Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-6, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

A New Method of Proposing Distribution and Its Application to Real Data

Sandeep K. Maurya, Arun Kaushik* & Rajwant K. Singh,
Sanjay K. Singh & Umesh Singh
[1]Department of Statistics, Banaras Hindu University, Varanasi, India.

***Abstract**: In this paper we develop a method for proposing new lifetime distribution which is parsimonious in parameter and also adds more flexibility in baseline distribution. Here one parameter exponential distribution has been taken here as baseline distribution. Here also derive the shape of the distribution mathematically, stochastic ordering and various statistical properties of like moments, conditional moments, mean deviation about mean and median, quantile, moment generating function, characteristics function, cumulant generating function, entropies and order statistics have been discussed. The method of moments and maximum likelihood estimation are used to obtain the estimate of the unknown parameter. At last, a real data set has been analyzed to show how the proposed models utilize in the practical situation.*

Introduction

Life time distribution are used to explain real life phenomenon in various field of life specially in medical, banking, marketing, engineering and others. There are a lot of life time distribution available in literature to explain such phenomenon. Some of the extensively used distribution are exponential, Weibull, gamma, normal, log normal, Rayleigh, Pareto, etc. In all of above distributions, exponential distribution is very useful lifetime distribution having several interesting properties but it is restricted to use only in situation of constant hazard rate. So that researchers are looking new life time distribution having various shapes of hazard rate. Many generalizations have been done by taking exponential distribution as the baseline distribution including Weibull and Lindley, which is also famous lifetime distributions. In the context of increasing flexibility in distribution, many generalization or transformation methods are available in the literature based on baseline distribution. Mudholkar et. al.(1993) propose three-parameter exponential Weibull distribution. Gupta et. al. (1998) propose exponentiated type distribution by adding one more shape parameter. Shaw and Buckley (2007) propose a new transformation method also by adding one parameter. Cordeiro et. al.(2013) proposed a new class of distribution by adding two more shape parameters. Kumaraswamy (1980) gives another method of proposing new distribution by taking baseline distribution. Here one point may be noted that all of the above transformation methods include few additional parameter in existing one. This leads to the addition of complexities in future inferences. In other sense, one can say that additional parameters provide flexibility but at the same time also add complexity in the estimation of parameters and others inferential procedure.

Keeping this point in mind, in this paper, we propose a new transformation method of proposing lifetime distribution by taking some baseline distribution and also having parsimonious in parameter along with increase flexibility of the baseline distribution. Let G(x) be a baseline cumulative distribution function (CDF) then our proposed transformation method is:

$$F(x) = 1 - \frac{\log(2 - G(x))}{\log 2}, \tag{1}$$

where F(x) be the new CDF and called this method, Logarithm transformed (LT) method. The associated probability distribution function (PDF) f(x) by taking g(x) as baseline pdf is:

$$f(x) = \frac{g(x)}{(2 - G(x))\log 2} \tag{2}$$

and hazard rate, h(x) is

$$h(x) = \frac{g(x)}{(2 - G(x)) * \log(2 - G(x))}. \tag{3}$$

Now for the cheeking of the flexibility of our proposed LT method we take here one parameter exponential distribution as baseline distribution, then CDF of new distribution by taking LT method define in equation (1) and say it logarithmic transformed exponential (LTE) distribution is:

$$F(x) = 1 - \frac{\log(1 + e^{-\theta x})}{\log 2} \tag{4}$$

and associated pdf is:

$$f(x) = \frac{\theta e^{-\theta x}}{(1 + e^{-\theta x})\log 2} \tag{5}$$

and hazard rate is:

$$h(x) = \frac{\theta e^{-\theta x}}{(1 + e^{-\theta x})\log(1 + e^{-\theta x})}. \tag{6}$$

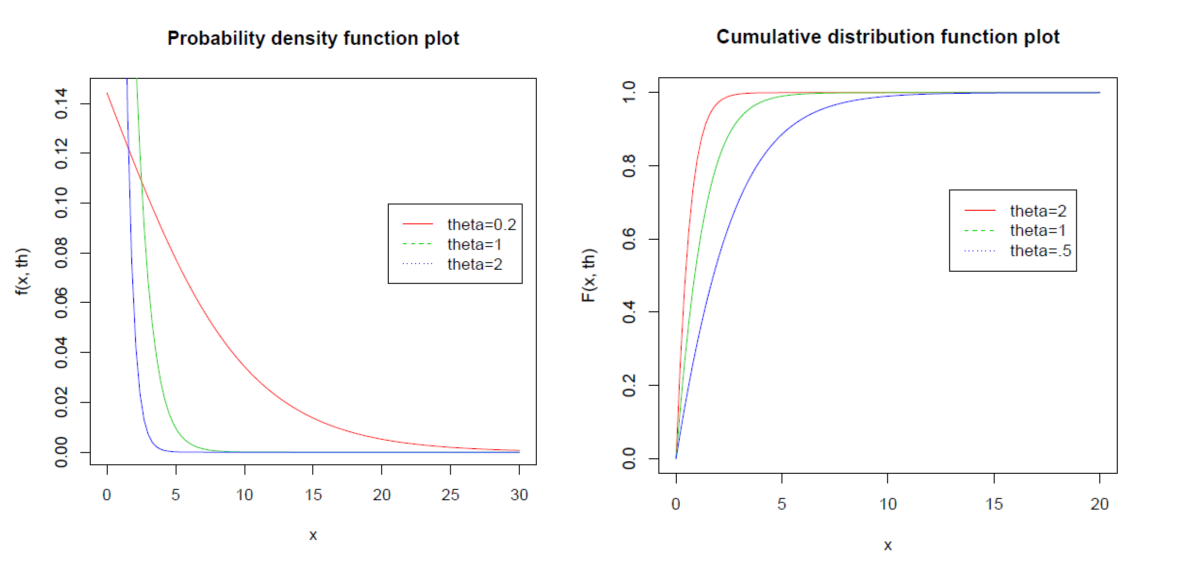

Figure 1 PDF and CDF plot of proposed distribution.

The proposed LTE distribution have increasing hazard rate. The rest of the paper is organized as follows: In section 2, we have discussed the shapes of CDF, PDF and hazard rates for various value of parameters. In section 3, we have discussed stochastic order relationship. In section 4, various statistical properties of the proposed distribution given in have been discussed. Section 5, discusses the basic inferential procedures for the distribution. In section 5, we have illustrated the suitability of proposed model for a real data set and finally the conclusions are summarized in section 6.

Shapes

The shape of distribution provides us an idea about the nature of the distribution. So that from equation (4) and (5), the CDF and PDF plots for various value of parameter are given in Figure 1. The hazard rate for different values of the parameter is plotted in Figure 2 and shows that proposed distribution have increasing hazard rate. It can be mathematically shown that the proposed distribution has increasing hazard rate by following the result of Glaser (1980). He defined the term $\eta(t)=\frac{-f'(t)}{f(t)}$, where $f(t)$ denotes density function and $f'(t)$ be the first derivative of $f(t)$ with respect to t and proof that-

1. If $\eta'(t)$>0 for all t, then distribution has increasing failure rate (IFR).

Since in case of our proposed distribution we see that;

$$\eta(t)=\frac{\theta}{(1+e^{-\theta t})}$$

and

$$\eta'(t)=\frac{\theta^2 e^{-\theta x}}{(1+e^{-\theta x})^2}. \qquad (7)$$

Since $e^{-\theta x}>0$ for all $t>0$, hence from equation (7) we can say that $\eta'(t)>0$ for all $\theta\geq 0$. Hence, our proposed distribution has increasing failure rate (IFR).

1. Stochastic Order

Let X_1 and X_2 be random variable having CDF $F_1(x)$ and $F_2(x)$ with parameter θ_1 and θ_2 respectively. Then X_1 is said to stochastically greater than X_2 if $F_1(x)\leq F_2(x)$ (see Gupta et al. (1998) for more detail).

Let $\theta_1<\theta_2$ and X_1 and X_2 be random variable follows proposed distribution, then

$$\frac{F_1(x)}{F_2(x)}=\frac{\log 2-\log(1+e^{-\theta_1 x})}{\log 2-\log(1+e^{-\theta_2 x})}.$$

This shows that X_1 is stochastically greater than X_2 for $\theta_1<\theta_2$.

2. Statistical Properties

Various statistical properties of our proposed distribution like moments, conditional moments, moment generating function (MGF), characteristics function (CHF), cumulant generating Function (CGF), mean deviation about mean and mean deviation about median, quantile, order statistic, Renyi entropy and Shannon entropy are discussed below.

2.1. Moments

Moments of distribution is an important property of distribution. However, in order to obtain derivation of the expression of moments, we shall derive the following lemma.

$$L_1(\theta, r, \delta) = \int_0^\infty \frac{x^r e^{-\delta x}}{(1+e^{-\theta x})} dx$$
$$= \sum_{k=0}^{\infty} (-1)^k \frac{r!}{(\delta+\theta k)^{r+1}}.$$

Proof. We know that infinite sum of geometric series $\sum_{k=0}^{\infty} (-1)^k e^{-\theta k x} = \frac{1}{(1+e^{-\theta x})}$ we get,

$$L_1(\theta, r, \delta) = \sum_{k=0}^{\infty} (-1)^k \int_0^\infty x^r e^{-(\delta+\theta k)x} dx$$

using the result of the gamma function, we get

$$= \sum_{k=0}^{\infty} (-1)^k \frac{r!}{(\delta+\theta k)^{r+1}}.$$

Using the above lemma 4.1, we get r^{th} moments as,

$$E(X^r) = \frac{\theta}{\log 2} L_1(\theta, r, \theta).$$

Also, we know that Hence mean of proposed distribution is,

$$E(X) = \frac{\theta}{\log 2} L_1(\theta, 1, \theta)$$
$$= \frac{1}{\theta \log 2} \sum_{k=0}^{\infty} \frac{(-1)^k}{(1+k)^2}.$$

Now using the result of Euler (1735) which was obtained by solving Basel problem and found that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \pi^2/6.$ Utilizing this result we get the infinite sum of the series $1 - 1/2^2 + 1/3^2 - 1/4^2 .... = \pi^2/12.$. Hence, the mean of proposed distribution is $E(X) = \frac{\pi^2}{12\theta \log 2}.$

Similarly, using the lemma 4.1, other moments can be obtained and thus the variance, skewness and kurtosis of distribution of X can be easily obtained.

2.2. Conditional Moments

Before going to expressions of conditional moments, we can be derived the following lemma. Let

$$L_2(\theta, r, \delta, t) = \int_t^\infty \frac{x^r e^{-\delta x}}{(1+e^{-\theta x})} dx$$
$$= \sum_{k=0}^{\infty} \sum_{l=0}^{r} \frac{r!}{l!} \frac{t^l e^{-(\delta+k\theta)t}}{(\delta+k\theta)^{r+1-l}}. \quad (8)$$

Proof. We use the similar presses as used in lemma 4.1, we get

$$L_2(\theta, r, \delta, t) = \sum_{k=0}^{\infty} (-1)^k \int_t^\infty x^r e^{-(\delta+\theta k)x} dx. \quad (10)$$

Now using complementary incomplete gamma function is defined as $\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$. Also, this function can be written as $(a-1)! e^{-x} \sum_{l=0}^{a-1} x^l / l!$. Then using the complementary incomplete gamma function in equation (10) and after simplifying, we have

$$L_2(\theta, r, \delta, t) = \sum_{k=0}^{\infty} \sum_{l=0}^{r} \frac{r!}{l!} \frac{t^l e^{-(\delta+k\theta)t}}{(\delta+k\theta)^{r+1-l}}.$$

(9)

Hence by the lemma 4.2, the r^{th} conditional moments can be obtained as follows,

$$E(X^r \mid X > x) = \frac{\theta}{(1-F(x))\log 2} L_2(\theta, r, \theta, x). \quad (11)$$

2.3. Moment generating function, characteristics function, and cumulant generating function

Let X denotes the random variable of our proposed distribution, then its moment generating function (MGF) is given as follows:

$$M_X(t) = \frac{\theta}{\log 2} L_1(\theta, 0, \theta - t) \qquad for \qquad t < \theta. \qquad (12)$$

Similarly, Characteristic function (CHF) of X can be found as,

$$\phi_X(t) = \frac{\theta}{\log 2} L_1(\theta, 0, \theta - it) \qquad (13)$$

where $i=\sqrt{-1}$ represents imaginary and cumulant generating function (CGF) of X found as,

$$K_X(t)=\log(\frac{\theta}{\log 2})+\log L_1(\theta,0,\theta-t). \qquad (14)$$

2.4. Mean Deviation

Expression of mean deviation about mean and mean deviation about median is defined by, $\delta_1(X)=\int_0^\infty |x-\mu|\, f(x)dx$ and $\delta_2(X)=\int_0^\infty |x-M|\, f(x)dx$ respectively, where μ stands for mean, and M stands for the median. Then mean deviation about mean is derived as,

$$\delta_1(X)=\int_0^\mu(\mu-x)f(x)dx+\int_\mu^\infty(x-\mu)f(x)dx.$$

Now utilizing integral by part we have,

$$\delta_1(X)=2\mu F(\mu)-2\mu+2\int_\mu^\infty xf(x)dx,$$

Where $F(\cdot)$ denotes the proposed CDF. Also, from Lemma 4.2,

$$\int_\mu^\infty f(x)dx=\frac{\theta}{\log 2}L_2(\theta,1,\theta,\mu),$$

and thus

$$\delta_1(X)=2\mu F(\mu)-2\mu+\frac{\theta}{\log 2}L_2(\theta,1,\theta,\mu).$$

By the similar process, the mean deviation about median is defined as,

$$\delta_2(X)=\int_0^M(M-x)f(x)dx+\int_M^\infty(x-M)f(x)dx$$

the remain steps are same as in mean deviation about the mean, we get

$$\delta_2(X)=-\mu+2\int_M^\infty xf(x)dx.$$

By Lemma 4.2,

$$\int_M^\infty xf(x)dx=\frac{\theta}{\log 2}L_2(\theta,1,\theta,M).$$

Hence, the mean deviation about the median can be easily obtained as,

$$\delta_2(X)=-\mu+\frac{\theta}{\log 2}L_2(\theta,1,\theta,M). \qquad (15)$$

2.5. Quantile Function

The p^{th} quantile of the quantile function $Q(p)$ is can be obtained by solving the equation $F(Q(p))=p$. Hence from equation (4), we get

$$Q(p)=\frac{-1}{\theta}\log[2^{1-p}-1] \qquad (16)$$

2.6. Bonferroni, Lorenz curves and Gini index

Lorenz curve was defined by Lorenz (1905), is used to describe the inequality in the distribution of quantity, especially in the field of economics as in term of income and wealth. It is a continuous function for probability distribution and defines as:

$$L(p)=\frac{1}{\mu}\int_0^q xf(x)dx,$$

$$\text{or}\quad L(p)=\frac{1}{\mu}\int_0^p F^{-1}(x)dx, \qquad (17)$$

where μ stands for the mean of the distribution and $q=F^{-1}(p)$. Bonferroni curve was determined by Bonferroni (1930), and it also measures the same. Bonferroni and Lorenz's curve is not only applicable in the field of economics but also other areas like reliability, medical, demography. It is defined as:

$$B(p)=\frac{1}{\mu p}\int_0^q xf(x)dx,$$

$$\text{or}\quad B(p)=\frac{1}{\mu p}\int_0^p F^{-1}(x)dx. \qquad (18)$$

A Lorenz curve cannot be define if the mean of distribution is zero or infinite. The amount of information in a Lorenz curve is given by Gini coefficient and Lorenz asymmetry coefficient (LAC). The Gini index is defined as:

$$G=1-2\int_0^1 L(p)dp. \qquad (19)$$

If X has proposed LTE distribution then by equation (5) and using the lemma 4.2 in equation (17) and (18), we get the Lorenz curve is

$$L(p)=\frac{1}{\mu}\left[\mu-\frac{\theta}{\log 2}L_2(\theta,1,\theta,q)\right]$$
$$=1-\frac{12}{\pi^2}\sum_{k=0}^{\infty}\frac{(2^{1-p}-1)^{1+k}}{1+k}\left[\frac{1}{1+k}-\log(2^{1-p}-1)\right] \qquad (20)$$

and Bonferroni curve is

$$B(p)=\frac{1}{\mu p}\left[\mu-\frac{\theta}{\log 2}L_2(\theta,1,\theta,q)\right]$$
$$=\frac{1}{p}-\frac{12}{\pi^2 p}\sum_{k=0}^{\infty}\frac{(2^{1-p}-1)^{1+k}}{1+k}\left[\frac{1}{1+k}-\log(2^{1-p}-1)\right] \qquad (21)$$

respectively. Now using the equation (20) in equation (19) we get, the Gini index

$$G = -1 + \frac{24}{\pi^2 \log 2}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^l}{(1+k)^2} * \left[\frac{1}{(k+l+2)} + \frac{k+1}{(k+l+2)^2}\right] \quad (22)$$

And LAC is defined as:

$$S = F(\mu) + L(\mu), \quad (23)$$

where $F(\cdot)$ defined in equation (4) and $L(\cdot)$ is defined in equation (20) and for $S = 1$, the Lorenz curve is symmetric.

2.7. Order Statistics

Let we take $X_1, X_2, \ldots, X_n$ be n random sample from the proposed distribution and the corresponding order statistics is, $X_{1:n} < X_{2:n} < \ldots < X_{n:n}$. It is well known that the pdf of r^{th} (for $r = 1,2,\ldots,n$) order statistics $X_{r:n}$, say $f_r(x)$, when the population CDF and PDF are F(x) and f(x) respectively, is given as,

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!}F^{r-1}(x) * [1-F(x)]^{n-r} f(x)$$

$$= \frac{n!}{(r-1)!(n-r)!}\sum_{i=0}^{n-r}(-1)^i\binom{n-r}{i}F^{r+i-1}(x)f(x). \quad (24)$$

Hence using equation (4) and (5) in (24) we get,

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!}\frac{\theta}{\log 2}\sum_{j=0}^{n-r}(-1)^j\binom{n-r}{j} * \frac{e^{-\theta x}}{(1+e^{-\theta x})}\left[1-\frac{\log(1+e^{-\theta x})}{\log 2}\right]^{r+j-1} \quad (25)$$

and corresponding r^{th} order statistic of CDF $F_r(x)$ is,

$$F_r(x) = \sum_{l=r}^{n}\binom{n}{l}F^l(x)(1-F(x))^{n-l} = \sum_{l=r}^{n}\sum_{m=0}^{n-l}\binom{n}{l}\binom{n-l}{m}(-1)^m F^{l+m}(x). \quad (26)$$

So that from equation (4), equation (26) can be written as,

$$F_r(x) = \sum_{l=r}^{n}\sum_{m=0}^{n-l}(-1)^m\binom{n}{l}\binom{n-l}{m} * \left[1-\frac{\log(1+e^{-\theta x})}{\log 2}\right]^{l+m} \quad (27)$$

2.8. Entropies

Entropy is measure of average amount of uncertainty removed upon revealing the outcome of X. Let X be a random variable having pdf f(x) then Renyi entropy is proposed by [Renyi(1961)] and defined as, $\mathrm{J}_R(\gamma) = \frac{1}{1-\gamma}\log[\int f^{\gamma}(x)dx]$ where $\gamma > 0$ and $\gamma \neq 1$. From equation (5) we get,

$$\int_0^{\infty} f^{\gamma}(x)dx = \left(\frac{\theta}{\log(2)}\right)^{\gamma}\int_0^{\infty}\left[\frac{e^{-\theta x}}{(1+e^{-\theta x})}\right]^{\gamma}dx \quad (28)$$

after simplification, we get,

$$\int_0^{\infty} f^{\gamma}(x)dx = \left(\frac{\theta}{\log 2}\right)^{\gamma}\sum_{i=0}^{\infty}\binom{-\gamma}{i}\frac{1}{\theta(i+\gamma)}$$

For more detail on the expansion of binomial series see Graham et. al.(1989). Hence, we get,

$$\mathrm{J}_R(\gamma) = \frac{\gamma}{1-\gamma}\log\left[\frac{\theta}{\log 2}\right] + \frac{1}{1-\gamma}\log[\sum_{i=0}^{\infty}\binom{-\gamma}{i}\frac{1}{\theta(i+\gamma)}] \quad (29)$$

Shannon entropy was proposed by Shannon (1951) and can be defined as $E[-\log f(x)]$. It behaves as a particular case of Re n′nyi's entropy for $\gamma \to 1$. Using L' Hospital Rule for $\gamma \to 1$ we get,

$$-\log f(x) = -\log\left(\frac{\theta}{\log 2}\right) + \theta x + \sum_{i=1}^{\infty}(-1)^{i-1}\frac{e^{-\theta x i}}{i} \quad (30)$$

and hence,

$$E[-\log f(x)] = -\log\left(\frac{\theta}{\log 2}\right) + \frac{\pi^2}{12\log 2} + \frac{1}{\log 2}\sum_{i=1}^{\infty}\sum_{k=0}^{\infty}\frac{(-1)^{i+k-1}}{(1+i+k)^i}. \quad (31)$$

Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-6, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

3. Estimation

In this section we consider two method for estimation of parameter first one by the method of moments and another is maximum likelihood estimate method and also derive associated Fisher information matrix along with confidence interval for parameter. Let $X_1, X_2, \ldots, X_n$ be n random sample from the proposed distribution. For the estimate of the parameter through the method of moments, we have to compare the theoretical mean of the proposed distribution, to sample mean. Let m be the sample mean. Then by equating theoretical moments with the sample mean, we have an equation,

$$\frac{\pi^2}{12\theta \log 2} = m$$

and hence through the method of moments, the estimate $\hat{\theta}$ of parameter θ is,

$$\hat{\theta} = \frac{\pi^2}{12 m \log 2}. \qquad (32)$$

Now consider the method of maximum likelihood estimation of the parameter θ of the proposed distribution. MLE is obtained by maximizing the log-likelihood function. The log-likelihood function is,

$$\log L = n \log\left(\frac{\theta}{\log 2}\right) + \sum_{i=1}^{n}\left(\theta x_i + \log(1+e^{-\theta x_i})\right). \qquad (33)$$

Differentiating it concerning the parameter θ we get,

$$\frac{\partial \log L}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} \frac{x_i}{1+e^{-\theta x_i}}. \qquad (34)$$

After equating the equation (34) to zero, we get a nonlinear equation and after solving the equation we get MLE $\hat{\theta}$ of parameters θ. Also, note that this equation cannot be solved analytically. However, one can use some numerical technique for the solution. Therefore, we propose the use of Newton-Raphson method. For the choice of initial guess, contour plot technique is used. Fisher Information matrix can be estimated by,

$$I(\hat{\theta}) = \left[\frac{-\partial^2}{\partial \theta^2} \log L\right]_{\hat{\theta}} \qquad (35)$$

where,

$$\frac{-\partial^2}{\partial \theta^2} \log L = \frac{n}{\theta^2} + \sum_{i=1}^{n} e^{-\theta x_i}\left(\frac{x_i}{1+e^{-\theta x_i}}\right)^2.$$

For large samples, we can obtain the confidence intervals based on Fisher information matrix $I^{-1}(\hat{\theta})$ which provides the estimated asymptotic variance for the parameter θ. Thus, a two-sided $100(1-\eta)\%$ confidence interval of θ can be defined as $\hat{\theta} \pm Z_{\eta}/2\sqrt{var(\hat{\theta})}$. Where $Z_{\eta}/2$ denotes the upper $\eta/2\%$ point of standard normal distribution.

(32)

4. Real Data Application

In this section, we use a real data set for model suitability of our proposed LTE. Further, to show the superiority of our proposed distribution, we take five well-known lifetime distributions as:

1-Exponential distribution with pdf:

(33)

$$f(x) = \theta e^{-\theta x} \qquad x > 0, \theta > 0.$$

2-Lindley distribution with pdf:

$$f(x) = \frac{\theta^2}{\theta+1}(1+x)e^{-\theta x} \qquad x > 0, \theta > 0.$$

(34)

3-Gamma distribution with pdf:

$$f(x) = \frac{1}{\theta^{\alpha}} \frac{x^{\theta-1}}{\Gamma\theta} e^{-\frac{x}{\theta}}$$

$$x > 0, \theta > 0, \alpha > 0.$$

Table 1: MLE, AIC, BIC and KS statistics with the p-value for fitted data sets.

Distribution	ML Estimate			KS- Test		AIC	BIC
	α	θ	Log-Likelihood	Statistics	p-value		
Exponential	-	0.0219	-241.090	0.1911	0.0520	484.179	486.091
Lindley	-	0.0429	-251.430	0.1990	0.0381	504.861	506.773
LTE	-	0.0259	-240.086	0.1839	0.0679	482.172	484.084
Weibull	0.9490	44.9126	-241.002	0.1928	0.0486	486.004	489.828
Gamma	0.7991	57.1716	-240.190	0.2022	0.0335	484.380	488.205
EE	0.7798	0.0187	-239.995	0.2042	0.0310	483.990	487.814

4-Gamma distribution with pdf:

$$f(x) = \frac{\alpha}{\theta}\left(\frac{x}{\theta}\right)^{\alpha-1} e^{-\left(\frac{x}{\theta}\right)^{\alpha}}$$

$$x > 0, \theta > 0, \alpha > 0.$$

5-Exponentiated exponential (EE) distribution with pdf:

$$f(x) = \alpha\theta e^{-\theta x}(1-e^{-\theta x})^{\alpha}$$

$$x > 0, \theta > 0, \alpha > 0.$$

The data set includes the lifetime of 50 devices and was taken by Aarset (1987) and analyzed by Sarhan and Kundu (2009) and so on and listed as:

0.1, 0.2, 1, 1, 1, 1, 1, 2, 3, 6,
7, 11, 12, 18, 18, 18, 18, 18, 21, 32,
36, 40, 45, 46, 47, 50, 55, 60, 63, 63,
67, 67, 67, 67, 72, 75, 79, 82, 82, 83,
84, 84, 84, 85, 85, 85, 85, 85, 86, 86.

We have used AIC (Akaike Information Criterion), BIC (Bayesian information criterion) and KS test and estimated log-likelihood value for model their suitability, regarding fitting. The AIC, BIC, and KS distance (D) are defined as,

$$AIC = 2*k - 2*\log \hat{L},$$

$$BIC = k*\log(n) - 2*\log \hat{L},$$

$$D = \underset{x}{Sup}|F_n(x) - F(x)|,$$

where $$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{x_i \le x}$$

is empirical distribution function, F(x) is CDF, n is sample size, k is the number of parameters and $\hat{L}$ is the likelihood value for the considered distribution. For an indication of better fit of distributions, we have a smaller value of AIC, BIC, and the K-S test statistic. We have also calculated the MLEs of parameters for various distributions and estimated log-likelihood value.

The value of AIC, BIC, and K-S statistic along with associated p-value and estimated log-likelihood value with maximum likelihood estimates of parameters from the considered real data set of compared distributions are present in Table 1. Description of fitted model is given below:

We found that out of all the previously discussed five distributions only exponential, LTE and Weibull fit the data set at 5% level of significance. However, from K-S statistics or associated p-value, we may say that LTE provides better fit than exponential and Weibull distribution. Also, the similar result is drawn by estimated log-likelihood value which is maximum than the other two. It is also noted that LTE has smaller AIC and BIC in comparison to other five considered distributions. Hence, we conclude that our proposed LTE distribution is a more suitable for the present data as compared to other five distributions.

The empirical cumulative distribution function (ECDF) and fitted CDF plot, for the considered data set, has been shown in Figure 5. That figure also implies that the proposed LTE distribution is a better fit to the considered real data in comparison to other competitive models.

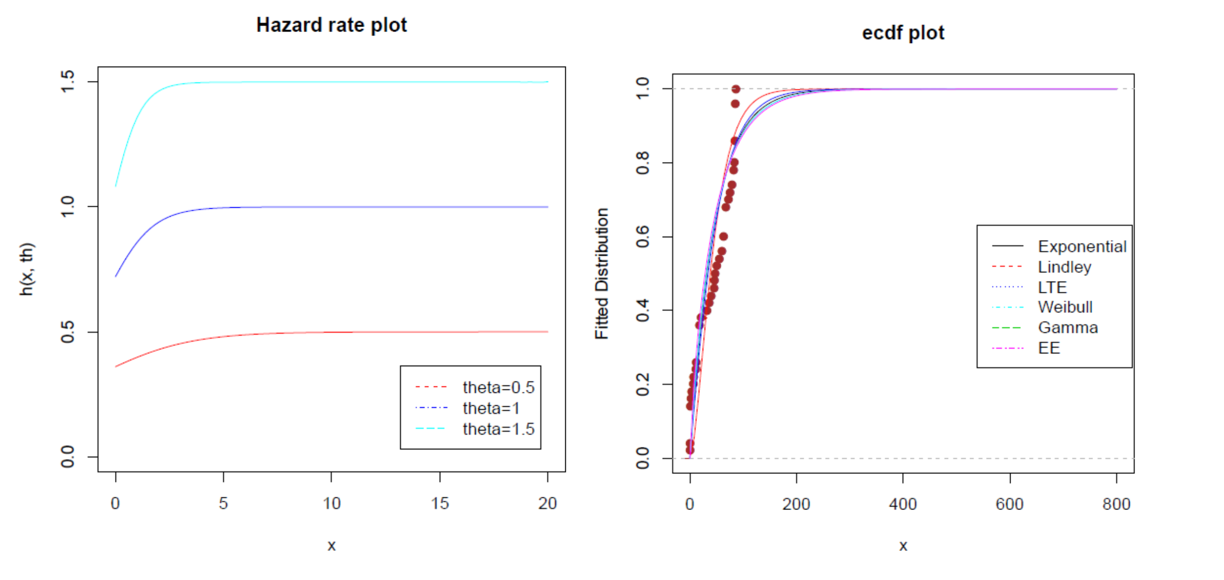

Figure 2. ECDF and fitted CDF plot of the real data set.

5. Conclusion

In the present paper we propose a new transformation, say LT transformation to generate new lifetime distribution and to show the suitability of transformation, we have taken exponential distribution as baseline distribution. The proposed LTE distribution has increasing hazard rate. Here also we derived its mean, r^{th} moments, r^{th} conditional moments, moment generating function, characteristics function, cumulant generating function, mean deviation about mean and median, quantile, Renyi and Shannon entropies, r^{th} order statistic for CDF and PDF for the proposed distribution have been derived. Along this, for estimation of the parameter, we have used the method of moments and method of maximum likelihood estimate, $100(1-\eta)\%$ confidence interval, and expected Fisher information matrix has also been discussed.

Lastly, we consider a real data set and five other distribution namely exponential, Lindley, Weibull, gamma, exponentiated exponential. It is shown that our proposed LTE distribution fits the considered real data set very well even better than other distributions. This providing good fit to the considered data set. Hence, we can easily conclude that our proposed LTE distribution may be considered as a suitable model for lifetime data.

References

[1] M. V. Aarset. How to identify a bathtub hazard rate. IEEE Transactions on Reliability, 360 (1):0 106–108, 1987.

[2] C. E. Bonferroni. Elementi di statistica generale. 1930.

[3] G. M. Cordeiro, E. M. M. Ortega, and C. C. da-Cunha Daniel. The exponentiated generalized class of distributions. Journal of Data Science, 11:0 1–27, 2013.

[4] R. E. Glaser. Bathtub and related failure rate characterizations. Journal of the American Statistical Association, 75:0 667–672, 1980.

[5] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete mathematics*. 1989.

[6] R .C. Gupta, P. L. Gupta, and R .D. Gupta. Modeling failure time data by Lehmann alternatives. Communications in Statistics-Theory and Methods, 27:0 887–904, 1998.

[7] P. Kumaraswamy. Generalized probability density-function for double random- processes. Journal of Hydrology, 462:0 79–88, 1980.

[8] M. O. Lorenz. Methods of measuring the concentration of wealth. Publications of the American statistical association, 90 (70):0 209–219, 1905.

[9] G. S. Mudholkar, D. K. Srivastava, and G. D. Kollia. Exponentiated weibull family for analyzing bathtub failure-rate data. IEEE Transactions on Reliability, 42:0 299–302, 1993.

[10] A. Renyi. On measures of entropy and information. In Proceedings of the 4th Berkeley symposium on mathematical statistics and probability. Berkeley University of California Press, 1:0 547–561, 1961.

[11] A. M. Sarhan and D. Kundu. Generalized linear failure rate distribution. Communications in Statistics: Theory and Methods, 380 (5):0 642–660, 2009.

[12] C .E. Shannon. Prediction and entropy of printed English. The Bell System Technical Journal, 30:0 50–64, 1951.

[13] W. T. Shaw and I. R. Buckley. The alchemy of probability distribution: Beyond gram-charlier cornish-fisher expansions, and skew-normal and kurtotic-normal distribution. Research report, 2007.