EOSC 213

Final Project

University of British Columbia

02 / May / 2020

Aaron Arcadi (std. # 36236339)
Cynthia Lee (std. # 52104395)
Jordan Myers (std. # 71273403)
Graham Perry (std. # 82106584)

Attached in our project submission is our version of the disease propagation simulation for Canada. We took two approaches to completing this project, both stemming from one method. We began coding the simulation from the data provided in the Diffusion on Graph notebook and example code found on the Epimodel notebook. We calculated the laplacian matrix L using Matlab, as such that code does not appear in our documents.

The equations and variables found in the three provided simulation equations were implemented into the Epimodel code which allowed us to visualize the disease propagation on graphs. From this, we decided to try modeling the graphs to represent the data we found for each of the thirteen provinces in Canada while assuming the variables were constant in time. We inevitably found this process very time consuming and impractical for large amounts of data. We researched and found how to import the data from stats Canada using Pandas. However, the code we wrote was unable to run the importation, thus we calculated the death rate of each province and manually entered the data into a list of floats in an array. We then attempted to make mu, the death rate, time dependent by calculating the mu everyday starting from around March 22 (when the borders closed) until present, and feeding it through a for loop so that theta can receive an updated mu value everyday:

$$\text{mu} = \frac{number\ of\ deaths}{total\ number\ of\ infected}$$

After many attempts, we were still unable to code a for loop using plot functions to correctly generate a graph that we visioned. The logic behind our for loop was straight-forward, and it is unfortunate that the functions we thought would work, did not. Our code of attempting to solve for this loop is attached in the ipynb file named *Loop Code Attempt.*

We then turned our attention to the derivatives with torch notebook to try and incorporate parameter estimation which would allow us to develop a more unified simulation for the whole country. Despite adding noise to the data, the graphs followed the same paths. Further following the notebook, we attempted to experiment with “loss” to estimate values for our parameters. Because we were estimating seven variables, we turned to automatic differentiation rather than the guess and check method as shown in the gradient plot on the notebook. Using the automatic differentiation methods the code returned values for the seven parameters which would be time dependent. However due to errors in the code we could not find a way to make these variables useful enough for accurate approximations. We then turned back to the beginning of the project to re-evaluate our approach. We then attempted to rework our code to take in vectors rather than float values which would allow us to simulate the disease propagation for all of Canada in a unified manner.

Our first approach interpreted the S, E, and I values as single numbers. This way we were able to develop a separate epimodel for each province. We used the recorded data from stats Canada and copied the values of “numtoday” from the Excel sheet to a Word document. We then removed all formatting and added commas in between each value to achieve a list formatting. We repeated this process for all 13 provinces and territories and combined each list/ array into a matrix of 13 rows and 87 columns, with each row representing a province, and column representing a day. However, as it was impossible to account for flux between provinces using isolated provincial epimodels, it quickly became clear that a different method was needed.

The second approach interpreted the S,E,I as vectors of length 13 that could be multiplied by the flux matrix. Using this method allowed us to model all 13 provinces at once as well as the flux of the S,E,I populations between them.

For initial S,E,I populations, we arbitrarily set S = 0.99, E = 0.01 and I = 0 for each of the provinces. This gave us relatively believable and unique plots for the provinces that did not differ too far from reality. However, each of these plots occured over the same time period and peaked at essentially the same time. This did not match the data from Statistics Canada CSV file where at one given point in time, each province was in a different stage of the plot. To resolve this, we changed our initial parameters to the first available data in the CSV file. BC started with 1 positive case and Ontario started with 3. From there the SEI spread across to the rest of the provinces, resulting in plots that were more accurate to the timeline of each province’s outbreak.

Uploaded Files

The Automatic Differentiation file is our experimentation with similar code from the notebook. It takes the equations containing the unknowns and returns a tensor with corrected values. Located at the bottom of the sheet.

~AutomaticDifferentiationUsingPytorch

This file is a simple experiment with importing excel data using Pandas:

~LearningToUploadDataUsingPandas

13 Province Model

~NationwideEpimodel

The following file contains our attempts in generating a graph that continuously updates mu to the most accurate value:

~Loop Code Attempt

The following files each contain a graph that models the curves of infection rate, exposed rate, and susceptible rate for each province/territory; a graph that models noise; and values of the parameter with use of gradient (but were unable to be used in calculations):

~Alberta
~BritishColumbia
~Manitoba
~NewBrunswick
~NewfoundlandandLabrador
~NorthWestTerritories
~NovaScotia
~Nunavut
~Ontario
~PrinceEdwardIsland
~Quebec
~Saskatchewan
~YukonTerritory

NOTE

A few points explaining the graphs in each document describing the spread of the virus within a given province:

- The time step is 0.05 meaning that every 0.05 units on the horizontal axis represents a day in the real world
- Each graph starts on March 22 because that is the day that we could say with confidence that regular international traffic into Canada had ceased and the only movement inside the country was domestic travel
- This means that the 0.5 marker on the horizontal axis represents 10 days of time corresponding with the first of April
- The peak of each infected graph represents the most cases recorded on a given day for a given province or in situations where the recorded cases had not yet peaked, the predicted peak
- The vertical axis of the graph represents the total population of the given province as a percentage
- For instance 0.01 on the vertical axis represents one percent of the overall population of the given province