Title: Epidemiologic Study of Prince George’s County COVID-19 Data over the period February 1, 2020 – September 30, 2020

Author: Robert Bull and Alexander Klein

University of Maryland

Date: March 11, 2023

Venue: MATH 420 Mid-semester Project Presentation

1. **Introduction**

Our project focuses on the Prince George’s County infection and death rates during the COVID-19 pandemic. A predominantly democratic county, it faced relatively typical infection rates for the virus, but also had an interestingly high mortality rate (1).

We implemented the SIR and SEIR models described in class; we optimized these models to more effectively model and subsequently predict the infection rates and fatalities of COVID within the county.

1. **Background/Methods**

The SIR (Susceptible, Infected, Recovered) Model is a deterministic model that is intended to simulate states of an individual who has or has not been exposed to the disease we wish to model. It uses the three states listed above and presents the transitions as the following system of differential equations that we solved using the Euler method for initial value problems:

The SEIR (Susceptible, Exposed, Infected, Removed) model provides a similar simulation as the aforementioned SIR model, with the inclusion of a fourth state (Exposed). The exposed state aims to capture the group of individuals who have been exposed to the disease, but are not yet infectious. It acts as an intermediate between the Susceptible and Infected states. The model is given by the following system of differential equations:

We used these models primarily because they are particularly inclined to model this exact situation: they are relatively simplistic and allow for the information gained from them to be optimized easily. By using the Euler method on the equations, we can quickly redo simulations when we need to change coefficients to better fit the measured data. We use the assumptions that S(0) = N = 909327, E(0) = I(0) = 31, and R(0) = 0.

We parameterized the data using four general parameters: alpha (α), beta (β), gamma (γ), and rho (ρ). In the case of SEIR, a fifth parameter, delta (δ) was also used. Alpha represents the infectious period of the disease, or how long it stays transmissible after a patient becomes contagious. The infectious period is equal to 1/ α. Beta represents the number of daily contacts an infected individual has with susceptible people. In our data, we used R0 to represent β / α, or the number of individuals one infected person interacts with over their infectious period. δ captures the incubation period of the virus, with the incubation period equaling 1/δ. γ represents the fatality rate after infection. That is, what percentage of people died after being infected instead of recovering after their infectious period. Finally, ρ determined the undercounting factor of infections. Often, people infected with COVID would not report the illness to authorities or counting officials. Thus, they would be data points we would not know about, and we would have to scale the modelled infections down to represent that. An Euler scheme was run for every (α,β) pair in SIR, and for every (α,β,δ) pair in SEIR. Optimal ρ and γ values were computed for each of these pairs, by optimizing the following p-norms:

, where and p varied by model, as described later. We then calculated the objective function J via:

Models were trained for and , each of which were trained on three p-values: p = {1, 2, inf}. For each of these 6 parameter combinations, one model was trained on the first 40-days of infection, and one model was trained on the first 120-days of infection. These sets of models were implemented for two datasets: a PG-County COVID dataset and a validation dataset. Thus, there were 24 models trained (12 40-day and 12 120-day), for each of the SIR and SEIR algorithms. The 40-day models were qualitatively measured against the true infection rates and the 120-day models to determine how effectively they acted as predictors.

(Presents Models and Algorithms designed for this problem.

Explanation of why the models/algorithms are the way they are.

Explain how these models are parametrized. What do these parameters mean.)

1. **Results**

For the 120-day SIR models trained on PG County data, we saw relatively uniform optimal parameters across all combinations.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| p-value |  | α | ß | γ | ρ |  | J |  |
| 1 | (0,1) | 0.39 | 0.546 | 0.0013891 | 0.13708 | 0.71 | 4661 | 0.462 |
| 1 | (1,1) | 0.38 | 0.513 | 0.001597 | 0.14437 | 0.74 | 132302 | 0.526 |
| 2 | (0,1) | 0.4 | 0.56 | 0.0013557 | 0.14199 | 0.71 | 281747 | 0.462 |
| 2 | (1,1) | 0.4 | 0.54 | 0.0015736 | 0.16732 | 0.74 | 187335583 | 0.526 |
| Inf | (0,1) | 0.37 | 0.5365 | 0.0012436 | 0.13055 | 0.689 | 83 | 0.406 |
| Inf | (1,1) | 0.4 | 0.56 | 0.0013747 | 0.15037 | 0.71 | 2187 | 0.462 |

Table 1. Results from the SIR simulations trained on 120 days of the PG County dataset. Values of the objective function J are rounded to the nearest whole number to conserve space.

|  |  |  |  |
| --- | --- | --- | --- |
| p-value |  | γ | ρ |
| 1 | (0,1) | 0.00063745 | 0.070779 |
| 1 | (1,1) | 0.0010196 | 0.099565 |
| 2 | (0,1) | 0.00064007 | 0.077424 |
| 2 | (1,1) | 0.00089627 | 0.090144 |
| Inf | (0,1) | 0.1965 | 1 |
| Inf | (1,1) | 0.00088873 | 0.10168 |

Table 2. Optimal gamma and rho parameters from SIR simulations trained on 40 days of the PG County dataset.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| p-value |  | α | ß | δ | γ | ρ |  | J |  |
| 1 | (0,1) | 0.39 | 0.799 | 0.37 | 0.000878 | 0.08104 | 2.05 | 5266 | 0.19 |
| 1 | (1,1) | 0.39 | 0.7215 | 0.39 | 0.000878 | 0.08104 | 1.85 | 146013 | 0.249 |
| 2 | (0,1) | 0.39 | 0.819 | 0.37 | 0.0008403 | 0.07955 | 2.1 | 364969 | 0.178 |
| 2 | (1,1) | 0.39 | 0.7605 | 0.39 | 0.0008403 | 0.07955 | 1.95 | 230767916 | 0.217 |
| Inf | (0,1) | 0.37 | 0.814 | 0.37 | 0.0007976 | 0.08013 | 2.2 | 94 | 0.156 |
| Inf | (1,1) | 0.39 | 0.819 | 0.37 | 0.0007976 | 0.08013 | 2.1 | 2365 | 0.178 |

Table 3. Results from the SEIR simulations trained on 120 days of the PG County dataset. Values of the objective function J are rounded to the nearest whole number to conserve space.

|  |  |  |  |
| --- | --- | --- | --- |
| p-value |  | γ | ρ |
| 1 | (0,1) | 0.23999 | 0.95292 |
| 1 | (1,1) | 0.00062517 | 0.061355 |
| 2 | (0,1) | 0.27636 | 1 |
| 2 | (1,1) | 0.0006608 | 0.062734 |
| Inf | (0,1) | 0.40318 | 1 |
| Inf | (1,1) | 0.00068269 | 0.076155 |

Table 4. Optimal gamma and rho parameters from SEIR simulations trained on 40 days of the PG County dataset.

A graph of a graph

Description automatically generatedA graph with green and black lines

Description automatically generated

Figure 1. SIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = 1.

A graph of a graph

Description automatically generatedA graph with green and black lines

Description automatically generated

Figure 2. SIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = 1.

A graph of a graph

Description automatically generatedA graph with green and black lines

Description automatically generated

Figure 3. SIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = 2.

A graph of a graph

Description automatically generatedA graph with green and black lines

Description automatically generated

Figure 4. SIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = 2.

A graph on a white background

Description automatically generatedA graph on a white background

Description automatically generated

Figure 5. SIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = inf.

A graph of a graph

Description automatically generatedA graph with green and black lines

Description automatically generated

Figure 6. SIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = inf.

A graph of a normal distribution

Description automatically generated A graph of a function

Description automatically generated

Figure 7. SEIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SEIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = 1.

A graph of a normal distribution

Description automatically generated A graph with a green line

Description automatically generated

Figure 8. SEIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SEIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = 1.

A graph of a graph

Description automatically generated A graph of a function

Description automatically generated

Figure 9. SEIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SEIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = 2.

A graph of a graph

Description automatically generated A graph with a green line

Description automatically generated

Figure 10. SEIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SEIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = 2.

A graph of a normal distribution

Description automatically generated A graph of a function

Description automatically generated

Figure 11. SEIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SEIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = inf.

A graph of a normal distribution

Description automatically generated A graph of a normal distribution

Description automatically generated

Figure 12. SEIR Infection rates (Left), plotting 120-day (solid red), 40-day (dashed red), and measured (blue). SEIR Cumulative deaths (Right), plotting 120-day (solid black), 40-day (dashed black), and measured (green). Parameters were and p = inf.

A graph of a mountain

Description automatically generated

Figure 13. SEIR 120-day (solid red), 60-day (dashed red), and measured (blue). SEIR 120-day (solid purple), 60-day (dashed purple), and measured (green). Parameters were and p = 1.

1. **Discussion**

Examination of the measured infection data shows a local peak at t=55, corresponding to March 17th – near the time that quarantine lockdown procedures were first initiated. Infection rates plateaued and began to fall after around 10 days, dropping to a local minimum at around 110 days.

Looking first at the SIR models, we can observe that from a qualitative perspective, the 120-day Euler scheme was a decent approximation for both and in all cases. While the 40-day wasn’t a very good approximation in any of the models, the 40-day was a decent predictor in all cases except for , p = inf (Figure 5). In this case, we can observe that the optimal value for ρ was 1. This would imply that there was no undercounting in the measured infection data. Given that this is both impractical, and contradicts the undercounting factors of our other models, we can assume this is in error. Interestingly, the value of ρ is irrelevant in this case, since , thus ρ zeros out in the calculation of the objective function J. However, we also have an abnormally high γ value of 0.1965. This implies that covid is lethal to nearly 20% of those who become infected. The true mortality rate of COVID-19 has been estimated to lie between 3-6%, which is far below the amount determined from this model (2). Across all of our 120-day models, we observe an average infectious period of slightly over 2 days, and around 0.5 close contacts per day. We also observe 1.3-1.6% mortality, and a 10-20% rate of undercounting infections. While the mortality rate is small, it is more accurate than the mortality rates determined from the 40-day models, which are generally below 1%. We also discovered that the herd immunity numbers were much larger than expected. Around 40% of the population being completely uninfected does not seem to mirror the real-world virus very well.

The 40-day models seem to predict a more rapid drop-off of the disease than what is experienced. This could imply that the chosen parameters/model failed to capture some element of the coronavirus. For instance, there is no parameter to determine how infectious the disease is, in other words, how likely it is for a “close contact” (ß) to convert a susceptible individual into an infected individual. In Figures 1-6, the measured data fails to reach a peak as high as the 120-day model predicted, even though the model was trained over the whole range. This is most likely because the SIR model failed to account for the quarantine protocols that came into effect around t = 50 (just before the 120-day model expects a peak of infection rates).

Looking next at the SEIR model, we can see that the 40-day models failed to accurately predict the measured trends, as did all of the 40-day models for . Like the 40-day (0,1) model from SIR, several of the optimal γ values in SEIR were > 0.200, well above the measured mortality rate of COVID. Our hypothesis for the inaccuracies is the difference in trends between the first 40 days of infection rates and the entire 120-day period of study. The first 40 days were essentially monotonically increasing, while the remaining 80 days had plateaus and periods of decrease. This could cause the model to overfit the first 40 days, leading to an expectation of further increase and a misestimation of peak location. In particular, the lockdown procedures were not accounted for in the model, as mentioned in the discussion on SIR, which could further lead to this consistent overcounting. Considering that the models were the only ones with such degree of error, we assume the inclusion of the norm term in the objective function is necessary to properly model the data. Figure 13 shows that we would achieve much better predictive results by training a model on the first 60 days, rather than the first 40 – particularly in the case of estimating deaths via . This lets us include the beginning of the plateau period in our training data, which leads to a more realistic prediction of future infection rates.

The optimal α,,δ parameters in the SEIR model were all near the upper bounds of their respective ranges (0.4, 2.2,0.4). Since α and were large, and ß = α , it follows that ß was also near the upper bound if its range. This implies a short infectious period and incubation period (both around 2.2 days in length), and average daily close contacts of ~0.8. The experimental values are consistent with prior literature on the values of COVID-19, although may have been slightly larger if the parameter range were larger (3). Herd immunity in the SEIR model was significantly lower than SIR, hovering around 20% compared to SIR’s 40%. Since the SEIR values were much closer to literature values than SIR, it makes sense SEIR herd immunity would be more accurate as well (since herd immunity is a function of ).

Future work on the topic would include exploring wider parameter ranges (particularly increasing the upper bounds, as the optimal values consistently bumped against the bounds we had in place). Increasing the upper bound of could allow us to get a more accurate prediction of herd immunity, as stated prior. Optimizing models on various values could also help isolate the problem we ran into with = 40 on SEIR. Finding the exact point at which the model begins to behave normally should give insight into the point at which the rapid increase of measured infection rates slowed down (likely coinciding with quarantine protocols).

**References**

Cite any source you used for your report.

(1) PG Demographics: <https://elections.maryland.gov/pdf/vrar/2022_10.pdf>

(2) [SARS-CoV-2 pandemic: An overview - PMC (nih.gov)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7832554/)

(3) [COVID-19 Compared to Other Pandemic Diseases - PMC (nih.gov)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7426550/)