MATH 371 - Forecasting Covid-19 Cases in the US

Jonas Smith

Monday 4th May, 2020

We have two goals for this project, firstly given a data set of new cases for the Covid-19 virus and forecast what the numbers will approximately resemble, secondly we will use a very basic decision tree and look at what would happen if we were to "open up" states and local businesses. We will be using the Center of Disease Control and Prevention's or cdc's numbers on new day to day Covid-19 cases in the United States. And also the Journal of Travel Medicine for determining the reproduction rate for the corona-virus which I use for the decision tree portion of the project.

Forecasts

To get started we need a data source, we will be using the cdc's website and look at the cases for the United states from 02/25/2020 to 05/01/2020. With this information we will build the casual linear regression table, given below, that will help us find our casual regression formula $Y = a + bX$, which we find on the next page.

Day (x)	Daily Cases (y)	x^2	y^2	xy
1	0	1	0	0
2	1	4	1	2
3	0	9	0	0
4	8	16	64	32
5	6	25	35	30
...	...	...	...	...
80	11219	6400	125865961	897520
81	16082	6561	258630724	1302642
82	15618	6724	243921924	1280676
83	13968	6889	195105024	1159344
84	30740	7056	944947600	2582160
85	76279	7225	5818485841	6483715
sum 3486	1340639	194054	35698560697	69927566

We now use values from the table to obtain,

$$a = \frac{\sum x^2 \sum y - \sum x \sum xy}{n \sum x^2 - (\sum)^2} = \frac{(194054)(1340639) - (3486)(69927566)}{(85)(194054) - (3486)^2} = 4144.582721128416$$

and,

$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum)^2} = \frac{(85)(69927566) - (3486)(1340639)}{(85)(194054) - (3486)^2} = 285.8974854120314$$

and it follows that,

$$Y = a + bX = 4144.582721128416 + 285.8974854120314(X)$$

Where X is the number associated with the next day. When we now try the next day we can see that it is forecasted to have approximately 28,160 new cases on day 86. We can automate this process very easily and forecast as many days as we want into the future, with lower and lower accuracy the further we get out. Shown below is the Forecasting function used for the remainder of the project

```
def getForecast(Stats, count):
    # init variables
    x_sum = y_sum = x_sqr = y_sqr = x_y = 0

    # Does all summations
    for stats in Stats:
        x_sum += stats.index
        y_sum += stats.newCases
        x_sqr += stats.index * stats.index
        y_sqr += stats.newCases * stats.newCases
        x_y += stats.index * stats.newCases

    # finds a & b
    a = ((x_sqr) * (y_sum) - (x_sum)*(x_y)) / ((count - 1)*(x_sqr) -
        (x_sum)*(x_sum))
    b = ((count - 1) * (x_y) - (x_sum)*(y_sum)) / ((count - 1)*(x_sqr) -
        (x_sum)*(x_sum))

    return(a + ( b * count ))
```

Opening back up

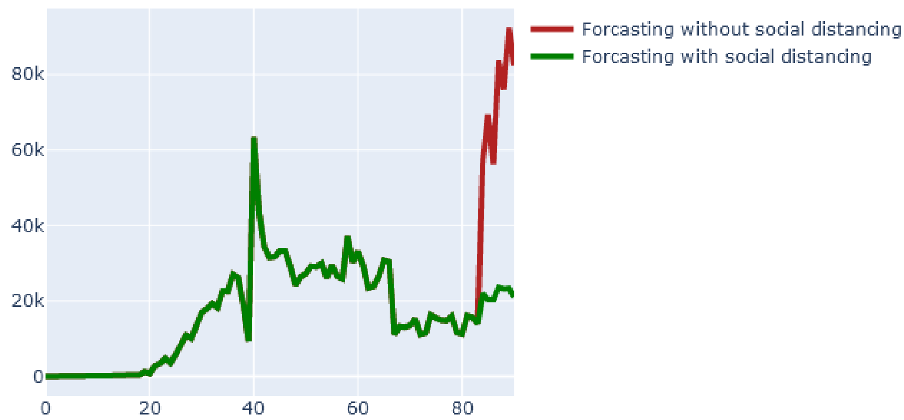
Using the method above, we can see that in a weeks time (05/10/2020) with the current system of social distancing the US will have nearly 1,562,841 cases of Covid-19, However what would happen if we opened all states in the US, meaning we allowed businesses and citizens to act pre-pandemic. Well we will first need to decide on a few variables.

Lets talk about reproduction rates of a virus, the reproduction rate of the seasonal flu is 1.28 which means that if we have 100 citizens that have the flu, we can expect them to spread it to 128 people. Now we need to look at the Journal of Travel Medicine to determine the reproduction rate of Covid-19, what we find is that Covid has a range of **1.4** to **6.49**!, depending on which countries you look at. We will use the median 2.79.

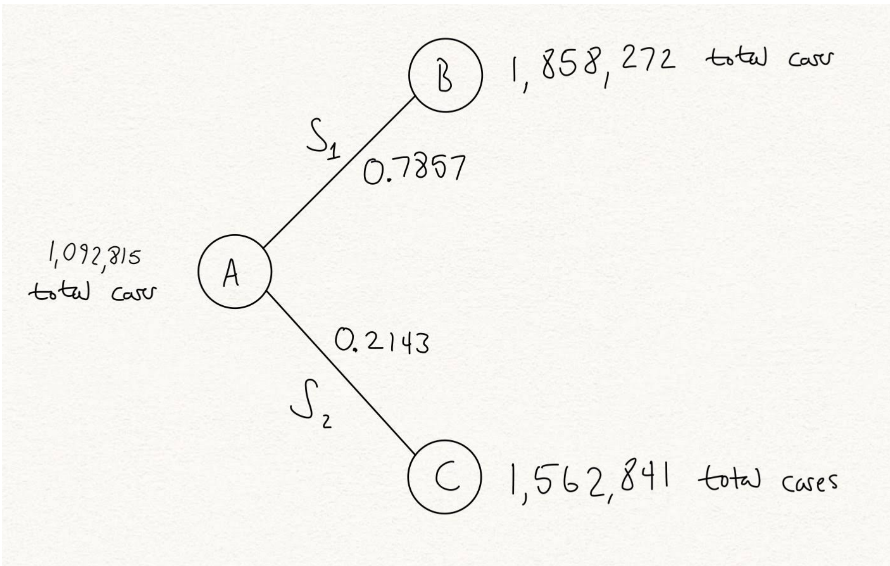
We now need to determine a type of probability of danger for opening or not opening the States. We have to take some liberties with some rough estimations as this information is not publicly available. What we are going to do is add a very basic probability of getting the virus, we are going to assume that people that are working and going about their weekly lives like normal are spending 55 hours out of the house a week, and people in quarantine are going to be spending roughly 15 hours out of there houses a week, so respectively the non quarantine people have a 78.57% multiplier chance, and people in quarantine have a 21.43% multiplier chance.

The below code iterates through the next seven days of the week forecasting what the next day will look like in regards to cases and adding those cases to the list. The prob variable is associated with the danger variable we created above giving us a value of 0.2143 for not opening states and a 0.7857 for opening them again. The $random.uniform(2.4, 2.9)$ is the value we use to add a very slight randomization to our outcome as humans are variable.

```
for x in range(7):
    forValue = getForecast(Stats, count)
    forValue = forValue * random.uniform(2.4, 2.9) * prob
    print("{0} : {1}".format(count, forValue))
    Stats.append(Stat(count, "new date", int(forValue)))
    count+=1
```



With this given information we can determine that opening the States back up after one week will have nearly 1,858,272 cases in the US, to look back out our results from earlier without coming out of quarantine the results were 1,562,841 cases! That means that roughly 295,431 new cases would occur over a week. Now we can set up a very simple decision tree, where we are trying to minimize the total number of cases.



And without too much input needed we can see that based on our current situation the best way to keep cases at a minimum is to continue to keep citizens social distancing until a vaccine has been released to the public.

References

[1] CDC
https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html

[2] Journal of Travel Medicine,
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7074654/

[3] Psychology Today
https://www.psychologytoday.com/us/blog/laugh-cry-live/202004/the-shocking-numbers-beh

[4] plotly
https://pythonbasics.org/plotly/