International Journal of Statistics and Applications 2015, 5(6): 338-348
DOI: 10.5923/j.statistics.20150506.08

Shanker Distribution and Its Applications

Rama Shanker

Department of Statistics, Eritrea Institute of Technology, Asmara, Eritrea

Abstract In this paper a new one parameter lifetime distribution named "Shanker distribution" for modeling lifetime data has been suggested. Various mathematical properties of the new distribution including its shape, moments, generating functions, hazard rate and mean residual life functions, stochastic ordering, order statistics, Renyi entropy measure, mean deviations, Bonferroni and Lorenz curves, and stress-strength reliability have been presented. The conditions of over-dispersion, equi-dispersion, and under-dispersion of Shanker distribution have been discussed. The method of maximum likelihood estimation and the method of moments have been discussed for estimating its parameter. The usefulness, applicability and superiority of the proposed distribution over exponential and Lindley distributions have been illustrated with three real lifetime data- sets.

Keywords Moments, Hazard rate function, Mean residual life function, Mean deviations, Order statistics, Stress-strength reliability, Estimation of parameter, Goodness of fit

1. Introduction

A number of continuous distributions for modeling lifetime data such as exponential, Lindley, gamma, lognormal, and Weibull are available in statistical literature. The exponential, Lindley and the Weibull distributions are more popular than the gamma and the lognormal distributions for modeling lifetime data-sets because the survival functions of the gamma and the lognormal distributions cannot be expressed in closed forms and both require numerical integration. Though each of exponential and Lindley distributions has one parameter, the Lindley distribution has one advantage over the exponential distribution that the exponential distribution has constant hazard rate whereas the Lindley distribution has monotonically decreasing hazard rate.

Lindley distribution, introduced in the context of Bayesian analysis as a counter example of fiducial statistics, having its probability density function (p.d.f.) and cumulative distribution function (c.d.f.)

$$f(x;\theta)=\frac{\theta^2}{\theta+1}(1+x)e^{-\theta x}\quad ;x>0,\ \theta>0 \qquad (1.1)$$

$$F(x;\theta)=1-\left[1+\frac{\theta x}{\theta+1}\right]e^{-\theta x}\ ;x>0,\theta>0 \qquad (1.2)$$

has been introduced by Lindley (1958). A detailed study about its mathematical properties, estimation of parameter and application showing the superiority of Lindley distribution over exponential distribution for the waiting times before service of the bank customers has been done by Ghitany *et al* (2008). The Lindley distribution has been generalized, extended and modified by different researchers including Zakerzadeh and Dolati (2009), Nadarajah *et al* (2011), Deniz and Ojeda (2011), Bakouch et al (2012), Shanker and Mishra (2013 a, 2013 b), Shanker *et al* (2013), Elbatal *et al* (2013), Ghitany *et al* (2013), Merovci (2013), Liyanage and Pararai (2014), Ashour and Eltehiwy (2014), Oluyede and Yang (2014), Singh *et al* (2014), Sharma *et al* (2015), Shanker *et al* (2015), Alkarni (2015), Pararai *et al* (2015), Abouammoh *et al* (2015) are some among others.

Shanker *et al* (2015) has detailed study about the comparative study of exponential, and Lindley distributions for modeling various real lifetime data-sets and found that in some data-sets Lindley is better than exponential while in others exponential is better than Lindley. In this paper our objective is to introduce a distribution which gives better fitting than both exponential and Lindley distributions for modeling real lifetime data-sets from various fields of knowledge.

In this paper, a new one parameter lifetime distribution having its probability density function (p.d.f.)

$$f(x;\theta)=\frac{\theta^2}{\theta^2+1}(\theta+x)e^{-\theta x}\quad ;x>0,\ \theta>0 \quad (1.3)$$

has been suggested for modeling various types of real lifetime data-set. We would call this distribution, "Shanker distribution". The p.d.f. (1.3) can be shown as a mixture of exponential (θ) and gamma $(2,\theta)$ distributions as follows:

* Corresponding author:
shankerrama2009@gmail.com (Rama Shanker)
Published online at http://journal.sapub.org/statistics

$$f(x;\theta) = pf_1(x,\theta) + (1-p) f_2(x,\theta) \qquad (1.4)$$

where $p = \frac{\theta^2}{\theta^2+1}$, $f_1(x,\theta) = \theta e^{-\theta x}$ and $f_2(x,\theta) = \theta^2 x e^{-\theta x}$.

The first derivative of (1.3) is obtained as

$$\frac{d}{dx} f(x;\theta) = \frac{\theta^2}{\theta^2+1}\left(1-\theta^2-\theta x\right)e^{-\theta x}$$

Thus $\frac{d}{dx} f(x;\theta) = 0$ gives $x_0 = \frac{1-\theta^2}{\theta}$. From this it follows that

(i) for $0<\theta<1$, $\frac{d}{dx} f(x;\theta) = 0$ implies that $x_0 = \frac{1-\theta^2}{\theta}$ is the unique critical point at which $f(x;\theta)$ is maximized.

(ii) for $\theta \ge 1$, $\frac{d}{dx} f(x;\theta) \le 0$ implies that $f(x;\theta)$ is decreasing in x.

Therefore, the mode of the distribution (1.3) is given by

$$\text{Mode} = \begin{cases} \frac{1-\theta^2}{\theta}, & 0<\theta<1 \\ 0, & \text{otherwise} \end{cases} \qquad (1.5)$$

The corresponding cumulative distribution function (c.d.f.) is given by

$$F(x,\theta) = 1 - \frac{(\theta^2+1)+\theta x}{\theta^2+1} e^{-\theta x};\ x>0, \theta>0 \qquad (1.6)$$

The graphs of the p.d.f. and the c.d.f. of Lindley and Shanker distributions for different values of θ are shown in **figures 1 and 2.**

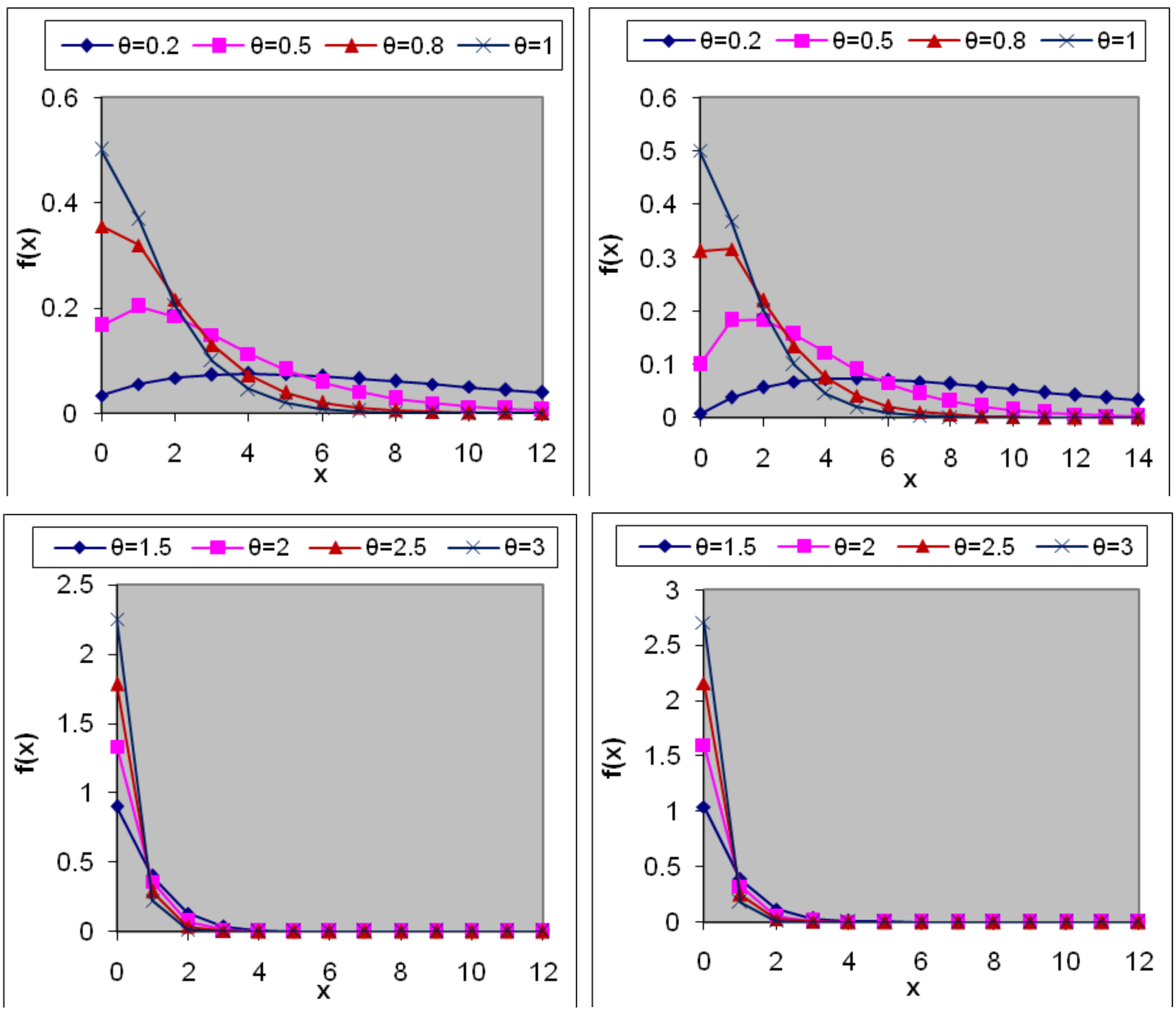

Figure 1. Graphs of the pdf of Lindley and Shanker distributions for different values of parameter θ. Left hand side graphs are for Lindley distribution and right hand side graphs are for Shanker distribution

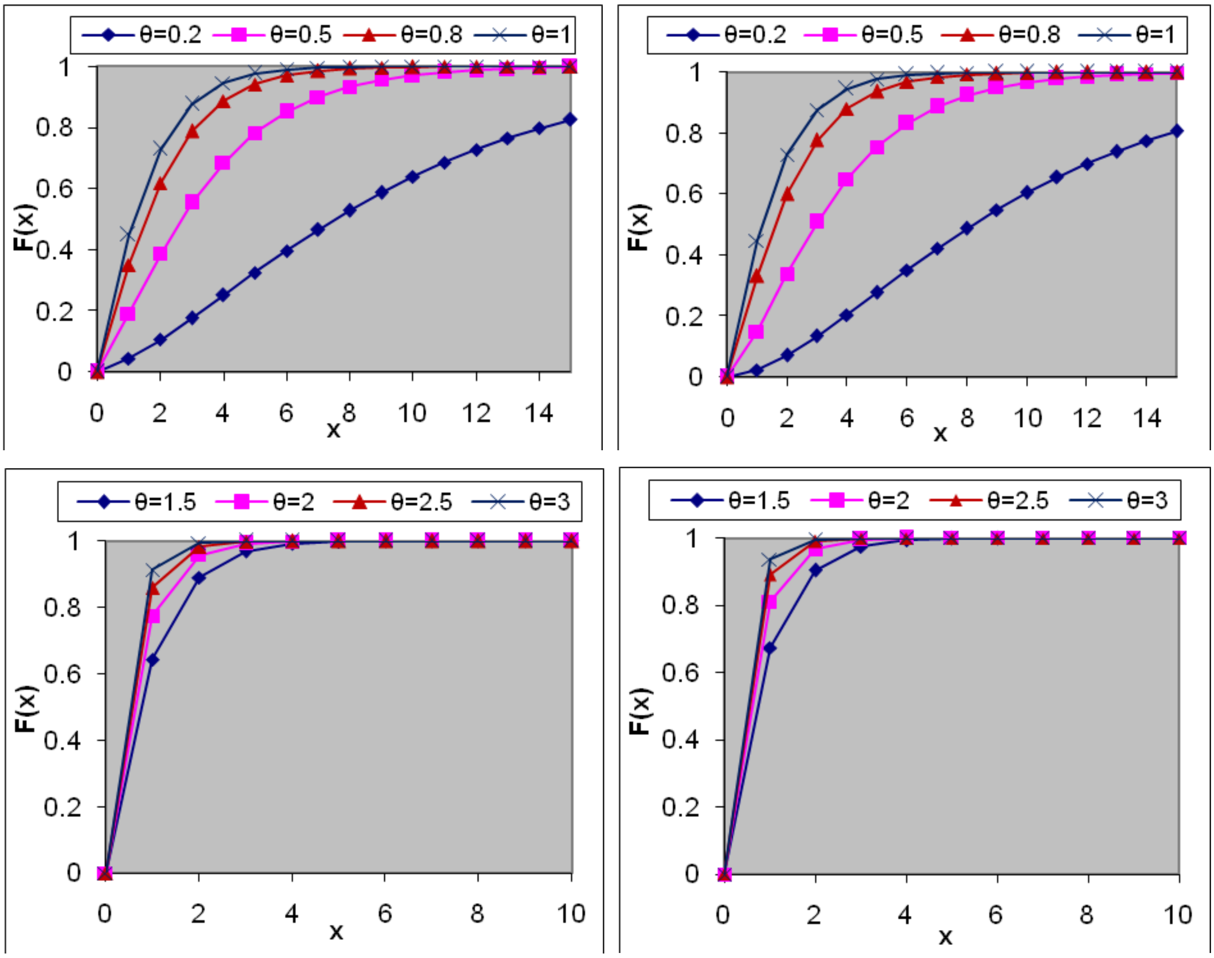

Figure 2. Graphs of the cdf of Lindley and Shanker distribution for different values of parameter θ. Left hand side graphs are for Lindley distribution and right hand side graphs are for Shanker distribution

2. Moments and Associated Measures

The r the moment about origin of Shanker distributon (1.3) has been obtained as

$$\mu_r' = \frac{r!\left(\theta^2 + r + 1\right)}{\theta^r\left(\theta^2 + 1\right)} \quad ; r = 1, 2, 3, ...$$

and so the first four moments about origin as

$$\mu_1' = \frac{\theta^2 + 2}{\theta\left(\theta^2 + 1\right)}, \qquad \mu_2' = \frac{2\left(\theta^2 + 3\right)}{\theta^2\left(\theta^2 + 1\right)},$$

$$\mu_3' = \frac{6\left(\theta^2 + 4\right)}{\theta^3\left(\theta^2 + 1\right)}, \qquad \mu_4' = \frac{24\left(\theta^2 + 5\right)}{\theta^4\left(\theta^2 + 1\right)}$$

Thus the moments about mean of the Shanker distribution are as follows

$$\mu_2 = \frac{\theta^4 + 4\theta^2 + 2}{\theta^2\left(\theta^2 + 1\right)^2}$$

$$\mu_3 = \frac{2\left(\theta^6 + 6\theta^4 + 6\theta^2 + 2\right)}{\theta^3\left(\theta^2 + 1\right)^3}$$

$$\mu_4 = \frac{3\left(3\theta^8 + 24\theta^6 + 44\theta^4 + 32\theta^2 + 8\right)}{\theta^4\left(\theta^2 + 1\right)^4}$$

The coefficient of variation $(C.V)$, coefficient of skewness $\left(\sqrt{\beta_1}\right)$, coefficient of kurtosis $\left(\beta_2\right)$ and Index of dispersion (γ) of Shanker distribution are thus obtained as

$$C.V = \frac{\sigma}{\mu_1'} = \frac{\sqrt{\theta^4 + 4\theta^2 + 2}}{\theta^2 + 2}$$

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{2\left(\theta^6 + 6\theta^4 + 6\theta^2 + 2\right)}{\left(\theta^4 + 4\theta^2 + 2\right)^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\left(3\theta^8 + 24\theta^6 + 44\theta^4 + 32\theta^2 + 8\right)}{\left(\theta^4 + 4\theta^2 + 2\right)^2}$$

$$\gamma = \frac{\sigma^2}{\mu_1'} = \frac{\theta^4 + 4\theta^2 + 2}{\theta\left(\theta^2 + 1\right)\left(\theta^2 + 2\right)}$$

It can be easily shown that Shanker distribution is over-dispersed $\left(\mu < \sigma^2\right)$, equi-dispersed $\left(\mu = \sigma^2\right)$ and under-dispersed $\left(\mu > \sigma^2\right)$ for $\theta < (=) > \theta^* = 1.171535555$. It would be recalled that Lindley distribution is over-dispersed $\left(\mu < \sigma^2\right)$, equi-dispersed $\left(\mu = \sigma^2\right)$ and under-dispersed $\left(\mu > \sigma^2\right)$ for $\theta < (=) > \theta^* = 1.170086487$ while exponential distribution is over-dispersed $\left(\mu < \sigma^2\right)$, equi-dispersed $\left(\mu = \sigma^2\right)$ and under-dispersed $\left(\mu > \sigma^2\right)$ for $\theta < (=) > \theta^* = 1$.

3. Generating Functions

The moment generating function $\left(M_X(t)\right)$, characteristic function $\left(\varphi_X(t)\right)$, and cumulant generating function $\left(K_X(t)\right)$ of Shanker distribution (1.3) are given by

$$M_X(t) = \left(1 - \frac{\theta t}{\theta^2 + 1}\right)\left(1 - \frac{t}{\theta}\right)^{-2}, \left|\frac{t}{\theta}\right| \le 1$$

$$\varphi_X(t) = \left(1 - \frac{\theta i t}{\theta^2 + 1}\right)\left(1 - \frac{it}{\theta}\right)^{-2}, i = \sqrt{-1}$$

$$K_X(t) = \log\left(1 - \frac{\theta i t}{\theta^2 + 1}\right) - 2\log\left(1 - \frac{it}{\theta}\right)$$

Using the expansion $\log(1-x) = -\sum_{r=0}^{\infty} \frac{x^r}{r}$, we get

$$K_X(t) = -\sum_{r=0}^{\infty} \frac{\left(\theta i t / \theta^2 + 1\right)^r}{r} + 2\sum_{r=0}^{\infty} \frac{\left(it/\theta\right)^r}{r}$$

$$= 2\sum_{r=0}^{\infty} \frac{1}{\theta^r} \frac{(it)^r}{r} - \sum_{r=0}^{\infty} \left(\frac{\theta}{\theta^2 + 1}\right)^r \frac{(it)^r}{r}$$

$$= 2\sum_{r=0}^{\infty} \frac{(r-1)!}{\theta^r} \frac{(it)^r}{r!} - \sum_{r=0}^{\infty} (r-1)! \frac{\theta^r}{\left(\theta^2 + 1\right)^r} \frac{(it)^r}{r!}$$

Thus the rth cumulant of Shanker distribution is given by

$$K_r = \text{coefficient of } \frac{(it)^r}{r!} \text{ in } K_X(t)$$

$$= (r-1)! \frac{2}{\theta^r} - (r-1)! \frac{\theta^r}{\left(\theta^2 + 1\right)^r}$$

$$= (r-1)! \left[\frac{2}{\theta^r} - \frac{\theta^r}{\left(\theta^2 + 1\right)^r}\right]; r = 1, 2, 3, \ldots$$

This gives

$$\mu_1' = K_1 = \frac{\theta^2 + 2}{\theta\left(\theta^2 + 1\right)}$$

$$\mu_2 = K_2 = \frac{\theta^4 + 4\theta^2 + 2}{\theta^2\left(\theta^2 + 1\right)^2}$$

$$\mu_3 = K_3 = \frac{2\left(\theta^6 + 6\theta^4 + 6\theta^2 + 2\right)}{\theta^3\left(\theta^2 + 1\right)^3}$$

$$\mu_4 = K_4 + 3K_2^2 = \frac{3\left(3\theta^8 + 24\theta^6 + 44\theta^4 + 32\theta^2 + 8\right)}{\theta^4\left(\theta^2 + 1\right)^4}$$

which are the same as obtained earlier.

4. Hazard Rate and Mean Residual Life Functions

Let X be a continuous random variable with p.d.f. $f(x)$ and c.d.f. $F(x)$. The hazard rate function (also known as the failure rate function) and the mean residual life function of X are respectively defined as

$$h(x) = \lim_{\Delta x \to 0} \frac{P\left(X < x + \Delta x \middle| X > x\right)}{\Delta x} = \frac{f(x)}{1 - F(x)} \quad (4.1)$$

and

$$m(x) = E\left[X - x \middle| X > x\right] = \frac{1}{1 - F(x)} \int_x^{\infty} \left[1 - F(t)\right] dt \quad (4.2)$$

The hazard rate function, $h(x)$ and the mean residual life function, $m(x)$ of Lindley distribution are given by

$$h(x) = \frac{\theta^2(1+x)}{(\theta+1) + \theta x} \quad (4.3)$$

And $$m(x) = \frac{\theta + 2 + \theta x}{\theta\left(\theta + 1 + \theta x\right)} \quad (4.4)$$

The corresponding hazard rate function, $h(x)$ and the mean residual life function, $m(x)$ of the Shanker distribution are obtained as

$$h(x)=\frac{\theta^2(\theta+x)}{(\theta^2+1)+\theta x} \tag{4.5}$$

and

$$m(x)=\frac{1}{\left[(\theta^2+1)+\theta x\right]e^{-\theta x}}\int_x^\infty\left[(\theta^2+1)+\theta t\right]e^{-\theta t}dt$$

$$=\frac{\theta^2+\theta x+2}{\theta(\theta^2+\theta x+1)} \tag{4.6}$$

It can be easily verified that $h(0)=\frac{\theta^3}{\theta^2+1}=f(0)$ and $m(0)=\frac{\theta^2+2}{\theta(\theta^2+1)}=\mu_1'$. It is also obvious from the graphs of $h(x)$ and $m(x)$ that $h(x)$ is an increasing function of x, and θ, whereas $m(x)$ is a decreasing function of x, and θ. The hazard rate function and the mean residual life function of the Shanker distribution show its flexibility over exponential and Lindley distributions.

The graphs of the hazard rate function and mean residual life function of Lindley and Shanker distributions are shown in figures 3 and 4.

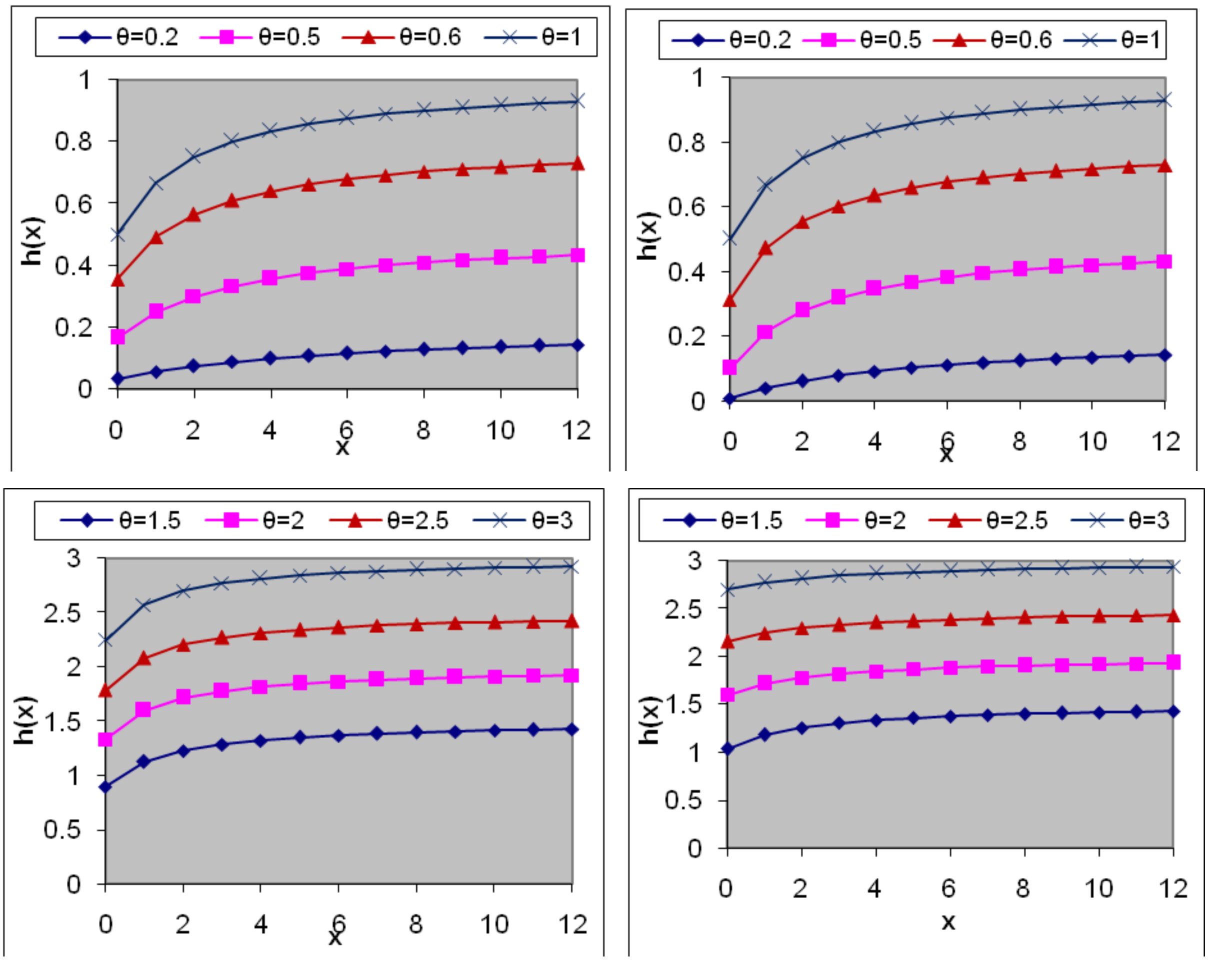

Figure 3. Graph of hazard rate function of Lindley and Shanker distributions for different values of parameter θ. Left hand side graphs are for Lindley distribution and right hand side graphs are for Shanker distribution

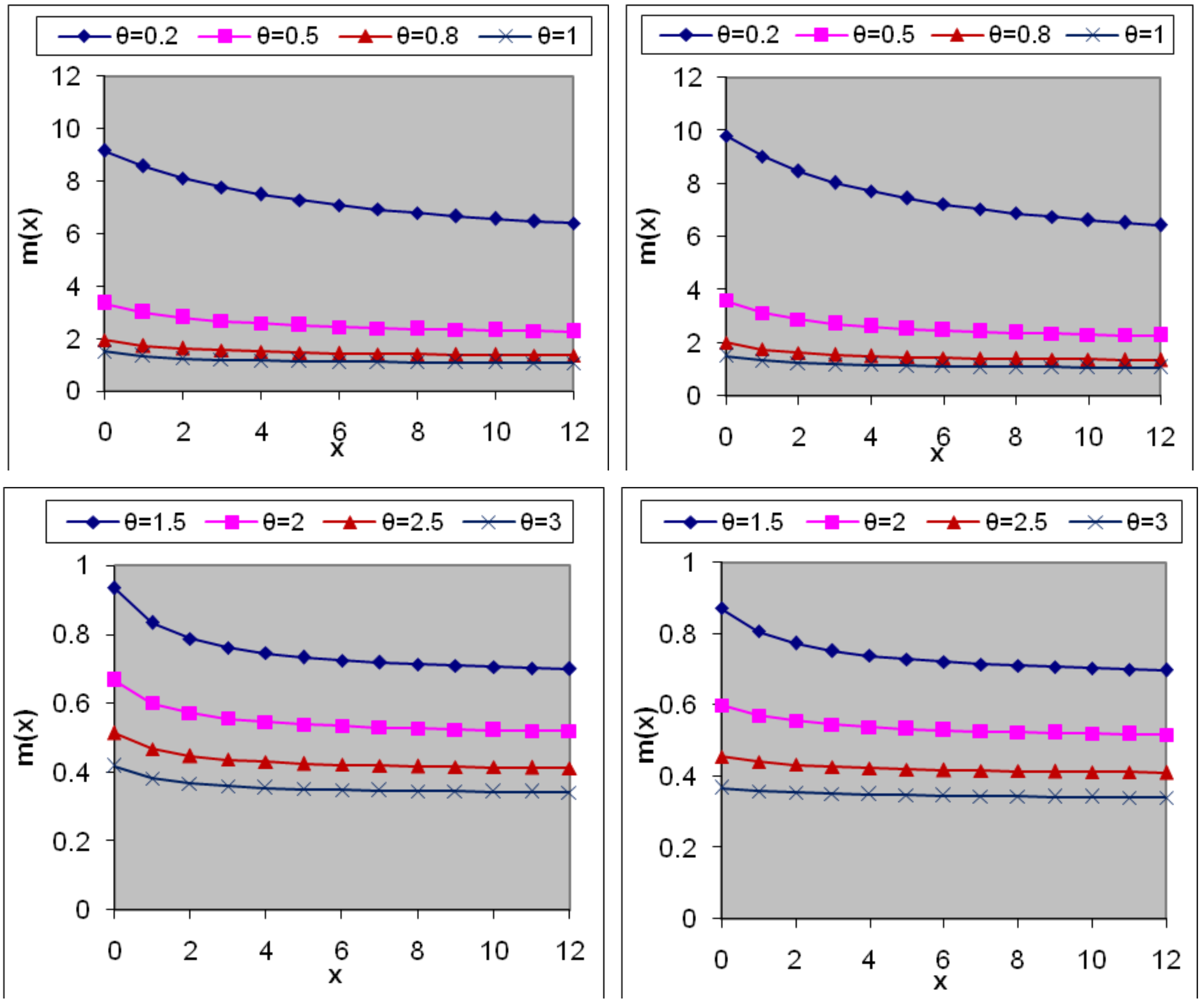

Figure 4. Graphs of mean residual life function of Lindley and Shanker distributions for different values of parameter θ. Left hand side graphs are for Lindley distribution and right hand side graphs are for Shanker distribution

5. Stochastic Orderings

The stochastic ordering of positive continuous random variables is an important tool for judging their comparative behavior. A random variable X is said to be smaller than a random variable Y in the

(i) stochastic order $\left(X \leq_{st} Y\right)$ if $F_X\left(x\right) \geq F_Y\left(x\right)$ for all x

(ii) hazard rate order $\left(X \leq_{hr} Y\right)$ if $h_X\left(x\right) \geq h_Y\left(x\right)$ for all x

(iii) mean residual life order $\left(X \leq_{mrl} Y\right)$ if $m_X\left(x\right) \leq m_Y\left(x\right)$ for all x

(iv) likelihood ratio order $\left(X \leq_{lr} Y\right)$ if $\dfrac{f_X\left(x\right)}{f_Y\left(x\right)}$ decreases in x.

The following stochastic ordering relationship given in Shaked and Shanthikumar (1994) are well known for stochastic ordering of distributions

$$X \leq_{lr} Y \Rightarrow X \leq_{hr} Y \Rightarrow X \leq_{mrl} Y$$
$$\Downarrow$$
$$X \leq_{st} Y$$

The Shanker distribution is ordered with respect to the strongest 'likelihood ratio' ordering which has been shown in the following theorem:

Theorem: Let X ~ Shanker distributon $\left(\theta_1\right)$ and Y ~ Shanker distribution $\left(\theta_2\right)$. If $\theta_1 \geq \theta_2$, then $X \leq_{lr} Y$ and hence $X \leq_{hr} Y$, $X \leq_{mrl} Y$ and $X \leq_{st} Y$.

Proof: We have

$$\frac{f_X(x)}{f_Y(x)} = \frac{\theta_1^{\,2}\left(\theta_2^{\,2}+1\right)}{\theta_2^{\,2}\left(\theta_1^{\,2}+1\right)}\left(\frac{\theta_1+x}{\theta_2+x}\right)e^{-(\theta_1-\theta_2)x}\ ;\ x>0$$

Now

$$\log\frac{f_X(x)}{f_Y(x)} = \log\left[\frac{\theta_1^{\,2}\left(\theta_2^{\,2}+1\right)}{\theta_2^{\,2}\left(\theta_1^{\,2}+1\right)}\right]+\log\left(\frac{\theta_1+x}{\theta_2+x}\right)-\left(\theta_1-\theta_2\right)x$$

and $\frac{d}{dx}\log\frac{f_X(x)}{f_Y(x)}=\frac{-(\theta_1-\theta_2)}{(\theta_1+x)(\theta_2+x)}-(\theta_1-\theta_2)$

Thus for $\theta_1 \ge \theta_2$, $\frac{d}{dx}\log\frac{f_X(x)}{f_Y(x)}<0$. This means that $X \le_{lr} Y$ and hence $X \le_{hr} Y$, $X \le_{mrl} Y$ and $X \le_{st} Y$.

6. Distribution of Order Statistics

Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from Shanker distribution (1.3). Let $X_{(1)} < X_{(2)} < ... < X_{(n)}$ denote the corresponding order statistics. The p.d.f. and the c.d.f. of the k th order statistic, say $Y = X_{(k)}$ are given by

$$f_Y(y)=\frac{n!}{(k-1)!(n-k)!}F^{k-1}(y)\{1-F(y)\}^{n-k}f(y)$$
$$=\frac{n!}{(k-1)!(n-k)!}\sum_{l=0}^{n-k}\binom{n-k}{l}(-1)^l F^{k+l-1}(y)f(y)$$

and

$$F_Y(y)=\sum_{j=k}^{n}\binom{n}{j}F^j(y)\{1-F(y)\}^{n-j}$$
$$=\sum_{j=k}^{n}\sum_{l=0}^{n-j}\binom{n}{j}\binom{n-j}{l}(-1)^l F^{j+l}(y),$$

respectively, for $k = 1, 2, 3, ..., n$.

Thus, the p.d.f. and the c.d.f of the k th order statistics of Shanker distribution are obtained as

$$f_Y(y)=\frac{n!\theta^2(\theta+x)e^{-\theta x}}{(\theta^2+1)(k-1)!(n-k)!}\sum_{l=0}^{n-k}\binom{n-k}{l}(-1)^l$$
$$\times\left[1-\frac{(\theta^2+1)+\theta x}{\theta^2+1}e^{-\theta x}\right]^{k+l-1}$$

and

$$F_Y(y)=\sum_{j=k}^{n}\sum_{l=0}^{n-j}\binom{n}{j}\binom{n-j}{l}(-1)^l\left[1-\frac{(\theta^2+1)+\theta x}{\theta^2+1}e^{-\theta x}\right]^{j+l}$$

7. Renyi Entropy Measure

The entropy of a random variable X is the measure of variation of uncertainty. There are various entropy measures available in statistics literature but a popular entropy measure is Renyi entropy (1961). If X is a continuous random variable having probability density function $f(.)$, then Renyi entropy is defined as

$$T_R(\gamma)=\frac{1}{1-\gamma}\log\left\{\int f^\gamma(x)dx\right\}$$

where $\gamma > 0$ and $\gamma \ne 1$.

Thus, the Renyi entropy for the Shanker distribution (1.3) can be obtained as

$$T_R(\gamma)=\frac{1}{1-\gamma}\log\left[\int_0^\infty\frac{\theta^{2\gamma}}{(\theta^2+1)^\gamma}(\theta+x)^\gamma e^{-\theta\gamma x}dx\right]$$
$$=\frac{1}{1-\gamma}\log\left[\int_0^\infty\frac{\theta^{3\gamma}}{(\theta^2+1)^\gamma}\left(1+\frac{x}{\theta}\right)^\gamma e^{-\theta\gamma x}dx\right]$$
$$=\frac{1}{1-\gamma}\log\left[\int_0^\infty\frac{\theta^{3\gamma}}{(\theta^2+1)^\gamma}\sum_{j=0}^{\infty}\binom{\gamma}{j}\left(\frac{x}{\theta}\right)^j e^{-\theta\gamma x}dx\right]$$
$$=\frac{1}{1-\gamma}\log\left[\sum_{j=0}^{\infty}\binom{\gamma}{j}\frac{\theta^{3\gamma-j}}{(\theta^2+1)^\gamma}\int_0^\infty e^{-\theta\gamma x}x^{j+1-1}dx\right]$$
$$=\frac{1}{1-\gamma}\log\left[\sum_{j=0}^{\infty}\binom{\gamma}{j}\frac{\theta^{3\gamma-j}}{(\theta^2+1)^\gamma}\frac{\Gamma(j+1)}{(\theta\gamma)^{j+1}}\right]$$
$$=\frac{1}{1-\gamma}\log\left[\sum_{j=0}^{\infty}\binom{\gamma}{j}\frac{\theta^{3\gamma-2j-1}}{(\theta^2+1)^\gamma}\frac{\Gamma(j+1)}{(\gamma)^{j+1}}\right]$$

8. Mean Deviations from Mean and Median

The amount of scatter in the population is measured to some extent by the totality of deviations usually from their mean and median and these are known as the mean deviation about the mean and the mean deviation about the median and are defined as

$$\delta_1(X)=\int_0^\infty|x-\mu|f(x)dx \quad\text{and}\quad \delta_2(X)=\int_0^\infty|x-M|f(x)dx,$$

respectively, where $\mu = E(X)$ and $M = \text{Median}(X)$. The measures $\delta_1(X)$ and $\delta_2(X)$ can be computed using the following simplified expressions

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

$$=\mu F(\mu)-\int_0^{\mu}x\,f(x)\,dx-\mu\left[1-F(\mu)\right]+\int_{\mu}^{\infty}x\,f(x)\,dx$$

$$=2\mu F(\mu)-2\mu+2\int_{\mu}^{\infty}x\,f(x)\,dx$$

$$=2\mu F(\mu)-2\int_0^{\mu}x\,f(x)\,dx \qquad (8.1)$$

and

$$\delta_2(X)=\int_0^{M}(M-x)f(x)\,dx+\int_{M}^{\infty}(x-M)f(x)\,dx$$

$$=M\,F(M)-\int_0^{M}x\,f(x)\,dx$$

$$-M\left[1-F(M)\right]+\int_{M}^{\infty}x\,f(x)\,dx$$

$$=-\mu+2\int_{M}^{\infty}x\,f(x)\,dx$$

$$=\mu-2\int_0^{M}x\,f(x)\,dx \qquad (8.2)$$

Using p.d.f. (1.3) and expression for the mean of Shanker distribution, we have

$$\int_0^{\mu}x\,f(x)\,dx=\mu-\frac{\left\{\theta^2\mu^2+\theta\left(\theta^2+2\right)\mu+\left(\theta^2+2\right)\right\}e^{-\theta\mu}}{\theta\left(\theta^2+1\right)} \qquad (8.3)$$

$$\int_0^{M}x\,f(x)\,dx=\mu-\frac{\left\{\theta^2M^2+\theta\left(\theta^2+2\right)M+\left(\theta^2+2\right)\right\}e^{-\theta M}}{\theta\left(\theta^2+1\right)} \qquad (8.4)$$

Using expressions from (8.1), (8.2), (8.3), (8.4) and (1.6), $\delta_1(X)$ and $\delta_2(X)$ of Shanker distribution are obtained as

$$\delta_1(X)=\frac{\left(\theta^2+2\theta\mu+2\right)e^{-\theta\mu}}{\theta\left(\theta^2+1\right)} \qquad (8.5)$$

and

$$\delta_2(X)=\frac{2\left[\theta^2M^2+\theta\left(\theta^2+2\right)M+\left(\theta^2+2\right)\right]e^{-\theta M}}{\theta\left(\theta^2+1\right)}-\mu \qquad (8.6)$$

9. Bonferroni and Lorenz Curves and Indices

The Bonferroni and Lorenz curves (Bonferroni, 1930) and Bonferroni and Gini indices have been used in economice to study income and poverty. Now a days, these curves and indices have many applications in other fields of knowledge including demography, reliability, insurance and medicine and engineering. The Bonferroni and Lorenz curves are defined as

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

$$=\frac{1}{p\mu}\left[\int_0^{\infty}x\,f(x)\,dx-\int_q^{\infty}x\,f(x)\,dx\right] \qquad (9.1)$$

$$=\frac{1}{p\mu}\left[\mu-\int_q^{\infty}x\,f(x)\,dx\right]$$

and

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

$$=\frac{1}{\mu}\left[\int_0^{\infty}x\,f(x)\,dx-\int_q^{\infty}x\,f(x)\,dx\right] \qquad (9.2)$$

$$=\frac{1}{\mu}\left[\mu-\int_q^{\infty}x\,f(x)\,dx\right]$$

respectively or equivalently

$$B(p)=\frac{1}{p\mu}\int_0^{p}F^{-1}(x)\,dx \qquad (9.3)$$

and

$$L(p)=\frac{1}{\mu}\int_0^{p}F^{-1}(x)\,dx \qquad (9.4)$$

respectively, where $\mu=E(X)$ and $q=F^{-1}(p)$.

The Bonferroni and Gini indices are thus defined as

$$B=1-\int_0^{1}B(p)\,dp \qquad (9.5)$$

and

$$G=1-2\int_0^{1}L(p)\,dp \qquad (9.6)$$

respectively.

Using p.d.f. (1.3), we have

$$\int_q^\infty x\,f(x)\,dx = \frac{\left\{\theta^3 q + \theta^2\left(q^2+1\right) + 2\theta q + 2\right\} e^{-\theta q}}{\theta\left(\theta^2+1\right)} \quad (9.7)$$

Now using equation (9.7) in (9.1) and (9.2), we have

$$B(p) = \frac{1}{p}\left[1 - \frac{\left\{\theta^3 q + \theta^2\left(q^2+1\right) + 2\theta q + 2\right\} e^{-\theta q}}{\theta^2+2}\right] \quad (9.8)$$

and $$L(p) = 1 - \frac{\left\{\theta^3 q + \theta^2\left(q^2+1\right) + 2\theta q + 2\right\} e^{-\theta q}}{\theta^2+2} \quad (9.9)$$

Now using equations (9.8) and (9.9) in (9.5) and (9.6), the Bonferroni and Gini indices of Shanker distribution are obtained as

$$B = 1 - \frac{\left\{\theta^3 q + \theta^2\left(q^2+1\right) + 2\theta q + 2\right\} e^{-\theta q}}{\theta^2+2} \quad (9.10)$$

$$G = -1 + \frac{2\left\{\theta^3 q + \theta^2\left(q^2+1\right) + 2\theta q + 2\right\} e^{-\theta q}}{\theta^2+2} \quad (9.11)$$

10. Stress-Strength Reliability

The stress- strength reliability describes the life of a component which has random strength X that is subjected to a random stress Y. When the stress applied to it exceeds the strength $(X < Y)$, the component fails instantly and the component will function satisfactorily till $X > Y$. Therefore, $R = P(Y < X)$ is a measure of component reliability and it is known as stress-strength parameter in statistics literature. It has wide applications in almost all areas of knowledge especially in engineering such as structures, deterioration of rocket motors, static fatigue of ceramic components, aging of concrete pressure vessels etc.

Let X and Y be independent strength and stress random variables having Shanker distribution (1.3) with parameter θ_1 and θ_2 respectively. Then, the stress-strength reliability R is obtained as

$$R = P(Y < X) = \int_0^\infty P(Y < X \mid X = x) f_X(x)\,dx$$

$$= \int_0^\infty f(x;\theta_1)\; F(x;\theta_2)\,dx$$

$$= 1 - \frac{\theta_1^2 \left[\begin{array}{l}\theta_1^3\left(\theta_2^2+1\right) + \theta_1^2\theta_2\left(2\theta_2^2+3\right) \\ +\theta_1\left(\theta_2^4+3\theta_2^2+1\right) + \theta_2\left(\theta_2^2+3\right)\end{array}\right]}{\left(\theta_1^2+1\right)\left(\theta_2^2+1\right)\left(\theta_1+\theta_2\right)^3}.$$

11. Estimation

11.1. Maximum Likelihood Estimate (MLE)

Let $(x_1, x_2, x_3, \ldots, x_n)$ be a random sample of size n from Shanker distribution (1.3). The likelihood function, L of (1.3) is given by

$$L = \left(\frac{\theta^2}{\theta^2+1}\right)^n \prod_{i=1}^n \left(\theta + x_i\right) e^{-n\theta\bar{x}}$$

and so its log likelihood function is thus obtained as

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

Now

$$\frac{d\log L}{d\theta} = \frac{2n}{\theta\left(\theta^2+1\right)} + \sum_{i=1}^n \frac{1}{\left(\theta + x_i\right)} - n\bar{x}$$

where $\bar{x}$ is the sample mean.

The maximum likelihood estimate (MLE), $\hat{\theta}$ of θ is the solution of the equation $\frac{d\log L}{d\theta} = 0$ and so it can be obtained by solving the following non-linear equation

$$\frac{2n}{\theta\left(\theta^2+1\right)} + \sum_{i=1}^n \frac{1}{\theta + x_i} - n\bar{x} = 0 \quad (11.1.1)$$

11.2. Method of Moment Estimate (MOME)

Equating the population mean of the Shanker distribution to the corresponding sample mean, the method of moment estimate (MOME), $\tilde{\theta}$ of θ is the solution of the following cubic equation

$$\bar{x}\,\theta^3 - \theta^2 + \bar{x}\,\theta - 2 = 0 \quad (11.2.1)$$

where $\bar{x}$ is the sample mean.

12. Applications

The Shanker distribution has been fitted to a number of real lifetime data-sets and in almost all data-sets it gives better fit than exponential and Lindley distributions. In this section, we present the fittings of Shanker, Lindley and exponential distributions to three real lifetime data- sets to show the applicability and superiority of Shanker distribution over Lindley and exponential distributions.

Data set 1: The first data - set represents the waiting times (in minutes) before service of 100 Bank customers and examined and analyzed by Ghitany *et al* (2008) for fitting the Lindley (1958) distribution. The data are as follows:

0.8, 0.8, 1.3, 1.5, 1.8, 1.9, 1.9, 2.1, 2.6, 2.7, 2.9, 3.1, 3.2, 3.3, 3.5, 3.6, 4.0, 4.1, 4.2, 4.2, 4.3, 4.3, 4.4, 4.4, 4.6, 4.7, 4.7, 4.8, 4.9, 4.9, 5.0, 5.3, 5.5, 5.7, 5.7, 6.1, 6.2, 6.2, 6.2, 6.3, 6.7, 6.9, 7.1, 7.1, 7.1, 7.1, 7.4, 7.6, 7.7, 8.0, 8.2, 8.6, 8.6, 8.6, 8.8, 8.8, 8.9, 8.9, 9.5, 9.6, 9.7, 9.8, 10.7, 10.9, 11.0, 11.0, 11.1, 11.2, 11.2, 11.5, 11.9, 12.4, 12.5, 12.9, 13.0, 13.1, 13.3, 13.6, 13.7, 13.9, 14.1, 15.4, 15.4, 17.3, 17.3, 18.1, 18.2, 18.4, 18.9, 19.0, 19.9, 20.6, 21.3, 21.4, 21.9, 23.0, 27.0, 31.6, 33.1, 38.5.

Data set 2: The second data - set represents the lifetime's data relating to relief times (in minutes) of 20 patients receiving an analgesic and reported by Gross and Clark (1975, P. 105). The data are as follows:

1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3.0, 1.7, 2.3, 1.6, 2.0

Data set 3: The third data set is the strength data of glass of the aircraft window reported by Fuller *et al* (1994)

18.83, 20.80, 21.657, 23.03, 23.23, 24.05, 24.321, 25.50, 25.52, 25.80, 26.69, 26.77, 26.78, 27.05, 27.67, 29.90, 31.11, 33.20, 33.73, 33.76, 33.89, 34.76, 35.75, 35.91, 36.98, 37.08, 37.09, 39.58, 44.045, 45.29, 45.381

In order to compare exponential, Lindley and Shanker distributions, $-2\ln L$, AIC (Akaike Information Criterion), AICC (Akaike Information Criterion Corrected), BIC (Bayesian Information Criterion), K-S Statistics (Kolmogorov-Smirnov Statistics) for three real lifetime data sets have been computed and presented in table 1. The formulae for computing AIC, AICC, BIC, and K-S Statistics are as follows:

$$AIC = -2\ln L + 2k, \qquad AICC = AIC + \frac{2k(k+1)}{(n-k-1)}, \qquad BIC = -2\ln L + k\ln n \text{ and}$$

$D = \underset{x}{\text{Sup}}\left|F_n(x) - F_0(x)\right|$, where k = the number of parameters, n = the sample size and $F_n(x)$ is the empirical distribution function.

The best distribution which gives much closer fit to lifetime data-sets corresponds to lower $-2\ln L$, AIC, AICC, BIC, and K-S statistics

Table 1. MLE's, - 2ln L, AIC, AICC, BIC, and K-S Statistics of the fitted distributions of data -sets 1, 2, and 3

Data-Sets	Model	Parameter estimate	$-2\ln L$	AIC	AICC	BIC	K-S statistic	p-value
1	Shanker	0.1983	635.3	637.3	637.4	639.9	0.0419	0.991
	Lindley	0.1866	638.1	640.1	640.1	642.7	0.0677	0.707
	Exponential	0.1012	658.0	660.0	660.1	662.6	0.1729	0.004
2	Shanker	0.8039	59.7	61.8	62.0	62.8	0.3151	0.005
	Lindley	0.8161	60.5	62.5	62.7	63.5	0.3410	0.002
	Exponential	0.5263	65.7	67.7	67.9	68.7	0.3895	0.000
3	Shanker	0.0647	252.3	254.3	254.5	255.8	0.3263	0.000
	Lindley	0.0630	254.0	256.0	256.1	257.4	0.3332	0.000
	Exponential	0.0325	274.5	276.7	276.7	277.9	0.4264	0.000

It can be easily verified from above table that the Shanker distribution gives better fitting than the Lindley and exponential distributions for modeling real lifetime data-sets and thus Shanker distribution should be preferred to the Lindley and exponential distributions. Further, in data-set 1 Shanker distribution also gives better fitting than a new generalized Lindley distribution (with $-2\ln L = 635.66$) introduced by Abouammoh *et al* (2015) and a two parameter Lindley distribution (with $-2\ln L = 635.75$) introduced by Shanker and Mishra (2013 a).

13. Concluding Remarks

In the present paper a one parameter lifetime distribution named, "Shanker distribution" has been introduced. Its various mathematical properties including shape, moments and associated measures, hazard rate and mean residual life functions, stochastic ordering, mean deviations, and order statistics have been discussed. Further, expressions for Bonferroni and Lorenz curves, Renyi entropy measure and stress-strength reliability of the proposed distribution have been derived. The conditions of over-dispersion, equi-dispersion, and under-dispersion of Shanker distribution have been discussed along with the Lindley and exponential distributions. The method of moments and the method of maximum likelihood estimation have also been discussed for estimating its parameter. Finally, the goodness of fit test using K-S Statistics (Kolmogorov-Smirnov Statistics) for three real lifetime data- sets have been presented to demonstrate its applicability and superiority over exponential and Lindley distributions.

ACKNOWLEDGEMENTS

The author is grateful to the anonymous reviewer for some helpful comments.

REFERENCES

[1] Abouammoh, A.M., Alshangiti, A.M. and Ragab, I.E. (2015): A new generalized Lindley distribution, *Journal of Statistical Computation and Simulation,* preprint http://dx.doi.org/10.1080/ 00949655.2014.995101.

[2] Alkarni, S. (2015): Extended Power Lindley distribution-A new Statistical model for non-monotone survival data, *European journal of statistics and probability*, 3(3), 19 – 34.

[3] Ashour, S. and Eltehiwy, M. (2014): Exponentiated Power Lindley distribution, *Journal of Advanced Research,* preprint http://dx.doi.org/10.1016/ j.jare. 2014.08.005.

[4] Bakouch, H.S., Al-Zaharani, B. Al-Shomrani, A., Marchi, V. and Louzad, F. (2012): An extended Lindley distribution, *Journal of the Korean Statistical Society*, 41, 75 – 85.

[5] Bonferroni, C.E. (1930): Elementi di Statistca generale, Seeber, Firenze.

[6] Deniz, E. and Ojeda, E. (2011): The discrete Lindley distribution-Properties and Applications, *Journal of Statistical Computation and Simulation*, 81, 1405 – 1416.

[7] Elbatal, I., Merovi, F. and Elgarhy, M. (2013): A new generalized Lindley distribution, *Mathematical theory and Modeling*, 3 (13), 30-47.

[8] Fuller, E.J., Frieman, S., Quinn, J., Quinn, G., and Carter, W. (1994): Fracture mechanics approach to the design of glass aircraft windows: A case study, SPIE Proc 2286, 419-430.

[9] Ghitany, M.E., Atieh, B. and Nadarajah, S. (2008): Lidley distribution and its Application, *Mathematics Computing and Simulation*, 78, 493 – 506.

[10] Ghitany, M., Al-Mutairi, D., Balakrishnan, N. and Al-Enezi, I. (2013): Power Lindley distribution and associated inference, *Computational Statistics and Data Analysis*, 64, 20 – 33.

[11] Gross, A.J. and Clark, V.A. (1975): Survival Distributions: Reliability Applications in the Biometrical Sciences, John Wiley, New York.

[12] Lindley, D.V. (1958): Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society, Series B*, 20, 102- 107.

[13] Liyanage, G.W. and Pararai. M. (2014): A generalized Power Lindley distribution with applications, *Asian journal of Mathematics and Applications*, 1 – 23.

[14] Merovci, F. (2013): Transmuted Lindley distribution, *International Journal of Open Problems in Computer Science and Mathematics*, 6, 63 – 72.

[15] Nadarajah, S., Bakouch, H.S. and Tahmasbi, R. (2011): A generalized Lindley distribution, *Sankhya Series* B, 73, 331 – 359.

[16] Oluyede, B.O. and Yang, T. (2014): A new class of generalized Lindley distribution with applications, *Journal of Statistical Computation and Simulation*, 85 (10), 2072 – 2100.

[17] Pararai, M., Liyanage, G.W. and Oluyede, B.O. (2015): A new class of generalized Power Lindley distribution with applications to lifetime data, *Theoretical Mathematics & Applications,* 5(1), 53 – 96.

[18] Renyi, A. (1961): On measures of entropy and information, in proceedings of the 4th berkeley symposium on Mathematical Statistics and Probability, 1, 547 – 561, Berkeley, university of California press.

[19] Shaked, M. and Shanthikumar, J.G. (1994): Stochastic Orders and Their Applications, Academic Press, New York.

[20] Shanker, R. and Mishra, A. (2013 a): A two-parameter Lindley distribution, *Statistics in Transition-new series*, 14 (1), 45-56.

[21] Shanker, R. and Mishra, A. (2013 b): A quasi Lindley distribution, *African Journal of Mathematics and Computer Science Research*, 6(4), 64 – 71.

[22] Shanker, R., Sharma, S. and Shanker, R. (2013): A two-parameter Lindley distribution for modeling waiting and survival times data, *Applied Mathematics*, 4, 363 – 368.

[23] Shanker, R., Hagos, F, and Sujatha, S. (2015): On modeling of Lifetimes data using exponential and Lindley distributions, *Biometrics & Biostatistics International Journal*, 2 (5), 1-9.

[24] Sharma, V., Singh, S., Singh, U. and Agiwal, V. (2015): The inverse Lindley distribution - A stress-strength reliability model with applications to head and neck cancer data, *Journal of Industrial &Production Engineering*, 32 (3), 162 – 173.

[25] Singh, S.K., Singh, U. and Sharma, V.K. (2014): The Truncated Lindley distribution-inference and Application, *Journal of Statistics Applications & Probability*, 3(2), 219 – 228.

[26] Zakerzadeh, H. and Dolati, A. (2009): Generalized Lindley distribution, *Journal of Mathematical extension*, 3 (2), 13 – 25.