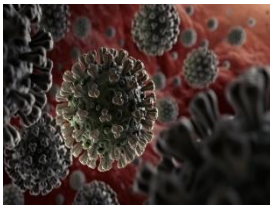
Statistical analysis, modelling and forecasting of the COVID-19 data

Abhasita Das Munshi, Rohit Thapa, Indranil Sadhya, Poushali Mallick

Department of Physics, Surendranath College,24/2 –MG Road, 700009



ABSTRACT: As the novel coronavirus (COVID-19) spreads internationally, understanding the growth and nature of its infection is crucial for estimating the true extent of the epidemic and fatality rate of the disease. In this project we estimate the publicly available data for different countries by numerically fitting those against two growth models of infection spreading. Our aim is to gain knowledge on the nature of evolution of this infection and predict the future scenarios. Datasets used in this work are taken from the public website of WHO (https://www.worldometers.info/coronavirus/).

MODEL AND ANALYSIS: We obtained data on confirmed cases for three different countries – China, USA, and India up to 5th April, 2020. Two models are used for the current analysis: the exponential function and the Logistic model. Each model has three parameters which will be estimated by a curve fitting calculation on the real data.

1.**Exponential Model:** this model describes an unstoppable infection growth. The most generic function is

$$F(x, a, b, c) = a.e^{b(x-c)}$$

Here the variable x represents the time since first detection and a, b, c are the three parameters

2.**Logistic Model:** this model has been widely used to describe the growth of a population. Unlike the exponential model, this is more realistic as it is very likely that the infection should end someday in the future. Even if everybody will be infected, they will develop the proper immunity defence to avoid a second infection, as long as the virus doesn't mutate too much. The most generic expression of a logistic function is:

$$F(x, a, b, c) = \frac{c}{1 + e^{-(x-b)/a}}$$

Again the variable x refers to the time and a, b, c are the three parameters. We use `python` programming with help of `numpy`, `scipy` and `matplotlib` libraries to analyse and visualize data.

RESULTS AND DISCUSSION: The plots for China, USA and India are shown in Figure 1. In the x-axis we show that the number of days from the first infection detected and in the y-axis, we plot the total no. of confirmed cases. The red and the blue lines represent the exponential and the logistic model respectively and the green points are the real data used in this analysis. Note that for the fitting some of the data points have been ignored to achieve a better fit using a proper analysis of residuals.

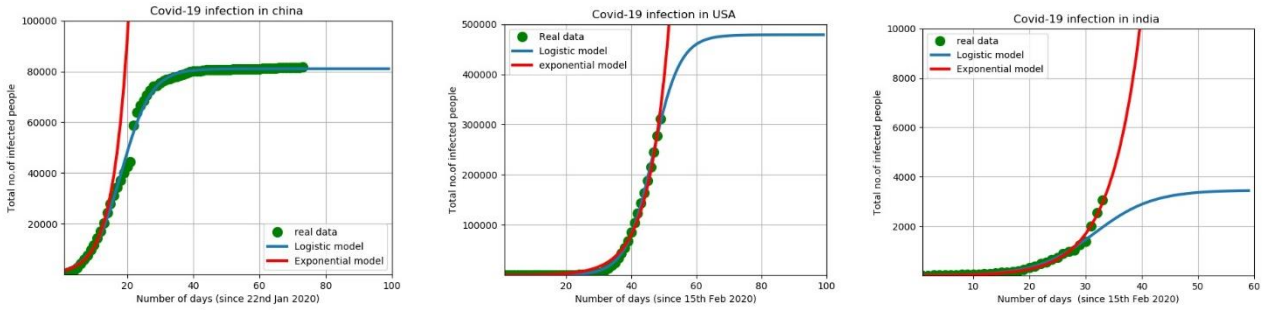


Figure 1. Analysis of COVID-19 data for China, USA, India.

In case of China one can see that, already a saturation has been achieved, so the valid model would be the logistic model. Whereas in case of USA, the present data cannot distinguish between the exponential and the logistic model. Both the models are valid and we cannot rule out the other model unless more data come in future. Finally, the plot for India indicates there is a clear departure from the logistic model – the present data strongly support the exponential one. That means we are still at a very high risk of an unstoppable growth of infection. If this trend continues we can expect about a total of 8000 confirmed cases in the next week or so.

CONCLUSION: Our result suggests that, the present situation in India is really alarming - we are still in a very dangerous regime where the infection is growing exponentially! As indicated by other works, sustained periods of lockdown (maybe with periodical relaxation), along with social distancing, will reduce the no. of cases to levels where individualised social contact tracing and quarantine may become feasible. So stay home, stay safe and follow the orders issued by the Government. Note that, these data for India and USA will be updated in the future and ultimately will support the logistic model like the case for china and as a result our fitting model will also get changed. We aim to revisit the problem in future with better availability of data to provide more accurate estimates of the progress of this epidemic.

This project has been done under supervision of Dr. Sourav Mitra, Assistant Professor, Physics Department, Surendranath College.