Dynamic Modeling of Reported Covid-19 Cases and Deaths with Continuously Varying Case Fatality and Transmission Rate Functions

Mingdong Lyu

Nov 7th, 2020
Introduction

Features of COVID-19

• a relatively long incubation period, with a mean time of 5 days
• transmissible while individuals are asymptomatic
• disease severity is widely variable, depending on age, comorbidities, baseline health and access to care

Prior research

• Zhao et al and Nishiura et al have used the SIR (symptomatic-infectious-recovered) model and SEIRD to estimate the number of undetected cases
• Some approaches also incorporate transportation information (such as human migration data and community mobility data) to analyze the impact of travel on disease transmission and thus the effect of travel restriction

Most studies using SEIRD or SIR assume the transmission rate and death rate to be constant over time.
Introduction

Features of COVID-19

• a relatively long incubation period, with a mean time of 5 days
• transmissible while individuals are asymptomatic
• disease severity is widely variable, depending on age, comorbidities, baseline health and access to care

Prior research

• Godio et al modify the new recovery rate to a sinusoidal function with six parameters and they adjust the transmission rate according to mobility trends.
• Piccolomini et al. compare two piecewise time-dependent infection rate functions and fit the infection rate function, incubation rate, and death rate for each uniformly divided time interval.

These approaches require introduction of many parameters to depict time dependency, thus risking overfitting and reducing model generality.
The Proposed Time Varying Model

SEIRD model

\[
\begin{align*}
\frac{\partial S(t)}{\partial t} &= -\beta(t) \cdot I(t) \cdot \frac{S(t)}{N} \\
\frac{\partial E(t)}{\partial t} &= \beta(t) \cdot I(t) \cdot \frac{S(t)}{N} - \sigma \cdot E(t) \\
\frac{\partial I(t)}{\partial t} &= \sigma \cdot E(t) - (1 - \alpha(t)) \cdot \gamma I(t) - \alpha(t) \cdot \rho \\
\frac{\partial R(t)}{\partial t} &= (1 - \alpha(t)) \cdot \gamma I(t) \\
\frac{\partial D(t)}{\partial t} &= \alpha(t) \cdot \rho \cdot I(t)
\end{align*}
\]

\( \beta(t) = \) transmission rate at time t
\( \sigma = \) transformation rate from exposed to infectious
\( \alpha(t) = \) likelihood of eventual death of a person who is infected at time t
\( \gamma = \) transformation rate from infectious to recovered, which is the reciprocal of the recovery time
\( \rho = \) transformation rate from infectious to death
The Proposed Time Varying Model

**Sigmoid function**

\[ S(x) = \frac{1}{1 + e^{k(x-a)}} \]

- \( k \) determines the slope of the function
- \( a \) determines the x value at the middle point

**Function for transmission rate and death rate**

\[ \beta(t) = \beta_{\text{end}} + \frac{\beta_{\text{start}} - \beta_{\text{end}}}{1 + e^{m \cdot (x-a)}} \]

\[ \alpha(t) = \alpha_{\text{end}} + \frac{\alpha_{\text{start}} - \alpha_{\text{end}}}{1 + e^{n \cdot (t-b)}} \]

- \( \beta_{\text{start}} \) is the starting reproduction number
- \( \beta_{\text{end}} \) is the ending reproduction number
- \( \alpha_{\text{start}} \) is the starting death rate, ranging from 0 to 1
- \( \alpha_{\text{end}} \) is the ending death rate, ranging from 0 to 1
- \( m, n, a, b \) are the shape parameters
The Proposed Time Varying Model

Parameter Estimation and Model Fitting

Data period: period from the day of first reported case in each state until 07/28/2020, across all 50 American states.

Data before the first reported cases: For each state of the USA, we chose a start date of 4 days prior to the date of the first confirmed case.

Parameters to be fitted: shape parameters $m, n, a, b$ and the starting/ending parameters $\beta_{start}, \beta_{end}, \alpha_{start}, \alpha_{end}$

Parameters Derived from Prior Research

<table>
<thead>
<tr>
<th>Resource</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>As mentioned in the CDC reports [17]–[19], the median incubation period is 4 days, ranging from 2 – 7.</td>
<td>$\sigma = 1/4$</td>
</tr>
<tr>
<td>Among 305 hospitalized patients and 10,647 recorded deaths, the median time of hospitalization was 8.5 days</td>
<td>$\gamma = 1/8.5$</td>
</tr>
<tr>
<td>the median interval from illness onset to death was 10 days (IQR = 6 - 15 days).</td>
<td>$\rho = 1/10.$</td>
</tr>
</tbody>
</table>
The Proposed Time Varying Model

Parameters Derived from Optimization

Model function: \( M(t; \beta_{\text{start}}, \beta_{\text{end}}, m, a, \alpha_{\text{start}}, \alpha_{\text{end}}, n, b) \): \( t \to \mathbb{R}^2 \)

\[
\begin{align*}
\min_P & \| M(t; P) - [\text{Cases}(t), \text{Deaths}(t)] \|_2^2 \\
\text{s.t.} & \quad 0.5 \leq \beta_{\text{start}} \leq 7.5 \\
& \quad 0.1 \leq \beta_{\text{end}} \leq 2.5 \\
& \quad 0 \leq \alpha_{\text{start}} \leq 1 \\
& \quad 0 \leq \alpha_{\text{end}} \leq 1 \\
& \quad 0.01 \leq m \leq 0.33 \\
& \quad 0.01 \leq n \leq 0.33 \\
& \quad 0 \leq a \leq 125 \\
& \quad 0 \leq b \leq 125
\end{align*}
\]
The Proposed Time Varying Model

Parameters Derived from Optimization

Adjusted model function: \( M(t; [\beta_{\text{start}}, \beta_{\text{end}}, m, a, \alpha_{\text{start}}, \alpha_{\text{end}}, n, b]): t \rightarrow R^2 \)

\[
\min_p \left\| \left( \hat{I}(t) + \hat{R}(t) + \hat{D}(t) - \text{Cases}(t) \right)^2 + w \left( \hat{D}(t) - \text{Deaths}(t) \right)^2 \right\|_2
\]

s.t.
- \( 0.5 \leq \beta_{\text{start}} \leq 7.5 \)
- \( 0.1 \leq \beta_{\text{end}} \leq 2.5 \)
- \( 0 \leq \alpha_{\text{start}} \leq 1 \)
- \( 0 \leq \alpha_{\text{end}} \leq 1 \)
- \( 0.01 \leq m \leq 0.33 \)
- \( 0.01 \leq n \leq 0.33 \)
- \( 0 \leq a \leq 125 \)
- \( 0 \leq b \leq 125 \)

the parameters set \( P = [\beta_{\text{start}}, \beta_{\text{end}}, m, a, \alpha_{\text{start}}, \alpha_{\text{end}}, n, b] \) is initialized as [0.75, 0.5, 0.25, 10, 0.4, 0.1, 0.25, 10]
The Proposed Time Varying Model

Data limitation

i. We recognize that reported data are not the same as actual cases and actual deaths, which are unknowable.

ii. Daily confirmed cases are influenced by widely varying testing rates and policies, which change over time. At the beginning of the epidemic, the limited test kits were restricted to those who suffer from severe symptoms and those who are in a higher risk of exposure.

iii. Death data are likely to be more accurate, but they too can suffer from reporting errors, due to how deaths are attributed to COVID-19 (or not), the timing of filing reports and the general accuracy of reporting.

For these reasons, our model represents estimation of reported data rather than the actual (but unknown) number of cases and number of deaths.

iv. Reporting has also shown a consistent day-of-week variation across many locations, with weekend data differing from weekday data.

To smooth out these effects, we model the moving 7-day average data instead of the daily reported data.
Results

Model Accuracy – Relative RMSE

\[ RRMSE = \left( \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N} \right)^{1/2} \]

The average and median RRMSEs for deaths are 1.61% and 1.33%; for cases the average and median values are 2.30% and 1.88%.
Results

Model Accuracy

New York case prediction
- Relative RMSE: 1.13%

California case prediction
- Relative RMSE: 1.18%

New York death prediction
- Relative RMSE: 0.64%

California death prediction
- Relative RMSE: 1.17%

Florida case prediction
- Relative RMSE: 2.91%

Hawaii case prediction
- Relative RMSE: 7.34%

Florida death prediction
- Relative RMSE: 1.10%

Hawaii death prediction
- Relative RMSE: 7.26%
 Effective Reproduction Number Calculation and Trends

$X$ as the vector of infected class (i.e. E, I)
$Y$ as the vector of uninfected class (i.e. S, R, D).

Let $\frac{dx}{dt} = F(X, Y) - V(X, Y)$,
where $F(X, Y)$ is the vector of new infection rates (flows from $Y$ to $X$) and $V(X, Y)$ is the vector of all other rates.

$$M = \begin{pmatrix} \frac{\partial F}{\partial X} & (\frac{\partial F}{\partial X})^{-1} \end{pmatrix}_{(N,0,0,0,0)} = \begin{bmatrix} 0 & \beta(t) \cdot [\sigma \ (1 - \alpha(t)) \cdot \gamma + \alpha(t) \cdot \rho]^{-1} \\ 0 & 0 \end{bmatrix}$$

$$R_t = \frac{\beta(t)}{(1 - \alpha(t)) \cdot \gamma + \alpha(t) \cdot \rho}$$

[1] Effective Reproduction Number Calculation and Trends
Fitted $\text{Rep}(t)$ at the start of the epidemic across all states

$\text{Rep}(t)$ ranges from 1.27 to 16.49, with a median value of 2.87
Results

Effective Reproduction Number Calculation and Trends

Fitted $Rep(t)$ on July 28th
Results

Effective Reproduction Number Calculation and Trends

History of effective reproduction number for New York, California, Florida and Hawaii
Results

Death Rate Trends

History of death rate for New York, California, Florida and Hawaii
Results

- Our model is premised on the assumption that transmission rates do not at first go down, and then later go up, it needs to be modified for states that exhibit multiple waves of the disease.

History of Hawaii’s intervention policy

1. The Hawaii Department of Health announced the first positive case on **March 6th**
2. Then they immediately enacted a stay at home order on **March 25th**
3. From **April 19th** to **May 7th**, the case curve flattened
4. The state announced on **May 7th** that Hawaii would embark on the first phase of reopening

Two-phase fitting methods

- We fit the first stage with the initialization of only one exposed people at the start.
- To initialize the second phase, we use the predicted number of exposed people, infectious people and recovered people from the first phase, combined with the reported deaths as of **May 7th**.
Results

Multi-Phase Model

Fitting results for the two phases Hawaii

Historical results of the effective reproduction number and death rate

The RRMSE for cases declines below 2.5% and the RRMSE for deaths declines below 2.7%
We have developed an extension of the SEIRD model that represents changing transmission and death rates over time as a continuous Sigmoid function, under the hypothesis that these rates change gradually, rather than immediately upon implementation of public health policies or treatments.

The model fits historical data for the United States well for most states. Using Hawaii as an example, a multi-phase extension of the model provides a more accurate fit.

Both the base model and multi-phase model provide a way to explain historical reported cases and deaths with a small set of parameters, which in the future can enable analyses of uncertainty and variations in disease progression across regions.
❑ Website: www.covid19datasource.com

Thanks for listening

Q&A