Journal of Data Science 549-566 , DOI: 10.6339/JDS.201807_16(3).0006

A NEW GENERALIZED TWO PARAMETER LINDLEY DISTRIBUTION

* N. Ekhosuehi , + F. Opone and F. Odobaire

Department of Mathematics, University of Benin, Benin City, Nigeria.

*email of authors: *nosakhare.ekhosuehi@uniben.edu; +festus.opone@physci.uniben.edu*

Abstract: A new generalized two-parameter Lindley distribution which offers more flexibility in modeling lifetime data is proposed and some of its mathematical properties such as the density function, cumulative distribution function, survival function, hazard rate function, mean residual life function, moment generating function, quantile function, moments, Renyi entropy and stochastic ordering are obtained. The maximum likelihood estimation method was used in estimating the parameters of the proposed distribution and a simulation study was carried out to examine the performance and accuracy of the maximum likelihood estimators of the parameters. Finally, an application of the proposed distribution to a real lifetime data set is presented and its fit was compared with the fit attained by some existing lifetime distributions.

Keywords: Lindley Distribution, Hazard rate, Renyi entropy, Moments, Stochastic ordering

1. Introduction

Lindley (1958) proposed the Lindley distribution which is a one parameter lifetime distribution in the context of Bayesian statistics, as a counter example of fiducial statistics. A random variable X is said to follow the Lindley distribution with scale parameter $\theta > 0$, if the density function is given by

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

and the corresponding cumulative distribution function is obtained as:

$$F(x,\theta) = 1 - (\tfrac{\theta+1+\theta X}{(\theta+1)})e^{-\theta x} \tag{2}$$

The probability density function (pdf) of the one-parameter Lindley distribution given in (1) is a two-component mixture of Exponential (θ) and Gamma $(2,\theta)$. Thus, equation (1) can be expressed as:

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

where $f_1(x)$ and $f_2(x)$ are the pdf of the Exponential (θ) and Gamma $(2,\theta)$ distribution respectively and $p = \frac{\theta}{\theta+1}$ is a mixing proportion.

The relevance of this distribution as one of the lifetime distributions was overlooked in the literature, as the popularity of the Exponential distribution gain more ground in modeling lifetime data until Ghitany et al. (2008) studied its mathematical properties and showed that some of the mathematical properties of the Lindley distribution exhibits more flexibility than the Exponential distribution in modeling lifetime data. In spite of its flexibility, the Lindley distribution only accounts for increasing failure rate property which has been identified as a major drawback in lifetime analysis. To address this situation, many generalizations of the Lindley distribution have been introduced in literature. Bakouch et al. (2012) introduced the extended Lindley distribution, Ghitanty et al. (2013) introduced the Power Lindley distribution, Shanker et al. (2013) introduced the two-parameter Lindley distribution, MirMostafaee et al. (2015) introduced the Beta Lindley distribution, Bhati et al. (2015) introduced the Lindley-Exponential distribution, Sharma and Khandelwal (2017) introduced an extension of the Inverse Lindley distribution and Korkmaz et al. (2018) introduced the Odd Lindley Burr XII distribution. These generalizations of the Lindley distribution exhibits an increasing and decreasing failure rate properties. However, none of them accounts for a constant or bathtub shape failure rate properties.

In this paper, we introduce a new generalized two-parameter Lindley distribution which spans through increasing, decreasing, constant and bathtub shape failure rate properties. As we shall see in the main work, the proposed distribution has the attractive feature of generalizing both the Lindley distribution and the Exponential distribution. The remaining sections of this paper are organized as follows: In Section 2, we introduce the density function and cumulative distribution function of the proposed distribution, in Section 3, we discussed the sub-models of the proposed distribution. In Section 4, we obtained the survival function, hazard rate function and the mean residual life function of the proposed distribution. Sections 5-9 cover the quantile function, moments and related measures, moment generating function, Renyi entropy and stochastic ordering of the proposed distribution. We estimated the parameters of the distribution using maximum likelihood method and conducted a simulation study to examine the performance and accuracy of the maximum likelihood estimators of the parameters in Section 10. Finally, in Section 11, we fit the proposed distribution to a real lifetime data set and compared its fit with the fit attained by some existing related lifetime distributions.

2. PDF and CDF of the Proposed Distribution (NG2PLD)

A random variable X is said to follow the new generalized two-parameter Lindley distribution, if its density function is defined by

$$f(x,\theta)=\frac{\theta^2}{\theta+1}\left(1+\frac{\theta^{\alpha-2}x^{\alpha-1}}{\Gamma(\alpha)}\right)e^{-\theta x}\,;\; x>0,\alpha,\theta>0 \quad (3)$$

The pdf of the NG2PLD is a two-component mixture of Exponential (θ) and Gamma (α,θ) distributions which can be expressed as, $f(x,\theta)=pf_1(x)+(1-p)f_2(x)$ where, $f_1(x)=\theta e^{-\theta x}, f_2(x)=\frac{\theta^2 x^{\alpha-1}}{\Gamma(\alpha)}e^{-\theta x}$ are the pdf of the Exponential (θ) and Gamma (α,θ) distributions respectively and $p=\frac{\theta}{\theta+1}$ is the mixing proportion. The corresponding cumulative density function of the NG2PLD is obtained as:

$$F(x)=\frac{\theta\Gamma(\alpha)\left(1-e^{-\theta x}\right)+\Gamma(\alpha)-\Gamma(\alpha,\theta x)}{(\theta+1)\Gamma(\alpha)} \quad (3)$$

where, $\Gamma(\alpha)$ and $\Gamma(\alpha,x)$are respectively the complete gamma function and upper incomplete gamma function.

$$\Gamma(\alpha)=\int_0^\infty t^{\alpha-1}e^{-t}dt=\int_0^x t^{\alpha-1}e^{-t}dt+\int_x^\infty t^{\alpha-1}e^{-t}dt$$

The graphical presentation of the density function of NG2PLD for some fixed values of the parameters is shown in Figure 1.

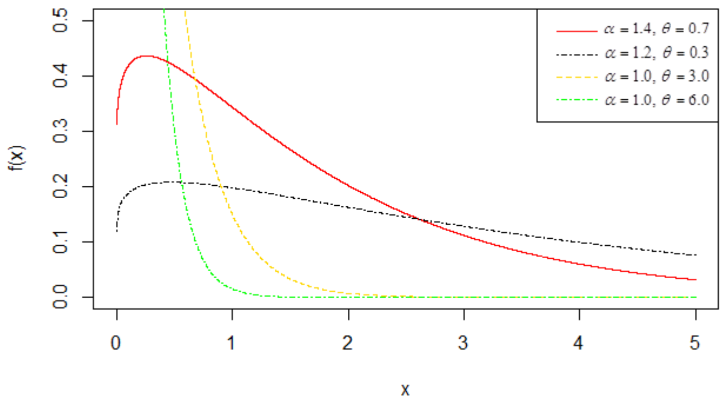

Figure 1: Probability density function of the NG2PLD

3. Sub-models of the NG2PLD

3.1. Exponential Distribution

For $\alpha = 1$, the new generalized two-parameter Lindley distribution reduces to an exponential distribution with pdf and cdf which can be obtained from equation (3) and (4) as:

$$f(x,\theta) = \frac{\theta^2}{\theta+1}\left(1 + \frac{\theta^{-1}x}{\Gamma(1)}\right)e^{-\theta x} \tag{5}$$

$$= \frac{\boldsymbol{\theta}^2}{\boldsymbol{\theta}+\mathbf{1}}\left(\mathbf{1} + \frac{\mathbf{1}}{\boldsymbol{\theta}}\right)\boldsymbol{e}^{-\boldsymbol{\theta x}}$$

$$= \frac{\boldsymbol{\theta}^2}{\boldsymbol{\theta}+\mathbf{1}}\left(\frac{\boldsymbol{\theta}+\mathbf{1}}{\boldsymbol{\theta}}\right)\boldsymbol{e}^{-\boldsymbol{\theta x}}$$

$$\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\theta}) = \boldsymbol{\theta e}^{-\boldsymbol{\theta x}}\,;\;\; \boldsymbol{x} > \mathbf{0}, \boldsymbol{\theta} > \mathbf{0}$$

and

$$\boldsymbol{F}(\boldsymbol{x}) = \frac{\boldsymbol{\theta\Gamma}(\mathbf{1})\left(\mathbf{1} - \boldsymbol{e}^{-\boldsymbol{\theta x}}\right) + \boldsymbol{\Gamma}(\mathbf{1}) - \boldsymbol{\Gamma}(\mathbf{1},\boldsymbol{\theta x})}{(\boldsymbol{\theta}+\mathbf{1})\boldsymbol{\Gamma}(\mathbf{1})}$$

$$\boldsymbol{where}\ \boldsymbol{\Gamma}(\mathbf{1}) = \mathbf{0}! = \mathbf{1}\ \boldsymbol{and}\ \boldsymbol{\Gamma}(\boldsymbol{\alpha},\boldsymbol{\theta t}) = (\boldsymbol{\alpha}-\mathbf{1})!\,\boldsymbol{e}^{-\boldsymbol{\theta t}}\sum_{\boldsymbol{k}=\mathbf{0}}^{\boldsymbol{\alpha}-\mathbf{1}}\frac{(\boldsymbol{\theta t})^{\boldsymbol{k}}}{\boldsymbol{k}!}\,;\ \boldsymbol{\alpha} > \mathbf{0}$$

$$i.e., \Gamma(1,\theta t) = e^{-\theta t}$$

Substituting $\Gamma(1) = 1\ and\ \ \Gamma(1,\theta t) = e^{-\theta t}\ in(3.11)$, we get

$$F(x) = \frac{\theta(1)\left(1-e^{-\theta x}\right)+\left(1-e^{-\theta x}\right)}{(\theta+1)(1)} \tag{6}$$

$$= \frac{(\boldsymbol{\theta}+\mathbf{1})(\mathbf{1}-\boldsymbol{e}^{-\boldsymbol{\theta x}})}{(\boldsymbol{\theta}+\mathbf{1})}$$

$$\boldsymbol{F}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{e}^{-\boldsymbol{\theta t}}$$

Equations (5) and (6) are respectively the density function and cumulative distribution function of the exponential distribution.

3.2. Lindley Distribution

For $\alpha = 2$, the NG2PLD reduces to the Lindley distribution with pdf and cdf similarly obtained from (3) and (4) as:

$$f(x,\theta) = \frac{\theta^2}{\theta+1}(1+\frac{x}{\Gamma(2)})e^{-\theta x} \quad (7)$$

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

And

$$\boldsymbol{F(X)} = \frac{\boldsymbol{\theta\Gamma(2)\left(1-e^{-\theta x}\right)+\Gamma(2)-\Gamma(2,\theta x)}}{\boldsymbol{(\theta+1)\Gamma(2)}}$$

but

$$\Gamma(2,\theta t) = (2-1)!\,e^{-\theta x}(1+\theta x) = e^{-\theta x}+\theta x e^{-\theta x} \quad (8)$$

$$\boldsymbol{F(x)} = \frac{\boldsymbol{\theta\left(1-e^{-\theta x}\right)+1-e^{-\theta x}-\theta x e^{-\theta x}}}{\boldsymbol{(\theta+1)}}$$

$$= \frac{\boldsymbol{\theta-\theta e^{-\theta x}+1-e^{-\theta x}-\theta x e^{-\theta x}}}{\boldsymbol{\theta+1}}$$

$$= \mathbf{1}-\left(\frac{\boldsymbol{\theta+1+\theta x}}{\boldsymbol{(\theta+1)}}\right)\boldsymbol{e^{-\theta x}}$$

Similarly, equations (7) and (8) show the pdf and cdf of the Lindley distribution respectively.

4. Survival, Hazard and Mean Residual Life Function

Let X be a continuous random variable with density function f(x) and cumulative distribution function F(x). The survival (reliability) function, hazard rate (failure rate) function and mean residual life function of the new generalized two-parameter Lindley distribution are defined by:

$$S(x) = \frac{\theta\Gamma(\alpha)e^{-\theta x}+\Gamma(\alpha,\theta x)}{(\theta+1)\Gamma(\alpha)} \quad (9)$$

$$h(x) = \frac{\theta^2(\Gamma(\alpha)+\theta^{\alpha-2}x^{\alpha-1})e^{-\theta x}}{\theta\Gamma(\alpha)e^{-\theta x}+\Gamma(\alpha,\theta x)} \quad (10)$$

and

$$m(x) = \frac{1}{[\theta+\sum_{k=0}^{\alpha-1}\frac{(\theta x)^k}{k!}]e^{-\theta x}}\left[e^{-\theta x}+\sum_{k=0}^{\alpha-1}\frac{\Gamma(k+1,\theta x)}{(\theta k!)}\right];\ \alpha = 1.2, \ldots \quad (11)$$

The graph of the hazard rate function of the NG2PLD for different values of α and θ is given in Figure 2.

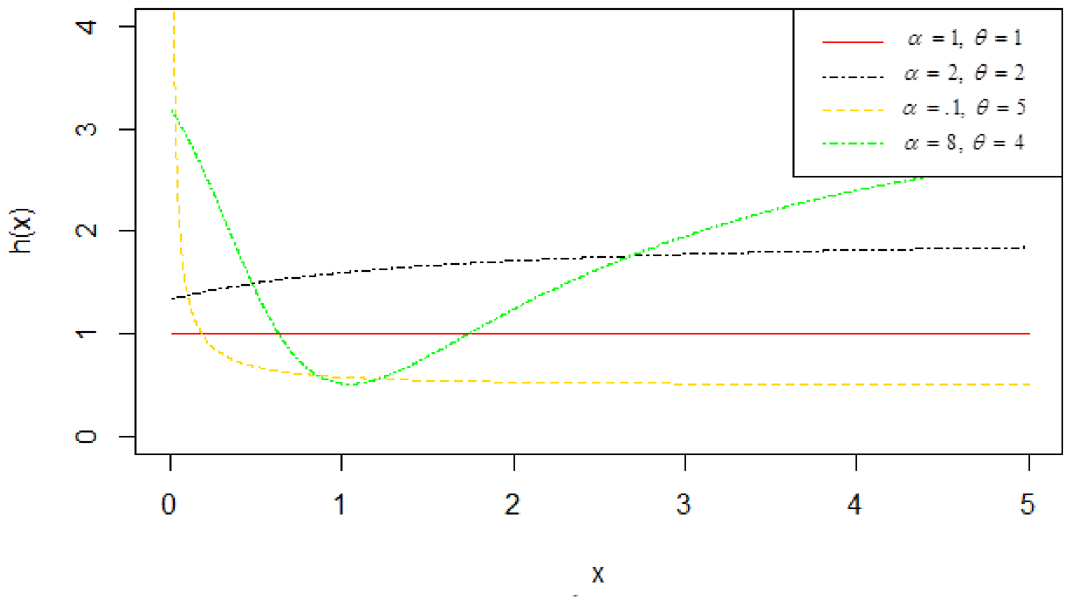

Figure 2: Hazard rate function of the NG2PLD

Remarks

From Figure 2, it is clearly observed that the hazard rate function of the NG2PLD at different values of α and θ can be constant, increasing, decreasing and bathtub-shaped, which offer more flexibility than the Lindley distribution and Exponential distribution.

5. Quantile Function of the NG2PLD

The quantile function of the NG2PLD cannot be expressed in a closed form, thus an explicit form of expressing the quantile function of a probability distribution using numerical methods suggested in the work Oluyede et al. (2016) is used. Hence, the quantile function of the NG2PLD can be derived by setting $F(x) = \mu$ to obtain a system of nonlinear equation which is given by:

$$\theta\Gamma(\alpha) - \theta\Gamma(\alpha)e^{-\theta x} + \Gamma(\alpha) - \Gamma(\alpha, \theta x) - \mu\Gamma(\alpha)(\theta + 1) = 0 \qquad (12)$$

where F(x) is the distribution function defined in equation (4) and $0 < \mu < 1$.

Subsequently, random samples from the new generalized two-parameter Lindley distribution can be generated using equation (12) for some fixed values of the parameters as shown in Table 1.

Table 1: Quantiles of the NG2PLD

U	$(\theta = 0.4, \alpha = 0.9)$	$(\theta = 0.6, \alpha = 1)$	$(\boldsymbol{\theta = 0.2, \alpha = 1.5})$	$(\boldsymbol{\theta = 0.1, \alpha = 0.2})$
0.1	0.210369	0.040292	1.204967	0.000105
0.2	0.469554	0.114301	2.192296	0.003361
0.3	0.773133	0.212332	3.194379	0.025493
0.4	1.130547	0.335459	4.272821	0.107270
0.5	1.559200	0.489266	5.483975	0.327678
0.6	2.089515	0.685022	6.905770	0.823208
0.7	2.779340	0.945100	8.674226	1.834076
0.8	3.759286	1.320798	11.085989	3.859637
0.9	5.447360	1.977307	15.073533	8.431106

6. Moments and Related Measures

Theorem 1: let X ~ NG2PLD, then the rth moment about the origin of X is given by

$$\mu'_r = \frac{1}{\theta^r(\theta+1)}\left[\theta\Gamma(r+1) + \frac{\Gamma(r+\alpha)}{\Gamma(\alpha)}\right]; \ r = 1,23, \ldots \tag{13}$$

Proof: refer to appendix

The first four raw moments about the origin of the NG2PLD are obtained using equation (13) as:

$$\boldsymbol{\mu'_1 = \frac{\theta+\alpha}{\theta(\theta+1)}; \quad \mu'_2 = \frac{\alpha^2+\alpha+2\theta}{\theta^2(\theta+1)}; \quad \mu'_3 = \frac{\alpha^3+3\alpha^2+\alpha+6\theta}{\theta^3(\theta+1)}}$$

$$\boldsymbol{\mu'_4 = \frac{\alpha^4+6\alpha^3+11\alpha^2+6\alpha+24\theta}{\theta^4(\theta+1)}}$$

Similarly, the k^{th} central moments of a given random variable X, can be defined as:

$$\mu_k = E\{(X-\mu)^k\} = E\{\textstyle\sum_{r=0}^{k}(-1)^r \binom{k}{r} X^{k-r}\mu^r\} \tag{14}$$

$$= \left\{\sum_{r=0}^{k}(-\mathbf{1})^r \binom{\boldsymbol{k}}{\boldsymbol{r}} \boldsymbol{E}\left(\boldsymbol{X}^{\boldsymbol{k-r}}\right)\boldsymbol{\mu}^r\right\}$$

$$= \sum_{r=0}^{k}(-\mathbf{1})^r \binom{\boldsymbol{k}}{\boldsymbol{r}} \boldsymbol{\mu_{k-r}\mu^r}\}$$

Where $\mu'_1 = \mu \ and \ \mu_0 = 1$.

Using equation (14), the $2^{nd}, 3^{rd} \ and \ 4^{th}$ central moments can be obtained as:

$$\boldsymbol{\mu_2 = \mu'_2 - \mu^2; \ \ \mu_3 = \mu'_3 - 3\mu'_2\mu + 2\mu^3; \quad \mu_4 = \mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 3\mu^4}$$

where the first raw moment and the second central moment corresponds to the mean and variance respectively.

Thus, the variance, coefficient of variation, measure of skewness and kurtosis of the NG2PLD are given by:

$$\boldsymbol{Variance(\sigma^2) = \mu_2' - \mu^2 = \frac{\theta^2 + \theta\alpha^2 - \theta\alpha + 2\theta + \alpha}{\theta^2(\theta+1)^2}}$$

$$\boldsymbol{Coefficient\ of\ variation(\gamma) = \frac{(\mu_2' - \mu^2)^{1/2}}{\mu} = \frac{\sqrt{\theta^2 + \theta\alpha^2 - \theta\alpha + 2\theta + \alpha}}{\theta + \alpha}}$$

$$\boldsymbol{Measure\ of\ Skewness(\sqrt{\beta_1}) = \frac{\mu_3}{\mu_2^{3/2}}}$$

$$\boldsymbol{= \frac{2\theta^3 + \theta^2(\alpha^3 - \alpha + 6) - \theta(\alpha^3 - 6\alpha^2 + 5\alpha - 6) + 2\alpha}{(\theta^2 + \theta\alpha^2 - \theta\alpha + 2\theta + \alpha)^2}}$$

$$\boldsymbol{Measure\ of\ kurtosis(\beta_2) = \frac{\mu_4'}{\mu_2^2} = \frac{9\theta^4 + \theta^3 A - \theta^2 B + \theta C + 3\alpha^2 + 6\alpha}{(\theta^2 + \theta\alpha^2 - \theta\alpha + 2\theta + \alpha)^2}}$$

where $A = \alpha^4 + 2\alpha^3 + 5\alpha^2 - 8\alpha + 36$, $B = \alpha^4 - 10\alpha^3 - 13\alpha^2 + 16\alpha - 48$ and $C = \alpha^4 - 4\alpha^3 + 29\alpha^2 - 14\alpha + 24$

Table 2: Summary statistics of simulation study on moments of the NG2PLD

μ_r'	$(\theta = 0.4, \alpha = 0.9)$	$(\theta = 1, \alpha = 0.6)$	$(\theta = 0.2, \alpha = 1.5)$	$(\theta = 0.1, \alpha = 0.2)$
$\mu_1' = \mu$	3.04221	0.78220	6.50824	2.61234
μ_2'	18.83344	1.36407	68.31628	29.68378
μ_3'	166.9087	3.19028	937.9325	475.5799
μ_4'	1792.359	8.61943	15495.3	9441.437
$\sigma^2 = \mu_2$	9.57838	0.75223	25.95915	22.85943
σ	3.09490	0.86731	5.09501	4.78115
γ	1.01732	1.10881	0.78286	1.83022
$\sqrt{\beta_1}$	1.73169	1.45079	1.75007	2.54910
β_2	5.99628	4.45734	4.53772	10.6164

Table 2 shows the simulation results of the first four moments, variance (σ^2), standard deviation (σ), coefficient of variation (γ), coefficient of skewness $(\sqrt{\beta_1})$ and coefficient of kurtosis (β_2) of the new generalized two-parameter Lindley distribution. A random data of size 100 was generated from the NG2PLD using equation (12) for different parameter values.

7. Moment Generating Function

The moment generating function of a random variable X following the NG2PLD is defined by:

$$M_X(t) = \int_0^\infty e^{tx} f(x)dx \tag{15}$$

$$= \int_0^\infty e^{tx}[\frac{\theta^2}{\theta+1}(1+\frac{\theta^{\alpha-2}x^{\alpha-1}}{\Gamma(\alpha)})e^{-\theta x}]dx$$

$$= \frac{\theta^2}{\theta+1}\int_0^\infty (1+\frac{\theta^{\alpha-2}x^{\alpha-1}}{\Gamma(\alpha)})\, e^{-(\theta-t)x}dx$$

$$= \frac{\theta^2}{\theta+1}[\int_0^\infty e^{-(\theta-t)x}dx + \frac{\theta^{\alpha-2}}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}\, e^{-(\theta-t)x}dx$$

using $y = (\theta - t)x$;

$$\frac{\theta^{\alpha-2}}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}\, e^{-(\theta-t)x}dx = \frac{\theta^{\alpha-2}}{\Gamma(\alpha)}\int_0^\infty \frac{y^{\alpha-1}e^{-y}}{(\theta-t)^\alpha}dy = \frac{y^{\alpha-2}}{(\theta-t)^\alpha} \tag{16}$$

$$M_X(t) = \frac{\theta^2}{\theta+1}[\frac{1}{\theta-t} + \frac{\theta^{\alpha-2}}{(\theta-t)^\alpha}]$$

8. Renyi Entropy

According to Renyi (1961), entropy of a random variable X is a measure of variation of uncertainty. Renyi entropy is defined by

$$\tau_R(\gamma) = \frac{1}{1-\gamma}\log\left[\int f^\gamma(x)dx\right],\ \ \gamma > 0,\ \ \gamma \neq 1 \tag{17}$$

Thus, the Renyi entropy of NG2PLD is given by:

$$\boldsymbol{\tau_R(\gamma) = \frac{1}{1-\gamma}log[\int_0^\infty \frac{\theta^{2\gamma}}{(\theta+1)^\gamma}(1+\frac{\theta^{\alpha-2}x^{\alpha-1}}{\Gamma(\alpha)})^\gamma e^{-\theta\gamma x}dx]}$$

$$\boldsymbol{= \frac{1}{1-\gamma}log[\frac{\theta^{2\gamma}}{(\theta+1)^\gamma}\int_0^\infty (1+\frac{\theta^{\alpha-2}x^{\alpha-1}}{\Gamma(\alpha)})^\gamma\, e^{-\theta\gamma x}dx]}$$

but, $(1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k$

$$\tau_R(\gamma) = \frac{1}{1-\gamma}log[\frac{\theta^{2\gamma}}{(\theta+1)^\gamma}\int_0^\infty \sum_{k=0}^\gamma \binom{\gamma}{k}(\frac{\theta^{\alpha-2}x^{\alpha-1}}{\Gamma(\alpha)})^k e^{-\theta\gamma x}dx]$$

$$\frac{1}{1-\gamma}\log[\frac{\theta^{2\gamma}}{(\theta+1)^{\gamma}}\sum_{k=0}^{\gamma}\binom{\gamma}{k}\frac{\theta^{(\alpha-2)^{k}}}{(\Gamma(\alpha))^{k}}\int_{0}^{\infty} x^{(\alpha-1)k}\, e^{-\theta\gamma x}dx]$$

but $\int_0^\infty x^{(\alpha-1)k}\, e^{-\theta\gamma x}dx = \frac{\Gamma(\alpha k-k+1)}{(\theta\gamma)^{\alpha k-k+1}}, hence$

$$=\frac{1}{1-\gamma}\log[\frac{\theta^{2\gamma}}{(\theta+1)^{\gamma}}\sum_{k=0}^{\gamma}\binom{\gamma}{k}\frac{\theta^{(\alpha-2)^{k}}}{(\Gamma(\alpha))^{k}}\frac{\Gamma(\alpha k-k+1)}{(\theta\gamma)^{\alpha k-k+1}}]$$

$$\tau_R(\gamma)=\frac{1}{1-\gamma}log[\frac{\theta^{2\gamma}}{(\theta+1)^{\gamma}}\sum_{k=0}^{\gamma}\binom{\gamma}{k}\frac{\theta^{-(k+1)}}{(\Gamma(\alpha))^{k}}\frac{\Gamma(\alpha k-k+1)}{\gamma^{\alpha k-k+1}}] \tag{18}$$

We can clearly observe that at $\alpha = 2$, the Renyi entropy of NG2PLD given by equation (18) is reduced to that of Lindley distribution defined by:

$$\tau_R(\gamma)=\frac{1}{1-\gamma}log[\frac{\theta^{2\gamma}}{(\theta+1)^{\gamma}}\sum_{k=0}^{\gamma}\binom{\gamma}{k}\frac{\Gamma(k+1)}{\gamma^{k+1}\theta^{k+1}}] \tag{19}$$

9. Stochastic Ordering

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

(i) Stochastic Order $(X \leq_{st} Y)$, if $F_Y(x)$ for all x

(ii) Hazard Rate Order $(X \leq_{hr} Y)$, if $h_X(x) \geq h_Y(x)$ for all x

(iii) Mean Residual Life Order $(X \leq_{mrl} Y)$, if $m_X(x) \leq m_Y(x)$ for all x

(iv) Likelihood Ratio Order $(X \leq_{lr} Y)$, if $\frac{f_X(x)}{f_Y(x)}$ is decreasing in x

Shaked and Shanthikumar (1994), suggests the following result for establishing stochastic ordering of distributions.

$$\begin{array}{c} X \leq_{lr} Y \Rightarrow X \leq_{hr} Y \Rightarrow X \leq_{mrl} Y \\ \Downarrow \\ X \leq_{st} Y \end{array}$$

The theorem below shows that the new generalized two-parameter Lindley distribution is ordered with respect to the strongest "likelihood ratio" ordering

Theorem 2:Let X ~ NG2PLD (α_1, θ_1) and Y ~ NG2PLD (α_2, θ_2). If $\alpha_1 = \alpha_2$ and $\theta_1 \geq \theta_2$ (or if $\theta_1 = \theta_2\ and\ \alpha_1 \leq \alpha_2$) then $X \leq_{lr} Y$ and hence $X \leq_{hr} Y, X \leq_{mrl} Y$ and $X \leq_{st} Y$.

Proof: *refer to appendix*

10. Parameter Estimation

10.1 Maximum likelihood estimates

Let $x_1, x_2, \dots, x_n$ be a random sample of size n from the NG2PLD, then the log-likelihood function is defined by:

$$\ell\,(x,\alpha,\theta) = \sum_{i=1}^{n} \log[\frac{\theta^2 e^{-\theta x_i}}{\theta+1} + \frac{\theta^\alpha x_i^{\alpha-1} e^{-\theta x_i}}{(\theta+1)\Gamma(\alpha)}] \tag{20}$$

$$= \sum_{i=1}^{\boldsymbol{n}} [\boldsymbol{log\theta^2} + \boldsymbol{log}(\boldsymbol{e^{-\theta x_i}}) - \boldsymbol{log}(\boldsymbol{\theta}+\mathbf{1}) + \boldsymbol{\alpha log\theta} + (\boldsymbol{\alpha}-\mathbf{1})\boldsymbol{log x_i} + \boldsymbol{log}(\boldsymbol{e^{-\theta x_i}})$$

$$-\boldsymbol{log}(\boldsymbol{\theta}+\mathbf{1}) - \boldsymbol{log\Gamma}(\boldsymbol{\alpha})]$$

$$= \sum_{i=1}^{\boldsymbol{n}} [\boldsymbol{log\theta^2} - \boldsymbol{\theta x_i} - \boldsymbol{log}(\boldsymbol{\theta}+\mathbf{1}) + \boldsymbol{\alpha log\theta} + (\boldsymbol{\alpha}-\mathbf{1})\boldsymbol{log\, x_i} - \boldsymbol{\theta x_i} - \boldsymbol{log}(\boldsymbol{\theta}+\mathbf{1})$$

$$-\boldsymbol{log\Gamma}(\boldsymbol{\alpha})$$

$$\ell\,(x,\alpha,\theta) = 2nlog\theta - 2\theta\sum_{i=1}^{n} x_i - 2nlog(\theta+1) + n\alpha log\theta \tag{21}$$

$$+(\boldsymbol{\alpha}-\mathbf{1})\sum_{i=1}^{\boldsymbol{n}} \boldsymbol{log\, x_i} - \boldsymbol{nlog}\boldsymbol{\Gamma}(\boldsymbol{\alpha})$$

The associated score function is obtained by equating the first derivative of equation (21) to zero.

$$\frac{\partial \boldsymbol{\ell}}{\partial \boldsymbol{\theta}} = \frac{\mathbf{2}\boldsymbol{n}}{\boldsymbol{\theta}} - \mathbf{2}\sum_{i=1}^{\boldsymbol{n}} \boldsymbol{x_i} - \frac{\mathbf{2}\boldsymbol{n}}{\boldsymbol{\theta}+\mathbf{1}} + \frac{\boldsymbol{n\alpha}}{\boldsymbol{\theta}} = \mathbf{0}$$

$$\frac{\partial \boldsymbol{\ell}}{\partial \boldsymbol{\alpha}} = \boldsymbol{nlog\theta} + \sum_{i=1}^{\boldsymbol{n}} \boldsymbol{log x_i} - \boldsymbol{n\Psi}(\boldsymbol{\alpha}) = \mathbf{0}$$

with $\Psi(\alpha) = \frac{\partial log\Gamma(\alpha)}{\partial \alpha}$, and $\Psi(\alpha)$ i s the digamma function. The maximum likelihood estimators $\hat{\theta}$ and $\hat{\alpha}$ can be achieved using the Newton Raphson's iterative method which is given by:

$$\hat{\emptyset} = \emptyset_k - H^{-1}(\emptyset_k)U(\emptyset_k);\quad \hat{\emptyset} = (\hat{\alpha}, \hat{\theta})^T \tag{22}$$

where,

$$H(\emptyset_k) = \frac{\partial^2\ \ell}{\partial\emptyset\partial\emptyset}[\begin{matrix} \frac{\partial^2\ \ell}{(\partial\theta)^2} & \frac{\partial^2\ \ell}{\partial\theta\partial\alpha} \\ \frac{\partial^2\ \ell}{\partial\alpha\partial\theta} & \frac{\partial^2\ \ell}{(\partial\alpha)^2} \end{matrix}]$$

$$\frac{\partial^2 \boldsymbol{\ell}}{(\partial \boldsymbol{\alpha})^2} = \frac{-\mathbf{2}\boldsymbol{n}}{\boldsymbol{\theta^2}} + \frac{\mathbf{2}\boldsymbol{n}}{(\boldsymbol{\theta}+\mathbf{1})^2} - \frac{\boldsymbol{n\alpha}}{\boldsymbol{\theta^2}}$$

$$\frac{\partial^2 \ell}{\partial\theta\partial\alpha} = \frac{\partial^2 \ell}{\partial\alpha\partial\theta} = \frac{n}{\theta}$$

$\frac{\partial^2 \ell}{(\partial\alpha)^2} = -n\,\Psi(\alpha)$; $\Psi'(\alpha) = \frac{\partial\Psi(\alpha)}{\partial\alpha}$ denote the tri-gamma function.

10.2 Interval estimate

The asymptotic confidence intervals (CIs) for the parameters of NG2PLD $(\boldsymbol{\alpha}, \boldsymbol{\theta})$ is obtained according to the asymptotic distribution of the maximum likelihood estimates of the parameters.

Let $\widehat{\emptyset} = (\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}})$ be MLE of $\emptyset$, the estimators are approximately bi-variate normal with mean $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ and the Fisher information matrix is given by:

$$\mathrm{I}(\emptyset_k) = -\mathrm{E}(\mathrm{H}(\emptyset_k)) \qquad (23)$$

The approximate $(\mathbf{1} - \boldsymbol{\delta})$ 100 CIs for the parameters θ and α, respectively, are

$$\widehat{\boldsymbol{\alpha}} \pm \boldsymbol{Z}_{\boldsymbol{\delta}/2}\sqrt{\boldsymbol{var}(\widehat{\boldsymbol{\alpha}})}\ \boldsymbol{and}\ \widehat{\boldsymbol{\theta}} \pm \boldsymbol{Z}_{\boldsymbol{\delta}/2}\sqrt{\boldsymbol{var}(\widehat{\boldsymbol{\theta}})}$$

where $var(\hat{\alpha})\ and\ var(\hat{\theta})$ are the variance of α and θ which are given by the first and second diagonal element of the variance-covariance matrix $I^{-1}(\emptyset_k)\ and\ Z_{\frac{\delta}{2}}$ is the upper $(\frac{\delta}{2})$ percentile of the standard normal distribution. It is important to note that for a given set of data, the matrix given by equation (23) is obtained after the convergence of the Newton-Raphson procedure.

10.3 Simulation Study

In this section, we investigate the performance and accuracy of the maximum likelihood estimates of the new generalized two-parameter Lindley distribution parameters through a simulation study for different parameter values as well as different sample sizes. Random data from the NG2PLD are generated using equation (12). The Monte Carlo simulation study is repeated 1000 times for different sample sizes n = 30,50,75,100,200 and parameter values (θ = 1, α = 0.6), (θ = 0.4, α = 0.9) and (θ = 0.2, α = 1.5). It is important to note that the choice of values for the parameter estimate is such that the new distribution deviates from the Lindley distribution and exponential distribution. Since the proposed distribution reduces to an exponential distribution when α = 1 and Lindley distribution when α = 2 An algorithm for the simulation study is given by the following steps.

1. Choose the value N (i.e. number of Monte Carlo simulation);
2. choose the values $\emptyset_0 = (\theta_0, \alpha_0)$ corresponding to the parameters of the NG2PLD (θ, α);

3. generate a sample of size n from NG2PLD;
4. compute the maximum likelihood estimates $\widehat{\emptyset}_0$ of $\emptyset_0$
5. repeat steps (3-4), N-times;
6. compute the average bias $= \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\emptyset}_i - \emptyset_0\right)$ and the mean square error $(MSE) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\emptyset}_i - \emptyset_0)^2$

Table 3: Monte Carlo Simulation Results for Average Bias and Mean Square Error of the MLE

Parameter	*n*	*Average Bias* (θ)	*Mean Square* Error(θ)	*Average Bias* (α)	*Mean Square* Error(α)
θ = 1, α = 0.6	30	-0.025437	0.027332	0.028064	0.007918
	50	-0.010501	0.018949	0.012783	0.005141
	75	-0.009189	0.011683	0.010397	0.003380
	100	-0.006388	0.007936	0.009098	0.002172
	200	-0.004868	0.003621	0.006296	0.001134
θ = 0.4, α = 0.9	30	-0.021087	0.009461	0.050699	0.017222
	50	-0.012415	0.004974	0.023227	0.009496
	75	-0.008595	0.003950	0.017636	0.006962
	100	-0.006974	0.003033	0.015692	0.004501
	200	-0.002584	0.001498	0.006307	0.002189
θ = 0.2, α = 1.5	30	-0.012186	0.003276	0.071504	0.037534
	50	-0.010743	0.002601	0.051776	0.026602
	75	-0.002626	0.001382	0.024609	0.012611
	100	-0.001278	0.001481	0.016418	0.012151
	200	-0.000859	0.000565	0.003967	0.004611

From Table 3, we clearly observe that the values of the average bias and the mean square error of the parameter estimates decreases as the sample size n increases.

11. Application of the NG2PLD

In this section, we fit the NG2PLD to a real lifetime data set and compare its fit with some existing lifetime distribution such as the Exponential distribution with density function $f(x,\theta) = \theta e^{-\theta k}$ Lindley distribution with density function $\mathrm{f}(\mathrm{x},\theta) = \frac{\theta^2}{\theta+\alpha}(1+\mathrm{x})\mathrm{e}^{-\theta\mathrm{x}}$ and Two-Parameter Lindley with density function $f(x,\alpha,\theta) = \frac{\theta^2}{\theta+\alpha}(1+\alpha x)e^{-\theta x}$. The estimates of the parameters of the distribution, $-2log(L)$, Akaike Information Criterion $[AIC = 2k - 2log(L)]$ and Kolmogorov-Smirnov Statistic $(K - S)$, were considered for the comparison. Where n is the number of observations, k is the number of estimated parameters and L is the value of the likelihood function evaluated at the parameter estimates.

Data Set: The data set represents the remission times (in months) of a random sample of 128 bladder cancer patients reported in Lee and Wang (2003). The data set is presented as follows:

0.08	2.09	3.48	4.87	6.94	8.66	13.11	23.63	0.20	2.23	3.52	4.98	6.97
9.02	13.29	0.40	2.26	3.57	5.06	7.09	9.22	13.8	25.74	0.50	2.46	3.64
5.09	7.26	9.47	14.24	25.82	0.51	2.54	3.70	5.17	7.28	9.74	14.76	26.31
0.81	2.62	3.82	5.32	7.32	10.06	14.77	32.15	2.64	3.88	5.32	7.39	10.34
14.83	34.26	0.90	2.69	4.18	5.34	7.59	10.66	15.96	36.66	1.05	2.69	4.23
5.41	7.62	10.75	16.62	43.01	1.19	2.75	4.26	5.41	7.63	17.12	46.12	1.26
2.83	4.33	5.49	7.66	11.25	17.14	79.05	1.35	2.87	5.62	7.87	11.64	17.36
1.40	3.02	4.34	5.71	7.93	11.79	18.1	1.46	4.40	5.85	8.26	11.98	19.13
1.76	3.25	4.50	6.25	8.37	12.02	2.02	3.31	4.51	6.54	8.53	12.03	20.28
2.02	3.36	6.76	12.07	21.73	2.07	3.36	6.93	8.65	12.63	22.69	-	-

Table 4: Comparison Criterion for the Data Set

Models	Estimates	-2logL	AIC	K-S
NG2PLD	$\theta = 0.2942$ $\alpha = 0.8303$	826.7324	830.7325	0.0684
LINDLEY	$\theta = 0.1960$	839.0858	841.0858	0.1165
TPLD	$\theta = 0.1067$ $\alpha = -0.00002$	828.7072	832.7073	0.0847
EXPONENTIAL	$\theta = 0.1068$	828.7072	830.7073	0.0846

The fits of the density and cumulative distribution fit and the probability-probability (P-P) plots of each distribution for the data set are given in the figures 3 and 4 respectively.

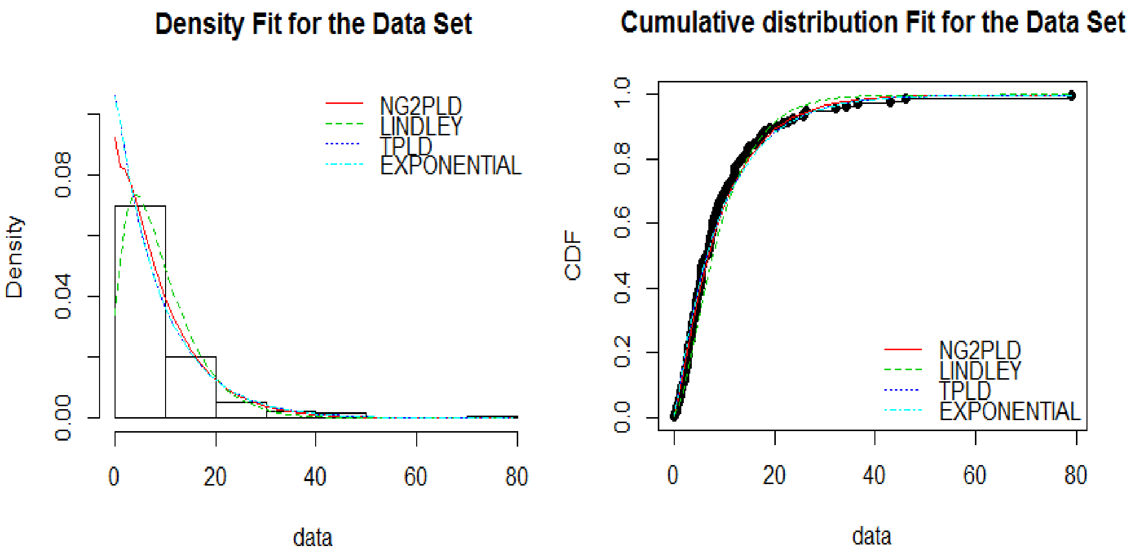

Figure 3: Density and Cumulative Distribution fit for the Data Set

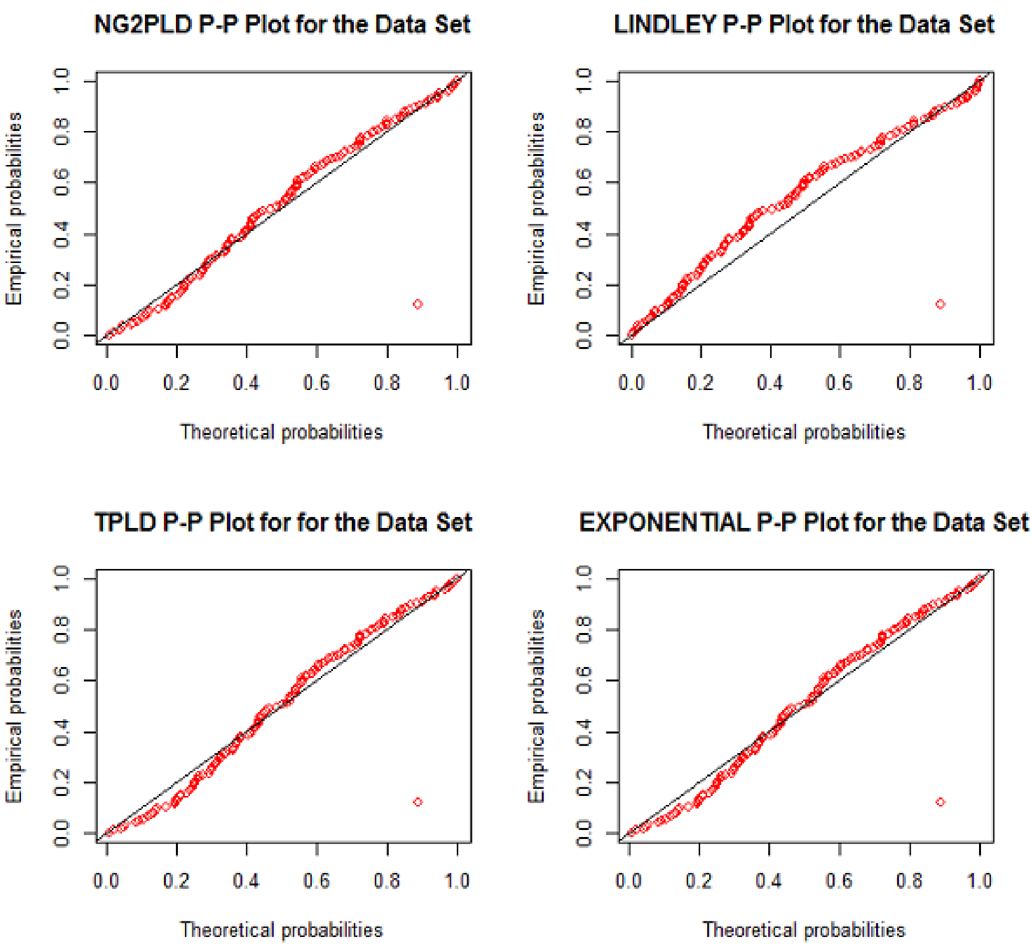

Figure 4: Fitted probability-probability plot for the data set

The best fitted distribution is considered by investigating the distributions with the minimum value of -2logL, AIC and K-S Statistic. Table 4 indicates that new generalized two-parameter Lindley distribution gives the best fit and thus demonstrates superiority over the examined lifetime distributions in modeling the lifetime data sets under study. This conclusion was further supported by inspecting the probability-probability plots, the density and cumulative distribution fit of the distributions for the real lifetime data sets.

Concluding Remark

In this paper, a new generalized two-parameter Lindley distribution is proposed and the mathematical properties such as the shape of the density, hazard rate function, mean residual life function, quantile function, moment generating function, moments, Renyi entropy and stochastic ordering have been discussed. The maximum likelihood estimation method for estimating its parameters have been discussed as well. A simulation study also indicates a good performance and accuracy of the maximum likelihood estimators of the parameters. The application of the proposed distribution to a real lifetime data set reveals its superiority over the Lindley distribution, Two-Parameter Lindley distribution and Exponential distribution.

References

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

[2] Bhati, D., Malik, M. A. and Vaman, H. J. (2015). Lindley-Exponential distribution: Properties and applications. METRON. 73(3): 335-357

[3] 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.

[4] Ghitany, M., Atieh, B. and Nadarajah, S.,(2008). Lindley distribution and its applications. Mathematics and Computers in Simulation. 78: 493-506.

[5] Korkmaz, C. M., H. M. Yousof, M. Rasekhi and G. G. Hamedani (2018). The Odd Lindley Bur XII Model. Bayesian Analysis, Classical Inference and Characterizations. Journal of Data Science. 16(2): 327-354.

[6] Lee, E. T. and Wang, J. (2003). Statistical Methods for Survival Data Analysis, New York, Wiley.

[7] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society. 20(1): 102-107.

[8] MirMostafaee, S. M. T. K, Mahdizadeh and S. Nadarajah (2015). The Beta Lindley distribution. Journal of Data Science. 13(3): 603-626.

[9] Oluyede, B.O., S. Foya, G. Warahena-Liyanage, S. Huang (2016). The Log-Logistic Weibull Distribution with Applications to Lifetime Data. Austrian Journal of Statistics. 45: 43-69.

[10] Rényi, A,(1961). On measure of entropy and information. Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability 1, University of California Press, Berkeley. 547-561.

[11] Shanked, M., Shanthikumar, J.G. (1994). Stochastic Order and Their Applications, Academic Press, New York.

[12] Shanker, R., Sharma, S., Shanker and Shanker, R.(2013). A Two-Parameter Lindley distribution for modeling waiting and survival times data. Applied Mathematics, 4 : 363-368.

[13] Sharma, V. K. and P. Khandelwal (2017). On the extension of the Inverse Lindley distribution. Journal of Data Science. 15(2): 205-220.

Appendix

Theorem 1: let X ~ NG2PLD, then the rth moment about the origin of X is given by

$$\mu_r' = \frac{1}{\theta^r(\theta+1)}\left[\theta\Gamma(r+1) + \frac{\Gamma(r+\alpha)}{\Gamma(\alpha)}\right]; \; r = 1,2,3,\ldots$$

Proof:

Let X be a random variable following the NG2PLD with parameter α and θ, then

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r f(x)dx$$

$$= \int_0^{\infty} x^r \left[\frac{\theta^2}{\theta+1}(1 + \frac{\theta^{\alpha-2}x^{\alpha-1}}{\Gamma(\alpha)})e^{-\theta x}\right]dx$$

$$= \frac{\theta^2}{\theta+1}\left[\int_{-\infty}^{\infty} x^r e^{-\theta x} dx + \frac{\theta^{\alpha-2}}{\Gamma(\alpha)}\int_0^{\infty} x^{r+\alpha-1} e^{-\theta x} dx\right]$$

using the transformation $y = \theta x$;

$$\mu_r' = \frac{\theta^2}{\theta+1}\left[\frac{1}{\theta^{r+1}}\int_0^{\infty} y^r e^{-y} dy + \frac{\theta^{\alpha-2}}{\theta^{r+\alpha}\Gamma(\alpha)}\int_0^{\infty} y^{r+\alpha-1} e^{-y} dy\right]$$

since, $\int_0^{\infty} y^{r+\alpha-1} e^{-y} dy = \Gamma(r+\alpha)$

$$\mu_r' = \frac{\theta^2}{\theta+1}\left[\frac{\Gamma(r+1)}{\theta^{r+1}} + \frac{\Gamma(r+\alpha)}{\theta^{r+2}\Gamma(\alpha)}\right]$$

$$\mu_r' = \frac{1}{\theta^r(\theta+1)}\left[\theta\Gamma(r+1) + \frac{\Gamma(r+\alpha)}{\Gamma(\alpha)}\right];$$

Theorem2: Let X~NG2PLD (α_1, θ_1) and Y~NG2PLD $(\alpha_2, \theta_2$ .If $\alpha_1 = \alpha_2$ and $\theta_1 \geq \theta_2$ (or if $\theta_1 = \theta_2$ and $\alpha_1 \leq \alpha_2$) then $X \leq_{lr} Y$ and hence $X \leq_{hr} Y, X \leq_{mrl} Y$ and $X \leq_{st} Y$.

Proof:

We have,

$$\frac{f_X(x)}{f_Y(x)} = \frac{\frac{\theta_1^2}{\theta_1+1}\theta(1+\frac{\theta^{\alpha_1-2}x^{\alpha_1-1}}{\Gamma(\alpha_1)})e^{-\theta_1 x}}{\frac{\theta_2^2}{\theta_2+1}(1+\frac{\theta_2^{\alpha_2-2}x^{\alpha_2-1}}{\Gamma(\alpha_2)})e^{-\theta_2 x}}$$

$$= \frac{(\theta_1^2\Gamma(\alpha_1) + \theta_1^{\alpha_1}x^{\alpha_1-1})(\theta_2+1)\Gamma(\alpha_2)}{(\theta_2^2\Gamma(\alpha_2) + \theta_1^{\alpha_2}x^{\alpha_2-1})(\theta_1+1)\Gamma(\alpha_1)}\, e^{(\theta_2-\theta_1)x}$$

$$log\frac{f_X(x)}{f_Y(x)} = log\left(\theta_1^2\Gamma(\alpha_1) + \theta_1^{\alpha_1}x^{\alpha_1-1}\right) + log(\theta_2+1) + log\Gamma(\alpha_2)$$

$$-\log\left(\theta_2^2\Gamma(\alpha_2) + \theta_2^{\alpha_2}x^{\alpha_2-1}\right) - \log(\theta_1+1) - \log\Gamma(\alpha_1) + (\theta_2-\theta_1)x$$

Case 1:

if $\alpha_1 = \alpha_2$ and $\theta_1 \geq \theta_2$ then $\frac{d}{dx}log\frac{f_X(x)}{f_Y(x)} < 0$. This means that $X \leq_{lr} Y$ and hence $X \leq_{hr} Y, X \leq_{mrl} Y$ and $X \leq_{st} Y$.

Case 2:

if $\theta_1 = \theta_2$ and $\alpha_1 \leq \alpha_2$ then $\frac{d}{dx}log\frac{f_X(x)}{f_Y(x)} < 0$. This means that $X \leq_{lr} Y$ and hence $X \leq_{hr} Y, X \leq_{mrl} Y$ and $X \leq_{st} Y$.