COVID-19 Pandemic Predictions

OBJECTIVE AND SCOPE

What will be the global impact of the novel coronavirus (COVID-19)? Answering this question requires accurate forecasting the spread of confirmed cases as well as analysis of the number of deaths and recoveries. Forecasting, however, requires ample historical data. At the same time, no prediction is certain as the future rarely repeats itself in the same way as the past. Moreover, forecasts are influenced by the reliability of the data, vested interest, and what variables are being predicted. Also, psychological factors play a significant role in how people perceive and react to the danger from the disease and the fear that it may affect them personally. This project introduces an objective approach to predicting the continuation of COVID-19 using a simple, but powerful method to do so. Assuming that the data used is reliable and that the future will continue to follow the past pattern of the disease, these forecasts suggest a continuing increase in the confirmed COVID-19 cases with sizable associated uncertainty. The risks are far from symmetric as underestimating its spread like a pandemic and not doing enough to contain it is much more severe than overspending and being over careful when it will not be needed.

PROCESS DISCRIPTION

The accuracy of traditional forecasting largely depends on the availability of data to base its predictions and estimates of uncertainly. In outbreaks of epidemics there is no data at all in the beginning and then limited as time passes, making predictions widely uncertain.

Besides, there are concerns that the data may not be reliable, as was the case of bird flu and SARS when the number of affected people and deaths were misreported to hide the extent of

the epidemic. Similarly, in the case of COVID-19, the reporting did not reflect the correct numbers as well when on the February 13 a new category of "clinically diagnosed" was added to "lab-confirmed" ones. Such problems decrease forecasting accuracy and increase uncertainty, making the drawing of definite conclusions more difficult.

Related to forecasting accuracy and uncertainty, there is a more severe problem that has to do the perception of epidemics and pandemics. Politicians are concerned with regards to the measures to be taken while the general population fears about the impact on the epidemic on their health/lives. Furthermore, the pharmaceutical firms are working on vaccinations for the new virus with considerable commercial interest.

The used data focus on the cumulative daily figures aggregated globally of the three main variables of interest: confirmed cases, deaths and recoveries. These were retrieved by the CSV files and from websites using web scraping script. The data refer to daily cumulative cases from January 22, 2020. I will use a model from a marketing paper by Emmanuelle le Nagard and Alexandre Steyer, that attempts to reflect the social structure of a diffusion process. Their application was the diffusion of innovations, not epidemics. However, there are commonalities in both domains, as the number of contacts each infected person/ innovation adopter has seems relevant. It also has the added benefit to allow fitting parameters to the beginning of a time series.

RESOURCES AND LIMITATIONS

I used python as a base programming language and anaconda environment for developing, anaconda environment supports various packages and notebooks environment, one of the notebook environments is jupyter notebook. The jupyter notebook is an open source web application that allows us to create and share documents that contain live code, equations,

visualizations and narrative text. It is used for data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning and much more.

CONCLUSION

I have investigated the problem of COVID-19 spread on global scale and also in various countries and their states/province. A mathematical model has been established, which follows the actual data trend of COVID-19 spread.