Transactions of the Nigerian Association of Mathematical Physics
Vol. 6 (Jan., 2018), pp. 335 -343
© Trans. of NAMP

In honour of Prof. Ekhaguere at 70

On the choice of two Lomax-based continuous probability distributions

T. G. Ieren[a*] and A. Yahaya[b]

[a,b]*Department of Statistics, Ahmadu Bello University Zaria, Nigeria*

Abstract. Probability distributions and their generalizations have contributed greatly to modeling and statistical analysis of random variables. However, due to the increased introduction of new distributions there has been a major problem with the applications of the several distributions in the literature which has to do with deciding the most appropriate distribution to be used for a given set of data. Most times it is discovered that the data set in question fits two probability distributions and hence one has to be chosen between the two. The Lomax-Weibull and Lomax-Log-Logistic distributions introduced by Cordeiro *et al.* (2014) using a Lomax-based generator were found to be positively skewed and may be victims of this situation when modeling positively skewed data sets. In this paper, we apply the two distributions to some selected data sets to compare their performance and provide useful solutions to the problem of selecting between them in real life situations. We used the value of the *AIC*, *CAIC*, *BIC*, *HQIC*, Cram'er-Von Mises*(W*)*, Anderson Darling*(A*)*and *K-S* statistics as test tools for selecting between these two distributions.

Keywords: Lomax-Weibull distribution, Lomax-Log-Logistic distribution, Lomax-based generator, *AIC*, *CAIC*, *BIC*, *HQIC*, Cram'er-Von Mises *(W*)*, Anderson Darling *(A*)* and *K-S* statistics.

1. Introduction

Selecting between two related probability distribution functions is of great importance and has been done by some researchers such as Atkinson (1969), Dumonceaux et al (1973), Atkinson (1970), Kundu and Manglick (2005), Dumonceaux and Antle (1973), and Kundu and Manglick (2004) e.t.c.

The Lomax distribution was pioneered to model business failure data by Lomax (1954). This distribution has found wide application in different fields such as income and wealth inequality, size of cities, actuarial science, medical and biological sciences, engineering, lifetime and reliability modeling.

A random variable X is said to follow a Lomax distribution with parameter α and β if its probability density function (*pdf*) is given by

$$f(x) = \frac{\alpha}{\beta}\left[1+\left(\frac{x}{\beta}\right)\right]^{-(\alpha+1)} \tag{1}$$

where the corresponding cumulative distribution function (*cdf*) is given as

[*] Corresponding author. Email: ternagodfrey@gmail.com

$$F(x) = 1 - \left[1 + \left(\frac{x}{\beta}\right)\right]^{-\alpha} \quad (2)$$

For $x > 0, \alpha > 0, \beta > 0$ where α and β are the shape and scale parameters respectively.

According to Cordeiro *et al.* (2014) the *cdf* and *pdf* of the Lomax-G family (Lomax-based generator) for any continuous probability distribution are given by:

$$F(x) = 1 - \left\{\frac{\beta}{\beta - \log[1 - G(x)]}\right\}^{\alpha} \quad (3)$$

and

$$f(x) = \alpha\beta^{\alpha} \frac{g(x)}{[1 - G(x)]\{\beta - \log[1 - G(x)]\}^{\alpha+1}}, \quad (4)$$

respectively, where g(x) and G(x) are the *pdf* and *cdf* of any continuous distribution to be generalized respectively and α>0 and β>0 are the two additional new parameters responsible for the scale and shape of the distribution respectively.

The rest of this paper is presented as follows: in Section 2 and section 3 we defined the Lomax-Weibull and Lomax-Log-Logistic distributions, respectively. In section 4 we present a description of the goodness-of-fittest, some data sets, their summary and analysis are provided in section 5. Finally, we provided remarks and conclusions in section 6.

2. The Lomax-Weibull Distribution (*LWD*)

The Weibull Distribution is a very popular continuous probability distribution named after a Swedish Engineer, Scientist and Mathematician, Waloddi Weibull (1887 – 1979). He proposed and applied this distribution in 1939 to analyze the breaking strength of materials. Since then, it has been widely used for analyzing lifetime data in reliability engineering. It is a versatile distribution that can take on the characteristics of other types of distributions, based on the value of the shape parameter. The Weibull distribution is a widely used statistical model for studying fatigue and endurance life in engineering devices and materials.

If a random variable *X* has the Weibull distribution with scale parameter α>0 and shape parameter β>0, then its *cdf* and *pdf* are respectively given by

$$F(x) = 1 - e^{-ax^{b}} \quad (5)$$

and

$$f(x) = abx^{b-1}e^{-ax^{b}} \quad (6)$$

For $x > 0, a > 0, b > 0$ where α *and* β are the scale and shape parameters respectively. Using equation (5) and (6) in (3) and (4) and simplifying, we obtain the *cdf* and *pdf* of the Lomax-Weibull distribution as:

$$F(x) = 1 - \frac{\beta^{\alpha}}{\left[\beta - \log\left[1 - \left(1 - e^{-ax^{b}}\right)\right]\right]^{\alpha}} \quad (7)$$

and

$$f(x)=\frac{\alpha\beta^{\alpha}abx^{b-1}}{\left(\beta-\log\left[1-\left(1-e^{-ax^{b}}\right)\right]\right)^{\alpha+1}} \quad (8)$$

respectively. Hence equation (7) and (8) are the *cdf* and *pdf* of the Lomax-Weibull distribution respectively. The following is a plot the *pdf* of the *LWD* at different parameter values.

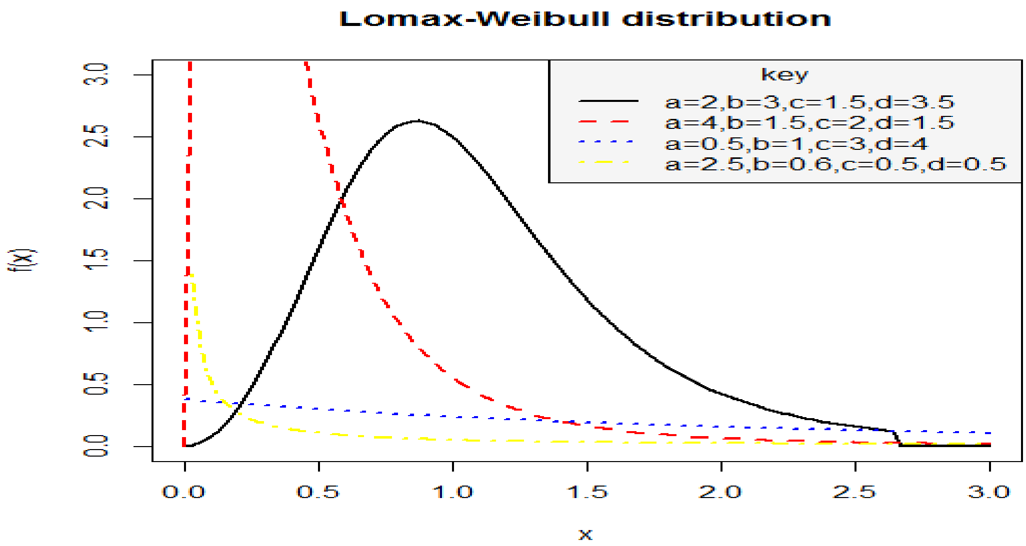

Figure 2: The graph of *pdf* of the *LWD* at different parameter values where $c=\alpha-and-d=\beta$.

Considering the plot above, we can rightly say that the *LWD* is skewed to the right with a very high degree of peakness and can be used for modeling data sets positively skewed with higher kurtosis.

3. The Lomax-Log-Logistic Distribution (*LLD*)

Log-Logistic Distribution: the log-logistic distribution also referred to as the fisk distribution in economics is a continuous probability distribution for a non-negative random variable. The log-logistic distribution is often used to model random lifetime data and hence has applications in reliability.

The *cdf* and *pdf* of the Log-Logistic distribution are given by

$$G(x)=1-\left[1+\left(\frac{x}{a}\right)^{b}\right]^{-1} \quad (9)$$

and

$$g(x)=\frac{b}{a^{b}}x^{b-1}\left[1+\left(\frac{x}{a}\right)^{b}\right]^{-2} \quad (10)$$

For $x>0$, where $a>0$ and $b>0$ are the scale and shape parameters respectively. Using equation (9) and (10) in (3) and (4) and simplifying, we obtain the *cdf* and *pdf* of the Lomax-Log-Logistic distribution as follows:

$$F(x)=1-\frac{\beta^{\alpha}}{\left[\beta+\log\left[1+\left(\frac{x}{a}\right)^{b}\right]\right]^{\alpha}} \tag{11}$$

and

$$f(x)=\frac{\alpha\beta^{\alpha}b}{a^{b}}x^{b-1}\frac{\left[1+\left(\frac{x}{a}\right)^{b}\right]^{-1}}{\left(\beta+\log\left[1+\left(\frac{x}{a}\right)^{b}\right]\right)^{\alpha+1}} \tag{12}$$

Hence equation (11) and (12) are the *cdf* and *pdf* of the Lomax-Log-Logistic distribution respectively. Below is a graph of the *pdf* of the *LLD* for some selected values of the model parameters.

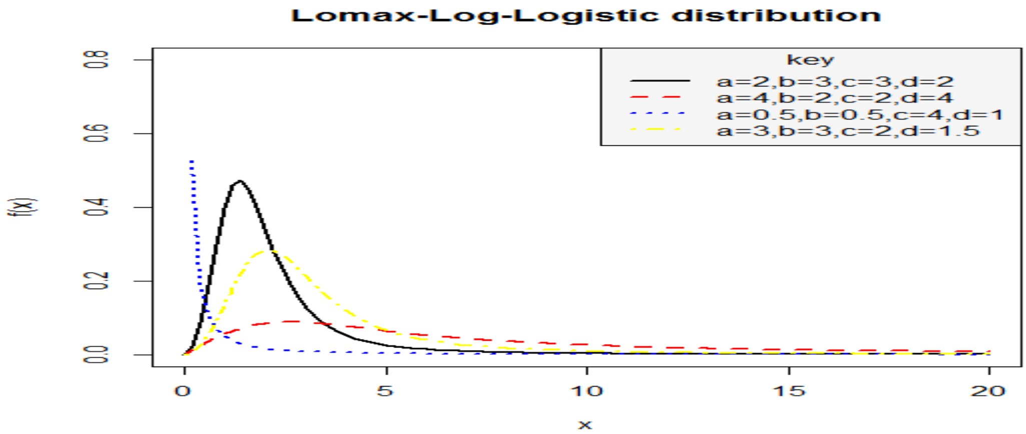

Figure 2: The graph of pdf of the LLD at different parameter values where $c=\alpha-and-d=\beta$.

The plot for the *pdf* shows that the *LLD* is positively skewed with a very low coefficient of kurtosis and therefore will only be good for data sets skewed to the right with moderate kurtosis.

4. Good-of-Fit Test

To compare these two distributions, we have considered some criteria: the value of the log-likelihood function evaluated at the MLEs (ll), *AIC* (Akaike Information Criterion), *CAIC* (Consistent Akaike Information Criterion), *BIC* (Bayesian Information Criterion), and *HQIC* (Hannan Quin Information Criterion). These statistics are given as:

$$AIC=-2ll+2k$$

$$BIC=-2ll+k\log(n),$$

$$CAIC=-2ll+\tfrac{2kn}{(n-k-1)} \text{ and}$$

$$HQIC=-2ll+2k\log\left[\log(n)\right]$$

where ll denotes the log-likelihood function evaluated at the *MLEs*, *k* is the number of model parameters and n is the sample size. We also used goodness-of-fit tests in order to know which distribution fits the data better, we apply the Cram'er-Von Mises *(W*)*, Anderson Darling *(A*)* and the Kolmogorov-Smirnov *(K-S)* statistics. Further information about these statistics can be obtained from Chen and Balakrishnan (1995).These statistics can be computed as:

$$K-S=D=\sup\left|F_n(x)-F_0(x)\right|$$

where $F_n(x)$ is the empirical distribution function and n is the sample size,

$$W^*=w^2\left(1+0.5/n\right)$$

and

$$A^*=A^2\left(1+0.75/n+2.25/n^2\right)$$

where

$$W^2=\sum_{i=1}^{n}\left\{u_i-(2_i-1)/(2n)\right\}^2+1/(12n),$$

$$A^2=-n-\frac{1}{n}\sum_{i=1}^{n}\left\{(2_i-1)\log(u_i)+(2n+1-2_i)\log(1-u_i)\right\},$$

$V_i=F\left(x_i,\hat{\theta}\right)$ is the known cdf with $\hat{\theta}$ (a k-dimensional parameter vector), $y_i=\Phi^{-1}(V_i)$ is the standard quantile function, $u_i=\Phi\left\{\left(y_i-\bar{y}/s_y\right)\right\}$, $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ and $s_y^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$.

Note: In decision making, model with the lowest values for these statistics would be chosen as the best model to fit the data set in question.

5. Analysis of Data

In this section, seven data sets have been used to fit both the Lomax-Weibull and Lomax-Log-Logistic distribution by applying the formulas of the test statistics in section 4 in order to discriminating between the two mentioned distributions. The available data sets and their respective summary statistics are provided in as follows;

Data set I: In this application we consider the data set on the remission times (in months) of a random sample of 128 bladder cancer patients.

Table 1: Summary Statistics for data set I

parameters	n	Minimum	Q_1	Median	Q_3	Mean	Maximum	Variance	Skewness	Kurtosis
Data set I	128	0.0800	3.348	6.395	11.840	9.366	79.05	110.425	3.3257	19.1537

Data Set II: This data set is the strength data of glass of the aircraft window reported by Fuller et al. (1994).

Table 2: Summary Statistics for data set II

parameters	n	Minimum	Q_1	Median	Q_3	Mean	Maximum	Variance	Skewness	Kurtosis
Data set I	31	18.83	25.51	29.90	35.83	30.81	45.38	52.61	0.43	2.38

Data set III: This data set represents the waiting times (in minutes) before service of 100 Bank customers and examined and analyzed by Ghitany *et al.* (2013) for fitting the Lindley distribution.

Table 3: Summary Statistics for data set III

parameters	n	Minimum	Q_1	Median	Q_3	Mean	Maximum	Variance	Skewness	Kurtosis
Data set I	31	0.80	4.675	8.10	13.020	9.877	38.500	52.3741	1.4953	5.7345

Data set IV: This data set represents the lifetime's data relating to relief times (in minutes) of 20 patients receiving an analgesic and reported by Gross*et al.* (1975) and has been used by Shanker *et al.* (2016).

Table 4: Summary Statistics for Data set IV

parameters	N	Minimum	Q_1	Median	Q_3	Mean	Maximum	Variance	Skewness	Kurtosis
Data set I	20	1.10	1.475	1.70	2.05	1.90	4.10	0.4958	1.8625	7.1854

Data Set V: This data represent the survival times in weeks formale rats. (Lawless, 2003).

Table 5: Summary statistics for Data set V

parameters	n	Minimum	Q_1	Median	Q_3	Mean	Maximum	Variance	Skewness	Kurtosis
Data set I	20	40.00	86.75	119.00	140.80	113.45	165.00	1280.892	-0.3552	2.2120

Data Set VI: The data set is from Lawless (1982). The data given arose in tests on endurance of deep groove ball bearings. The data are the number of million revolutions before failure for each of the 23 ball bearings in the life tests. Its summary is given as follows:

Table 6: Summary Statistics for Data Set VI

parameters	N	Minimum	Q_1	Median	Q_3	Mean	Maximum	Variance	Skewness	Kurtosis
Data set I	23	17.88	47.20	67.80	95.88	72.23	173.40	1404.78	1.0089	3.9288

Data set VII: The second data set represents 66 observations of the breaking stress of carbon fibres of 50mm length (in GPa) given by Nicholas and Padgett (2006). The descriptive statistics for this data are as follows:

Table 7: Descriptive statistics for data set VII

parameters	N	Minimum	Q_1	Median	Q_3	Mean	Maximum	Variance	Skewness	Kurtosis
Data set I	66	0.390	2.178	2.835	3.278	2.760	4.900	0.795	-0.1285	3.2230

From the summary statistics of the seven data sets, we found that data sets I, II, III, IV and VI are positively skewed, while V is approximately normal. Also, data sets I, III and IV have higher Kurtosis while others have moderate level of peakness.

Table 8 shows parameter MLEs to each one of the two fitted distributions for the seven data sets (Data sets I-VII), the table alsoshows the corresponding values of*AIC, BIC,CAIC and HQIC* for each model. The values in Table 8 show that the Lomax-Weibull distribution (*LWD*) performs better for five data sets while the Lomax-Log-Logistic distribution (*LLD*) performs better for just two data sets.

Table 8: Performance of the distribution using their AIC, CAIC, BIC and HQIC values of the models MLEs based on data sets I-VII.

Data sets	Models	Log-likelihood value	Statistics	Model Ranks
Data set I	*LWD*	420.7675	*AIC=849.5355* *CAIC=849.8607* *BIC=860.9437* *HQIC=854.1707*	*2*
	LLD	411.4727	*AIC=830.9454* *CAIC=831.2707* *BIC=842.3536* *HQIC=835.5806*	*1*
Data set II	*LWD*	146.435	*AIC=300.8701* *CAIC=302.4085* *BIC=306.606* *HQIC=302.7398*	*1*
	LLD	148.548	*AIC=305.096* *CAIC=306.6345* *BIC=310.832* *HQIC306.9658*	*2*
Data set III	*LWD*	342.2547	*AIC=692.5095* *CAIC=692.9305* *BIC=702.9302* *HQIC=696.7269*	*2*
	LLD	319.8772	*AIC=647.7543* *CAIC=648.1754* *BIC=658.175* *HQIC=651.9718*	*1*
Data set IV	*LWD*	10.3037	*AIC=28.6075* *CAIC=31.2741* *BIC=32.5904* *HQIC=29.3849*	*1*
	LLD	15.7405	*AIC=39.4809* *CAIC=42.1476* *BIC=43.4639* *HQIC=40.2585*	*2*
Data set V	*LWD*	132.1458	*AIC=272.2916* *CAIC=274.9582* *BIC=276.2745* *HQIC=273.0691*	*1*
	LLD	138,5343	*AIC=285.0687* *CAIC=287.7354* *BIC=289.0516* *HQIC=285.8462*	*2*
Data set VI	*LWD*	128.6364	*AIC=265.2728* *CAIC=267.495* *BIC=269.8148* *HQIC=266.4151*	*1*
	LLD	138.7535	*AIC=285.507* *CAIC=287.7292* *BIC=290.0489* *HQIC=286.6492*	*2*
Data set VII	*LWD*	83.5572	*AIC=175.1145* *CAIC=175.7702* *BIC=183.8731* *HQIC=178.5754*	*1*

	LLD	86.7655	*AIC=181.531* *CAIC=182.1867* *BIC=190.2896* *HQIC=184.9917*	*2*

We also notice that the five data sets for which the *LWD* performs better than *LLD* are those with low degree of kurtosis and the two data sets for which the *LLD* performs better are the ones with higher degree of kurtosis. Hence, we can say at this point that the *LWD* should be used for modeling positively skewed data sets most especially those with moderate or low kurtosis while the *LLD* should be applied when the data sets are skewed to the right with higher degree of peakness.

Table 9: Performance of the distributions using the W, A* and K-S values of the models based on data set II.*

Data sets	Models	W*	A*	K-S	Model Ranks
Data set I	*LWD*	0.0312	0.2084	0.2639	*1*
	LLD	0.0382	0.2856	0.0491	*2*
Data set III	*LWD*	0.0212	0.1663	0.2452	*1*
	LLD	0.0536	0.3742	0.0746	*2*
Data set V	*LWD*	0.0872	0.5949	0.6934	*1*
	LLD	0.1136	0.7662	0.4346	*2*
Data set VI	*LWD*	0.0302	0.1867	0.6401	*1*
	LLD	0.0521	0.3937	0.4333	*2*

Table 9 displays the values of goodness-of-fit statistics *W**, *A** and *K-S* for the two distributions under four selected data sets (I, III, V and VI). The results from table 9 confirm that irrespective of the coefficient of kurtosis, the Lomax-Weibull distribution performs better than the Lomax-Log-Logistic distribution. Based on the values of these statistics in table 9, we can confidently say that the *LWD* is better than the *LLD* and hence should be used for analyzing positively skewed data sets. Hence, the statement above is in line with Cordeiro *et al.* (2014) whom also said that the LWD is better than the Beta-Weibull, Kummaraswamy-Weibull, Weibull and the Burr distributions etc.

6 Conclusion

In this paper, we discriminated between two Lomax-based continuous probability distributions called the Lomax-Weibull and Lomax-Log-Logistic distributions. We considered seven real life data sets of different status and used the value of the log-likelihood function, *AIC,CAIC*, *BIC*, *HQIC*, Cram'er-Von Mises *(W*)*, Anderson Darling *(A*)* and *K-S* statistics as test tools for selecting between these two distributions. Our analysis and results proved that the Lomax-Weibull distribution has better performance compared to the Lomax-Log-Logistic distribution irrespective of the level of skewness and kurtosis.

References

[1] Atkinson, A. (1969). A Test of Discriminating between Models, Biometrica, 56, 337-341.
[2] Atkinson, A. (1970). A Method for Discriminating between Models (with Discussion), Journal of Royal Statistical Society, Ser. B, 32, 323-353.
[3] Chen, G and Balakrishnan, N. (1995). A general purpose approximate goodness of –fit test. Journal of Quality Technology, 27,154-161.

[4] Cordeiro, G. M., Ortega, E. M. M., Popovic, B. V and Pescim, R. R. (2014). The Lomax generator of distributions: Properties, minification process and regression model. *Applied Mathematics and Computation*, 247:465-486

[5] Dumonceaux, R. and Antle, C. E. (1973). Discriminating between the Log-Normal and Weibull Distribution, Technometrics,15, 923-926.

[6] 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.

[7] 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.

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

[9] Kundua, D. and Manglick, A. (2004). Discriminating between the Log-Normal and Gamma Distributions, Noval Research Logistic, 51, 893-905.

[10] Kundua, D. and Manglick, A. (2005). Discriminating between the Log-Normal and Gamma Distributions, Journal of Applied Statistical Sciences, 14, 175-187.

[11] Lawless, J. F. (1982). Statistical models and methods for lifetime data. John Wiley and Sons, New York, USA

[12] Lawless, J. F. (2003). Statistical Models and Methods for Lifetime Data, 2nd Ed., Wiley, New Jersey.

[13] Lomax, K.S. (1954) Business failures: Another example of the analysis of failure data, *Journal of the American Statistical Association*, 49: 847–852

[14] Nichols, M. D. and Padgett, W. J. (2006). A bootstrap control chart for Weibull percentiles. *Quality Reliability Engineering Int.*, 22: 141-151

[15] Shanker R, Fesshaye H, and Sharma S (2016). On Two - Parameter Lindley Distribution and its Applications to Model Lifetime Data. *Biometrics and Biostatics International Journal* 3(1): 00056. DOI: 10.15406/bbij.2016.03.00056