Estimation of Parameters for the SEIQRD Model

Bhatt Himanshu
Department of Mathematics, University of Delhi, 110007, INDIA
du.himanshu@gmail.com

Abstract

Accurately quantifying uncertainty in data-driven mechanistic models is crucial for public health applications. COVID-19, a complex disease with significant global health and economic ramifications, underscores this necessity. The pandemic's widespread infections, mortality and economic disruptions highlight the critical importance of understanding viral behavior and generating reliable short- and long-term forecasts of daily new cases. Machine learning and mathematical models are actively deployed in this effort.

To guide disease management strategies, researchers have employed diverse mathematical models to analyze the intricate transmission dynamics of COVID-19 under varying assumptions. This study presents the application of a six-compartment SEIQRD epidemiological model for estimating active COVID-19 cases and deaths. Parameter estimation is achieved through Approximate Bayesian Computation (ABC), leveraging the M-nearest neighbour Sequential Monte Carlo ABC method which delivers the estimated parameter values.

Keywords: SEIQRD model, Approximate Bayesian computation, Parameter estimation.

Introduction

Originating in China in December 2019, the novel coronavirus, COVID-19, has rapidly spread across the globe, impacting millions of individuals. Its inherent characteristics, including a high propensity for mutation and efficient transmission, have established COVID-19 as a significant threat to global public health[16]. Those infected, frequently experienced severe respiratory distress and are susceptible to developing serious health complications, particularly if they possess pre-existing chronic illnesses. The World Health Organization's declaration of COVID-19 as a pandemic on March 11, 2020, underscored the substantial challenge in effectively controlling its spread. The difficulties in containment are largely attributed to the phenomenon of asymptomatic transmission and the prolonged incubation period, wherein individuals may not exhibit symptoms until a considerable time after infection.

Mathematical modeling has proven to be an indispensable tool in projecting the scale and severity of disease outbreaks, thereby guiding the formulation of effective intervention strategies[5]. Models such as the Susceptible-Infected-Recovered (SIR) and its more complex variations, notably the Susceptible-Exposed-Infected-Removed (SEIR) model, have been extensively employed to analyze the COVID-19 pandemic[7]. For instance, Tang et al[13] utilized a traditional SEIR model to evaluate the infectivity of COVID-19, while Wu et al[15] developed an enhanced SEIR model to forecast the virus's dissemination. However, it is important to note that these studies often operate under the assumption that individuals in the exposed phase are non-infectious, a premise that may not accurately reflect the actual transmission dynamics of COVID-19.

The SEIQRD model, an advanced adaptation of the traditional SIR framework, systematically represents infectious disease dynamics by categorizing the population into susceptible, exposed, infected, quarantined, recovered and deceased groups[3]. It achieves a practical balance between detailed modeling and analytical clarity, avoiding over-parameterization while ensuring result interpretability. In this study, the model's parameters are carefully chosen to focus on essential epidemiological factors such as transmission rates, incubation periods, quarantine efficiency, recovery rates and mortality offering valuable insights into COVID-19's progression without adding unnecessary complexity.

SEIQRD Model

The substantial viral load in COVID-19 patients makes them highly infectious. Early in the disease, the virus is readily detectable in throat swabs, allowing for easy transmission through respiratory droplets. This poses a significant risk to those nearby, as even individuals without symptoms during the incubation phase can spread the virus. After infection, the virus multiplies within cells with symptoms typically appearing within 14 days. To mathematically model the spread of COVID-19, several simplifying assumptions are made.

These include treating the population as closed, excluding births from the susceptible group, allowing for reinfection after recovery, assuming all contacts lead to infection and ignoring deaths unrelated to COVID-19. Consider that the total population (N) of a certain area is divided into six compartments at time (t) as Susceptible $\mathrm{S}(t)$, Exposed $\mathrm{E}(t)$, Infected $\mathrm{I}(t)$, Quarantined $\mathrm{Q}(t)$, Recovered $\mathrm{R}(t)$ and Deaths $\mathrm{D}(t)$.[1,2] Then the population can be modeled with the system of differential equations:

$$\frac{d\mathrm{S}}{dt} = -\beta \mathrm{S}(t)\mathrm{I}(t) + \mu \mathrm{R}(t)$$

$$\frac{d\mathrm{E}}{dt} = \beta \mathrm{S}(t)\mathrm{I}(t) - \gamma \mathrm{E}(t)$$

$$\frac{dI}{dt} = \gamma E(t) - (\sigma + \lambda + \delta)I(t)$$
$$\frac{dQ}{dt} = \lambda I(t) - \xi Q(t) - \delta Q(t)$$
$$\frac{dR}{dt} = \sigma I(t) - \mu R(t) + \xi Q(t) \quad (1)$$
$$\frac{dD}{dt} = \delta I(t) + \delta Q(t)$$

where $\beta, \mu, \gamma, \sigma, \lambda, \delta, \xi$ represent the rate of transmission from susceptible to exposed, reinfection rate, (rate at which an exposed person becomes infected or incubation rate) rate of recovery, quarantine rate, death rate and quarantine period respectively and S(t) + E(t) + I(t) + Q(t) + R(t) + D(t) = N.

Approximate Bayesian Computation Sequential Monte Carlo M nearest Neighbours

Approximate Bayesian Computation (ABC) methods were developed to infer posterior distributions when evaluating likelihood functions which are computationally difficult or prohibitively expensive[14]. They leverage the efficiency of modern simulation techniques by substituting the likelihood calculation with a comparison between observed and simulated data[9]. The main idea is to use Baye's theorem to approximate the posterior distribution with the knowledge of the observed data. Let $\pi(\theta)$ denote the prior distribution for the parameter vector θ and the likelihood function for the data D be denoted by $p(D|\theta)$, then using the Bayes theorem, posterior distribution of the parameter θ is proportional to $\pi(\theta)p(D|\theta)$. Now we describe the steps for the ABC SMC MNN method to estimate the posterior distribution for the parameter vector θ and get the samples from the posterior distribution as follows[10,12] :

(i) Fix the number of generations S and the number of samples to be simulated from the posterior distribution as N.
(ii) Fix the tolerance value for each of the generation as $t_1, t_2, \dots, t_S$. Now set the generation variable $s = 1$.
(iii) If the generation variable is $s > 1$, select the M nearest neighbours of the parameter vectors θ_{s-1}^i, then evaluate the covariance matrix $\Sigma(\theta_{s-1}^i, M)$ for all the samples $i = 1,2, \dots, N$.
(iv) Set the sample indicator i=1.
(v) For the first generation i.e. for $s = 1$, sample θ^{**} is selected from the prior distribution $\pi(\theta)$. For generation s > 1, select θ^* from the previous generation samples $\{\theta_{t-1}^1, \theta_{t-1}^2, \dots, \theta_{t-1}^N\}$ with the weights $\{w_{t-1}^N, w_{t-1}^N, \dots, w_{t-1}^N\}$ and using θ^*, select a random sample θ^{**} from the multivariate normal distribution so that $\theta^{**} \sim N(\theta^*, \Sigma(\theta^*, M))$.
(vi) If $\pi(\theta^{**}) = 0$, return to step (iv). Else generate the data $\bar{D}^{**}$ from the model using the parameter vector θ^{**}.
(vii) If $(\bar{D}|\bar{D}^{**}) < t_s$, set $\theta_s^{(i)} = \theta^{**}$ and set the weight of the accepted sample corresponding to i as $w_s^i = \begin{cases} \pi(\theta^{**}), & if\ s = 1 \\ \frac{\pi(\theta^{**})}{\sum_{j=1}^N w_s^j \mathrm{N}(\theta_s^i|\theta_{s-1}^j, \Sigma(\theta_s^i,\ M))}, & if\ s > 1 \end{cases}$
(viii) If $i < N$, increase $i = i + 1$ and go to step (iv).
(ix) Normalize the weights so that $\sum_{i=1}^N w_s^{(i)} = 1$.
(x) If s < G, set $t = t + 1$, go to step (iv).

where N(.,.) represents multivariate normal distribution which is used as perturbation kernel in general. Uniform distribution is used as non-informative perturbation kernel and multivariate normal distribution as informative perturbation kernel. Effective exploration of the parameter space is dependent on choices made regarding tolerance levels, generational counts, simulation repetitions and the perturbation kernel.

Data and Simulations

To analyze the evolving nature of COVID-19 over time, we need parameter estimation that accounts for variability[6]. Consequently, we utilize a time-dependent approach to track these fluctuating trends. In this study we have divided the data in different time windows. We modeled the COVID-19 disease using the SEIQRD model. The data for this analysis came from Saudi Arabia. It covers the period from March 2, 2020, to June 21, 2020. The data was obtained from Kaggle[18].

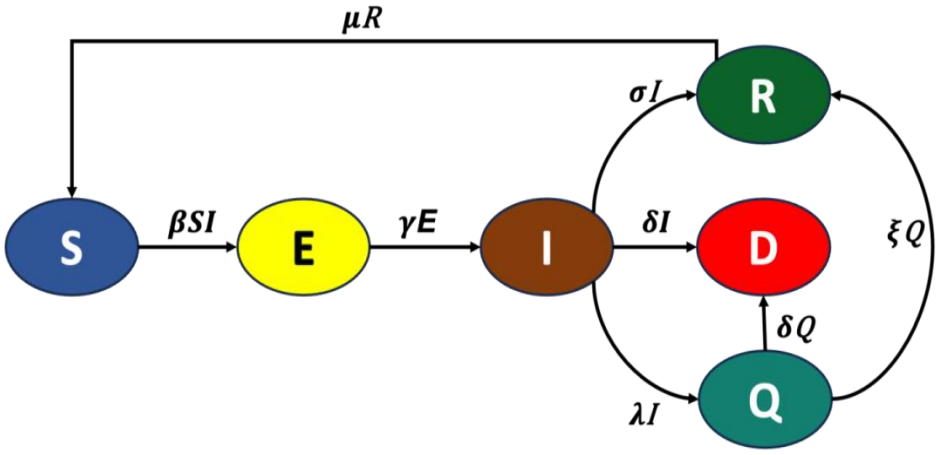

Figure 1: Representing the compartmental diagram for the SEIQRD model

It includes daily information on active, cumulative and new cases, as well as deaths and recoveries. Since the parameters involved in the SEIQRD model are time dependent and varying with respect to time. we have divided data into five time windows. First time window starts from day 1 to day 15 and second time window starts from day 16 to day 40, third starts from day 41 to day 65 and fourth starts from day 66 to day 85 and fifth time window starts from day 86 to day 110[11]. We used active cases and deaths as measurable data for the model. For the first time window, we used data for infected cases as $\mathrm{I}(t)$ and deaths $\mathrm{D}(t)$ from day 1 to day 15 in the Approximate Bayesian Computation sequential monte carlo M-nearest neighbour sampling method for the samples of the parameters in the SEIQRD model.

The distance metric is $d(\bar{D}|\bar{D}^{**}) < t_s$ if and only if $\sqrt{\left(\mathrm{I}(t) - \mathrm{I}^{**}(t)\right)^2} < t_s^I$ & $\sqrt{\left(\mathrm{D}(t) - \mathrm{D}^{**}(t)\right)^2} < t_s^I$ is used in the sampling method. The number of generation S=3 is used in each time window. Table 1 represents the prior distribution used for the parameters in the ABC SMC MNN sampling method. M=200 is used for the M-nearest neighbours.

The initial values of the tolerance sequence for each time window is given in table 2 which is used to sample the parameters using ABC SMC MNN sampling method. The total number of population of Saudi Arabia on March 2 , 2020 was $N = 35755176$. Initial values for infected and deaths compartment are taken from the available data and for the first time window, initial values of quarantined, recovered were zero. Initial value for Exposed compartment is taken as $\mathrm{E}(0) = 1000$ and for susceptible $\mathrm{S}(0) = N - \mathrm{I}(0) - \mathrm{R}(0) - \mathrm{Q}(0) - \mathrm{D}(0) - \mathrm{E}(0)$ and For subsequent time windows, the initial values for the exposed, quarantined and recovered compartments were set to the estimated values of these compartments from the preceding time window. These estimated values were derived from the SEIQRD model, utilizing the mean of the sampled parameters from the final generation of the ABC-SMC MNN sampling method's posterior distribution.

The ABC SMC-MNN sampling method was applied for each time window to get the samples of the parameters from the posterior distribution. For each time window, the samples obtained from the posterior distribution in the last generation taking their mean value as true value of the parameters for the corresponding time window and using them in the SEIQRD model, we get the estimated value of the infected and deaths compartment. Table 3 represents the estimated values of the parameters involved in the SEIQRD model for each time windows.

Table 1
Representing prior distribution of the parameters used in SEIQRD model.

Parameter	Prior distribution (Truncated Normal distribution)
β (transmission rate)	Truncnorm(lower = -0.01, upper = 0.1 , mean = 1.5e-8, sd = 1e-8)
μ (reinfection rate)	Truncnorm(lower = 0.0001, upper = 0.1, mean = 1.6e-4, sd = 1e-4)
γ (incubation rate)	Truncnorm(lower = 0.000001, upper = 0.01, mean = 0.015, sd = 0.001)
σ (recovery rate)	Truncnorm(lower = 0.01, upper = 0.1, mean = 0.05, sd = 0.01)
λ (quarantine rate)	Truncnorm(lower = 0, upper = 0.1, mean = 0.021, sd = 0.01)
δ (death rate)	Truncnorm(lower = 0.001, upper = 0.5, mean = 0.1, sd = 0.01)
ξ (quarantine period)	Truncnorm(lower = 0.001, upper = 0.1, mean = 0.07, sd = 0.01)

Table 2
Representing the sequence of tolerance used in the ABS SMC MNN

Time Window	Tolerance sequence (t_1^I, t_2^I, t_3^I) Infected	Tolerance sequence (t_1^D, t_2^D, t_3^D) Deaths
Day $1-15$	$(1e4, 8e3, 5e2)$	$(1e3, 8e2, 5e2)$
Day $16-40$	$(1e4, 8e3, 7e3)$	$(1e3, 8e2, 7e2)$
Day $41-65$	$(1e5, 9e4, 8e4)$	$(1e3, 7e2, 5e2)$
Day $65-85$	$(1e6, 8e5, 1e5)$	$(1e3, 7e2, 5e2)$
Day $86-110$	$(1e6, 8e5, 1e5)$	$(1e3, 7e2, 5e2)$

Table 3
Presenting the estimated values of the parameters involved in the SEIQRD model.

Parameter	Day 1-15	Day $16-40$	Day $41-65$	Day $65-85$	Day $86-110$
β	2.607926e-08	3.137336e-08	1.906792e-08	2.477616e-08	2.005747e-08
μ	1.404689e-04	2.180625e-04	1.836082e-04	2.233315e-04	1.230543e-04
γ	1.025117e-02	1.031941e-02	9.927145e-03	1.021641e-02	1.033123e-02
σ	4.881843e-02	4.086454e-02	4.920895e-02	4.590827e-02	5.012280e-02
λ	2.548353e-02	1.895990e-02	2.071212e-02	2.370133e-02	1.144448e-02
δ	1.061898e-02	9.215500e-02	1.053984e-02	1.083446e-02	9.866096e-03
ξ	7.048349e-02	7.409066e-02	7.051973e-02	6.572828e-02	7.009715e-02

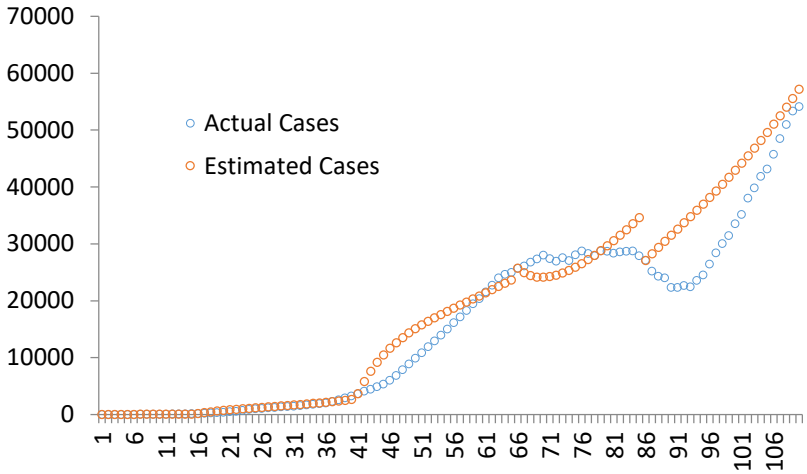

Figure 2: Depicting the Active and Estimated cases

Table 4
Represents the basic reproduction number each time window.

Time Window	Day 1-15	Day 16 – 40	Day 41 – 65	Day 65 – 85	Day 86 – 110
R_0	10.98043	7.380998	8.473392	11.01232	10.03954

Figure 2 represents the estimated and actual infected. Estimated cases are obtained by using the parameter values for each time window in the SEIQRD model. Using the parameter values of the last window, infected values can be estimated[8].

Reproduction Number: In infectious disease modeling, the basic reproduction number (R_0) plays a key role in analyzing outbreak patterns. It represents the average number of secondary infections caused by a single case in a fully susceptible population, helping to guide public health strategies[17]. However, R_0 varies due to factors such as behavioral shifts, vaccination efforts and viral mutations. When R_0 exceeds 1, the infection spreads widely whereas a value below 1 indicates a natural decline. Since R_0 is highly sensitive to model parameters, precise estimation is essential. Many COVID-19 studies have employed the Next Generation Matrix (NGM) method for this purpose[4]. Basic reproduction number for the SEIQRD model can be given by $R_0 = \frac{\beta N}{\sigma+\delta+\lambda}$. Table 4 represents the basic reproduction number each time window.

Conclusion

In conclusion, this study successfully demonstrated the strong integration of the SEIQRD model with the ABC-SMC MNN sampling method for precise parameter estimation in the complex dynamics of the COVID-19 pandemic in Saudi Arabia. This study established a robust framework for analyzing and forecasting the disease's trajectory by carefully computing the time-varying basic reproduction number (R_0), a key indicator of disease transmission. It also accurately estimates essential parameters such as infection, recovery and mortality rates. Given the pandemic's changing nature, adaptive models are critical and the time-varying technique used here adequately reflects these shifts. The findings of this study will have major practical implications, providing Governments and public health authorities with a data-driven tool for informed decision-making. These results, which provide a more accurate and thorough picture of the pandemic's course, can help direct targeted actions, improve resource allocation and strengthen public health policies to reduce the disease's impact on the population.

Acknowledgement

The author is supported with a senior research fellowship under the file number 09/0045(11788)/2021-EMR-I from the Council of Scientific and Industrial Research, Ministry of Science and Technology, Government of India, New Delhi.

References

1. Alleman T.W., Rollier M., Vergeynst J. and Baetens J.M., A stochastic mobility-driven spatially explicit SEIQRD COVID-19 Model with VOCs, seasonality and vaccines, *Applied Mathematical Modelling*, **123**, 507-525 **(2023)**

2. Alyami L. and Das S., State Estimation of the Spread of COVID-19 in Saudi Arabia using Extended Kalman Filter, 2022 Sensor Signal Processing for Defence Conference (SSPD), London, United Kingdom, 1-5 **(2022)**

3 Alyami L., Das S. and Townley S., Bayesian model selection for COVID-19 pandemic state estimation using extended Kalman filters: Case study for Saudi Arabia, *PLOS Global Public Health*, **4(7)**, e0003467 **(2024)**

4. Diekmann O., Heesterbeek J.A.P. and Roberts M.G., The construction of next-generation matrices for compartmental epidemic models, *Journal of the Royal Society Interface*, **7(47)**, 873-885 **(2010)**

5. Fanelli D. and Piazza F., Analysis and forecast of COVID-19 spreading in China, Italy and France, *Chaos, Solitons & Fractals*, **134**, 109761 **(2020)**

6. Garg P., Arora R., Madan S. and Singh D.K., Estimation of Parameters and Stability Analysis of Coronavirus Pandemic, *Advances in Differential Equations and Control Processes*, **28**, 119 - 134 **(2022)**

7. Hethcote H.W., The mathematics of infectious diseases, *SIAM Review*, **42(4)**, 599-653 **(2000)**

8. Madan S., Arora R., Garg P. and Singh D.K., Analysis and Prediction of Covid-19 Spread using Numerical Method, *Advances in Differential Equations and Control Processes*, **27**, 97–114 **(2022)**

9. McKinley T., Vernon I., Andrianakis I., McCreesh N., Oakley J., Nsubuga R. and White R., Approximate Bayesian Computation and simulation-based inference for complex stochastic epidemic models, *Statistical Science*, **33(1)**, 4-18 **(2018)**

10. Minter A. and Retkute R., Approximate Bayesian Computation for infectious disease modeling, *Epidemics*, **29**, 100368 **(2019)**

11. Singh D.K. and Bhatt H., Parameter estimation for Covid-19 cases using Bayesian model, *Montes Taurus J. Pure Appl. Math.*, **6(2)**, 99-108 **(2024)**

12. Sisson S.A., Fan Y. and Tanaka M.M., Sequential Monte Carlo without likelihoods, Proc. Natl. Acad. Sci. USA, **104**, 1760–1765 **(2007)**

13. Tang B., Wang X., Li Q., Bragazzi N.L., Tang S., Xiao Y. and Wu J., Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions, *J. Clin. Med.*, **9(2)**, 462 **(2020)**

14. Toni T., Welch D., Strelkowa N., Ipsen A. and Stumpf M.P., Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, *Journal of the Royal Society Interface*, **6(31)**, 187-202 **(2009)**

15. Wu J.T., Leung K. and Leung G.M., 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 **(2000)**

16. Pang L., Liu S., Zhang X., Tian T. and Zhao Z., Transmission dynamics and control strategies of COVID-19 in Wuhan, China, *Journal of Biological Systems*, **28(3)**, 543-560 **(2020)**

17. Zuhairoh F., Rosadi D. and Effendie A.R., Determination of basic reproduction numbers using transition intensities multi-state SIRD Model for COVID-19 in Indonesia, *Journal of Physics*, Conference Series, **1821(1)**, 012050 **(2021)**

18. https://www.kaggle.com/datasets/lamaabdullah11/saudi-arabia-covid19-data.

(Received 04th April 2025, accepted 05th May 2025)