A Data-driven Markov Chain Model for COVID-19 Transmission in South Korea

Sujin Ahn and Minhae Kwon
Brain and Machine Intelligence Lab.
Soongsil University
Seoul, Republic of Korea
asujin331@soongsil.ac.kr, minhae@ssu.ac.kr

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; COVID-19) has rapidly transmitted between people. We introduce epidemic model with the additional states such as a vaccinated and isolated state. A data-driven epidemic model based on the Markov chain would be a desirable approach to overcome the challenges that infer the latent states. To this aim, we take advantage of the reported data and underlying Markov chain dynamics. To verify our model, we set initial values of each states and estimated the state values by fitting COVID-19 dataset of South Korea. Throughout the investigation, it is confirmed that the proposed models can successfully estimate all states.

1 Introduction

COVID-19 has rapidly spread across the globe since 2019. As of August 4, 2021, approximately 200 million confirmed cases were reported worldwide [1]. Under the circumstance of various viral variants that have emerged, the importance of mathematical modeling of infectious diseases is highlighted to understand the ongoing outbreak.

There have been several approaches to model the overall course of the epidemic such as Susceptible-Infectious-Recovered (SIR) [2], and Susceptible-Exposed-Infectious-Recovered (SEIR) [3]. Such approaches are helpful to review a course of an epidemic once it is terminated, however, they have several limitations as follows. First, the conventional models provide only a macroscopic view of the epidemic by reaching a theoretical conclusion such that everyone is recovered or dead at the end of the epidemic.The conventional models explain the whole process of disease spread using time-independent epidemic parameters [4, 5]. Thus, these works are less helpful to understand and respond to the ongoing outbreak because the time-independent epidemic parameters missing the context of social policy. Next, the conventional models partition the population simply into a few groups. To model the real-world situation of COVID-19, however, we need a more complicated model that includes a vaccinated and an isolated population [6]. Lastly, the conventional epidemic models do not use data-driven approach to estimate the number of latent groups such as individuals who are infected but do not have symptoms yet [7]. However, estimating the number of individuals in the latent group is critical to plan the national-wide policy such as social distancing, travel restriction, and lockdown.

In this work, we propose a data-driven epidemic model based on the Markov chain. To reflect the real-world population, we design 7 states where each state indicates a group of populations including a vaccinated [8] and isolated group [9]. Our model uses daily reported data (e.g., daily new confirmed cases, daily new vaccinated cases, the number of isolated individuals, etc) and fits them into our model to find underlying dynamics parameters of the Markov chain. Thus, it can provide the estimation of the latent state, which could be a piece of critical information to policymakers.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

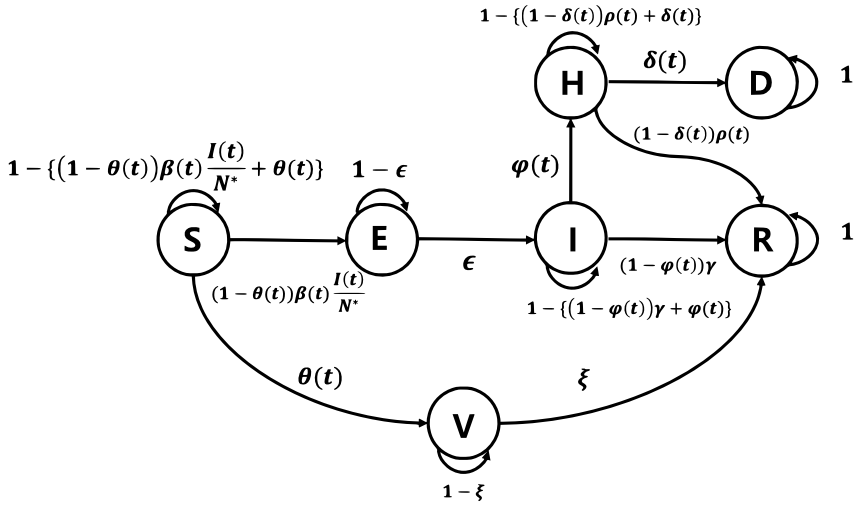

Figure 1: The state transition diagram of the proposed model

2 Proposed model

In this section, we introduce the Markov chain based transmission model to explain COVID-19 transmission. We consider a discrete-time Markov chain for grouping population and explain transitions between them. The considered state transition diagram is provided in Figure 1.

2.1 States

The proposed model includes 7 states; susceptible, exposed, infectious, recovered, vaccinated, hospitalized, and death. The population is partitioned into non-overlapping states so that an individual must belong to a state at a time.

- **Susceptible S**: Individuals who have the possibility of being infected. They have never gotten infected or vaccinated.
- **Vaccinated V**: Individuals who are vaccinated at least one dose.
- **Exposed E**: Individuals who are infected but not capable of transmitting the infection to another person.
- **Infectious I**: Individuals who are infected and capable of transmitting the infection to another person. They are not isolated from the susceptible group.
- **Hospitalized H**: Individuals who are confirmed the disease and isolated.
- **Recovered R**: Individuals who have immunity. They are recovered from the infection or have achieved immunity from the vaccine.
- **Death D**: Individuals who passed away because of the infection.

Suppose $\mathbf{S}(t), \mathbf{V}(t), \mathbf{E}(t), \mathbf{I}(t), \mathbf{H}(t), \mathbf{R}(t)$ and $\mathbf{D}(t)$ be state values, meaning the number of people that belongs to the state at time t and $N = \mathbf{S}(t) + \mathbf{V}(t) + \mathbf{E}(t) + \mathbf{I}(t) + \mathbf{H}(t) + \mathbf{R}(t) + \mathbf{D}(t)$ be total population. The total population N is assumed to be the close number which is the same over time as births and natural deaths are balanced.

2.2 State transition probabilities

The proposed Markov chain model satisfies the Markov property which means that the probability of moving to the next state depends only on the present state and not on the previous states [10]. The underlying dynamics of Markov chain can be described by the state transition probabilities. Let $J, J' \in \{\mathbf{S}, \mathbf{V}, \mathbf{E}, \mathbf{I}, \mathbf{H}, \mathbf{R}, \mathbf{D}\}$ be states and $p_{JJ'}(t)$ be the state transition probability from the state

J at time t to the state J' at time $t+1$. We allow the state transition probability to be time-varying. If $J' = J$, $p_{JJ}(t)$ denotes the probability of remaining in the state J at time t and time $t+1$.

Suppose $X_{JJ'}(t) \sim \mathcal{B}(J(t), p_{JJ'}(t))$ be a random variable at time t that follows the Binomial distribution, i.e.,

$$\Pr\left(X_{JJ'}(t) = x; J(t), p_{JJ'}(t)\right) = \frac{J(t)!}{x!(J(t)-x)!} p_{JJ'}(t)^x \left(1 - p_{JJ'}(t)\right)^{J(t)-x} \tag{1}$$

which means the number of people who moves from J to J'. Thus, the expected number of people is $\mathbb{E}(X_{JJ'}(t)) = J(t)p_{JJ'}(t)$. Figure 1 shows a digraph that describes all possible state transitions. All unmarked edges are forbidden transitions, i.e., $p_{JJ'}(t) = 0$.

In Figure 1, the state $\mathbf{S}$ has two outflows. The one goes to $\mathbf{V}$ with $p_{\mathbf{SV}}(t) = \theta(t)$, where $\theta(t)$ denotes the daily vaccination ratio of the first dose. The second outflow goes to $\mathbf{E}$ with $p_{\mathbf{SE}}(t) = (1-\theta(t))\,\beta(t)\frac{\mathbf{I}(t)}{N^*}$, where $\beta(t)$ denotes the transmission rate at time t and $N^* = N - \mathbf{H}(t) - \mathbf{D}(t)$ denotes the number of individuals who are not isolated or dead. The remaining ratio of the susceptible at time t is denoted as $1-\theta(t)$, and $\beta(t)\frac{\mathbf{I}(t)}{N^*}$ is the likelihood of the disease being transmitted to the susceptible. Thus, $p_{\mathbf{SE}}(t)$ indicates the probability of being infected from susceptible state. The state $\mathbf{E}$ has a single outflow heading to $\mathbf{I}$, meaning an exposed individual is now capable of transmitting the infection. The state transition occurs with the probability of $p_{\mathbf{EI}}(t) = \epsilon$, where ϵ denotes the inverse of the latent period. The state $\mathbf{I}$ has two outflows. The one goes to $\mathbf{H}$, with $p_{\mathbf{IH}}(t) = \varphi(t)$, where $\varphi(t)$ denotes the daily detected case ratio at time t. The second one goes to $\mathbf{R}$ with $p_{\mathbf{IR}}(t) = (1-\varphi(t))\gamma$, which means the undetected infectious individuals are immune. Here, $1-\varphi(t)$ indicates the portion of individuals who are infected but undetected, and γ indicates the recovery rate. The state $\mathbf{H}$ has two outflows. The first outflow is the population that goes to state $\mathbf{D}$ with $p_{\mathbf{HD}}(t) = \delta(t)$, where $\delta(t)$ denotes the fatality rate at time t. The second outflow is the population that goes to $\mathbf{R}$ with the probability of $p_{\mathbf{HR}}(t) = \rho(t)(1-\delta(t))$, where $\rho(t)$ denotes the recovery rate for hospitalized individuals at time t and $1-\delta(t)$ indicates the portion of individuals who managed to survive. The state $\mathbf{V}$ goes to $\mathbf{R}$ with $p_{\mathbf{VR}}(t) = \xi$, where ξ denotes the inverse of mean duration of immunity build after getting the first dose. We assumed that ξ is the $\frac{1}{49}$ [11] and ϵ,γ is the $\frac{1}{2}$, $\frac{1}{9}$, respectively [12]. The states $\mathbf{R}$ and $\mathbf{D}$ are absorbing states such that there is no outflow. Based on the model, we will estimate each state values in the next section.

3 Results

In this section, we aim to estimate state values by using the proposed model. The challenging points are how to estimate the latent state values such as $\mathbf{S}(t), \mathbf{E}(t), \mathbf{I}(t)$ and $\mathbf{R}(t)$ and how to set the initial state values.

To combat the challenges, first of all, we take advantage of the known data and underlying Markov chain dynamics to estimate the latent states. The real data of $X_{\mathbf{SV}}, X_{\mathbf{IH}}, X_{\mathbf{HR}}, \mathbf{H}(t)$, and $\mathbf{D}(t)$ are known. Thus, we could get clues to estimate $\mathbf{S}(t)$ and $\mathbf{I}(t)$ since they are the origins of $X_{\mathbf{SV}}$ and $X_{\mathbf{IH}}$, respectively.

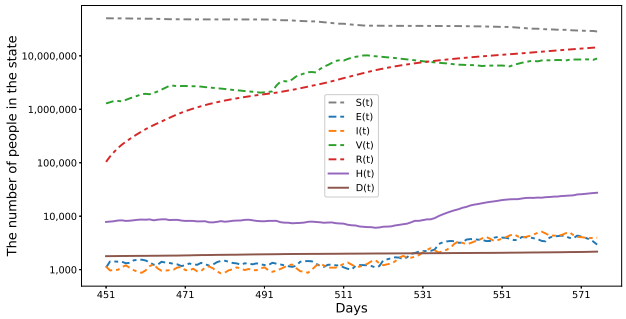

Figure 2: The state values over days. The solid lines are the known data, and the broken lines are the estimated values.

Next, we set the initial state values of $\mathbf{E}$ and $\mathbf{I}$. $\mathbf{E}(t-1)$, the initial value of $\mathbf{E}$ was estimated by tracing back the future population of infectious state since time t. This is because the population in the $\mathbf{I}$ state have stayed in previous state $\mathbf{E}$ before. We initialize $\mathbf{E}(t-1) = \sum_{s=1}^{t} w(s)\mathbf{I}(t+s-1)$, the sum of the population of infectious state from time t, weighted by $w(s)$. Here, $w(s)$ denotes the infectivity profiles, the probability of individuals that capable of transmitting the infection to another person at a

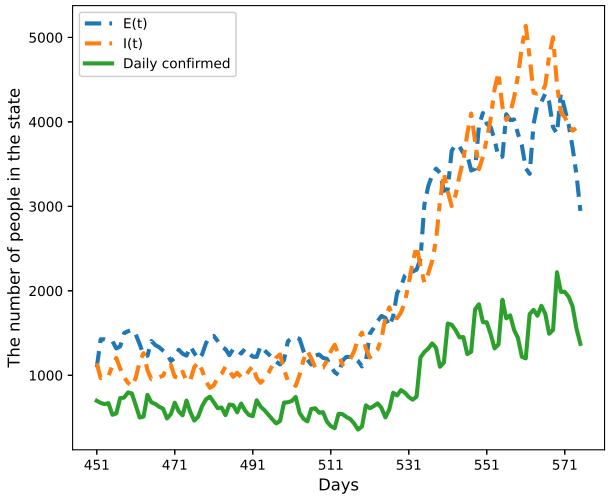

Figure 3: The estimation of $\mathbf{E}(t)$ and $\mathbf{I}(t)$ over days. $X_{\mathbf{IH}}(t)$ is provided in solid line

day s relative to the onset of symptoms [13]. According to Pan American Health Organization, $w(s)$ represented as $\Gamma(s; \frac{\mu^2}{\sigma^2}, \frac{\mu}{\sigma^2})$ where $\mu = 4.8$, $\sigma = 2.3$ [14].

The initial value of $\mathbf{I}$ was estimated by tracing back $X_{\mathbf{IH}}$. This is because $X_{\mathbf{IH}}$ was originated from the population in the $\mathbf{I}$ state. We initialize $\mathbf{I}(t-1)$ as

$$\mathbf{I}(t-1) = \alpha^{-1} X_{\mathbf{IH}}(t-1) + \sum_{s=2}^{t} \left(\alpha^{-1} - 1\right) X_{\mathbf{IH}}(t-s)(1-\gamma)^{s-1} \tag{2}$$

$\mathbf{I}(t-1)$ indicates the states of undetected infectious individuals that are still not recovered until time $t-1$. Here, weighted by α, the assumed portion of detected case. Thus, $\alpha^{-1} - 1$ enables tracking the undetected portions of $\mathbf{I}$ states. We can also infer the state values of $\mathbf{E}$ and $\mathbf{I}$ based on its initial value.

To confirm the proposed model, we estimate the state values using COVID-19 dataset of South Korea. In Figure 2, the estimation results are provided using data between April 15, 2021 and August 17, 2021. We count the days based on the beginning of the COVID-19 outbreak. The solid lines are the real data that are used for estimation, and the broken lines are the estimated results. We confirm that the proposed solution is able to estimate all states. It is easily observed that $\mathbf{S}(t)$ is decreasing and $\mathbf{R}(t)$ is increasing. This is because individuals in $\mathbf{S}(t)$ are moving toward $\mathbf{R}(t)$ and accumulated in $\mathbf{R}(t)$, which is the absorbing state.

The microscopic views of $\mathbf{I}(t)$ and $\mathbf{E}(t)$ are provided in Figure 3. It is shown that changing shape of $\mathbf{I}(t)$ and $\mathbf{E}(t)$ are similar to that of $X_{\mathbf{IH}}$. This is because $X_{\mathbf{IH}}$ is a portion of $\mathbf{I}(t)$ (i.e., detected individuals) so that similar changing shapes are expected unless the detection rate changes.

4 Discussions

In this paper, we proposed 7 states in our model corresponding to control measures in South Korea to capture more aspects of real-world transmission. Combining our epidemiological model with discrete-time Markov chain allows us not only effectively to fit discrete real-data, but also to estimate the number of individuals at any state based on underlying dynamics. Meanwhile, we introduced time-varying parameters in Markov chain dynamics such as $\beta(t), \rho(t), \delta(t), \phi(t), \theta(t)$, which could be real-time indicators. For future works, we will forecast the future values of time-varying parameters in Markov chain dynamics based on RNN[15]. The following values of time-varying parameters will be clues to infer the future values of the latent states.

Finally, these open questions are under current investigation:

- How to evaluate the estimation of state to show validity of the latent state's value?
- How to apply data and modify the proposed model to fit other countries' outbreak?
- How to implement a deep learning method related to short-term forecast to grasp the future transmission patterns?

References

[1] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534, 2020.

[2] Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.

[3] Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical models in epidemiology*, volume 32. Springer, 2019.

[4] Chris Groendyke and Adam Combs. Modifying the network-based stochastic seir model to account for quarantine: an application to covid-19. *Epidemiologic Methods*, 10(s1), 2021.

[5] D Pal, D Ghosh, PK Santra, and GS Mahapatra. Mathematical analysis of a covid-19 epidemic model by using data driven epidemiological parameters of diseases spread in india. *medRxiv*, 2020.

[6] Sunhwa Choi and Moran Ki. Estimating the reproductive number and the outbreak size of covid-19 in korea. *Epidemiology and health*, 42, 2020.

[7] Jian Chen, Michael C Fu, Wenhong Zhang, and Junhua Zheng. Predictive modeling for epidemic outbreaks: A new approach and covid-19 case study. *Asia-Pacific Journal of Operational Research*, 37(03):2050028, 2020.

[8] Zhe Xu, Bo Wu, and Ufuk Topcu. Control strategies for covid-19 epidemic with vaccination, shield immunity and quarantine: A metric temporal logic approach. *PloS one*, 16(3):e0247660, 2021.

[9] Carlos Balsa, Isabel Lopes, Teresa Guarda, and José Rufino. Computational simulation of the covid-19 epidemic with the seir stochastic model. *Computational and Mathematical Organization Theory*, pages 1–19, 2021.

[10] Mark Pinsky and Samuel Karlin. *An introduction to stochastic modeling*. Academic press, 2010.

[11] World Health Organization et al. Guidance on conducting vaccine effectiveness evaluations in the setting of new sars-cov-2 variants: interim guidance, 22 july 2021. addendum to evaluation of covid-19 vaccine effectiveness: interim guidance. Technical report, World Health Organization, 2021.

[12] Joseph T Wu, Kathy Leung, and Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, 395(10225):689–697, 2020.

[13] Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology*, 178(9):1505–1512, 2013.

[14] Pan American Health organization. Covid-19 modeling exercise, 2021.

[15] Alex Sherstinsky. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404:132306, 2020.