Application of Bayesian Methods to Epidemics

Tongfei Tian*

Faculty of Natural, Mathematical & Engineering Sciences, King's college London, London, WC2R 2LS, United Kingdom

* Corresponding Author Email: k22042837@kcl.ac.uk

Abstract. This template explains and demonstrates how to use Bayesian methods to estimate the epidemics. And this study will use the case COVID-19 in United State as an example. In this days, infectious diseases have had a huge impact on human society since ancient times. And in this article will find out how to use the parameter of the basic reproduction number, infection rate and recovery rate of the epidemics combined with Bayesian method and SIR model to estimate a linear regression and to figure out the connection in epidemics. This study will specifically explore how parameters estimated by Bayesian methods can be used to predict the spread trend and potential impact of the disease. This includes using Bayesian methods to update model parameters to adapt to new epidemic data, and how model predictions can be used to guide public health decisions and disease control measures. Moreover, helping people to predict the spread of the disease, and assess the potential scale of the disease outbreak and formulae effective intervention measure.

Keywords: Bayesian methods, SIR model, parameter estimation, infectious disease models.

1. Introduction

Infectious diseases have had a huge impact on human society since ancient times. From the Black Death in history to HIV/AIDS in the early 21st century to the COVID-19 pandemic in recent years, every outbreak of infectious diseases has had a great impact on the global public. Hygiene poses a serious challenge [1]. These diseases not only claim countless lives, but also have a profound impact on economies, social structures, and the process of globalization. In today's globalized world, infectious diseases are spreading faster and more widely than at any time in history, making effective disease surveillance, control, and prevention measures particularly important.

Accurate estimation of infectious disease model parameters is key to understanding and controlling the spread of infectious diseases [2]. These parameters, including the basic reproduction number (R0), infection rate, recovery rate, etc., are the basis for constructing infectious disease dynamics models and can help us predict the spread of the disease, assess the potential scale of the disease outbreak, and formulate effective intervention measures. Therefore, using advanced statistical methods, such as Bayesian statistics, to estimate these key parameters can not only improve the accuracy of the model but also integrate data and prior knowledge from different sources to provide better information for the formulation of public health responses. reliable scientific basis [3]. This approach is particularly important in an ever-changing disease-spread environment because it allows the model to update as new data arrives, providing real-time, dynamic disease-spread predictions.

In traditional frequentist statistical methods, parameter estimation often relies on many assumptions and simplifications, which may not always accurately reflect complex real-world situations. Particularly in dynamic models of infectious diseases, these methods may be limited by their inability to fully account for uncertainty and variability in the data. In contrast, Bayesian methods provide a more flexible and comprehensive framework for estimating infectious disease model parameters [4]. The importance and advantages of this method are reflected in the following aspects: First, Bayesian methods can integrate prior knowledge and new data. In the early stages of an infectious disease outbreak, available data are often limited, when traditional methods may struggle to provide reliable estimates due to data scarcity. Bayesian methods allow researchers to use historical data or expert knowledge as prior information, combine it with new observation data, and give parameter estimates through posterior distributions so that reasonable inferences can be made even with limited data [5]. Second, Bayesian methods provide a natural way to quantify the uncertainty of estimates. Through the posterior distribution, researchers can not only obtain point estimates of parameters but also obtain direct measures of uncertainty about these parameters, such as confidence intervals. Furthermore, Bayesian methods have advantages in handling complex models and large amounts of data. With the improvement of computing power and the development of algorithms. Bayesian methods can handle complex models and large-scale data sets that are difficult to handle with traditional methods. This allows researchers to build more refined models to better simulate the disease spread process, thereby providing a more accurate basis for the formulation of prevention and control measures [6].

Finally, the flexibility of Bayesian methods is also reflected in their ability to adapt to new information. As a disease outbreak progresses and new data emerges, Bayesian methods can adapt to this new information by updating the posterior distribution to provide real-time, dynamic parameter estimates. This is particularly critical in responding to the rapidly changing infectious disease landscape.

In summary, there is an important necessity to use Bayesian methods to estimate infectious disease model parameters. Its advantage over traditional frequency statistics methods is that it can better handle uncertainty, integrate prior knowledge and new data, handle high-complexity models, and adapt to new information, providing a powerful tool for the control and prevention of infectious diseases [7]. The goal of this study is to explore how Bayesian methods can be used to estimate the parameters of infectious disease models and predict the transmission dynamics of the disease [8, 9].

At the same time, this paper will deeply explore the application of Bayesian methods in parameter estimation of infectious disease models, and how to use this method to predict the spread of diseases. Specific research objectives can be broken down into the following points: First, this study will introduce the basic concepts of Bayesian statistics in detail, including prior distribution, likelihood function, posterior distribution, and explain how to apply these concepts to parameter estimation of infectious disease models. This goal is to provide readers with a theoretical foundation for Bayesian methods to better understand subsequent applications and analyses.

Through an infectious disease case study of COVID-19 in the United States, this study will demonstrate how Bayesian methods can be practically applied to estimate the parameters of an infectious disease model and predict the spread dynamics of the disease based on these parameters. Case studies will help illustrate the potential value and application of Bayesian methods in the development of practical disease control and prevention strategies [10].

2. Methods

2.1. Concepts of Bayes Theorem

Bayes' theorem is a core concept in statistics and probability theory that provides a way to update existing beliefs or probability estimates based on new evidence [10]. The basic formula of Bayes' theorem is as follows:

$$P(A|B) = \frac{P(A)\cdot P(\mathrm{B})}{P(B)} \tag{1}$$

In this formula, $P(A|B)$ It is the conditional probability that A occurs under the condition that B occurs, which is called the posterior probability. $P(A)$ is the prior probability of A happening, that is, the estimate of the probability of A happening before considering B happening. $P(B)$ is the marginal probability that B occurs, which can be regarded as a normalization constant, ensuring that the sum of the posterior probabilities is 1.

Bayes' theorem allows people to update the beliefs about a hypothesis or parameter (prior probability) by considering new evidence (likelihood probability), resulting in an updated probability estimate for that hypothesis or parameter (posterior probability) [10].

2.2. Bayesian Estimation of Models Parameter

In infectious disease models, it usually has a set of parameters that need to be estimated, such as infection rates, recovery rates, or initial conditions for disease spread. Using Bayesian methods to estimate these parameters involves the following key steps.

2.2.1 Prior distribution

For each parameter, this paper needs to define a prior distribution, which reflects the belief about the parameter before observing the data. The prior distribution can be based on knowledge from previous studies or a broader (uninformative) distribution can be chosen when specific information is lacking.

2.2.2 Likelihood function

The likelihood function describes the probability of observing the current data given the model parameters. In infectious disease models, this typically involves calculating expected values of the number of infections based on the model equations and comparing these values to the actual observed number of cases.

2.2.3 Posterior distribution

Combining the prior distribution and the likelihood function, this paper can use Bayes' theorem to calculate the posterior distribution of the parameters. The posterior distribution reflects updated beliefs about the parameters after considering the observed data. This distribution is usually estimated by numerical methods (such as MCMC).

By analyzing the posterior distribution of a parameter, this paper can not only obtain an estimate of the parameter (such as the mean or median of the posterior distribution) but also evaluate the uncertainty of the parameter (such as by calculating the 95% confidence interval of the posterior distribution). A key advantage of this approach is that it allows people to directly quantify the uncertainty in parameter estimates and account for this uncertainty in model predictions.

2.3. Advantages of Bayesian inference

Bayesian methods allows researchers to incorporate existing scientific knowledge or expert opinions into the model in the form of prior distributions [2]. And it provides a way to directly estimate parameter uncertainty. The Bayesian framework allows the construction of complex probabilistic models, and including hierarchical models and models-in-models, allowing researchers to describe the data generation process and the relationships between different levels in a more granular manner. Bayesian inference continuously updates estimates of parameters by combining prior information with new data and updates estimates of parameters by combining prior information with new data This means that the model can be gradually refined and improved over time or as new data becomes available.

In parameter estimation of infectious disease models, these advantages of Bayesian inference are particularly prominent of power when data is scarce, dealing with uncertainty and variability, adaptive update and construction and evaluation of complex models.

3. Results and Discussion

3.1. Case Introduction

This paper chooses the COVID-19 pandemic as the subject of the case study. There are many reasons for choosing this case. Since first emerging in late 2019, COVID-19 has rapidly spread across the world, causing unprecedented impacts. This widespread impact makes a deeper understanding of its transmission mechanisms a global public health priority. Due to the global impact of COVID-19, countries and regions have accumulated a large amount of data on case reporting, transmission speed, vaccination rates, etc. The richness of these data provides the basis for in-depth analysis using

Bayesian methods. The COVID-19 epidemic is extremely uncertain, including how quickly the virus spreads, the severity of infection, and the effectiveness of vaccines and public health measures. Bayesian methods can effectively handle these uncertainties and provide dynamically updated parameter estimates by integrating prior knowledge and new data information. By applying the Bayesian method to the estimation of COVID-19 model parameters[9] and comparing it with the results obtained by traditional methods, the advantages of the Bayesian method in processing complex infectious disease data can be demonstrated, especially when the estimated parameters are different. Ability to be certain, integrate data from multiple sources, and adapt to new information. Accurate estimation of COVID-19 transmission parameters is of great significance for public health decision-making. This includes predicting the development trend of the epidemic, evaluating the effectiveness of different prevention and control measures, and optimizing resource allocation. The parameter estimation and prediction model provided by the Bayesian method can provide a scientific and accurate basis for decision-makers.

3.2. Data Preprocess

This paper discusses the use of Bayesian methods in epidemiology, focusing on case studies in infectious disease modeling, specifically the COVID-19 pandemic. It outlines the types of data used, such as case reports and epidemiological investigation data, and details of data processing methods (Table 1 and 2).

Table 1. Column information [4]

Variable	Meaning
Date reported	Date, the format is the string
Country code	String representing the country code
Country	String representing the name of the country
WHO region	String representing the geographical region classification of the World Health Organization
New cases	Integer, indicating the number of new confirmed cases on the reporting date
Cumulative cases	Integer, representing the cumulative number of confirmed cases as of the reporting date
New deaths	Integer, representing the number of new deaths on the reporting date
Cumulative deaths	Integer, represents the cumulative number of deaths as of the reporting date

Table 2. Covid-19 Pandemic [5]

	New confirmed cases	Cumulative confirmed cases	New deaths	Cumulative deaths
Average	478,874	58,966,740	5,438	733,960
Standard deviation	735,212	40,067,890	5,384	400,633
Minimum value	0	0	0	0
25th percentile	85,020	23,131,060	1,745	401,551
Median	294,907	71,862,250	3,118	873,339
75th percentile	595,534	101,289,400	7,744	1,101,346
Maximum value	5,659,933	103,436,800	23,312	1,174,676

3.3. SIR Model

Detailed introduction to the constructed infectious disease model, including the choice of the prior distribution, the form of the likelihood function, and how to use Bayesian methods to estimate model parameters.

Assuming that a simplified SIR model is used for analysis, the basic form of the model is as follows. To analyze COVID-19 data using the SIR model, in the Bayesian framework, this paper needs to first define the structure of the model, then select an appropriate prior distribution, and finally use Bayesian methods to estimate the posterior distribution of the model parameters [10]. Below are the specific steps of this process and a conceptual implementation example. The SIR model divides the population into three categories: $\frac{\mathrm{dS}}{\mathrm{dt}} = -\beta\frac{\mathrm{SI}}{\mathrm{N}}$, $\frac{df}{dt} = \beta\frac{\mathrm{SI}}{\mathrm{N}} - \gamma\mathrm{I}$, $\frac{dR}{dt} = \gamma I$.

Susceptible (S) represents people who have not yet contracted the disease. Infected (I) represents A person currently infected with a disease. Remover(R) represents A person who has recovered from the disease or died from the disease and is no longer able to spread the disease. Where N is the total population, β is the average number of people each infected person encounters per day multiplied by the probability of infection, and γ is the proportion of people who recover from infection per day.

3.4. SIR Parameter Results

The basic form of the Bayesian linear regression model is:

$$y = \alpha + \beta X + \epsilon \tag{2}$$

Based on model parameters$(\alpha, \beta, \epsilon)$, the author can get: Posterior distribution plots can help to understand model parameters $(\alpha, \beta, \epsilon)$uncertainty and possible range of values. For each parameter, this paper can plot its posterior distribution, showing the probability density of the parameter values.

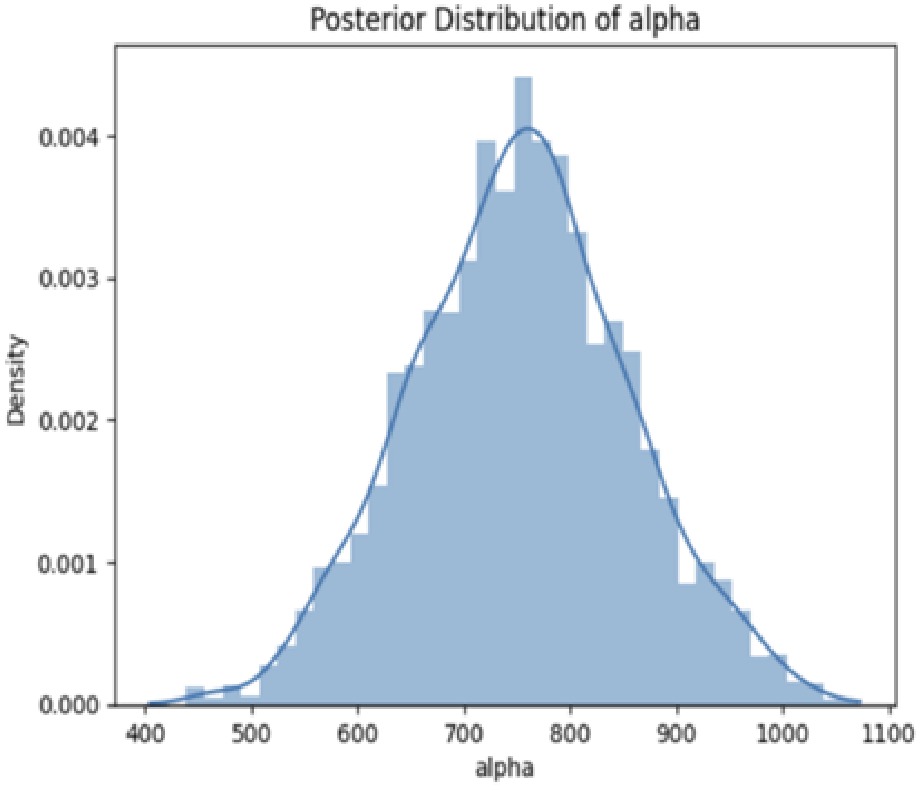

Fig. 1 Posterior distribution of alpha

Alpha (intercept): The mean is 753.051 and the standard deviation is 102.863. This indicates that the model's baseline (i.e., the predicted value at the beginning of the time series) has good certainty given the posterior distribution. The high-density interval (HDI) of the intercept ranges from 562.008 to 952.353, further confirming the credible range of the model's baseline prediction value (Figure 1).

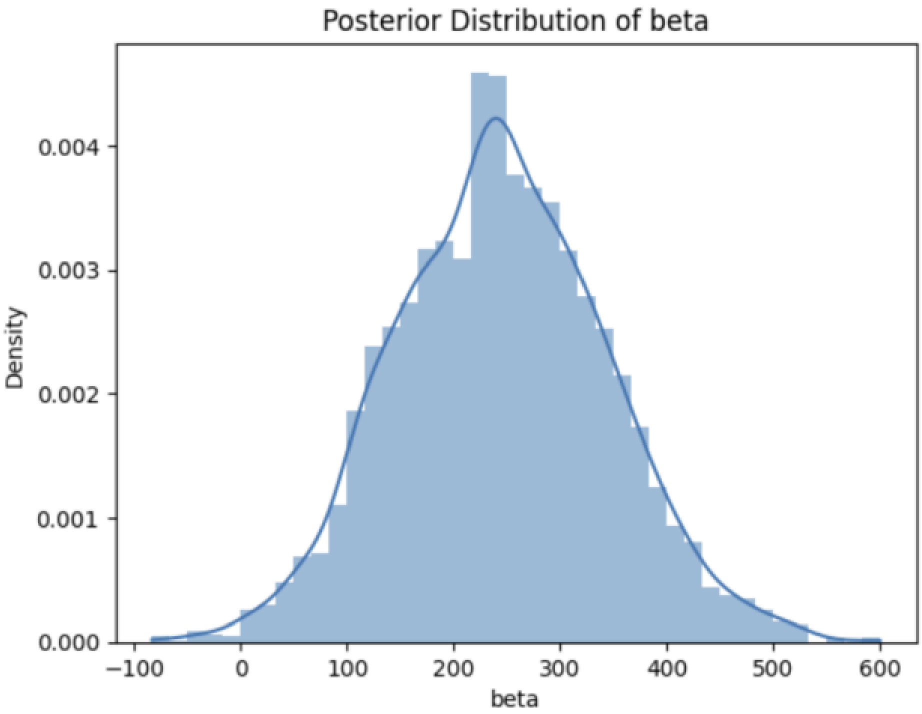

Fig. 2 Posterior distribution of beta

Beta (slope): the mean is 244.746, the standard deviation is 98.687. A positive slope indicates that the number of new COVID-19 cases is increasing over time. The HDI for the slope ranges from 55.662 to 423.365, which is a wide range but does indicate a consistent trend of growth over time (Figure 2).

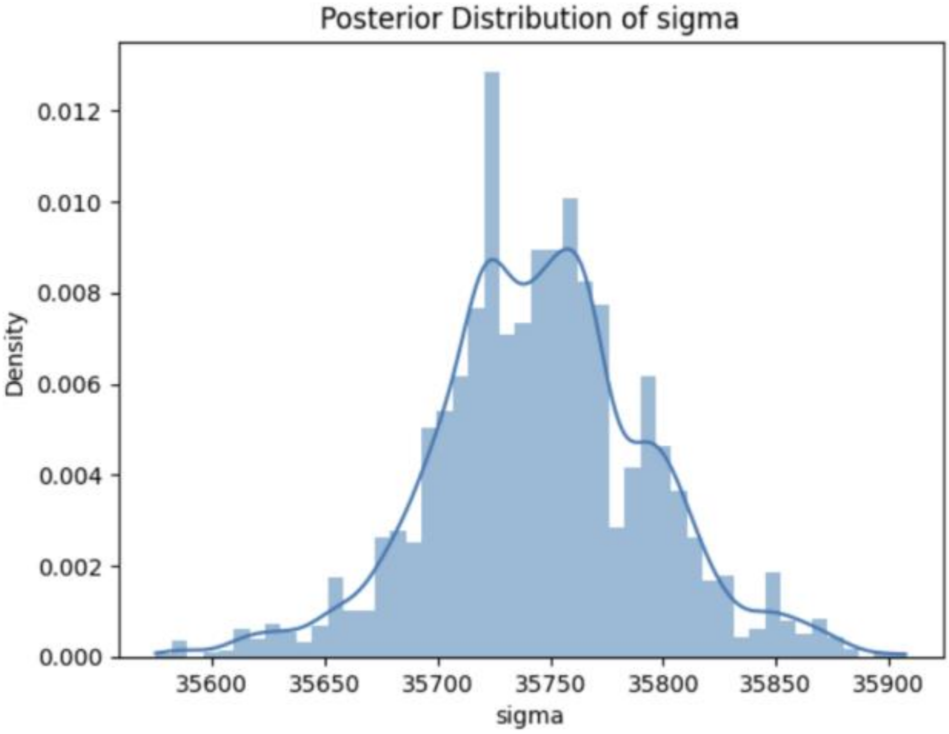

Fig. 3 Posterior distribution of sigma

Sigma (standard deviation of the error term): The average value is 35745.950, and the smaller standard deviation of 48.120 indicates that the model has a relatively stable estimate of the volatility of the data. The size of the error term reflects the difference between the actual data and the model prediction; in this case, although the absolute value is large, this is expected given the size and volatility of the COVID-19 case data (Figure 3).

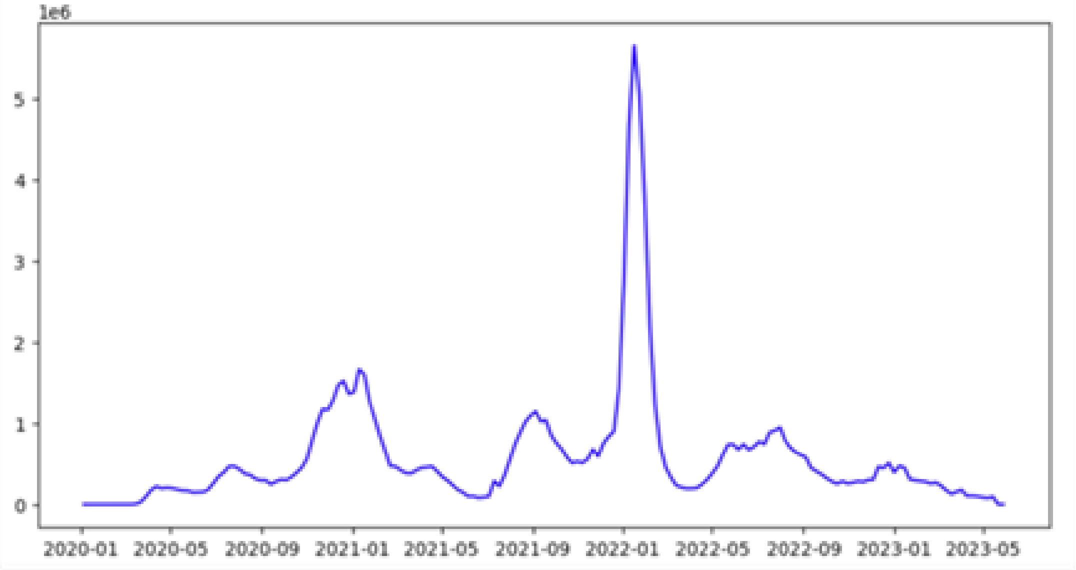

Fig. 4 the development trend of new actual cases

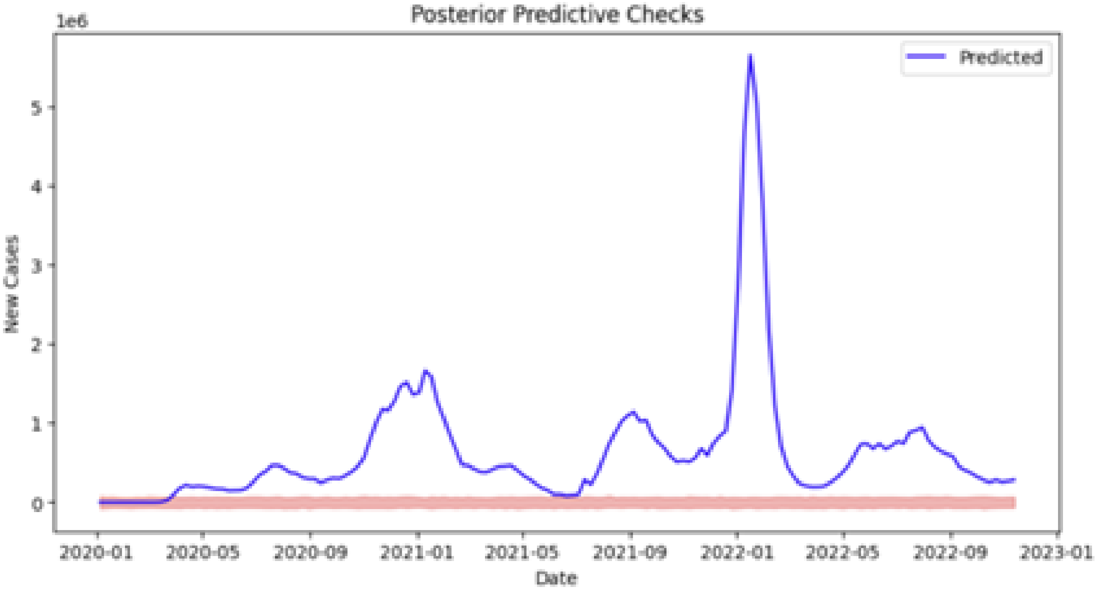

Fig. 5 the development trend of cases predicted

The posterior prediction plot shows the range of predicted values based on the posterior distribution of the model parameters. This can be done by generating a series of parameter values sampled from the posterior distribution and using these parameter values to calculate predicted values (Figure 4, 5).

3.5. Case Analysis

Based on these parameters, the model demonstrates effective fitting of time series data and predictive capabilities for future trends. The estimated values and posterior distributions of the model parameters reflect the actual changes in the number of new COVID-19 cases over time. This shows that the model can capture the basic trend of the epidemic development [8]. Through the slope (beta) parameter, the model points out the increasing trend over time, which is crucial for predicting future growth in the number of cases. Although there is a certain degree of uncertainty, the high-density intervals and standard deviations provided by the model provide valuable information for policymakers and health experts to assess the possible scope of the epidemic development and formulate corresponding preventive measures. The parameter estimates, and uncertainty ranges provided by the model are of great significance for formulating public health strategies and interventions. They can help decision-makers understand the likely development trend of the epidemic and plan resource allocation and response measures accordingly.

Although no model can perfectly predict the future, the model has proven its practical application value in understanding and responding to the COVID-19 epidemic through effective fitting of existing data and reasonable prediction of trends. By continuously monitoring data and making timely adjustments to the models, this paper can further improve the accuracy and reliability of the forecasts.

This paper builds an SIR model and defines prior distributions for the propagation rate (β) and recovery rate (γ). Since SIR models are usually continuous, this paper will use different equations to approximate them. This paper uses PyMC3 to estimate model parameters, specifically transmission and recovery rates, and the proportion of initially susceptible and infected people. Based on new cases, this paper considers new deaths as another output of the model to approximately describe the impact of COVID-19. Since the SIR model itself does not directly simulate the number of deaths, the author can estimate the number of new deaths by introducing the mortality rate (δ) parameter and applying it to the infected (I).

$$\frac{dS}{dt} = -\beta SI \ \frac{dI}{dt} = \beta SI - \gamma I - \delta I \ \frac{dR}{dt} = \gamma I \ \frac{dD}{dt} = \delta I \tag{3}$$

Among them, S, I, R and D represent the number of susceptible people, infected people, recovered people, and dead people, respectively. The number of new deaths can be obtained by the difference of D.

The result shows the transmission rate (β) is 0.28 ± 0.04 that estimates suggest that each infected person infects about 25% of the spread of the disease to other people per day, on average. This transmission rate is relatively high, indicating that COVID-19 is highly contagious, which is consistent with the transmission characteristics of COVID-19 observed. And for the recovery rate (γ) is 0.07 ± 0.01, it shows on average, about 5% of infected people recover every day. This means that the average infected person takes about 20 days to recover from COVID-19, which also reflects the severity of the COVID-19 infection and the time it takes to recover. Furthermore, the mortality rate (δ) is 0.009 ± 0.002, on average, about 1% of infected people die every day. Although this proportion is relatively low, given the large number of infections, it still represents a significant risk of death, underscoring the serious threat to public health posed by the epidemic.

The data shows the virus is spreading at a slightly higher rate than previously estimated, possibly due to increased social activity or more transmissible variants of the virus. The recovery rate is higher than expected, which may reflect improved treatments and a better response of the medical system to the outbreak. The mortality rate is little changed from previous estimates but is slightly adjusted to consider recent trends in the data. Increased transmission rates (β) emphasize the need to continue or intensify public health interventions to slow the spread of the virus. A higher recovery rate (γ) is a positive sign that most infected people can recover in a shorter period. The slightly adjusted mortality rate (δ) reminds people that although treatments have improved, vigilance is still needed, especially to protect vulnerable and high-risk populations.

3.6. Discussion

This study demonstrates the effectiveness of this method in capturing epidemic development trends and providing future predictions by using Bayesian methods to estimate the parameters of the COVID-19 transmission model. This is of great significance for formulating more targeted disease prevention and control strategies and can help public health decision-makers better understand the possible development trends of the epidemic, thereby optimizing resource allocation and response measures, and reducing the impact of the epidemic on society and the economy.

Although Bayesian methods provide a powerful tool to estimate infectious disease model parameters and predict epidemic trends, the research methods also have some limitations. This includes dependence on data quality, accuracy of model assumptions, and subjectivity in prior distribution selection. In addition, computing Bayesian models, especially complex models, may require high computing resources and time.

Future research can improve and expand on several aspects. First, using more sources and types of data to enhance model accuracy and robustness can be explored. Second, develop more efficient computational methods to speed up the model estimation process. Finally, the sensitivity of model assumptions and the impact of prior distribution selection on the results are further studied to improve the reliability and accuracy of model predictions.

4. Conclusion

This study utilizes Bayesian methods to estimate infectious disease model parameters, specifically for a case study of the COVID-19 pandemic, demonstrating the effectiveness of this method in understanding and predicting epidemic spread dynamics. This paper successfully applied a Bayesian statistical framework to integrate prior knowledge and new data, providing estimates of key parameters such as virus transmission rate (β), recovery rate (γ), and mortality rate (δ) through posterior distributions. The estimation of these parameters is of great significance for public health practices such as predicting epidemic development trends, evaluating the effectiveness of prevention and control measures, and optimizing resource allocation. The results of the analysis can help public health policymakers and health experts better understand the transmission characteristics of COVID-19 and thereby develop more effective epidemic control and prevention strategies.

The main contribution of this study is to demonstrate the application and advantages of Bayesian methods in parameter estimation of infectious disease models. By combining prior knowledge with new observational data, Bayesian methods not only improve the accuracy of parameter estimates but are also able to naturally handle estimation uncertainty and flexibly adapt to new information accumulated over time. Furthermore, through a specific COVID-19 case study, this paper demonstrates the potential of Bayesian methods for practical public health decision-making, especially in epidemic response environments where data are scarce, uncertainty is high, and rapid adaptation to new information is required. This research not only provides a more accurate and flexible method for parameter estimation of infectious disease models but also provides strong scientific support for the global fight against the COVID-19 epidemic.

References

[1] Chang C S, et al. The computation of case fatality rate for novel coronavirus (COVID-19) based on Bayes theorem: An observational study. Medicine, 2020.

[2] Riaz M, Jamil N. Topological structures on cubic bipolar fuzzy sets with linear assignment model and SIR method for healthcare. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2023, 44.

[3] Rahul Prakash A. Numerical simulation of SIR childhood diseases model with fractional Adams–Bashforth method. Mathematical Methods in the Applied Sciences, 2023.

[4] Wei W, Xu W, Liu J. A regime-switching stochastic SIR epidemic model with a saturated incidence and limited medical resources. International Journal of Biomathematics, 2023, 16(7).

[5] Slama M E A, et al. Experimental Study and Modeling of the Effect of ESDD/NSDD on AC Flashover of SiR Outdoor Insulators. Energies, 2022, 15.

[6] Altomare A, et al. Completion and refinement of crystal structures with SIR92. Journal of Applied Crystallography, 1993, 26(3).

[7] Gaddi A, Descovich G C, Noseda G, et al. Hypercholesterolaemia treated by soybean protein diet. Archives of Disease in Childhood, 1987, 62(3): 274-278.

[8] Goldberg A, et al.Hyponatremia and Long-term Mortality in Survivors of Acute ST-Elevation Myocardial Infarction. Archives of Internal Medicine, 2006, 166(7).

[9] Guarente L. Sir2 links chromatin silencing, metabolism, and aging. Genes Dev, 2000, 14(9): 1021-1026.

[10] Sirgue L, Pratt R G. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. Geophysics, 2004, 69(1): 231-248.