COVID-19 Forecasting

Predicting Mortality and Morbidity Rates

Elias Ghantous[1] Efran Himel[1] Derrick Liu[1]
Tiffany Quang[1] and Aubrey Wiederin[1]

[1]Department of Mathematics and Statistics
University of Massachusetts Amherst

April 22, 2020

Outline

Outline

Basic SIR Model

The Basis of Comparison for Our SIR Model

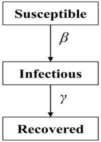
- A SIR model is a simple compartmental model that helps us predict the dynamics of an infectious disease in a closed population (i.e. people not traveling outside of stationed area).
- The name, SIR, is a representation of the three compartments the model is split into: **S**usceptible, **I**nfectious, **R**ecovered.
- It describes how the number of people in each compartment can change over time by using ordinary differential equations.
- Claim: We believe a SIR model with geographic dependant components can outperform the standard SIR Models. We believe these geographic components can be optimized with neural nets to produce the best parameters for the SIR Model.

Outline

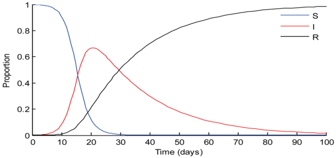
Basic SIR Model

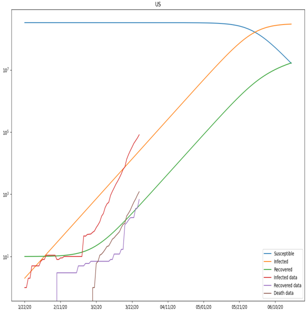
The Basis of Comparison for Our SIR Model

- Here is an overview of what an SIR Model looks likes as well as a generic SIR for the United States



$$\frac{dS}{dt} = -\frac{\beta IS}{N},$$
$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I,$$
$$\frac{dR}{dt} = \gamma I,$$





Basic SIR Model

The Basis of Comparison for Our SIR Model

- For our baseline, we used the model available at https://github.com/Lewuathe/COVID19-SIR, with some minor modifications.
- We used the model to predict the next 150 days of our country from the start date of January 22, 2020. Our initial parameters included:

$$S = 328200000 \quad I = 2 \quad R = 0$$

- The model predicted a beta (β), the infection rate of the disease, of 0.1496 and a gamma (γ), the recovery rate, of 0.002533.
- Together, both our β and γ produced an $R_0 = 59.078$, which is the basic reproduction rate.
- Our R_0 is about ten times higher than it should be. This is due to the model under-estimating the γ value due to a sparsity of recovery rate data.

Outline

Regression Models

Finding Useful Predictor Variables

- A SQL database on the various characteristics of each state and the impacts of COVID was created from data provided by various reliable sources such as the CDC and U.S. Census Bureau.
- The initial predictor variables in our data set were: population density, land area, state GDP, number of airports, number of automobiles, number of buses, number of hospitals, number of active doctors, average income, and proportion of people uninsured. Variables were added and removed during during the modeling process.

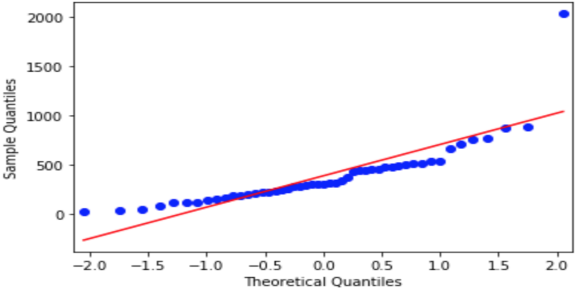
Outline

Regression Models

Finding Useful Predictor Variables

- After the "no multicollinearity" assumption was checked and collinear variables were removed, the two multiple linear regression models for predicting number of deaths due to COVID-19 and number of COVID-19 cases showed statistically significant coefficients for population density and number of airports.
- As expected, the coefficients for population density and number of airports were positive (21.4757 and 12.5072) for the model predicting COVID-19 cases and were positive (0.7089 and 0.3736) for the model predicting deaths.

- The only predictor variable that satisfied the linearity and normality assumptions and was not collinear with any other predictor variable was "number of airports".



Regression Models

Finding Useful Predictor Variables

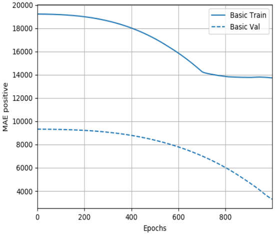
- Multiple linear regression is perhaps not very useful to run on our data set for Coronavirus forecasting purposes.
- Perhaps logistic regression could work better, with the outcome variable being survival or death ($1 =$ survival, death $= 0$). We couldn't find a data set that was conducive to logistic regression.
- The number of airports nonetheless could be an important predictor variable for Coronavirus forecasting.

Outline

Neural Nets

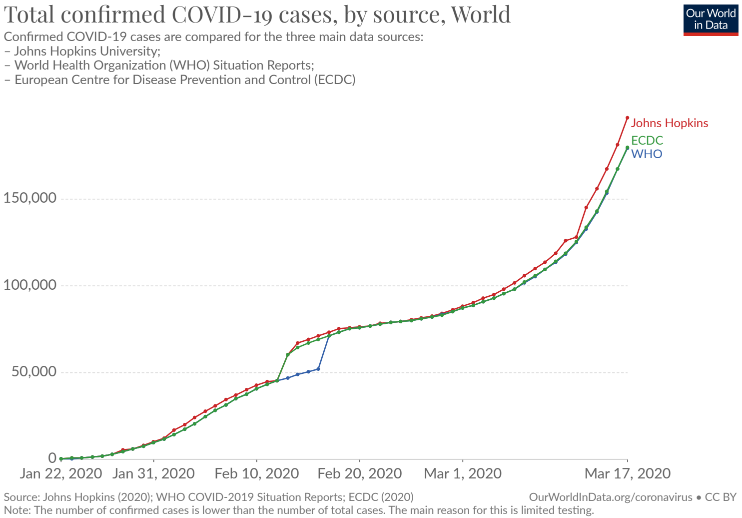
Using Data to Predict Outcomes

- Trained neural nets to predict various outcomes such as the number of positive Coronavirus cases and the number of deaths.
- Initial goal was to use predicted values to approximate the β and γ variables for SIR Model but the neural net's predicted values had high residuals.
- Various neural net configurations were implemented to reduce the high errors but were unsuccessful. We were unable to reintroduce the predictions back into the SIR Model.



Outline

Cases over time

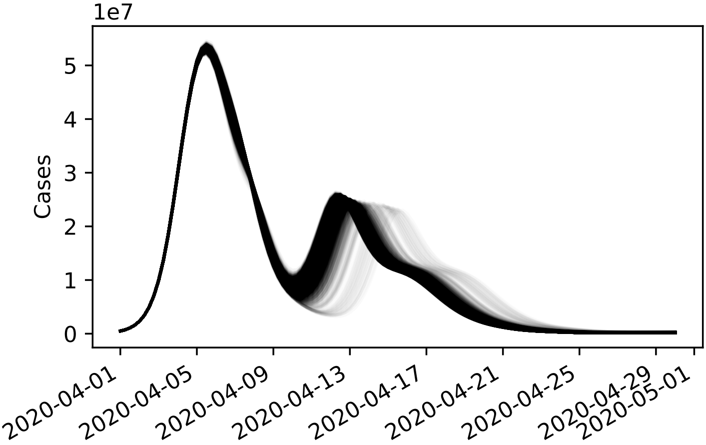


Our SIR Model

Improvements Relative to Basic SIR Model

- The basic SIR model does not account for geography, and so predicts smooth, sinusoidal growth.
- In contrast, as the actual virus spreads to new regions (cities, states, countries, etc.), it causes sinusoidal "jumps" in the infection curve as a new population is infected.
- We modeled this as a set of coupled SIR models on a graph, with the coupling being transport of infected and uninfected between regions.
- In addition, our model allows for independent parameters at each region, allowing for differences in population density and response to the epidemic.

Example simulation trajectories (placeholder parameters)



Outline

Our SIR Model

Other possible improvements

- More complicated models than SIR are possible, such as with including an incubation period (SEIR).
- Given the appropriate data, it is also simple for us to model the effects that social distancing and quarantines have on the spread and on different segments of the population.
- Modeling of vaccination is another change to the basic model that would be very useful.
- Though very much outside the scope of this project, a more in-depth model could come from building an analytic estimate for spread in each region based on an Ornstein-Uhlenbeck process, to account for the geometric limitations of the spread.

Potential Issues

- Both the regression model and the SIR models do not take into account many of the factors that play a role in the spreading of COVID-19, such as effects of social distancing.
- Multiple linear regression does not work well for coronavirus forecasting. Logistic regression could potentially work well if there is a data set conducive to this type of regression.
- The Basic SIR Model is useful but does not provide a believable R_0 value.
- Data collection was done through various sources and would need to be cross-checked or confirmed to be accurate data.

Summary and Findings

- Through multiple linear regression we found that population density and number of airports are significant in predicting COVID-19 deaths.
- Neural nets on our database could help us predict outcomes for each day so that we gain the rates from those predictions to implement into the SIR model.
- Unlike the basic SIR model, our model shows "jumps" in the infection curve as a new population is infected.
- Outlook
  - The main factor that limited our final outcome for our SIR model was the inaccuracy of our neural nets and thus they will needs reconfiguration. This limitation prevented us from fully exploring our SIR model.
  - We will proceed in estimating the infection and recovery rates for the model with our current neural nets and compare it with real world outcomes. Also, we will create secondary model that uses actual data provided by state databases to create another SIR model based on real world values instead of predicted ones.

For Further Reading I

K. Sasaki.
COVID-19 dynamics with SIR model.
The First Cry of Atom, 2020.

J. Michaud, J. Kates, L. Levitt.
COVID-19 Models: Can They Tell Us What We Want to Know?
Global Health Policy, 2020.

R.T.Q. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud.
Neural Ordinary Differential Equations.