



Statistical Inferences: Based on Exponentiated Exponential Model to Assess Novel Corona Virus (COVID-19) Kerala Patient Data

Anurag Pathak¹ · Manoj Kumar¹ · Sanjay Kumar Singh² · Umesh Singh²

Received: 29 October 2020 / Revised: 12 June 2021 / Accepted: 25 June 2021 /
Published online: 3 August 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

In this article, we use exponentiated exponential distribution as a suitable statistical lifetime model for novel corona virus (covid-19) Kerala patient data. The suitability of the model has been followed by different statistical tools like the value of logarithm of likelihood, Kolmogorov–Smirnov distance, Akaike information criterion, Bayesian information criterion. Moreover, likelihood ratio test and empirical posterior probability analysis are performed to show its suitability. The maximum-likelihood and asymptotic confidence intervals for the parameters are derived from Fisher information matrix. We use the Markov Chain Monte Carlo technique to generate samples from the posterior density function. Based on generated samples, we can compute the Bayes estimates of the unknown parameters and can also construct highest posterior density credible intervals. Further we discuss the Bayesian prediction for future observation based on the observed sample. The Gibbs sampling technique has been used for estimating the posterior predictive density and also for constructing predictive intervals of the order statistics from the future sample.

Keywords MLE · Bayes estimate · Bayes prediction · LR test · Empirical posterior probability

1 Introduction

In December 2019, a novel corona virus was recognized from China, in Wuhan city, Hubei province, see Arshad Ali et al. [4], Li et al. [20]. More importantly, the epidemic of covid-19 has occurred in most of the countries. At the international

✉ Manoj Kumar
manustats@gmail.com

¹ Department of Statistics, Central University of Haryana, Mahendergarh 123031, India

² Department of Statistics, Banaras Hindu University, Varansi 221005, India

level, WHO has declared the covid-19 as a epidemic, it has affected more than 215 countries, areas or territories as a pandemic, see WHO [31]. But in India the first covid-19 case was reported on 30-01-2020, from a student, and then spread gradually in each state and increased the number of covid-19 cases. The virus has spread through respiratory droplets. The coughing, fever and sneezing are found to be the main symptom of covid-19, and for more see Chen et al. [9] and Paules et al. [23], etc. Further the transmission patterns of covid-19 based on age-specific social contact characterization and countries with different social structures and lifestyles effects have been studied by Liu et al. [21] and Shi et al. [25] respectively. Chen et al. [9] has been discussed in the form of global health concern. Recently, Kumar [18] has discussed cluster analysis for covid-19 data set of different states and union territories (UTs) in India. While Khakharia et al. [17] have developed an outbreak prediction system for covid-19 of the top ten high densely populated countries in worldwide. Using machine learning and a suitable statistical model for predicting the rise in the new cases has highlighted for most of the countries. Moreover, on the outbreak of covid-19, many researchers have started investigating and using the covid-19 data for various perspectives. Therefore for statistical inference purpose, we have considered covid-19 data during 09-03-2020 to 23-04-2020 from Kerala state in India. The data have shown in further section. The collected data has been modeled with distribution and analyzed. Authors refer to see in detail, for modeling, prediction, big data analytics by Shi et al. [25], longitudinal data analysis by Temesgen et al. [29] and real-time decision analysis by Tien [30]. We know that the exponential model has the most exploited distribution for lifetime data analysis. But its suitability has been restricted with a constant hazard rate, which is very difficult to verify in many practical problems. This led to the development of other distributions. The Weibull and gamma distribution have been extensively used for analysing lifetime data, particularly, in those situations where the hazard rate has monotonically increasing, decreasing and constant. The exponentiated exponential distribution was developed by Gupta et al. [13], which has an alternative to two parameters (shape and scale) as the Weibull and gamma distribution, see Gupta and Kundu [14].

1.1 The Model

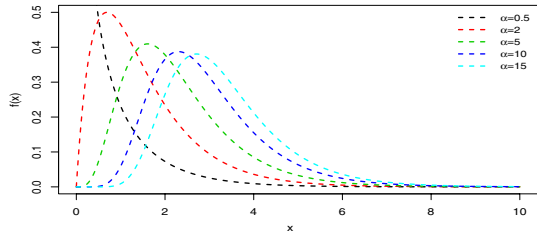
The probability density function (pdf) of the exponentiated exponential distribution or model (EED) is given by

$$f(x|\alpha, \lambda) = \alpha\lambda(1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}; \quad x > 0, \alpha > 0, \lambda > 0. \quad (1)$$

where α and λ are the shape and scale parameters, when $\alpha = 1$, it reduces to the one parameter exponential distribution. Figure 1 shows EED for $\alpha = 0.5, 2, 5, 10, 15; \lambda = 1$. The cumulative density function (cdf) of EED with two parameters are given as

$$F(x|\alpha, \lambda) = (1 - e^{-\lambda x})^\alpha; \quad x > 0, \alpha > 0, \lambda > 0. \quad (2)$$

Fig. 1 The probability model of the exponentiated exponential for different values of α and λ



The survival and hazard function for the EED are given by

$$S(x|\alpha, \lambda) = 1 - (1 - e^{-\lambda x})^\alpha; \quad x > 0, \alpha > 0, \lambda > 0 \tag{3}$$

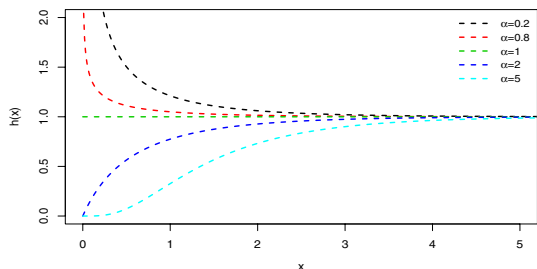
and

$$h(x|\alpha, \lambda) = \frac{\alpha \lambda (1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}}{(1 - e^{-\lambda x})^\alpha}; \quad x > 0, \alpha > 0, \lambda > 0, \tag{4}$$

respectively. The hazard function shown in Fig. 2 for $\alpha = 0.2, 0.8, 1, 2, 5; \lambda = 1$.

It may be noted here that the hazard function is monotone increasing if $\alpha > 1$ and $\lambda > 0$; monotone decreasing if $\alpha < 1$ and $\lambda > 0$; and constant shaped if $\alpha = 1$ and $\lambda = 1$. The EED can be reduced to exponential distribution (ED) having a shape and scale parameter one. For more detailed about the EED, readers are referred to see Gupta et al. [13], Gupta and Kundu [14]. Most of the studies on the EED and ED have focused on estimation of the parameters, reliability and hazard functions see Singh [26]. In Bayesian paradigm, the prior distribution of the parameters have used to derive the posterior density of the parameters. This method has also used to obtain predictive inference for future observations. To learn more about the Bayesian inferences and model selections, readers are referred to Gelman et al. [11], Baklizi [5], Bernardo and Smith [7], Berger [6] and Geisser [10]. But no one has attempted to work on empirical posterior probability based inference for EED. Thus, in this manuscript our main goals are (i) to obtain the Classical and Bayes estimate of the parameter; (ii) to utilize the predictive models with a Kerala covid-19 data to obtain the predictive inference for future observations; (iii) to justify that the given sample follow the EED by using AIC, BIC and empirical posterior probability of model selection criteria and goodness-of-fit tests.

Fig. 2 Hazard function on the probability model of the exponentiated exponential for different values of α and λ



Hence, the purpose of this paper is to present a systematic inference of EED. In Sect. 2, the maximum likelihood estimation of the parameter, asymptotic confidence interval of EED. Bayes estimators of EED parameter using non informative prior is derived in Sect. 2.4. The Bayes prediction of EED for future observations explained in Sect. 3. In Sect. 4 mainly focused on the model selection criteria. Kerala covid-19 real data analysis report and discussion are mentioned in Sect. 5. In Sect. 6 shown the conclusion of this study.

2 Estimation of Parameters

2.1 The Likelihood Function

Let x_1, x_2, \dots, x_n be a random sample of size n are drawn from EED, and $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ forms an observed sample. Then for given a set of data $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ from Eq. (1), the likelihood function is given by

$$L(\alpha, \lambda | \mathbf{x}) = \alpha^n \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \prod_{i=1}^n (1 - e^{-\lambda x_i})^{\alpha-1}. \quad (5)$$

The log-likelihood function of EED specified in Eq. (5) is given by

$$\ln(L(\alpha, \lambda | \mathbf{x})) = n \ln(\alpha) + n \ln(\lambda) + (\alpha - 1) \sum_{i=1}^n \ln(1 - e^{-\lambda x_i}) - \lambda \sum_{i=1}^n x_i. \quad (6)$$

2.2 MLE of α and λ

The maximum likelihood estimate (MLE) of α and λ are the solution of the following normal equation (obtained by equating the partial derivative of Eq. (6) w. r. t. α and λ to zero). Now, obtain the MLE of α and λ from above Eq. 5. We can solve the two non linear normal equation given as,

$$\frac{n}{\alpha} + \sum_{i=1}^n \ln(1 - e^{-\lambda x_i}) = 0, \quad (7)$$

$$\frac{n}{\lambda} + (\alpha - 1) \sum_{i=1}^n \frac{x_i e^{-\lambda x_i}}{1 - e^{-\lambda x_i}} - \sum_{i=1}^n x_i = 0. \quad (8)$$

It may be noted here that the normal equations thus obtained do not have explicit or closed solutions. Therefore, we have used the Newton–Raphson method for its numerical solution.

2.3 CI of α and λ

The confidence interval (CI) of α and λ for EED can be derived by the Fisher’s information matrix, which is given by

$$\mathbf{I} \begin{bmatrix} \alpha \\ \lambda \end{bmatrix} = E \begin{bmatrix} -\frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \alpha^2} & -\frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \alpha \partial \lambda} \\ -\frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \lambda \partial \alpha} & -\frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \lambda^2} \end{bmatrix},$$

where

$$\begin{aligned} \frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \alpha^2} &= -\frac{n}{\alpha^2}, \\ \frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \alpha \partial \lambda} &= \frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \lambda \partial \alpha} = \sum_{i=1}^n \frac{x_i e^{-\lambda x_i}}{1 - e^{-\lambda x_i}}, \\ \frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \lambda^2} &= -\frac{n}{\lambda^2} - (\alpha - 1) \sum_{i=1}^n x_i^2 e^{-\lambda x_i} \frac{(1 - 2e^{-\lambda x_i})}{(1 - e^{-\lambda x_i})^2}. \end{aligned}$$

The observed Fisher’s information matrix is

$$\mathbf{I} \begin{bmatrix} \hat{\alpha} \\ \hat{\lambda} \end{bmatrix} = E \begin{bmatrix} -\frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \alpha^2} & -\frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \alpha \partial \lambda} \\ -\frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \lambda \partial \alpha} & -\frac{\partial^2 \ln(L(\alpha, \lambda|\mathbf{x}))}{\partial \lambda^2} \end{bmatrix}_{\alpha=\hat{\alpha}, \lambda=\hat{\lambda}} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{bmatrix}.$$

Therefore, the observed variance–covariance matrix becomes $I^{-1}(\hat{\alpha}, \hat{\lambda})$. Here,

$$\mathbf{I} \begin{bmatrix} \hat{\alpha} \\ \hat{\lambda} \end{bmatrix} \sim N \left(\begin{bmatrix} \alpha \\ \lambda \end{bmatrix}, \mathbf{I}^{-1} \begin{bmatrix} \hat{\alpha} \\ \hat{\lambda} \end{bmatrix} \right),$$

and where the elements of

$$\mathbf{I}^{-1}(\hat{\alpha}, \hat{\lambda}) = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} \\ \mathbf{I}_{21} & \mathbf{I}_{22} \end{bmatrix}^{-1}.$$

A two-sided normal approximate $100(1 - \psi)\%$ CI for the parameters α and λ are $(\hat{\alpha} \pm z_{\psi/2} \sqrt{Var(\hat{\alpha})})$ and $(\hat{\lambda} \pm z_{\psi/2} \sqrt{Var(\hat{\lambda})})$ respectively, and where $z_{\psi/2}$ is the upper $(\psi/2)^{th}$ percentile of the standard normal distribution.

2.4 Bayes Estimate

In order to obtain the Bayes estimate of the parameters α and λ based on Kerala covid-19 data set. We assume that these parameters are independently uniformly distributed over the interval $(0, \alpha)$ and $(0, \lambda)$ respectively. Berger [6] discussed an estimation theory under uncertain prior information, see Jeffreys et al. [16] and Jeffreys [15]. The joint prior distribution of α and λ are given by

$$g(\alpha, \lambda) = \frac{1}{\alpha\lambda}; \quad \alpha > 0, \lambda > 0. \quad (9)$$

Using the joint prior density of Eq. (9) and likelihood of Eq. (5), the posterior density of α and λ is given by

$$\pi(\alpha, \lambda|\mathbf{x}) = \Psi_0(\mathbf{x})\alpha^{n-1}\lambda^{n-1}e^{-\lambda\sum_{i=1}^n x_i} \prod_{i=1}^n (1 - e^{-\lambda x_i})^{\alpha-1}, \quad (10)$$

where $\Psi_0(\mathbf{x})$ is a normalizing constant. Hence, the respective marginal posterior pdfs of α and λ are given by

$$\pi_1(\alpha|\mathbf{x}) = \int_{\lambda=0}^{\infty} \pi(\alpha, \lambda|\mathbf{x})d\lambda \quad (11)$$

and

$$\pi_2(\lambda|\mathbf{x}) = \int_{\alpha=0}^{\infty} \pi(\alpha, \lambda|\mathbf{x})d\alpha. \quad (12)$$

The Bayes estimator $\hat{\alpha}_B$ and $\hat{\lambda}_B$ of α and λ comes out to be

$$\hat{\alpha}_B = \int_0^{\infty} \alpha\pi_1(\alpha|\mathbf{x})d\alpha \quad (13)$$

and

$$\hat{\lambda}_B = \int_0^{\infty} \lambda\pi_2(\lambda|\mathbf{x})d\lambda, \quad (14)$$

respectively, provided the posterior expectation exist. The above integrals in Eqs. (13) and (14) are not obtainable in closed form. Therefore, Markov Chain Monte Carlo (MCMC) method are used for their calculation. In this method, first we generate samples from posterior distributions, through the Gibbs sampler via the Metropolis–Hastings algorithms, then this sample is used to compute Bayes estimates. The Gibbs algorithm are used for simulating the samples from the full conditional posterior distributions while the Metropolis–Hastings algorithm generate samples from an arbitrary proposal distribution (i.e. a Markov transition kernel). For details about this method, see Singh et al. [28]. Hence, conditional posterior distributions of the parameters α and λ can be written as

$$\pi_1^*(\alpha|\mathbf{x}) \propto \alpha^{n-1} \prod_{i=1}^n (1 - e^{-\lambda x_i})^{\alpha-1} \quad (15)$$

and

$$\pi_2^*(\lambda|\mathbf{x}) \propto \lambda^{n-1}e^{-\lambda\sum_{i=1}^n x_i} \prod_{i=1}^n (1 - e^{-\lambda x_i})^{\alpha-1}, \quad (16)$$

respectively. The Bayes estimate for α and λ , the required sample from Eqs. (15) and (16) are generated using MCMC procedure. For the details of the algorithm used, see Singh et al. [27].

3 Bayes Prediction

Let us consider a future sample $\mathbf{y} = (y_1, y_2, \dots, y_m)'$ of size m , independent of the informative sample $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ and let $y_{(1)} < y_{(2)} < \dots < y_{(r)} < \dots < y_{(m-1)} < y_{(m)}$ be the sample order statistics. Suppose that we are interested in the posterior predictive density of the future order statistic $y_{(r)}$ given the informative set of data x_1, x_2, \dots, x_n . If the pdf of the r^{th} order statistic in the future sample is denoted by $h_{(r)}(\cdot|\alpha, \lambda)$, then

$$h_{(r)}(y|\alpha, \lambda) = \frac{1}{\text{Beta}(r, m-r+1)} [F(y|\alpha, \lambda)]^{r-1} [1 - F(y|\alpha, \lambda)]^{m-r} f(y|\alpha, \lambda), \quad (17)$$

here $f(\cdot|\alpha, \lambda)$ and $F(\cdot|\alpha, \lambda)$ are same as above Eqs. (1) and (2), respectively. The predictive density of $y_{(r)}$ is denoted as $h_{(r)}^*(\cdot|\mathbf{x})$, then

$$h_{(r)}^*(y|\mathbf{x}) = \int_{\alpha=0}^{\infty} \int_{\lambda=0}^{\infty} h_{(r)}(y|\alpha, \lambda) \pi(\alpha, \lambda|\mathbf{x}) d\alpha d\lambda, \quad (18)$$

where $\pi(\alpha, \lambda|\mathbf{x})$ is the joint posterior density of α and λ as given Eq. (10). It is note that $h_{(r)}^*(y|\mathbf{x})$ cannot be expressed in closed form and hence it cannot be evaluated analytically. When $r = 1$ & $m = 1$ in Eq. (18), it becomes the one sample prediction given below

$$h^*(y|\mathbf{x}) = \int_{\alpha=0}^{\infty} \int_{\lambda=0}^{\infty} f(y|\alpha, \lambda) \pi(\alpha, \lambda|\mathbf{x}) d\alpha d\lambda. \quad (19)$$

Therefore, the consistent solution of $h_{(r)}^*(y|\mathbf{x})$ can be obtained by using the Gibbs sampler via the Metropolis–Hastings algorithm. Hence, the consistent estimator of $h_{(r)}^*(y|\mathbf{x})$ is given by

$$\hat{h}_{(r)}^*(y|\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M h_{(r)}(y|\alpha_i, \lambda_i), \quad (20)$$

where $(\alpha_i, \lambda_i), i = 1, \dots, M$ are obtained from $\pi(\alpha, \lambda|\mathbf{x})$ using the MCMC technique. Along the same line, if we want to estimate the predictive distribution of $y_{(r)}$, say $H_{(r)}^*(\cdot|\mathbf{x})$, then a consistent estimator of $H_{(r)}^*(y|\mathbf{x})$ can be obtained as

$$\hat{H}_{(r)}^*(y|\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M H_{(r)}(y|\alpha_i, \lambda_i). \quad (21)$$

Where $H_{(r)}(y|\alpha, \lambda)$ denotes the distribution function of $h_{(r)}(y|\alpha, \lambda)$, which is given by

$$H_{(r)}(y|\alpha, \lambda) = \frac{1}{\text{Beta}(r, m - r + 1)} \int_0^{F(y|\alpha, \lambda)} [F(t|\alpha, \lambda)]^{r-1} [1 - F(t|\alpha, \lambda)]^{m-r} f(t|\alpha, \lambda) dt. \tag{22}$$

It may also noted here that the same MCMC sample $(\alpha_i, \lambda_i), i = 1, 2, \dots, M$ can be used to compute $\hat{h}_{(r)}^*(y|\mathbf{x})$ or $\hat{H}_{(r)}^*(y|\mathbf{x})$ for all y . Further an important problem to establish a two sided predictive interval of the r^{th} order statistic $Y_{(r)}$ drawn from a future sample $\{Y_1, \dots, Y_m\}$ of size m , and it does not depend on the informative sample $\{x_1, \dots, x_n\}$.

Now we construct a $100\psi\%$ predictive interval for $Y_{(r)}$. In symmetric case, the $100\psi\%$ predictive interval for $Y_{(r)}$ can be obtained by

$$\frac{1 + \psi}{2} = P(Y_{(r)} > L|\mathbf{x}) = 1 - H_{(r)}^*(y|\mathbf{x}) \Rightarrow H_{(r)}^*(y|\mathbf{x}) = \frac{1}{2} - \frac{\psi}{2} \tag{23}$$

and

$$\frac{1 - \psi}{2} = P(Y_{(r)} > U|\mathbf{x}) = 1 - H_{(r)}^*(y|\mathbf{x}) \Rightarrow H_{(r)}^*(y|\mathbf{x}) = \frac{1}{2} + \frac{\psi}{2} \tag{24}$$

for the lower bound L and upper bound U respectively, see Al-Jarallah and Al-Hussaini [3]. While in asymmetric case, the predictive interval of the form (L, ∞) and $(0, U)$ with the coverage probability ψ can be obtained by

$$P(Y_{(r)} > L|\mathbf{x}) = \psi \Rightarrow H_{(r)}^*(y|\mathbf{x}) = 1 - \psi \tag{25}$$

and

$$P(Y_{(r)} > U|\mathbf{x}) = 1 - \psi \Rightarrow H_{(r)}^*(y|\mathbf{x}) = \psi \tag{26}$$

for L and U respectively. It is very difficult to obtain the analytical solutions. Therefore, we may apply some right numerical techniques for solving these nonlinear equations.

4 Model Selection Methods

4.1 Information Criterion

Entropy as an information criteria in the statistics and tells us that estimate of parameter is the disorder or lack of information about the sampling of distribution. The modeling process can be judged by Akaike’s information criterion (AIC) and Bayesian information criterion (BIC).

4.2 AIC

This is a tool for model selection, it was developed by Akaike [2]. The estimated statistical models and the goodness of fit are measured through AIC. For the considered

data, all the considered competing models may be ranked according to their AIC, with the one having the lowest AIC being the best. This can be evaluated by

$$AIC = 2k - 2 \ln(L(\alpha, \lambda|\mathbf{x})).$$

4.3 BIC

BIC has developed by Schwarz et al. [24], for model selection among the considered statistical models with different number of parameters. The MLE method is used to estimate the model parameter, it is possible to increase the likelihood by adding parameters, which may result in over fitting. The BIC resolves this problem by introducing a penalty per parameter term for the number of parameters in the model. This led to the AIC, and can be evaluated by

$$BIC = k \ln(n) - 2 \ln(L(\alpha, \lambda|\mathbf{x})).$$

The lower value of BIC is preferred among the estimated considered models. It may also noted that, k is the number of parameters in the model and n is the sample size.

4.4 Nonparametric Empirical Posterior Probability

In Bayesian setup, we want to choose a most appropriate model from a bunch of r fully specified models. Let $f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_r(\mathbf{x})$ be densities or models with underlying observation vector $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ for $\mathbf{x} \in S$. We want to extract most important and informative model for $\mathbf{x} \in S$. For this, we assume that $\phi_1, \phi_2, \dots, \phi_r$ be the subjective probabilities of each models with $\sum_{j=1}^r \phi_j = 1$, where ϕ_j denotes the probability of the j^{th} model. Although if there is no any further information to distinguish models, then the unbiased choices $\phi_j = 1/r$, for $j = 1, 2, \dots, r$ are reasonable, and ϕ_j is referred as prior probabilities for j^{th} model, see Leonard and Hsu [19]. Hence, empirical posterior probability of the j^{th} model is given by

$$\phi_j^* = \frac{\phi_j f_j(\mathbf{x})}{\sum_{i=1}^r \phi_i f_i(\mathbf{x})}, \quad j = 1, 2, \dots, r. \quad (27)$$

The posterior mean value function of sampling density is

$$p^*(\mathbf{u}) = \frac{\sum_{j=1}^r \phi_j(\mathbf{u}) f_j(\mathbf{x})}{\sum_{j=1}^r \phi_j f_j(\mathbf{x})}, \quad \mathbf{u} \in S, \quad (28)$$

where S denotes the set of all possible realization of the observation vector $\mathbf{x} = (x_1, x_2, \dots, x_n)'$.

Fig. 3 TTT plot for Kerala covid-19 data

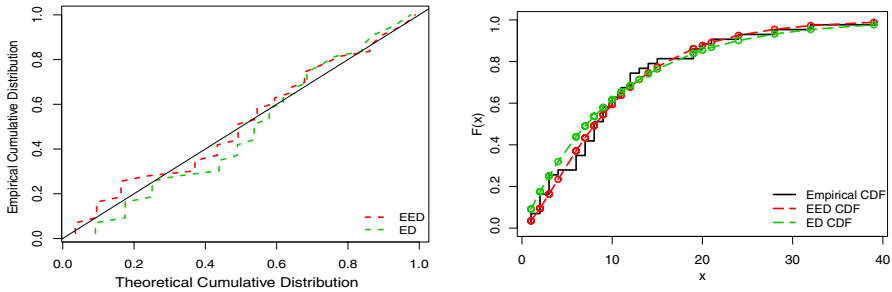
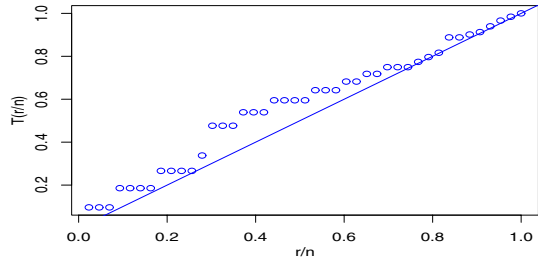


Fig. 4 PP plot and CDF plot for Kerala covid-19 data

5 Kerala Covid-19 Real Data Analysis

The covid-19 positive confirmed number of patients data are taken from Kerala state between 09-03-2020 to 23-04-2020. The daily bulletin of data has been issued by covid-19 outbreak control and prevention state cell, health and family welfare department, directorate of health services, government of Kerala. Available at <https://dhs.kerala.gov.in/>. The covid-19 confirmed number of patients data are reported as below

3, 6, 8, 2, 3, 2, 3, 1, 12, 12, 15, 28, 14, 9, 19, 39, 6, 20, 32, 7, 24, 21, 9,
11, 8, 13, 8, 9, 12, 7, 10, 2, 3, 8, 1, 7, 1, 4, 2, 6, 19, 11, 10.

Between these date the total 447 persons have been tested covid-19 positive, and 129 persons are considered under treatment.

According to an idea with associated failure rate, we considered, a graphical method based on total time on test (TTT) plot as a crude indicator see Aarset [1]. The empirical TTT is given as

$$T(r/n) = \frac{\sum_{i=1}^r x_{(i)} + (n - r)x_{(r)}}{\sum_{i=1}^n x_{(i)}}$$

where $r = 1, 2, \dots, n$ and $x_{(r)}$ is the order statistics of the sample.

Figure 3 shows the TTT plot, which is concave indicating that data relates to an increasing failure rate. Thus, it can be properly accommodated by a EED. The fitting of EED was checked using PP plot and CDF plot given in Fig. 4.

5.1 Goodness of Fit Test

5.2 LR Test

By testing the hypothesis that $\alpha = 1$, we can see if the EED provides a better prediction of the Kerala covid-19 data of the component compared with one parameter ED. Note that the graphical presentation of empirical and theoretical cdf can be verified in case that EED can explain the data well enough. The likelihood ratio (LR) test will be discussed here for theoretical justification purpose, see a recent work on LR test by Pathak et al. [22]. The test on the shape parameter α , the null hypothesis is $H_0 : \alpha = 1$, and the test statistic

$$\Lambda = -2 \ln \left(\frac{L(1, \hat{\lambda}|\mathbf{x})}{L(\hat{\alpha}, \hat{\lambda}|\mathbf{x})} \right),$$

is asymptotically with χ_1^2 . Here $L(\hat{\alpha}, \hat{\lambda}|\mathbf{x})$ is the likelihood function of the EED, $\hat{\alpha}, \hat{\lambda}$ is the MLE, and $1, \hat{\lambda}$ is the MLE under condition of H_0 . The MLE of the parameters for Kerala covid-19 data are $\hat{\alpha} = 1.572179, \hat{\lambda} = 0.126561$ and $\ln(L(\hat{\alpha}, \hat{\lambda}|\mathbf{x})) = -141.6486$. Hence, the observed Fisher’s information matrix \mathbf{I} is

$$\mathbf{I} = \begin{bmatrix} 17.39312 & -187.3144 \\ -187.31445 & 3963.9184 \end{bmatrix}.$$

Thus, the inverse of \mathbf{I} is given by

$$\mathbf{I}^{-1} = \begin{bmatrix} 0.117074044 & 0.0055323187 \\ 0.005532319 & 0.0005137046 \end{bmatrix}^{-1}.$$

Now, we can obtained the asymptotic 95% intervals for the parameters as summarized in Table 1.

Under $H_0 : \alpha = 1$, the MLE of λ is 0.096196, and $\ln(L(1, \hat{\lambda}|\mathbf{x})) = -143.6784$. Thus, the ratio test statistic is $\Lambda = 0.04391913$, and corresponding p value is 0.9560809. Hence, we have enough evidence (i.e., more than 95% level of

Table 1 The 95% confidence intervals for α and λ based on MLE

Parameters	MLE	Confidence interval
α	1.572179	(0.9015562, 2.242802)
λ	0.1265606	(0.0821379, 0.1709833)

confidence) to reject our null hypothesis and conclude that the sub model ED (when $\alpha = 1$) cannot give a good enough fit to Kerala covid-19 data.

5.3 K–S Test

It is well known that Kolmogorov–Simirnov (K–S) test is applicable even in the case of very small samples as well as large samples. This test assumes continuous of the distribution function, to check difference between $F_n(x)$ and $F(x)$, see Gibbons and Chakraborti [12]. Since, to test

$$H_0 : F_n(x) = F(x),$$

$$H_1 : F_n(x) \neq F(x),$$

where $F_n(x)$ is the sample (empirical) distribution function, $F(x)$ is specified EED for all x . The test statistic

$$D_n = \sup|F_n(x) - F(x)|$$

is less then tabulated value of K–S distance $D_{n,\psi}$ then accept H_0 otherwise reject. The statistic D_n is used to obtain the confidence bands on $F_n(x)$ for all x , where $F_n(x)$ is a consistent estimator for cdf $F(x)$. The number $D_{n,\psi}$ is obtain from the K–S table (critical value), such that

$$P[\sup|F_n(x) - F(x)| < D_{n,\psi}] = 1 - \psi,$$

where, $0 \leq F(x) \leq 1, \forall x$. Thus we define

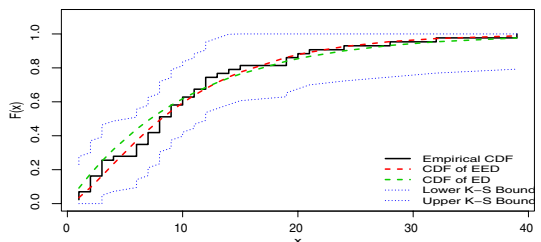
$$L_n(x) = \max[F_n(x) - D_{n,\psi}, 0],$$

and

$$U_n(x) = \min[F_n(x) + D_{n,\psi}, 1],$$

where $L_n(x)$ and $U_n(x)$ are lower and upper confidence band for the cdf $F(x)$, with $100(1 - \psi)\%$ confidence coefficient. Of course, the $F(x)$ lies completely within the limits if and only if the hypothesis cannot be rejected at $\psi\%$ level of significance. Hence, the K–S test has been applied on Kerala covid-19 data, and value of the test statistics $0.093023 < 0.207398$ (critical value) and the K–S bound plot in Fig. 5, which shows that EED provides a satisfactory fit to the Kerala covid-19 data.

Fig. 5 The K–S bound plot for Kerala covid-19 data



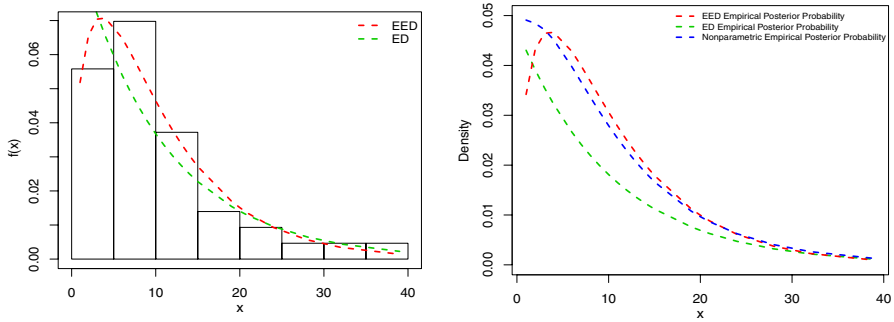


Fig. 6 The fitted density and empirical posterior probability plot for Kerala covid-19 data

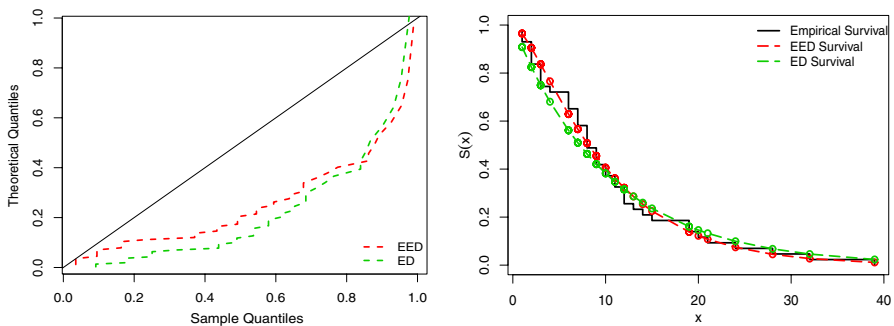


Fig. 7 The fitted sample quantile and survival plot for Kerala covid-19 data

5.4 Discussions

By all the above criteria of goodness of fit as well as discussed above model suitability for considered Kerala covid-19 data, we see that from the fitted density and empirical posterior probability plots in Fig. 6 and sample quantile plot in Fig. 7 of EED provide better fit than ED. Also, from Table 2, it can be seen that the quantiles of the EED and ED at various points of Kerala covid-19 data are very less similar to each. The quantiles value of the EED were found to be more than ED, and hence it covers significantly more information than ED. Moreover, the main advantage of taking EED over ED is that it has accommodate different types of hazard rate.

According to the results shown in Table 3, we observed that the value of $-\log(\text{Likelihood})$, p value, and empirical posterior probability values of EED are larger than ED for Kerala covid-19 data. Also, the value of K–S distance, AIC and BIC are found lesser than ED. Therefore, on the basis of Table 3 we conclude that the EED satisfactorily modeled the Kerala covid-19 data.

Now, we obtained the MLE of α , λ and corresponding Bayes estimates with non-informative priors for the Kerala covid-19 data. The MLE of α and λ can be obtained from above Eqs. 7 and 8 respectively. The computation of MLE of α and

Table 2 The EED and ED quantiles of the Kerala covid-19 data

<i>Quantiles Model</i>	2.5%	5%	10%	25%	50%	75%	90%	95%	97.5%
<i>EED</i>	0.794988	1.272541	2.077198	4.223451	8.148912	14.1312455	21.61929	27.17217	32.68593
<i>ED</i>	0.200045	0.405286	0.832491	2.273078	5.476801	10.9536023	18.19354	23.67034	29.14714

λ for Kerala covid-19 data samples, we used the Newton-Raphson method with the initial guess for α and λ have been taken from contour plots, which is shown in Fig. 8 and their MLE are verified from Fig. 9.

For the calculation of Bayes estimate under the prior in Eq. 9, and are obtained by the expected value of α and λ with respect to their marginal posteriors as stated above Eqs. (11) and (12) respectively. The Bayes estimate of α and λ are obtained as $\hat{\alpha}_B$ and $\hat{\lambda}_B$, respectively, and corresponding 95% HPD intervals are summerise in Table 4.

The highest predictive density interval (HPDI) is the interval that includes the most probable values of a given predictive density at a given significance level, subject to the condition that the density function has the same value at the end points. For more about HPD intervals, the reader is referred to Box and Tiao [8]. We have reported certain HPD intervals by using the Kerala covid-19 data, see Table 5.

Figure 10 represents the posterior predictive density and distribution function for the future observations. The graphs are displayed based on future observations.

6 Conclusion

The EED studied in this article is flexible in modeling with increasing failure rate function. Based on different types of goodness of fit tests, we saw that EED provide better fit for Kerala covid-19 data. The MLE, Bayes estimates and nonparametric

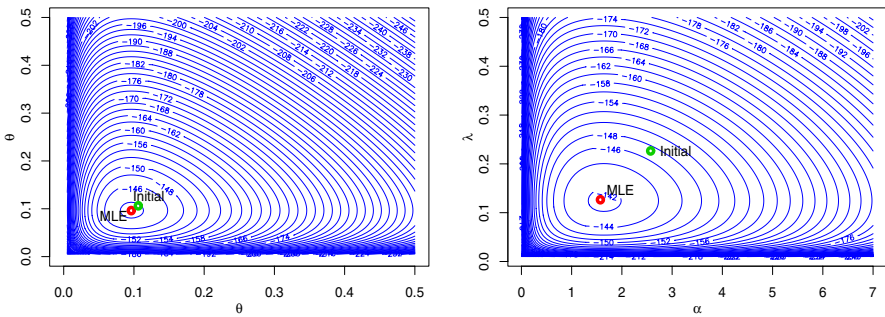


Fig. 8 The contour plot of α and λ for Kerala covid-19 data

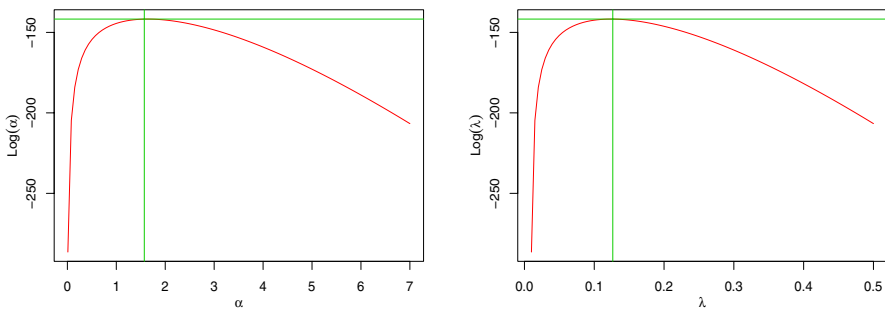


Fig. 9 The likelihood plot of α and λ for Kerala covid-19 data

Table 3 The MLE of parameters, $-\log(\text{Likelihood})$, CI, K–S distance, K–S critical value, p value, AIC, BIC and empirical posterior probability values for different models

Distribution	$-\log(\text{Likelihood})$	MLE	CI	K–S	Critical value	p value	AIC	BIC	Empirical posterior probability
<i>EED</i>	141.6486	$\hat{\alpha} = 1.572179$	(0.90156, 2.2428)	0.093023	0.207398	0.992337	287.2972	290.8196	0.50738
		$\hat{\lambda} = 0.126561$	(0.08214, 0.17098)						
<i>ED</i>	143.6784	$\hat{\alpha} = 0.096196$	(0.06741, 0.12498)	0.139535	0.207398	0.796649	289.3568	291.118	0.49262

empirical posterior probability are derived. Among various types of model selection criterion as well as proposed empirical posterior probability, we observe that EED has perform better than ED. The MLE, AIC, BIC, CIs and HPDs are reported in Tables 3 and 4, respectively. Since we have no prior information about data, the Bayesian estimation is done under non-informative prior. Moreover, we have also verified with the help of Table 5, the Bayes predictive inference for one sample future observations of a real data set that are positive skewed. Also, we have seen in Table 5, for next 43 days future covid-19 patients will be detected positive

Table 4 The Bayes estimate and 95% HPD intervals for α and λ

Parameters	Bayes estimate	HPD Interval
$\hat{\alpha}_B$	1.576098	(1.25246, 1.889621)
$\hat{\lambda}_B$	0.1262713	(0.08057642, 0.1699846)

Table 5 Summary results of the Bayes predictive inference for a future observations

Predictive results	Row moments	Corrected moments
Kerala covid-19 data		
n = 43	$\mu'_1 = 12.70755$	$\mu_2 = 116.4683$
Mean = 12.70755	$\mu'_2 = 277.9502$	$\mu_3 = 1846.827$
SD = 10.91977	$\mu'_3 = 8338.947$	$\mu_4 = 68950.5$
$\hat{\alpha} = 1.576098$	$\mu'_4 = 301746.6$	$\beta_1 = 1.469314$
$\hat{\lambda} = 0.126271$		$\beta_2 = 5.083021$
$\Psi(y) = 1.653819e-16$		
HPD Intervals	Coefficient of Skewness and Kurtosis	
90% HPDI: (1.575177, 27.8505)	$\gamma_1 = \sqrt{\beta_1} = 1.212153$	
95% HPDI: (1.054455, 28.6958)	$\gamma_2 = \beta_2 - 3 = 2.083021$	
97% HPDI: (1.054455, 41.36516)		
99% HPDI: (1.054455, 49.81872)		

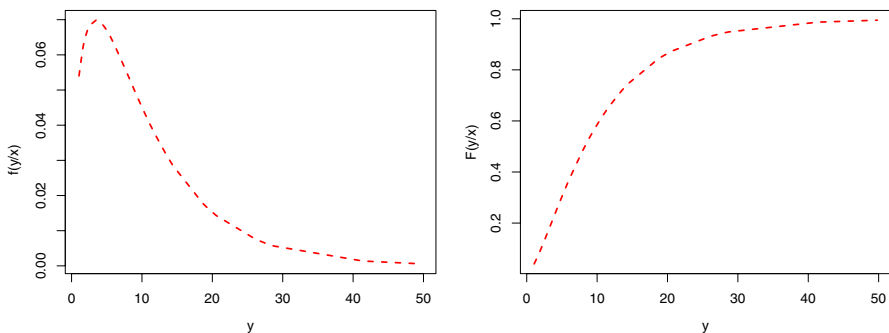


Fig. 10 The Posterior predictive density and distribution function based on future observations

approximately 546 persons. Thus we recommend that EED may be used in future research for Bayesian inference with positive skewed data.

Acknowledgements Authors are very heartfelt thank to the editor-in-chief, associate editors and referees for their valuable suggestions which make the path to improvements in this paper.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest and no funding was received for conducting this study. If any of the sections are not relevant to the manuscript, kindly mention as ‘Not applicable’ under the heading ‘Conflict of Interest’ statement.

References

1. Aarset MV (1987) How to identify a bathtub hazard rate. *IEEE Trans Reliab* 36(1):106–108
2. Akaike H (1978) A Bayesian analysis of the minimum AIC procedure. *Ann Inst Stat Math* 30(1):9–14
3. Al-Jarallah R, Al-Hussaini E (2007) Bayes inference under a finite mixture of two-compound Gompertz components model. *J Stat Comput Simul* 77(11):915–927
4. Arshad Ali S, Baloch M, Ahmed N, Arshad Ali A, Iqbal A (2020) The outbreak of coronavirus disease 2019 (covid-19)-an emerging global health threat. *J Infect Public Health* 13(4):644–646. <https://doi.org/10.1016/j.jiph.2020.02.033>
5. Baklizi AS (2004) Bayesian prediction intervals for ranges and waiting times. *J Interdiscip Math* 7(2):165–171
6. Berger JO (2013) *Statistical decision theory and Bayesian analysis*. Springer, New York
7. Bernardo JM, Smith AF (2009) *Bayesian theory*, vol 405. Wiley, New York
8. Box GE, Tiao GC (2011) *Bayesian inference in statistical analysis*, vol 40. Wiley, New York
9. Chen Y, Liu Q, Guo D (2020) Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol* 92(4):418–423
10. Geisser S (1993) *Predictive inference*, vol 55. CRC Press, Boca Raton
11. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian data analysis*. Chapman and Hall/CRC, London
12. Gibbons JD, Chakraborti S (2014) *Nonparametric statistical inference: revised and expanded*. CRC Press, London
13. Gupta RC, Gupta PL, Gupta RD (1998) Modeling failure time data by Lehman alternatives. *Commun Stat Theory methods* 27(4):887–904
14. Gupta RD, Kundu D (2001) Exponentiated exponential family: an alternative to gamma and Weibull distributions. *Biomet J J Math Methods Biosci* 43(1):117–130
15. Jeffreys H (1946) An invariant form for the prior probability in estimation problems. *Proc R Soc Lond A* 186(1007):453–461
16. Jeffreys H et al (1939) *Theory of probability*. The Clarendon Press, Oxford
17. Khakharia A, Shah V, Jain S, Shah J, Tiwari A, Daphal P, Warang M, Mehendale N (2021) Outbreak prediction of covid-19 for dense and populated countries using machine learning. *Ann Data Sci* 8(1):1–19
18. Kumar S (2020) Monitoring novel corona virus (covid-19) infections in India by cluster analysis. *Ann Data Sci* 7:417–425
19. Leonard T, Hsu JS (2001) *Bayesian methods: an analysis for statisticians and interdisciplinary researchers*, vol 5. Cambridge University Press, Cambridge
20. Li Q et al (2020) An outbreak of NCIP (2019-nCoV) infection in China-Wuhan, Hubei province, 2019–2020. *China CDC Wkly* 2(5):79–80
21. Liu Y, Gu Z, Xia S, Shi B, Zhou XN, Shi Y, Liu J (2020) What are the underlying transmission patterns of covid-19 outbreak? An age-specific social contact characterization. *EClinicalMedicine* 22:100354

22. Pathak A, Kumar M, Singh SK, Singh U (2020) Bayesian inference: Weibull Poisson model for censored data using the expectation–maximization algorithm and its application to bladder cancer data. *J Appl Stat* 1–23. <https://doi.org/10.1080/02664763.2020.1845626>
23. Paules CI, Marston HD, Fauci AS (2020) Coronavirus infections—more than just the common cold. *JAMA* 323(8):707–708
24. Schwarz G et al (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464
25. Shi Y, Tian Y, Kou G, Peng Y, Li J (2011) Optimization based data mining: theory and applications. Springer, New York
26. Singh SK (2011) Estimation of parameters and reliability function of exponentiated exponential distribution: Bayesian approach under general entropy loss function. *Pak J Stat Oper Res* 7(2):217–232
27. Singh SK, Singh U, Kumar M (2013) Estimation of parameters of generalized inverted exponential distribution for progressive type-II censored sample with binomial removals. *J Probab Stat* 2013:1–12
28. Singh SK, Singh U, Kumar M (2016) Bayesian estimation for Poisson-exponential model under progressive type-II censoring data with binomial removal and its application to ovarian cancer data. *Commun Stat Simul Comput* 45(9):3457–3475
29. Temesgen A, Gurmesa A, Getchew Y (2018) Joint modeling of longitudinal CD4 count and time-to-death of HIV/TB co-infected patients: a case of Jimma University specialized hospital. *Anna Data Sci* 5(4):659–678
30. Tien JM (2017) Internet of things, real-time decision making, and artificial intelligence. *Ann Data Sci* 4(2):149–178
31. WHO (2020) Coronavirus disease 2019 (covid-19) situation report-71. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200331-sitrep-71-covid-19.pdf?sfvrsn=4360e92b_8/. Online; Accessed 31 Mar 2020

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.