

코로나-19 진행에 따른 SIR 기반 예측모형적용 연구

김 훈* · 조상섭** · 채동우***

Research on Application of SIR-based Prediction Model According to the Progress of COVID-19

Hoon Kim* · Sang Sup Cho** · Dong Woo Chae***

Abstract

Predicting the spread of COVID-19 remains a challenge due to the complexity of the disease and its evolving nature. This study presents an integrated approach using the classic SIR model for infectious diseases, enhanced by the chemical master equation (CME). We employ a Monte Carlo method (SSA) to solve the model, revealing unique aspects of the SARS-CoV-2 virus transmission. The study, a first of its kind in Korea, adopts a step-by-step and complementary approach to model prediction. It starts by analyzing the epidemic's trajectory at local government levels using both basic and stochastic SIR models. These models capture the impact of public health policies on the epidemic's dynamics. Further, the study extends its scope from a single-infected individual model to a more comprehensive model that accounts for multiple infections using the jump SIR prediction model. The practical application of this approach involves applying these layered and complementary SIR models to forecast the course of the COVID-19 epidemic in small to medium-sized local governments, particularly in Gangnam-gu, Seoul. The results from these models are then compared and analyzed.

Keywords : COVID-19 Prediction Based on Scenarios, Deterministic SIR Model, Stochastic SIR Model, Jump SIR Model, Comparative Analysis of Predictive Accuracy

Received : 2024. 01. 26. Revised : 2024. 02. 15. Final Acceptance : 2024. 02. 19.

* First Author, Associate professor, Konyang Univ. Hospital, e-mail: kimhoongoing@hanmail.net

** Co Author, Professor, Hoseo University, Department of Digital Technology Management, e-mail: choss@hoseo.edu

*** Corresponding Author, Ph. D in Economics, Hoseo University, 20, Hoseo-ro79beon-gil, Baebang-eup, Asan-si, Chungcheongnam-do, 31499, Korea, Tel : +82-41-540-5071, e-mail: walras@hoseo.edu.

1. 서 론

2019년 말에 SARS-CoV-2 바이러스로 인한 코로나-19 대유행이 발생한 이후, 보건 당국은 확진자 추적, 바이러스 탐지 및 전염병 확진자 예측을 위해 다양한 노력을 진행해왔다. 현재까지 코로나-19 예측모형 접근 방식은 다음 범주 중 하나에 속한다. 먼저 가장 기본적이고 편리한 예측을 위한 단순 통계 모형이 그것이며, 둘째, SARS-CoV-2의 확산 진행을 분석하는 상황 기반모형으로 미래에 가능한 시나리오를 조사하는 모형이 주를 이루고, 마지막으로 바이러스 확산을 제어하기 위한 개입을 시뮬레이션하기 위한 미분방정식(예: 구획 모형) 또는 에이전트 기반의 기계론적 모형이 바로 그것이다[Keeling, and Ross, 2008 참조].

Kermack and McKendrick(1927)은 코로나-19 방역 예측 및 관리 모형으로 표준적인 SIR(Susceptible-Infected-Recovered) 모델을 제안하였다. 이 모형은 질병의 진행을 예측하는 데에 일반적으로 사용되는 구획 결정론적 모형이다. 그러나 코로나-19가 진행되면서 예측의 정확성 문제가 발생하였으며, 본 모형의 특성이 코로나-19 전염병을 장기적으로 예측하는 데에는 효율적이지 않음이 밝혀졌다[Roda, 2020 참조]. 이에 따라 예측모형의 불확실성을 효율적으로 처리하고, 질병 완화 및 억제 전략의 효과를 예측하기 위해 모형 예측 제어 접근법이 제안되었다. 이 접근법은 단순한 결정론적인 측면과 함께, 코로나-19 진행 상황에 따른 확률론적인 모형을 통해 전염병 확산의 예측 불가능한 특성을 반영하고 있다.

본 연구에서는 코로나-19 질병 예측모형을 다루고 있다. 이번 모형은 Pajaro et al.(2017)에서 제안한 확률론적 접근을 기반으로 한다. 본 분석은 기존의 전염병 구획 모형(SIR 유형 모형)을 CME(Chemical Master Equations)에 기반한 화학 공정 수식으로 확장하였다. CME는 확률 분포의 시간적 변화를 설명하며, 확률론적 시뮬레이션 알고리즘(SSA)을 통해 CME의 정확한 시간 경과 궤적을 계산할 수 있는 장점이 있다.¹⁾ 또한, 공중 보건 시스템에서 제공하는 새로운 감염

자에 대한 데이터 계열은 실제로 확률론적 프로세스 또는 양의 정수에 대한 무작위 보행을 구현한 것이다. 이러한 접근 방식은 SSA 알고리즘을 활용하여 해결된 CME 방정식을 기반으로 하여, 코로나-19 진행을 정확하게 예측하고 빠르게 수정하는 데에 특히 유용하다는 것이 알려져 있다.

〈Table 1〉 Usefulness of Determinism, Stochasticity, and Jump SIR Prediction Models

Model	Forecast of confirmed cases	Predictive power (%)
$SIR(\alpha = 1/7)$	187	70.6
$SIR(\alpha < 1/7)$	54	20.4
Stochastic SIR	11	4.1
Jump SIR	13	4.9
Total Confirmed Cases	265	100

Source : Pajaro et al.[2022, p. Modified from 4].²⁾

본 연구에서 제안한 코로나-19 예측모형은 다음과 같은 특성을 갖는다.

- 본 모형은 기본적으로 확률론적 SIR이다. 중요한 것은 확률론적 기계론 예측모형은 미래로 확장함에 따라 불확실성이 증가한다는 것이다.
- 본 모형은 상태와 매개변수 개수 측면에서 단순하며, 매우 적은 매개변수로도 탁월한 예측 능력을 보장한다. 이전에 다른 복잡한 모형보다 단순한 SIR 모형이 더 우수한 성능을 발휘한다는 보고도 있다.
- 본 모형은 기본적으로 높은 예측 능력을 갖춘 결정론적 모형에서 코로나 팬데믹의 바이러스 진행의 복잡성을 포함하고 있다.
- 본 모형은 매우 견고한 특성을 보인다. 이는 다양한 시나리오에서 일관된 예측 결과를 제공하며 모델의 신뢰성을 높인다.
- 본 모형은 코로나-19의 특징인 슈퍼 확산 사건을 포괄하여 분석할 수 있다. 이는 전염병의 급속한 확산 현상을 모델링하여 예측에 반영할 수 있음을 의미한다.

1) 이 문제를 해결하기 위하여 SSA방법 이전에 다른 대안적 방법으로 Sidje et al.[1999]가 제안한 반복 알고리즘을 생성하는 부분 공간 근사(KSA) 방법도 사용되었다.

2) 회복 모수는 격리 기간을 결정하는 α 이며, 기존 연구는 1/14을 제안하고 있으나, 본 실증분석 결과 우리나라의 경우 1/7이 더 적합한 것으로 나타났다.

- 본 모형은 다양한 시나리오를 검증하고 자유로운 자가 격리 기간을 반영한 중소 규모 자치단체 단위의 전염병 예측에 사용할 수 있다. 이는 예측모형의 유연성과 적용 범위를 확대시킨다.

기존의 연구 중에서 Sang-Sup Cho et al. [2021]은 COVID-19의 슈퍼감염에 대한 통계적 분석을 수행했다. 마찬가지로, Sang-Sup Cho et al. (2020)은 몬테카를로 기법을 이용하여 대면 수업 재개가 2차 감염에 미치는 영향을 분석했다. 그러나 최근 연구 중에서 본 연구와 가장 유사한 연구는 Jin-oh Kim et al. (2020)에서 SIR 예측모형과 LSTM 기계학습모형을 이용한 예측 비교연구이다. 하지만 심각한 감염이 발생하고 빠르게 관리해야 하는 상황에서는 예산과 시스템이 취약한 중소 지자체 단위에서 LSTM 방법을 적용하기는 어려운 상황이다. 이에 반해, 본 연구에서는 장기 시나리오와 단기 예측에 모두 유효한 확률론적이고 기계론적인 모형을 소개하고, 실증적으로 우리나라 소규모 지방자치단체인 서울특별시 강남구에 적용했다. 이러한 예측모형은 지자체 수준에서 전염병의 진행을 분석하고 예측하기 위해 소개되었으며, 적용 사례로는 서울특별시에서 2021년 동안 발생한 코로나-19에 대한 보건 시스템의 측정값을 활용하였다. 특히 SARS-CoV-2에 적용하기 위해 제안된 이 모형은 다른 병원체의 분석에도 쉽게 적용할 수 있어, 제시된 방법론은 질병 역학의 발전에 큰 기여를 할 것으로 기대된다.

본 연구의 구성은 다음과 같이 전개된다. 후술하는 장에서는 다양한 확장 기능을 갖춘 모형을 기술하였다. 먼저, 간단한 결정론 SIR(중소 규모 자치단체의 의료 시스템에서 감염된 개인의 진화를 비교적 정확하게 예측되는 것으로 입증됨)을 소개했다. 그런 이후에 다음 진행단계에 보완적으로 사용될 확률적 SIR 예측모형으로 확대했다. 마지막 진행 단계로 SARS-CoV-2는 한 명의 감염된 개인이 동시에 여러 사람의 감염을 허용하는 점프가 있는 질병 특성(슈퍼 감염자)을 반영하는 점프가 있는 SIR 모형으로 확대하여 분석하였다. 실증적 적용 사례로 기본적인 SIR 모형과 그 확장모형을 활용하여 2020년 1월부터 2022년 12월까지 우리나라 서울특별시 강남구 단위 자치단체에 대한 코로나-19 예측 결과와 상호 비교

분석하였다.³⁾

2. 코로나-19 SIR 예측모형

이런 장에서는 코로나-19 팬데믹의 단계별 진행을 예측하기 위해 개발된 예측모형을 기술했다. 본 장의 목표는 초기에 가능한 가장 단순한 모형으로 분석을 시도한 후, 최종적으로는 가장 우수한 예측 능력을 얻는 모형을 제시하는 것이다. 중요한 것은 예측모형이 국가 수준보다 지역 수준(자치단체)과 단계별 단기적 기간(최대 7~10일)의 정확한 예측이다. 이를 위해서 기본적 예측모형을 중심인 결정론적 SIR 모형으로부터 최종 복잡한 진행을 코로나-19를 예측하는 다중 감염 가능성을 고려한 점프 SIR 예측모형까지 단계적으로 간략하게 서술하였다.

2.1 결정론적 SIR 예측모형: 기본 모형

해당 총인구 N 은 잠재적 감염 가능자 S , 감염자 I 그리고 회복자(사망을 포함) R 로 구성된다. S, I, R 의 관계는 다음과 같다. $S+I \Rightarrow 2I$ 이며 $I \Rightarrow R$ 이다. 이때 감염률 β 와 회복률 α 가 모수로 작용된다. 표준적 SIR 모형과 달리 본 연구에서는 확진자의 동태적 행위 반응을 위해서 CME(Chemical Master Equation)을 고려했다 [Jahnke and Huisinga, 2007; Sunkara, 2009]. 이는 다음 서술하는 식 (1)과 같다.

$$\begin{aligned} \frac{dP(x,t)}{dt} = & \beta \frac{S+I}{N} IP(S+1, I-1, R, t) \\ & - \beta \frac{S}{N} IP(x,t) + \alpha(I+1)P(S, I+1, R-1, t) \\ & - \alpha IP(x,t) \end{aligned} \quad (1)$$

여기서 $x = (S, I, R)$ 이며, $P: [0, N]^3 \times R^+ \Rightarrow [0, 1]$ 로 변환하는 확률밀도함수이다. 기존 SIR 모형에 의한 예측 방법과 다른 점은 SSA(Stochastic Simulation Algorithm)를 적용하여 SIR 모형 과정을 구현한다는 것이다. 단일 SSA 시뮬레이션(또는 실현)의 궤적은

3) 본 코로나-19 예측모형은 서울특별시 강남구 뿐만 아니라 다른 중소 규모 지자체로 적용할 수 있다.

CME의 해인 확률 함수의 정확한 표본이므로 CME의 해는 현실적인 수의 실현으로 근사화된다. 이러한 방식으로, 의료 시스템에 의해 매일 보고되는 감염자 수는 SIR 프로세스에 대한 SSA의 단일 실현으로 간주하고, 데이터를 바로잡기 위하여 주어진 시간 범위에 대한 예측 값으로 10^3 번 반복한 SSA 실현의 평균 및 표준 편차를 사용했다.

2.2 확률적 SIR 모형: 바이러스 변형 기간

결정론적 SIR 예측모형에서 중요한 모수인 β 는 예측 기간에 일정하게 취급한다. 일정한 매개변수를 사용하는 SIR 모형은 감염 능자가 변동하지 않으면 팬데믹 행위의 전환점을 포착할 수 없다. 그러나 많은 지역에서 당국은 전염병의 다양한 변이단계에서 감염을 줄이기 위해 일련의 정책(봉쇄, 여행 금지, 사고 모임 수용 인원 제한 및 기타 제한 포함)을 적용한다. 따라서 보정에 사용되는 감염자의 시간 경과 데이터는 기본적 SIR 모형으로 정확하게 포착할 수 없는 경향의 변화를 보여줄 수 있다. 본 연구에서는 총감염자 수의 기울기 부호의 변화가 감지되는 시점을 전환점(T_{TP})으로 정의했다. 보정 목적으로 확진자에 전환점이 표시되면 다음과 같이 정의된 가변 감염률 β 을 사용하여 기존 SIR 모형을 확장한다[Seo and Choi, 2015].

$$\beta = \begin{cases} \beta_0, & t < T_{TP} \\ \beta_1, & t \geq T_{TP} \end{cases} \quad (2)$$

만일 $\beta_0 = \beta_1$ 이면 이 예측모형은 결정론적 SIR 모형과 같다.

2.3 점프 SIR 모형: 슈퍼 확산 기간

기본적 SIR 모형은 한 번에 한 사람의 감염을 고려한다. 이 가정이 SIR 예측모형이 SARS-CoV-2 감염자의 시간 경과 자료에서 발견되는 급격한 점프를 포착할 수 없는 이유이다. 감염된 개인 한 명이 같은 상호 작용 이벤트의 결과로 두 명 이상의 취약한 사람을 감염시킬 수 있다. 이러한 확산(슈퍼 확산) 상황을 고려하기 위해 기본적 SIR 모형과 관련된 반응을 다음(3)과 같이 일반화할 수 있다.

$$\begin{aligned} nS + I &\Rightarrow (n+1)I \\ I &\Rightarrow R \end{aligned} \quad (3)$$

본 연구에서 보진 당국이 설정한 대규모 집회 제한과 호환되는 $n=5$ (5인 이상 집합 규제)을 적용했다. 따라서 $n=1$ 인 경우 점프 SIR 모형은 가장 간단한 SIR 예측모형으로 회귀한다. 다음 장에서는 위에서 기술한 SIR 예측모형들을 실제 자료에 적용하고 그 유용성을 평가하였다. 이미 앞에서 살펴보았듯이, 기본적으로 표준적 SIR 예측모형에 보완적으로 확률적 그리고 점프 SIR 확장모형을 적용할 것을 제안한다.

3. 코로나-19 예측모형 적용 사례

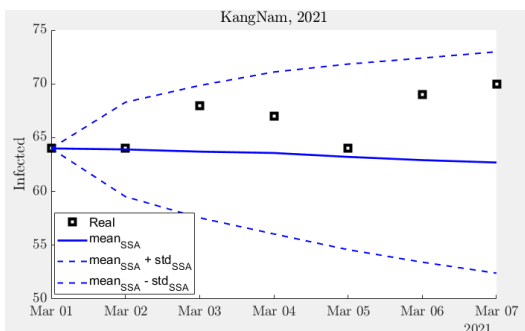
코로나-19의 발생과 퍼짐을 예측하는 주요 모형의 적용은 초기에는 표준적인 SIR 모형으로 시작되었다. 이 SIR 모형은 감염병의 전파를 수학적으로 모델링한 것으로, 감염되지 않은 사람들, 감염된 사람들, 그리고 회복한 사람들의 세 가지 상태를 기반으로 한다. 이를 바탕으로, 본 연구에서는 결정론적 SIR 모형, 확률적 SIR 모형, 그리고 슈퍼 확산 기간에 점프하는 SIR 모형으로 확장하여 적용하였다. 각각의 모형은 서로 다른 특성이 있지만, 그 구조는 동일하게 유지되며, 이는 모든 모형이 동일한 핵심 가정에 기반하고 있다는 것을 의미한다. 이러한 모형들의 유용성은 국가 수준에서의 질병 예측뿐만 아니라, 중소 지방 자치 단체에 대한 분석에서도 중요한 역할을 한다. 본 연구에서는 서울특별시에서 제공하는 '서울 열린 데이터 광장'의 보건 시스템에서 얻은 데이터를 사용하였다. 이 데이터는 감염자 수와 7일의 격리기간 동안 누적된 감염자 수를 포함하고 있다. 이를 통해, 감염자의 수와 감염 사태의 진행 상황을 정확하게 파악할 수 있다.

더불어, 본 연구는 서울특별시 강남구를 중심으로 분석하였다. 이 지역은 서울특별시 내에서도 인구 밀도가 높고, 높은 생활수준에 따라 보건수준도 높으며, 사회적 거리두기의 중요성이 크게 강조되는 지역으로서, 코로나-19의 전파 경로 및 속도를 이해 하는데 매우 중요한 역할을 하기 때문이다. 이로써, 본 연구는 강남구를 중심으로 한 감염병 전파 모델을 제시하며, 이를 통해 효과적인 감염병 관리 및 예방에 대한 준거 값을 제시하고자 한다.

본 연구에서 사용한 각 모형의 자유 매개변수는 다음과 같다. 기본적 SIR 모형에서 매개변수 α (복구율 상수) 및 β (감염률 상수)는 기존 연구를 참조하여 사전적으로 지정하였다. 자기상관 문제를 회피하기 위해서 매개변수를 $\alpha=1/7$ (누적 발생률은 7일의 시간 간격으로 계산됨)로 고정한다. 중요한 것은 완화 정책이 개인 회복률 상수(매개변수 α 고정과 연관됨)에 영향을 미치지 않고 감염률 상수 β (따라서 데이터에서 보정해야 함)에 영향을 미친다는 것이다. 확률적 SIR 모형의 경우 매개변수 β_0 , β_1 및 T_{change} 을 자유롭게 수정할 수 있다. T_{change} 은 역학 기술기 부호의 변화가 발생하는 시점이다. 점프 SIR 모형의 경우 매개변수 β 을 급격한 진행을 보완할 목적으로 사용할 수 있다.

3.1 SIR 모형 예측 결과

결정론 SIR 예측모형은 공중 보건 시스템에서 제공한 감염자 데이터를 기준으로 보정된다. 앞에서 설명된 SIR 예측모형의 결정변수 β 는 7일 누적 감염자의 궤적에 맞추기 위해 공중 보건 시스템에서 제공하는 감염자 데이터로부터 추정된다. 수정은 매주 이루어진다. 앞 장에서 설명한 대로 회복률 상수는 $\alpha=1/7$ 로 고정된다. 의료 시스템에 의해 감지된 감염자 수의 변화는 고정된 $\alpha=1/7$ 를 사용하는 가장 간단한 SIR 예측모형이며 연구 기간에 대부분 주(총 265명의 70% 이상) 동안 정확하게 예측할 수 있었다. $\alpha=1/7$ 및 감염률 상수 β 의 경우 기존 연구에 따라서 다음 값을 사용했다. 구체적으로 $\beta \in \beta_{v,j} = \{0.14\}$ 로 적용했다. 예측에 따른 결과는 <Figure 1>에 나타났다.



<Figure 1> SIR Prediction Model for COVID-19 in Gangnam-gu, Throughout the Entire Period

이 분석 결과는 두 가지 주요한 관점에서 예측 모형의 평가와 예측의 유용성에 중요한 연구적 의미를 지닌다.

첫째로, 본 기본 모형은 코로나-19 완화 정책의 정량적 평가에 활용될 수 있는 강력한 도구가 된다. 이 모형은 감염률 상수의 영향을 간편하게 정량화하여, 정책기관이나 연구자들이 코로나-19의 확산상황에 대한 정확한 이해를 바탕으로 효과적인 완화 정책을 수립하고 실행할 수 있게 해준다. 이를 통해 향후 발생할 팬데믹 상황에 대한 통제력을 높일 수 있을 것이다.

둘째로, 이 모형은 예측 목적으로도 사용될 수 있으며, 해당 값을 얻기 위해 필요한 것은 하나의 매개변수(이전 주의 값으로 고정 가능)와 시작 날짜의 초기 조건이다. 이는 본 모형이 데이터를 기반으로 한 굉장히 실용적인 예측 도구라는 것을 의미한다. 여기서 말하는 매개변수는 감염률, 회복률 등 전염병의 전파에 영향을 미치는 요인들을 의미하며, 이를 통해 미래의 감염 상황을 예측할 수 있다.

이미 <Table 1>에서 확인한 바와 같이, 표준적인 SIR 예측 모형은 의료 시스템에서 감지된 감염 사례의 시간적 변화를 1 표준 편차 내에서 비교적 정확하게 예측하는 것으로 나타났다. 예측의 정확도는 초기 예측과 후속 예측 사이에 4.1%의 변동성이 나타났으며, 특히 강남구에서의 감염 점프 모형은 4.9%의 보완적인 적용이 필요함을 보여준다. 이는 각 지역의 사회 경제적 특성에 따른 감염률의 차이를 반영하기 위한 보완적인 측정 방법이 필요하다는 것을 시사하고 있다.

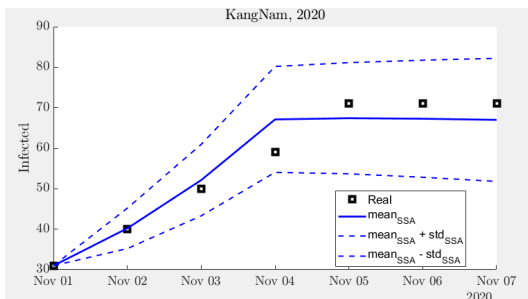
3.2 확률적 SIR 모형 예측 결과

연구기간 동안 전환점을 보여주는 전체 연구 기간에 대한 예외적인 주기가 존재한다. 이는 갑작스러운 감염자의 수의 증가를 의미한다. 급작스러운 감염자 확산상황이 그것이다. 확률론적 SIR 모형으로 얻은 예측을 비교한 적용된 예시가 <figure 2>에 표현 되었다. <Figure 2>에서 확인할 수 있는 것처럼, 두 가지 감염률 β_0 과 β_1 을 사용하는 확률론적 SIR 예측모형은 누적 감염 수준에 대한 훨씬 더 정확한 예측력이 향상되었다. 이를 통해 변화하는 감염 트렌드의 역학적 추이에 확률론적 SIR 모형을 적용함으로써, 전환점이 발생하는 조건들을 분석하는 데 도움이 된다는 점을 확인할 수 있다. 이는 감염의 확산 패턴을 이해하고, 예방 조치를 세우는 데 중요한

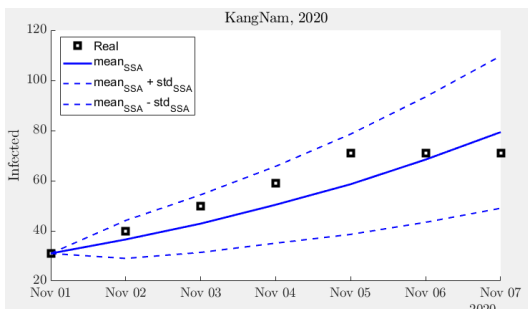
역할을 한다. 그러나 이러한 모형의 한계점은 불가피하게 존재한다. 전환점의 발생 시기에 대한 정확한 정보가 없다면, 예측모형의 정확도는 제한적이게 된다는 것이다. 이러한 한계를 극복하고 예측력을 향상 시키기 위해서는 실시간 데이터에 대한 보정 작업과 기계 학습 방법의 도입이 필요하다. 이를 통해 예측모형의 정확성을 높이고 감염 확산을 예방하는 데 효과적인 방안을 제시할 수 있을 것이다.

3.3 점프 SIR 모형 예측 결과

코로나-19의 가장 큰 속성인 슈퍼 확산 사건에 관해서 결정론 및 확률적 SIR 예측모형은 정확한 예측을 하지 못한다. 따라서 앞에서 설명한 점프 SIR 예측모형을 사용하여 누적 확진 사례의 역학을 살펴보고, 이를 적용함으로써 앞의 두 가지 SIR 예측모형을 사용하는 것 이상의 더 나은 예측을 얻을 수 있다. <Figure 3>에서 보듯이 점프 SIR 예측모형을 적용하면 일정 시점의 슈퍼 확산 감염자를 예측하는 데 도움이 된다.



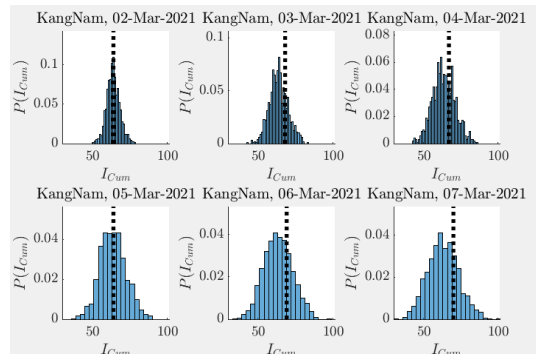
<Figure 2> Gangnam-gu SIR Prediction Model during the COVID-19 Virus Transformation Period



<Figure 3> Gangnam-gu SIR Prediction Model during the COVID-19 Super Spread Period

3.4 SIR 모형의 예측력 측정

SSA 알고리즘의 적용은 확률론적 SIR 모형 예측의 정확성을 평가하기 위해 평균 감염 사례 수와 표준 편차를 계산하는 데 사용된다. 이 외에도 상기 모든 SIR 모형에 대해 CME 솔루션의 근사치로 10^3 개의 SSA 시뮬레이션으로부터 감염 사례(신규 및 누적)의 확률 밀도 분포를 얻을 수 있다. 제공된 <Figure 4>는 이러한 SSA 알고리즘의 적용을 통해 얻은 결과를 보여준다. 이 그림은 특정 주 동안 주어진 누적 사례 수의 확률을 제시하는 것으로, 감염의 진행을 시각적으로 이해하는데 도움을 준다. <Figure 4>에서 볼 수 있듯이, 보건 시스템에 의해 보고된 사례는 시뮬레이션으로 얻은 확률 밀도 분포의 중위에 가까운 것으로 나타났다. 이는 제안된 예측 모형이 전염병의 행위를 잘 포착하고 있음을 시사하며, 이 모형의 유효성과 신뢰성을 확인하는 강력한 증거로 작용한다4).



<Figure 4> Comparison of Predicted Probability Distribution and Actual Value of Deterministic SIR Prediction Model

4. 결론 및 시사점

지금까지의 연구는 코로나바이러스 고전적 구획 모델(SIR 유형)을 기반으로 하는 확률론적 모형을 활용하여 코로나 바이러스 확산의 결정론적 특성을 통합하였다. 이러한 모형에 전체 인구를 세 가지 주요그룹, 즉 세분화한 각 그룹(예: 민감성, 감염자, 회복자)사이의 전환을 화학과정으로 해석했다. 즉 CME(Chemical Master

4) <Figure 4>는 결정론 SIR 예측모형의 예측 확률 분포만 보였으나 다른 두 예측모형에서도 유사한 결과를 보인다.

Equation)모형은 화학 반응 시스템의 고유한 확률론을 통합한 모델이므로, 이 새로운 공식을 사용하여 해당 CME를 제안함으로써 바이러스 감염 전파의 잡음을 자연적인 형태로 통합할 수 있는 장점을 보였다.

또한, 이 분석과정에서는 Gillespie(1976)의 SSA(확률적 시뮬레이션 알고리즘)를 사용하여 제안된 다양한 모형을 해결하여 CME(감염자의 확률 밀도 함수의 시간 진화)에 의한 감염자 수의 시간 변화 궤적 해법을 실현하였다. 결정론적 SIR 모형은 일련의 매개변수에 대한 고유한 궤적만 캡처하는 반면, 두 가지 확률론적 SIR 모형은 전염병 확산의 잡음 측정 값을 자동으로 제공하는 결과를 생성한다. 이로부터 시간이 지남에 따라 보고된 표준 편차 또는 얻은 분포의 꼬리가 높아짐에 따라 모형의 불확실성이 어떻게 증가하는지 관찰할 수 있었다.

이를 근거로 본 연구에서는 제시된 표준적 SIR 모형과 두 가지 확률론적 모형을 중심으로 공중 보건 데이터를 통합하여, 우리나라 서울특별시 강남구의 지방자치단체 수준에서 코로나 19 전염병의 확산 진행을 분석하고 예측하는 방법을 소개하였다. 이러한 예측모형 적용 사례에서 보듯이 다양한 지방자치단체에서 SARS-CoV-2 감시를 위해 보정되고, 성공적 감염자 예측모형을 통하여 전염병 확산감염 방지에 도움을 줄 수 있다. 아울러 이를 통해 본 예측 모형들은 미래 발생 가능한 전염병 관리와 보건정책 수립에 필수적인 도구를 구성할 것으로 보인다.

본 연구 결과를 요약하면 다음과 같다. 먼저 전통적이고 가장 기본적인 SIR 예측모형은 전반적으로 감염자 예측에 기본 모형으로 평가된다. 특히 확진자 예측의 강건성이 높기 때문이다. 둘째, 그럼에도 불구하고 감염병의 특성을 보완해야 하는 부분이 있다. 질병의 성격이 진행 과정에 따라 변이 등을 통하여 바뀌는 상황이며, 또 다른 하나는 슈퍼 감염자 발생사례이다. 이를 위하여 두 가지 방향에서 기본적 SIR 모형을 보완하면 코로나-19 감염자 예측 정확도가 증가하였다.

이러한 예측모형은 보정이 쉽고(시간 경과 데이터에서 보정을 위한 매개변수가 2개뿐임), 모든 지방자치단체 수준의 감염자 수를 4개의 서로 다른 감염률 매개변수만 사용하여 1년 동안을 예측하기에 매우 견고하다. 특히 확률론적 SIR 모델에 대한 4가지 매개변수 세트를 얻었는 바, 이는 고려된 5개 자치단체의 기간에 바이러

스 감염을 설명하기에 충분했다. 이러한 매개변수는 펜데믹 기간에 부과된 제한(봉쇄)으로 설명될 수 있다. 따라서 본 연구는 코로나 19 확산의 경우 전염병의 바이러스 확산을 예측하기 위해 CME를 기반으로 한 이러한 종류의 확률론적 모델의 유용성에 대한 증거를 제공한다. 향후에는 본 분석을 토대로 전환점 예측을 크게 촉진할 수 있는 매개변수의 자동화된 실시간 보정을 위한 인공지능 기술 통합이 필요하다.

References

- [1] Cho, S. and Kim, J., "Intervention analysis for spread of COVID-19 in South Korea using SIR model", *The Korean Journal of applied Statistics*, Vol. 34, No. 3, 2021, pp. 477-489.
- [2] Cho, S. S., Chae, D. W., and Lim, S. J., "Resumption of School Face-to-Face Classes and Analysis of Secondary Infected Persons in COVID 19 : Applying the Monte-Carlo Method", *Journal of Information Technology Applications & Management*, Vol. 28, No. 1, 2021, pp. 33-41.
- [3] Cho, S. S., Lee, J. M., and Sawng, Y. W., "Data Analysis on COVID-19 Infection Heterogeneity", *Journal of Technology Management*, Vol. 6, No. 4, 2021, pp. 219-233.
- [4] Gillespie, D. T., "A general method for numerically simulating the stochastic time evolution of couple chemical reactions", *J. Comput. Phys.*, Vol. 22, 1976, pp. 403-434.
- [5] Jahnke, T. and Huisinga, W., "Solving the chemical master equation for monomolecular reaction systems analytically", *Journal of mathematical biology*, Vol. 54, 2007, pp. 1-26.
- [6] Keeling, M. J. and Ross, J. V., "On methods for studying stochastic disease dy-

- namics”, *J. R. Soc. Interface*, Vol. 5, 2008, pp. 171-181.
- [7] Kermack, W. O. and McKendrick, A. G., “Contributions to the Mathematical Theory of Epidemics-I”, *Proceedings of the Royal Society*, 115A, 1927, pp. 700-721.
- [8] Kim, J., Kim, J., Kim, H., Park, H., and Kim, J., “Comparison and analysis of COVID-19 confirmed case prediction based on SIR model and LSTM”, *Korea Intelligent Information Systems Society Conference Proceedings*, 2020.
- [9] Pajaro, M., Alonso, A. A., Otero-Muras, I., and Vazquez, C., “Stochastic modeling and numerical simulation of gene regulatory networks with protein bursting”, *J. Theoret Biol*, Vol. 421, 2017, pp. 51-70.
- [10] Pajaro, M., Noelia, F., Antonio, A. A., and Irene, O. M., “Stochastic SIR model predicts the evolution of COVID-19 epidemics from public health and waste water data in small and medium-sized municipalities: A one year study”, *Chaos, Solitons and Fractals*, Vol. 164, 2022, pp. 1-10.
- [11] Roda, W. C., Varughese, M. B., Han, D., and Li, M. Y., “Why is it difficult to accurately predict the COVID-19 epidemic?”, *Infec Dis Model*, Vol. 5, 2020, pp. 271-81.
- [12] Seo, M. and Choi, B. S., “A Estimation Method for Stochastic Epidemic Model”, *Journal of The Korean Data Analysis Society*, Vol. 17, No. 3, 2015, pp. 1247-1259.
- [13] Sidje, R. B., and Stewart, W. J., “A numerical study of large Sparse Matrix Exponential’s Arising in Markov Chains”, *Comput. Stat. Data An*. Vol. 29, 1999, pp. 345-368.
- [14] Sunkara, V., “The chemical master equation with respect to reaction counts”, In *Proc. 18th World IMACS/MODSIM Congress* Vol. 703, 2009, p. 707.

■ 저자소개



Hoon Kim

Hoon Kim obtained his B.A. from Chosun University and his Ph.D from Chungnam National University College of Medicine. He served at Daejeon Eulji University

Medical Center as a professor. He mainly have two research fields. In medical field, he has focused on oral and maxillofacial surgery (included implant). In social science field, his research has focused on the areas of health-care service and social service based on public healthcare. He is a professor at the Konyang University medical center.



Sang-sup Cho

Sang-sup Cho earned his Ph.D. in Economics from Saint Louis University and subsequently served as a chief researcher at ETRI (Electronics and Telecommunicat-

ions Research Institute). His research primarily revolves around the economics of technology innovation in the fields of Electronics and Telecommunication, Artificial Intelligence, and the Fourth Industrial Revolution, with a foundation in Econometrics. Currently, he holds a position as a professor in the Department of Management of Digital Technology at Hoseo University.



Dong-woo Chae

Dong-woo Chae is currently pursuing his Ph.D. in Economics at Hoseo University, having graduated from Sung-Kyun-Kwan University. His research focuses on the com-

plexities of Economics, specifically in the areas of Environmental Economics, utilizing econometric methodologies. Previously, he held the position of Manager of the Accounting Team at Hoseo University. Presently, his academic pursuits involve the study of economic phenomena within the framework of administrative organizational structures.