REGULAR PAPER

Generalization of the Lindley distribution with application to COVID-19 data

C. S. Rajitha[1] · **A Akhilnath**[1]

Abstract

Creating new distributions with more desired and flexible qualities for modeling lifetime data has resulted in a concentrated effort to modify or generalize existing distributions. In this paper, we propose a new distribution called the power exponentiated Lindley (PEL) distribution by generalizing the Lindley distribution using the power exponentiated family of distributions, that can fit lifetime data. Then the main statistical properties such as survival function, hazard function, reverse hazard function, moments, quantile function, stochastic ordering, MRL, order statistics, etc., of the newly proposed distribution have been derived. The parameters of the distribution are estimated using the MLE method. Then, a Monte Carlo simulation study is used to check the consistency of the parameters of the PEL distribution in terms of MSE, RMSE, and bias. Finally, we implement the PEL distribution as a statistical lifetime model for the COVID-19 case fatality ratio (in %) in China and India, and the new cases of COVID-19 reported in Delhi. Then we check whether the new distribution fits the data sets better than existing well-known distributions. Different statistical measures such as the value of the log-likelihood function, K-S statistic, AIC, BIC, HQIC, and p-value are used to assess the accuracy of the model. The suggested model seems to be superior to its base model and other well-known and related models when applied to the COVID-19 data set.

Keywords Lindley distribution · Maximum likelihood estimation · Monte Carlo simulation study · Power exponentiated family · Quantile function · Stochastic ordering

1 Introduction

Now China is experiencing the most significant increase in COVID-19 cases because of the new variant of coronavirus known as the 'stealth' sub-variant of omicron, also known as BA.2. The first case of this variant was reported in March 14*th* of 2022 in China. This outbreak occurred when corona virus infections are declining in several other nations, and limitations are being eased. The BA.2 variant is anticipated to have the same severity as the 'original version', according to the World Health Organization (WHO). Same time in India, cases are reducing, and restrictions are removed. All public places are opened to the public. The vaccination process is progressing in every country. Following the outbreak of COVID-19, a slew of researchers have begun to investigate and use the virus data for various purposes. Liu et al. [18] proposed the arc-sine-modified Weibull distribution to model the mortality rates of COVID-19 cases in China. Pathak et al. [23] used the exponentiated exponential distribution as a suitable statistical lifetime model for Kerala COVID-19 patient data. Nagy et al. [22] introduced a superior discrete statistical model for the COVID-19 mortality numbers in Saudi Arabia and Latvia. Riad et al. [26] introduced a new lifetime distribution named the power Bilal (PB) distribution to model COVID-19 data and investigated its statistical properties. Ahsan-ul-Haq et al. [3] used different probability distributions to model COVID-19 cases in Pakistan; these are some of the recent studies based on COVID-19 data. As a result, in this study, we used COVID-19 data from Delhi, India, and China.

Once all distributional assumptions are satisfied, the accuracy of parametric inference and data set modeling is mostly influenced by how well the presented data fits the probability

---

C. S. Rajitha and A. Akhilnath have authors contributed equally to this work.

✉ C. S. Rajitha
rajitha.sugun@gmail.com; cs_rajitha@cb.amrita.edu

A Akhilnath
akhilnathakhi888@gmail.com

[1]   Department of Mathematics, Amrita School of Physical Sciences, Amrita Vishwa Vidyapeetham, Coimbatore, India

distribution. Several studies have been conducted in order to create probability distributions with significantly more accurate and adaptable characteristics which can model real-life data sets of different kinds. The need to produce new distributions arises from theoretical concerns, actual applications, or both. There has been significant development in the generalization of some well-known distributions and their practical application to challenge other well-known distributions. The exponential distribution is ideal for portraying the lifetime data, like for many types of manufactured items. The main feature of the exponential distribution is that it may be used to simulate the behavior of things with a fixed failure rate. It has a straightforward mathematical structure that makes it simple to manipulate. While discussing the sampling of standard deviation (SD), Kondo [16] in 1930 referred to the exponential distribution as Pearson's X Type distribution. Then, Jowett [14] proposed the exponential distribution and its applications in 1958. Due to analytical simplicity, the exponential distribution was commonly employed to describe failure time data in the early days of reliability theory research. However, in most modern reliability engineering issues, the exponential distribution's constant failure rate property is not necessarily appropriate. Hence, many generalized forms of this distribution came into existence. Some of them are the generalized exponential distributions by Gupta and Kundu [13], the k-generalized exponential distribution proposed by Rather and Rather [25], the extended exponentiated exponential distribution and its properties by Abu et al. [2].

Also, numerous lifetime distributions have been proposed as an alternative to exponential distribution for modeling lifetime data in the literature. Lindley [17] presented the Lindley distribution to demonstrate that the Bayes' and fiducial distributions are not always identical. Later, Ghitany et al. [11] investigated the properties of the Lindley distribution and found that it is a better alternative to the exponential distribution. If a continuous random variable X follows the Lindley distribution, then cumulative distribution function (cdf) with parameter q is provided by

$$G(x) = 1 - \frac{e^{-qx}(1 + q + qx)}{1 + q}; \quad x > 0, q > 0 \qquad (1)$$

and probability density function (pdf) as

$$g(x) = \frac{q^2}{(q+1)}(1 + x)e^{-qx}; \quad x > 0, q > 0 \qquad (2)$$

Ghitany et al. [10] also introduced the power Lindley distribution as a new generalization of the Lindley distribution. Later, Nadarajah et al. [21] introduced a new generalization of the Lindley distribution called the generalized Lindley distribution. Sakthivel et al. [27] introduced two parameter

cubic rank transmutation of Lindley distribution. Ashly and Rajitha [7] proposed the negative binomial improved second degree Lindley distribution. Ashour and Eltehiwy [8] proposed the Lindley distribution with three parameters known as the exponentiated power Lindley distribution, which has got a wide range of applications in fields including survival analysis, reliability, biology, and others. The power XLindley (PXL) distribution was introduced by Meriem et al. [19]. It is a two-parameter distribution that extends the XLindley distribution. Recently the negative binomial Akash distribution and its applications was proposed by Rajitha and Ashly [24]. Besides, some heavy-tailed distributions have also been developed by many researchers. Teamah et al. [29]. introduced a relatively new heavy-tailed statistical model using alpha power transformation and exponentiated log-logistic distribution, called alpha power exponentiated log-logistic distribution.

Furthermore, researchers have shown a growing interest in developing new families of distributions to improve the accuracy of fitting complicated forms of data. The new families were established using a variety of ways that included adding extra location, scale, shape, and transmuted characteristics. Some of the newly proposed families are the Chen-G family by Anzagra et al. [6], the sine Topp-Leone-G family by Al-Babtain et al. [4], the Poisson-G family by Abouelmagd et al. [1], and the odd Lomax trigonometric generalized family of distributions by Alshanbari et al. [5] etc. Recently, Klakattawi et al. [15] proposed the Marshall–Olkin Weibull generated family. With the help of the power exponentiated family of distributions proposed by Modi [20], we propose a new generalization of the Lindley distribution in this study. The cdf and pdf of the power exponentiated family are given by

$$F(x) = \frac{(v^{(G(x))^{\alpha}} - 1)}{(v - 1)}; \quad x > 0, \alpha > 0, v > 0, v \neq 1 \qquad (3)$$

and

$$f(x) = \frac{(\alpha v^{(G(x))^{\alpha}} ln(v)(G(x))^{\alpha-1} g(x))}{(v - 1)};$$
$$x > 0, \alpha > 0, v > 0, v \neq 1 \qquad (4)$$

where $G(x)$ and $g(x)$ are pdf and cdf of any baseline distribution.

The main motivation of this article is to introduce a new superior model capable of modeling and fitting different types of data. We also want to show the superiority of the new model in beating all its competitors and recommend the proposed distribution as a strong and novel candidate for modeling real data sets. When modeling a situation with a known distribution is challenging, we may use generalization to account for extra data variation. The challenges that are currently present

are evolving significantly along with our world. Because of this, we require additional generalizations of probability distributions to capture more complex data. As COVID-19 is the new problem faced by almost every country, we tried to model those data using generalized form of different distributions, but the result was not that convincing. As a result, we proposed a new distribution to solve this. The following things impacted the creation of this work:

- Although the PEL distribution is limited to the closed form expression of the characteristics of the distribution, it is simple to implement.
- Explicitly defining the statistical characteristics is straightforward.
- PEL distribution is suitable for modeling, decreasing, increasing, and upside-down lifetime data. Its density can be symmetric or right skewed and many various shapes.
- Also, its hazard rate can be increasing, upside-down, increasing constant.
- The novel PEL distribution is adequate for skewed data that may not be adequately fitted by other distributions.
- Also, it may be used pretty well, to analyse a large number of real-life data sets and fits them quite well, also it can be used in various problems in applied areas such as medicine, engineering, industrial reliability, COVID-19 analysis, and survival analysis.
- The PEL distribution has simple closed forms for both the cdf and hazard rate functions. Hence it will be useful to work with censored and complete samples as well.

This study aims to develop a new distribution called the power exponentiated Lindley (PEL) distribution and derive some of its characteristics. In Sect. 2, we introduce the pdf and cdf of the PEL distribution along with its mixture representation. Then in Sect. 3, some of the essential statistical properties are derived, such as hazard rate, survival function, reverse hazard function, moments, moment generating function, mean residual life, mean past lifetime, stochastic ordering, and order statistics. In Sect. 4, we find the maximum likelihood estimates (MLE) for the parameters of the PEL distribution using the log-likelihood function. Section 5 is about the Monte Carlo simulation study which is used to check the consistency of the parameters of the PEL distribution in terms of bias, mean squared error (MSE), and root mean squared error (RMSE). Then in Sect. 6 we discuss about the application of the PEL distribution using COVID 19 case fatality ratio (in %) in China and India, and the new cases of COVID-19 reported in Delhi. Further, we prove that the PEL distribution is more efficient in fitting data than other well-known existing distributions. In Sect. 7, the results of the simulation study and the application of real-life data are addressed. Finally, in Sect. 8, conclusion is presented.

2 The power exponentiated Lindley distribution

With the help of the power exponentiated family of distributions from Eqs. (3) and (4), we introduce a new generalization of the Lindley distribution called the PEL distribution. The cdf for the PEL distribution with α and v as shape parameters and q as scale parameter can be obtained as:

$$
F(x) = \frac{v^{\{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} - 1}{v - 1};
$$
$$
x > 0, \alpha > 0, q > 0, v > 0, v \neq 1 \tag{5}
$$

By differentiating this cdf, the corresponding pdf is obtained as,

$$
f(x) = \frac{\alpha v^{\{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} ln(v)\{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha-1} q^2(1+x)e^{-qx}}{(v-1)(q+1)};
$$
$$
x > 0, \alpha > 0, q > 0, v > 0, v \neq 1 \tag{6}
$$

2.1 Expansion of the PEL distribution

Expansion of pdf and cdf of the PEL distribution is useful while deriving their properties. For this purpose, we use the following two lemmas:

Lemma 2.1.1 *If λ is a positive real non integer and $\mid y \mid \leq 1$, from Gradshteyn et.al ([12] p.25) Equation (1.110) we get binomial series expansion as;*

$$
(1 - y)^{\lambda-1} = \sum_{i=0}^{\infty} (-1)^i \binom{\lambda - 1}{i} y^i
$$

Lemma 2.1.2 *If b and y are any real numbers, then from Gradshteyn et.al ([12] p.26) Equation (1.211.2),*

$$
b^y = \sum_{j=0}^{\infty} \frac{(y \ln b)^j}{j!}
$$

Using Lemma 2.1.1 and Lemma 2.1.2, expansion of cdf of the PEL distribution can be derived as,

$$
F(x) = \frac{1}{v-1} \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^j (\ln v)^i}{i!(1+q)^j} \binom{\alpha i}{j} \right.
$$
$$
\left. (e^{-qx})^j (1 + q + qx)^j - 1 \right];
$$
$$
x > 0, \alpha > 0, q > 0, v > 0, v \neq 1 \tag{7}
$$

By applying Lemma 2.1.2 in Eq. (6), we get the expansion of pdf of the PEL distribution as;

$$f(x) = \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} \sum_{i=0}^{\infty} \frac{(\ln v)^i}{i!}$$
$$\left(1 - \frac{e^{-qx}(1+q+qx)}{1+q}\right)^{\alpha i + \alpha - 1} (1+x)(e^{-qx})$$

By binomial series expansion (Lemma 2.1.1), we derive the mixture representation of pdf of the PEL distribution as,

$$f(x) = \left\{ \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} \sum_{i,j=0}^{\infty} \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} \alpha i + \binom{\alpha-1}{j} \right. $$
$$(1+q+qx)^j$$
$$\left. \times (1+x)(e^{-qx})^{j+1} \right\}; \tag{8}$$
$$x > 0, \alpha > 0, q > 0, v > 0, v \neq 1$$

Figures. 1 and 2 show the graphical representation of cdf and pdf of the PEL distribution for different combination of parameter values of α, v and q. The pdf plots show that as the values of all parameter increase, the peak of graph decreases.

$$Z(x) = \frac{\alpha v^{\{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} \ln(v)\{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha-1} q^2(1+x)e^{-qx}}{\left(v - v^{\{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha}\right)(q+1)} \tag{11}$$

Also all plots of pdf are positively skewed. In cdf plots, when value of parameter q increases, graph reaches its maximum value for small value of x. In case of increase in values of other two parameters (v and α), graph reaches maximum value for higher value of x.

3 Statistical properties of the PEL distribution

Various statistical properties of the PEL distribution such as hazard function, reverse hazard function, survival function, median, mode, moments, MGF, inverted moments, incomplete rth moments, MRL, stochastic ordering, quantile function, order statistics etc, are discussed in this section.

3.1 Survival function, hazard function and, reverse hazard function

The survival function is a function that calculates the probability of a patient, device, or other object of interest surviving after a certain period. It is given by,

$$S(x) = 1 - F(x)$$

The survival function of the PEL distribution can be written as,

$$S(x) = \frac{v - v^{\{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha}}{v-1};$$
$$x > 0, \alpha > 0, q > 0, v > 0, v \neq 1 \tag{9}$$

Survival function with expansion of cdf of the PEL distribution is,

$$S(x) = \frac{1}{(v-1)}\left[v - \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^j (\ln v)^i}{i!(1+q)^j} \binom{\alpha i}{j}(e^{-qx})^j \right.$$
$$\left. (1+q+qx)^j \right]; \tag{10}$$
$$x > 0, \alpha > 0, q > 0, v > 0, v \neq 1$$

The hazard function of a distribution is given by $Z(x) = \frac{f(x)}{1-F(x)} = \frac{f(x)}{S(x)}$, is a useful measure for describing life phenomena. It calculates the conditional likelihood of a failure based on the current state of the system. For the PEL distribution hazard function can be written as,

where $x > 0, \alpha > 0, q > 0, v > 0, v \neq 1$.
Using expansion of pdf and cdf of the PEL distribution, hazard function is:

$$Z(x) = \frac{\alpha q^2 \ln(v)}{(q+1)} \sum_{i,j=0}^{\infty}$$
$$\frac{\frac{(\ln v)^i (-1)^j}{i!(1+q)^j}\binom{\alpha i + \alpha - 1}{j}(1+q+qx)^j(1+x)(e^{-qx})^{j+1}}{\left(v - \frac{(-1)^j(\ln v)^i}{i!(1+q)^j}\binom{\alpha i}{j}(e^{-qx})^j(1+q+qx)^j\right)} \tag{12}$$

where $x > 0, \alpha > 0, q > 0, v > 0, v \neq 1$ And reverse hazard rate function for the PEL distribution can be written as, $r(x) = \frac{f(x)}{F(x)}$
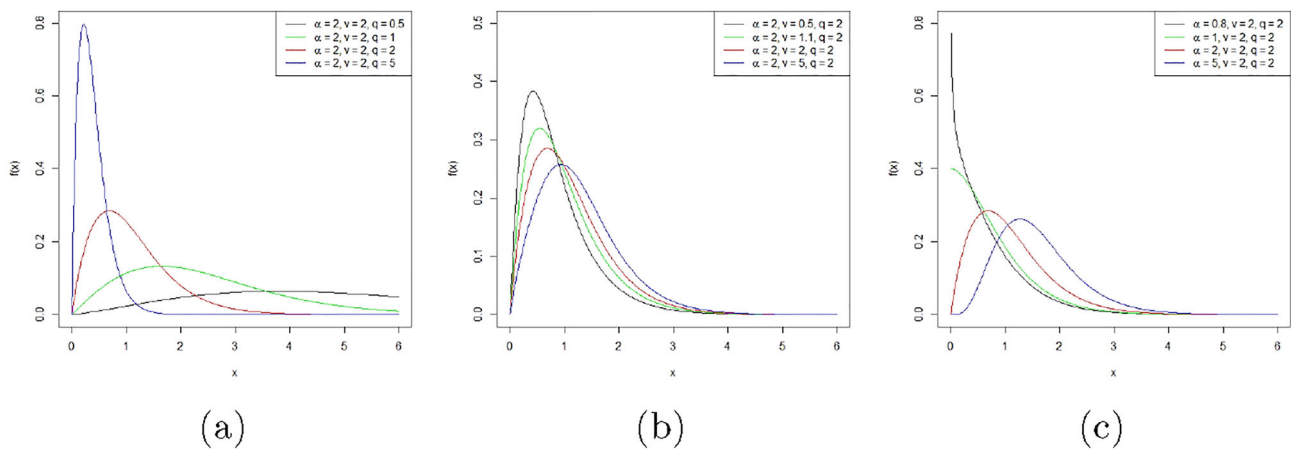
Fig. 1 Plots for pdf of the PEL distribution for distinct parameter values
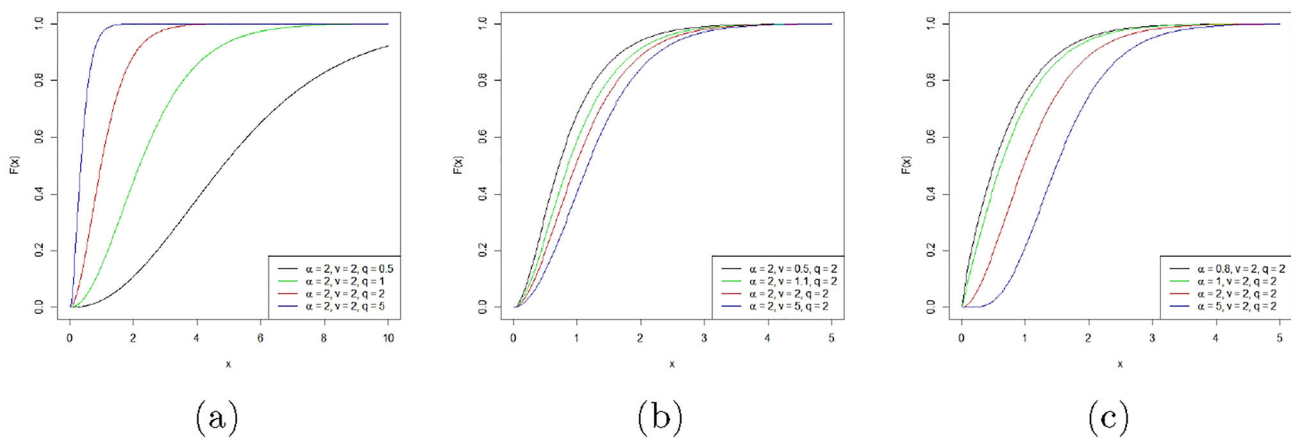

Fig. 2 Plots for cdf of the PEL distribution for distinct parameter values

$$r(x) = \frac{\alpha v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} ln(v)\{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha-1} \frac{q^2}{q+1}(1+x)e^{-qx}}{v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} - 1}$$

(13)

where $x > 0, \alpha > 0, q > 0, v > 0, v \neq 1$.
When expansion of pdf and cdf is taken into consideration,
reverse hazard function is,

$$r(x) = \frac{\alpha q^2 \ln(v)}{(q+1)} \sum_{i,j=0}^{\infty} \frac{\frac{(\ln v)^i (-1)^j}{i!(1+q)^j}\binom{\alpha i+\alpha-1}{j}(1+q+qx)^j(1+x)(e^{-qx})^{j+1}}{\left(\frac{(-1)^j(\ln v)^i}{i!(1+q)^j}\binom{i}{j}(e^{-qx})^j(1+q+qx)^j - 1\right)}$$

(14)

where $x > 0, \alpha > 0, q > 0, v > 0, v \neq 1$.

From the plot of hazard function (Fig. 4), it is evident that for the PEL distribution, as value of x increases hazard rate increases and after certain values for x, hazard rate becomes constant. Also if α is less than 1, hazard rate decreases
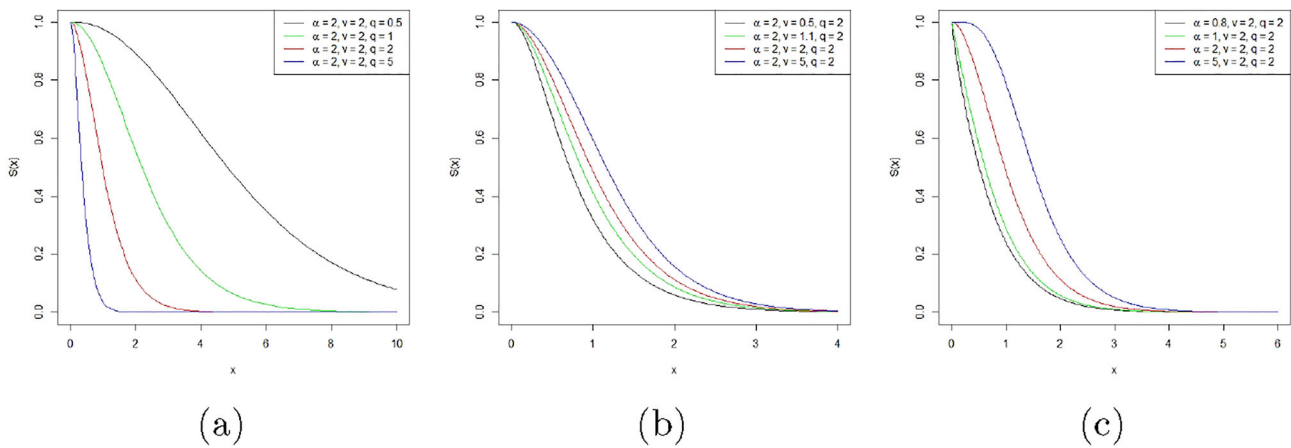
Fig. 3 Plots for survival function of the PEL distribution for different parameter values
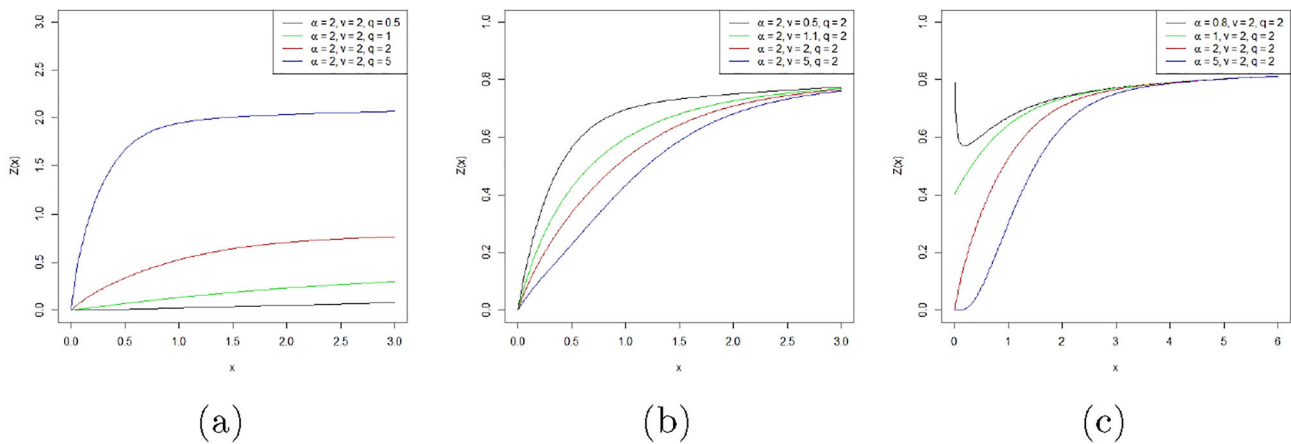

Fig. 4 Plots for hazard function of the PEL distribution for distinct parameter values

initially and then increases. The survival and reverse hazard function (Figs. 3 and 5) value decreases as value of x increases.

3.2 Median and mode

The median for a distribution is obtained by

$$\int_0^m f(x)dx = \frac{1}{2}$$

Therefore in case of the PLE distribution, the median can be obtained as,

$$\int_0^m \frac{\alpha v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} \ln(v)\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha-1}\frac{q^2}{q+1}(1+x)e^{-qx}}{v-1}$$

$$dx = \frac{1}{2}$$

$$e^{-qm}(1+q+qm) = \left[1-\left(\frac{\ln\left(\frac{(v+1)}{2}\right)}{\ln(v)}\right)^{\frac{1}{\alpha}}\right](1+q);$$

$$x > 0, \alpha > 0, q > 0, v > 0, v \neq 1 \tag{15}$$

Thus, the median of the PEL distribution can be obtained by solving Eq. (15).

As we have the pdf of the PEL distribution with random variable X, then corresponding mode is obtained by equating $f'(x) = 0$. That is,

$$\frac{q^2\alpha\ln(v)}{(q+1)(v-1)}\frac{d}{dx}v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha}$$

$$\left\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\right\}^{\alpha-1}$$

$$(1+x)e^{-qx} = 0 \tag{16}$$

$$\left\{\left(\left[1-\frac{e^{-qx}(1+q+qx)}{1+q}\right]^\alpha \alpha\ln(v)+(\alpha-1)\right)-x+\frac{1}{q}-1\right.$$

$$\left.\times\left(\frac{q}{q+1}e^{-qx}\left[1-\frac{e^{-qx}(1+q+qx)}{1+q}\right]^{-1}(1+x)^2\right)\right\}=0$$

where $x > 0, \alpha > 0, q > 0, v > 0, v \neq 1$. Thus, the mode of the PEL distribution can be obtained by solving Eq. (16).
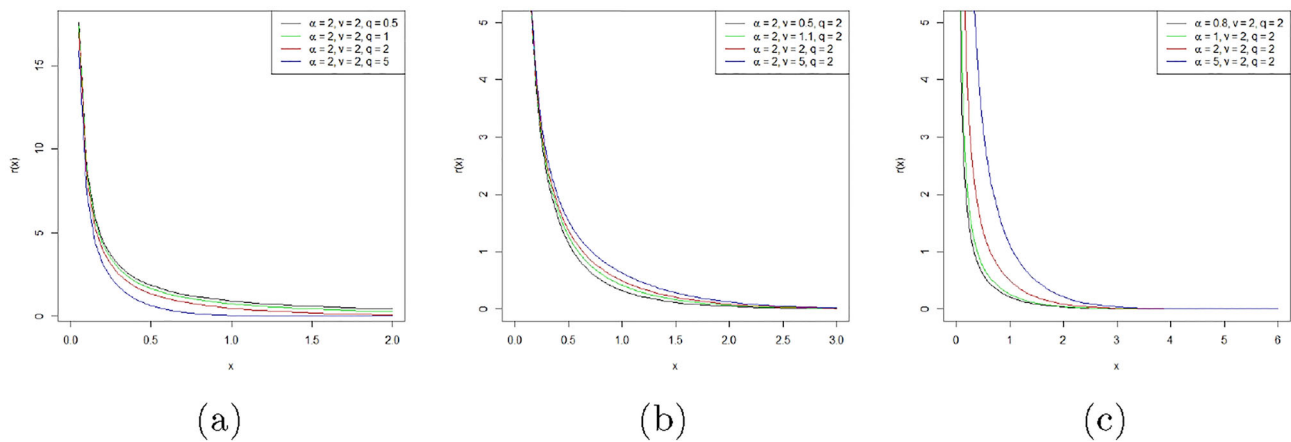
Fig. 5 Plots for the reverse hazard rate function of the PEL distribution for distinct parameter values

3.3 Moments

Moments are popularly used to describe the characteristic of a distribution. The rth moment of the random variable X can be written as,

$$\mu'_r = E[X^r] = \int_0^\infty x^r f(x) dx$$

The rth moment of the PEL distribution can be derived as,

$$\mu'_r = \frac{q^2 \ln(v)\alpha}{(q+1)(v-1)} \sum_{i,j=0}^{\infty} \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} \binom{\alpha i + \alpha - 1}{j}$$
$$\int_0^\infty x^r (1+q+qx)^j$$
$$\times (1+x)(e^{-qx})^{j+1} dx$$

Above equation is not having an anti-derivative, let

$$E_r = \sum_{i,j=0}^{\infty} \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} \binom{\alpha i + \alpha - 1}{j}$$
$$\int_0^\infty x^r (1+q+qx)^j (1+x)(e^{-qx})^{j+1} dx \qquad (17)$$
$$= E_r(x; q, v, \alpha);$$
$$x > 0, \alpha > 0, r > 0, q > 0, v > 0, v \neq 1$$

Then rth moment of the PEL distribution can be written as,

$$\mu'_r = \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} E_r;$$
$$x > 0, \alpha > 0, r > 0, q > 0, v > 0, v \neq 1 \qquad (18)$$

The different moments is found by substituting $r=1,2,3..,r$. So first 4 raw moments are as follows;

$$\mu'_1 = \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} E_1$$

$$\mu'_2 = \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} E_2$$

$$\mu'_3 = \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} E_3$$

$$\mu'_4 = \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} E_4$$

where $E_1 = \sum_{i,j=0}^{\infty} \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} \binom{\alpha i + \alpha - 1}{j} \int_0^\infty x(1+q+qx)^j(1+x)(e^{-qx})^{j+1} dx$, $E_2 = \sum_{i,j=0}^{\infty} \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} \binom{\alpha i + \alpha - 1}{j} \int_0^\infty x^2(1+q+qx)^j(1+x)(e^{-qx})^{j+1} dx$,
$E_3 = \sum_{i,j=0}^{\infty} \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} \binom{\alpha i + \alpha - 1}{j} \int_0^\infty x^3(1+q+qx)^j(1+x)(e^{-qx})^{j+1} dx$,
$E_4 = \sum_{i,j=0}^{\infty} \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} \binom{\alpha i + \alpha - 1}{j} \int_0^\infty x^4(1+q+qx)^j(1+x)(e^{-qx})^{j+1} dx$

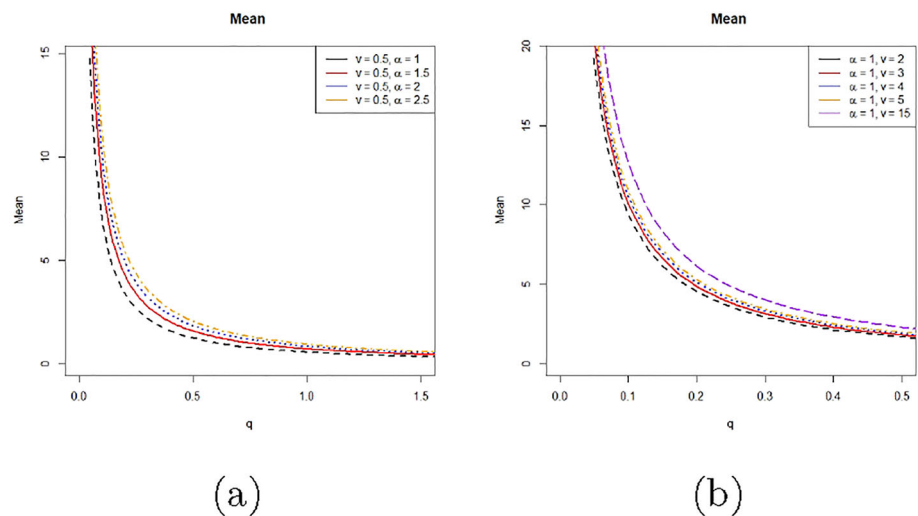#### 3.3.1 Mean

For the PEL distribution mean is obtained as,

$$\mu_1 = \mu'_1 = \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} E_1;$$
$$x > 0, \alpha > 0, r > 0, q > 0, v > 0, v \neq 1 \qquad (19)$$

Figure 6 shows the plot of mean of the PEL distribution and mean decreases as values of parameter increase.

3.3.2 Variance

Variance, $\mu_2 = \mu'_2 - (\mu'_1)^2$

Fig. 6 Plots of mean for different parameter values

(a)

(b)

The variance for the PEL distribution it is obtained as,

$$\mu_2 = \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} \left[E_2 - \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} (E_1)^2 \right];$$
$$x > 0, \alpha > 0, r > 0, q > 0,$$
$$v > 0, v \neq 1 \tag{20}$$

Figure 7 shows that as values of parameters increase, variance decreases.

3.3.3 Skewness

The skewness is a metric for the asymmetry of the probability distribution of a real-valued random variable with respect to its mean. It is defined as:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

where,

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$

So, the skewness of the PEL distribution is obtained as;

$$\beta_1 = \frac{E_3}{E_2} \left[\left(\frac{(q+1)(v-1)}{\alpha q^2 \ln(v)} \right)^2 + 2(E_3)^2 \right]$$
$$- \frac{3(q+1)(v-1)E_1}{\alpha q^2 \ln(v)}; \tag{21}$$
$$x > 0, \alpha > 0, r > 0, q > 0, v > 0, v \neq 1$$

From Fig. 8, it is clear that as values of parameters increase, skewness decreases. Also all values are positive, so its positively skewed.

3.3.4 Kurtosis

The kurtosis is a statistical term used to characterize how much data clusters in the tails or the peak of a frequency distribution. It is defined as,

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

where,

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 + 3(\mu_1)^4$$

For the PEL distribution, kurtosis can be obtained as,

$$\beta_2 = \left\{ \frac{6q^2 \ln(v)\alpha(E_1)^2}{(q+1)(v-1)E_2} + \frac{1}{(E_2)^2} \left(\frac{(q+1)(v-1)}{q^2 \ln(v)\alpha E_4} \right. \right.$$
$$+ \frac{3(E_1)^4 q^4 (\ln(v))^2 \alpha^2}{(q+1)^2 (v-1)^2}$$
$$\left. \left. - 4E_3 E_1 \right) \right\}; \tag{22}$$
$$x > 0, \alpha > 0, r > 0, q > 0, v > 0, v \neq 1$$

3.4 Moment generating function

The moment generating function (MGF) creates a single function from which all the moments of a random variable can be recovered at a later time. A probability distribution is uniquely determined by its MGF. The MGF of a random variable X is given by,

$$M_X(t) = \int_0^\infty e^{tx} f(x) dx$$

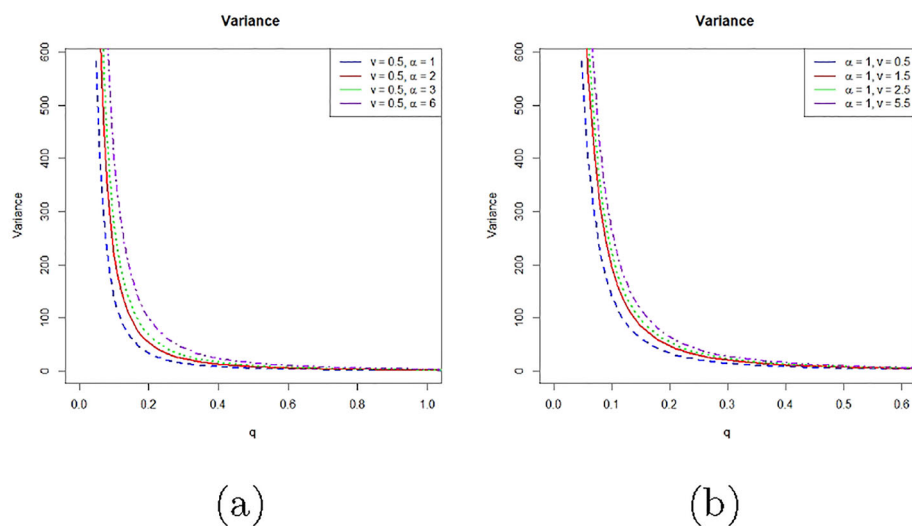**Fig. 7** Plots for variance for different parameter values

(a)　　　　　　　　(b)

Fig. 8 Plots of skewness for different parameter values
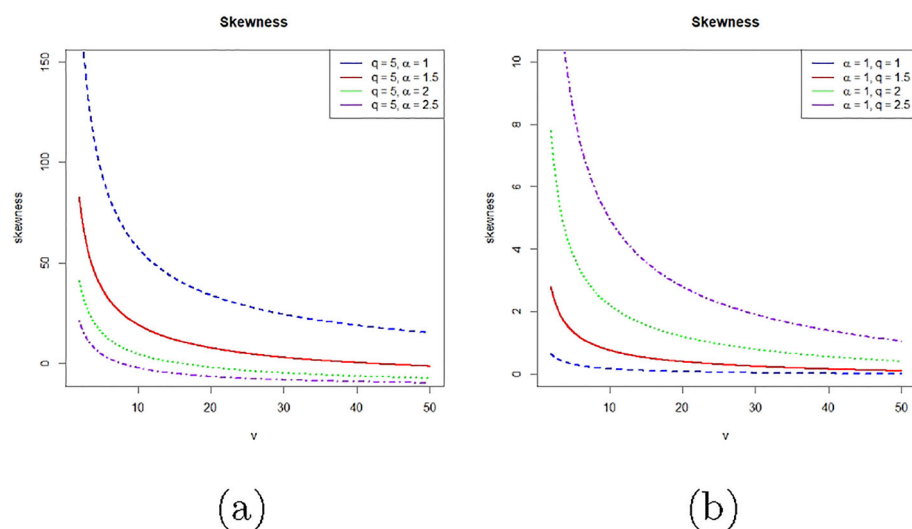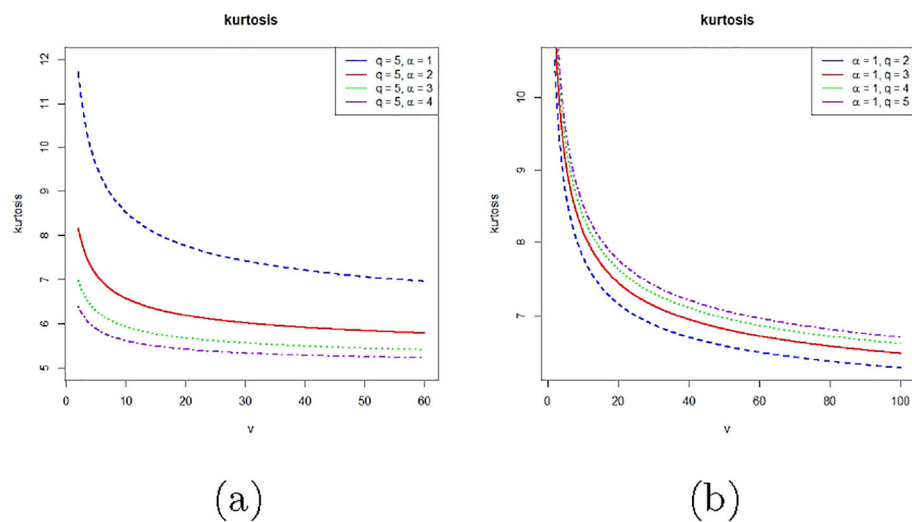

(a)　　　　　　　　(b)

Fig. 9 Plots of kurtosis for different parameter values

(a)　　　　　　　　(b)

For the PEL distribution, MGF is obtained as;

$$
M_X(t) = \left\{ \int_0^\infty e^{tx} \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} \sum_{i=0}^\infty \sum_{j=0}^\infty \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} (e^{-qx})^{j+1} \right.
$$

$$
\left. \times (1+q+qx)^j (1+x) dx \right\}
$$

$$
M_X(t) = \left\{ \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} \int_0^\infty \sum_{i,j=0}^\infty \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} \right. \tag{23}
$$

$$
(e^{-qx})^{j+1} (1+q+qx)^j
$$

$$
\left. \times e^{tx} (1+x) dx \right\};
$$

$$
x > 0, \alpha > 0, r > 0, q > 0, v > 0, v \neq 1
$$

3.5 Inverted moments

The rth inverted moments of a distribution is derived by

$$
\mu_r^* = \int_0^\infty x^{-r} f(x) dx
$$

For the PEL distribution, using Eq. (17) inverted moment is obtained as,

$$
\mu_r^* = \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} E_{-r}(x; q, v, \alpha); \tag{24}
$$
$$
x > 0, \alpha > 0, r > 0, q > 0, v > 0, v \neq 1
$$

3.6 Incomplete moment

The incomplete rth moment is defined by

$$
m_r(x) = \int_0^x x^r f(x) dx
$$

For the PEL distribution, incomplete rth moment is obtained by;

$$
m_r(x) = \left\{ \frac{\alpha q^2 \ln(v)}{(q+1)(v-1)} \int_0^x \sum_{i,j=0}^\infty \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} x^r (e^{-qx})^{j+1} \right.
$$

$$
\left. (1+q+qx)^j \times (1+x) dx \right\}; \tag{25}
$$

$$
x > 0, \alpha > 0, r > 0, q > 0, v > 0, v \neq 1
$$

3.7 Mean residual life function

In reliability and survival analysis, the mean residual life (MRL) function is important. It specifies how long a system will function, starting at a point of time x. The MRL function of a life time random variable X' is given as,

$$
\phi(x) = \frac{1}{s(x)} \int_x^\infty x f(x) dx - x
$$

where, $s(x)$ is the survival rate function and $f(x)$ is pdf of the distribution.

For the PEL distribution, the MRL function is obtained as,

$$
\phi(x) = \frac{\alpha q^2 \ln(v) \sum_{i,j=0}^\infty \frac{(\ln v)^i (-1)^j}{i!(1+q)^j} \binom{\alpha i + \alpha - 1}{j} \int_x^\infty (1+q+qx)^j (1+x)(e^{-qx})^{j+1} dx - x}{(q+1)\left(v - \sum_{i,j=0}^\infty \frac{(-1)^j (\ln v)^i}{i!(1+q)^j} \binom{\alpha i}{j}(e^{-qx})^j (1+q+qx)^j \right)}; x, q, v, \alpha > 0, v \neq 1 \tag{26}
$$

3.8 Order statistics

David [9] given the pdf of the kth order statistics $X_{k:n}$ where $k = 1, 2, ..., n$ as;

$$
f_{k:n}(x) = C_{k:n}[F(x)]^{k-1}[1 - F(x)]^{n-k} f(x); x > 0
$$

where, $C_{k:n} = \frac{n!}{(k-1)!(n-k)!}$

Theorem 3.8.1 *Let $X_1, X_2, ..., X_n$ be a simple random sample from a PEL distribution, with cdf and pdf, respectively, provided by (5) and (6). The order statistics acquired from this sample be, $X_{(1:n)} \leq X_{(2:n)} \leq ... \leq X_{(n:n)}$. Then k^{th} order statistic of the PEL distribution is given by*

$$
f_{k:n}(x) = \left\{ \frac{C_{k:n} \alpha q^2 \ln(v)}{(q+1)} \sum_{i,j=0}^\infty \binom{n-k}{j} \right.
$$

$$
\binom{k+i-1}{j} \frac{(-1)^{k+j+2i-1}}{(v-1)^{(k+i)}} \tag{27}
$$

$$
\times v^{\{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}\alpha(j+1)}
$$

$$
\left. \{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha-1} (1+x) e^{-qx} \right\}
$$

where, $C_{k:n} = \frac{n!}{(k-1)!(n-r)!}$ and $x, q, v, \alpha > 0, v \neq 1$

Proof When $F(x)$ and $f(x)$ are from Eq. (5) and (6) respectively, and by applying Lemma 2.1.1, we get

$$
\begin{aligned}
f_{k:n}(x) &= C_{k:n} \sum_{i=0}^{\infty} (-1)^i \binom{n-k}{j} [F(x)]^{k+i-1} f(x) \\
&= \left\{ \frac{C_{k:n}\alpha q^2 \ln(v)}{(q+1)(v-1)} \sum_{i=0}^{\infty} (-1)^i \binom{n-k}{j} \right. \\
&\quad \left[\frac{v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} - 1}{v-1} \right]^{k+i-1} \\
&\quad \times v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} \{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha-1} \\
&\quad \left. (1+x)e^{-qx} \right\}
\end{aligned}
$$

Hence kth order statistic of the PEL distribution is obtained as,

$$
\begin{aligned}
f_{k:n}(x) &= \left\{ \frac{C_{k:n}\alpha q^2 \ln(v)}{(q+1)} \sum_{i,j=0}^{\infty} (-1)^{k+j+2i-1} \binom{n-k}{j} \right. \\
&\quad \binom{k+i-1}{j} \frac{1}{(v-1)^{(k+i)}} \\
&\quad \times v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha(j+1)}} \{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha-1} \\
&\quad \left. (1+x)e^{-qx} \right\}
\end{aligned}
\tag{28}
$$

where, $x, q, v, \alpha > 0, v \neq 1$.
Corresponding $1st$ and nth order statistics of the PEL distribution are:

$$
\begin{aligned}
f_{1:n}(x) &= \left\{ \frac{n\alpha q^2 \ln(v)}{(q+1)} \sum_{i,j=0}^{\infty} (-1)^{j+2i} \binom{n-1}{i} \binom{i}{j} \frac{(1+x)e^{-qx}}{(v-1)^{(1+i)}} \right. \\
&\quad \left. \times v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha(j+1)}} \{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha-1} \right\}
\end{aligned}
\tag{29}
$$

$$
\begin{aligned}
f_{n:n}(x) &= \left\{ \frac{n\alpha q^2 \ln(v)}{(q+1)} \sum_{i,j=0}^{\infty} (-1)^{n+j+2i-1} \binom{0}{i} \binom{n+i-1}{j} \right. \\
&\quad \frac{(1+x)e^{-qx}}{(v-1)^{(n+i)}} \\
&\quad \left. \times v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha(j+1)}} \{1 - \frac{e^{-qx}(1+q+qx)}{1+q}\}^{\alpha-1} \right\}
\end{aligned}
\tag{30}
$$

$\square$

3.9 Quantile function

Theorem 3.9.1 *Let a random sample X be from the PEL distribution, then the quantile function of X is given by*

$$
\begin{aligned}
Q_X(u) &= -1 - \frac{1}{q} - \frac{1}{q} W_{-1} \left[-e^{-(1+q)} \left(1 - \right. \right. \\
&\quad \left. \left. \left[\frac{\ln(u(v-1)+1)}{\ln v} \right]^{\frac{1}{\alpha}} \right) \times (1+q) \right]; \\
&\quad x, q, v, \alpha > 0, v \neq 1
\end{aligned}
\tag{31}
$$

Proof To find the quantile function for the PEL distribution, we equate $F_X(x) = u$ where $u \in (0, 1)$.

$$
\frac{v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} - 1}{v-1} = u
$$

$$
v^{\{1-\frac{e^{-qx}(1+q+qx)}{1+q}\}^\alpha} = u(v-1) + 1
$$

Taking logarithm on both sides

$$
\left[1 - \frac{e^{-qx}(1+q+qx)}{1+q} \right]^\alpha \ln v = \ln(u(v-1)+1)
$$

$$
e^{-qx}(1+q+qx) = \left(1 - \left[\frac{\ln(u(v-1)+1)}{\ln v} \right]^{\frac{1}{\alpha}} \right)(1+q)
$$

Multiplying, $-e^{-(1+q)}$

$$
-(1+q+qx)e^{-(1+q+qx)} = -e^{-(1+q)} \left(1 - \left[\frac{\ln(u(v-1)+1)}{\ln v} \right]^{\frac{1}{\alpha}} \right)(1+q)
$$

We can see from the equation above that the Lambert W function of the real argument $-e^{-(1+q)} \left(1 - \left[\frac{\ln(u(v-1)+1)}{\ln v} \right]^{\frac{1}{\alpha}} \right)$ $(1+q)$ is $-(1+q+qx)$. Then we have;

$$
\begin{aligned}
W[&-e^{-(1+q)} \left(1 - \left[\frac{\ln(u(v-1)+1)}{\ln v} \right]^{\frac{1}{\alpha}} \right)(1+q)] \\
&= -(1+q+qx)
\end{aligned}
\tag{32}
$$

Moreover, it is clear that for each $q > 0$ and $x > 0$, $(1+q+qx)$ is greater than 0 and it may be verified that $-(1+q+qx)e^{-(1+q+qx)} = -e^{-(1+q)} \left(1 - \left[\frac{\ln(u(v-1)+1)}{\ln v} \right]^{\frac{1}{\alpha}} \right)(1+q) \in (\frac{-1}{e}, 0)$, since $u \in (0, 1)$. In account of the characteristics of negative branch of the Lambert W function, Eq. (32) becomes

$$
\begin{aligned}
W_{-1}[&-e^{-(1+q)} \left(1 - \left[\frac{\ln(u(v-1)+1)}{\ln v} \right]^{\frac{1}{\alpha}} \right)(1+q)] \\
&= -(1+q+qx)
\end{aligned}
\tag{33}
$$

Therefore, the quantile function of the PEL distribution can be written as,

$$Q_X(u) = -1 - \frac{1}{q} - \frac{1}{q} W_{-1}[-e^{-(1+q)}\left(1 \right.$$

$$\left. - \left[\frac{\ln(u(v-1)+1)}{\ln v}\right]^{\frac{1}{\alpha}}\right)(1+q)]$$

□

3.10 Stochastic ordering

In many practical problems, it becomes necessary to compare two life distributions with reference to some of their characteristics and there stochastic ordering plays an important role. A random variable Y is greater than X in the following way:

(i) Stochastic order $Y \geq_{st} X$ if $F_Y(x) \leq F_X(x)$
(ii) Hazard rate order $Y \geq_{hr} X$ if $h_Y(x) \leq h_X(x)$
(iii) Mean residual life order $Y \geq_{mrl} X$ if $M_Y(x) \geq M_X(x)$
(iv) Likelihood ratio order $Y \geq_{lr} X$ if $\frac{f_X(x)}{f_Y(x)}$ decreasing in x

The following are the relationship between the above-mentioned properties of a distribution.

(a) $Y \geq_{lr} X \implies Y \geq_{hr} X \implies Y \geq_{mrl} X$ and
(b) $Y \geq_{lr} X \implies Y \geq_{st} X$

Theorem 3.10.1 *Let* $X \sim PEL(v_1, q_1, \alpha_1)$ *and* $Y \sim PEL(v_2, q_2, \alpha_2)$ *if* $v_1 \leq v_2, q_1 \leq q_2$ *and* $\alpha_1 \leq \alpha_2$, *we have* $Y \geq_{lr} X$ *then* $Y \geq_{hr} X, Y \geq_{mrl} X$ *and* $Y \geq_{st} X$.

Proof To prove $\frac{f_X(x)}{f_Y(x)}$ decreasing in x we have to show that the derivative of $\frac{f_X(x)}{f_Y(x)}$ is less than 0.

$$\frac{f_X(x)}{f_Y(x)} = \left\{\left(\frac{v_1^{\{1-\frac{e^{-q_1 x}(1+q_1+q_1 x)}{1+q_1}\}^{\alpha_1}}}{v_2^{\{1-\frac{e^{-q_2 x}(1+q_2+q_2 x)}{1+q_2}\}^{\alpha_2}}}\frac{\{1-\frac{e^{-q_1 x}(1+q_1+q_1 x)}{1+q_1}\}^{\alpha_1-1}}{\{1-\frac{e^{-q_2 x}(1+q_2+q_2 x)}{1+q_2}\}^{\alpha_2-1}}\right)\right.$$

$$\left. \times\left(\frac{\alpha_1 ln(v_1)q_1^2(v_2-1)(q_2+1)}{\alpha_2 ln(v_2)q_2^2(v_1-1)(q_1+1)}\right)e^{-(q_1-q_2)x}\right\}$$

Taking ln (log to base e) on both sides:

$$\ln\frac{f_X(x)}{f_Y(x)} = \left\{\ln\left(\frac{\alpha_1 ln(v_1)q_1^2(v_2-1)(q_2+1)}{\alpha_2 ln(v_2)q_2^2(v_1-1)(q_1+1)}\right)\right.$$

$$+\ln\left(\frac{v_1^{\{1-\frac{e^{-q_1 x}(1+q_1+q_1 x)}{1+q_1}\}^{\alpha_1}}}{v_2^{\{1-\frac{e^{-q_2 x}(1+q_2+q_2 x)}{1+q_2}\}^{\alpha_2}}}\right)$$

$$\left. \times\ln\left(\frac{\{1-\frac{e^{-q_1 x}(1+q_1+q_1 x)}{1+q_1}\}^{\alpha_1-1}}{\{1-\frac{e^{-q_2 x}(1+q_2+q_2 x)}{1+q_2}\}^{\alpha_2-1}}\right) - (q_1-q_2)x\right\}$$
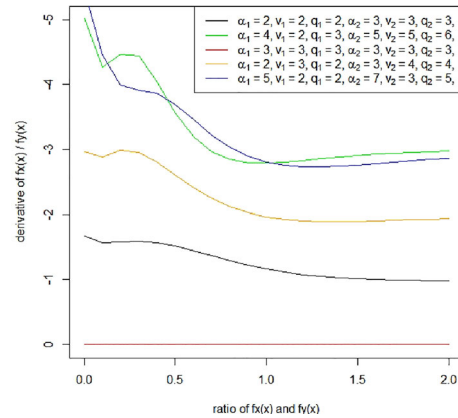

Fig. 10 Plot of the derivative of ratio of pdfs

Differentiating both sides with respect to x, we get

$$\frac{d}{dx}ln\left(\frac{f_X(x)}{f_Y(x)}\right) = \left\{q_2 - q_1 - \alpha_2 \ln(v_2)(1-\right.$$

$$\frac{(q_2 x + q_2 + 1)e^{-q_2 x}}{q_2+1})^{\alpha_2-1}(\frac{e^{-q_2 x}}{q_2+1}$$

$$\times\frac{q_2(q_2 x + q_2 + 1)}{q_2+1} - \frac{q_2 e^{-q_2 x}}{q_2+1}) + \ln(\alpha_2-1)(\frac{q_2 e^{-q_2 x}}{q_2+1}$$

$$-\frac{q_2(q_2 x + q_2 + 1)e^{-q_2 x}}{q_2+1})$$

$$+ \alpha_1 \ln(v_1)(1 - \frac{(q_1 x + q_1 + 1)e^{-q_1 x}}{q_1+1})^{\alpha_1-1} \qquad (34)$$

$$(\frac{q_1(q_x + q_1 + 1)e^{-q_1 x}}{q_1+1}$$

$$-\frac{q_1 e^{-q_1 x}}{q_1+1}) + \ln(\alpha_1-1)(\frac{q_1(q_1 x + q_1 + 1)e^{-q_1 x}}{q_1+1}$$

$$\left. - \frac{q_1 e^{-q_1 x}}{q_1+1})\right\};$$

$$x, q, v, \alpha > 0, v \neq 1$$

Here, we are using graph to show that all values of derivative of $f_X(x)/f_Y(x)$ (Eq. (34)) are less than 0, when $v_1 \leq v_2$, $q_1 \leq q_2$ and $\alpha_1 \leq \alpha_2$.

From Fig. 10, it is clear that all derivative values of likelihood function are negative for all x. So we can conclude that $f_X(x)/f_Y(x)$ is decreasing in x. Hence, we proved $Y \geq_{lr} X$ so we can say that $Y \geq_{hr} X, Y \geq_{mrl} X$ and $Y \geq_{st} X$ when Y and X follows the PEL distribution. □

4 Maximum likelihood estimation method

To find the maximum likelihood estimates (MLE) for parameters α, v and q of the PEL distribution, we follow the following steps:

The likelihood function of the PEL distribution is obtained as,

$$\prod_{i=1}^{n} f(x_i : \alpha, v, q)$$

$$= \left\{ \prod_{i=1}^{n} \left[\frac{\alpha v^{\{1 - \frac{e^{-qx_i}(1+q+qx_i)}{1+q}\}^{\alpha}} ln(v) \{1 - \frac{e^{-qx_i}(1+q+qx_i)}{1+q}\}^{\alpha-1}}{(v-1)(q+1)} \right. \right.$$

$$\left. \left. \times q^2(1+x_i)e^{-qx_i} \right] \right\}$$

The log-likelihood function is,

$$L = \ln\left[\prod_{i=1}^{n} f(x_i : \alpha, v, q) \right] = \left\{ n \ln \alpha + n \ln(\ln v) \right.$$

$$- n \ln(v-1) - n \ln(q+1)$$

$$+ n \ln q^2 - \sum_{i=1}^{n} q x_i + \sum_{i=1}^{n} \ln(1+x_i) + \ln(v) \sum_{i=1}^{n}\{1$$

$$- \frac{e^{-qx_i}(1+q+qx_i)}{1+q}\}^{\alpha}$$

$$+ (\alpha - 1) \sum_{i=1}^{n} \ln\{1 - \frac{e^{-qx_i}(1+q+qx_i)}{1+q}\} \right\}$$

The derivative of the log-likelihood function with respect to v can be written as:

$$\frac{\partial L}{\partial v} = \frac{n}{v \ln(v)} - \frac{n}{v-1}$$

$$+ \frac{1}{v} \sum_{i=1}^{n} \left[\left(1 - \frac{e^{-qx_i}(1+q+qx_i)}{1+q}\right)^{\alpha} \right] \quad (35)$$

The derivative of the log-likelihood function with respect to α can be written as:

$$\frac{\partial L}{\partial \alpha} = \left\{ \frac{n}{\alpha} + \sum_{i=1}^{n} \left[1 - \frac{e^{-qx_i}(1+q+qx_i)}{1+q} \right]^{\alpha} \right.$$

$$\ln\left(1 - \frac{e^{-qx_i}(1+q+qx_i)}{1+q}\right) \quad (36)$$

$$\left. \times \ln v + \sum_{i=1}^{n} \ln\left(1 - \frac{e^{-qx_i}(1+q+qx_i)}{1+q}\right) \right\}$$

The derivative of the log-likelihood function with respect to q can be written as:

$$\frac{\partial L}{\partial q} = \left\{ \frac{2n}{q} - \frac{n}{q+1} + \alpha \ln v \sum_{i=1}^{n} \left[\left(1 - \frac{e^{-qx_i}(1+q+qx_i)}{1+q}\right)^{\alpha-1} \right] \right.$$

$$- \sum_{i=1}^{n} x_i \times \left[\left(\frac{e^{-qx_i}(x_i^2 q + x_i q - 1)}{q+1} \right) \right.$$

$$+ \left(\frac{e^{-qx_i}(1+q+qx_i)}{(1+q)^2} \right) \right]$$

$$+ (\alpha - 1) \sum_{i=1}^{n} \left[\frac{\left(\frac{e^{-qx_i}(x_i^2 q + x_i q - 1)}{q+1} \right) + \left(\frac{e^{-qx_i}(1+q+qx_i)}{(1+q)^2} \right)}{1 - \frac{e^{-qx_i}(1+q+qx_i)}{1+q}} \right] \right\};$$

$$x, q, v, \alpha > 0, v \neq 1 \quad (37)$$

We can estimate the unknown parameters using the MLE method by assigning these non-linear Eqs. (36)–(37) to zero and solving them simultaneously. But due to the complexity of these equations, numerical methods like the Newton–Raphson method is used for solving the equations with the help of R software.

5 Simulation study

This section presents a Monte Carlo simulation study to assess the performance of the parameters of the PEL distribution when the MLE method is applied. Consistency of the parameters is measured in terms of MSE, RMSE, and bias. With sample sizes of $n = 10, 40, 70$, and 100 with 7 distinct parameter combinations, the simulation study was repeated 1000 times. The steps to carry out a simulation study are as follows:

- *Step* 1: Random number generation of the PEL distribution
  - (i) A random variable is generated from the uniform distribution $U(0, 1)$.
  - (ii) The created random variable is then fed into the quantile function to produce random numbers of the PEL distribution.
  - (iii) steps (i) and (ii) are repeated to obtain observations of the desired sample sizes n.

- *Step* 2: For each sample, the MLEs are computed.
- *Step* 3: Repeat steps 1 and 2, 1000 times.
- *Step* 4: Then, the MLEs and corresponding MSEs, RMSEs and bias are calculated using the equations,

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\Omega_i - \hat{\Omega})^2$$

$$RMSE = \sqrt{MSE}$$

$$Bias = \frac{1}{n} \sum_{i=1}^{n} (\Omega_i - \hat{\Omega})$$

where Ω_i = observed values of the parameter, $\hat{\Omega}$ = predicted MLEs of parameter, n = number of data points

Table 1 Result of simulation study with different parameter values

($q = 2$, $v = 6$, $\alpha = 7$)

Sample size	Parameter	MLE values	MSE	RMSE	Bias
10	q	2.229891	0.05284965	0.229891	−0.229891
	v	6.44686	0.1996834	0.4468595	−0.4468595
	α	8.257543	1.581416	1.257543	−1.257543
40	q	2.078754	0.006202	0.078754	−0.078754
	v	6.438407	0.192200	0.438407	-.438407
	α	8.115525	1.244396	1.115525	−1.115525
70	q	2.096669	0.009344812	0.09666857	−0.09666857
	v	6.35277	0.124446	0.35277	−0.35277
	α	7.969238	0.939422	0.969238	−0.969238
100	q	2.084044	0.007063391	0.08404398	−0.08404398
	v	6.327033	0.1069504	0.3270328	−0.3270328
	α	7.5781	0.3341996	0.5781	−0.5781

($q = 3$, $v = 5$, $\alpha = 5$)

Sample size	Parameter	MLE values	MSE	RMSE	Bias
10	q	3.301988	0.09119683	0.3019881	−0.3019881
	v	4.380859	0.3833354	0.6191409	0.6191409
	α	6.326143	1.758655	1.326143	−1.326143
40	q	3.252458	0.063735	0.252458	−0.252458
	v	4.56723	0.187289	0.43277	0.43277
	α	6.09412	1.197098	1.09412	−1.09412
70	q	3.039756	0.001580571	0.03975639	−0.03975639
	v	4.79342	0.042304	0.20568	0.20568
	α	5.8346	0.6965566	0.8345997	−0.8345997
100	q	3.068626	0.004709587	0.06862643	−0.06862643
	v	4.922712	0.005973372	0.07728759	0.07728759
	α	5.167415	0.02802767	0.1674147	−0.1674147

($q = 2$, $v = 4$, $\alpha = 5$)

10	q	2.362413	0.1313435	0.3624134	−0.3624134
	v	5.448296	2.097562	1.448296	−1.448296
	α	6.472191	2.167348	1.472191	−1.472191
40	q	2.366537	0.1343496	0.3665373	−0.3665373
	v	5.006866	1.013779	1.006866	−1.006866
	α	6.131707	1.280761	1.131707	−1.131707
70	q	2.021464	0.0004607113	0.02146419	−0.02146419
	v	4.50431	0.25432	0.50431	−0.50431
	α	5.162059	0.026263	0.162059	−0.162059
100	q	2.01071	0.0001146987	0.01070975	−0.01070975
	v	3.900007	0.009998551	0.09999275	0.09999275
	α	5.008271	0.00006840	0.008270939	−0.008270939

Table 1 continued

($q = 3, v = 5, \alpha = 5$)					
Sample size	Parameter	MLE values	MSE	RMSE	Bias
($q = 2, v = 6, \alpha = 5$)					
10	q	2.143071	0.02046922	0.1430707	−0.1430707
	v	4.655974	1.806405	1.344026	1.344026
	α	5.881638	0.7772855	0.881638	−0.881638
40	q	2.047812	0.002286014	0.04781228	−0.04781228
	v	4.690861	1.713845	1.309139	1.309139
	α	5.499703	0.249703	0.4997029	−0.4997029
70	q	2.082811	0.006857659	0.08281098	−0.08281098
	v	5.25277	0.55835	0.74723	0.74723
	α	5.369238	0.136336	0.369238	- 0.369238
100	q	2.074615	0.005567403	0.07461503	−0.07461503
	v	5.47779	0.2727033	0.52221	0.52221
	α	5.309179	0.0955919	0.3091794	−0.3091794
($q = 2, v = 5, \alpha = 4$)					
10	q	2.460609	0.2121606	0.460609	−0.460609
	v	6.049118	1.100648	1.049118	−1.049118
	α	6.192997	4.809234	2.192997	−2.192997
40	q	2.435673	0.1898112	0.4356732	−0.4356732
	v	6.056851	1.116934	1.056851	−1.056851
	α	6.18177	4.760122	2.18177	−2.18177
70	q	2.108643	0.01180321	0.1086426	−0.1086426
	v	5.799162	0.63865	0.799162	−0.799162
	α	5.65016	2.723027	1.65016	−1.65016
100	q	2.000951	9.052476e−07	0.000951445	−0.000951445
	v	5.3867261	0.1495571	0.3867261	−0.3867261
	α	4.436273	0.1903345	0.4362734	- 0.4362734

($q = 2, v = 5, \alpha = 5$)					
Sample size	Parameter	MLE values	MSE	RMSE	Bias
10	q	2.299243	0.08954612	0.2992426	−0.2992426
	v	5.517957	0.2682797	0.5179573	−0.5179573
	α	6.257853	1.257853	1.18177	−1.257853
40	q	2.150361	0.02260835	0.1503607	−0.1503607
	v	4.50711	0.2429405	0.4928899	0.4928899
	α	5.567977	0.3225978	0.567977	−0.567977
70	q	2.087306	0.007622	0.087306	−0.087306
	v	4.51032	0.239786	0.48968	0.48968
	α	5.563583	0.317625	0.563583	−0.563583
100	q	2.055663	0.003098397	0.05566325	−0.05566325
	v	4.518508	0.2318341	0.4814916	0.4814916
	α	5.503536	0.2535486	0.5035361	−0.5035361

Table 1 continued

(q = 2, v = 5, α = 5)					
Sample size	Parameter	MLE values	MSE	RMSE	Bias
(q = 1, v = 5, α = 5)					
10	q	1.104757	0.01097399	0.1047568	−0.1047568
	v	5.454657	0.206712	0.454657	−0.454657
	α	5.674442	0.4548726	0.6744424	−0.6744424
40	q	0.9401529	0.003581671	0.05984706	0.05984706
	v	5.40569	0.16458	0.40569	−0.40569
	α	5.529449	0.280316	0.529449	−0.529449
70	q	1.036275	0.001315878	0.03627504	−0.03627504
	v	4.550159	0.2023571	0.4498411	0.4498411
	α	5.441113	0.1945803	0.4411126	−0.4411126
100	q	0.9874188	0.0001582864	0.01258119	0.01258119
	v	4.94588	0.002928959	0.05411985	0.05411985
	α	4.740304	0.06744196	0.2596959	0.2596959

Table 2 Summary statistics of case fatality ratio (in %) of COVID-19 in China

Minimum	First quartile	Median	Mean	Third quartile	Maximum
1.000	1.750	2.580	2.358	3.080	3.610

Table 3 Parameter estimation of different distributions for case fatality ratio (in %) of COVID-19 in China

Distribution	Parameter estimates
PEL	q= 1.684938
	v = 19.384176
	α= 5.615004
PEE	p = 1.592209
	v = 34.266258
	β= 9.256132
EPL	θ = 1.326197
	β = 1.430201
	α= 8.830552
EG	λ = 2.103211
	θ = 10.214999
L	θ = 0.6768592
E	λ = 0.4240158

Numerical results are given in Table 1 for the PEL distribution with parameter values (q = 2, v = 5, α = 5), (q = 1, v = 5, α = 5), (q = 3, v = 5, α = 5), (q = 2, v = 4, α = 5), (q = 2, v = 6, α = 5), (q = 2, v = 5, α = 4) and (q = 2, v = 6, α = 7).

As the sample size (n) increases, Table 1 clearly shows that the estimated parameter values tends to approach the actual parameter values, and all errors and bias are decreasing. This indicates that the precision and consistency of the MLEs are attained as the sample size increases. Thus, we can logically conclude that the MLE approach is really effective at estimating the parameters of the PEL distribution.

6 Real data analysis

The performance of the proposed PEL distribution is evaluated in this section, using three real data sets and comparing the PEL distribution with certain well-known existing dis-

Table 4 Fitting of case fatality ratio (in %) for COVID-19 in China

Distribution	Log-likelihood	D	AIC	BIC	HQIC	p-value
PEL	−31.39583	0.13715	68.79167	72.44829	69.80586	0.7348
PEE	−31.89439	0.15297	69.78878	73.44541	70.80297	0.6022
EG	−34.11301	0.30605	72.22602	74.66378	72.90215	0.01849
EPL	−38.63597	0.40619	83.27195	86.92857	84.28614	0.05227
L	−42.84234	0.28665	87.68468	88.90356	88.02274	0.03287
E	−46.44959	0.34559	94.89917	96.11805	95.23724	0.005101

Fig. 11 Graphical representation of each distribution for case fatality ratio (in %) of China
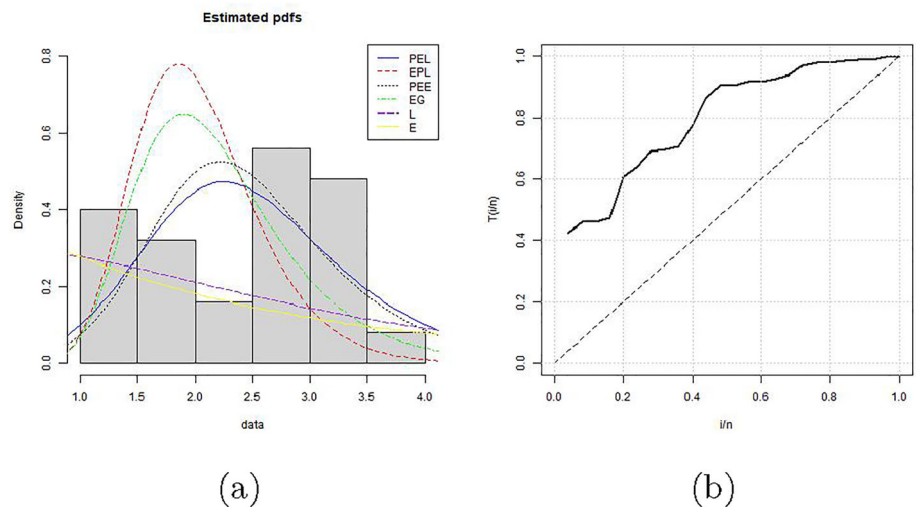

(a)　　　　　　　　　　　　　(b)

Table 5 Summary statistics of case fatality ratio (in %) of COVID-19 in India

Minimum	First quartile	Median	Mean	Third quartile	Maximum
1.014	1.717	2.328	2.723	3.162	10.480

Table 6 Parameter estimation of different distributions for case fatality ratio (in %) of COVID-19 in India

Distribution	Parameter Estimates
PEL	$q = 1.225578$
	$v = 2.532038$
	$\alpha = 3.797981$
PEE	$p = 1.147063$
	$v = 5.743843$
	$\beta = 6.642685$
EPL	$\theta = 1.670529$
	$\beta = 0.765471$
	$\alpha = 16.820413$
EG	$\lambda = 1.531871$
	$\theta = 3.938690$
L	$\theta = 0.639152$
E	$\lambda = 0.3672219$

tributions. To assess the performance of the considered distributions, we calculate the Akaike information criteria (AIC), Bayesian information criteria (BIC), Hannan–Quinn information criterion (HQIC) and p-value. Meanwhile, the best distribution is the one with the highest log-likelihood value, p-value, and lowest AIC, BIC, and the HQIC values. Distributions which are used to compare are; the power exponentiated exponential (PEE) distribution proposed by Modi [20], the exponentiated power Lindley (EPL) distribution proposed by Ashour and Eltehiwy [8], the exponentiated gamma (EG) distribution proposed by Shawky and Bakoban [28], the Lindley (L) distribution proposed by Lindley [17] and the exponential (E) distribution by Kondo [16]. The pdf of the distributions are as follows:

- PEE:

$$f(x) = \frac{\beta p v^{\{1-e^{-px}\}^{\beta}} \ln(v)\{1 - e^{-px}\}^{\beta-1} e^{-px}}{v - 1};$$
$$x, \beta, p, v > 0, v \neq 1$$

- EPL:

$$f(x) = \left(\frac{\alpha \theta^2 \beta x^{\beta-1}}{\theta + 1}\right)(1 + x^{\beta})e^{-\theta x^{\beta}}$$
$$\left[1 - \left(1 + \frac{\theta x^{\beta}}{1 + \theta}\right)e^{(-\theta x^{\beta})}\right]^{\alpha-1};$$
$$x, \theta, \beta > 0$$

- EG:

$$f(x) = \theta \lambda^2 x e^{(-\lambda x)}\left[1 - e^{(-\lambda x)}(\lambda x + 1)\right]^{\theta-1};$$
$$\theta, \lambda, x > 0$$

- L:

$$f(x) = \frac{q^2}{(q+1)}(1 + x)e^{-qx}; \quad x, q > 0$$

- E:

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

Data set 1: The data represent the COVID-19 case fatality ratio (in %) of China, from 8*th* March to 1*st* April of year

Table 7 Fitting of case fatality ratio (in %) for COVID-19 in India

Distribution	Log-likelihood	D	AIC	BIC	HQIC	p-value
PEL	−88.90882	0.095222	183.8176	189.7846	186.1189	0.7117
PEE	−89.20371	0.14562	184.4074	190.3744	186.7086	0.2023
EG	−91.46156	0.10876	186.9231	190.9011	188.4573	0.5454
L	−101.9173	0.27143	205.8345	207.8235	206.6016	0.0007005
EPL	−102.6339	0.36951	211.2678	217.2347	213.569	7.887e−07
E	−108.0896	0.3109	218.1792	220.1682	218.9463	5.854e−05

2022 due to new variant of COVID-19 (stealth omicron). The data is collected from official site of World Health Organization (WHO) [https://covid19.who.int/]. The data are as follows:

1.09, 1.00, 1.08, 1.12, 1.50, 1.60, 1.77, 1.81, 2.07, 1.75, 2.58, 2.59, 2.65, 3.09, 3.20, 3.47, 3.21, 2.77, 3.17, 2.65, 3.00, 3.61, 3.08, 2.70, 2.41.

From Table 4, PEL distribution is having high *p*-value, log-likelihood value and low AIC, BIC, and HQIC values when compared to other distributions. So, the PEL distribution provides a better fit for the above data-set of case fatality ratio (in %) in China due to COVID-19.

From Fig. 11, its evident that the PEL distribution fits to histogram of data better than any other distributions. Also the PEE distribution has a good fit but not up to the PEL distribution. And the total time on test (TTT) plot is concave down and monotonically increasing, so we can say that hazard rate is increasing.

Data set 2: The data represent the case fatality ratio (in %) of India, from 1*st* February to 1*st* April of year 2022 due to COVID-19. Covid cases in India started decreasing from February of 2022. The data is collected from official site of World Health Organization (WHO) [https://covid19.who.int/]. The data are as follows:

1.067, 1.757, 1.705, 1.849, 1.131, 1.595, 1.524, 1.014, 1.266, 1.678, 1.758, 1.898, 1.459, 3.370, 1.283, 1.753, 1.840, 2.134, 2.293, 2.217, 2.365, 1.485, 2.603, 2.952, 2.164, 3.142, 4.880, 2.885, 1.513, 2.704, 3.169, 2.485, 6.080, 2.462, 1.508, 1.078, 3.777, 3.407, 2.363, 5.893, 3.421, 7.211, 2.001, 2.087, 3.487, 3.457, 4.925, 2.469, 10.48, 2.514, 2.779, 2.514, 2.285, 3.895.

When compared to other distributions in Table 7, the PEL distribution is having high *p*-value, log-likelihood value and low AIC, BIC, and HQIC values. So, the PEL distribution provides the best fit for the data-set of case fatality ratio (in %) in India due to COVID-19.

From Fig. 12, it is clear that the PEL distribution performs better than any other distributions. Also the EG distribution has a good fit but not up to the PEL distribution. And the TTT plot is concave down and monotonically increasing, so we can say that hazard rate is increasing.

Data set 3: The data represent the count of new cases of COVID-19 reported in capital city of India from 11*th* of

Table 8 Summary statistics of new cases of COVID-19 in Delhi

Minimum	First quartile	Median	Mean	Third quartile	Maximum
71.0	111.8	126.5	125.1	140.2	176.0

Table 9 Parameter estimation of different distributions of new cases of COVID-19 in Delhi

Distribution	Parameter estimates
PEL	q= 0.02726960
	v = 0.01210532
	α= 11.66903275
PEE	p = 0.02874461
	v = 19.79049937
	β= 8.76642866
EPL	θ = 1.9637799
	β = 0.1106876
	α= 9.5688528
EG	λ = 0.002413117
	θ = 0.562750362
L	θ = 0.01277392
E	λ = 0.009069704

March 2022 to 11*th* of April 2022. Data is collected from JHU CSSE COVID-19 data (https://github.com/CSSEGISandData/COVID-19). Dataset is:

174, 161, 132, 136, 131, 144, 148, 140, 127, 122, 108, 104, 132, 111, 112, 120, 71, 90, 95, 123, 113, 131, 114, 85, 82, 112, 126, 176, 146, 160, 141, 137

When compared to other distributions in Table 10, the PEL distribution is having high *p*-value, log-likelihood value and low AIC, BIC, and HQIC values. As a result, for modeling the data-set of new cases of COVID-19 in Delhi, the PEL distribution is the best choice.

Figure 13 also shows that the PEL distribution provide best fit to the data set compared with other existing distributions. And the TTT plot is concave down and monotonically increasing, so we can say that hazard rate is increasing.

Fig. 12 Graphical representation of each distribution for case fatality ratio (in %) of COVID-19 in India
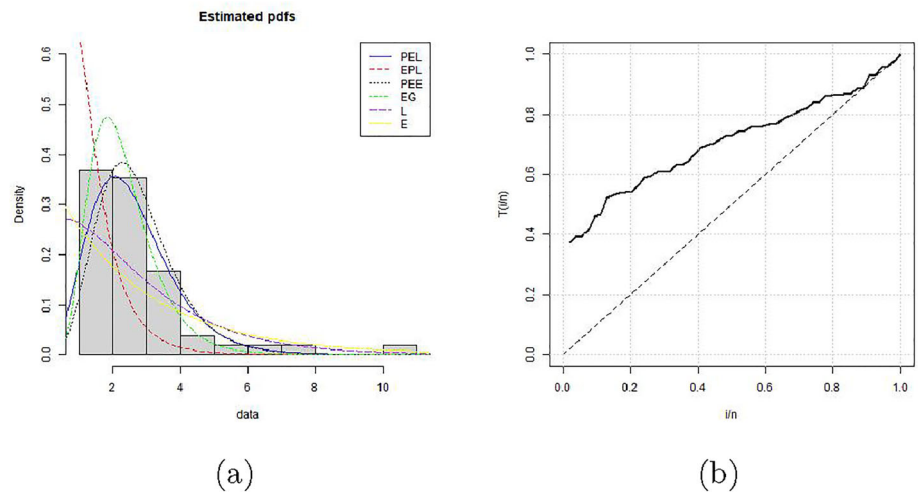

(a) (b)

Table 10 Fitting of new cases of COVID-19 in Delhi

Distribution	Log-likelihood	D	AIC	BIC	HQIC	p-value
PEL	-149.8621	0.094787	305.7242	310.1214	307.1818	0.936
PEE	-155.7403	0.18741	317.4805	321.8777	318.9381	0.211
L	-176.4869	0.33998	354.9738	356.4396	355.4597	0.001226
E	-186.8052	0.49341	375.6104	377.0761	376.0962	3.423e$-$07
EG	-213.3347	0.77917	430.6694	433.6009	431.6411	2.2e$-$16
EPL	-223.2282	0.48475	452.4564	456.8536	453.9139	5.884e$-$07

Fig. 13 Graphical representation of each distribution of new cases of COVID-19 in Delhi

(a) (b)
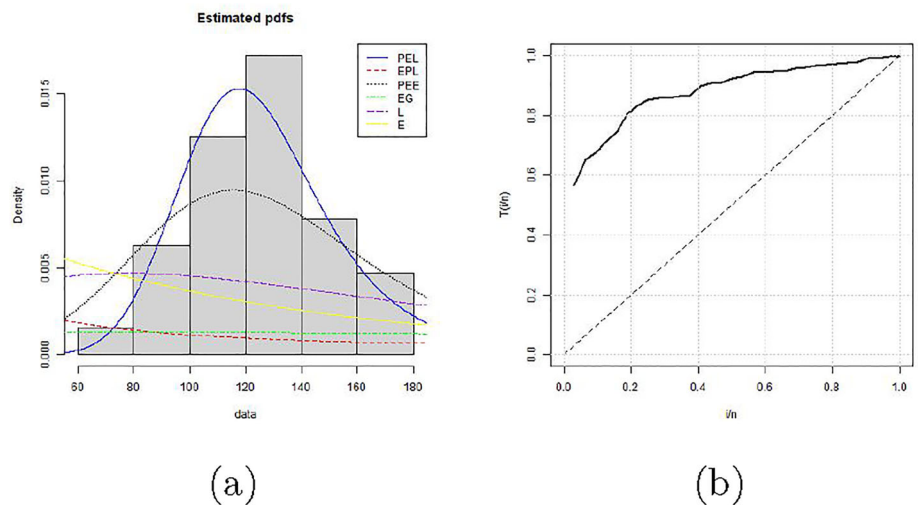
7 Result and discussion

In Sect. 5, we conducted a Monte Carlo simulation study to test the consistency of the MLEs of the PEL distribution. We considered varying sample sizes along with different combinations of parameters and for 1000 times we replicated the generated random samples. For each combination of parameters for different sample sizes, we calculated error measures like standard error, MSE, RMSE and bias. We can see that the MLEs converge to the respective parameter values as sample

size increases, also all errors and bias decrease as sample size increases.

Then to confirm the efficiency of the proposed distribution, we took three lifetime data sets based on COVID-19 and compared it with the existing distributions. To assess the fitting of the considered distributions, we calculated the AIC, BIC, HQIC, k-s statistic, log-likelihood value and p-value. Meanwhile, the best distribution is the one with the highest log-likelihood value or the lowest AIC, BIC and HQIC values. The values in Tables 4, 7 and 10 show that the PEL

distribution has highest log likelihood and *p*-value, also lowest AIC, BIC, and HQIC values. To find these values, we used the following equations;

$$AIC = 2p - 2\ln L$$
$$BIC = p\ln m - 2\ln L$$
$$HQIC = 2L + 2p\ln(\ln m)$$

where, L is the largest value of the likelihood function for the model, p is the number of predicted parameter estimates, and m is the number of observations.

Also, to get a good clarity we used pdf diagram against histogram plot of data sets and TTT-plot in Figs. 11, 12 and 13. From these figures, we can see that the PEL distribution fits better to the histogram plots of COVID-19 data sets than other considered distributions.

8 Conclusion

This study proposes a new distribution known as the PEL distribution. Different statistical properties such as hazard rate, survival function, reverse hazard function, moments, median, and mode of the new distribution are derived. Also, order statistics, stochastic ordering, and quantile function are derived. Parameters of the new distribution are estimated using the MLE method. The performance of the parameters is then tested with a Monte Carlo simulation study. We considered three real lifetime data sets to analyze the performance and superiority of the PEL distribution by comparing it with existing five distributions. From both simulation study and real data analysis, we are able to conclude that the newly proposed PEL distribution fits better to all data sets when compared to other well-known distributions and also the parameters are consistent with low error measures such as MSE, RMSE and bias.We hope that this model will be used for data analysis in many different fields such as economics, engineering, and medicine.

9 Future work

We will extend this work with regression model of PEL distribution in the next article, utilizing a categorical data and several classical and Bayesian approach estimators. Also, we will perform probabilistic machine learning for modeling the COVID-19 data. We may assess the proposed adaptability to the data by comparing it to the competitive regression models and machine learning models.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Abouelmagd, T.H.M., Hamed, M.S., Ebraheim, Abd El Hadi N.: The poisson-G family of distributions with applications. Pak. J. Stat. Oper. Res. **13**(2), 313–326 (2017)
2. Abu-Youssef, S.E., Mohammed, B.I., Sief, M.G.: An extended exponentiated exponential distribution and its properties. Int. J. Comput. Appl. **121**(5), 1–6 (2015)
3. Ahsan-ul-Haq, M., Ahmed, M., Zafar, J., Ramos, P.L.: Modeling of COVID-19 cases in Pakistan using lifetime probability distributions. Ann. Data Sci. **9**(1), 141–152 (2022)
4. Al-Babtain, A., Elbatal, I., Chesneau, C., Elgarhy, M.: Sine Topp-Leone-G family of distributions: theory and applications. Open Phys. **18**(1), 574–593 (2020)
5. Alshanbari, H.M., Gemeay, A.M., El-Bagoury, A.A.A.H., Khosa, S.K., Hafez, E.H., Muse, A.H.: A novel extension of Frechet distribution: application on real data and simulation. Alex. Eng. J. **2022**(1), 7917–7938 (2022)
6. Anzagra, L., Sarpong, S., Nasiru, S.: Chen-G class of distributions. Cogent Math. Stat. **7**(1), 1–8 (2020)
7. Ashly, R., Rajitha, C.S.: Negative binomial improved second degree Lindley distribution and its application. Adv. Math. Sci. J. **9**(2), 569–581 (2020)
8. Ashour, S.K., Eltehiwy, M.A.: Exponentiated power Lindley distribution. J. Adv. Res. **6**(6), 895–905 (2015)
9. David, H.A.: Order Statistics. Wiley, New York (1981)
10. Ghitany, M.E., Al-Mutairi, D.K., Balakrishnan, N., Al-Enezi, L.J.: Power Lindley distribution and associated inference. Comput. Stat. Data Anal. **64**(C), 20–33 (2013)
11. Ghitany, M.E., Atieh, B., Nadarajah, S.: Lindley distribution and its application. Math. Comput. Simul. **78**(4), 493–506 (2008)
12. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series and Products. Academic Press, San Diego (2007)
13. Gupta, R.D., Kundu, D.: Generalized exponential distributions. Aust. N. Z. J. Stat. **41**(2), 173–188 (1999)
14. Jowett, G.H.: The exponential distribution and its applications. Inc. Stat. **8**(2), 89–95 (1958)
15. Klakattawi, H., Alsulami, D., Elaal, M.A., Dey, S., Baharith, L.: A new generalized family of distributions based on combining Marshal-Olkin transformation with T-X family. PLoS One **17**(2), 1–18 (2022)
16. Kondo, T.: A theory of the sampling distribution of standard deviation. J. Biom. **22**(2), 34–64 (1930)
17. Lindley, D.V.: Fiducial distributions and Bayes' theorem. J. R. Stat. Soc. **20**(1), 102–107 (1958)
18. Liu, X., Ahmad, Z., Gemeay, A.M., Abdulrahman, A.T., Hafez, E.H., Khalil, N.: Modeling the survival times of the COVID-19

patients with a new statistical model: a case study from China. PLoS One **16**(7), 1–31 (2021)

19. Meriem, B., Gemeay, A.M., Almetwally, E.M., Halim, Z., Alshawarbeh, E., Abdulrahman, A.T., Abd El-Raouf, M.M., Hussam, E.: The power XLindley distribution: statistical inference, fuzzy reliability, and COVID-19 application. J. Funct. Spaces. **2022**(2), 1–21 (2022)

20. Modi, K.: Power exponentiated family of distributions with application on two real-life datasets. Thail. Stat. **19**(3), 536–546 (2021)

21. Nadarajah, S., Bakouch, H.S., Tahmasbi, R.A.: Generalized Lindley distribution. Sankhyā Indian J. Stat. Ser. B. 7 **3**(2), 331–359 (2011)

22. Nagy, M., Almetwally, E.M., Gemeay, A.M., Mohammed, H.S., Jawa, T.M., Sayed-Ahmed, N., Muse, A.H.: The new novel discrete distribution with application on COVID-19 mortality numbers in kingdom of Saudi Arabia and Latvia. Complexity **2021**(1), 1–20 (2021)

23. Pathak, A., Kumar, M., Singh, S.K., Singh, U.: Statistical inferences: based on exponentiated exponential model to assess novel corona virus (COVID-19) Kerala patient data. Ann. Data Sci. **9**(1), 101–119 (2022)

24. Rajitha, C.S., Ashly, R.: The negative binomial Akash distribution and its applications. Reliab. Theory Appl. **17**(2), 482–491 (2022)

25. Rather, N.A., Rather, T.A.: New generalizations of exponential distribution with applications. J. Prob. Stat. **2017**, 1–9 (2017)

26. Riad, F.H., Alruwaili, B., Gemeay, A.M., Hussam, E.: Statistical modeling for COVID 19 virus spread in Kingdom of Saudi Arabia and Netherlands. Alex. Eng. J. **61**(12), 9849–9866 (2022)

27. Sakthivel, K.M., Rajitha, C.S., Dhivakar, K.: Two parameter cubic rank transmutation of Lindley distribution. In: AIP Conference Proceedings Vol. 2261(1), p. 030086. (2020)

28. Shawky, A.I., Bakoban, R.A.: Exponentiated gamma distribution: different methods of estimations. J. Appl. Math. **2012**(1), 1–23 (2012)

29. Teamah, A.A.M., Elbanna, A.A., Gemeay, A.M.: Ahmed: heavy-tailed log-logistic distribution properties, risk measures and applications. Stat. Optim. Inf. Comput. **9**(4), 910–941 (2021)