Journal of the Nigerian Association of Mathematical Physics
***Volume 35,* (*May, 2016*), pp 451 – 460**
© J. of NAMP

A Study of Lindley Distribution and its Associated Distributions

N. Ekhosuehi

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

Abstract

In this paper, we compare some properties of the Lindley distribution with two of its associated distributions; the Exponential and a special Gamma distribution which have some relationship with the Lindley distribution. The properties considered are the Survival function, Hazard function, moments and maximum likelihood estimation. The Lindley distribution is found to have more flexibility than the other two distributions in terms of certain Statistic such as the coefficients of variation, skewness and Kurtosis. We used real lifetime data which comprises of the length of time it takes a University graduate to get a job to make comparison among these three distributions. The estimates of the parameter of these models were obtained using the maximum likelihood. The Kolmogorov-Smirnov test Statistic, P-P plot and the Q-Q plot show that the three models fit the data well but the Lindley distribution has the best fit since it has the smallest value in terms of the Kolmogorov-Smirnov test Statistic for the observed sample.

Key words: Lindley distribution; Survival function; hazard function; moments; maximum Likelihood; Kolmogorov test Statistic

1.0 Introduction

The Lindley distribution was originally introduced such that the distribution was used to illustrate the difference between fudicial distribution and posterior distribution [1]. The distribution which is a mixture of Exponential distribution and Gamma distribution with scale parameter 2 and location parameter θ has attracted a wide range of applicability in the literature. Variants of this distribution have been proposed. Some of these families of Lindley distributions are: Discrete Poisson Lindley distribution for modelling count data [2]. The properties of the Lindley distribution and its usefulness [3]. A two-parameter weighted Lindley distribution which is useful for modelling mortality data [4]. A generalised Lindley distribution that is superiority over the popular Gamma, Weibull and Lognormal models [5]. The Power Lindley distribution model [6]. A Quasi Lindley distribution which can be applied in the social sciences [7]. Estimation of the reliability of a stress-strength system from power Lindley [8]. The Lindley-Exponential distribution model with applications to Biological data [9]. Extended Power Lindley distribution with Statistical application for non-monotone survival data [10]. A new class of generalized power Lindley distribution with application to different set of lifetime data [11].

However, in this paper, we compare some properties of the Lindley distribution with the Exponential distribution and the Gamma ($2,\theta$) where the Lindley distribution is a mixture of both exponential and gamma ($2,\theta$) distributions. We also demonstrate this using a real lifetime data which comprises of the length of time it takes a graduate in a particular locality of Delta State of Nigeria to get a job.

2.0 Exponential, Gamma and Lindley Distributions

The exponential density function having the parameter θ can be defined by

$$g_1(t,\theta)=\theta e^{-\theta t}, t>0, \theta>0 \tag{1}$$

The Gamma distribution with scale parameter α and location parameter θ, can be defined by:

$$g_\alpha(t,\theta)=\frac{\theta^\alpha t^{\alpha-1}}{\Gamma(\alpha)}e^{-\theta t}, t>0 \tag{2}$$

Corresponding author: N. Ekhosuehi, E-mail: nosakhare.ekhosuehi@uniben.edu, Tel.: +2347037819078

The Gamma distribution has a special case of exponential distribution when α =1.
For proper a survey on equations (1) and (2), see Lawless [12].
The probability density function of a random variable T having the Lindley distribution [1] is given by:

$$f(t,\theta)=\frac{\theta^2}{\theta+1}(1+t)e^{-\theta t}, t>0, \theta>0 \tag{3}$$

As demonstrated by [7], if $f(t,\theta)$ is written as:

$$f(t,\theta)=p_1 g_1(t,\theta)+p_2 g_2(t,\theta), \quad t>0, \theta>0$$

where $g_1(t,\theta)=\theta e^{-\theta t}$, $g_2(t,\theta)=\theta^2 t e^{-\theta t}$ and $p_1=\theta/(\theta+1),\ p_2=1/(\theta+1)$ are mixing proportions, hence, $f(t,\theta)$ is a mixture of Exponential (θ) distribution and Gamma (2, θ) distribution where Gamma (2, θ) is a special case of α = 2, in the general case of the Gamma distribution defined in equation (2).

2.1 Failure rate and Hazard Rate Functions

The probability distribution of the time-to-failure of an item can be characterized by the failure rate [13]. It is defined by the function.

$$h(t)=\frac{f(t)}{S(t)} \tag{4a}$$

where $h(t)$ is also called the Hazard rate [14]. The function, $f(t)$ is called a probability density function (pdf) and $S(t)$ is the probability of failure after time t or the survival function or simply the reliability. Typically, $S(t)$ has the probabilistic expression given by:

$$S(t)=pr(T\geq t).$$

$$=\int_t^\infty f(x)dx$$

Hence, $S(t)=1-F(t)$ where $F(t)$ is the cumulative distribution function (CDF) of $f(t)$
The CDF of the pdf can be obtained using the following:

$$F(t) = \int_0^t f(x)dx \tag{4b}$$

In many applications of lifetime distributions, $h(t)$ is monotone. If $h(t)$ increases monotonically, the distribution is said to have increasing failure rate (IFR) and if $h(t)$ decreases monotonically, it is said to have decreasing failure rate (DFR). The IFR property is characterized by a device that can consistently deteriorate with age, whereas DFR property is characterized by device that improves with age. Many physical phenomena exhibit failure rate that are non-monotonic such as the study of 'human life', whose failure rate is termed Bathtub (BT) shaped [13].

Using equation (4b), the CDF, Survival function and Hazard function of the exponential, Gamma and Lindley distributions are obtained and tabulated in Table 1 respectively. The shape of the density functions, survival function and the Hazard functions for some fixed parameter θ are given in Figures 1(a, b, c), 2(a, b, c) and 3(a, b, c) respectively.

Table 1: The CDF, $F(t)$, the Survival function, $S(t)$ and Hazard function, $h(t)$ of the exponential, Gamma(2, θ) and Lindley distribution.

f(t)	F(t)	S(t)	h(t)
Exponential	$1 - e^{-\theta t}$	$e^{-\theta t}$	θ
Gamma(2, θ)	$1 - [1 + \theta t]e^{-\theta t}$	$[1 + \theta t]e^{-\theta t}$	$\dfrac{\theta^2 t}{1 + \theta t}$
Lindley	$1 - \dfrac{\theta + 1 + \theta t}{\theta + 1} e^{-\theta t}$	$\dfrac{\theta + 1 + \theta t}{\theta + 1} e^{-\theta t}$	$\dfrac{\theta^2 (1 + t)}{\theta + 1 + \theta t}$

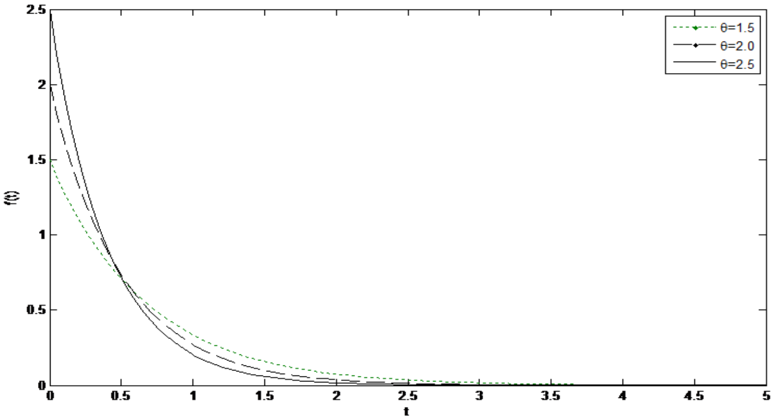


Fig. 1a: Density function of the Exponential distribution for varying parameter θ = 1.5, 2.0 and 2.5

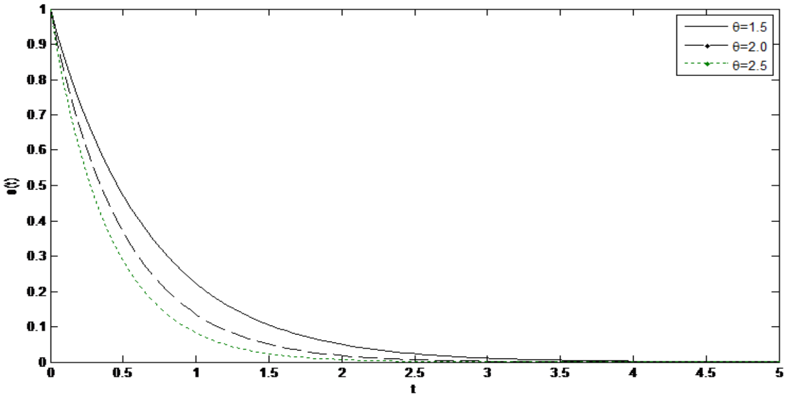


Fig. 1b: Survival function of the Exponential distribution for varying parameter θ = 1.5, 2.0 and 2.5

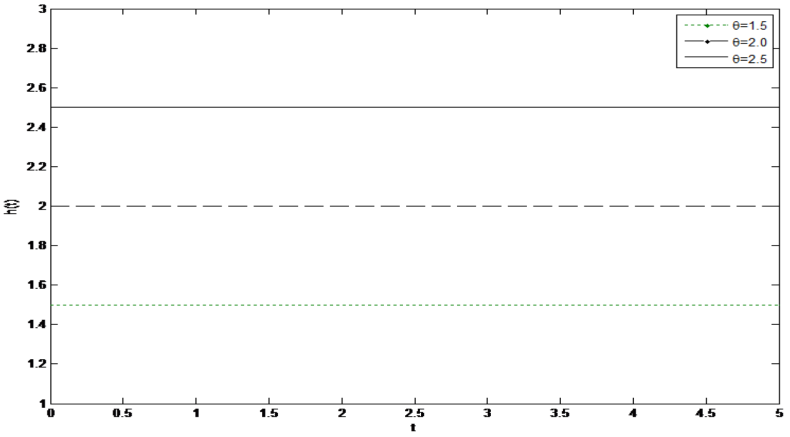


Fig. 1c: Hazard function of the Exponential distribution for varying parameter θ = 1.5, 2.0 and 2.5

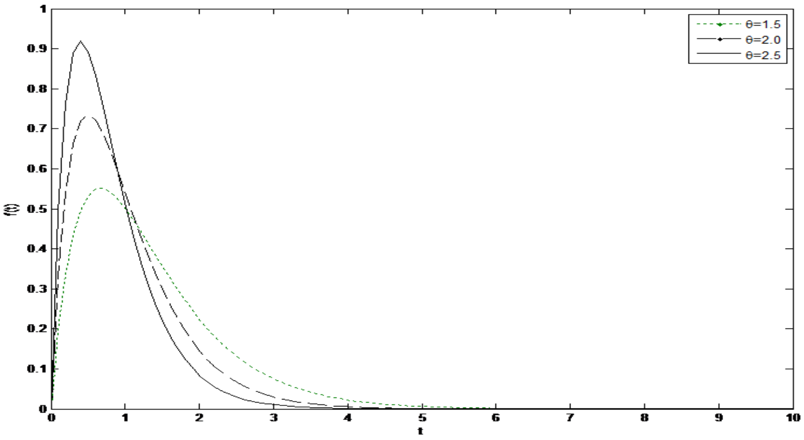


Fig. 2a: Density function of Gamma(2, θ) distribution for varying parameter θ = 1.5, 2.0 and 2.5

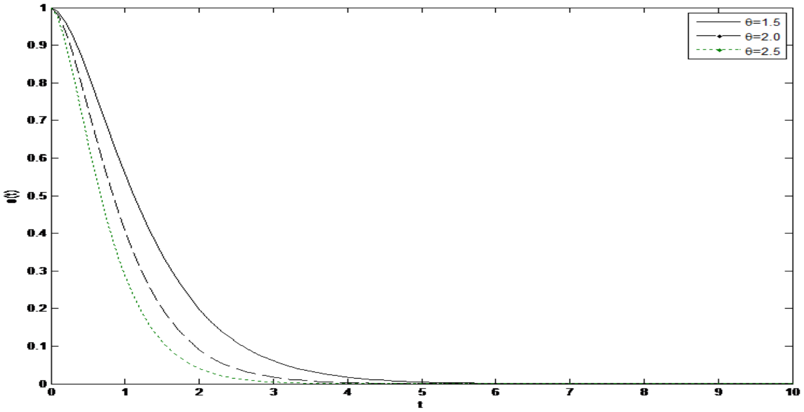


Fig. 2b: Survival function of gamma(2, θ) distribution for varying parameter θ = 1.5, 2.0 and 2.5

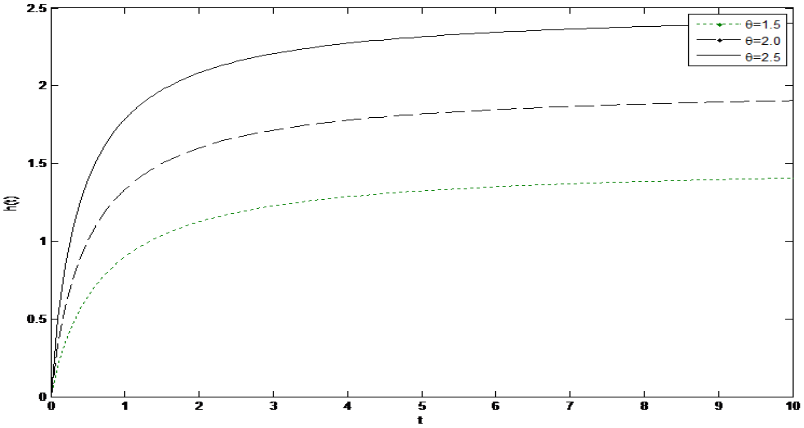


Fig. 2c: Hazard function of Gamma(2, θ) distribution for varying parameter θ = 1.5, 2.0 and 2.5

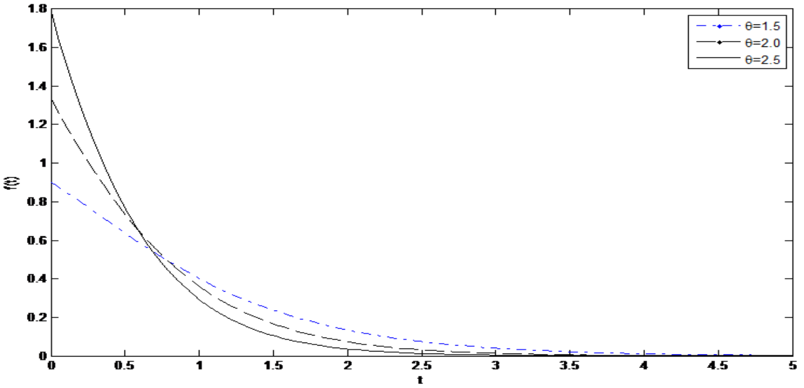


Fig. 3a: Lindley Probability density distribution for varying parameter θ = 1.5, 2.0 and 2.5

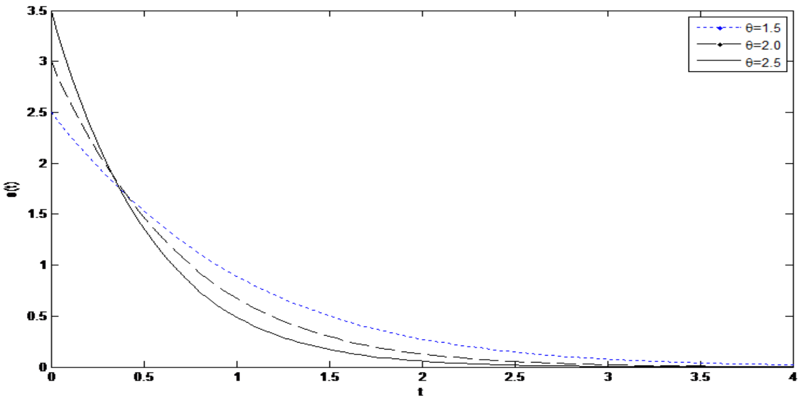


Fig. 3b: Survival function of the Lindley distribution for varying parameter θ = 1.5, 2.0 and 2.5

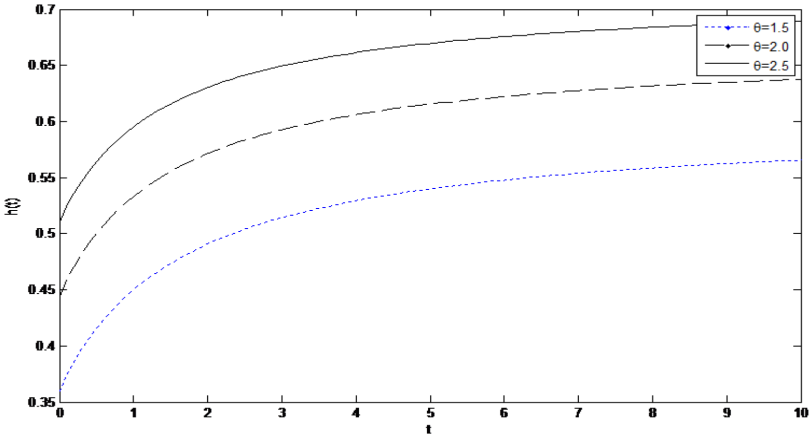


Fig. 3c: Hazard function of the Lindley distribution for varying parameter θ = 1.5, 2.0 and 2.5

2.2 Moments of the Exponential, Gamma(2, θ) and Lindley Distributions

The *r*th moment about the origin of a random variable, say T of a given probability density function, $f(t;\theta)$ is defined by:

$$\mu_r' = E(T^r) = \int_{-\infty}^{\infty} t^r f(t,\theta)dt \tag{5}$$

Similarly, the kth central moments of a given probability density function can be defined as:

$$\mu_k = E\{(T-\mu_1')^k\} = E\left\{\sum_{r=0}^{k}(-1)^r\binom{k}{r}T^{k-r}\mu^r\right\}$$

since $\mu_1' = \mu$ is the mean of the distribution.

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

$$= \sum_{r=0}^{k} (-1)^r \binom{k}{r} \mu'_{k-r} \mu^r \qquad (6)$$

The 2nd, 3rd and 4th central moments can be deduced from equation (6) where the 2nd central moment correspond to the variance while the 3rd and 4th central moments can be used for computing the coefficient of skewness and coefficient of kurtosis respectively. Using equation (5), the *r*th moment about the origin of the Exponential, Gamma(2, θ) and Lindley distributions are obtained in equations (7), (8) and (9) respectively:

$$\mu'_r = \frac{\Gamma(r+1)}{\theta^r}, \quad r = 1,2,\cdots \qquad (7)$$

$$\mu'_r = \frac{\Gamma(r+2)}{\theta^r}, \quad r = 1,2,\cdots \qquad (8)$$

$$\mu'_r = \frac{r!(\theta + r + 1)}{\theta^r (\theta + 1)}, \quad r = 1,2,\cdots \qquad (9)$$

Similarly, the coefficient of variation (CV), coefficient of Skewness (S_k) and coefficient of Kurtosis (K_s) are defined as follows:

$$CV = \frac{(\mu_2)^{1/2}}{\mu}$$

$$S_k = \frac{\mu_3}{(\mu_2)^{3/2}}$$

$$K_s = \frac{\mu_4}{(\mu_2)^2}$$

Thus the first four moments ($\mu'_r; r = 1,2,3,4$), the variance (μ_2), CV, S_k and K_s of the three distributions are obtained and tabulated in Table 2. The results obtained in Table 2, reveals the flexibility of the Lindley distribution over the exponential and Gamma(2, θ) distributions whose CV, S_k and K_s are constant values while the Lindley are all functions of the parameter which can take on any positive value.

Table 2: Summary Statistics of the first four moments with the variance, coefficient of Skewness and coefficient of Kurtosis of the Exponential, Gamma(2, θ) and Lindley distributions

μ'_r	Exponential(θ)	Gamma(2, θ)	Lindley(θ)

$\mu_1^{/}$	$\frac{1}{\theta}$	$\frac{2}{\theta}$	$\frac{\theta+2}{\theta(\theta+1)}$
$\mu_2^{/}$	$\frac{2}{\theta^2}$	$\frac{6}{\theta^2}$	$\frac{2(\theta+3)}{\theta^2(\theta+1)}$
$\mu_3^{/}$	$\frac{6}{\theta^3}$	$\frac{24}{\theta^3}$	$\frac{6(\theta+4)}{}$
$\mu_4^{/}$	$\frac{24}{\theta^4}$	$\frac{120}{\theta^4}$	$\frac{24(\theta+5)}{}$
μ_2	$\frac{1}{\theta^2}$	$\frac{2}{\theta^2}$	$\frac{\theta^2+4\theta+2}{\theta^2(\theta+1)^2}$
CV	1	$\frac{\sqrt{2}}{2}$	$\frac{\theta^2+4\theta+2}{}$
S_k	2	$\frac{2}{\sqrt{2}}$	$\frac{2(\theta^3+6\theta^2+6\theta+2)}{(\theta^2+4\theta+2)^{3/2}}$
K_s	6	9	$\frac{3(3\theta^4+24\theta^3+44\theta^2+32\theta+8)}{(\theta^2+4\theta+2)^2}$

2.3 Method of Moments and Maximum Likelihood

Given a random sample $T=\{t_1,t_2,\cdots,t_n\}$ of size n from a density function $f(t;\theta)$, then an estimator $\hat{\theta}$ of the parameter θ can be obtained by equating the sample mean with the raw mean given in equation (5). That is

$$\bar{t}=\frac{\sum_{i=1}^{n} t_i}{n}=\mu \tag{10}$$

If equation (10) is used to estimate the parameter θ where the mean is a function of θ, then, this method is called method of moment. Similarly, the log-likelihood function of a density function is defined by

$$\ell(T;\theta)=\sum_{i=1}^{n}\log f(t_i;\theta)$$

The maximum likelihood estimate (MLE) of θ is found by minimizing the log-likehood function $\ell(T,\theta)$ and solving for θ when the score function ($\frac{\partial \ell(T;\theta)}{\partial \theta}$) is set equal to zero, that is:

$$\frac{\partial \ell(T;\theta)}{\partial \theta} = 0 \tag{11}$$

Using method of moments given in equation (10) and MLE given in equation (11) to estimate the parameter of the three distributions under study gives the estimated $\hat{\theta}$. Using the method of moment and maximum likelihood for estimating the parameter of the Lindley distribution has been showed to give the same result [3]. These similar results also holds for the Exponential and Gamma(2, θ) distributions. The results obtained for the three distributions understudy, using the method of moments and MLE are given in Table 3 while the estimated parameter values are given in Table 4.

Table 3: Summary results of the method of moment and MLE for Exponential, Gamma(2, θ) and Lindley distributions

Distribution	Method of moment	Log-likelihood	Score function
Exponential	$\bar{t} = \frac{1}{\theta}$	$n\log\theta - \theta\sum_{i=1}^{n} t_i$	$\frac{n}{\theta} - \sum^{n} t_i = 0$
Gamma(2, θ)	$\bar{t} = \frac{2}{\theta}$	$2n\log\theta + \sum_{i=1}^{n}\log(t_i) - \theta\sum_{i=1}^{n} t_i$	$\frac{2n}{\theta} - \sum^{n} t_i = 0$
Lindley	$\bar{t} = \frac{\theta+2}{\theta(\theta+1)}$	$2n\log(\theta) - n\log(\theta+1) + \sum_{i=1}^{n}\log(1+t_i) - \theta\sum_{i=1}^{n} t_i$	$\frac{2n}{\theta} - \frac{n}{\theta+1} - \sum_{i=1}^{n} t_i = 0$

Table 4: Summary Statistics of the estimated parameter values using the method of moments and the MLE for the Exponential, Gamma(2, θ) and Lindley distributions

Distribution	Estimated parameter
Exponential	$\hat{\theta} = \frac{1}{\bar{t}}$
Gamma(2, θ)	$\hat{\theta} = \frac{2}{\bar{t}}$
Lindley	$\hat{\theta} = \frac{-(\bar{t}-1) + \sqrt{(\bar{t}-1)^2 + 8\bar{t}}}{2\bar{t}}$

3.0 Application

The data used in this paper, comprises of the length of time it takes a graduate before getting a job. The sample was obtained from a particular locality in Nigeria, Ukwuani Local Government Area of Delta State. The data was obtained by administering questionnaires to each respondent who were required to simply specify how long (in months) they were unemployed before getting a job. The length of time starts from when an individual passed out of the National youth service Corps (NYSC) before getting a job. This information is in Table 5 and the Histogram is given in figure 4 while the basic summary Statistics is given in Table 6.

From Table 6, we observe that the distribution is positively skewed. Thus the distribution points to the right since coefficient of skewness is greater than 0^{0}, indicating that the sample is not from a normal distribution while the kurtosis value is greater than 3, indicating excess kurtosis which is also non-normal. The variance is also very large. The reason for this high variability could not be accounted for, as it could be due to many factors such as poor quality of education, lack of skills, economic recession and so on.

We decided to fit the three distributions to the data considering the following: Firstly, the three distributions fit into the exponential family which is exhibited by the histogram. Secondly, the distributions understudy can be used to explain the large variance in the sample data, since the parameter of each model is a function of the respective variances; which can attain large values for some parameter space.

The estimated model parameter, the Akaike information criterion (AIC), the Bayesian information criterion and the

Kolmogorov-Smirnov test for the three models are given in Table 7, the density histogram with the theoretical density of the three distributions is given in Figure 5, the empirical CDF and theoretical CDF is given in Figure 6 while the estimated hazard functions for the three models are given in Figure 7.

Table 5: The length of time (in months) it takes 108 University graduates to get employed.

9, 9, 13, 15, 214, 219, 23, 27, 331, 30, 32, 32, 33, 18, 19, 19, 21, 27, 44, 44, 44, 44, 35, 36, 4, 41, 42, 6, 54, 55,

57, 46, 47, 47, 48, 49, 68, 70, 71, 71, 57, 61, 62, 62, 63, 83, 87, 86, 71, 71, 71, 74, 77, 8, 98, 99, 107, 86, 88, 88,

89, 95, 95, 120, 125, 125, 109, 11, 11, 111, 112, 115, 142, 155, 154, 129, 13, 131, 133, 137, 139, 200, 207, 213,

173, 173, 181, 182, 189, 19, 385, 9, 36, 47, 61, 71, 88, 11, 131, 181, 23, 271, 284, 299, 15, 200, 19, 55

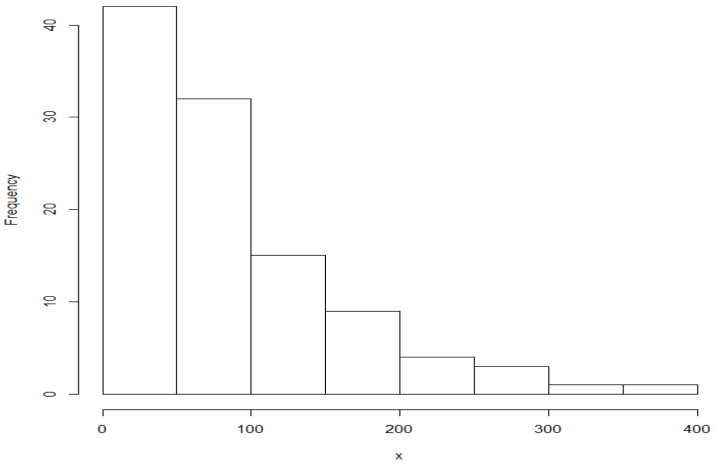


Fig. 4: Histogram of Unemployment Duration in Months

Table 6: Basic Summary Statistics for the sample data.

Mean	Median	Mode	Variance	Minimum	Maximum	Range	CV	Skewness	Kurtosis
87.9907	70	71	5692	4	385	381	0.8574	1.5103	5.4108

Table 7: Basic summary Statistics of the estimated parameter values, Log-likelihood, Akaike information criterion (AIC), Bayesian information criterion (BIC) and Kolmogorov Smirnov (K-S) test of the

Exponential, Gamma(2, θ) and Lindley distributions

Distribution	Estimated parameter	Log-likehood	AIC	BIC	K-S
Exponential	0.0114	-502.3100	1006.60	1009.30	0.094436
Gamma(2, θ)	0.0227	-591.9872	1186.00	1188.70	0.09907
Lindley	0.0225	-586.2747	1174.55	1177.22	0.09396

The parameters of each distribution can be estimated using Table 4 while the log-likelihood is estimated using Table 3. The AIC and BIC are computed respectively, using the following:

AIC = -2(Log-likelihood) + 2k

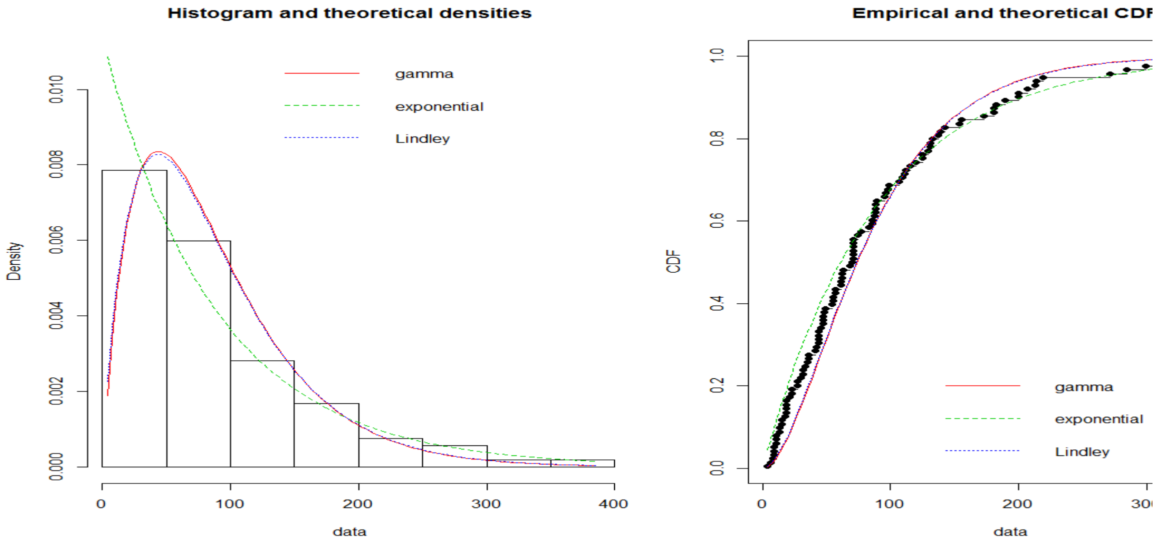
BIC = -2(Log-likelihood) + 2log(n)

where k is the number of parameters in the model and n is the sample size. AIC and BIC are measures of goodness of fit of a model where the model with minimum AIC or BIC value is usually preferred among competing models.

The Kolmogorov-Smirnov (K-S) goodness of fit test was accepted at 5% level of significance, for the three models where K-S test statistic D_n, is defined by:

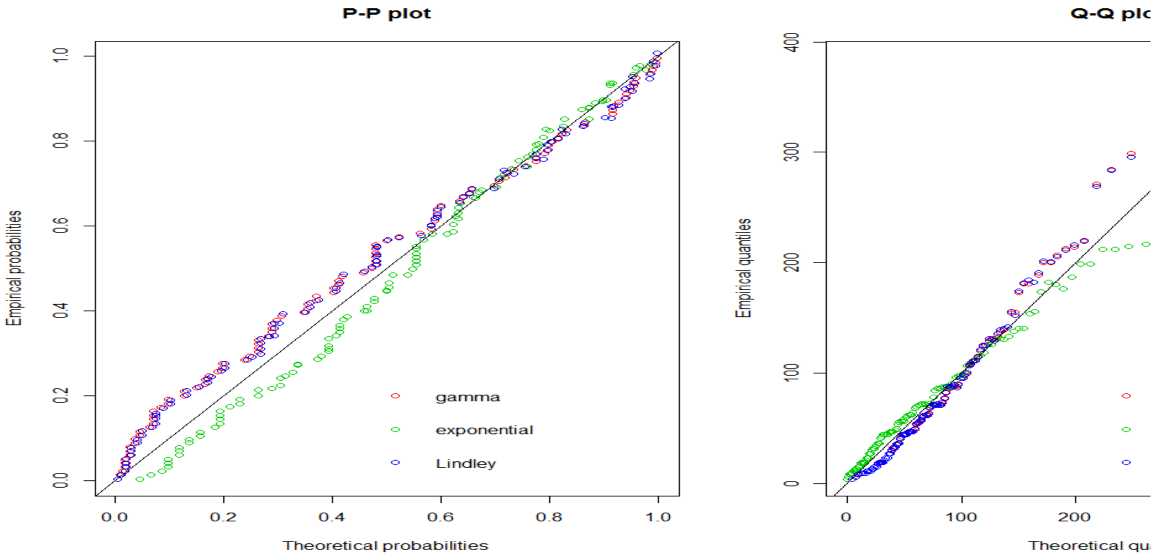
$$D_n = \sup_x \left| F_n(x) - \widehat{F}_n(x) \right|$$

Such that $\widehat{F}_n(x)$ is the cumulative distribution (CDF) and $F_n(x)$ is an empirical CDF [15].



• (b)

Figure 5: (a) Estimated density histogram of the real data and fitted density functions of the Exponential, Gamma(2, θ) and Lindley. (b) The empirical CDF (thick line) and theoretical CDF of the Exponential, Gamma(2, θ) and Lindley.



(a) (b)

Fig 6: (a) Probability (P-P) plot of the estimated Exponential, Gamma(2, $\theta)$ and Lindley. (b) Quantile (Q-Q) plot of the Exponential, Gamma(2, $\theta)$ and Lindley.

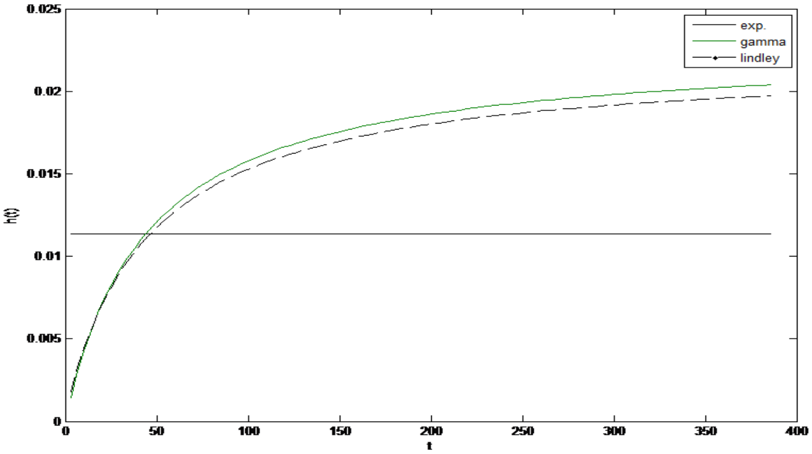


Figure 7: Estimated hazard functions for the Exponential(horizontal line), Gamma(2, θ)(smooth line) and Lindley(broken line)

4.0 Interpretations of Results

The AIC and BIC suggests that the exponential has the best fit of the three models while the K-S test Statistic for Lindley distribution has the smallest value of the test. The estimated densities of the Gamma(2, θ) and Lindley are indistinguishable from each other and both appears to fit the densities more than the Exponential distribution. The P-P plot, Q-Q plot and hazard function for the Gamma(2, θ) and Lindley also follow similar shape.

On the overall assessment, the Lindley distribution remains the best choice in terms of K-S test Statistics. More importantly, the Exponential distribution cannot be used to explain the sample data as it does not suggest an improvement or deterioration in the system in terms of the hazard rate function. This was the case displayed in Figure 7, where the Hazard function for the Exponential distribution is a constant value represented by the horizontal line while the Gamma(2, θ) and Lindley Hazard functions are monotone increasing.

5.0 Conclusion

In this paper, we have studied some properties of the Lindley distribution and its associated distributions; the Exponential and Gamma(2, θ) distributions. We found that the Lindley has more flexibility than the other two distributions in terms of the coefficient of variation, coefficient of skewness and coefficient of kurtosis. We also used a real lifetime data to fit these three distributions to make comparison among these models. Some measure of goodness of fit test such as the AIC, BIC, K-S test, P-P plot and Q-Q plot where used. Our findings was that the Exponential distribution has the minimum AIC and BIC but could not compete with the Lindley distribution in terms of the K-S test Statistics, the P-P plot and Q-Q plot. However, the Lindley distribution is the best choice of the three models in terms of the K-S test Statistic, P-P plot, Q-Q plot and the estimated Hazard function.

6.0 References

[1] Lindley, D. V. (1958) Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society, Series B, Methodological, 20, 102–107.

[2] Sankaran, M. (1970). The discrete Lindley distribution. Biometrics. 26(1): 145-149

[3] Ghitany, M. E., Atieh, B. and Nadarajah, S. (2008) Lindley distribution and its application. Math. Comput. Simul., 78, 493–506.

[4] Ghitany, M. E. Al-qallaf, F., Al-Mutairi, D. k. and Hussain, H. A. (2011). A new two parameter weighted Lindley distribution and its applications to survival data. Mathematics and Computer in simulation. 81(6): 1190-1201

[5] Zakerzadah, H., and Dolati, A. (2010). Generalised Lindley distribution. J. Math. Ext. 3(2): 13-25

[6] Ghitany M.E., Al-Mutairi D. K., Balakrishnan N. and AlEnezi L.J.(2013), Power Lindley distribution and associated inference. Computatinal Statistics and Data Analysis. 64, 20-33.

[7] Rama, S. and Mishra, A. (2013). A quasi-lindley distribution. African Journal of mathematics and Computer Science Research. 6(4), 64-71

[8] Ghitany M. E., Al-Mutairi D. K. and Aboukhamseen S. M. (2015). Estimation of the reliability of a stress-strength system from power lindley distributions, Communications in Statistics Simulation and Computation, 44(1), 118-136.

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

[10] Alkarni, S. H. (2015). Extended Power Lindley distribution: A new Statistical model for non-monotone survival data. European Journal of Statistics and Probability, Vol.3, No.3, 19-34

[11] Pararai, M., Warahena-Liyanage, G. and Oluyede, B. O. (2015). A New Class of Generalized Power Lindley Distribution with Applications to Lifetime Data. Theoretical Mathematics & Applications, vol. 5, no. 1, 53-96

[12] Lawless, J. F. (2003). Statistical Models and Methods for Lifetime Data, 2nd edition, John Wiley & Sons.

[13] Glaser, R. E (1980), Bathtub and Related Failure Rate Characterizations. Journal of the American Statistical Association. 75,667-672

[14] Barlow, R. E. and Marshall, A. W. (1965). Tables of Bounds for Distributions with Monotone Hazard Rate Journal of the American Statistical Association, Vol. 60, No. 311, 872-890

[15] Evans, D. L., Drew, J. H. and L. M. Leemis (2008). The Distribution of the Kolmogorov-Smirnov, Cramer–von Mises and Anderson–Darling Test Statistics for Exponential Populations with Estimated Parameters. Communications in Statistics—Simulation and Computation, 37: 1396–1421