Proceedings of the 2023 International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/28/20230400

Using logistic function to predict epidemic trend of COVID-19 in China

Yihao Xu

Department of mathematics, The Chinese University of Hong Kong, 999077, Hong Kong

1155173913@link.cuhk.edu.hk

Abstract. As the novel coronavirus continues to spread and mutate, there has been growing concern over public health. Multiple measures have been enacted to mitigate the transmission of the disease, resulting in varying infection scenarios across different countries. To achieve timely and effective control of the epidemic, we note that predicting the future course of an epidemic plays an important role. The logistic function, a continuous-time demographic model, may be a suitable mathematical tool for estimating the trend of the epidemic. This paper aims to evaluate the accuracy of the logistic map in estimating the future trend of the COVID-19. We collect the most recent COVID-19 epidemiological data prior to January 30, 2023, and subsequently integrate figures into the curve fitting tool in MATLAB to generate an epidemic curve. By comparing the actual numbers and the predicted figures, the accuracy of logistic map can be properly assessed.

Keywords: logistic function, infectious disease, COVID-19, modeling, prediction.

1. Introduction

The COVID-19 pandemic, also known as the coronavirus pandemic, is a continual worldwide disease outbreak caused by the SARS-CoV-2 virus, resulting in severe acute respiratory syndrome [1]. Along with a diverse set of symptoms like fever, cough, fatigue, and difficulty breathing, the virus can induce anxiety, sleep disorders, cognitive decline, and has substantial effects on the well-being, mental health, and employment of individuals experiencing long COVID [2].

As of 11:31am CET on 16 March 2023, the World Health Organization has reported 760,360,956 confirmed cases of COVID-19 globally, with 6,873,477 deaths [3]. Despite strict efforts to control it since its initial identification, the COVID-19 disease has now become a global pandemic, presenting a considerable threat and challenge to both global health and the economy [1,4]. Moreover, as the virus continues to spread, new variants have emerged with varying degrees of virulence and infectivity, posing a challenge to efforts to control the pandemic [5]. Therefore, to determine whether the pandemic will reach a peak or diminish, the trajectory of the global pandemic plays a pivotal role, as it motivates the accurate forecasting of the pandemic's infection trends.

Over time, different predictive models have been developed and utilized to forecast and anticipate the trends and patterns of COVID-19. For instance, researchers in Spain used ARIMA (Autoregressive Integrated Moving Average), a mathematical statistical model, to forecast the number of hospitalizations and deaths due to COVID-19 and found that their models had good predictive performance [6]. An additional illustration could be provided by the implementation of machine learning models. In India, scholars utilized machine learning models to anticipate the number of COVID-19 infections and fatalities in their nation, and their models showed reliable forecasting accuracy [7]. Furthermore, infectious disease models such as the SIR (susceptible-infected-removed) model were used to forecast the impact of different interventions such as quarantine measures and contact tracing on the spread of COVID-19, and it was found that these interventions were effective in reducing transmission [8]. It's important to note that the accuracy of predictions based on mathematical models depends on the ability of the models to capture the complexity of the epidemic and the quality of the available data.

The logistic function is a first-order ordinary differential equation which was first introduced to represent natural population growth and species diffusion [9]. The equation describes the continuous growth of a population in terms of its current size and the rate of growth. It is commonly used in regression fitting primarily because of its simple formula and efficient computational capabilities.

The logistic function has been applied in various fields such as ecology, medicine, economics and sociology. A typical application could be given by A. G. McKendrick, who experimentally tested the equation in modeling the growth of bacteria in broth by utilizing an approach to accurately estimate and validate the parameters involved in the logistic equation[10]. The logistic equation could also provide a useful framework for understanding and forecasting the spread of infectious diseases. Yang et al. [11] conducted a comparative analysis of the disease fitting capabilities between the logistic differential equation (LDE) model and the generalized logistic differential equation (GLDE) model and established a comprehensive disease early warning system for different types of infectious diseases.

This study will mainly discuss the models based on the logistic function. We will construct a model for the epidemic trend of COVID-19 in China by inputting integrated data to the curving fitting tool in MATLAB. The predictive capability of the logistic function enables proactive measures to be taken, such as implementing timely interventions, allocating healthcare resources, and adjusting public health strategies. By leveraging the valuable information derived from the logistic function, policymakers can make informed decisions to mitigate the spread of the virus, protect public health, and minimize the societal impact of the epidemic.

2. Methodology

The logistic function is a mathematical model that is often used to study the dynamics of populations, including the spread of infectious diseases. It is a simple, non-linear equation that can capture the behavior of complex systems.

This experiment will record the weekly number of fresh COVID-19 cases in China for a period of 13 weeks starting from November 7, 2022. And by adopting the mathematical estimation constructed by logistic map, another set of data can be calculated. If the error between the two sets of the data is acceptable, then in this case, the logistic map estimation can be applied to predict the future trends.

Chen et al. [12] conducted a fitting analysis of the daily cumulative number of reported COVID-19 cases in Hubei, China. Figure 1 [12] illustrates that the findings consistently demonstrated an S-shaped curve pattern, effectively captured by employing a logistic fitting method.

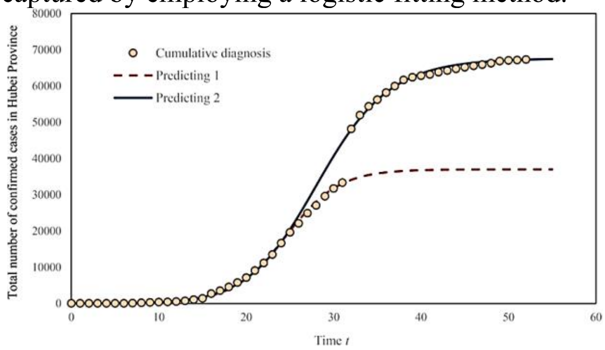

Figure 1. The sigmoid growth curve of total number of confirmed cases of COVID-19 in Hubei Province [12].

2.1. Data collection

The latest updated information on the global COVID-19 outbreak is obtained from WHO, which provides a country-level dashboard [3]. Consequently, we collected data on the COVID-19 case count from November 7, 2022, to January 30, 2023, for periodic analysis.

Before November 7, 2022, there have been 9,159,106 confirmed cases of COVID-19 reported to WHO [3]. We set the week of November 7 to be Week 0, and the week of January 30 to be Week 12. The weekly increment and the number of accumulated confirmed cases can be observed in Table 1 and Figure 2 intuitively.

Table 1. The weekly increment and the number of accumulated confirmed cases from Week0 to Week12.

Week	Weekly increment	Accumulate cases
0	167,652	9,326,758
1	157,203	9,483,961
2	142,745	9,626,706
3	146,919	9,773,625
4	1,845,892	11,619,517
5	14,419,505	26,039,022
6	41,174,900	67,213,922
7	21,483,225	88,697,147
8	6,982,790	95,679,937
9	1,975,433	97,655,370
10	658,339	98,313,709
11	158,680	98,472,389
12	188,554	98,660,943

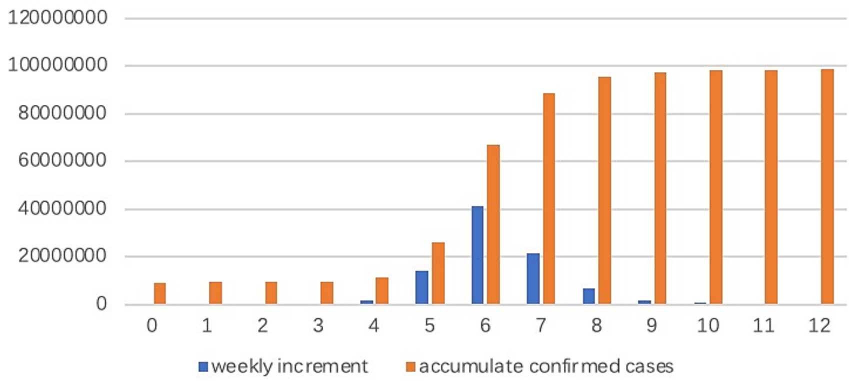

Figure 2. The histogram of China situation.

2.2. Logistic function

The logistic equation was first introduced by Pierre François Verhulst in 1838 [13]. The function is a generalization of the equation for exponential growth but with a ceiling on population size, and the differential equation is given by

$$\frac{dP}{dt} = rP(1 - \frac{P}{K}) \tag{1}$$

In this equation, r represents the intrinsic growth rate, K represents the carrying capacity of the environment, and P represents the population size. As time progresses, the value of P changes, leading to the emergence of an S-shaped curve in the logistic equation.

There are many previous examples using logistic map to predict the trends of the COVID-19. For instance, in 2020, Wang et al. [14] suggested a forecasting strategy using the Prophet and logistic models to analyze the COVID-19. By identifying the fastest-growing point, the logistic model fits the cap value, which is subsequently sent into the prophet model for forecasting. Studies were carried out to forecast the peak of the pandemic, the point at which it grew most quickly, and the time at which recovery would turn. The results showed how well the model predicts the COVID-19 epidemic turning point and size by plotting predicted trends of the world and five specific nation.

Another example could be given by Giuseppe Consolini and Massimo Materassi [15]. The researchers proposed a relatively simple enhancement to the logistic dynamics used for modeling the spread of a pandemic. This enhancement incorporated a power-law dependence on the duration of infection, which may result from at least two separate methods, including the simple elimination of interpersonal connections and/or the isolation of afflicted individuals. The findings clearly indicated the necessity for additional theoretical and numerical investigation, as well as for applying the same methodology to the spread of COVID-19 in other countries.

From Figure 1 and Figure 2, we can observe that the accumulated number of COVID-19 is S-shaped, which can support the hypothesis that the epidemic trend of COVID-19 could be well fitted into the logistic function. Initially, during the outbreak, when precautionary measures were not stringent and the number of infected individuals was low, the rate of infection increased slowly. However, as the proportion of infected individuals reached a certain threshold, the spread of the disease witnessed a rapid and exponential growth, as evidenced by the escalating numbers. Subsequently, through government regulations and public cooperation, the spread of the disease gradually slowed down, eventually reaching the maximum number of cumulative infections.

One of the key observations is the turning point in the cumulative curve, indicating a shift from rapid to slower increases in the number of cases. This turning point signifies a transition in the trajectory of the cumulative curve and is a significant milestone in the progression of the outbreak. If the dataset encompasses this turning point and a subsequent time interval, predictions of future case numbers can be conducted with a reasonable level of accuracy.

2.3. Parameter estimation

By solving Eq. (1), we have

$$P(t) = \frac{K}{1+ce^{-Krt}} \tag{2}$$

where c is a constant. To find out the turning point, we calculate the second derivative of $P(t)$, which is given by

$$\frac{d^2\mathrm{P}}{dt^2} = r^2\mathrm{P}(1-\frac{\mathrm{P}}{\mathrm{K}})(1-\frac{2\mathrm{P}}{\mathrm{K}}) \tag{3}$$

Note that $P(t)$ starts with $P(0) = \frac{K}{1+c}$. By assumption, P is a positive number between $\frac{K}{1+c}$ and K, therefore $\frac{d\mathrm{P}}{d\mathrm{t}}$ increases before $P(t_*) = \frac{K}{2}$ and decreases after this turning point, where $t_* = \frac{\ln c}{Kr}$ represents the time point indicating a transition from rapid increases in the number of cases to slower increases.

It's worth noting that prior to Week 0, a substantial number of cases had already been accumulated, which cannot be neglected. Therefore, we decide to improve the model by adding another parameter h. For convenience, we further set a=Kr, hence $t_* = \frac{\ln c}{a}$ and the revised version is given by

$$P(t) = \frac{K}{1+ce^{-at}} + h \tag{4}$$

In this article, we will choose Eq. (4) to capture the dynamics of the epidemic. We first plot the number of cases over time using a graph, which can provide a visual display of how the number of cases is changing over time. Then we utilize the Nonlinear Least Squares method [16] by using MATLAB to fit the logistic model to the data. Subsequently, we can employ the logistic function to model the progression of the virus and estimate the parameters K, c, a, and h. With the derived parameters obtained from the logistic growth model, we can make predictions about the future trend of the epidemic and the maximum number of cases within a specific time interval.

Different countries have implemented various measures to mitigate the transmission of the disease, resulting in varying infection situations [17]. For instance, COVID-19 growth in China has already stabilized [18], whereas India is still experiencing an upward trend [19]. Hence, the initial task is to determine whether the virus is still undergoing rapid growth or has transitioned beyond the exponential growth phase, approaching the maximum number of cases. If the current day is not greater than t_*, it signifies that the turning point is yet to be reached, and the growth may still be in an exponentially increasing phase. Otherwise, it indicates that growth is approaching its endpoint and the spread of the virus has been contained.

3. Results

Since the number of accumulated confirmed cases is quite large, we decide to represent it in the form of ×10^6, and by plotting the statistics on MATLAB, we obtain Figure 3.

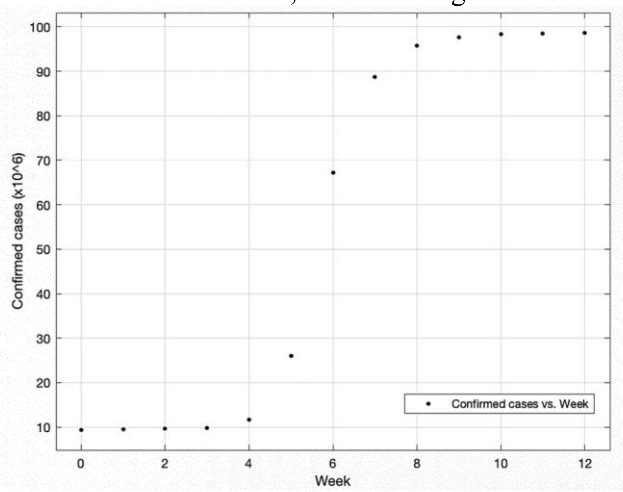

Figure 3. Confirmed cases (×10^6) vs. Week.

Four parameters are utilized in this estimation. We set the upper bound for K and h to be 100×10^6 and 10×10^6, respectively. Applying logistic model to the curve fitting tool, we obtain Figure 4 and the Table 2. There is a remarkable congruence observed between the actual number of confirmed cases and the predicted curve during the period from November 7 to January 30. This congruence highlights the ability of the model to accurately capture the dynamics of the epidemic and generate reliable predictions.

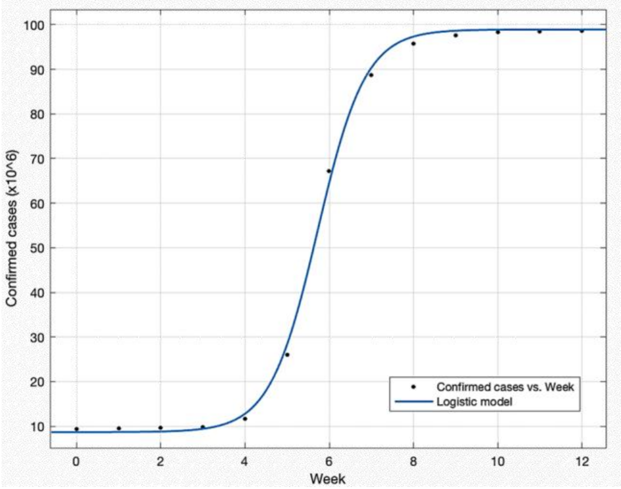

Figure 4. Fitting curve of the logistic function generated by MATLAB.

DOI: 10.54254/2753-8818/28/20230400

Table 2. The estimated values.

	K (×10^6)	*c*	*a*	*h* (×10^6)
Estimated value	90.2177	24400	1.7650	8.7995

Subsequently, we employed the generated model to produce dynamic forecasts 3 weeks ahead, spanning from February 7, 2023, to February 20, 2023.

Table 3. Comparison between the actual number of confirmed cases and the predicted curve.

Week	Actual number	Predicted number	Relative error (‰)
13 (Feb 6)	98,809,986	98,916,961	1.083
14 (Feb 13)	98,904,475	98,917,159	1.282
15 (Feb 20)	98,982,145	98,917,193	0.656

As is shown in Table 3, the predicted values are quite close to the actual numbers and relative errors are acceptable. From the figures and tables above, we can discover that the turning point has already passed and the total size of the COVID-19 epidemic, which can be represented as K+h, is projected to reach 99.0172×10^6. These findings contribute to the scientific understanding of the epidemic dynamics and provide valuable insights for public health planning and interventions.

4. Conclusion

In this paper, we used the logistic function to model the pandemic trend of COVID-19 in China. The estimated curve generated by MATLAB was applied to predict the weekly increment and the total epidemic size. The Results in Section 3 demonstrated that the logistic map provides a valuable and accurate tool for understanding and predicting the epidemic trend, as showcased by the plotted predictive trend.

However, it's essential to note that the precision of the predictions relies on the validity of the assumptions underlying the models. For instance, when using the logistic function, we have already assumed that there will be a maximum outbreak size. However, in the real world, achieving complete eradication of the virus is a challenging task, making the exact prediction of the total epidemic size an elusive goal. Also, this approximation is valid if the growth rate is continuous and smoothly varying. When the growth rate is subject to abrupt changes or other nonlinearities, a discrete map may be a more accurate and efficient. Furthermore, factors such as the efficacy of public health measures and the emergence of new variants should be also considered. Therefore, it's essential to interpret the predictions with caution and to update the models regularly as new data becomes available. By acknowledging these limitations and considering the dynamic nature of the pandemic, we can use the logistic function as a guide while incorporating other relevant factors to improve our understanding of the course of epidemic and make more informed decisions in managing the outbreak.

References

[1] Al-Rohaimi, A. H., & Al Otaibi, F. (2020). Novel SARS-CoV-2 outbreak and COVID19 disease; a systemic review on the global pandemic. *Genes & diseases, 7*(4), 491–501. https://doi.org/10.1016/j.gendis.2020.06.004

[2] Aiyegbusi, O. L., Hughes, S. E., Turner, G., Rivera, S. C., McMullan, C., Chandan, J. S., Haroon, S., Price, G., Davies, E. H., Nirantharakumar, K., Sapey, E., Calvert, M. J., & TLC Study Group (2021). Symptoms, complications and management of long COVID: a review. *Journal of the Royal Society of Medicine, 114*(9), 428–442. https://doi.org/10.1177/01410768211032850

[3] WHO Coronavirus (COVID-19) Dashboard. (n.d.). WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. https://covid19.who.int/

[4] Naseer, S., Khalid, S., Parveen, S., Abbass, K., Song, H., & Achim, M. V. (2023). COVID-19 outbreak: Impact on global economy. *Frontiers in public health, 10*, 1009393. https://doi.org/10.3389/fpubh.2022.1009393
[5] Otto, S. P., Day, T., Arino, J., Colijn, C., Dushoff, J., Li, M., Mechai, S., Van Domselaar, G., Wu, J., Earn, D. J. D., & Ogden, N. H. (2021). The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Current biology : CB, 31*(14), R918–R929. https://doi.org/10.1016/j.cub.2021.06.049
[6] Carrión-García, A., Jabaloyes, J., & Grisales, A. (2021). The growth of COVID-19 in Spain. A view based on time-series forecasting methods. *Data Science for COVID-19*, 643–660. https://doi.org/10.1016/B978-0-12-824536-1.00020-4
[7] Hota, H. S., Handa, R., & Shrivas, A. K. (2021). COVID-19 pandemic in India: Forecasting using machine learning techniques. *Data Science for COVID-19*, 503–525. https://doi.org/10.1016/B978-0-12-824536-1.00030-7
[8] Cooper, I., Mondal, A., & Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, solitons, and fractals, 139*, 110057. https://doi.org/10.1016/j.chaos.2020.110057
[9] J.S. Cramer (2004). The early origins of the logit model. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 35*(4), 613-626. https://doi.org/10.1016/j.shpsc.2004.09.003.
[10] McKendrick, A. G., and Pai, M. K. (1912). XLV.—The Rate of Multiplication of Micro-organisms: A Mathematical Study. *Proceedings of the Royal Society of Edinburgh, 31*, 649–655. http://doi.org/10.1017/S0370164600025426
[11] Yang, T., Wang, Y., Yao, L., Guo, X., Hannah, M. N., Liu, C., Rui, J., Zhao, Z., Huang, J., Liu, W., Deng, B., Luo, L., Li, Z., Li, P., Zhu, Y., Liu, X., Xu, J., Yang, M., Zhao, Q., Su, Y., ..Chen, T. (2022). Application of logistic differential equation models for early warning of infectious diseases in Jilin Province. *BMC public health, 22*(1), 2019. https://doi.org/10.1186/s12889-022-14407-y
[12] Chen, Y., Li, Y., Feng, S., Man, X., & Long, Y. (2021). Gravitational scaling analysis on spatial diffusion of COVID-19 in Hubei Province, China. *PloS one, 16*(6), e0252889. https://doi.org/10.1371/journal.pone.0252889
[13] Bacaër, N. (2011). A Short History of Mathematical Population Dynamics. *Springer*, London. https://doi.org/10.1007/978-0-85729-115-8_6
[14] Wang, P., Zheng, X., Li, J., & Zhu, B. (2020). Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, solitons, and fractals, 139*, 110058. https://doi.org/10.1016/j.chaos.2020.110058
[15] Consolini, G., & Materassi, M. (2020). A stretched logistic equation for pandemic spreading. *Chaos, solitons, and fractals, 140*, 110113. https://doi.org/10.1016/j.chaos.2020.110113
[16] David, P., & Peter, R. (2006). Nonlinear least-squares estimation. *Journal of Multivariate Analysis, 97*(2), 548-562. https://doi.org/10.1016/j.jmva.2005.04.002
[17] Rajan, S., McKee, M., Hernández-Quevedo, C., Karanikolos, M., Richardson, E., Webb, E., & Cylus, J. (2022). What have European countries done to prevent the spread of COVID-19? *Lessons from the COVID-19 Health system response monitor. Health policy (Amsterdam, Netherlands), 126*(5), 355–361. https://doi.org/10.1016/j.healthpol.2022.03.005
[18] Zhu, S., Feng, S., Ning, X., & Zhou, Y. (2022). Analysis of China's fight against COVID-19 from the perspective of policy tools-policy capacity. *Frontiers in public health, 10*, 951941. https://doi.org/10.3389/fpubh.2022.951941
[19] Patel, R. B., & Patel, B. B. (2022). An Analysis of the COVID-19 Situation in India in Terms of Testing, Treatment, Vaccine Acceptance and National Economic Performance. *International journal of public health, 67*, 1604975. https://doi.org/10.3389/ijph.2022.1604975