CS 7A - Spring 2020 - Modeling Epidemics with the SIR Model. Due 4/13/20

As the first step in the modeling process, we identify the independent and dependent variables. As is often true, it makes sense to take the independent variable as time t, which we will measure in days.

Initially, we look at the following dependent variables:

$$\begin{aligned} S(t) &= \text{the number of susceptible individuals.} \\ I(t) &= \text{the number of infectious individuals.} \\ R(t) &= \text{the number of recovered individuals.} \end{aligned}$$

All of these dependent variables are measured in the unit, "individuals", or people...humans. It's desirable in modeling to have non-dimensional measure, which we're going to obtain by dividing each of these variables by the total size of the population in question, N. This will result in a dimensionless number between 0 and 1; a percentage of the population. The new variables will be lower-case letters:

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

The next step in the modeling process is to make a bunch of simplifying assumptions. We'll assume that $I(0)$, the number of originally infected people, which is small, but positive. We'll assume that all other people are susceptible, so that $S(0) = N - I(0)$. Obviously, $R(0) = 0$, since, initially, no one has recovered yet. We'll also assume that for all t, $S(t) + I(t) + R(t) = N$ is the whole population. No can one enter, no one can leave and if somebody dies...well...that's a form of recovery? Ok, nobody dies. That simplifies things a lot!

Now we make simplifying assumptions that no one is ever added to the susceptible group, and the only way to leave that group is to become infected. So the time rate of change of $S(t)$ depends on the size of $S(t)$, and the amount of contact between susceptibles and infecteds. To simplify matters, suppose that each infected person has a fixed number b of contacts per day that are sufficient to spread the disease. Not all of these contacts are with susceptible individuals, so, if we assume (again, for simplicity's sake) that there is a homogeneous mixing of the population, the fraction of these contacts of the whole population that are with $S(t)$ is $S(t)/N = s(t)$. Therefore, on average, each infection individual generates $b \cdot s(t)$ new infected individuals per day. This is huge simplification that leads to the equation for the time rate of change of $S(t)$:

$$\frac{dS(t)}{dt} = -b \cdot s(t) \cdot I(t) \Leftrightarrow \frac{ds(t)}{dt} = -b \cdot s(t) \cdot i(t)$$

Since we are aiming to program a computer to implement this model, and computers don't have the capability to embrace the infinite or the infinitesimal, we'll make do with the approximation

$$\frac{\Delta s(t)}{\Delta t} \approx -b \cdot s(t) \cdot i(t)$$

$$\boxed{\Rightarrow \Delta s \approx -b \cdot s(t) \cdot i(t) \Delta t}$$

A simple model for the recovered population is to assume a fixed fraction k of the infected group will recover on any given day. If, for example, the average duration of infection is three days, then, on average, one-third of the currently infected population recovers each day. This leads to an equation for the time rate of change of $R(t)$:

$$\frac{dR(t)}{dt} = k \cdot I(t) \Leftrightarrow \frac{dr(t)}{dt} = k \cdot i(t)$$

$$\boxed{\Rightarrow \Delta r(t) \approx k \cdot i(t) \Delta t}$$

These two equations, together with $s(t)+i(t)+r(t)=1$ allow us to get and expression for the time rate of change of $i(t)$. Since the sum of these percentages is 100%, a constant, we have that

$$\begin{aligned}
&\frac{d}{dt}\left(s(t)+i(t)+r(t)\right)=0\\
\Leftrightarrow\ &\frac{ds(t)}{dt}+\frac{di(t)}{dt}+\frac{dr(t)}{dt}=0\\
\Rightarrow\ &\frac{\Delta s}{\Delta t}+\frac{\Delta i}{\Delta t}+\frac{\Delta r}{\Delta t}\approx 0
\end{aligned}$$

We can solve this last equation to get Δi approximately in terms of $\Delta s, \Delta r$ and Δt, and then substitue for Δs and Δr from the forgoing analysis to get

$$\boxed{\Delta i \approx (b \cdot s(t)i(t) - k \cdot i(t))\Delta t}$$

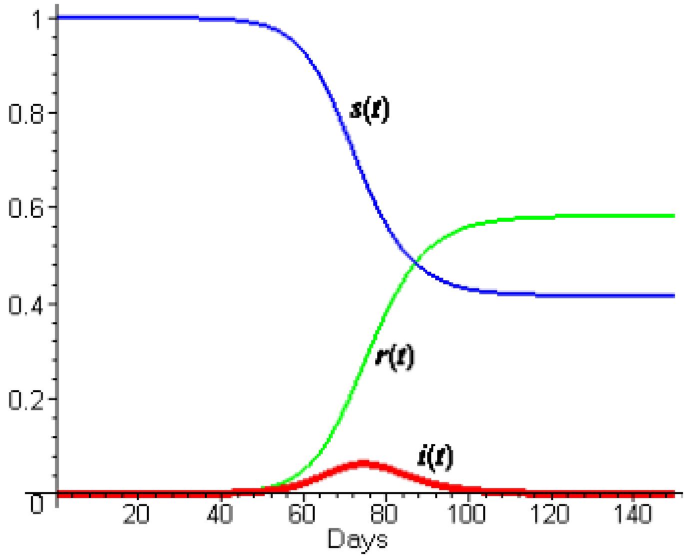
To implement the model, we need to give each differential equation an initial condition. Let's take, for instance, the flue epidemic in New York City in the late 1960's. Hardly anyone was immune at the beginning of the epidemic, so almost everyone was susceptible. We will assume that there was a trace level of infection in the population, say, 10 people. Thus, our initial values for the population variables are

$$\boxed{S(0)\approx 7,900,000}\boxed{I(0)\approx 10}\boxed{R(0)=0}$$

Normalizing (non-dimensionalizing) these we have

$$\boxed{s(0)=1}\boxed{i(0)\approx 1.27e-6}\boxed{r(0)=0}$$

We don't know values for the parameters b (transmission rate) and k (recovery rate) yet, but we can estimate them, and then adjust them as necessary to fit the excess death data. We have already estimated the average period of infectiousness at three days, so that would suggest $k = 1/3$. If we guess that each infected would make a possibly infecting contact every two days, then b would be $1/2$. This is just a guess. The following plot shows the solution curves for these choices of b and k.



1. Write a c++ program that meets the specifications below for NYC flu.

```
// Pseudocode to produce data for the plot
// precondition: initial values s(0), i(0), r(0) and
//                  parameters b and k.
// postcondition: Three vector<double> objects: susceptibles, infectious
//             and recovered that are produce by the iterative processes
//
//                  s(n) = s(n-1) + change in s
//                  i(n) = i(n-1) + change in i
//                  r(n) = r(n-1) + change in r
//
//             following the model described above.
```