Final Project: Compartmental Model Simulation Project

STAT 244NF: Infectious Disease Modeling

General Description

This project is a continuation of the work that you did in the Epidemiological Characteristics project. If you worked with a partner, you should continue working with that person. If you did not work with a partner and would like to, some students have overlapping topics - let me know if you would like me to match you with someone working on the same topic.

Primary Aims

Simulation Aims

The primary aim of this project is to set up and carry out compartmental model simulations on the infection/disease you researched in your first project. This should be done in the context of a single population, which likely will be one of the populations you identified in your first project (this is relevant for things like population size, demographics, vaccination, time period, etc.). To the extent it is possible with what we have learned in this class, the model should incorporate information on all the parameters you researched your first project (e.g., reproductive number, infectious period, etc.) to generate predictions of disease outbreaks for the infection/disease you are studying. I expect multiple simulations here, as detailed below.

For all simulations, you will need to implement your simulation via one or more functions, and provide all code and resulting output in a well-organized, knitted R Markdown file. **Do not include extraneous output like data frames, etc. Your document should only include relevant text, code, figures/tables. Check with me if you are unsure about your finished product.**

- **Deterministic simulation**: You will conduct a deterministic compartmental model simulation. In your previous project, you should have identified ranges associated the various parameters of interest. For this simulation, you should use the average value for each parameter. As part of your final submission, you should include a well-organized plot of your simulation results.

- **Stochastic simulation**: You will conduct a stochastic compartmental model simulation using random variables distributed according to Binomial distributions, as we have done in class. In your previous project, you should have identified ranges associated the various parameters of interest. For this simulation, you should use the average value for each parameter. You should carry out at least 500 simulations (with different seeds) with the same initial conditions. As part of your final submission, information on results from all 500 simulations should be presented in a well-organized plot. (We will go over how to do this in class.)

- **Stochastic simulation incorporating parameter uncertainty** (Two options, read on for details): You will conduct a stochastic compartmental model simulation using random variables distributed according to Binomial distributions, as we have done in class. There are two options for how to complete this portion, with suggestions on which you should choose, depending on how experienced you are with statistics and programming in R.

- *Option 1* (appropriate for those with less experience): In addition to the simulations you carried out in the Stochastic simulation above, also carry out simulations for the minimum and maximum parameter values in your ranges (this will result in two more simulations). For each combination, carry out at least 500 simulations (with different seeds) with the same initial conditions. As part of your final submission, information on results from all 500 simulations should be presented in well-organized plots (one for each set of parameter values). This should give you two plots in addition to the one made for the Stochastic simulation above.

- *Option 2* (appropriate for those with more experience, particularly if you have taken probability or 300-level statistics): In our stochastic models, the individuals moving between compartments are represented by random variables, and at each simulation we draw outcomes from their probability distributions. These probability distributions have the same transition probabilities each time (the average from the Stochastic simulation, for example). However, we can also treat these parameters as random variables, and assign probability distributions to them that reflect the range of possible parameter values. This forms a hierarchical model. *For example*, in an SEIR model for a non-sexually transmitted infection,

$$E_t^{(SE)} \sim Binomial\left(S_{t-1}, \lambda_t = 1 - \exp\left\{-\frac{R_0}{D} \times \frac{I_{t-1}}{N}\right\}\right)$$

but there may be some uncertainty about R_0 and D, so we can assign probability distributions to these, too, like:

$$R_0 \sim Uniform(\min(R_0), \max(R_0))$$

where the parameters for the Uniform distribution are the minimum and maximum values you found for the reproductive number in your research. There are other distribution possibilities, too, which are are welcome to look into. In your function, prior to running your for loop, you will need to make random draws from the probability distributions for *all* the parameters and then save them for use in the for loop. This will introduce more variability into your simulations. If you plan to pursue this option, we should talk about your distributional choices well in advance.

Data Comparison Aim

While most of this project is focused on simulation, it is nice to have some data with which to compare simulations, and as statisticians, it is always important to be thinking about data as well as models and methods. In your first project, you were asked to find a data set related to the infection/disease you chose to investigate. (If you didn't find one for that, you need to find one now - CDC, WHO, and other government health sites are usually good options.) The data you are able to find will likely be information on infections, like the data we were looking at in the beginning of this class. Usually we do not have information on the whole infection process, which we are modeling through the compartmental model. Once you have your data set, you should plot it versus time and perform an informal comparison with your simulation results. For instance, are the cases increases at similar rates in the real data and in the simulation? If not, why do you think this might be the case? (Were the data collected in an endemic population where not everyone was susceptible? What are other factors that might influence this, like high vaccination rates?)

Final Project Submission

Your final project submission should be compiled in a well-organized R Markdown file that is fully self-contained. This document should include the following:

- A brief, 1-2 paragraph description of the infection/disease you are studying that includes information about the natural history of the disease (relevant for choosing your compartmental model) and information about the population in which you are specifically planning to simulate the process.

- The simulations discussed above, including all functions, code, and plots. You should also include a paragraph description at the beginning of each simulation section to describe what you are doing. Include the difference equations for the simulation. These should be specific to the compartmental model you are using and should reflect whether the simulation is deterministic or stochastic (you need a different set of difference equations for the deterministic versus the stochastic simulations. You should define all notation (any of the symbols and random variables) in words so that it is clear to someone not in the class what you mean, and also include information on each of the probability distributions for the random variables so that someone familiar with statistics but not familiar with this specific class could understand). Make sure this section is legible, includes informative and attractive figures, and adequately summarizes your simulation results. This will probably be 3-5 pages with figures, depending on how much you write and how many figures you include.

- A 1-2 page discussion of your simulation results, including what you observed and how well these results seem to reflect the actual disease process. Is there anything unusual in your results? What are the limitations of your simulations? For example, were you assuming everyone was susceptible when the disease is endemic in the population? These models assume homogeneous (uniform) mixing in the population - is this reasonable (probably not - think about why)? Did you incorporate age differences? If not, how would this impact your model? Did you include anything about vaccination? Is your transmission through a vector?

- References - these can be the same as those in your first project, plus any course notes (including this class) and other resources that you used, either for data, for code, or for the infectious disease background.

- A link to the data set you used.

In total, your report will probably be between 5-10 pages, including figures (but excluding references). If it is a lot longer or a lot shorter then we should talk about it.

Other Important Information

I will include a template for the project, which you will be able to pull from Github (just like we do the labs and homework).

There will be one formal project check-in for this project, in the form of a homework assignment.

This project is due no later than Monday, December 13 at 12 noon. Any extensions need to be approved in advance by your academic dean, so do not leave this project until the last minute!