BIOGLOBIA Vol. 7 (1): 31 –37, 2020 **ISSN: 2349-5626**

Prediction Strategy for the Total Confirmed Cases of COVID-19 in India

HIMANSHU SHEKHAR

Sc 'G', Director (Admin), O/O DG(ACE), Pashan, Pune-411021, Maharashtra, India
Author's e-mail: himanshudrdo@rediffmail.com
Received: 02.06.2020 **Revised: 21.06.2020** **Accepted:23.06.2020**

ABSTRACT

Total or cumulative confirmed cases for COVID-19 in India is progressive addition of daily confirmed cases. Both daily and total confirmed cases are rising at a very high rate and in this paper, attempt made earlier to predict the confirmed cases from available data for 10-15 days ahead is compiled. The exponential variation diverges as overestimated values after 10-15 days and long duration prediction could not be ascertained. In addition to this, exponential variation does not have any method to predict the turnaround or lowering of curve or logistic variation. So, normal distribution curve with chi-square minimization is fitted to the available data and attempt is made to match the results up to 21.06.2020. Using this match, extrapolation is made up to 15.07.2020, which predicts a value of 8.50 lakhs by 15.07.2020. Another attempt is also made to fit a linear variation using data from 01.06.2020 to 21.06.2020. This is also extrapolated to 15.07.2020 to give a total confirmed case of 6.60 lakhs. It is predicted that by 15.07.2020, total confirmed cases in India will be between 6.60 lakhs and 8.50 lakhs.

Keywords: *Pandemic, COVID-19, Modelling, Prediction, Normal Distribution, Growth Period, Doubling Period, Mean, Amplitude, Turnaround*.

INTRODUCTION

COVID-19 has gripped the world and is spreading at such a speed with more potency that prevention, mitigation and control measures are falling flat. Although recoveries are reported, but the pandemic is still not under control. The current article is also an attempt to predict the spread of COVID-19 in India from the available statistical data and tools.

Global Prediction Attempts

Many researchers tried to understand the pattern of spread, based on available data. In China one model is developed to map the spread of COVID-19 and the same is assumed to be accurate enough for both backward confirmation and forward extrapolation (Lixiang *et al*., 2020). In yet another article, mathematical tools are exploited to understand the behaviour of infections and establishment of prediction tools in case of epidemics (Julien, 2020). The prediction strategy is also extended to spread of COVID-19 in Western World and major parameters like turning point, attack rate and durations are captured through modelling (Xiaolei *et al*., 2020). The paper discusses 6 countries including France, USA, Canada, Italy, Germany and UK. It is predicted that a turning point of 69 days and recovery by first week of June 2020. Overall, data available is modelled by various exponential factors and the same is used to predict the nature of disease in future. However, a realistic and deterministic prediction strategy in Indian Context is missing and is not available. This paper is an attempt to predict the outburst of COVID-19 in India using exponential variation, normal distribution and linear variation. The data available for every 30 days from 15.03.2020 at an interval of 15 days is analysed and attempt is made to establish a prediction strategy for next 15 days. The variation of total confirmed cases as exponential variation is assumed in most of the cases.

Normal distribution, with chi-square optimization is applied for the assessment of turnaround in daily confirmed cases, which can augment the prediction strategy.

Prior Prediction Strategy

Initially, for the data of pandemic in India from 15.03.2020 to 30.04.2020 is considered. Daily confirmed cases are assumed to be following exponential cumulative confirmed cases are derived from the daily confirmed cases (Shekhar, 03 May 2020). The prediction made on 30.04.2020 is reproduced, as Fig. 1. The actual number of cases are monitored at regular interval and for different stated values of total confirmed cases (Shekhar, 05 May 2020 and 10 May 2020). The actual cases were matching to the prediction till 10.05.2020, but deviation started from 15.05.2020 (Shekhar, 15 May 2020) and actual number of confirmed cases were found to be lower than prediction using exponential variation. The deviation continued further and it demanded revising the prediction strategy and numerical values.

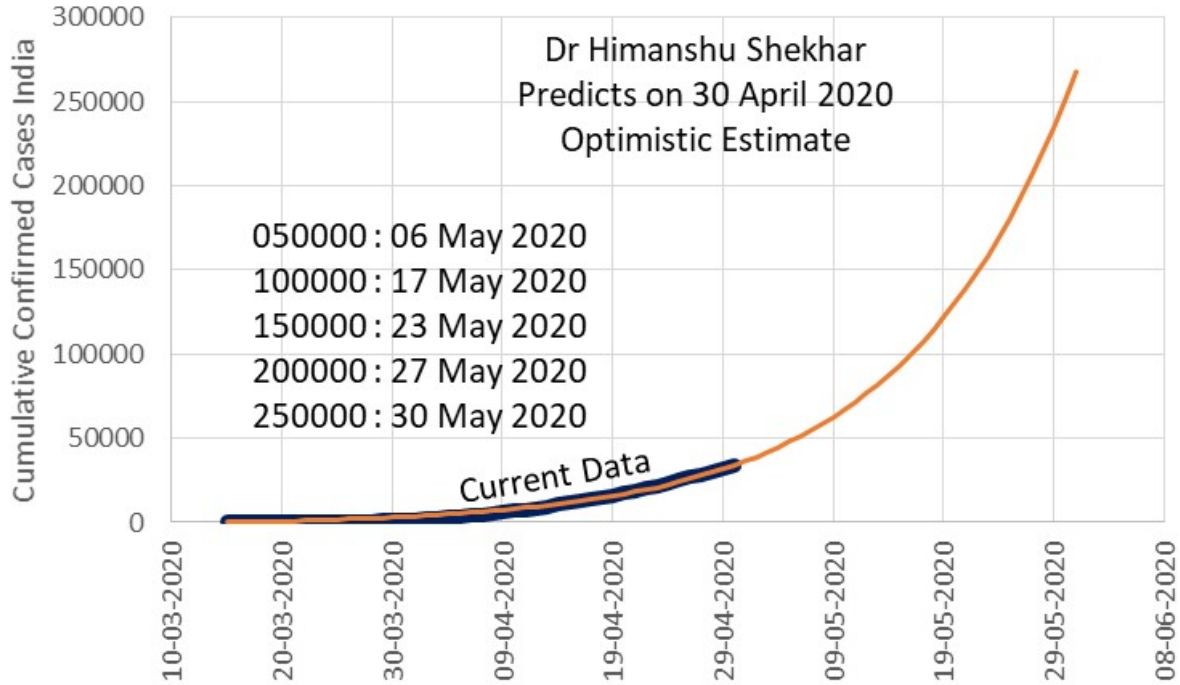

Fig. 1. Prediction Made on 30.04.2020 using 45 days data

A fresh prediction is made on 20.05.2020, after India crossed the dubious distinction of 1.0 lakh confirmed cases (Shekhar, 21 May 2020). The input data available now was from 15.03.2020 to 20.05.2020 and fresh prediction was made again as given in Fig. 2. It was expected to be followed for next 15 days again. In fact, the curve was again an exponential distribution and the variation matched the actual data for next 15 days, but again, the prediction started giving overestimated values as compared to the actual number of total confirmed cases. This required a fresh prediction again. The assumption of exponential distribution was always deviating from actual cases after some time and every time, it resulted in an overestimated values. Clearly, the variation is not at all exponential and piece-wise linear variation can be used for prediction. In addition to this normal distribution curve with chi-square minimization of error can also be applied to predict a turnaround of daily number of cases.

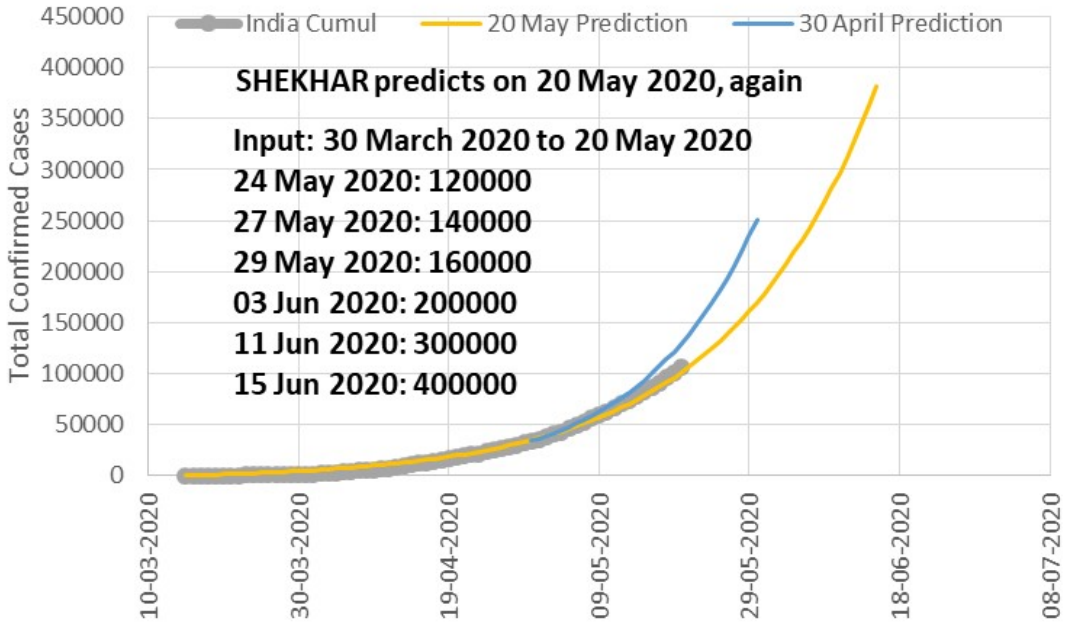

Fig. 2. Fresh Prediction made on 20.05.2020 for Confirmed Cases in India

Mathematical Formulation

The daily confirmed cases of the pandemic add to cumulative confirmed cases. The ratio of daily confirmed cases on a particular date to that at the previous date gives the growth factor. The value of growth factor should be less than 1 for quick control. The time taken to double the total confirmed cases is called doubling period and this is also one of the indicator of the rapidity of rise of confirmed cases. If doubling period is less, rapid rise is contemplated.

Normal distribution curve is a bell-shaped curve with symmetric limbs on either side of a peak value. For normal distribution curve, generally two specified parameters mean and standard deviation are there. Mean value indicates the location of peak along horizontal axis and standard deviation is a measure of spread of the curve on either side. Most of the normal distribution curve is normalized to a peak value of 1, so additional term called amplitude is used for simulating the daily confirmed cases.

If, Daily confirmed cases = C, Mean = H, Standard Deviation = S, Amplitude = G, Time in days lapsed from a reference date = t, then the mathematical formulation is shown below.

$$C = G\ x\ e^{-\frac{(t-H)^2}{S}}$$

The reference date for all calculation is taken as 31.12.2019, so that 01.01.2020 is 1 for variable t. A quick description of the nature of such variations is illustrated in book entitled 'India May COVID Math' (Shekhar, 2020). If actual daily confirmed cases is 'A', then chi-square is defined as $(C-A)^2/A$. For any arbitrary value of 'H', 'S' and 'G', chi-square is minimized and the best fit curve is obtained, so that prediction strategy can be made and location of peak and then reduction in daily confirmed cases can be predicted. Cumulative confirmed cases are also calculated from daily confirmed cases, as predicted values of confirmed cases at any time instance.

Similarly linear variation is applied to last 30 days and optimized using 0.999 value of regression coefficient. Based on the best fit curve for 30 days, prediction for next 15 days is attempted to establish the effectiveness of the method.

Data Analysis

For the calculation, data of total and daily confirmed cases till 21.06.2020 in India is taken. The daily confirmed cases are obtained and growth factor is calculated. The growth factor is plotted against time as Fig. 3. The value is mostly oscillating around 1 and it has an average value of 1.030729 for the last 21 days. In June 2020, the values are mostly above 1. Similarly, doubling period is also calculated for the total confirmed cases in days. The doubling period is continuously increasing and on 21.06.2020, it has a value of 19 days (Fig. 4). The continuous rise in doubling period, indicates that there is some controlling factor acting on the confirmed cases, which may be restricting the unabated growth of COVID-19 cases in India. Although the data is purely a random number and a clear dependence of data of a date cannot be correlated with data of any previous occasion. However, assuming the variation as a continuous variation, further analysis is executed.

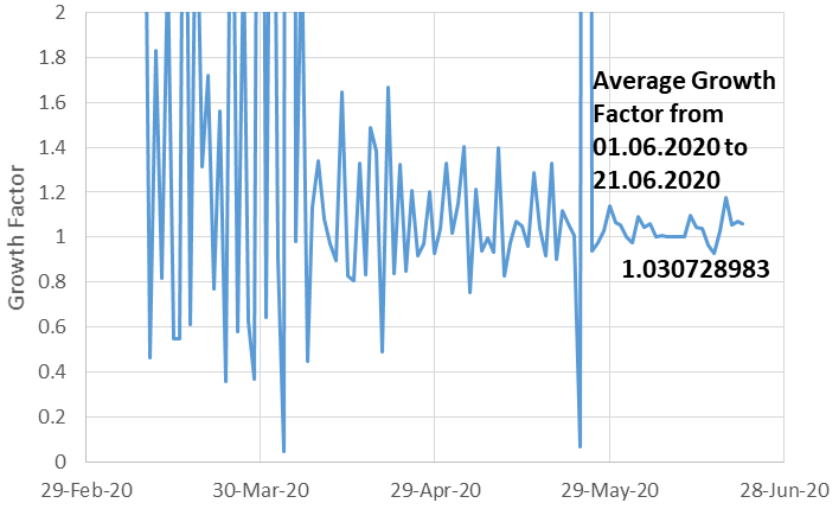

Fig. 3. Growth Factor for the Daily Confirmed Cases in India

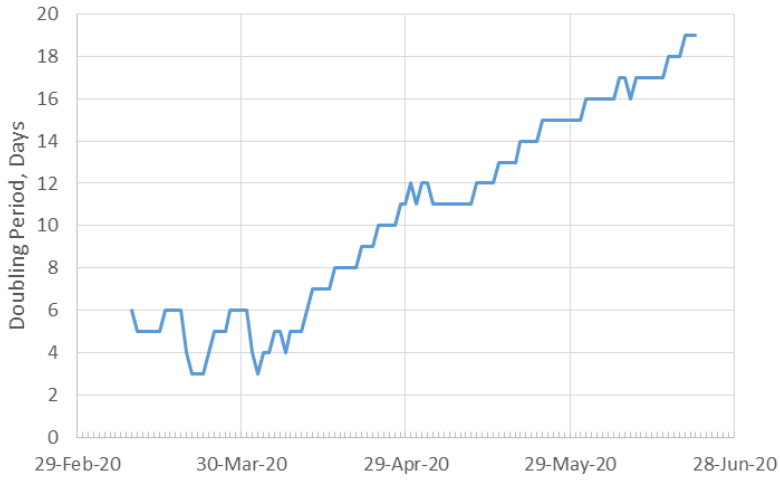

Fig. 4. Variation of Doubling Period of Pandemic COVID-19 over Time

The daily confirmed cases are plotted as Fig.5. A normal distribution curve is fitted and the chi-square minimization is applied to get best fit curve. The chi-square minimized normal distribution curve is also superimposed in Fig. 5.

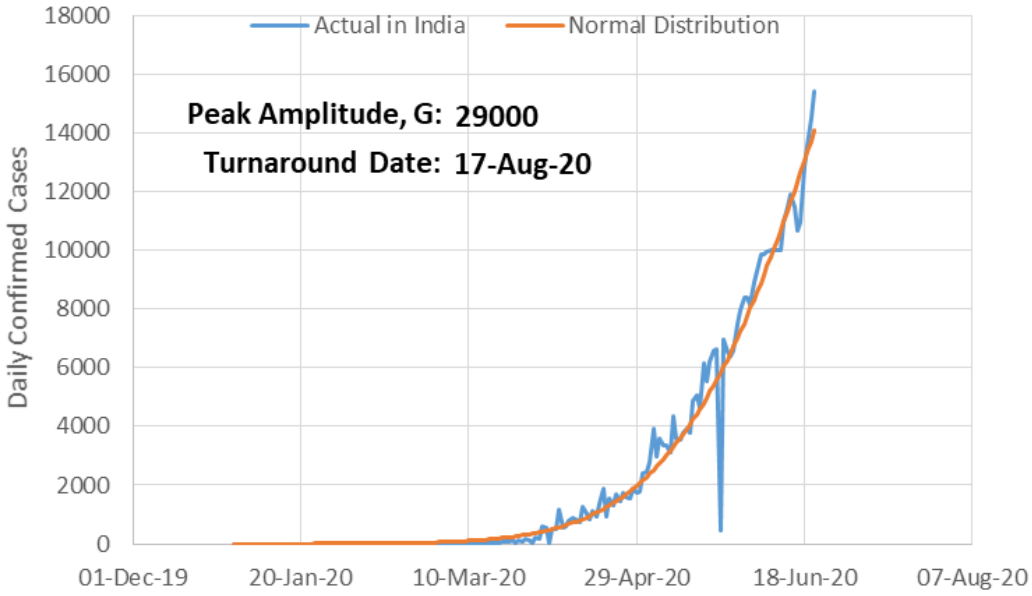

Fig. 5. Variation of Actual Daily Confirmed Cases and Normal Distribution

From this daily distribution, cumulative distribution is generated and superimposed over the total confirmed cases curve to get the match between actual and simulated condition till 21.06.2020. The variation of cumulative confirmed cases is shown as Fig. 6.

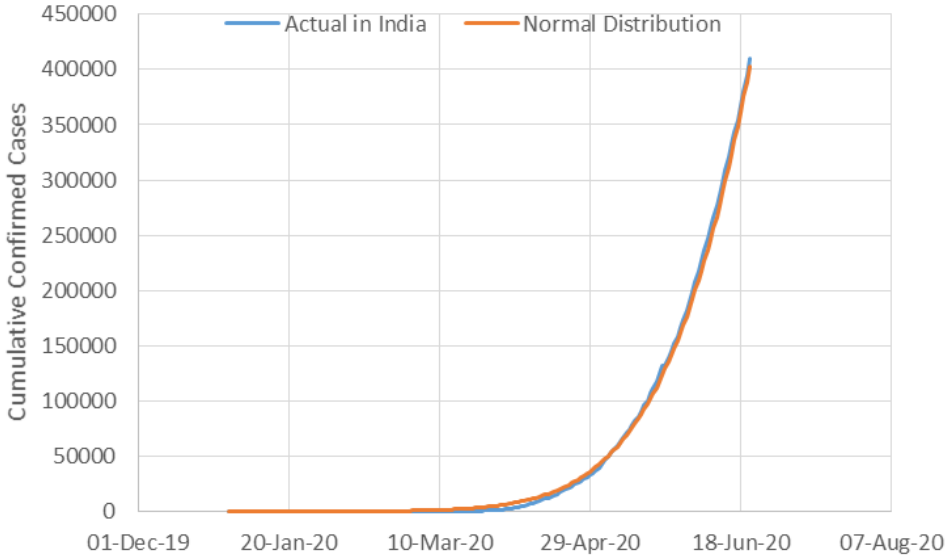

Fig. 6. Cumulative Confirmed Cases Superimposed with Normal Distribution Simulation

As in earlier occasions, the exponential variation, failed to match with the actual confirmed cases for longer duration, the attempted is restricted to take only last 21 days data and extrapolated as linear variation. Based on normal distribution and linear extrapolation, prediction is made up to 15.07.2020, as Fig.7.

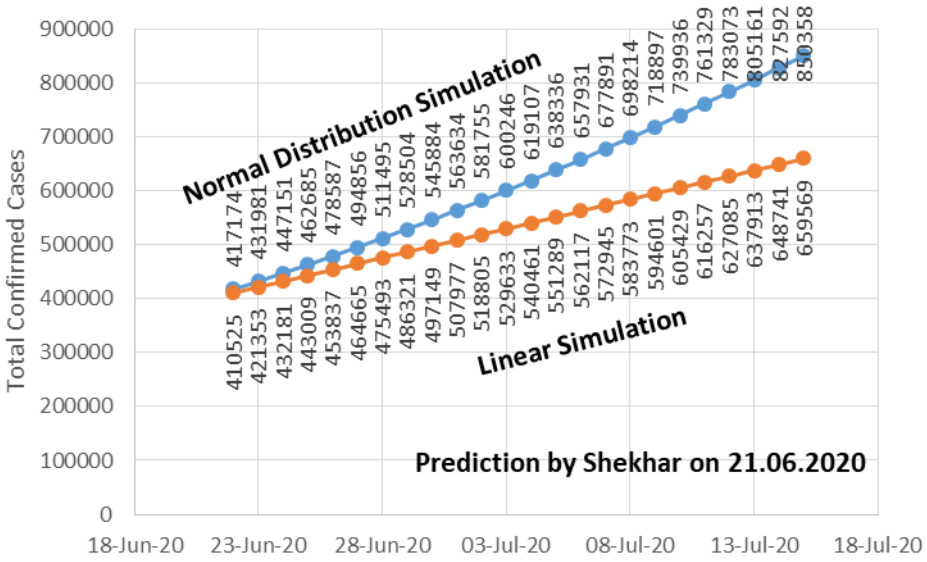

Fig. 7. Prediction of Total Confirmed Cases in India up to 15.07.2020

As per normal distribution simulation, 5.0 lakhs total confirmed cases will be attained by 28.06.2020 and 6.0 lakhs cases by 03.07.2020. The estimate indicates a value of 8.50 lakhs by 15.07.2020. However, linear simulation gives a lower value of confirmed cases. As per linear simulation 5.0 lakhs total confirmed case will be attained by 01.07.2020 and 6.0 lakh cases by 10.07.2020. By 15.07.2020, total 6.60 lakhs cases are predicted. The exponential variation and linear simulation can be taken as upper and lower bond for prediction and total number of confirmed cases in India on 15.07.2020 may be lying between 6.60 lakhs and 8.50 lakhs.

CONCLUSION

The cumulative or daily confirmed cases for pandemic COVID-19 in India is a random number, which is fitted as a continuous variation. Growth factor and doubling period for India is showing contradicting trend. Growth factor is always higher than 1, indicating a faster increase in number of confirmed cases. Contrary to this, rise in doubling period results in a slower rate of rise. However, the current situation indicates a rapid growth and turnaround is predicted in August 2020, using minimized chi-square variation applied to the normal distribution curves. Similarly a linear variation is also fitted and by 15.07.2020, the total number of confirmed cases will be between 6.60 lakhs and 8.50 lakhs.

REFERENCE

1. Julien, A. 2020. Mathematical Epidemiology in Data Rich World, *Infectious Disease Modelling*, 5:161-168.
2. Lixiang, Li., Zihang, Yang., Zhongkai, Dang., Cui, Meng., Jingze, Huang., Haotian, Meng., Deyu, Wang., Guanhua, Chen., Jiaxuan, Zhang., Haipeng, Peng and Yiming, Shao. 2020. Propagation Analysis and Prediction of the COVID-19. *Infectious Disease Modelling*, 5:282-292.
3. Shekhar, H. 2020. https://himanshushekharscience.blogspot.com/2020/05/prediction-of-covid-19-confirmed-case.html (3May).
4. Shekhar, H. 2020. https://himanshushekharscience.blogspot.com/2020/05/prediction-of-covid-19-in-india-05-may.html (5 May).

5. Shekhar, H. 2020. https://himanshushekharscience.blogspot.com/2020/05/prediction-of-covid-19-in-india-upto-10.html (10 May).
6. Shekhar, H. 2020. https://himanshushekharscience.blogspot.com/2020/05/predictionof-covid-19-in-india-15-may.html (15 May).
7. Shekhar, H. 2020, https://himanshushekharscience.blogspot.com/2020/05/india-crossing-1-lakh-leads-to-fresh.html (21 May).
8. Shekhar, H. 2020. 'India May COVID Maths'. ISBN-13: 979-8650355373, ASIN: B089HZLXP3.
9. Xiaolei, Zhang., Renjun, Ma and Lin Wang. 2020. Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries, Chaos, Solition and Fractals, 135, 2020, 109829.