Original Research Article

A Model Fit Comparative Study of K-Component Mixture of One Parameter Univariate Distributions

ABSTRACT

Many a researcher develops distributions with the aim to have one-such that models a set of unprecedented data. Mixture distribution as a development model accommodates k-number of distributions to form a new one. This is a comparative review on mixture distribution; where the study seeks to ascertain whether higher number of k-component mixtures could result to development of models that show better fits. In the performance comparison, special consideration was given to univariate one parameter distributions derived using mixture models, and the results show thatdistributions of higher k-mixture components($k \geq 3$)relatively have greater propensity to exhibit better fit than the lesser mixture component distributions ($k < 3$).

Keyword: Mixture distribution,Component mixtures, AIC, Gamma distribution, Model fit.

Comment [L1]: Abstract should be more precise to the findings, not some general overview like this.

1. INTRODUCTION

Comment [L2]: Introduction may accompany a bit about the models compared in the manuscript

Researchers are tirelessly on the look-out for new distributions that model data as events unfold. This, they do by developing a more suitable probability models, instead of subjecting a data to transformation procedures as done in the past. Some methods used in distribution developmentare convolution, parameterization, and mixture models or distribution among many others. Fader and Bruce (2013) developed gamma-gamma distribution, employing the convolution method in the modeling of monetary value. By parameterization, we imply family extension of baseline distributions, as in generalizations. Generalization models are necessary as they meet the rising need of wide applications on data emanating from different real-life events. Cordeiro and de Castro (2011) developed and studied a family of generalized distributions based on the Kumaraswamy distribution proposed by Kumaraswamy (1980). Gupta et al., (1999) first proposed a generalization of the standard exponential distribution, called the exponentiated exponential (EE) distribution, to mention but a few.

Mixture distribution as would be considered in this comparative study is a model that combines parameters and weights ofa weighted summation of distributions to form a new distribution.Ghojogh et al., (2019) detailed in his work "*fitting a mixture distribution to data*", the k-mixing characteristics and model formats. These k-distributions can come from different families of distributions to form the new. However, this leads to derivation complexities; therefore, most mixtures are preferably independent and identically distributed. In other words, they are taken from one family of distribution (e.g., all normal distributions, or exponential or gamma distributions etc.). In order to reduce these complexities, distributions that share likely support or range can be mixed to have the new; for example exponential, pareto, weibull and gamma distributions share likely support in their parameters and variables.In like manner, beta distribution, truncated normal, uniform distribution etc. share the same support range.

Furthermore, complexities are reduced when discrete is mixed with other discrete distributions and continuous with other continuous distribution(s).

Lindley, akash, shanker, ishita, odoma, amarendra, akshaya, shambhu, pranavdistributions etc., are examples where mixture distribution is used in derivations. These exponential-gamma mixtures share the same support:$x > 0\ and\ \theta > 0$. Some common properties among them are: the unimodality and positive skewness of their shapes; the biasedness, consistency and asymptotic normality of their maximum likelihood estimators. In addition, they are all univariate one parameter distributionsranging from $k = 2\ to\ 6$ component mixtures; with an increasing failure rate (IFR) monotonic hazard shapes. This paper is intended to examine whether greater flexibility is achieved across the increasing k – number of mixtures regardless of the different distribution's weight and shape parametric constructions.

2. Methods and Materials

A distribution $f(x)$is a mixture of k-component distributions $g_1, g_2, .., g_k$, if

$$f(x) = \sum_{i=1}^{k} d_i g_i \tag{1}$$

$$f(x,\theta) = d_1 g_1(x,\theta) + d_2 g_2(x,\theta) + d_3 g_3(x,\theta) + \cdots + d_k g_k(x,\theta) \tag{2}$$

whereθ is the vector of parameters as used by Lindsay (1995)and Ghojogh(2019);subject to

$$\sum_{i=1}^{k} d_i = 1; \quad d_i > 0 \tag{3}$$

where$d_i$is the mixing probability (Friedman et al., 2009) and derived as the only weight that complements each of the distributions.

This paper will adopt the following distributions: Lindley, Ishita, Akash, Sujatha, Odoma, Aradhana, Akshaya, Amerendra, Devya and Shambhu distributions among many others. These are all products of mixture distribution; which are combinations of two or more distributions. Lindley (1958) proposedlindley distribution; Shanker R., (2015) proposed Akash distributions; Shanker R., (2016) proposed Aradhana, Devya, Amarendra,Sujatha, Akshaya and Shambhu distributions; and OdomC., (2019) proposed Odoma distribution.The combinations of varying shape parameters and different mixing weights or probabilities and a constant scale make up the gamma-construction of these pdfs. Lindley combined 1 and 2 shape parameters with $\frac{\theta}{\theta+1}$&$\frac{1}{\theta+1}$ as probability weights; for Odoma distribution, it is 1, 3 and 5 shape parameters with $\frac{\theta^5}{\theta^5+\theta^3+6}$, $\frac{\theta^2}{\theta^5+\theta^3+1}$&$\frac{\theta^3+6\theta^5+6}{(\theta^5+\theta^3+6)(\theta^5+\theta^3+1)}$ as probability weights. In Devya distribution, it is a gamma composition of 1, 2, 3, 4 and 5 shape parameters with mixing weights as: $\frac{\theta^4}{\theta^4+\theta^3+2\theta^2+6\theta+24}$,$\frac{\theta^3}{\theta^4+\theta^3+2\theta^2+6\theta+24}$,$\frac{2\theta^2}{\theta^4+\theta^3+2\theta^2+6\theta+24}$,$\frac{6\theta}{\theta^4+\theta^3+2\theta^2+6\theta+24}$&$\frac{24}{\theta^4+\theta^3+2\theta^2+6\theta+24}$andthe rest unmentioned here follow similar procedures.

Table 1: Probability and cumulative distribution function for the selected distributions

Comment [L3]: The distribution may be written in equation form rather than a table form.Tables and figures are a must to cite in the section.

Distribution	k-mixtures	PDF $\theta > 0, \alpha > 0, x > 0$	CDF
Exponential	1	$\theta e^{-\theta x}$ or $\frac{1}{\theta} e^{-\frac{x}{\theta}}$	$1 - \theta e^{-\theta x}$
Lindley	2	$\frac{\theta^2}{\theta + 1}(1 + x)e^{-\theta x}$	$1 - \left(1 + \frac{\theta x}{\theta + 1}\right)e^{-\theta x}$
Akash	2	$\frac{\theta^3}{\theta^2 + 2}(1 + x^2)e^{-\theta x}$	$1 - \left(1 + \frac{\theta x(\theta x + 2)}{\theta^2 + 2}\right)e^{-\theta x}$
Shanker	2	$\frac{\theta^2}{\theta^2 + 1}(\theta + x)e^{-\theta x}$	$1 - \left(1 + \frac{\theta x}{\theta^2 + 1}\right)e^{-\theta x}$
Ishita	2	$\frac{\theta^2}{\theta + 1}(1 + x)e^{-\theta x}$	$1 - \left(1 + \frac{\theta x}{\theta + 1}\right)e^{-\theta x}$
Odoma	3	$\frac{\theta^5}{2(\theta^5 + \theta^3 + 24)}(2x^4 + \theta x^2 + 2\theta)e^{-\theta x}$	$1 - \left[1 + \frac{\theta^2 x^2(\theta^2 x^2 + 4\theta x + 12)}{\theta^5 + \theta^3 + 24} + \frac{\theta x(\theta^4 x + 2\theta^3 + 48)}{2(\theta^5 + \theta^3 + 24)}\right]e^{-\theta x}$
Sujatha	3	$\frac{\theta^3}{\theta^2 + \theta + 2}(1 + x + x^2)e^{-\theta x}$	$1 - \left(1 + \frac{\theta x[\theta x + \theta + 2]}{\theta^2 + \theta + 2}\right)e^{-\theta x}$
Aradhana	3	$\frac{\theta^3}{\theta^2 + 2\theta + 2}(1 + x)^2 e^{-\theta x}$	$1 - \left(1 + \frac{\theta x[\theta x + 2\theta + 2]}{\theta^2 + 2\theta + 2}\right)e^{-\theta x}$
Amarendra	4	$\frac{\theta^4}{\theta^3 + \theta^2 + 2\theta + 6}(1 + x + x^3)e^{-\theta x}$	$1 - \left(1 + \frac{\left[\theta^3 x^3 + \theta^2(\theta + 3)x^2 + \theta(\theta^2 + 2\theta + 6)x\right]}{\theta^3 + \theta^2 + 2\theta + 6}\right)e^{-\theta x}$
Akshaya	4	$\frac{\theta^4}{\theta^3 + 3\theta^2 + 6\theta + 6}(1 + x)^3 e^{-\theta x}$	$1 - \left(1 + \frac{\left[\theta^3 x^3 + 3\theta^2(\theta + 1)x^2 + 3\theta(\theta^2 + 2\theta + 2)x\right]}{\theta^3 + 3\theta^2 + 6\theta + 6}\right)e^{-\theta x}$
Devya	5	$\frac{\theta^5}{\theta^4 + \theta^3 + 2\theta^2 + 6\theta + 24}(1 + x + x^2 + x^3 + x^4)e^{-\theta x}$	$1 - \left(1 + \frac{\left[\theta^4(x^4 + x^3 + x^2 + x) + \theta^3(4x^3 + 3x^2 + 2x) + 6\theta^2(2x^2 + x) + 24\theta x\right]}{\theta^4 + \theta^3 + 2\theta^2 + 6\theta + 24}\right)e^{-\theta x}$
Shambhu	6	$\frac{\theta^6}{\theta^5 + \theta^4 + 2\theta^3 + 6\theta^2 + 24\theta + 120}(1 + x + x^2 + x^3 + x^4 + x^5)e^{-\theta x}$	$1 - \left(1 + \frac{\left[\theta^5(x^5 + x^4 + x^3 + x^2 + x) + \theta^4(5x^4 + 4x^3 + 3x^2 + 2x) + 2\theta^3(10x^3 + 6x^2 + 3x) + 12\theta^2(5x^2 + 2x) + 120\theta x\right]}{\theta^5 + \theta^4 + 2\theta^3 + 6\theta^2 + 24\theta + 120}\right)e^{-\theta x}$

3. Performance comparison

We have in literature thatthe flexibility of a distribution can be improved on,by adding extra parameters through extensions and generalizations. However, we examine in this comparative review, whether goodness of fit depends on increase in the k-number of mixture components. This is achieved with six different data sets, over the performance comparisons (Akaike Information Criterion - AIC) of the selected mixture derived models. Exponential distribution, lindley distribution, ishita distribution and akash distribution represent the $k = 1 \& 2$ component mixtures; Odoma distribution, aradhana and sujatha distribution represent the $k = 3$ component mixtures; Amarendra and Akshaya distribution represent for the $k = 4$ mixture components; devya and shambhu represents $k = 5 \& 6$ mixture components respectively.

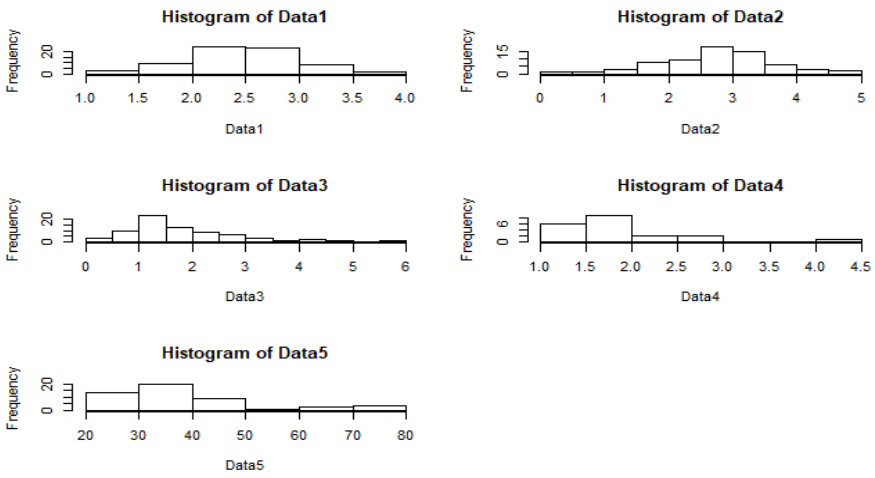

Figure1. The histogram represents Data set I – V.

Comment [L4]: The font should be unique

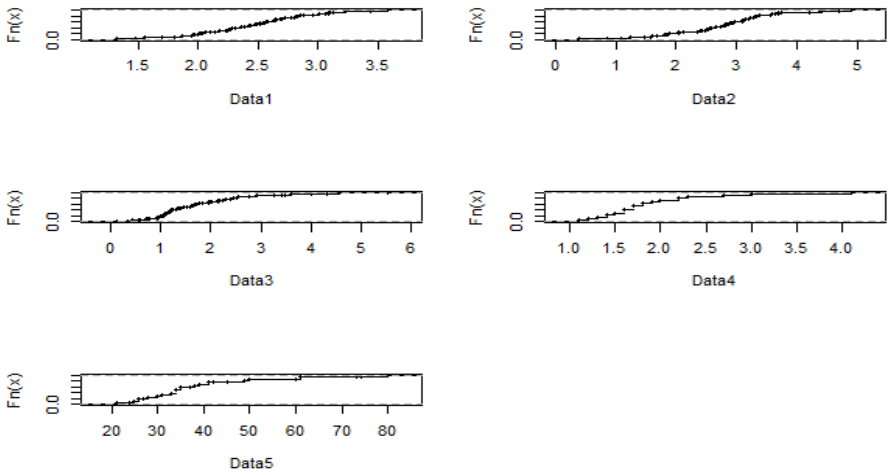

Figure 2. Empirical Distribution Function for Data I-V

Comment [L5]: The font should be unique

Data 1 -Ghitany (2013); Data 2 -Nichol (2006); Data 3 -Elbatal (2013); Data 4 - Gross and Clark (1975); and Data5 is originally collected as the remission time (in months) of 50 breast cancer women subjected

to treatment, using trastzuzumab as medication. Cancer Registry Department, University of Benin Teaching Hospital, Benin, Edo State, Nigeria.

Table 2: k-Mixture Component Performance Comparison

Distribution	k-Mixtures	Data1	Rank	Data2	Rank	Data3	Rank
Exponential	1	263.7	11	267.9	11	228.0	10
Lindley	2	240.4	10	246.8	10	215.9	7
Ishita	2	225.1	8	230.6	6	218.6	9
Akash	2	226.3	9	232.7	9	216.7	8
Odoma	3	211.3	5	215.6	3	228.7	11
Aradhana	3	221.9	6	230.8	7	207.1	2
Sujatha	3	223.6	7	231.6	8	210.4	5
Amarendra	4	209.9	4	219.5	5	208.5	3
Akshaya	4	206.9	3	218.6	4	200.9	1
Devya	5	198.2	2	209.5	2	208.9	4
Shambhu	6	187.7	1	201.2	1	211.3	6

Table 3: Continuation of Table 2

Distribution	k-Mixtures	Data4	Rank	Data5	Rank
Exponential	1	67.7	11	469.9	11
Lindley	2	62.5	10	439.2	10
Ishita	2	62.2	9	421.4	6
Akash	2	61.5	8	421.6	7
Odoma	3	64.0	7	406.9	3
Aradhana	3	58.5	5	423.3	9
Sujatha	3	59.5	6	422.5	8
Amarendra	4	57.6	4	412.8	4
Akshaya	4	55.0	1	413.8	5
Devya	5	56.5	3	406.5	2
Shambhu	6	55.9	2	403.2	1

Comment [L6]: Should follow the same in all the tables

DISCUSSION AND CONCLUSION

Comment [L7]: Put a section number.

In table 3, we can observe explicitly in the ranking of the AIC, that there are clusters of numbersranging from 1-11,matched according to their performance. Now, the smaller, the AIC ranking, the more a distribution could be rated as having better fit.As observed and regardless of the data used in the comparisons, it can be seen that the AIC ranking generally indicates that distributions of higher k-mixture components*relatively*have greater propensity to exhibit better fit than the lesser mixture component

distributions. Unfortunately, the higher k-mixture component distributions, say $k \geq 3$is yet to receive massive attention in the field of statistics, due to the many complexities in the derivation of its properties. However, following the results obtained in this study,higher order mixtures seems to have better fit than the mathematically tractable ones. By implication, while we use the$k < 3$ component mixtures for classroom examples, the development and application of higher mixturesshould be encouraged and recommended, at least for industrial purposes.

REFERENCES

Elbatal, I., Merovci, F., and Elgarhy, M. (2013). A new generalized Lindley distribution, *Mathematical Theory and Modeling*, 3(13), 30-47.

Fader, P. S., &Hardie, B. G. (2013). The Gamma-Gamma model of monetary value.

Friedman, Jerome, Hastie and Robert, 2009. The elements of Statistical learning, volume 2. *Springer series in statistics* New York, NY, USA.

Ghitany, M. E., Atieh, B., &Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and computers in simulation*, *78*(4), 493-506.

Ghitany, M. E., Al-Mutairi, D. K., Balakrishnan, N., & Al-Enezi, L. J. (2013). Power Lindley distribution and associated inference. *Computational Statistics and Data Analysis*, *64*, 20-33.

Ghojogh, B., Ghojogh, A., Crowley, M., &Karray, F. (2019). Fitting a mixture distribution to data: tutorial. *arXiv preprint arXiv:1901.06708*.

Gross A.J., Clark VA., 1975. Survival Distributions: Reliability Applications in the Biometrical Sciences, John Wiley, New York.

Lindsay B. G., 1995. Mixture models: theory, geometry and applications, NSF-CBMS Regional Conference Series in Probability and Statistics, Hayward, CA, USA: Institute of Mathematical Statistics, ISBN 0-940600-32-3, JSTOR 4153184.

Lindley D.V., 1958. Fiducial distributions and Bayes' Theorem. Journal of the Royal Statistical Society. Series B.; 20(1):102–107.

Nichols M.D., & Padgett, W. J., (2006). A bootstrap control chart for Weibull percentiles. *Quality and Reliability Engineering International*, 22(2), 141-151.

Odom, C.C., &Ijomah, M. A. (2019). Odoma Distribution and Its Application. *Asian Journal of Probability and Statistics*, 4(1), 1-11.

Shanker R., (2015). Akash distribution and its applications. *International Journal of Probability and Statistics*, 4(3)65-75.

Shanker R., (2015). Shanker Distribution and its applications. *International Journal of Statistics and Application*, 5(6)338-348.

Shanker R., (2016). Aradhana distribution and its applications. *International Journal of Statistics and Application*, 6(1)23-34.

Shanker R., (2016). Devya distribution and its applications. *International Journal of Statistics and Application*, 6(4)189-202.

Shanker R., (2016). Amarendra distribution and its applications. *American Journal of Mathematics and Statistics*, 6(1)44-56.

Shanker R., (2016). Shambhu distribution and its applications. *International Journal of Probability and Statistics*, 5(2)48-63.

Shanker R., (2016). Sujatha distribution and its applications. *Statistics in Transition New Series*, 17(3)391-410.

Comment [WU8]: Please carefully go through all reference and maintain same style and edition. Citation generator as a tool may help.

Shanker R., (2017). Akshaya distribution and its applications. *American Journal of Mathematics and Statistics*, 7(2)51-59.

Shanker R., Shukka K., (2017). Ishita distribution and its applications. *BiomBiostatInternational Journal* 5(2)39-46.

Comment [L9]: Check the reference properly and cite them in that line. Also match the font with the other references.