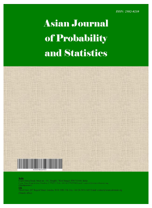


Asian Journal of Probability and Statistics

11(1): 52-61, 2021; Article no.AJPAS.65197

ISSN: 2582-0230

Nwikpe Probability Distribution: Statistical Properties and Goodness of Fit

Barinaadaa John Nwikpe[1*], Isaac, Didi Essi[1] and Amos Emeka[1]

[1]Department of Mathematics, Rivers State University, Port Harcourt, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. Author BJN designed the study, derived the distribution, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors IDE and AE managed the analyses of the study and the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v11i130260
Editor(s):
(1) Dr. Manuel Alberto M. Ferreira, Lisbon University, Portugal.
Reviewers:
(1) Hemanta K. Baruah, The Assam Royal Global University, Gauhati University, India.
(2) Bachioua Lahcene, University of Hail, Saudi Arabia.
Complete Peer review History: http://www.sdiarticle4.com/review-history/65197

Original Research Article

Received: 25 November 2020
Accepted: 30 January 2021
Published: 01 March 2021

Abstract

In this paper, we introduce a new continuous probability distribution developed from two classical distributions namely, gamma and exponential distributions. The new distribution is called the Nwikpe distribution. Some statistical properties of the new distribution were derived. The shapes of its probability density function have been established for different values of the parameters. The moment generating function, the first four raw moments, the second moment about the mean, Renyi's entropy and the distribution of order statistics were derived. The parameter of the new distribution was estimated using maximum likelihood method. The shape of the hazard function of the new distribution is increasing. The flexibility of the distribution was shown using some real life data sets, the goodness of fit shows that the new distribution gives a better fit to the data sets used in this study than the one parameter exponential, Shanker, Lindley, Akash, Sujatha and Amarendra distributions.

Keywords: Probability distribution; probability density function; statistical properties and maximum likelihood.

1 Introduction

Due to advancement in science and technology, the volume of data obtainable for analysis is fast growing. Lately, research in distribution theory shows that most classical distributions do not give satisfactory fit to

**Corresponding author: E-mail: nwikpe4real@gmail.com;*

some of these real life data sets, calling for the formation of new flexible probability distributions that will enhance quality output. Consequently, a lot of efforts have been geared towards making existing probability distributions more flexible. Thus, several techniques now exist for the generalization and extension of classical distribution. For example, the flexibility of a probability distribution could be enhanced by means of generalization, using the available generalized family of distributions. In this work, we develop a new distributions using the mixture of classical distributions.

In search for more flexible distributions, generalized families (G-families) of distributions have gained prominence in recent times. With the use of G-families such as exponentiated G-families by Gupta el tal., [1], Transmuted G-families by Shawn and Buckley [2], Marshall-Olkin-G-families by Marshall and Olkin [3], Beta Generalised (Beta-G) family of distributions by Eugene et al., [4] amongst others, new distributions have been derived thus, increasing the number of distributions available in probability theory. When classical distributions are generalized, the resultant distributions are more flexible than the baseline distribution. For instance, Ghitany [5] generalized the Pareto type 1 distribution using the Marshall Olkin family of distributions, the result of the study revealed the that generalized distribution performed better than the baseline distribution for the data set used in the study. Similarly, Ristic et al. [6] developed Marshall Olkin Gamma distribution, the shape of it density and hazard rate were established, the distribution was also found to be very flexible. Jose and Krishna [7] derived the PDF of Marshall Olkin extended Uniform distribution, the distribution was also found to be more flexible than its baseline distribution.

Recent studies have shown that mixing classical distributions is another efficient way of realizing new flexible distributions. This is owing to the fact that two or more distributions could be compounded to obtain a new distribution. Mixed models have been widely applied in statistics because of their flexibility in terms of real life data sets. For example, the Shanker distribution which is a mixed model proposed by Shank [8] has been shown to be more flexible than the classical exponential distribution, its baseline distributions for some data sets. The probability density function and the cumulative distribution function of Sujatha distribution was also derived by Shanker [9] as a three-component mixture of exponential (θ), gamma(2 θ) and gamma(3, θ) distributions with mixing proportions $\frac{\theta^2}{\theta^2+\theta+2}, \frac{\theta}{\theta^2+\theta+2}, \frac{2}{\theta^2+\theta+2}$. The sujatha distribution also gave a better fit to some real life data set than the classical exponential distribution. Other distributions have been proposed using this method. In this study, we propose a new one-parameter continuous probability distribution using the mixture technique.

2 The Nwikpe Distribution

A random variable X is said to follow the Nwikpe distribution if its probability density function (PDF) is given as:

$$f(x;\theta) = \frac{\theta^3}{12(\theta^2+10)}(\theta^3 x^5 + 12)e^{-\theta x} \qquad (1)$$
$$x > 0, \ \theta > 0$$

Proof

Generally, the probability density function (PDF) of a kth components additive mixture distribution is defined for a random variable y as follows:

$$f(y,\theta) = \sum_{j=1}^{k} f_j(y,\theta_j)\omega_j, \sum_{j=1}^{k}\omega_j = 1 \qquad (2)$$

Where,

θ_j = the vector of parameters for the mixture model
ω_j =Mixture proportion

The probability density function in (1) is a two component mixture of $Gamma(\theta, 6)$ and exponential distributions with mixture ratios $\omega_1 = \frac{\theta^2}{(\theta^2+10)}$ and $\omega_2 = \frac{10}{(\theta^2+10)}$

From equation (2) we get

$$f(x,\theta) = \frac{\theta^3}{(\theta^2+10)}e^{-\theta x} + \frac{10}{(\theta^2+10)}\frac{\theta^6 x^5}{\Gamma(6)}e^{-\theta x}$$

$$= \frac{\theta^3}{12(\theta^2+10)}(\theta^3 x^5 + 12)e^{-\theta x}$$

Which is the PDF of the Nwikpe distribution

The corresponding cumulative distribution of the Nwikpe distribution in (1) is given as

$$F(x;\theta) = \left[1 - \left(1 + \frac{\theta^5 x^5 + 5\theta^3 x^3(4+x) + 60x\theta(\theta x + 2)}{12(\theta^2+10)}\right)e^{-\theta x}\right] \tag{3}$$

Theorem: The Nwikpe distribution is a proper pdf.

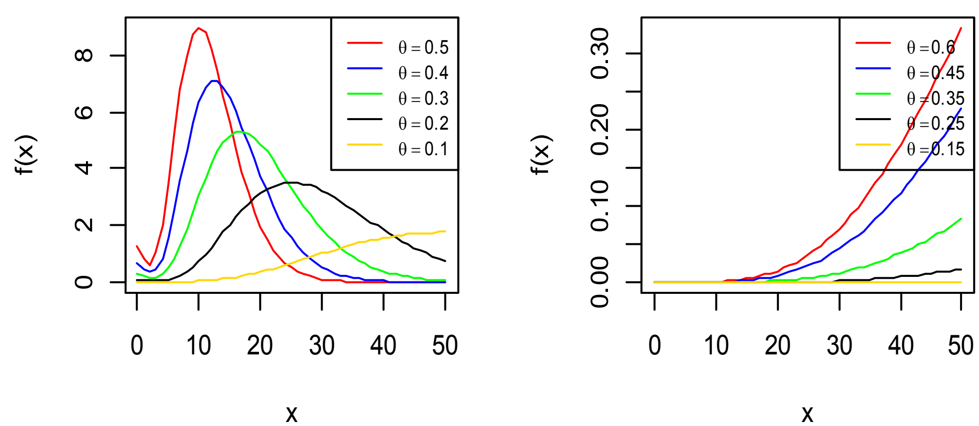
This suffices that

$$\int_{-\infty}^{\infty} f(x;\theta)dx = 1$$

$$= \frac{\theta^3}{12(\theta^2+10)}\int_0^{\infty}(\theta^3 x^5 + 12)e^{-\theta x}dx$$

$$= \frac{\theta^3}{12(\theta^2+10)}\left(\frac{120 + 12\theta^2}{\theta^3}\right) = 1$$

3 The Graph of PDF of the Nwikpe Distribution



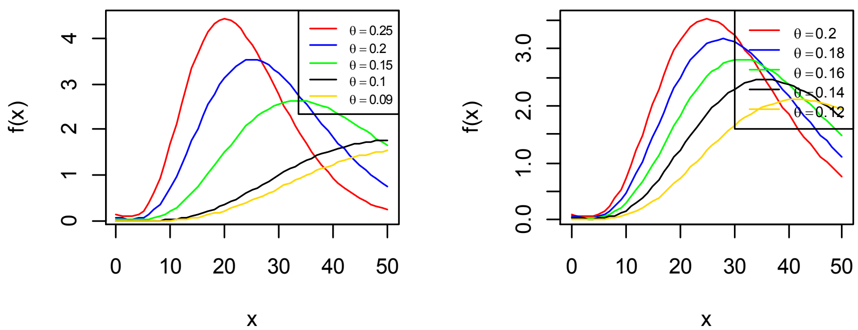


Fig. 1. Graph of the pdf of the Nwikpe distribution

4 The Graph of CDF of the Nwikpe Distribution

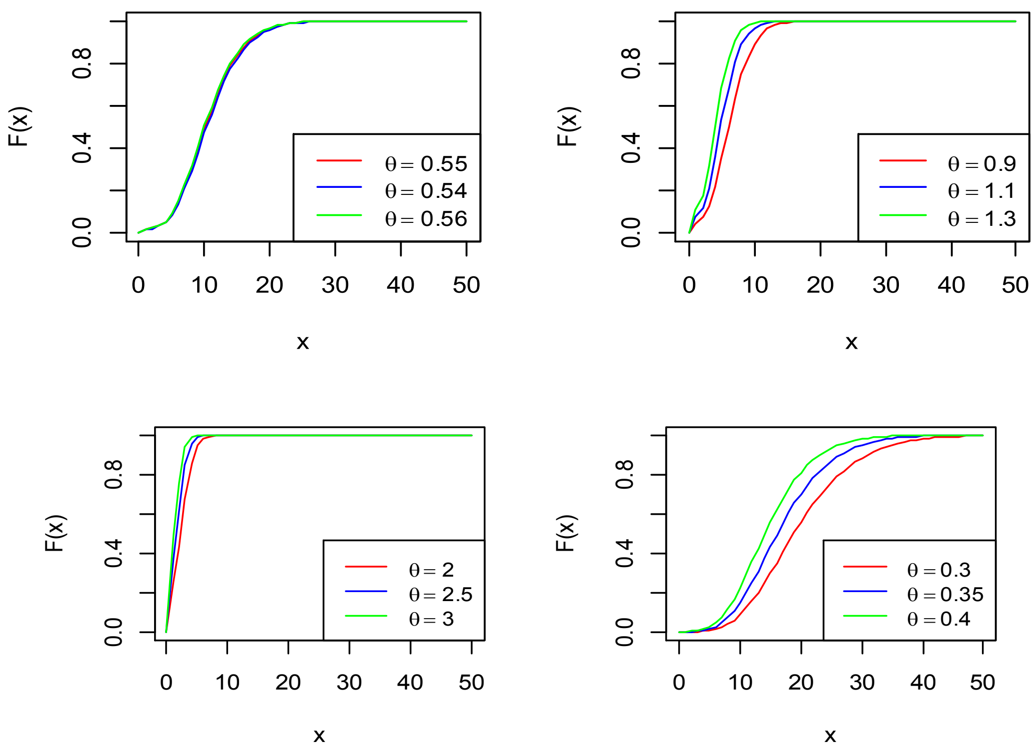


Fig. 2. Graph of the CDF of the Nwikpe distribution

5 Hazard Function of the Nwikpe Distribution

In general, the hazard function of a random variable Y is defined as:

$$h(y) = \frac{f(y)}{1 - F(y)} \qquad = \frac{f(y)}{S(y)} \tag{4}$$

The hazard function of a random variable X which follow follows the Nwikpe distribution is given as:

$$h(x) = \frac{\frac{\theta^3(\theta^3 x^5+12)e^{-\theta x}}{12(\theta^2+10)}}{1 - \left[1 - \left(1 + \frac{\theta^5 x^5+5\theta^3 x^3(4+x)+60x\theta(\theta x+2)}{12(\theta^2+10)}\right)e^{-\theta x}\right]}$$

$$= \frac{\theta^3(\theta^3 x^5 + 12)}{12(\theta^2 + 10) + \theta^5 x^5 + 5\theta^3 x^3(4 + x) + 60x\theta(\theta x + 2)} \tag{5}$$

6 Graph of the Hazard Function of the Nwikpe Distribution

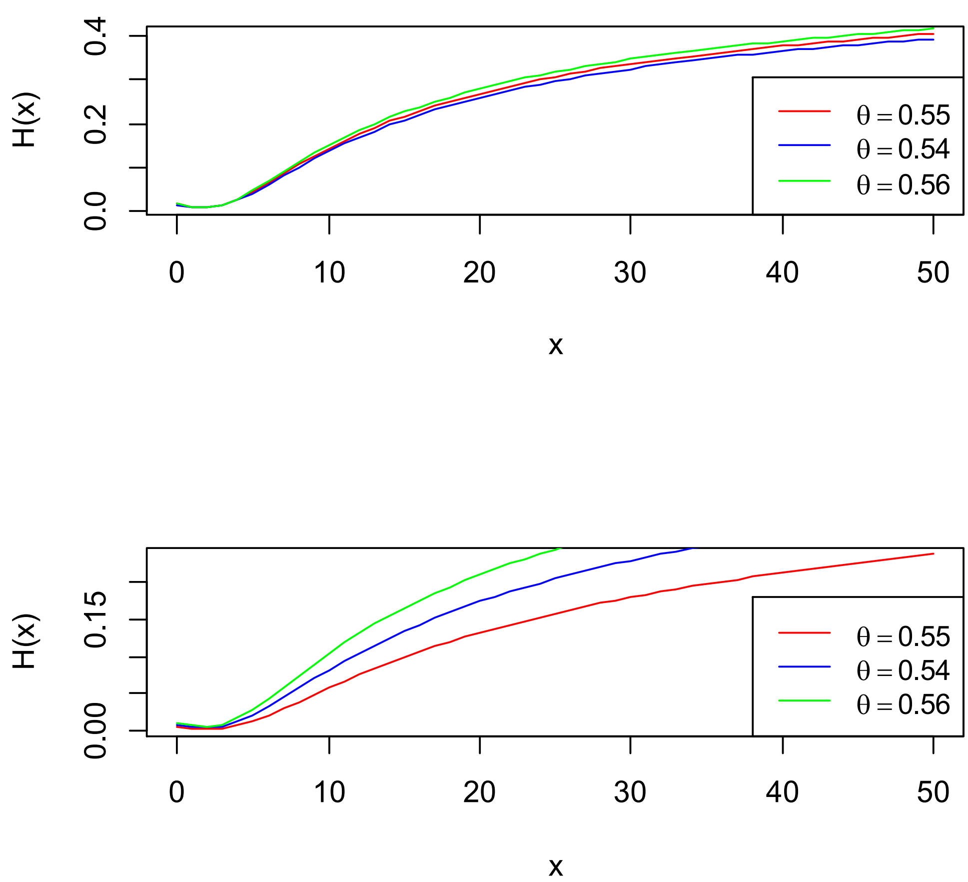


Fig. 3. Graph of the harzad function of the N-E type2 distribution

7 The Moment Generating Function of the New Nwikpe Distribution

By definition, the moment generating function of a random variable X is given as

$$M_X(t) = E(e^{xt}) = \int_{-\infty}^{\infty} e^{xt} f(x)dx$$

If X follows the Nwikpe distribution, its moment generating function is given by

$$M_X(t) \;= \frac{\theta^3}{12(\theta^2+10)}\int_0^{\infty} e^{xt}(\theta^3x^5+12)e^{-\theta x}dx$$

$$= \frac{\theta^3}{12(\theta^2+10)}\left(\frac{\theta^3\Gamma(6)}{(\theta-t)^6}+\frac{12}{(\theta-t)}\right)$$

$$= \frac{\theta^3}{12(\theta^2+10)}\left\{\frac{120\theta^3}{\theta^6}\sum_{r=0}^{\infty}\binom{r+5}{r}\left(\frac{t}{\theta}\right)^r+\sum_{r=0}^{\infty}\frac{12}{\theta}\left(\frac{t}{\theta}\right)^r\right\}$$

$$= \frac{\theta^3}{12(\theta^2+10)}\sum_{r=0}^{\infty}\left(\frac{(r+5)(r+4)(r+3)(r+2)(r+1)+12\theta^2}{\theta^3}\right)\left(\frac{t}{\theta}\right)^r$$

$$= \sum_{r=0}^{\infty}\left(\frac{(r+5)(r+4)(r+3)(r+2)(r+1)+12\theta^2}{12(\theta^2+10)}\right)\left(\frac{t}{\theta}\right)^r$$

The kth raw moment is given by the coefficients of θ^k

$$\mu_k^{'} = \left(\frac{(k+5)(k+4)(k+3)(k+2)(k+1)+12\theta^2}{12\theta^k(\theta^2+10)}\right) \tag{6}$$

Using equation (5) we obtain the first four raw moments of the Nwikpe distribution are given as follows:

$$E(X) = \mu_1^{'} = \frac{720+12\theta^2}{12\theta\,(\theta^2+10)} = \frac{\theta^2+60}{(\theta^2+10)}\;, E(X^2) = \mu_2^{'} = \frac{420+2\theta^2}{\theta^2\,(\theta^2+10)}$$

$$E(X^3) = \mu_3^{'} = \frac{30240+24\theta^2}{\theta^3\,(\theta^2+10)}, E(X^4) = \mu_4^{'} = \frac{3360+6\theta^2}{\theta^4\,(\theta^2+10)}$$

8 The Second Central Moment of the Distribution of the Nwikpe Distribution

By definition, the second central moment of a random variable X is given as

$$\mu_2 = E(x-\mu)^2 = E(X^2) - [E(X)]^2 \tag{7}$$

Thus, the second central moment of the Nwikpe distribution is

$$\mu_2 = \frac{420+2\theta^2}{\theta^2\,(\theta^2+10)} - \left(\frac{\theta^2+60}{(\theta^2+10)}\right)^2 = \frac{\theta^4+320\theta^2+600}{\theta^2\,(\theta^2+10)^2}$$

9 Distribution of Order Statistics of the Nwikpe Distribution

The pdf of the kth order statistics of an n-dimensional random $x_1\, x_2, \ldots\, x_n$ is given by

$$f_{ord:k}(x;\theta) = \frac{n!}{(k-1)!\,(n-k)!}\sum_{i=0}^{n-k}\binom{n-k}{i}(-1)^i\ [F(x)]^{n-k+i}\ f(x)$$

$$f_{ord:k}(x;\theta) = \frac{n!\,\theta^3(\theta^3x^5+12)e^{-\theta x}}{12(\theta^2+10)(k-1)!\,(n-k)!}\sum_{i=0}^{n-k}\binom{n-k}{i}(-1)^i \times$$

$$\left[1-\left(1+\frac{\theta^5x^5+5\theta^3x^3(4\ +x)+60x\theta(\theta x+2)}{12(\theta^2+10)}\right)e^{-\theta x}\right]^{k+i-1}$$

$$= \frac{n!\,\theta^3(\theta^3x^5+12)e^{-\theta x}}{12(\theta^2+10)(k-1)!\,(n-k)!}\sum_{i=0}^{n-k}\sum_{j=0}^{p}\binom{n-k}{i}\binom{p}{j}(-1)^{i+j} \times$$

$$\left(1+\frac{\theta^5x^5+5\theta^3x^3(4\ +x)+60x\theta(\theta x+2)}{12(\theta^2+10)}\right)^p e^{-\theta xp}$$

Where $p = k+i-1.$

Recall $(1+x)^r = \sum_{m=0}^{r}\binom{r}{m}x^m$

$$f_{ord:k}(x;\theta) = \frac{n!\,\theta^3(\theta^3x^5+12)e^{-\theta x}}{12(\theta^2+10)(k-1)!\,(n-k)!}\sum_{i=0}^{n-k}\sum_{j=0}^{p}\sum_{m=0}^{p}\binom{n-k}{i}\binom{p}{j}\binom{r}{m}(-1)^{i+j} \times$$

$$\left(\frac{\theta^5x^5+5\theta^3x^3(4\ +x)+60x\theta(\theta x+2)}{12(\theta^2+10)}\right)^m e^{-\theta xp} \tag{8}$$

10 Renyi's Entropy of the Nwikpe Distribution

The entropy of a random variable is the average amount of uncertainty or randomness inherent in the realizations of the random variable. For any continuous random variable $\boldsymbol{X}$, the Renyi's entropy of order $\boldsymbol{\varphi}$ is defined as:

$$\tau_\varphi = \frac{1}{1-\varphi}log\int_0^\infty f(x)^\varphi dx$$

The Renyi's entropy of a continuous random variable *X* whose pdf is $f(x)$ which follows the Nwikpe distribution is defined as:

$$\tau_\varphi = \frac{1}{1-\varphi}log\left\{\int_0^\infty\left(\frac{\theta^3}{12(\theta^2+10)}(\theta^3x^5+12)e^{-\theta x}\right)^\varphi dx\right\}$$

$$= \frac{1}{1-\varphi}log\left\{\left(\frac{\theta^3}{12(\theta^2+10)}\right)^\varphi\int_0^\infty\left[12\left(1+\frac{\theta^3x^5}{12}\right)\right]^\varphi e^{-\varphi\theta x}\right\}$$

$$= \frac{1}{1-\varphi}log\left\{\left(\frac{12\theta^3}{12(\theta^2+10)}\right)^\varphi\int_0^\infty\sum_{i=0}^{\infty}\binom{\varphi}{i}\left(\frac{\theta^3x^5}{12}\right)^i e^{-\varphi\theta x}\right\}$$

$$= \frac{1}{1-\varphi} log\left\{\left(\frac{12\theta^3}{12(\theta^2+10)}\right)^{\varphi} \sum_{i=0}^{\infty} \left(\frac{\theta^3}{12}\right)^i \int_0^{\infty} x^{5i} e^{-\varphi\theta x}\right\}$$

$$= \frac{1}{1-\varphi} log\left(\frac{12\theta^3}{12(\theta^2+10)}\right)^{\varphi} \sum_{i=0}^{\infty} \left(\frac{\theta^3}{12}\right)^i \frac{\Gamma(5\mathrm{i}+1)}{(\varphi\theta)^{5\mathrm{i}+1}} \tag{9}$$

11 Estimation of Parameter for the Nwikpe Distribution

Let $x_1, x_2, x_2, \ldots, x_n$ be independent and identically distributed random sample of size n from the Nwikpe distribution with parameter $\theta,$ the likelihood function is defined as;

$$l(f(x;\theta)) = \prod_{i=1}^{n} (f(x_i;\theta))$$

$$= \left(\frac{\theta^3}{12(\theta^2+10)}\right)^n \prod_{i=1}^{n} (\theta^3 x_i{}^5 + 12) e^{-\theta x}$$

By taken the log of both sides we get

$$l(f(x_i;\theta)) = nlog\theta^3 - nlog(12(\theta^2+10)) + \sum_{i=1}^{n} log(\theta^3 x_i{}^5 + 12) - \theta \sum_{i=1}^{n} x_i$$

$$\frac{\partial l(f(x_i;\theta))}{\partial(\theta)} = \frac{3n\theta^2}{\theta^3} - \frac{24\theta\, n}{12(\theta^2+10)} + \sum_{i=1}^{n} \frac{3\theta^2 x_i{}^5}{(\theta^3 x_i{}^5 + 12)} - \sum_{i=1}^{n} x_i$$

$$= \frac{3n}{\theta} - \frac{2\theta\, n}{(\theta^2+10)} + \sum_{i=1}^{n} \frac{3\theta^2 x_i{}^5}{(\theta^3 x_i{}^5 + 12)} - \sum_{i=1}^{n} x_i = 0 \tag{10}$$

The solution of equation (9) gives the maximum likelihood estimates of the parameters (θ) for the Nwikpe distribution. However, the equation cannot be solved analytically thus, was solved numerically using R programming with some data set.

12 Application of the Nwikpe Distribution

To determine the flexibility and supremacy of the Nwikpe distribution over some competing models, the distribution was fitted to three data sets; The performance of the distribution was compared with exponential, Lindley, Akash, Amarendra, Sujatha and Shanker distributions for the data sets using Akaike Information Criterion(AIC), Bayesian Information Criterion (BIC), Akaike Information Criterion Corrected (AICC) and -2ln*L*. The distribution that is associated with the smallest AIC, BIC AICC and -2InL is regarded as the most flexible and superior distribution for a given data set. The results are shown in the tables below.

Data Set 1: The first data set is the length of time (in years) that 81 randomly selected Nigerian graduates stayed without job before been employed by the universal basic education commission in 2011.

2,5,7,5,6,7,7,6,6,9,9,6,6,7,5,4,5,2,9,8,5,9,6,6,7,2,8,3,6,6,2,8,5,7,4,5,6,8,8,9,3,7,6,2,6,8,9,7,6,6,9,5,9,5,5,5,3,9,
8,6,6,6,7,9,4,4,6,9,7,8,8,9,4,6,3,5,4,7,6,6,5

Data Set 2: The second data set is the strength data of glass of the aircraft window given by Fuller et al 1994) Studied by Shanker (2015). The data set is given below

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

Data Set 3: The third data is the tensile strength, measured in GPa, of sixty-nine (69) carbon fibres tested under tension at gauge lengths of 20 mm. The data was reported by Bader and Priest in Shanker [9]. The data have been used by Ogutunde [10] to fit the generalized inverse exponential distribution and Shanker (2016) to fit the sujatha distribution. The data is given below;

1.312, 1.314, 1.479, 1.552 1.700 1.803 1.861 1.865 1.944 1.958 1.966 1.997 2.006 2.021 .027 2.055 2.063 2.098 2.140 2.179 2.224 2.240 2.253 2.270 2.272 2.274 2.301 2.301 2.359 2.382 2.382 2.426 2.434 2.435 2.478 2.490 2.511 2.514 2.535 2.554 2.566 2.570 2.586 2.629 2.633 2.642 2.648 2.684 2.697 2.726 2.770 2.773 2.800 2.809 2.818 2.821 2.848 2.880 2.954 3.012 3.067 3.084 3.090 3.096 3.128 3.233 3.433 3.585 3.585

Table 1. Goodness of fit the Nwikpe distribution for data set 1

Model	Parameter Estimate	-2inL	AIC	BIC	AICC
Nwikpe	0.9450	361.4104	363.4103	363.3189	363.461
Exponential	0.1640	454.9130	456.91	459.3044	456.959
Shanker	0.3084	408.9220	410.9216	410.83	410.9722
Lindley	0.2910	418.5780	420.578	420.478	420.6286
Sujatha	0.4403	392.3864	394.3863	394.2949	394.4369
Akash	0.4605	388.6078	390.6073	390.5162	390.56683
Amerandra	0.6015	373.9708	375.9707	375.8792	376.021

Table 2. Goodness of fit of the Nwikpe distribution for data set 2

Model	Parameter Estimate	-2inL	AIC	BIC	AICC
Exponential	0.0324	274.53	276.53	277.96	276.67
Nwikpe	0.1945	223.305	225.305	224.7964	225.44
Shanker	0.0647	252.30	254.30	255.80	254.50
Amerandra	0.1283	233.41	235.4088	236.843	253.55
Sujatha	0.09561026	241.5032	243.5031	244.9432	243.64301
Akash	0.09706217	240.6818	242.6818	244.100	242.80
Lindley	0.0629322	252.9932	255.9942	257.42331	256.1332

Table 3. Goodness of fit of the Nwikpe distribution for Data set 3

Model	Parameter estimate	-2inL	AIC	BIC	AICC
Exponential	0.407941	261.7432	263.7411	265.9655	263.80112
Nwikpe	2.155523	198.80528	200.8053	200.644	200.86
Shanker	0.6580296	233.0054	235.0054	237.2376	235.01
Amerandra	1.244256	207.947	209.947	209.775	210.01
Sujatha	0.9361194	221.6088	223.6088	225.8355	223.6688
Akash	0.9647255	224.2798	226.2797	228.51323	226.34234
Lindley	0.65900001	238.3667	240.3659	242.6134	240.44

13 Conclusion

In this paper, a novel probability distribution has been derived. The new distribution is a single parameter continuous distribution named Nwikpe distribution. The density function of the novel model was derived using a two component additive mixture of gamma and distributions with varying mixing weights. The graphs of the probability density function in Fig. 1 reveal that the distribution could be used to model heavily skewed data sets. Fig. 3 shows that the hazard function of the distribution is increasing. Some statistical properties of the distribution such as the distribution of order statistics, crude and raw moments, Renyi's entropy and moment generating function have been derived. The parameters of the distribution were estimated using the method of maximum likelihood. The performance of the new model was determined by fitting it to three real data set, using goodness of fit criterial such as AIC, BIC AICC and -2inL it was established that the Nwikpe distribution gives a better fit to the data sets used in this work compared with the competing distributions.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Gupta RC, Gupta RD, Gupta PL. Modeling failure time data by Lehman alternatives. Communications in Statistics-Theory and Methods. 1998;27(4):887-904.

[2] Shaw W, Buckley I. The alchemy of probability distributions: beyond gram-charlier expansions and a skew-kurtotic-normal distribution from a rank transmutation map. Research Report; 2007.

[3] Marshall AW, Olkin I. A new method for adding a parameter to a family of distributions with application to the exponential and weibull families. Biometrika. 1997;84(3):641-652.

[4] Eugene N, Lee C, Famoye F. Beta-normal distribution and its applications. Communications in Statistics: Theory and Methods. 2002;31:497-512.

[5] Ghitany M. Marshall-olkin extended pareto distribution and its application. Journal of Applied Mathematics. 2005;18(1):17-32.

[6] Ristic MM, Jose KK, Ancy J. A marshall-olkin gamma distribution and mini_cation process. STARS: International Journal. 2007;1(2):107-117.

[7] Jose KK, Krishna E. Marshall-olkin extended uniform distribution, Prob Stat Forum. 2011;4:78-88.

[8] Shanker R. Shanker distribution and its applications. American Journal of Mathematics and Statistics. 2015;6(1):44-56.

[9] Shanker R. Sujather distribution and its applications. American Journal of Mathematics and Statistics. 2016;6(1):44-56.

[10] Oguntunde PE. The generalised inverted generalized exponential distribution with an application to a censored data. Journal of Statistics Applications and Probability. 2017;4(2):223-230.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/65197