i = i_{i-1} + \left. \frac{di}{dt} \right|_{i-1} \cdot \Delta t$$

$$r_i = r_{i-1} + \left. \frac{dr}{dt} \right|_{i-1} \cdot \Delta t$$

Given the SIR differential equations, we have:

$$\begin{aligned}s_i &= s_{i-1} - b s_{i-1} i_{i-1} \cdot \Delta t \\ i_i &= i_{i-1} + (-k i_{i-1} + b s_{i-1} i_{i-1}) \cdot \Delta t \\ r_i &= r_{i-1} + k i_{i-1} \cdot \Delta t\end{aligned}$$

6 Finding Optimal Parameters for a Dataset

There are a few ways to go about this. The first way is to estimate b and k based on the recovery rate and R value typically associated with an SIR outbreak. The other way, which is far more involved, is to define a loss function between the predictions of the model and the data. Then, compute the sum of that loss function over all of the observations. This method gives us a way to quantify the difference between our model's predictions and the data. Using this method, we can then perform a grid search over all of the possible parameters to obtain the optimal model.