East Asian Math. J.
Vol. 39 (2023), No. 5, pp. 493–503
http://dx.doi.org/10.7858/eamj.2023.035

EARLY WARNING FORECASTS FOR COVID-19 IN KOREA USING BAYESIAN ESTIMATION OF THE TRANSMISSION RATE

Byul Nim Kim

Abstract. Tendency prediction of daily confirmed cases is an important issue for public health authorities. To protect the tendency, we estimate the transmission rate of stochastic SEIR model for COVID-19 in Korea using particle Markov chain Monte Carlo method. The results show that the increasing and decreasing tendency of estimated transmission rate appear one or two days in advance compared to daily incidence cases, and as time evolves the standard deviation of the estimates of transmission rate reduces. Since ten months have passed since the first incident case of COVID-19 in Korea, we expect to forecast the tendency of daily confirmed cases for the next one or two days more accurately using our method.

1. Introduction

Mathematical models for the transmission dynamics of infectious diseases aim at the understanding epidemiological patterns and predicting the consequences of public health interventions. In most classical disease transmission models, the transmission rate plays a role in ensuring that the model gives a reasonable qualitative description of the disease dynamics. Instead of assuming the transmission rate as a constant, assuming time-varying transmission rate might be more realistic approach. In this paper, we analyze a stochastic SEIR model with time-varying incidence rate. Accurate estimation of the transmission rate is an essential issue because it has a significant influence on model predictions and conclusions.

Models for predicting time-variance have been studied in many ways. The sequential Monte Carlo methods have successfully been applied to a range of problems requiring rapid online analysis of data, such as target tracking, and data streaming [1, 2, 3]. Typically, the Matingale methods are based on counting process of data of infections and recoveries [6, 7, 8, 9]. The approximation methods have been the useful for diffusion process [10] and the Gaussian process [11]

Received August 14, 2023; Accepted August 30, 2023.
2010 *Mathematics Subject Classification.* 11A11.
Key words and phrases. Epidemiology, Mathematical modeling, Bayesian, Transmission rate.

(pISSN 1226-6973, eISSN 2287-2833)

is based on the Markov jump process. Simulation-based models included the approximate Bayesian computation methods [12, 13], pseudo-marginal methods [14] and sequential particle filter methods [13, 15, 16, 17, 18]. Traditional agent–based data augmentation methods have been target the joint posterior distribution of the missing data and model parameters to obtain a tractable complete data likelihood [19, 20, 21, 22, 23].

Recently, sequential Monte Carlo methods of the parameter estimation, which combines Bayesian estimation method with Markov chain Monte Carlo methods, were developed. Camacho et al. model the time-varying transmission rate, β_t, by a Wiener process (also known as standard Brownian motion) with positive constraints using a stochastic SEIR framework [24]. To estimate future cases in real time, 5000 stochastic trajectories are simulated by sampling a set of parameters and states from the joint posterior distribution for the last fitted data point. This method is a kind of particle Markov chain Monte Carlo. Funk et al. reconstructed a time-varying transmission rate, β_t, by reforming the Camacho et al.'s Wiener process model. They adopted that the observation process was modelled to operate on the weekly incidence, given by the number of infectious individuals entering the quarantine compartment. They assumed that the observed incidence followed a normal approximation to the negative binomial distribution with reporting probability and overdispersion [25]. Thompson et al. proposed a two-step procedure to estimate the time-dependent reproduction number from data informing the serial interval and from data on the incidence of cases. They adopted Bayesian MCMC method. The first step uses data on known pairs of index and secondary cases to estimate the serial interval distribution; the second step estimates the time-varying reproduction number estimated jointly from incidence data and from the posterior distribution of the serial interval obtained in the first step. They have revised the approach of Cori et al. for estimating with the time-dependent method [26, 27].

Unlike the traditional method of adding the Markov chain Monte Carlo methods to the ordinary differential equation methods, we combine the sequential Monte Carlo methods with the Markov chain Monte Carlo methods to the SEIR model in which the distribution of events with probabilistic mechanisms have a binomial distribution. And in parameter estimation, we assume the number of infectious patients is proportional to the number of quarantine patients, unlike the existing method using conceptually ambiguous incidence data. As a result we have an estimation of the transmission rate with standard deviation for each time. The estimations of the transmission rate give us to early forecast the tendency of the incidence.

2. Data and Method

2.1. Data

To estimate the transmission parameter, the daily data from the Korea Disease Control and Prevention Agency (KDCA) were analyzed and the cases confirmed between February 8, 2020 to April 9, 2020 were used [41]. The data provide daily domestic confirmed cases and daily confirmed inflow-cases from foreign countries and cases discharged from quarantine authorities. We excluded the inflow-cases from the confirmed cases, in that those are immediately quarantined once they are confirmed at the National Quarantine Station in Korean territory so they have no chance to transmit the disease in Korea. As shown in Figure 1(A), the occurrence in Korea increased rapidly in the early stages, and then gradually calmed down after about 60 days. The number of quarantined or isolated cases is displayed in Figure 1(B), which taken as the cumulative confirmed cases minus the cumulative discharged by quarantine authorities and deaths.

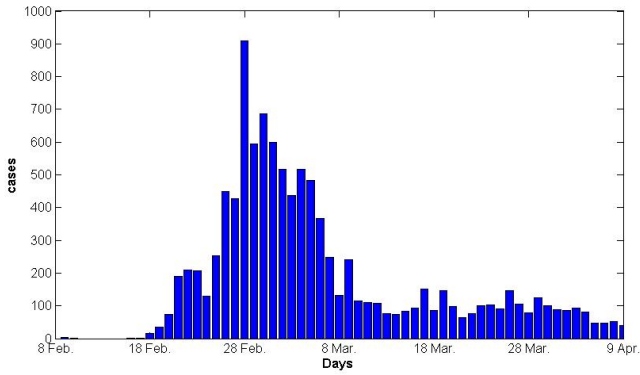
2.2. Methods

The underlying mathematical model for COVID-19 infection is the deterministic Kermack and McKendric SEIR model, where the population is divided into four groups, including susceptible (S), exposed (E), infectious (I), and recovered or death (R):

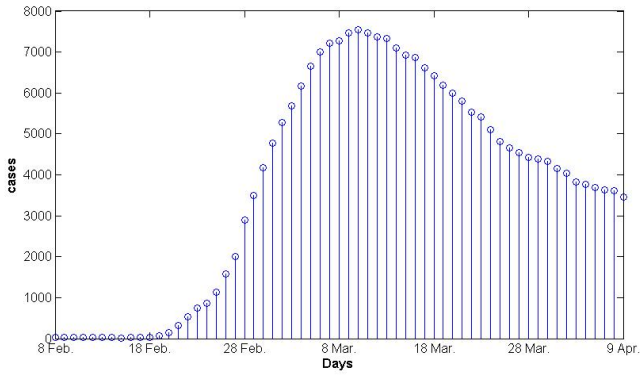
$$\begin{aligned}\frac{dS}{dt} &= -\beta IS/N \\ \frac{dE}{dt} &= \beta IS/N - \kappa E \\ \frac{dI}{dt} &= \kappa E - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

Transmission rate (β) explains how many effective contacts occur in susceptible class and $\beta IS/N$ is the number of individuals newly infected and κ is the per capita rate of being infectious. Usually, the length of the latent period is obtained $1/\kappa$, and the mean time spent in the infectious class is used for $1/\gamma$.

In the real world where these processes are applied, each compartment changes to an increment over time, so we can formulate the dynamic deterministic parts as continuous time Markov chain (CTMC) in a discrete time model, i.e., we can convert the preceding deterministic parts into counting processes as follows [32].



(A) The number of daily confirmed domestic cases of COVID-19 in Korea



(B) The number of quarantine or isolated cases of COVID-19 in Korea

FIGURE 1. KDCA Data from February 8 to April 9 in Korea [41].

$$
\begin{aligned}
S_{t+1} &= S_t - \Delta_{SE} \\
E_{t+1} &= E_t + \Delta_{SE} - \Delta_{EI} \\
I_{t+1} &= I_t + \Delta_{EI} - \Delta_{IR} \\
R_{t+1} &= R_t + \Delta_{IR}
\end{aligned}
$$

The numbers of individuals within each compartment change through time in increments as discrete variation amounts between compartments occur. The notation Δ_{AB} represents the number of individuals moving from A to B during the time interval $[t, t + \Delta t)$. For example, Δ_{SI} is the number of people infected

or infectious from susceptible class in Δt. Two events Δ_{SE}, Δ_{EI}, Δ_{IR} of "outcome" are possible: becoming infected or recovering. Each event has the binomial distribution with the outcome probability of an event, induced by its hazard function, as follows [20].

$$\Delta_{SE} \sim B(S_t, 1 - e^{-\beta_t I_t/N})$$
$$\Delta_{EI} \sim B(E_t, 1 - e^{-\kappa})$$
$$\Delta_{IR} \sim B(I_t, 1 - e^{-\gamma})$$

For estimating the transmission rate β_t, we adopt the statistical technique, so called Particle filter (PF) or sequential Monte Carlo (SMC) [36, 35, 37, 38, 39], to recursively explore conditional densities in state-space models. For given values of θ, N particles $(\tilde{x_i}^j)$ are sequentially propagated from t_0 to t_n. In each step t_i, the trajectories that best fit the data $y_{1:i}$ are given more weight through importance sampling techniques. Given the observation data $y_{1:k}$ up to time k, the posterior distribution $\pi(x_{1:n}, \theta|y_{1:n})$ is defined using the PF methods. And in order to sample from $\pi(x_{1:n}, \theta|y_{1:n})$, we use the popular Bayesian method, so called particle Markov chain Monte Carlo (PMCMC), which combines PF and MCMC [15]. In this paper, in particular, we used the Bootstrap filter method, which is the simplest method of PF, and the Metropolis-Hastings algorithm, which is the classical system, as the MCMC algorithm [40]. The procedures for these two methods are as follows, where the observed data y_t is quarantine patients, the latent variables x_t are exposed (E) and recoveries (R), and the parameters we should estimate are time-varying β_T. Here we assume that the number of infectious patients (I_t) is proportional to the numbers of quarantine patients The reason we use the assumption will be discussed in Discussion and Conclusions.

3. Results

In the simulation of PMCMC, we assume the hierarchical prior distribution of the transmission parameter β_t to be $Beta(\alpha_1, \alpha_2)$, the prior of α_1 and α_2 to be $Beta(1, 1)$, and the parameters κ, γ for COVID-19 in Korea to be $\frac{1}{7}$, $\frac{1}{14}$, respectively [33].

Figure 2 represents the relation between the estimate of β_t and daily incidence cases. The increasing and the decreasing tendency of estimated β_t appear one or two days in advance compared to the daily incident cases. The increasing (decreasing) tendency of β_t from time t to $t+1$ shows the increasing (decreasing, respectively) tendency of daily incidence cases from time t to $t+1$ or (from time $t+1$ to $t+2$). For close look, we show the cross-correlation in Figure 2b and 2c. The highest one occurs at lag +1 (Cross-correlation: 0.952 with 95 % CI) and the second highest one occurs at lag +2 (Cross-correlation: 0.930 with 95 % CI), implying that the estimation of β_t is highly related to the number of daily incidence cases on later one or two days.

Procedure: Particle MCMC

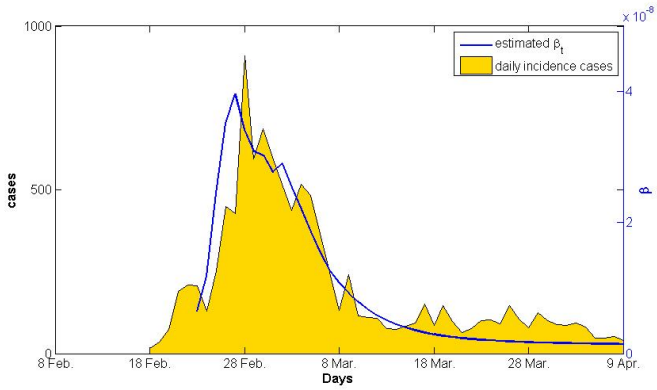
Choose an initial value $\beta_T^{(0)}$;
for $m \leftarrow 1$ **to** M **do**
  Generate $x_0^{(m)} \sim f(x_0)$;
  Set $\pi_0^{(m)} = \frac{1}{M}$;
end
for $t \leftarrow 1$ **to** T **do**
  for $m \leftarrow 1$ **to** M **do**
    Generate $\tilde{x}_t^{(m)} \sim q(x_t | x_{t-1}^{(m)}, y_t)$;
    Calculate unnormalized weight $w_t^{(m)} = \frac{f(\tilde{x}_t^{(m)} | x_{t-1}^{(m)}) g(y_t | \tilde{x}_t^{(m)})}{q(\tilde{x}_t^{(m)} | x_{t-1}^{(m)}, y_t)} \pi_{t-1}^{(m)}$;
  end
  for $m \leftarrow 1$ **to** M **do**
    Normalize $w_t^{(m)}$ as $\pi_t^{(m)} = \frac{w_t^{(m)}}{\sum_{i=1}^{M} w_t^{(i)}}$;
  end
  for $m \leftarrow 1$ **to** M **do**
    Sample an index j_m from the set $1, \ldots, M$ with probabilities $\{\pi_t^{(m)}\}_{m=1}^{M}$;
    Set $x_t^{(m)} = \tilde{x}_t^{(j_m)}$;
    Set $\pi_t^{(m)} = \frac{1}{M}$;
  end
  Calculate $\tilde{p}(y_{t|1:t-1}) = \frac{1}{M} \sum_{m=1}^{M} w_t^{(m)}$;
end
Draw $x_{1:T}^{(0)} \sim p(x_{1:T} | y_{1:T}, \beta_T^{(0)})$ by sampling with its full state history;
for $i \leftarrow 1$ **to** I **do**
  Generate $\beta_T^* \sim q(\beta_T | \beta_T^{(i-1)})$;
  Calculate unnormalized weight $w_t^{(m)} = \frac{f(\tilde{x}_t^{(m)} | g(y_t | \tilde{x}_t^{(m)})}{q(\tilde{x}_t^{(m)} | x_{t-1}^{(m)}, y_t)} \pi_{t-1}^{(m)}$;
  Obtain the marginal likelihood estimate $\tilde{p}(y_{1:T} | \beta_T^*)$;
  Draw $x_{1:T}^* \sim p(x_{1:T} | y_{1:T}, \beta_T^*)$ by sampling with its full state history;
  Compute $a^* = \min \left[1, \frac{\tilde{p}(y_{1:T} | \beta_T^*) p(\beta_T^*)}{\tilde{p}^{(i-1)} p(\beta_T^{(i-1)})} \frac{q(\beta_T^{(i-1)} | \beta_T^*)}{q(\beta_T^* | \beta_T^{(i-1)})} \right]$;
  Generate $r \sim U(0, 1)$;
  if $a^* > r$ **then**
    Set $\beta_T^{(i)} = \beta_T^*$, $x_{1:T}^{(i)} = x_{1:T}^*$, and $\tilde{p}^{(i)} = \tilde{p}(y_{1:T} | \beta_T^*)$;
  else
    Set $\beta_T^{(i)} = \beta_T^{(i-1)}$, $x_{1:T}^{(i)} = x_{1:T}^{(i-1)}$, and $\tilde{p}^{(i)} = \tilde{p}^{(i-1)}$;
  end
end

We represent the posterior distribution of β_t between February 23 and March 13 in Figure 3. The posterior distributions of the first five days in the period are positioned sparsely and have a large standard deviation compared to the other days. As the estimation of β_t increases, we have a larger standard deviation; after the highest peak of the estimation of β_t, the standard deviation and β_t decrease simultaneously. During the small fluctuation of β_t after the highest peak, the standard deviation also fluctuates. As time evolves, the standard deviations are getting smaller.

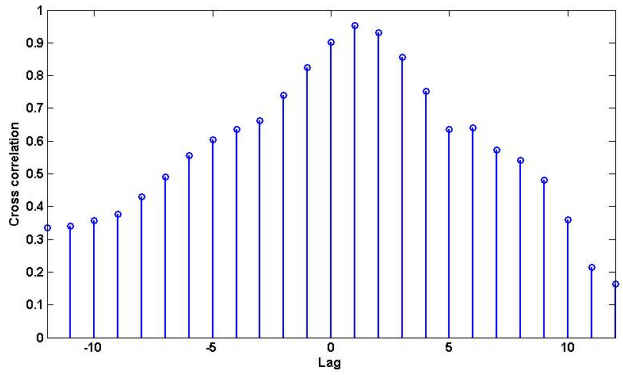
4. Discussions

The preceding tendency of estimated β_t to daily incidence cases as shown in Figure 2 is useful in practice. We see that the increasing and decreasing tendency of estimated β_t appear one or two days in advance compared to daily incidence cases. Therefore, if we can estimate β_t in COVID-19 at the current time, we may predict the increasing or decreasing daily incidence cases for the next one or two day in advance. As we see in Figure 3, as time evolves, the standard deviation of β_t is getting small such that the estimate of β_t will be more accurate than the early stage, The estimation of β_t is a vital tool for predicting daily incidence cases, because we have data for ten months since the first incident case to inform our model.

Once the Korean governmental surveillance system detects a person, there is no more chance to transmit. Hence, when we apply the data of daily confirmed cases to the SEIR compartment model, we should not identify the confirmed date data as the data of onset date on the same day. M. Ki [34] analyzed the KCDA raw data of 28 confirmed patients in Korea and could discriminate the onset date and the date detected by surveillance. The results showed that there is an average of 4 days difference. Such analysis could be possible, because the number of those data is small, and such confidential data are available to the author. However, the onset date information is not available to researchers but also too big to figure the onset date out for each case. The promising candidate based mathematical model for COVID-19 in Korea to estimate of the transmission rate might be in Figure 4. In Section 2, we assumed that the number of infectious patients (I_t) is proportional to the numbers of quarantine patients (Q_t) to apply I_t to the SEIR model, not to the SEIQR model in Figure 4. The government should provide more detailed information for each incidence case so that the researcher can estimate the transmission rate β_t with small standard deviation, which predicts the tendency for daily incidence cases more accurately and earlier in advance.



(A) The estimated β_t vs. daily incidence cases

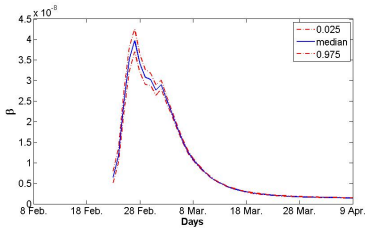


(B) The cross-correlation plot between estimated β_t and daily incidence cases.

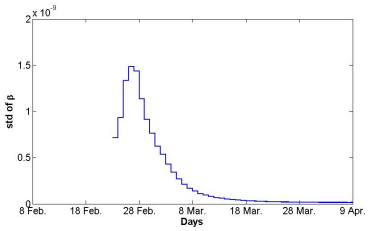
Lag	-10	-9	-8	-7	-6	-5	-4
Cross-correlation	0.357	0.377	0.429	0.491	0.555	0.603	0.635
Lag	-3	-2	-1	0	1	2	3
Cross-correlation	0.661	0.739	0.824	0.901	0.952	0.930	0.855
Lag	4	5	6	7	8	9	10
Cross-correlation	0.751	0.635	0.639	0.574	0.542	0.481	0.359

(C) The cross-correlation between estimated β_t and daily incidence cases.

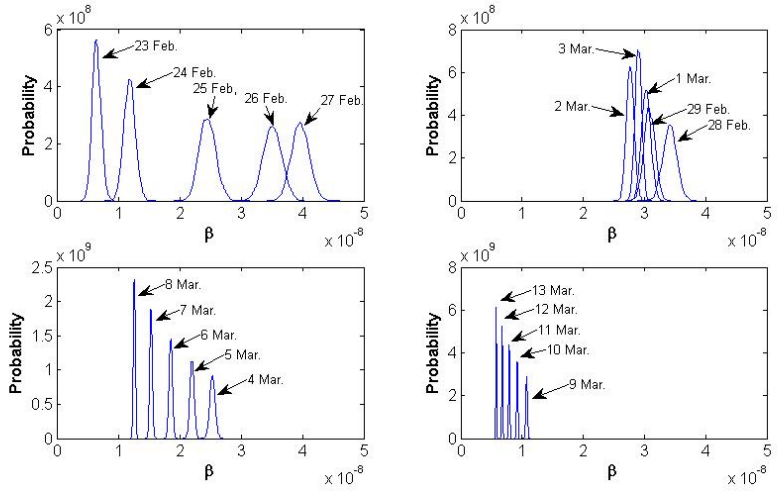
FIGURE 2. The preceding tendency of estimated β_t to daily incidence cases



(A) The percentile estimates of β_t



(B) standard deviation of β_t



(C) The posterior distribution of β_t

FIGURE 3. The daily density estimations of β_t

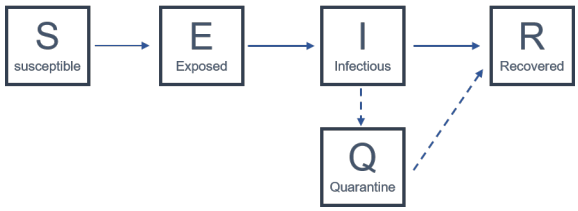


FIGURE 4. Flow chart for COVID-19 transmission dynamics.

Acknowledgement

This research was supported by Kyungpook National University Development Project Research Fund, 2020

References

[1] Welding, J., and Neal, P. (2019). Real time analysis of epidemic data. arXiv preprint arXiv:1909.11560.

[2] Nemeth, C. (2014). Parameter estimation for state space models using sequential Monte Carlo algorithms (Doctoral dissertation, Lancaster University).

[3] Zhu, J., Chen, J., Hu, W., and Zhang, B. (2017). Big learning with Bayesian methods. National Science Review, 4(4), 627-651.

[4] Adams, BM. Banks, HT. Davidian, M. dae Kwon, H. Tran, HT. Wynne, SN. and Rosenberg, ES. Hiv dynamics: modeling, data analysis, and optimal treatment protocols. J. Comput. Appl. Math, 184 (2005), pp. 10–49.

[5] Adda, P. Dimi, JL. Iggidr, A. Kamgang, JC. Sallet, G. and Tewa, JJ. General models of host-parasite systems. Global analysis, Discrete Contin. Dyn. Syst. Ser. B, 8 (2007), pp. 1–17 (electronic).

[6] Becker, Niels G. "On a general stochastic epidemic model." Theoretical Population Biology 11.1 (1977): 23-36.

[7] Watson, Ray. "An application of a martingale central limit theorem to the standard epidemic model." Stochastic Processes and Their Applications 11.1 (1981): 79-89.

[8] Sudbury, Aidan. "The proportion of the population never hearing a rumour." Journal of applied probability (1985): 443-446.

[9] Andersson, H., and T. Britton. "Lecture notes in statistics." Stochastic epidemic models and their statistical analysis 151 (2000).

[10] Roberts, Gareth O., and Osnat Stramer. "On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm." Biometrika 88.3 (2001): 603-621.

[11] Jandarov, Roman, et al. "Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease." Journal of the Royal Statistical Society: Series C: Applied Statistics (2014): 423-444.

[12] McKinley, Trevelyan, Alex R. Cook, and Robert Deardon. "Inference in epidemic models without likelihoods." The International Journal of Biostatistics 5.1 (2009).

[13] Toni, Tina, et al. "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems." Journal of the Royal Society Interface 6.31 (2009): 187-202.

[14] McKinley, Trevelyan J., et al. "Simulation-based Bayesian inference for epidemic models." Computational Statistics & Data Analysis 71 (2014): 434-447.

[15] Andrieu, Christophe, Arnaud Doucet, and Roman Holenstein. "Particle markov chain monte carlo methods." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72.3 (2010): 269-342.

[16] Ionides, Edward L., et al. "Iterated filtering." The Annals of Statistics 39.3 (2011): 1776-1802.

[17] Dukic, Vanja, Hedibert F. Lopes, and Nicholas G. Polson. "Tracking epidemics with Google flu trends data and a state-space SEIR model." Journal of the American Statistical Association 107.500 (2012): 1410-1426.

[18] Koepke, Amanda A., et al. "Predictive modeling of cholera outbreaks in Bangladesh." The annals of applied statistics 10.2 (2016): 575.

[19] Auranen, Kari, et al. "Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data." Journal of the American Statistical Association 95.452 (2000): 1044-1053.

[20] Höhle, Michael, and Erik Jørgensen. Estimating parameters for stochastic epidemics. [The Royal Veterinary and Agricultural University], Dina, 2002.

[21] Cauchemez, Simon, et al. "A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data." Statistics in medicine 23.22 (2004): 3469-3487.

[22] Neal, Peter J., and Gareth O. Roberts. "Statistical inference and model selection for the 1861 Hagelloch measles epidemic." Biostatistics 5.2 (2004): 249-261.
[23] O'Neill, Philip D. "Bayesian inference for stochastic multitype epidemics in structured populations using sample data." Biostatistics 10.4 (2009): 779-791.
[24] Camacho, Anton, et al. "Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study." PLoS currents 7 (2015).
[25] Funk, Sebastian, et al. "Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model." Epidemics 22 (2018): 56-61.
[26] Thompson, R. N., et al. "Improved inference of time-varying reproduction numbers during infectious disease outbreaks." Epidemics 29 (2019): 100356.
[27] Cori, Anne, et al. "A new framework and software to estimate time-varying reproduction numbers during epidemics." American journal of epidemiology 178.9 (2013): 1505-1512.
[28] Dureau, Joseph, Konstantinos Kalogeropoulos, and Marc Baguelin. "Capturing the time-varying drivers of an epidemic using stochastic dynamical systems." Biostatistics 14.3 (2013): 541-555.
[29] Endo, Akira, Edwin van Leeuwen, and Marc Baguelin. "Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers." Epidemics 29 (2019): 100363.
[30] Bretó, Carles, and Edward L. Ionides. "Compound markov counting processes and their applications to modeling infinitesimally over-dispersed systems." Stochastic Processes and their Applications 121.11 (2011): 2571-2591.
[31] Martcheva, M. (2015). An introduction to mathematical epidemiology (Vol. 61). New York: Springer.
[32] King, Aaron A., Dao Nguyen, and Edward L. Ionides. "Statistical inference for partially observed Markov processes via the R package pomp." arXiv preprint arXiv:1509.00503 (2015).
[33] Choi, S., Ki, M. (2020). Estimating the reproductive number and the outbreak size of COVID-19 in Korea. Epidemiology and health, 42.
[34] Ki, M. (2020). Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV) disease in Korea. Epidemiology and health, 42.
[35] Doucet, Arnaud, and Adam M. Johansen. "A tutorial on particle filtering and smoothing: Fifteen years later." Handbook of nonlinear filtering 12.656-704 (2009): 3.
[36] Schön, Thomas B., et al. "Sequential Monte Carlo methods for system identification." IFAC-PapersOnLine 48.28 (2015): 775-786.
[37] Salmond, David. "Introduction to Particle Filters for Tracking and Guidance." Advances in Missile Guidance, Control, and Estimation 20121297 (2012).
[38] Gustafsson, Fredrik. "Particle filter theory and practice with positioning applications." IEEE Aerospace and Electronic Systems Magazine 25.7 (2010): 53-82.
[39] Chen, Zhe. "Bayesian filtering: From Kalman filters to particle filters, and beyond." Statistics 182.1 (2003): 1-69.
[40] Michaud, Nicholas, et al. "Sequential Monte Carlo methods in the nimble R package." arXiv preprint arXiv:1703.06206 (2017).
[41] KDCA briefing report, http://www.cdc.go.kr/npt/biz/npp/nppMain.do
[42] Jeyanathan, Mangalakumari, et al. "Immunological considerations for COVID-19 vaccine strategies." Nature Reviews Immunology (2020): 1-18.

Byul Nim Kim
Institute for Mathematical Convergence, Kyungpook National University, Daegu 41566, South Korea
Finance Fishery Manufacture Industrial Mathematics Center on Big Data, Busan National University, Busan, Republic of Korea 46241
Email address: air1227@pusan.ac.kr