Frontiers in Health InformaticsISSN-Online: 2676-7104 www.healthinformaticsjournal.com
2024; Vol 13: Issue 3 Open Access

Predictive modelling approach for cumulative fatality rate of COVID-19 in India

Sachin Tomer[1], Birjesh Kumar[1], Ashish Kumar Shukla[1*]
[1]Department of Statistics, Ramanujan College, University of Delhi, Delhi-110019.
***Corresponding Author: Ashish Kumar Shukla**
Ramanujan College, University of Delhi, Delhi-110019.

Cite this paper as: Sachin Tomer, Birjesh Kumar, Ashish Kumar Shukla (2024) Predictive modelling approach for cumulative fatality rate of COVID-19 in India. *Frontiers in Health Informatics, 13* (3), 5924-5942

Abstract: *Background: COVID-19 outbreak was first reported in Wuhan, China. In India, COVID-19 has 101139 laboratory-confirmed, 42,309 recovered and 3,163 deaths as on 19 May 2020. The study is conducted to select the best fit model for cumulative fatality rate due to COVID -19 in India from first death day to 19 May, 2020 and the predicted amount of cumulative fatality rate with respect to time is also presented. A comparison of different countries (where there are 50000+ cases registered by 19 May, 2020) has been prepared towards the average of cumulative fatality rate and average of cumulative recovery per death. Method: The goodness-of-fit tests: Kolmogorov-Smirnov test, Anderson-Darling test, Root Mean Square Error, and Coefficient of Determination are considered to select the best fit model for cumulative fatality rate. The probability distribution models which are frequently used in survival/death analysis are considered in the study. Results: On the behalf of the result of goodness-of-fit test, it is found that extreme value distribution model is selected as a best fit model on cumulative fatality rate due to COVID-19 in India. Conclusion: The average of cumulative fatality rate in United Kingdom is the highest and Russia has lowest. In terms of average of cumulative recovery per death, United Kingdom stands in the bottom and China is performing best. According to best fit model, the maximum value of cumulative fatality rate due to COVID-19 in India has been predicted as 4.2 and it is also predicting that the predicted value of cumulative fatality rate would be greater than 3.61, 3.65 and 3.7 after the 50, 80 and 110 days respectively.*

Keywords: *Covid-19, Probability distributions; Fatality rate; Goodness of fit test***.**

1. Introduction

COVID-19 is transferable virus which roots severe breathing problem. Its scientific name is Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), which came in to existence in the world from mid-November, 2019 with its first reported case in Wuhan, China [1,2,3]. After this ongoing outbreak which has spread to all over world, WHO declares this as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020 and a pandemic on 11th march 2020 [4,5]. This corona virus disease is named as COVID-19 in a report of WHO dated 11th February 2020. A rising infectious due to COVID-19 involves fast spreading, endangering the health of large numbers of people, which requires immediate action to prevent the disease at the community level.

The World is facing a tough corner due to COVID-19 with more than 47.50 million confirmed cases of infection and more than 3.17 Million confirmed deaths, as of 19th May 2020 in 209 countries and territories. On the other hand more than 18.25 million people have reportedly recovered from this disease till mentioned date**.** In total no cases of COVID-19 around the world the United States is covering approximate 30% cases itself i.e. highest no of cases (1508598) with first rank and Lesotho country in Southern Africa stands at last place as only one case has been reported till the study date [6](i.e. 19 May 2020).

In Indian perspective the first case of the COVID-19 reported on 30 January 2020, as a student returning from Wuhan, China where the virus originated. A total of 101139, cases, 42309 recoveries and 3,163 deaths in the country have been confirmed on 19th May 2020 [6]. India holds the fourth place in largest number of confirmed cases in Asia **.** Out of the total cases of COVID-19 around half of all reported cases was covered only by the five metro-cities, Mumbai, Delhi, Ahmedabad, Chennai and Pune of the country but as a majority of the confirmed cases were linked to other countries. According to a report of Oxford COVID-19 Government Response Tracker (OxCGRT) [7] which is based on data from 73 countries, it has been found that the Indian Government has responded stringently than other countries in tackling this pandemic.

Mathematical and Statistical analysis are in today need to develop some probabilistic models for the prediction of various infectious diseases as well as to predict the development trend of the virus infections. Many authors have contributed in this regard using some statistical models like, SEIR epidemic model, logistic model, bertalanffy model, gompertz model general linear model, poisson model and generalized logistic model [8,9,10,11].

The goodness-of-fit is the process to analyze the pattern and models fitting of a dataset, which plays a commendable role in statistical analysis. The methods of goodness-of-fit the Anderson–Darling test, the coefficient of determination test, Akaike information criterion (AIC) and Bayesian information criterion (BIC) play an essential role to select the best fit model [12,13]. A lot of works have been carried out by various authors in this area: the proportion of unreported accidents to develop a trace accident model and the serial interval of pandemic COVID-19 dataset have been estimated using different types of distributions like log normal, gamma, beta, and weibull distributions [14,15,16].

Death is the harshest reality of human race which is beyond comprehension. It can be happened due to in general or some infections or serious pandemic. However, one can estimate all mortality estimators like death rate, survival rate etc. using vital statistics and census data. But COVID-19 pandemic data is rarely available in natural disaster scenario or where mass displacement has taken place. This data unavailability is more peculiar for developing countries. The case, however, is different in the scenario of COVID-19 where the fatality rate is proceeding on a large scale. Fatality rate is measured for a particular population as the number of deaths per thousand individuals per year in a particular population. It is typically expressed in units of deaths. The epidemiologists use few ways of estimation to predict mortality. Some authors also presented their study on mortality dataset for ex. fitting of survival data using gompatrz, weibull and logistic function, and extreme value analysis of advanced age mortality data [17,18].

The study focuses on three important issues.

- To compare the different countries towards the average of cumulative fatality rate and average of cumulative recovery per death.
- To select the best fit model for cumulative fatality rate due to COVID-19 in India for the period from first death day to 19 May, 2020.
- To predict the maximum amount of cumulative fatality rate due to virus in India using best

fitted model and also predict the cumulative fatality rate with respect to days.

2. Materials and Method

The study of frequency investigation is most important to obtain the utmost fitted model that could anticipate extreme event of any epidemic phenomena for ex. death, recovery etc.. In this section, data and study area of study, the probability distribution models which are frequently used in survival/death analysis are discussed, the procedure of goodness-of-fit to select best fit model, procedure to find cumulative fatality rate of different countries and also cumulative recovery per death are presented too. A predictive approach to estimate the maximum amount of cumulative fatality rate due to COVID-19 in India also has been mentioned.

2.1. Data and study area

In this study, the countries where there are more than 50,000 infected cases have been reported due to COVID-19 till 19th May 2020, are considered to compare their average of cumulative fatality rate and their average of cumulative recovery per death. According to this criterion, 17 countries are taking into consideration: United Kingdom, France, Belgium, Italy, Iran, Spain, Maxico, Brazil, United States, China, Canada, India, Peru, Germany, Turkey, Saudi Arabia, and Russia.

The data on cumulative fatality rate in India, which occurred due to COVID-19 virus for the period from the first death day to 19th May, 2020, are considered to select the best fit model. The data of cumulative fatality rate and cumulative recovery per death has been obtained using the equation (21) and (22) respectively.

2.2. Commonly used probability distributions

The probability distribution models which are discussed below, are taking into consideration to select the best fit model for fatality rate due to COVID-19 in India. The estimation of parameters of the considered models is obtained using the method of maximum likelihood. For obtaining the values of parameters, the mean $\bar{X}$ and standard deviation σ are obtained using the following equations.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (1)$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \quad (2)$$

In the given probability distribution model, $F_X(x)$ and $f_X(x)$ represent cumulative distribution function (CDF) and probability density function (PDF) of a random variable X repectively.

2.2.1. Normal distribution

Normal distribution is the one of the most commonly found distribution in the real world, for a random variable X, cumulative distribution function and probability density function of the distribution are given in equation (3) and (4) respectively.

$$F_X(x) = \frac{1}{2}\left\{1 + erf\left(\frac{x-\mu}{\theta\sqrt{2}}\right)\right\}; x > 0 \quad (3)$$

$$f(x \mid \mu, \theta) = \frac{1}{\theta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\theta}\right)^2} \; ; \; -\infty < x < \infty ; -\infty < \mu < \infty ; \sigma > 0 \qquad (4)$$

The parameters μ and θ represent the mean and the standard deviation respectively.

2.2.2. Log-normal distribution

The distribution of exponential function of a normally distributed random variable is defined as lognormal distribution. Lognormal distribution has been extensively used in reliability, survival and stock prices analysis. The two-parameter lognormal distribution of a continuous random variable has the following Cumulative probability distribution and probability density function:

$$F_X(x) = \frac{1}{2} + \frac{1}{2} erf\left(\frac{In(x) - \mu}{\lambda\sqrt{2}}\right); x > 0 \qquad (5)$$

$$f(x \mid \mu, \lambda) = \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - \mu}{\lambda}\right)^2} ; -\infty < x < \infty \qquad (6)$$

where, μ is the location parameter and λ is the shape parameter.

2.2.3. Birnbaum-saunders distribution

This distribution introduced by Z. W. Birnbaum S. C. Saunders in 1969, is a continuous lifetime probability distribution. This distribution is extensively used in reliability studies. It has been used to model failure times and failures due to cracks. The Birnbaum-Saunders distribution has been extensively used for modeling in several fields namely medical sciences, engineering and biomedicine. The cumulative distribution and probability density function are giving in equation (7) and (8) respectively.

$$F_T(t) = 1 - \exp\left\{-\left(\frac{t}{\eta}\right)^{\omega}\right\}; t > 0 \qquad (7)$$

$$f(t \mid \omega, \eta) = \frac{1}{2\omega\eta\sqrt{\pi}} \left[\left(\frac{\eta}{t}\right)^{1/2} + \left(\frac{\eta}{t}\right)^{3/2}\right] \exp\left\{\frac{-1}{2\omega^2}\left(\frac{t}{\eta} + \frac{\eta}{t} - 2\right)\right\}; t > 0 \qquad (8)$$

Here, ω and η are shape and scale parameters respectively.

2.2.4. Logistic distribution

This distribution with similar shape to the normal distribution belongs to the family of continuous distributions. The distribution is mainly used in logistics regression, modelling growth, survival and reliability analysis.

The cumulative distribution and probability density function are:

$$F_X(x) = \frac{1}{1+\exp\left\{-\left(\frac{x-\alpha}{\gamma}\right)\right\}}\ ; x > 0 \qquad (9)$$

$$f(x \mid \alpha, \gamma) = \frac{\exp\left\{-\left(\frac{x-\alpha}{\gamma}\right)\right\}}{\gamma\left[1+\exp\left\{-\left(\frac{x-\alpha}{\gamma}\right)\right\}\right]^2}\ ; -\infty < x < \infty, \gamma > 0 \qquad (10)$$

where α and γ are location and scale parameters respectively.

2.2.5. Log-logistic distribution

Log-logistic distribution is a continuous probability distribution for a positive random variable whose logarithm has a logistic distribution. It is used in survival analysis, hydrology to model stream ow and precipitation, economics etc.. Equations (11) and (12) represent the cumulative distribution and probability density function respectively.

$$f(x \mid \theta, \xi) = \frac{\left(\frac{\xi}{\theta}\right)\left(\frac{x}{\theta}\right)^{(\xi-1)}}{\left[1+\left(\frac{x}{\theta}\right)^{\xi}\right]^2}\ ; x > 0, \theta, \xi > 0 \qquad (11)$$

$$F_X(x) = \frac{1}{1+\left(\frac{x}{\theta}\right)^{-\xi}}\ ; x > 0, \theta, \xi > 0 \qquad (12)$$

where θ and ξ are scale and shape parameters respectively

2.2.6. Weibull distribution

Weibull distribution is used to describe various types of observed failures of components and phenomena. They are widely used in reliability and survival analysis. The cumulative distribution and probability density function of the two parameter Weibull distribution are:

$$f(x \mid \lambda, a,) = \left(\frac{a}{\lambda}\right).\left(\frac{x}{\lambda}\right)^{(a-1)} \exp\left[\left\{-\left(\frac{x}{\lambda}\right)\right\}^{a}\right] \qquad ; x > 0; \lambda, a > 0 \qquad (13)$$

$$F_X(x) = 1 - \exp\left\{-\left(\frac{x}{a}\right)^{\lambda}\right\}; x > 0 \qquad (14)$$

where λ is the scale parameter and a is the shape parameter.

2.2.7. Extreme value distribution

The distribution often referred to as the Extreme Value Distribution (Type I) is the limiting distribution of the minimum of a large number of unbounded identically distributed random variables. The probability density function and cumulative distribution function are given by:

$$f(x \mid \mu, \beta) = \frac{1}{\beta} \exp\left(\frac{x-\mu}{\beta}\right) \exp\left\{-\exp\left(\frac{x-\mu}{\beta}\right)\right\} \quad ; -\infty < x < \infty, \mu, \beta > 0 \quad (15)$$

$$F_X(x) = 1 - \exp\left[-\left\{e^{\left(\frac{x-\mu}{a}\right)^{\lambda}}\right\}\right] \quad ; -\infty < x < \infty \quad (16)$$

where β is the scale parameter and μ is the location parameter.

2.3. Goodness-of-fit test

For checking the validity of a specified probability distribution, goodness-of-fit test statistic has been used. In procedure of statistical testing, there are so many test are available for checking the normality of the data. One of the most well-known tests, Empirical Distribution Function (EDF) test is used to serve this purpose. EDF presents the discrepancy between the empirical and theoretical distributions [19]. The Kolmogorov-Smirnov (K-S) test, the Anderson-Darling test and the Cramer Von-Mises test, these three most common methods are used under EDF test [20, 21]. The K-S test and the A-D tests have been applied to obtain the best fitting model on fatality rate dataset. The root mean square error (RMSE) and Coefficient of determination (*R^2*) are also used to assess the best fit model. A graphical method is also presented to demonstrate the result.

2.3.1. Kolmogorov-smirnov (K-S) test

The K-S test provides the maximum vertical difference between empirical distribution function and theoretical distribution function [22]. The K-S test is established to compare the empirical distribution function $F_n(x)$ and theoretical distribution function $F(x)$. It provides the maximum distance between $F_n(x)$ and $F(x)$. The empirical distribution function $F_n(x)$ for n independently and identically distributed observations X_i is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i} \leq x \quad (17)$$

Where $I_{X_i} \leq x$ is the indicator function, equal to 1 if $I_{X_i} \leq x$, otherwise 0.

The K-S test for a given cumulative distribution function $F(x)$ is:

$$D_n = Sup\left|F_n(x) - F(x)\right| \quad (18)$$

2.3.2. Anderson-darling (A-D) test

The Anderson-Darling test was introduced by Anderson and Darling [23]. The A-D test is also one of the most suitable procedures, which is used to select the best fit model. The A-D test is expressed as

$$A^2 = -\sum_{i=1}^{n}\left[(2i-1)\ln F_X(x_i) + \ln\{1 - F_x(x_{n+1-i})\}/n\right] - n \qquad (19)$$

Where, $F_X(x_i)$ is the cumulative distribution function of proposed distribution model at x_i; $i = 1,2,.....n$.

2.3.3. Root mean square error (RMSE)

Root mean square error plays an essential role in goodness-of-fit. It is also known as fit standard error and error of the regression. The RMSE can be defined as:

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - X)^2} \qquad (20)$$

where x_i denotes the estimated value and X denotes the observed value.

2.3.4. The coefficient of determination (R^2)

The coefficient of determination, expressed R^2 is used in procedure of testing of hypothesis and infer about the goodness-of-fit of a model. R^2, is generally in the range between 0 and 1. 1 indicates that the regression predictions perfectly fit the data.

2.4. Cumulative fatality rate (CFR) and cumulative recovery per death

If total number of cumulative deaths per day, cumulative infected cases per day, and cumulative recovery per day are denoted by D_t, I_t, and R_t respectively. The cumulative fatality rate (μ_t) and cumulative recovery per death (δ_t) are shown by equation (21) and (22) respectively.

$$\mu_t = \frac{D_t}{I_t} \qquad (21)$$

$$\delta_t = \frac{R_t}{D_t} \qquad (22)$$

2.5. A predictive approach

2.5.1. A cumulative distribution function using best fit model on the cumulative fatality rate due to COVID-19 in India is obtained to predict the maximum value of cumulative fatality rate.

2.5.2. If a random variable *X* is greater than or equal to an event of magnitude x_T, happened once in *T* days, The probability of happening this event in a given time *T* is as follows [24]

$$P(X \geq x_T) = \frac{1}{T}$$

$$P(X \leq x_T) = 1 - \frac{1}{T} \qquad (23)$$

3. Results and discussion

The objectives of this study are to compare the different countries towards the average of cumulative fatality rate and average of cumulative recovery per death, to select the best fit model for cumulative fatality rate due to COVID-19 in India from first death day to 19 May, 2020 and to predict the maximum amount of cumulative fatality rate due to virus in India using best fitted model.

It has been observed that the cumulative fatality rate showing the continuous change in death rate as the number of day increases for existence of COVID-19 in India. The changes in cumulative fatality rate due to COVID-19 in India are picturized in figure 1. Form this Figure it can be observed that the cumulative fatality rate is changing rapidly from the first death day to 09th April, 2020 in a range of 1.333333 to 2.895012. From onward 10th April, 2020 cumulative fatality rate is fluctuating about a particular value with a range of 3.103556 to 3.429,775. It is also noticed that the deviation of the second phase (on and after 10th April) is lesser than the deviation of the first phase (before 10th April). The average of cumulative fatality rate of second phase is found to be 3.241515. It shows that from the last 40 days CFR is fluctuating around 3.241515.

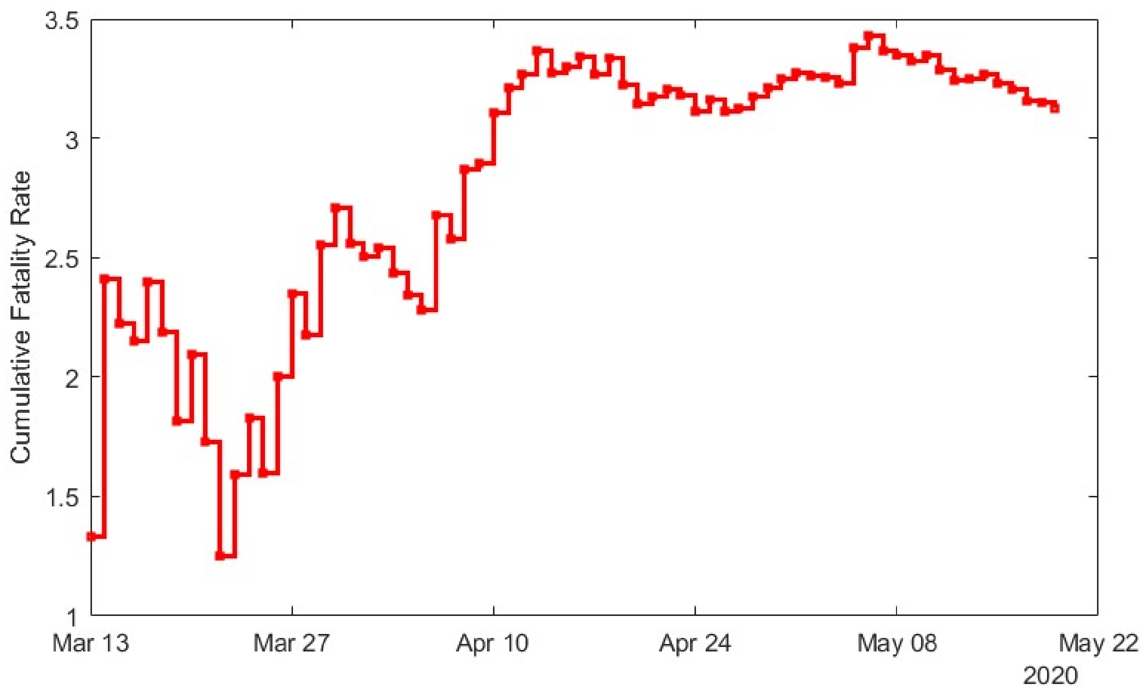

Figure 1. Graphic display of cumulative fatality rate from 1st day to 19th May 2020 in India.

3.1. Comparison towards average of cumulative fatality rate and average of cumulative recovery per death

Here comparison the average of cumulative fatality rate which is the average of ratio of cumulative death by the total cumulative number of people diagnosed with the COVID-19 of various countries is presented in Figure 2. In this Figure, it can be seen that average of cumulative fatality rate in United Kingdom with 11.07921 is the highest and lowest average of cumulative fatality rate is in Russia with 0.83527. India is ranked 12th and seem to performing better than USA, China and other 11 countries with average of cumulative fatality rate of 2.82005.

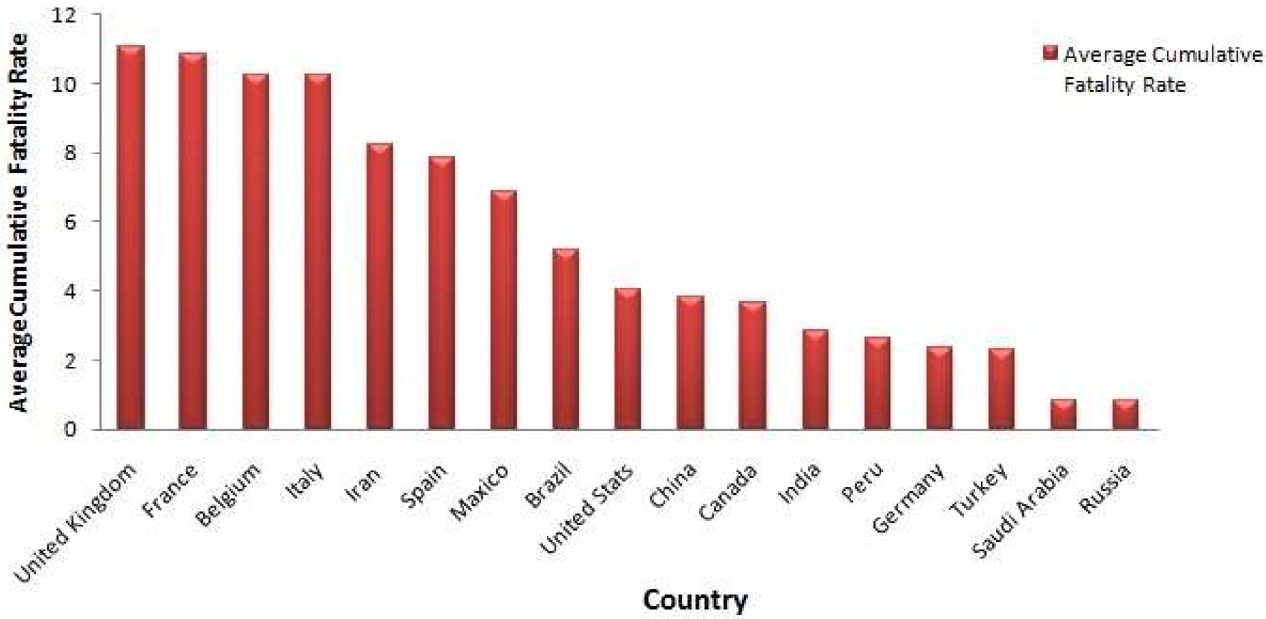

Figure 2. Comparison of various countries on Average of Cumulative Fatality Rate

While Looking at average of cumulative recovery per death which is the average of number of recovered cases per at the time of per death seen. From Figure 3, it can be seen that United Kingdom stands in the bottom with 1.06429 which is shocking and alarming as well. India's rate does not showcase a good rate i.e. 19.29292 which brings it to 14th position whereas China has presented a good performance being the best in terms of recovery with the average of recovery per death being 64.14031.

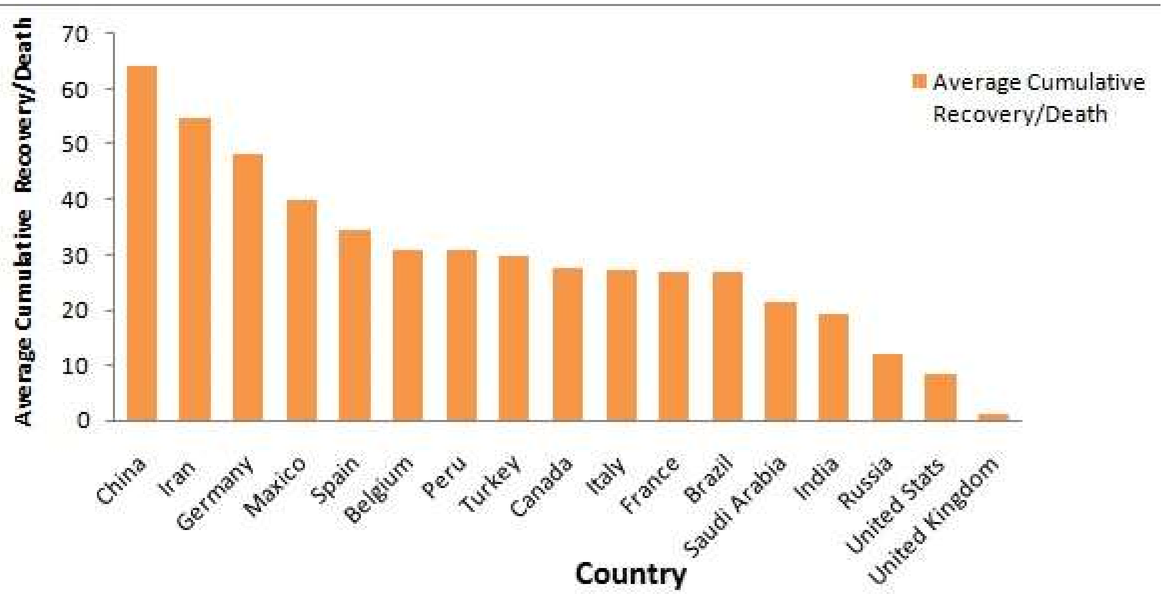

Figure 3. Comparison of various countries on Average of Cumulative Recovery per Death

3.2. Selecting the best-fit results

Estimation of parameters of considered probability distribution models is needed to assess the result of goodness-of-fit test statistic on cumulative fatality rate. Method of maximum likelihood estimator is used to estimate the parameters of the models. The result of mean$(\bar{X})$, standard deviation (σ), estimated values of parameters along with their standard error, and variance covariance matrix of the parameters with respect to the probability distribution model has been presented in table 1.

Table 1. Estimated value of the parameters of the considered distributions using death rate data on COVID-19 in India

Model	Mean $(\bar{X})$ and SD (σ)	Parameter	S.E.	Variance Covariance Matrix
Extreme Value	$\bar{X}$ = 3.07422 σ = 0.622283	μ = 3.07422 β = 0.387236	μ =0.0490253 β =0.0409801	μ β μ 0.00240348 -0.000577085 β -0.000577085 0.00167937
Logistic	$\bar{X}$ =2.90237 σ =0.596719	α = 2.90237 γ = 0.328989	α =0.0705996 γ = 0.0331573	α γ α 0.0049843 -0.000335915 γ -0.000335915 0.0010994
Log Logistic	$\bar{X}$ = 2.93978 σ = 0.710446	θ = 1.0509 ξ = 0.128797	θ = 0.027182 ξ = 0.0132244	θ ξ θ 0.000738861 -5.44104e-05 ξ -5.44104e-05 0.000174885
Normal	$\bar{X}$ = 2.82005 σ =0.581661	μ = 2.82005 θ= 0.581661	μ = 0.0705368 θ= 0.0504364	μ θ μ 0.00497544 -3.62996e-18 θ -3.62996e-18 0.00254383
Weibull	$\bar{X}$ = 2.83745 σ = 0.495393	λ = 3.03987 a = 6.7165	λ = 0.0570667 a = 0.720602	λ a λ 0.00325661 0.0112603 a 0.0112603 0.519267
Log Normal	$\bar{X}$ = 2.83007 σ = 0.697230	μ = 1.01084 λ = 0.242744	μ = 0.0294371 λ = 0.0210486	μ λ μ 0.000866541 6.79948e-20 λ 6.79948e-20 0.000443044 η ω

Birnbaum-Saunders	$\bar{X}$ = 2.81985	η = 2.73867	η = 0.0802655	η	0.00644256	-7.02448e-07
	σ = 0.691110	ω = 0.243491	ω =0.0208792	ω	-7.02448e-07	0.000435941

The obtained results of goodness-of-fit test statistic K-S test, coefficient of determination (R^2), root mean square error (RMSE), and A-D test under the seven considered distribution models have been exposed in table 2. The best fit result is taken as the smallest goodness-of-fit result among all the considered probability distribution models. It is observed that four goodness-of-fit tests produced different types of distribution models as a best fit; therefore, a ranking system is used to select the best fit model [24]. In this system, all results of goodness-of-fit tests are ranked for every probability distribution model, as rank 1 being the best, 2 the second best and so on. A probability distribution model is selected as the best fit model if it has the lowest sum of the four rankings. The rankings for the results are presented in table 3. On the basis of the ranking system, it has been observed that extreme value distribution model and logistic distribution model have the same lowest sum of rankings. (i.e. a tie between both the models). Therefore, two more tests of goodness-of-fit is taken into consideration to select the best fit model from extreme value distribution and logistic distribution, named Akaike information criterion (AIC) and Bayesian information criterion (BIC). These tests are also used to select the best fit model [25, 26]. A smallest value of the AIC and BIC are taken as a best fit model.

The results of AIC and BIC for extreme value distribution are100.2404 and 104.6794 respectively and that for logistic distribution are 123.4158 and 127.8548 respectively. It has been observed that extreme value distribution model has lesser result of AIC and BIC than logistic distribution model. Hence extreme value distribution is considered as a best fit model on cumulative fatality rate due to COVID-19 in India, while the logistic and log logistic distribution models are considered the second and third amount of best fit model respectively on the cumulative fatality rate.

Table 2. Statistical results and best fit results of the death rate data due to COVID-19 in India

Model	**K-S**	$\boldsymbol{R^2}$	**RMSE**	**A-D**
Extreme Value	0.2334923	0.8936727	0.1070618	4.2361
Logistic	0.2218186	0.8770564	0.1049617	4.301
Log Logistic	0.2269098	0.8591083	0.1084625	4.7304
Normal	0.2605446	0.8622518	0.1107778	4.5748
Weibull	0.2567092	0.8793352	0.1119165	4.7544
Log Normal	0.2654797	0.823231	0.1178861	5.2803
Birnbaum-Saunders	0.2699081	0.8190368	0.1195523	5.4162

Table 3. Statistical results and best fit results of the death rate data on COVID-19 in India

Models	**K-S**	R^2	**RMSE**	**A-D**
Extreme Value	3	1	2	1
Logistic	1	3	1	2
Log Logistic	2	5	3	4
Normal	5	4	4	3
Weibull	4	2	5	5
Log Normal	6	6	6	6
Birnbaum-Saunders	7	7	7	7

To see the goodness-of-fit, the fitted cumulative probability function verses empirical distribution function of considered probability distribution models on cumulative fatality rate have been displayed in Figure 4.1 to 4.7. However, to see the goodness-of-fit not only graphical display but numerical results are also important.

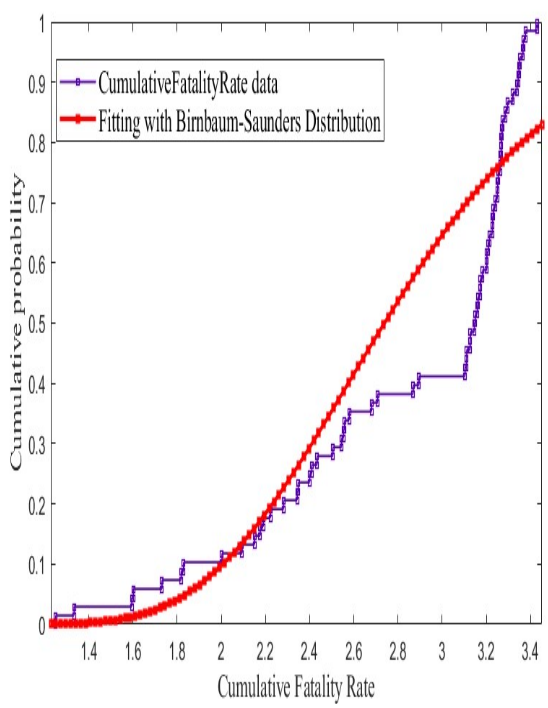

Figure 4.1

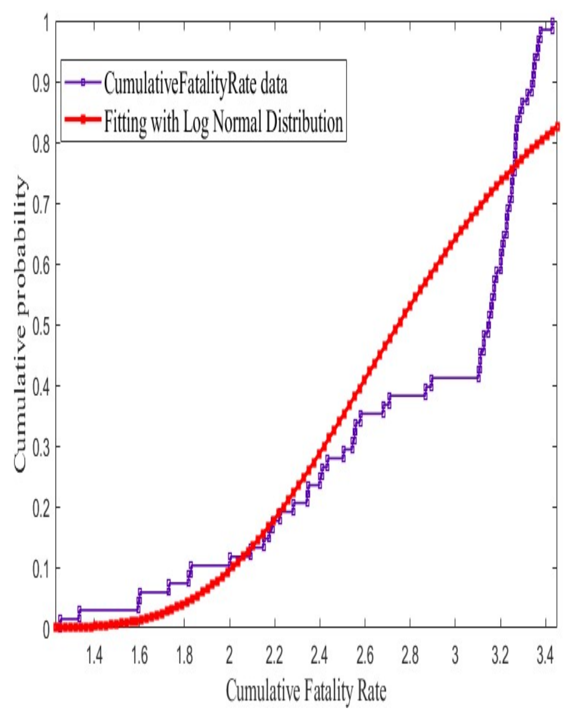

Figure 4.2

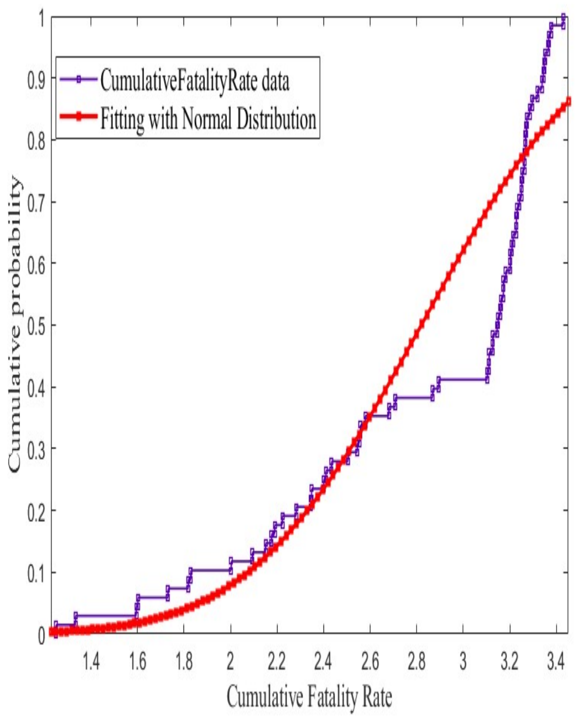

Figure 4.3

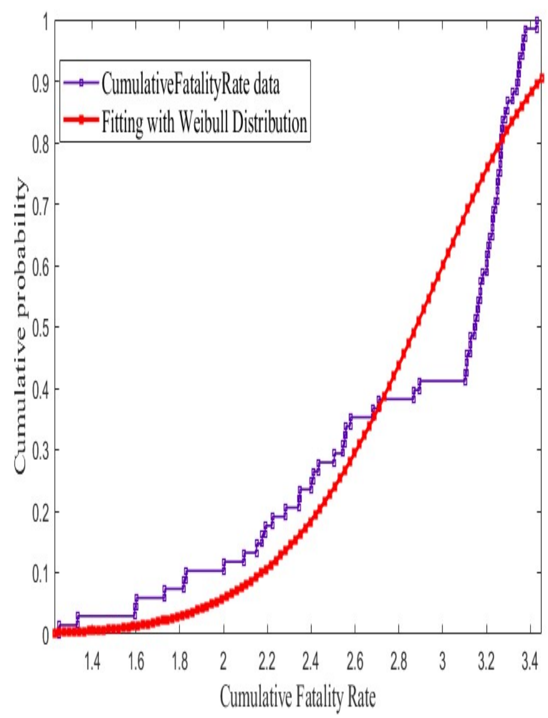

Figure 4.4

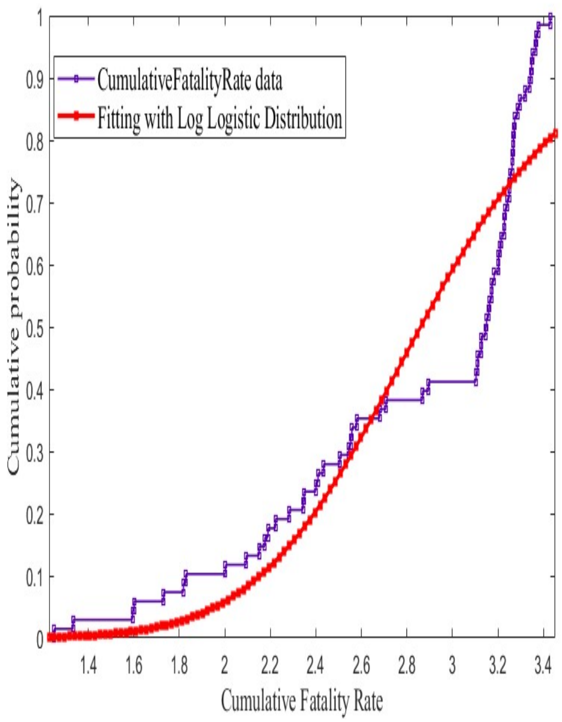

Figure 4.5

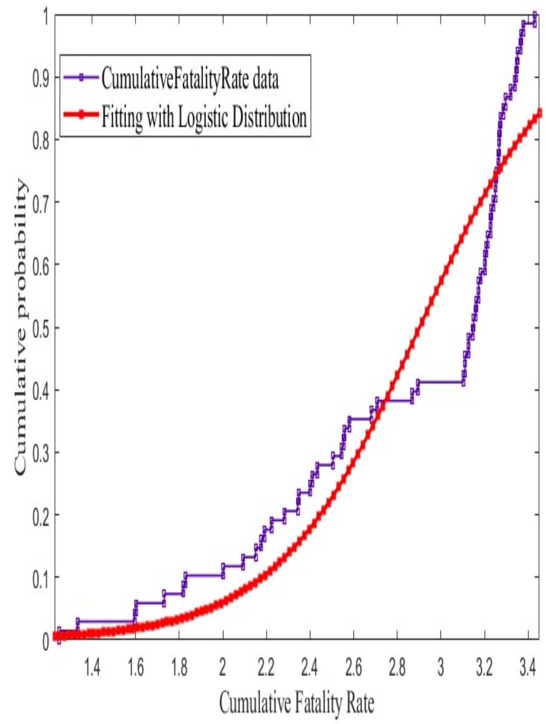

Figure 4.6

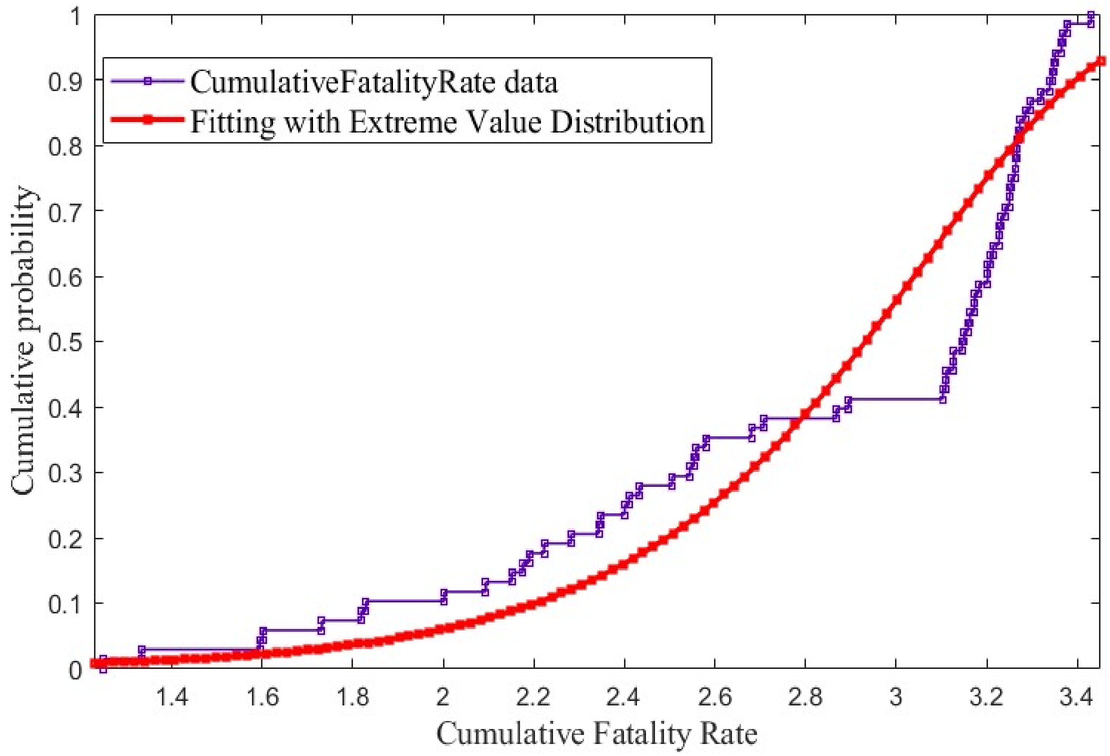

Figure 4.7

3.3. Predictive approach

Figure 5 display the cumulative probability plot of uppermost three best fitted models on cumulative fatality rate (i.e. extreme value distribution model, logistic distribution model and log logistics distribution model). Form figure 5, It can be seen that the cumulative probability plot of extreme value distribution model touches the certainty with the cumulative fatality rate of 4.2. It represents that the maximum amount of cumulative fatality rate due to COVID-19 in India has been predicted as 4.2. The maximum value of cumulative fatality rate has also been predicted as 6.2 and 10.2 in case of logistic and log logistic distribution models.

A prediction of cumulative fatality rate with respect to number of days has been displayed in figure 6. For extreme value distribution model, it has been observed that the values of cumulative fatality rate would be greater than 3.61, 3.65 and 3.7 after the 50, 80 and 110 days respectively. Similarly the amount of cumulative fatality rate would be greater than 4.34 and 4.44 after the 80 and 110 days respectively in case of logistic distribution model, while in case of log logistic distribution model, it would be greater than 5.02 and 5.24 after the same period respectively.

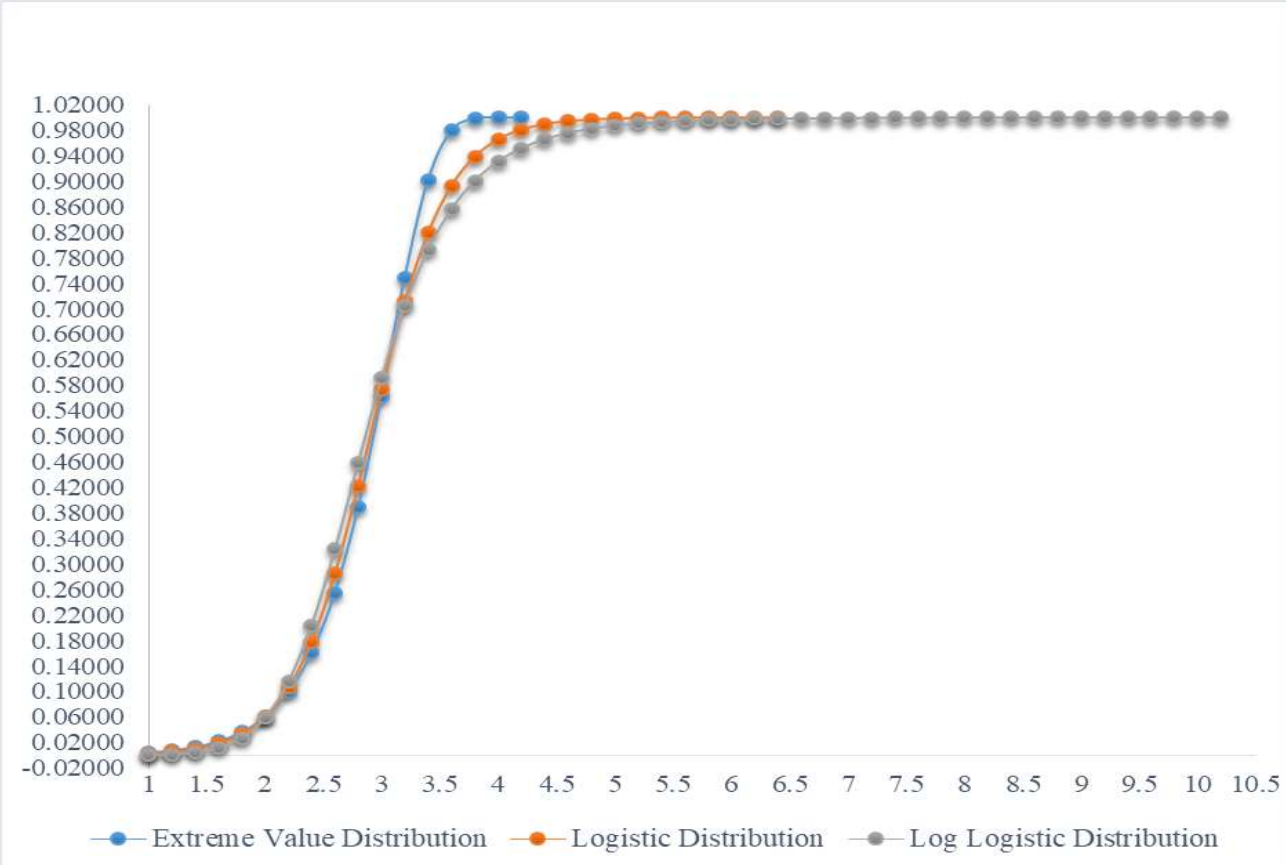

Figure 5. Predicted death rate using best fitted distribution

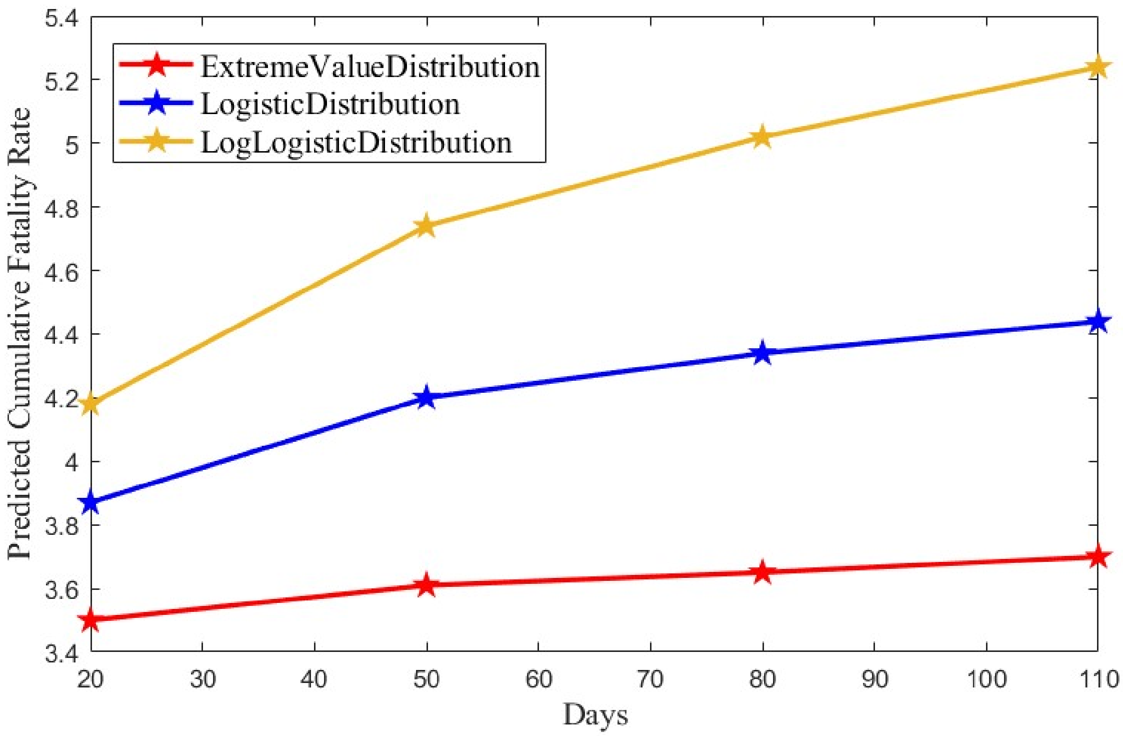

Figure 6. Predicted death rate using best fitted distribution

4. Conclusions

The study has been conducted on cumulative fatality rate and cumulative recovery per death due to COVID-19. To achieve the objectives of the study, the dataset have been analyzed with the help of goodness-of-fit tests and Graphical demonstration.

In terms of comparing of average of cumulative fatality rate of different countries it has been observed that from Figure 1, average of cumulative fatality rate of United Kingdom is highest with 11.07921 and lowest has been recorded for Russia with 0.83527. India is ranked 12th and seem to performing better than USA, China and other 11 countries with average of cumulative fatality rate of 2.82005.

Towards the average of cumulative recovery per death, it is observed that United Kingdom stands in the bottom with 1.06429 which is shocking and alarming as well. India's rate does not showcase a good rate i.e. 19.29292 which brings it to 14^{th} position whereas China is performing best with 64.14031.

On the basis of goodness-of-fit tests, it has been concluded that extreme value distribution model is selected as a best fit model on cumulative fatality rate due to COVID-19 in India. According to this best fit model, the maximum value of cumulative fatality rate due to COVID-19 in India has been predicted as 4.2.

A prediction of cumulative fatality rate with respect to number of days has also been presented. The predicted value of cumulative fatality rate due. to virus in India would be greater than 3.61, 3.65 and 3.7 after the 50, 80 and 110 days respectively if cumulative fatality rate follows the extreme value distribution model.

Conflict of Interest

The authors declare no conflicts of interest.

References

1. "Naming the coronavirus disease (COVID-19) and the virus that causes it". World Health Organization(WHO).Availablefrom: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causesit.

2. Hui DS, I Azhar E, Madani TA, et al. (2020) The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis.* 91:264-266. DOI:10.1016/j.ijid.2020.01.009.

3. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395:497-506. DOI:10.1016/s0140-6736(20)30183-5.

4. World Health Organization, Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV),(2020)Availablefrom: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov).

5. World Health Organization, WHO Director-General's Remarks at the Media Briefing on 2019-nCoV on 11 February 2020, (2020).Available from: https://www.who.int/dg/speeches/ detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11 february-2020.

6. Data Sources. Available from: https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset?select=covid_19_data.csv.

7. https://www.indiatoday.in/india/story/india-scores-high-on-COVID-19-response-tracker-made-by-oxford-university-1665573-2020-04-10.

8. Jia L, Li K, Jiang Y, Guo X, Zhao T. Prediction and analysis of Coronavirus Disease 2019. Available from: https://arxiv.org/ftp/arxiv/papers/2003/2003.05447.pdf.

9. Currie ID (2014) On fitting generalized linear and non-linear models of mortality. *Scand Actu Jour* 2016(4):356-383. dOI:10.1080/03461238.2014.928230

10. Wu K, Darcet D, Wang Q, Sornette D. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world.Available from: https://www.medrxiv.org/content/10.1101/2020.03.11.20034363v1.article-info. doi: https://doi.org/10.1101/2020.03.11.20034363.

11. Bentout S, Chekroun A, Kuniya T (2020) Parameter estimation and prediction for coronavirus disease outbreak 2019 (COVID-19) in Algeria. *AIMS Public Health* 7(2): 306-318.DOI: 10.3934/publichealth.2020026.

12. Lenart A, Missov TI (2016) Goodness-of-fit tests for the Gompertz distribution. *Commu Stat-Theo & Meth* 45(10):2920-2937.

13. Datta R, Trivedi PK, Kumawat A, Kumar R, Bhardwaj I, Kumari N. Agiwal V, Kumar S, Kumar A, Shukla A, Kumar J (2020) Statistical Modeling of COVID-19 Pandemic Stages Worldwide. Preprints 2020; 2020050319. doi: 10.20944/preprints 202005.0319.v1.

14. Nishiura H, Linton NM, Akhmetzhanov AR (2020) Serial interval of novel coronavirus (COVID-19) infections. *Inter Jour Infec Dise* 93: 284-286.

15. Du Z, Xu X, Wu Y, et al. (2020) Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerg Infec Disea* 26(6):1341-1343. dOI:10.3201/eid2606.200357.

16. Sarah N, Zamri N, Zamzuri ZH. (2019) Estimating the Proportion of Non-Fatality Unreported Tracre Accidents in Malaysia. *ASM Sc. J.* 12(1) for IQRAC2018: 239-245.

17. Wilson DL. The analysis of survival (mortality) data: Fitting Gompertz, Weibull, and logistic functions. Mechanisms of Ageing and Development. 1994; 74(1-2): 15-33.

18. Bebbington M, Lai CD, Zitikis R. "An Extreme Value Analysis of Advanced Age Mortality Data," Kathryn A. Watts, Debbie J. Dupuis, and Bruce L. Jones, October 2006. North American Actuarial Journal. 2007;11(3):173-174, DOI: 10.1080/10920277.2007.10597478.

19. Dufour JM, Farhat A, Gardiol L, Khalaf L (1998) Simulation-based Finite Sample Normality Tests in Linear Regressions. *Econ. J.* 1: 154-173.

20. Arshad M, Rasool MT, Ahmad MI (2003) Anderson Darling and Modified Anderson Darling tests for Generalized Pareto Distribution. *Pak. J. Appl. Sci.* 3: 85-88.

21. Seier E (2002) Comparison of Tests for Univariate Normality. *Intet. Stat. J*. 1: 117.

22. Conover WJ (1999) Practical Nonparametric Statistics, 3rd ed. John Wiley & Sons, Inc.: New York, NY, USA 428-433. ISBN 0471160687.

23. Anderson TW, Darling DA (1954) A Test of Goodness-of fit. *J. Am. Stat. Assoc.* 49: 765-769.

24. Alam MA, Emura K, Farnham C, Yuan J (2018) Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh. *Climate*.6(9).

25. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Auto Cont*. 19 (6): 716–723. DOI:10.1109/TAC.1974.1100705.

26. Schwarz, Gideon E (1978) Estimating the dimension of a model, *Annals of Statistics* 6 (2): 461–464. DOI:10.1214/aos/1176344136.