

Intervention analysis for spread of COVID-19 in South Korea using SIR model

Sumin Cho^a, Jaejik Kim^{1,a}

^aDepartment of Statistics, Sungkyunkwan University

Abstract

COVID-19 has spread seriously around the world in 2020 and it is still significantly affecting our whole daily life. Currently, the whole world is still undergoing the pandemic and South Korea is no exception to it. During the pandemic, South Korea had several events that prevented or accelerated its spread. To establish the prevention policies for infectious diseases, it is very important to evaluate the intervention effect of such events. The susceptible-infected-removed (SIR) model is often used to describe the dynamic behavior of the spread of infectious diseases through ordinary differential equations. However, the SIR model is a deterministic model without considering the uncertainty of observed data. To consider the uncertainty in the SIR model, the Bayesian approach can be employed, and this approach allows us to evaluate the intervention effects by time-varying functions of the infection rate in the SIR model. In this study, we describe the time trend of the spread of COVID-19 in South Korea and investigate the intervention effects for the events using the stochastic SIR model based on the Bayesian approach.

Keywords: COVID-19, intervention effect, ODE, SIR model

1. 서론

2019년말 발생한 코로나19 바이러스가 전 세계적으로 확산되어 2021년 1월 현재 총 9,500만여 명이 넘는 확진자 수와 200만여 명이 넘는 사망자 수를 기록하고 있다. 이와 같은 코로나19의 유행은 전 세계의 정치, 경제, 문화 등의 패러다임을 바꿀 정도로 인간 사회에 미치는 영향이 지대하고, 한국 역시 이 감염병에 대해 예외가 아니다.

코로나19 바이러스와 같은 감염병의 선제적인 방역대책 수립에 있어 감염률의 추정은 매우 중요하며 이를 통해 질병의 확산 정도를 가늠할 수 있다. 감염률의 추정은 감염병 모형의 가정 하에서 가능한데 이를 위해 Kermack과 McKendrick (1927)은 감염률과 제거율(혹은 회복률)의 두 개의 모수를 갖는 확정적 SIR 모형(deterministic sir model)을 제안하였다. SIR 모형은 기본적으로 각 개체가 특정한 감염병에 대해 몇 가지 상태를 거칠 수 있으며, 각 상태들 사이의 상호작용으로 인해 질병의 확산과 소멸의 정도가 결정될 수 있다는 가정 하에 성립한다. 구체적으로 SIR 모형은 전체 인구를 아직 감염되지 않은 상태인 그 위험성이

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2018R1D1A1B07049818).

¹ Corresponding author: Department of Statistics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea. E-mail: jaejik@skku.edu

있는 집단 S (susceptible), 이미 감염되어 다른 개체들에게 전염시킬 수 있는 집단 I (infected), 그리고 증세가 호전되어 면역력이 생기거나 혹은 사망하여 더 이상 다른 개체들과 상호작용이 이루어질 수 없는 집단 R (removed 또는 recovered)로 분류한다. 그리고 각 그룹에 속하는 인구수의 동태적인 변화를 상미분방정식 (ordinary differential equation)을 이용해 묘사하고, 그 해를 구하여 감염률과 제거율을 계산한다. SIR 모형은 다양한 상태를 갖는 감염병에 적용될 수 있다. 그 중 자주 쓰이는 모형 중 SEIR과 SIRS 모형이 있다. SEIR은 바이러스에 노출되어 감염되었지만 겉으로 증세가 나타나지 않고 있는 상태인 E (exposed)가 추가된 모형이고, SIRS는 한 번 회복되었지만 다시 감염될 수 있는 전염병을 분석하기 위한 모형으로 R 그룹에 속한 개체가 다시 S 그룹으로 이동한다는 가정이 추가된 모형이다.

한편 확정적 SIR 모형에 불확실성을 반영하기 위해 확률적 SIR 모형(stochastic sir model)이 개발되었다. 확률적 SIR 모형은 집단들의 동태적인 변화가 어떻게 측정되는지에 따라 연속적(continuous)인 경우와 이산적(discrete)인 경우로 나뉘는데, 후자의 경우 Bailey (1975)는 현재 시점의 감염자 수는 직전 시점의 감염 정보에 영향을 받는다고 가정하는 연쇄-이항 모형(chain-binomial model)을 제안하였다. 이것이 반영된 SIR 모형의 분석은 첫째, 전염병 자료가 정부나 특정한 조사 기관 등에 의해 이산적으로 수집될 수밖에 없는 점과, 둘째, 마코프 연쇄(Markov chain)와 유사한 단순화된 가정으로 인해 가능도 함수(likelihood function)를 도출하기 쉽다는 점에서 현실적으로 용이하다. 그리고 최근에는 컴퓨터의 발달과 더불어 베이지안(Bayesian) 방법을 통한 마코프 연쇄 몬테 카를로(Markov-chain Monte Carlo; MCMC) 기법을 도입해 이들 모형을 적합하려는 시도들이 이어져 왔다. Lekone과 Finkenstadt (2006)는 1995년 콩고민주공화국에서 발생한 에볼라바이러스의 전파 양상을, Kucharski 등 (2016)은 프랑스에서 유행한 지카바이러스를 확률적 SEIR 모형으로 수립하고 감염률 등의 모수를 MCMC를 통해 추정하였다. 또한 최근의 코로나19와 관련해서는 He 등 (2020)이 중국의 봉쇄 정책이 코로나19 확산에 얼마나 영향을 미쳤는가를 확률적 SEIR 모형과 MCMC를 이용하여 분석하였다. 이러한 베이지안 접근법은 자료가 갖는 불확실성을 고려하여 감염률을 추정하는데 이점을 갖는다.

본 연구에서는 베이지안 기법에 기반한 SIR 모형을 이용하여 한국에서 확산되고 있는 코로나19의 전파 양상을 전국과 코로나19가 심각했던 주요 5개 지역에 집중하여 분석한다. 또한, 코로나19가 진행되는 동안 있었던 집단 감염 사태나 정부의 대응 정책과 같은 사건들이 감염률의 증가 또는 감소에 얼마나 기여했는가를 통계적으로 검증해보고자 한다.

본 논문의 구성은 다음과 같다: 2절에서는 외부 개입효과를 고려한 확률적 SIR 모형을 소개하고, 3절에서는 대한민국의 전국 및 각 지역 단위의 코로나19 전파 양상을 베이지안 기법에 기반한 확률적 SIR 모형에 적합하고, 그러한 사건들의 개입효과를 베이지안 관점에서 설명한다.

2. 모형

2.1. 확률적 SIR 모형

본 절에서는 확정적 SIR 모형과 확률적 SIR 모형을 소개하고, 확률적 SIR 모형 하에서 개입효과를 고려하기 위한 시간변이함수(time-varying function)로 표현되는 감염률을 소개한다. 먼저 확정적 SIR 모형은 감염병의 확산과 소멸을 설명하는 수학적 모형으로 다음과 같은 가정 하에서 성립한다:

첫째, 전체 인구는 개별적으로 반드시 S (susceptible), I (infected), 또는 R (removed) 상태 중 하나에 속해야 한다. 둘째, 감염병 이전의 전체 인구는 모두 S 상태에 있다고 가정하고, I 와 R 의 순서로 이동할 가능성이 존재한다. 셋째, 전체 인구수는 고정되어있고, 개개의 구성원들은 각 상태로 변할 가능성이 동일하다. 이 가정들을



Figure 1: States in SIR model.

기반으로 다음과 같은 상미분방정식을 세울 수 있고, 이 방정식이 확정적 SIR 모형이다.

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta \frac{S(t)}{N} I(t), \\ \frac{dI(t)}{dt} &= \beta \frac{S(t)}{N} I(t) - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t).\end{aligned}\quad (2.1)$$

여기서 $S(t)$, $I(t)$, $R(t)$ 는 시점 t 에서 아직 감염되지 않은 개체수, 감염된 개체수, 회복 또는 사망한 개체수를 각각 나타내고, N 은 관찰의 대상이 되는 전체 모집단의 개체수를 의미한다. 또한, β , $\gamma > 0$ 는 각각 감염률과 제거율을 의미하고, $1/\beta$ 과 $1/\gamma$ 은 각각 한 개체가 감염까지 걸리는 시간과 회복 또는 사망까지 걸리는 시간을 각각 의미한다. β/γ 를 기초감염 재생산수(basic reproduction number)라고 하는데 감염병의 심각성 여부를 판단할 때 사용하는 중요한 지표 중 하나이다. 만일 이 값이 1보다 크면 감염병이 빠르게 확산되고 있음을 의미하고, 1보다 작으면 감염병이 빠르게 진정되는 양상을 보인다고 할 수 있다.

확률적 SIR 모형은 식 (2.1)의 확정적 SIR 모형에 각 상태의 개체수를 확률변수로 고려함으로써 임의성을 부여한다. Mode와 Sleeman (2000)은 충분히 작은 시간의 구간 내에서 상태가 변하는 개체의 수가 이산확률분포(discrete probability distribution)를 따른다고 가정하는 SIR 모형을 제안하였다. 구체적인 모형을 설명하기 위해 먼저 단위시간 h 에 대해 각 상태의 개체 수 변화량을 다음과 같이 정의한다.

$$\begin{aligned}S(t+h) - S(t) &= -B(t), \\ I(t+h) - I(t) &= B(t) - C(t), \\ R(t+h) - R(t) &= C(t).\end{aligned}\quad (2.2)$$

식 (2.2)에서 $B(t)$ 는 시점 t 에서 단위시간 h 동안 상태 S 에서 상태 I 로 이동한 개체수를 나타내고, $C(t)$ 는 상태 I 에서 상태 R 로 이동한 개체수를 나타낸다. 개별 개체는 서로 독립이고 개별 개체가 상태 S 와 I 에서 I 와 R 로 각각 이동할 확률이 동일하다고 가정하면, $B(t)$ 와 $C(t)$ 는 각각 다음과 같은 이항분포(binomial distribution)를 따르는 확률변수로 정의될 수 있다.

$$\begin{aligned}B(t) &\sim \text{Bin}(S(t), p_\beta), \\ C(t) &\sim \text{Bin}(I(t), p_\gamma).\end{aligned}\quad (2.3)$$

여기서 이항분포의 모수 p_β 와 p_γ 는 한 개체가 상태 S 와 I 에서 I 와 R 로 각각 이동하는데 걸리는 시간이 지수 분포를 따른다는 가정 하에서 $p_\beta = 1 - \exp(-\beta I(t)/Nh)$ 와 $p_\gamma = 1 - \exp(-\gamma h)$ 로 도출될 수 있다. 단위시간 h 가 충분히 작을 때, 테일러 전개(Taylor expansion)에 의해 $p_\beta \approx \beta I(t)/Nh$ 가 되고 $p_\gamma \approx \gamma h$ 가 된다. 이때 두 확률변수 $B(t)$ 와 $C(t)$ 의 기댓값은 각각 $E[B(t)] \approx \beta S(t)I(t)/Nh$ 와 $E[C(t)] \approx \gamma I(t)h$ 가 되고, h 에 대한 두 기댓값의 기울기는 식 (2.1)의 $dS(t)/dt$ 와 $dR(t)/dt$ 와 각각 일치한다.

일반적인 SIR 모형에서는 감염률 β 와 제거율 γ 가 시간에 따라 변하지 않는 고정된 상수로 가정한다. 대부분의 감염병 모형에서 고정된 제거율을 가정하는 것은 감염병에서 회복 또는 사망하는 것이 시기와 상황에 따라 크게 변하는 것이 아니므로 자연스러운 가정이다. 그러나, 코로나19와 같이 장기간 지속되는 감염병으로

부터 나온 자료에 고정된 감염률을 갖는 SIR 모형을 적용하면 감염병의 전파 양상을 표현하는데 적절치 않을 수 있다. 그 이유는 감염병 초기의 감염률이 감염병의 전파가 진행되는 전체 기간 동안 일정하지 않은 것이 일반적이기 때문이다. 예를 들어 정부의 적절한 방역정책에 의해 감염율이 떨어질 수도 있고, 반대로 예상치 못한 사건으로 인해 감염병의 확산이 가속화될 수도 있다. 이렇듯 장기간 유행하는 감염병에 대해 시간에 따라 변하는 감염률을 가정하는 것은 일반적으로 주어진 자료를 더 잘 설명할 가능성이 높다. 이 경우 감염률은 시간의 함수인 시간변이 감염률 함수(time-varying infection rate function) $\beta(t)$ 로 표현될 수 있다.

이러한 시간변이 감염률 함수는 감염병 기간 동안 정부의 방역정책이나 어떠한 사건들이 감염병의 전파에 미친 영향을 분석하는 개입효과 분석(intervention effect analysis)에 활용될 수 있다. 사건들의 개입효과 분석을 위해 본 연구에서는 일정한 형태의 감염률이 사건의 전후로 변하는 양상을 나타내는 계단함수(step function) 형태의 시간변이 감염률 함수를 가정한다.

계단 형태의 시간변이 감염률 함수를 정의하기 위해 우선 전체 감염병의 기간 동안 감염률에 영향을 미칠 만한 K 개의 사건들이 있었다고 가정하자. 또한, 분석의 대상이 되는 감염병 기간을 $[0, T]$ 라 하고, 그 사건들의 발생시간을 t_1, \dots, t_K ($0 = t_0 < t_1 \leq \dots \leq t_K < t_{K+1} = T$)라고 하자. 이때 계단 형태의 시간변이 감염률 함수는 다음과 같이 정의된다.

$$\beta(t) = \sum_{k=0}^K \beta_k I(t_k \leq t \leq t_{k+1}), \quad 0 < t < T, \quad (2.4)$$

여기서 $\beta_k > 0$ 는 k 번째 사건 이후의 감염률이고, $I(\cdot)$ 은 지시함수(indicator function)이다. k 번째 사건이 감염률의 변화에 영향을 미쳤는가 여부는 β_{k-1} 과 β_k 를 비교함으로써 가능하다.

2.2. 베이지안 추정 방법

본 연구에서는 계단형 시간변이 감염률 함수를 갖는 확률적 SIR 모형을 고려한다. 이 모형의 추정을 위해 자료의 불확실성(uncertainty)을 고려하기 용이한 베이지안(Bayesian) 방법을 사용한다.

시점 t 에서 상태 S 인 개체수와 상태 I 인 개체수를 각각 $S(t) = s_t$ 와 $I(t) = i_t$, $t = 1, \dots, T$ 라고 하자. 또한, 상태 S 에서 I 로 이동한 개체수 (즉, 시점 t 에 감염병에 걸린 개체수) $B(t) = b_t$, 상태 I 에서 R 로 이동한 개체수 (즉, 시점 t 에 회복 또는 사망한 개체수) $C(t) = c_t$, $t = 1, \dots, T$ 가 t 시점에서 관찰된다고 하자. 식 (2.2)의 확률적 SIR모형에서 확률변수 $B(t)$ 와 $C(t)$ 는 이항분포를 따르고 서로 독립이라고 가정된다. 식 (2.4)의 계단형 감염률 함수에 대해 모수벡터는 $\theta = (\beta_0, \beta_1, \dots, \beta_K, \gamma)^T$ 가 되고, 이때 가능도 함수(likelihood function)는 다음과 같이 정의될 수 있다.

$$L(\theta) = \prod_{t=1}^T \left[\binom{s_t}{b_t} p_{\beta(t)}^{b_t} (1 - p_{\beta(t)})^{s_t - b_t} \binom{i_t}{c_t} p_{\gamma}^{c_t} (1 - p_{\gamma})^{i_t - c_t} \right], \quad (2.5)$$

여기서 $p_{\beta(t)}$ 는 식 (2.4)의 계단형 감염률 함수 하에서 개별 개체가 상태 S 에서 I 로 이동할 확률을 의미하고, p_{γ} 는 개별 개체가 상태 I 에서 R 로 이동할 확률로 전체 시간에 대해 고정이다.

시간변이 감염률 함수 $\beta(t)$ 가 계단 함수의 형태일 때, 모수 β_k , $k = 0, 1, \dots, K$ 의 사전분포로서 Dehning 등 (2020)은 감염률이 양인 조건을 만족하는 로그-정규분포(log-normal distribution)를 사용하였고, 제거율 γ 의 사전분포 역시 로그-정규분포를 고려하였다. 따라서, 본 연구에서도 β_k , $k = 0, 1, \dots, K$ 와 γ 의 사전분포로서 로그-정규분포를 사용하고, 계단형 시간-변이 감염률 함수를 갖는 확률적 SIR 모형의 사후분포(posterior

distribution)는 다음과 같이 표현될 수 있다.

$$\begin{aligned}
 p(\boldsymbol{\theta}|\mathbf{b}, \mathbf{c}) &\propto L(\boldsymbol{\theta})\pi(\boldsymbol{\beta})\pi(\gamma), \\
 &\propto \prod_{t=1}^T \left[\binom{s_t}{b_t} p_{\beta(t)}^{b_t} (1 - p_{\beta(t)})^{s_t - b_t} \binom{i_t}{c_t} p_{\gamma}^{c_t} (1 - p_{\gamma})^{i_t - c_t} \right] \left[\prod_{k=0}^K \pi(\beta_k) \right] \pi(\gamma), \\
 &\propto \prod_{t=1}^T \left[p_{\beta(t)}^{b_t} (1 - p_{\beta(t)})^{s_t - b_t} p_{\gamma}^{c_t} (1 - p_{\gamma})^{i_t - c_t} \right] \left[\prod_{k=0}^K \pi(\beta_k) \right] \pi(\gamma), \tag{2.6}
 \end{aligned}$$

여기서 $\mathbf{b} = (b_1, \dots, b_T)^\top$, $\mathbf{c} = (c_1, \dots, c_T)^\top$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K)^\top$ 이다. $\pi(\beta_k)$, $k = 0, 1, \dots, K$ 는 k 번째 사건 이후의 감염률 β_k 의 사전분포이고 모수 μ_k 와 σ^2 를 갖는 로그-정규분포이다 (즉, $\beta_k \sim LN(\mu_k, \sigma^2)$). 또한, $\beta_0, \beta_1, \dots, \beta_K$ 들은 독립으로 가정한다. 마찬가지로 $\pi(\gamma)$ 는 제거율 γ 의 사전분포로 모수가 τ 와 δ^2 인 로그-정규분포이다 (즉, $\gamma \sim LN(\tau, \delta^2)$). Dehning 등 (2020)은 감염률과 제거율이 주어진 자료에 의존적일 수 있도록 로그-정규 사전분포의 분산을 충분히 크게 줄 것을 제안하였다. 이에 따라 본 연구에서도 충분히 큰 값의 분산을 갖는 감염률과 제거율의 사전분포를 고려한다. 식 (2.6)의 사후분포를 구하기 위해 MCMC 방법의 하나인 Metropolis-Hastings 알고리즘을 이용한다.

2.3. 모형선택과 개입효과의 유효성 판단

연구의 관심이 되는 감염병의 기간동안 감염률에 영향을 미치는 외부 개입에 해당하는 사건들이 K 개 있다고 가정하자. 주어진 자료에 대해 이 K 개의 사건들의 전체 혹은 부분집합이 식 (2.4)의 시간변이 감염률 함수에 적용되어 확률적 SIR 모형을 적합한다면 우리는 총 2^K 개의 후보 모형들(candidate models)을 고려할 수 있다. 이 2^K 개의 후보 모형 중 주어진 자료에 대해 최소의 모수로 감염병의 전파 양상을 가장 잘 설명할 수 있는 모형이 최적의 모형(optimal model)이라 할 수 있다. 그러한 최적의 모형을 찾기 위해서는 모형선택의 기준이 필요한데 본 연구에서는 deviance information criterion (DIC)를 사용한다. DIC는 다음과 같이 정의된다:

$$DIC = D(E[\boldsymbol{\theta}]) + 2\lambda. \tag{2.7}$$

여기서 $D(\boldsymbol{\theta}) = -2\log L(\boldsymbol{\theta})$ 이고, $\lambda = E[D(\boldsymbol{\theta})] - D(E[\boldsymbol{\theta}])$ 이다. DIC 값이 작을수록 주어진 자료에 대해 더 좋은 모형으로 판단된다. MCMC 방법으로 부터 추출된 표본을 이용하면 식 (2.7)의 기댓값 계산이 용이하므로 베이지안 모형의 비교에 있어 DIC가 자주 사용된다.

그러나, DIC를 기준으로 최적의 모형을 찾기 위해 2^K 개의 후보 모형 모두를 적합하는 것은 계산적으로 비효율적이며, 만일 K 가 큰 경우 계산적으로 실행 불가능하다. 따라서, 본 연구에서는 효율적으로 DIC가 가장 작은 최적의 모형을 찾기 위해 다음과 같은 전진 선택법(forward selection method)을 제안한다:

1. 외부 개입 사건들의 인덱스(index)들의 집합 $\mathcal{A} = \{1, \dots, K\}$ 로 놓는다.
2. 외부 개입 사건들의 초기 인덱스들의 집합 \mathcal{A}_0 를 공집합으로 놓고, 주어진 자료에 대해 외부 개입의 사건이 전혀 없는 모형 \mathcal{M}_0 를 적합한다 (고정된 감염률 β 가정).
3. $j = 1, \dots, K$ 에 대해,
 - 3-1. 여집합 \mathcal{A}_{j-1}^c 에서 사건 인덱스를 하나씩 \mathcal{A}_{j-1} 에 추가하여 모형을 적합하고 각 모형에 대해 DIC 값을 계산한다.
 - 3-2. 3-1로 부터 구한 \mathcal{A}_{j-1}^c 의 원소의 갯수 만큼의 모형 중 가장 작은 DIC 값을 갖는 모형을 \mathcal{M}_j 라 하고, 그 모형의 사건 인덱스 집합을 \mathcal{A}_j 로 놓는다.

4. M_0, \dots, M_k 의 모형 중 가장 작은 DIC 값을 갖는 모형을 최적의 모형으로 판단한다.

만일 외부 개입에 해당하는 어떤 사건이 DIC 기준에서 최적의 모형에 포함되지 않았다고 가정하자. 이 경우 그 사건이 감염률에 미치는 영향은 DIC 기준에서 최적의 모형에 포함될 만큼의 효과가 분명하지 않았다고 간주할 수 있다. 이는 그 사건의 감염률에 대한 개입 효과가 유효한지 여부를 판단하는 간접적인 기준이 될 수 있다. 또 다른 방법으로 사건 전후의 감염률의 유효한 변화를 β_{k-1} 과 β_k 의 credible 구간들을 비교함으로써 판단할 수 있다. Credible 구간은 사후분포를 기반으로 모수가 그 안에 있을 것이라고 믿어지는 구간을 의미한다. 만일 β_{k-1} 과 β_k 의 $(1 - \alpha)(100)\%$ credible 구간들이 서로 겹치지 않는다면 k 번째 사건의 외부 개입효과가 존재한다고 믿을 수 있다.

3. 코로나19 자료분석

3.1. 한국 코로나19 자료 소개

한국에서는 2020년 1월 20일에 코로나19 확진자가 처음 발생하였고 이에 질병관리본부(현 질병관리청)에서 중앙방역대책본부가 가동되기 시작했다. 이후 코로나19 관련 자료 수집은 각 지방자치단체에서 이루어졌으며, 중앙방역대책본부는 각 지역에서 수집된 자료를 종합하여 매일 정례 브리핑을 실시하였다. 중앙방역대책본부에서 발표하는 코로나19 국내 발생동향 자료는 매일 0시 기준으로 집계된 신규 확진자 수, 격리해제자 수, 사망자 수, 그리고 검사자 수가 지역별, 감염 유형별 등으로 구분되어 제공된다. 본 연구에서는 2020년 1월 20일부터 2020년 10월 20일까지 9개월여 간 전국에서 집계된 일일 확진자 수, 격리해제자 수, 사망자 수를 분석 대상으로 고려한다. 다만, 본 연구는 국내에서 발생한 코로나19 전파와 외부 개입 효과에 관심이 있으므로 분석 대상에서 해외유입 확진자와 관련한 정보는 제외하였다. 또한 크게 전국 단위 분석과 5개 주요 지역들에 대한 분석으로 나뉜다. 5개 주요 지역들은 위 분석 기간 동안 코로나19 확산이 심했었던 서울, 경기, 인천, 대구, 경북을 대상으로 하였다.

첫 확진자가 발생한 2020년 1월 20일을 $t = 1$ 로 설정하였고, 분석 기간의 마지막 날에 해당하는 2020년 10월 20일은 $t = 275$ 이다. 시점 t 에서의 일일 확진자 수는 $B(t) = b_t$ 가 되고, 일일 격리해제자 수와 사망자 수의 합을 $C(t) = c_t$ 로 정의하였다. $S(t) = s_t$ 와 $I(t) = i_t$, $t = 1, \dots, 275$ 는 식 (2.2)에 의해 계산될 수 있고, 초기 시점에서의 분석 단위별 인구수 $S(0) = s_0$ 는 2020년 1월 20일에 시기적으로 가장 가깝게 이용 가능한 2020년 6월 행정안전부에서 발표한 시도별 주민등록 인구수를 기준으로 하였다.

분석의 기간 중 감염률에 영향을 미칠만한 외부 개입에 해당하는 사건으로 Table 1에 제시된 5개의 사건을 고려하였다. ‘신천지 교회’, ‘이태원 클럽’, ‘사랑제일 교회’는 집단감염사태로 코로나19의 확산을 가속화하여 감염률이 사건 이전보다 증가했을 것으로 추측되고, 반대로 ‘사회적 거리두기’와 같은 정부의 방역정책은 감염률을 감소시킨 외부 개입으로 추측될 수 있다. 본 연구에서는 앞서 언급한 집단감염 사태들이 감염률을 유의미하게 증가시킨 사건들이었는지 그리고 사회적 거리두기 정책이 방역대책으로 감염률을 감소시키는데 효과적이었는지에 집중하여 분석을 진행하였다. 또한, 이러한 집단감염 사태와 방역대책이 5개 주요 지역에서 감염률의 변화를 유발시켰는지 아니면 어느 특정 지역에 유효한 영향을 미쳤는지 여부 역시 분석하였다.

3.2. 한국 코로나19 자료 분석

먼저 전국 단위 자료를 계단형 시간변이 감염률 함수를 갖는 확률적 SIR 모형들에 적용하였다. 2.3절에서 설명한 전진 선택법을 이용하여 구한 모형 M_0, \dots, M_5 에 대한 각 모수들의 사후평균(posterior mean)과 95%

Table 1: External intervention events

k	Date	t	Type	Event
1	2020/02/18	30	Mass infection	Shincheonji church
2	2020/03/22	63	Government policy	Social distance
3	2020/05/06	108	Mass infection	Itaewon club
4	2020/08/03	197	Mass infection	Sarangjeil presbyterian church
5	2020/08/30	224	Government policy	Social distance (2.5 stage)

Table 2: Posterior means, credible intervals, and DIC for national data

\mathcal{M}	β_0	β_1	β_2	β_3	β_4	β_5	γ	DIC
\mathcal{M}_0	0.046 (0.046,0.047)						0.045 (0.044,0.046)	38456.3
\mathcal{M}_1	0.039 (0.038,0.040)			0.061 (0.060,0.062)		0.045 (0.044,0.046)		37311.1
\mathcal{M}_2	0.063 (0.062,0.064)		0.022 (0.021,0.023)		0.061 (0.060,0.062)		0.045 (0.044,0.046)	33847.6
\mathcal{M}_3	0.063 (0.062,0.064)		0.022 (0.021,0.023)		0.126 (0.122,0.130)	0.042 (0.041,0.043)	0.045 (0.044,0.046)	30898.8
\mathcal{M}_4	0.063 (0.062,0.064)		0.012 (0.012,0.013)	0.041 (0.040,0.042)	0.126 (0.122,0.130)	0.042 (0.041,0.043)	0.045 (0.044,0.046)	29333.0
\mathcal{M}_5	0.085 (0.059,0.113)	0.063 (0.062,0.064)	0.012 (0.012,0.013)	0.041 (0.039,0.042)	0.126 (0.122,0.129)	0.042 (0.041,0.043)	0.045 (0.044,0.046)	29333.6

* (., .): 95% credible interval.

credible 구간, DIC 값들이 Table 2에 나타나 있다.

Table 2를 보면 모형 \mathcal{M}_1 ($DIC = 37311.1$)과 ‘사회적 거리두기’를 외부 개입으로 넣은 모형 \mathcal{M}_2 ($DIC = 33847.6$)의 DIC 값 차이가 가장 컸다. 즉, ‘사회적 거리두기’를 모형에 추가했을 때 DIC 값이 가장 크게 감소하였다. 이는 정부의 사회적 거리두기 방역대책이 시행 이전의 높은 감염률을 낮추는데 효과가 있었음을 의미한다. 두 번째로 DIC 값이 크게 감소한 부분은 \mathcal{M}_2 에 ‘사회적 거리두기(2.5단계)’의 외부 개입을 추가한 \mathcal{M}_3 였다. 이 또한, 2020년 8월 30일 이전 상대적으로 높았던 감염률이 ‘사회적 거리두기(2.5단계)’ 시행 이후 크게 낮아졌음을 의미한다.

가장 DIC 값이 낮은 모형은 \mathcal{M}_4 로 ‘신천지 교회’ 집단감염을 제외한 모든 외부 개입이 들어간 모형이었다. 이는 DIC의 기준에서 ‘이태원 클럽’과 ‘사랑제일교회’ 집단감염 사태가 최적의 모형에 반영될 정도로 효과가 있었던 반면, ‘신천지 교회’ 집단감염은 최적의 모형에 반영될 정도는 아닌 것으로 보인다. 그러나, ‘신천지 교회’ 사건이 감염병의 초기에 일어난 사건임을 상기할 필요가 있다. 감염병 초기에는 감염자의 수가 상대적으로 적기 때문에 고정된 감염률 하에서도 전체 감염자 수의 증가 속도가 느려보인다. 그러나 감염자 수가 증가할수록 똑같은 감염률이라 하더라도 전체 감염자 수는 급격하게 증가하게 된다. 이러한 감염병의 특성으로 인해 코로나19의 초기 사건인 ‘신천지 교회’ 집단감염은 유효한 외부 개입 효과로 보이지 않는 것으로 판단된다.

Figure 2는 Table 2에서 DIC 값이 가장 작았던 모형 \mathcal{M}_4 의 적합선, 실제 전국 단위 자료, 추정된 계단형 시간변이 감염률 함수를 나타낸 그림표이다. 더불어, MCMC를 통해 구한 사후분포의 표본들의 2.5%와 97.5% 표본 분위수(sample quantile)를 이용하여 구한 적합선과 추정된 감염률 함수에 대한 95% credible band들을

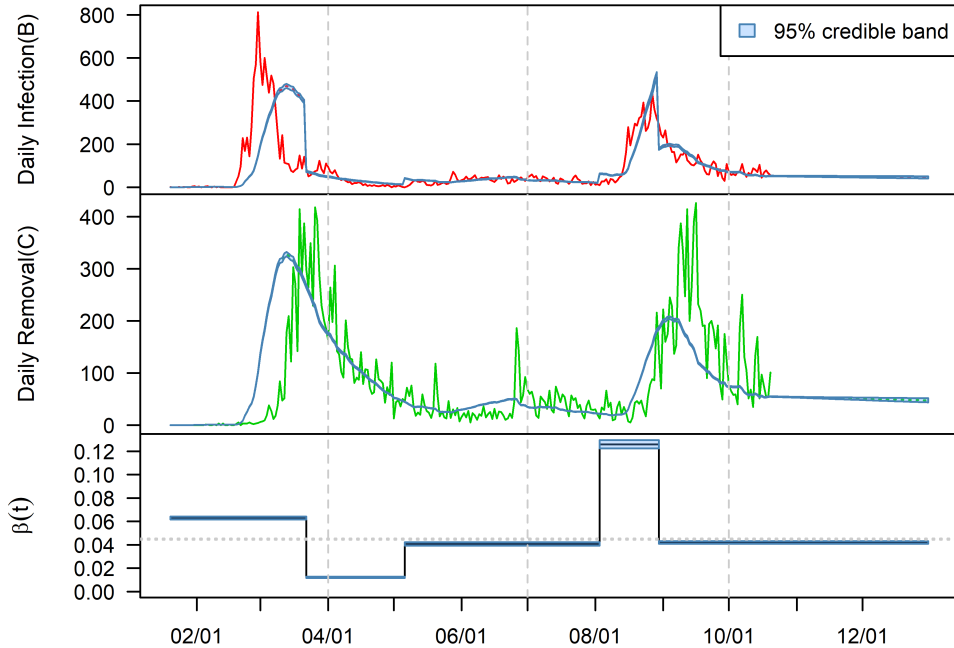


Figure 2: Fitted line and estimated infection rate function of model M_4 for national data; upper panel: Daily confirmed cases and fitted line (red: daily confirmed cases); middle panel: Daily isolation release cases and deaths and fitted line (green: daily isolation release cases and deaths); lower panel: Estimated infection rate.

보여준다. 전국 단위 자료로 전체 인구수 N 이 매우 크기 때문에 95% credible band의 폭이 매우 좁음을 알 수 있다. Figure 2를 통해 일일 확진자 수와 일일 격리해제자 및 사망자 수의 양상을 모형 M_4 가 잘 설명하고 있는 것으로 보인다. 계단형 감염률 함수를 보면 ‘신천지 교회’ 사건을 제외한 나머지 4개의 사건에서 명백하게 감염률이 변했음을 알 수 있고 그 4개의 변화하는 지점에서 95% credible band가 겹치지 않고 구분됨을 알 수 있다. 이를 통해 우리는 2020년 3월 20일에 시행된 ‘사회적 거리두기’가 감염률을 낮추는 외부 개입 효과가 있었으며, ‘이태원 클럽’과 ‘사랑제일 교회’ 집단감염이 감염률 상승에 영향을 미쳤음을 알 수 있다. 특히 2020년 8월 30일에 시행된 ‘사회적 거리두기(2.5단계)’는 ‘사랑제일 교회’ 집단감염으로 급격히 상승된 감염률을 거의 ‘사랑제일 교회’ 사건 이전 수준으로 낮출 만큼 큰 효과가 있었던 것으로 보인다.

추가적으로 Figure 2는 분석 기간인 2020년 10월 20일 이후의 예측선 또한 보여준다. 이 예측선은 8월 30일에서 10월 20일까지의 감염률인 β_5 의 추정값이 10월 20일 이후에도 유지된다는 가정 하에서 구한 일일 확진자 수와 일일 격리해제자 및 사망자 수이다. 만일 10월 20일까지의 감염률이 그 이후에도 유지된다면 일일 확진자 수가 서서히 감소하는 것으로 예측된다는 것을 알 수 있다. 그러나, 2021년 1월 현재 코로나 바이러스의 전파가 용이해지는 조건인 춥고 건조한 날씨로 인해 확진자 수가 다시 증가하는 양상을 보이고 있다.

본 연구에서는 코로나19가 상대적으로 심각했던 서울, 경기, 인천, 대구, 경상북도의 5개 지역을 선정하여 확률적 SIR 모형을 적용해 보았다. 각 지역별로 DIC 기준으로 전진 선택법을 이용하여 최적의 모형을 선정할 결과를 Table 3이 보여주고 있다. Table 3으로 부터 서울과 경기도가 비슷한 감염률 양상을 보여주고 있고, 이 양상은 전국 단위 감염률의 변화와 매우 유사하다는 것을 알 수 있다. 서울, 경기 수도권에 인구가 집중되어 있는 탓에 전국 단위의 감염률의 변화가 수도권의 변화에 의존하고 있는 듯한 인상을 준다. Figure 3과 Figure

Table 3: Posterior means and credible intervals of the optimal model for each region

Region	\mathcal{M}	β_0	β_1	β_2	β_3	β_4	β_5	γ
Seoul	\mathcal{M}_4	0.093 (0.083,0.102)		0.024 (0.021,0.026)	0.045 (0.042,0.048)	0.125 (0.120,0.131)	0.035 (0.034,0.037)	0.048 (0.046,0.049)
Kyeonggi	\mathcal{M}_4	0.095 (0.084,0.106)		0.030 (0.027,0.033)	0.047 (0.044,0.050)	0.117 (0.111,0.123)	0.049 (0.047,0.052)	0.055 (0.053,0.057)
Incheon	\mathcal{M}_5	0.087 (0.037,0.174)	0.118 (0.087,0.152)	0.035 (0.027,0.044)	0.045 (0.040,0.050)	0.169 (0.151,0.188)	0.040 (0.035,0.045)	0.052 (0.049,0.055)
Daegu	\mathcal{M}_3	0.061 (0.060,0.063)		0.007 (0.006,0.007)	0.011 (0.009,0.012)	0.066 (0.057,0.075)		0.038 (0.037,0.039)
Gyeongbuk	\mathcal{M}_3	0.061 (0.058,0.064)		0.008 (0.007,0.010)	0.018 (0.014,0.023)	0.063 (0.054,0.072)		0.038 (0.036,0.040)

* (· , ·): 95% credible interval.

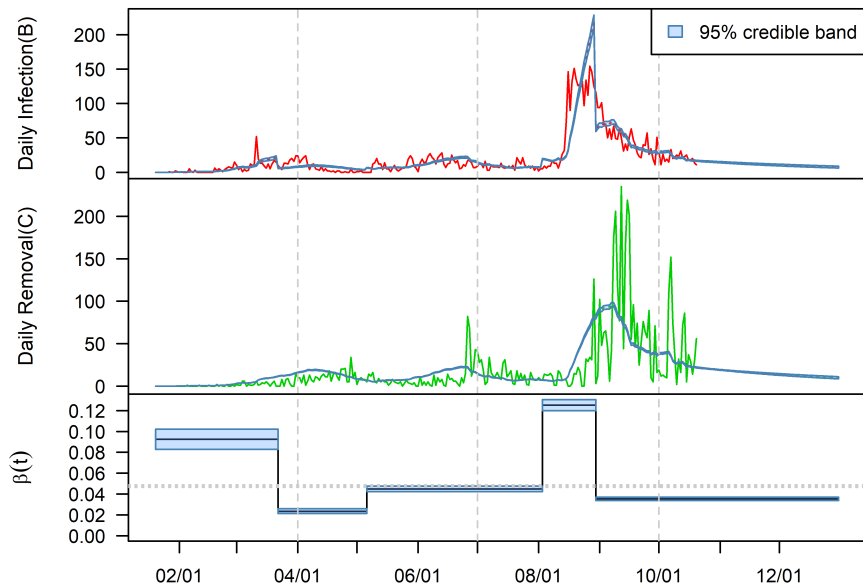


Figure 3: Result of model \mathcal{M}_4 for Seoul.

4는 서울과 경기도 지역의 일일 확진자 수, 일일 격리해제자 및 사망자 수, 추정된 감염률 함수의 추세와 양상을 보여준다. 이 그림표로 부터 서울과 경기도의 일일 확진자 수와 일일 격리해제자 및 사망자 수의 추세가 매우 유사하고, 그에 따른 감염률 함수의 양상 또한 매우 비슷하다는 것을 알 수 있다. Figure 5는 인천 지역의 코로나19 추세를 보여주고 있고, 인천 역시 서울, 경기와 비슷한 추세와 감염률 함수를 가지고 있음을 알 수 있다. 다만, DIC 기준으로 인천 지역에 대한 최적의 모형은 ‘신천지 교회’ 사건의 외부 개입 효과를 포함하나 Figure 5의 추정된 감염률 함수를 보면 95% credible band가 ‘신천지 교회’ 사건 전후로 겹쳐져 있음을 알 수 있다. 따라서, 인천 지역 역시 서울, 경기 지역과 비슷한 감염 양상을 보이고 있는 것으로 판단된다.

대구와 경북 지역은 서울, 경기, 인천의 수도권 지역과는 매우 다른 감염 양상을 보인다. 수도권 지역은 초반에 상대적으로 적은 코로나19 확진자가 발생했고, ‘이태원 클럽’과 ‘사랑제일 교회’와 같은 사건 이후

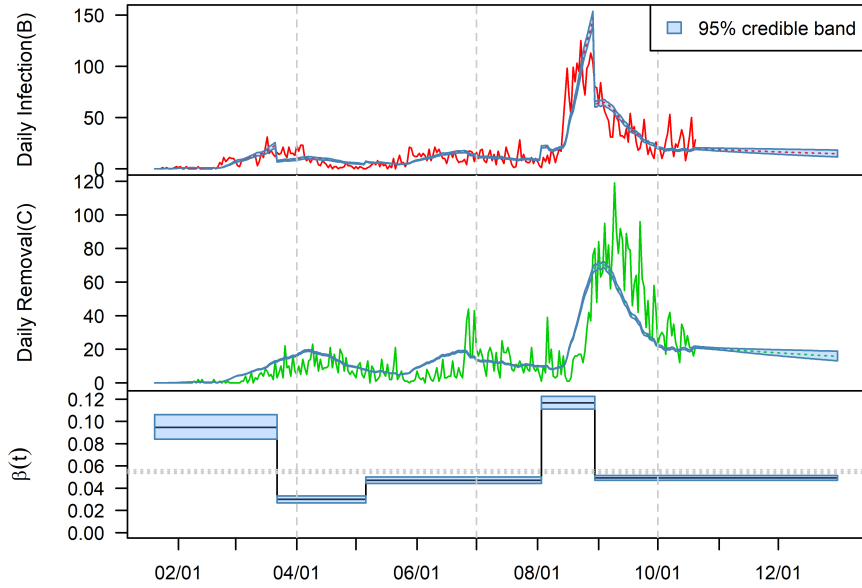


Figure 4: Result of model M_4 for Kyeonggi.

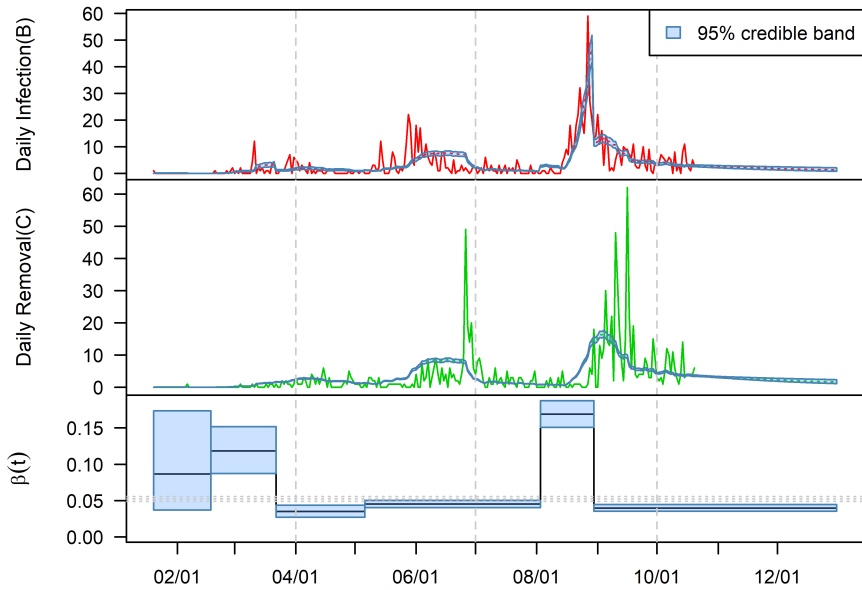


Figure 5: Result of model M_5 for Incheon.

급격히 증가하는 감염 양상을 보인다. 이에 반해 대구와 경북 지역은 코로나19 초기에 확진자 수가 급격히 증가한 이후 ‘사회적 거리두기’에 의해 확진자 수가 감소한 이후 급격한 변화를 보이고 있지 않다. 이는 ‘신천지 교회’ 집단감염이 대구와 경북을 중심으로 발생하는데 기인한 것으로 보인다.

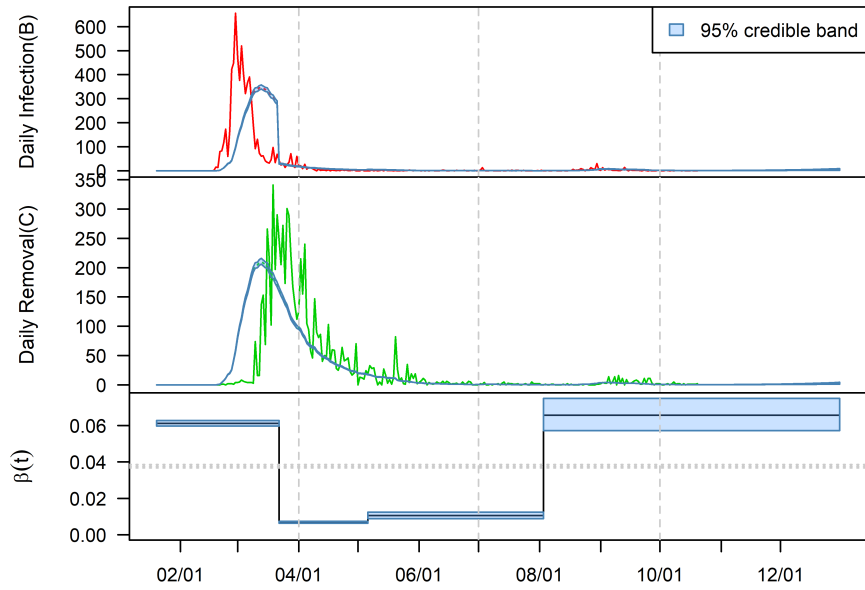


Figure 6: Result of model M_3 for Daegu.

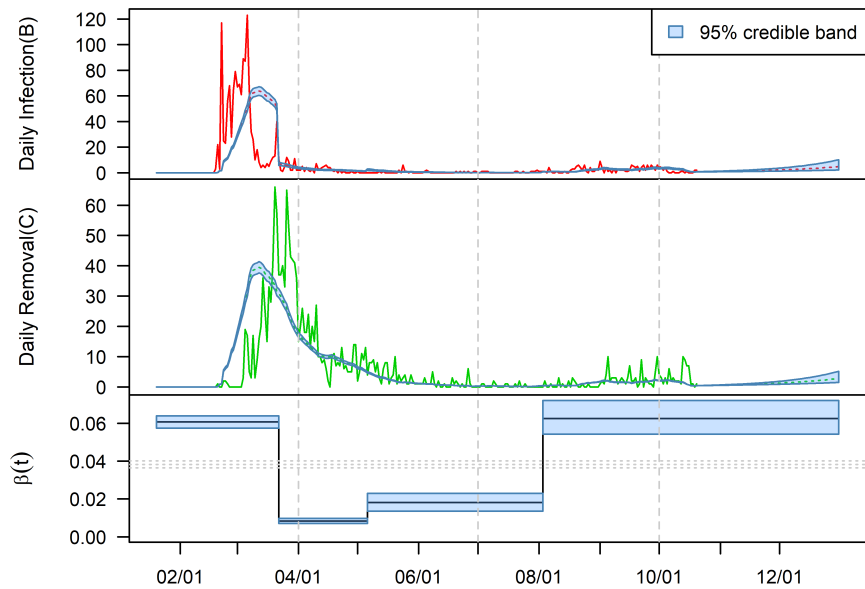


Figure 7: Result of model M_3 for Gyeongbuk.

4. 결론 및 토의

2021년 1월 현재 코로나19 바이러스의 전 세계적 유행이 지속되고 있고 여전히 심각한 상황이다. 이에 본 연구에서는 한국에서 발생한 집단감염과 이에 대한 정부의 방역정책과 같은 사건들이 코로나19의 전파에 유

효한 영향을 미쳤는지 통계적으로 살펴보기 위해 계단형 시간변이 감염률 함수를 가정한 확률적 SIR 모형을 적용하여 분석을 시도하였다. 그 결과 전국 단위의 분석에서 수도권 지역의 집단감염 사태가 유의미한 감염률의 증가를 가져왔고, 정부의 방역대책인 ‘사회적 거리두기’가 감염률을 낮추는데 기여를 하였음을 확인하였다.

그러나, 본 연구의 분석은 몇 가지 한계점을 갖는다. 첫째, 본 연구는 해외로 부터 유입된 코로나19 확진자를 고려하지 않았다. 이는 한국 내에서의 전파 양상에 초점을 맞추기 위해 제외된 것이나, 확진자의 해외 유입으로 인한 감염률 증가를 배제할 근거가 없다는 문제가 존재한다. 둘째, 코로나19는 완치가 되어 격리해제가 되었다 하더라도 일정 시간이 지나면 면역력을 잃어 코로나19에 재감염될 우려가 있다. 이를 고려한 모형이 SIRS 모형이다. 그러나, 재감염자 수에 대한 자료가 없는 관계로 본 연구에서는 재감염에 대한 부분을 고려하지 않았다. 마지막으로 코로나19의 감염률과 전파 속도는 계절적 요인에 대한 의존성이 높다. 일반적으로 코로나 바이러스는 습하고 더운 여름에 확산 속도가 떨어지는 반면, 춥고 건조한 겨울에 확산 속도가 증가하는 양상을 보인다. 그러나, 본 연구에서는 계절적인 요인을 고려하지 않았다. 만일 계절적인 요인까지 포함한 모형을 고려한다면 좀 더 정교한 예측과 분석이 가능할 것이다.

본 연구에서는 외부적인 사건이 감염률에 영향을 미치는 개입 효과에 집중하여 분석을 진행하였고, 이를 위해 계단형 시간변이 감염률 함수를 고려하였다. 그러나, 감염병 기간동안 전반적인 감염률 변화의 패턴을 추정하기 위해서는 계단형 감염률 함수를 고려하기 보다는 평활 스플라인(smoothing spline)과 같은 부드러운 비선형 함수의 형태를 갖는 감염률 함수를 고려하는 것이 예측력을 높이는데 더 유리할 것이다.

References

- Bailey NTJ (1975). *The Mathematical Theory of Infectious Diseases and its Applications*(2nd ed), Charles Griffin & Company Ltd, London.
- Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, and Priesemann V (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions, *Science*, **369**.
- He S, Tang S, and Rong L (2020). A discrete stochastic model of the COVID-19 outbreak: Forecast and control, *Mathematical Biosciences and Engineering*, **17**, 2792–2804.
- Kermack WO and McKendrick AG (1927). A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society A: Mathematical Physical and Engineering Science*, **115**, 700–721.
- Kucharski A, Russell T, Diamond C, Liu Y, Edmunds J, Funk S, and Eggo RM (2020). Early dynamics of transmission and control of COVID-19: A mathematical modelling study, *The Lancet Infectious Diseases*, **20**, 553–558.
- Lekone P and Finkenstadt B (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study, *Biometrics*, **62**, 1170–1177.
- Mode CJ and Sleeman CK (2000). *Stochastic Processes in Epidemiology: HIV/AIDS, Other Infectious Diseases, and Computers*, River Edge, Singapore.

Received January 29, 2021; Revised February 3, 2021; Accepted February 4, 2021

SIR 모형을 이용한 한국의 코로나19 확산에 대한 개입 효과 분석

조수민^a, 김재직^{1,a}

^a성균관대학교 통계학과

요약

코로나19 바이러스는 2020년에 전세계적으로 심각하게 확산되었고, 우리의 일상생활 전체에 상당한 영향을 미치고 있다. 현재 전세계는 이 유행병 사태 아래에 여전히 있고 한국 또한 이 상황에 대해 예외가 아니다. 이 유행병 기간동안 한국에서는 이 바이러스 확산을 방지하기 위한 또는 가속화시킨 몇 가지 사건들이 있었다. 감염병에 대한 방역 정책을 세우기 위해 이러한 사건들의 감염병 확산에 대한 개입 효과를 조사하는 것은 매우 중요하다. SIR 모형은 미분방정식을 통해 감염병 확산의 동적 행태를 파악하기 위해 자주 사용되는 방법이다. 그러나, SIR 모형은 관찰된 데이터의 불확실성을 고려하지 않는 결정적인 모형이다. 따라서 SIR 모형에서 데이터의 불확실성을 고려하기 위해 베이지안 접근법이 사용될 수 있고, 이러한 접근법은 SIR 모형에서 감염률에 대한 시간변이함수에 근거한 개입효과분석을 가능하게 한다. 본 연구에서는 베이지안 접근법에 근거한 확률적 SIR 모형을 이용하여 한국에서의 코로나19 바이러스의 확산 추세를 설명하고 그러한 사건들에 대한 개입효과를 조사한다.

주요용어: 개입효과, 미분방정식, 코로나19, SIR 모형

이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2018R1D1A1B07049818).

¹교신저자: (03063) 서울시 중로구 성균관로 25-2, 성균관대 통계학과, 부교수. E-mail: jaejik@skku.edu