Shako Kumaraswamy distributions with properties and applications in different fields

Shoaib Iqbal[1], Muhammad Hashim[1], Farrukh Jamal[1], Syeda Raheen Ayesha Saeed[1], Muzna Sarwar[1], Muhammad Ali Ibrahim[2]

[1] Department of Statistics, Faculty of Computing, The Islamia University of Bahawalpur, Pakistan

[2] IBMAS, The Islamia University of Bahawalpur, Pakistan

* Corresponding Email: Farrukh.jamal@iub.edu.pk

ABSTRACT

The Shako Kumaraswamy distribution is a novel and adaptable extension of the Kumaraswamy distribution that is presented in this paper. By adding a shape parameter, the suggested model expands on the traditional Kumaraswamy distribution and improves its ability to present a variety of distributional shapes and tail behaviours. The probability density function, quantile function, cumulative distribution function, moments, and other reliability metrics are all derived as part of a thorough mathematical analysis. Additionally, limiting behaviours and special situations are analysed to demonstrate the generality of the suggested distribution. The behaviour of the estimators under various sample sizes and parameter settings is evaluated in a thorough Monte Carlo simulation study to evaluate the effectiveness of the parameter estimation techniques. Additionally, the Shako Kumaraswamy distribution's practicality is illustrated using two real-world data sets from different domains.

Keywords: NEK Distribution; Maximum Likelihood Method

1. Introduction

This article proposes a new distribution, the Shako Kumaraswamy distribution, which represents a new extension in the Kumaraswamy class (NEK). We do this by using the LAII technique, similar to that utilised by [1]. In order to improve the classical Kumaraswamy distribution's capacity to capture a variety of distributional shapes and tail behaviours, this model

adds a new shape parameter. The Shako Kumaraswamy distribution shows improved adaptability and multi-field applicability with the use of the LAII technique.

The Kumaraswamy distribution is useful in many fields and applications. There is a need for a new form of distribution that will be more adaptable and expressive. To establish the discussed distribution, we propose a new version known as the NEK distribution. This newly proposed distribution is defined by four parameters a, b, β and θ decide the shape and behaviours of the distribution. The expanded model becomes more flexible with the introduction of β and θ parameters since probability density functions can be much more varied; thus, the model becomes capable of encompassing a wide array of real-world representations that reflect real-life probabilities. Typically, the most prevalent problem in statistics is analysing the actual datasets with the right kind of models. Therefore, issues to do with model selection present a very significant challenge, especially when the goal is to achieve high accuracy of included parameters. In the year 1980, Kumaraswamy came up with a distribution model, which is a two-parametric distribution. The Kumaraswamy distribution is a continuous probability distribution. It is often used to describe random variables with bounded support, which makes it helpful in a wide range of applications. It has been identified as an alternative to the Beta distribution because they both have the same basic shape properties (unimodal, uniantimodal, increasing, decreasing, monotone or constant). The shape is controlled by two parameters (a and b). As we put a=1 or b=1 if both are equal, then it is called the beta distribution. So Kumaraswamy's distribution is similar to the beta distribution. However, it may be easier to use because its probability density function and cumulative distribution function have simpler analytical expressions.

The probability density function of the Kumaraswamy distribution is

$$f(x) = abx^{a-1}\left(1 - x^a\right)^{b-1} \qquad (1)$$

Where a >0 and b>0 are shape parameters

The cumulative distribution function is

$$F(x) = 1 - \left(1 - x^a\right)^b \qquad (2)$$

For $0 \leq x \leq 1$

The study of Exponentiated distributions in statistics has garnered significant attention since 1996, with a multitude of writers contributing to the development of diverse classes of such

distributions. The Exponentiated Weibull Distribution was proposed by [2]. The Exponentiated Exponential distribution was developed by [3]. [4] Presented an extension of the standard gamma, standard Weibull, standard Gumbel, and standard Frechet distributions by introducing four additional exponentiated type distributions, building upon the same conceptual framework. The Exponentiated Exponential-Poisson distribution was introduced by [5]. While the Exponentiated Exponential-geometric distribution was proposed by [6]. The Exponentiated generalised inverse Gaussian distribution was recently presented by [7]. In this particular instance, we extend the Kumaraswamy distribution similarly. [8], [9], [10] employed the LA2 methodology to introduce exponentiated Fréchet, exponentiated Gumbel, and exponentiated log-logistic distributions. [11] Present other examples of the LA2 methodology's practical implementation.

2. Development for the new Exponentiated Kumaraswamy Distribution

In this paper, we develop a new probability distribution called the NEK distribution. It is the generalised form of the normal Kumaraswamy distribution. Using the mathematical approach, we determine the moments of the proposed probability distribution, probability density, and cumulative distribution. A few examples are applied to illustrate service quality improvement concerning the NEK distribution, while simulated experiments are conducted to explain its functioning. We show that in its current form, it lends itself well to a variety of data sets by simulating data for different data sets. In this case, the NEK distribution is useful in supplementing the statistical arsenal due to its ability to accommodate the variation in an extra set of parameters for a given dataset.

As compared to the standard Kumaraswamy distribution, the NEKD with four parameters (a, b, β and θ) offer more flexibility. It is important to comprehend how these elements engage with one another to best optimise the distribution according to one's modelling needs.

The four parameters are described as;

- Parameters a and b are shape parameters as well as the Kumaraswamy distribution.
- β increases the distribution's ability to capture asymmetric properties in real-world data.
- θ is called the location parameter, enabling the distribution to shift along the horizontal axis

$$f(x) = \frac{\theta\beta abx^{a-1}\left(1-x^a\right)^{b-1}}{\left[1-\left(1-x^a\right)^b\right]}\left[-\theta\log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta-1} e^{-\left(-\theta\log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta}} \qquad (3)$$

And

$$F(x) = e^{-\left[-\theta \log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta}} \tag{4}$$

The survival function, hazard rate function, reversed hazard rate function and cumulative hazard rate function are given as respectively,

$$S(x) = 1 - e^{-\left[-\theta \log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta}}$$

$$h(x) = \frac{\theta\beta abx^{a-1}\left(1-x^a\right)^{b-1}\left[-\theta\log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta-1} e^{-\left(-\theta\log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta}}}{\left[1-e^{-\left(-\theta\log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta}}\right]\left[1-\left(1-x^a\right)^b\right]}$$

$$H(x) = -\log\left(1-1-e^{-\left[-\theta\log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta}}\right)$$

$$r(x) = \frac{\theta\beta abx^{a-1}\left(1-x^a\right)^{b-1}\left[-\theta\log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta-1}}{\left[1-\left(1-x^a\right)^b\right]}$$

2.1. Shape of the pdf and hrfs

From the plots in Figure 1 and Figure 2, it can be seen that the family is very flexible and can offer many different types of density shapes. In Figure 1 the shapes are a bell shape, u-shape, J-shape, increasing, decreasing, symmetry, left skewed and right skewed. In Figure 2 an upside-down bath tub, decreasing-increasing, increasing-decreasing of the hazard rate function. Figure 3 shows the shapes of the CDF and the Survival function.

3. Development of the new Exponentiated Kumaraswamy Distribution

In this article, we derive the survival properties of the NEK models of the failure time data effectively. The NEK distribution, by extending the basic Kumaraswamy distribution, gains importance in the studies in reliability, engineering, and economics. Thus, the properties may include: survival functions, hazard rate functions, cumulative hazard rate functions, reversed hazard

rate functions, as well as other features of the distribution. The most common mathematical properties, such as the Quantile function, rth moments, moment generating function (mgf) and many others are discussed below.

3.1. Expansions and approximations

Using a PDF's series representation makes it simpler to calculate some distributional properties. The result below examines this factor for the exponentiated Pdf of the NEBIII distribution. (x) can be expressed as

$$f(x) = \sum_{i=0}^{\infty} \varphi_i \Psi(x; a.b, \beta, \theta)$$

Where $\varphi_i = \dfrac{ab\theta(-1)^i}{i!}$

And $\Psi(x; a.b, \beta, \theta) = \dfrac{\left[-\theta \log\left(1 - \left(1 - x^a\right)^b\right)\right]^{\beta i + \beta - 1}}{\left(1 - \left(1 - x^a\right)^b\right)}$
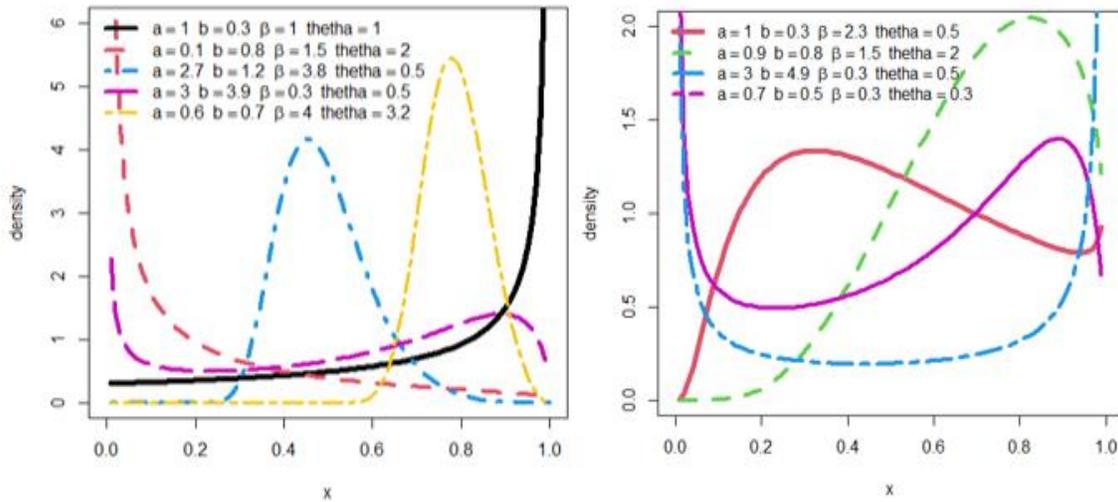
Proof. Using exponential series expansion, we have

$$f(x) = \frac{ab\theta\beta e^{-\left[-\theta \log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta}} \left[-\theta \log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta-1}}{\left(1-\left(1-x^a\right)^b\right)},$$

$$= \frac{c\theta k \sum_{i=0}^{\infty} \dfrac{(-1)^i}{i!} \left[-\theta \log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta i + \beta - 1}}{1-\left(1-x^a\right)^b} \qquad (5)$$

$$= \sum_{i=0}^{\infty} \varphi_i \Psi(x; a.b, \beta, \theta)$$

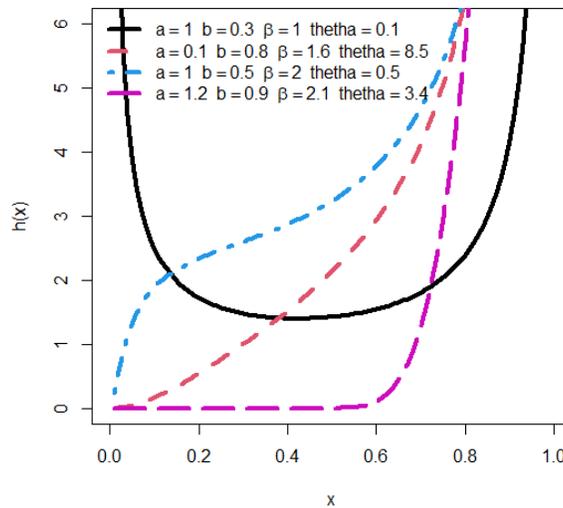## 3.2. Expansions and approximations

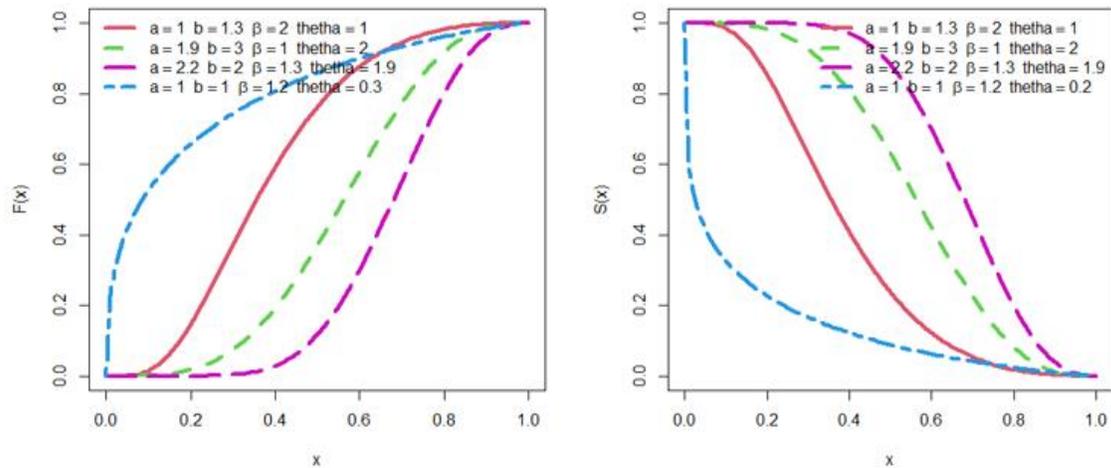The quantile function of the new extended Kumaraswamy distribution is given as;

$$Q(x) = \left[1 - \left\{ 1 - esp\left(-\frac{1}{\theta}\left(\log\left(\frac{1}{q}\right)\right)^{\frac{1}{\beta}} \right) \right\}^{\frac{1}{b}} \right]^{\frac{1}{a}} \qquad (6)$$

Figure 1: The pdf plot of the NEK distribution.

Figure 2: The hrf plot of the NEK distribution.

Figure 3: The cdf & SF plot of the NEK distribution.

3.3. The rth moment of NEK distribution

The expression of the rth moment is given as;

$$\mu_r = \int_{-\infty}^{\infty} x^r f(x) dx = \sum_{i=0}^{\infty} \varphi_i \int_0^1 x^r \Psi(x; a, b, \beta, \theta) dx$$

Finally, we have, using a weight function w_i and the Gamma function (·), we can further simplify the expression μ_r as follows:

$$\mu_r = \sum_{i=0}^{\infty} w_i \Gamma(\beta i + \beta) \qquad (7)$$

Where the series wi is defined as:

$$w_i = \beta \sum_{k=0}^{r/a} \sum_{j=0}^{k/a} \frac{(-1)^{i+j+k}}{i!} \binom{\frac{r}{a}}{k} \binom{\frac{k}{b}}{j} \left(\frac{j}{\theta}\right)^{-(\beta i + \beta)}$$

Where $\dfrac{r}{a} > 0$ and $\dfrac{k}{a} > 0$ real number

3.4. Moment-generating function

The moment-generating function of the new extended Kumaraswamy distribution is given by

$$E\left(e^{tx}\right) = \sum_{i=0}^{\infty} \varphi_i \int_0^1 e^{tx} \Psi\left(x; a, b, \beta, \theta\right) dx = \sum_{i=0}^{\infty} \varphi_i \int_0^1 \sum_{m=0}^{\infty} \frac{\left(tx\right)^m}{m!} \Psi\left(x; a, b, \beta, \theta\right) dx$$

Using a weight function v_i and the Gamma function $(\cdot)$, we can further simplify the expression of $E\left(e^{tx}\right)$ as follows:

$$E\left(e^{tx}\right) = \sum_{m=0}^{\infty} v_i \Gamma\left(\beta i + \beta\right) \tag{8}$$

Where the series v_i is defined as:

$$w_i = \beta \sum_{i=0}^{\infty} \sum_{k=0}^{m/a} \sum_{j=0}^{k/a} \frac{\left(-1\right)^{i+j+k} t^m}{m! i!} \binom{m}{a}{k} \binom{k}{b}{j} \left(\frac{j}{\theta}\right)^{-\left(\beta i + \beta\right)}$$

4.   Entropies

Entropy is a measurement of randomness or unpredictability in a system. Numerous definitions and uses for entropy may be found in various fields, such as physics, information theory, and statistics. Each entropy measure captures a different aspect of uncertainty or unpredictability in a system.

4.1.   Renyi entropy

The Renyi entropy quantifies the degree of uncertainty or randomness in a probability distribution. [12] first introduced it as a generalisation of the Shannon entropy.

$$l_r = \frac{1}{1-r} \log\left[\int f(x)^r dx\right] \tag{9}$$

4.2.   Shannon Entropy

Shannon Entropy was defined by [13] as a way to quantify uncertainty, disorder, or unpredictability in a physical system.

$$H(x) = \int_0^x \left[\log f(x)\right] f(x) dx$$

$$S(x) = \sum_{i=0}^{N} p(x_i) \log\left[p(x_i)\right] \tag{10}$$

4.3. Maximum Likelihood Function

Maximum Likelihood Estimation (MLE) is a technique for estimating statistical model parameters. It works on the assumption of maximising the likelihood function, which assesses how well the model describes the observed data. MLE is a technique for estimating the parameters of a distribution using observed data. It is a statistical approach for estimating the parameters of a probability distribution that best fit a given dataset.

$$L = \prod_{i=0}^{n} f(x)$$

$$= \prod_{i=0}^{n} \frac{\theta \beta a b x^{a-1} \left(1-x^a\right)^{b-1}}{\left[1-\left(1-x^a\right)^b\right]} \left[-\theta \log\left(1-\left(1-x^a\right)^b\right)\right]^{\beta-1} e^{-\left(-\theta \log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta}}$$

$$= \frac{\theta^n \beta^n a^n b^n \prod_{i=0}^{n} x^{a-1} \left(1-x^a\right)^{b-1} \left\{-\theta \log\left(1-\left(1-x^a\right)^b\right)\right\}^{\beta-1} e^{-\left(-\theta \log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta}}}{\prod_{i=0}^{n}\left[1-\left(1-x^a\right)^b\right]}$$

Taking log, we get;

$$\log L = \left[\begin{array}{l} n \log \theta + n \log \beta + n \log a + n \log b + (a-1)\sum_{i=0}^{n}\log x + (b-1)\sum_{i=0}^{n}\log\left(1-x^a\right) - \\ \sum_{i=0}^{n}\left(-\theta \log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta} + (\beta-1)\sum_{i=0}^{n}\log\left(-\theta \log\left(1-\left(1-x^a\right)^b\right)\right) - \sum_{i=0}^{n}\log\left(1-\left(1-x^a\right)^b\right) \end{array}\right]$$

Differentiate w.r.t θ, β, a, b **we get;**

$$\frac{\partial l}{\partial a} = \frac{n}{a} + \sum_{i=0}^{n}\log x - (b-1)\frac{\sum_{i=0}^{n} x^a \log x}{\left(1-x^a\right)} + \frac{\beta b \theta \sum_{i=0}^{n}\left(-\theta \log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta-1}\left(1-x^a\right)^{b-1} x^a \log x}{\left(1-\left(1-x^a\right)^b\right)} +$$

$$\frac{(\beta-1)b\sum_{i=0}^{n}\left(1-x^a\right)^{b-1} x^a \log x}{\left(1-\left(1-x^a\right)^b\right)\log\left(1-\left(1-x^a\right)^b\right)} - \frac{b\sum_{i=0}^{n}\left(1-x^a\right)^{b-1} x^a \log x}{\left(1-\left(1-x^a\right)^b\right)}$$

$$\frac{\partial l}{\partial b} = \frac{n}{b} + \sum_{i=0}^{n} \log\left(1-x^a\right) + \frac{\beta\theta\sum_{i=0}^{n}\left(-\theta\log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta-1}\left(1-x^a\right)^b\log\left(1-x^a\right)}{\left(1-\left(1-x^a\right)^b\right)} -$$

$$\frac{(\beta-1)\sum_{i=0}^{n}\left(1-x^a\right)^b\log\left(1-x^a\right)}{\left(1-\left(1-x^a\right)^b\right)\log\left(1-\left(1-x^a\right)^b\right)} + \frac{b\sum_{i=0}^{n}\left(1-x^a\right)^b\log\left(1-x^a\right)}{\left(1-\left(1-x^a\right)^b\right)}$$

$$\frac{\partial l}{\partial \theta} = \frac{n}{\theta} + \beta\sum_{i=0}^{n}\left(-\theta\log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta-1}\log\left(1-\left(1-x^a\right)^b\right) - \frac{(\beta-1)n}{\theta}$$

$$\frac{\partial l}{\partial \beta} = \frac{n}{\beta} + \theta\sum_{i=0}^{n}\left(-\theta\log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta}\log\left(-\theta\log\left(1-\left(1-x^a\right)^b\right)\right) - \sum_{i=0}^{n}\left(1-x^a\right)^b\log\left(1-x^a\right)$$

4.4. Order Statistics

Order statistics is a statistical discipline that examines the properties and applications of data elements that are organised in a predetermined sequence. The kth The statistic of a statistical sample is synonymous with its kth smallest value. Order statistics, in addition to rank statistics, are essential tools in non-parametric statistics and inference.

$$f_i(x) = \frac{n!}{(i-1)!(n-i)!}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{n-i}f(x)$$

$$f_i(x) = \frac{n!}{(i-1)!(n-i)!}\sum_{j=1}^{\infty}\binom{n-i}{j}(-1)^j\left[e^{-\left(-\theta\log\left(1-\left(1-x^a\right)^b\right)^{\beta}\right)}\right]^{i+j-1}\left(\frac{\theta\beta abx^{a-1}\left(1-x^a\right)^{b-1}}{\left[1-\left(1-x^a\right)^b\right]}\right) \quad (11)$$

$$\left(-\theta\log(1-\left(1-x^a\right)^b\right)^{\beta-1}e^{-\left(\theta\log\left(1-\left(1-x^a\right)^b\right)\right)^{\beta}}$$

Table 1:
Moments, Skewness and Kurtosis of *X* for selected parameter values for NEK,(a, b, β, θ).

$\mu_k^{/}$	NEK (1,4,3,2)	NEK (2,3,4,5)	NEK (3,2,5,6)
$\mu_1^{/}$	0.3112	0.6429	0.8387
$\mu_2^{/}$	0.1051	0.4144	0.7038

$\mu_3^/$	0.0384	0.2678	0.5909
$\mu_4^/$	0.0151	0.1736	0.4963
Variance	0.0082	0.0010	0.0003
S. D	0.0908	0.0330	0.0177
Skewness	0.7919	0.7313	0.6592
Kurtosis	831.2596	864161.1	29681113
CV	0.2918	0.05133	0.0212

Table 1 shows that the mean of the NEK distribution exhibits a steady increasing trend as the parameter values increase, yet the variance and standard deviation decline, suggesting a decrease in the spread overall. As the skewness values decrease, the distribution becomes more symmetrical, and the fast-growing kurtosis shows that the peaks are sharper and the tails are heavier. A more concentrated distribution around the mean is also suggested by the considerable drop in the coefficient of variation. These results demonstrate how well the NEK distribution captures a variety of data characteristics in terms of shape, dispersion, and central tendency.

5. Simulation Study

Simulation analysis is a vital tool in statistics that is used to evaluate the performance of estimates over partial replication at various sample sizes. Thus, the primary focus of this section is the simulation-based performance parameter estimations of the proposed NEK distribution model. A simulation process is replicated 1000 times, with varying sample sizes (n = 100,200,300,500).

The parameter estimates and associated MSE's for the NEK distribution that were derived using the maximum likelihood (ML) and least squares (LS) estimation techniques are shown in Table 2. Two parameter settings are used for the evaluation: Set 1 with (a, b, β, θ) = (0.4,1.2,1.6,0.3) and Set 2 with (a, b, β, θ) = (0.1,0.8,1.0,0.6).

Table 2:
MSEs, MLEs, LSE, Estimates for many scenarios.

| n | SET 1 | | | | | | SET 2 | | | | | |
| | MLE | | | LSE | | | MLE | | | LSE | | |
	Initial values	Estimates	MSEs	Initial values	Estimates	MSEs	Initial values	Estimates	MSEs	Initial values	Estimates	MSEs
100	a=0.4	0.2431	0.0018	a=0.4	0.431	0.0519	a=0.1	0.2312	0.3679	a=0.1	0.2931	0.0192
	b=1.2	0.1612	0.0135	b=1.2	1.329	0.0012	b=0.8	0.3591	0.4287	b=0.8	0.1322	0.0026
	β=1.6	0.7931	0.0273	β=1.6	1.642	0.0015	β=1.0	0.4812	0.2481	β=1.0	0.2331	0.0042
	θ=0.3	0.6521	0.0015	θ=0.3	1.341	0.0021	θ=0.6	0.5621	0.0315	θ=0.6	0.6721	0.0013
200	a=0.4	0.2431	0.1284	a=0.4	0.761	0.0023	a=0.1	0.3217	0.2561	a=0.1	0.4131	0.0136
	b=1.2	0.4512	0.5232	b=1.2	1.264	0.0014	b=0.8	0.5321	0.5312	b=0.8	1.423	0.002
	β=1.6	0.3931	0.1321	β=1.6	2.981	0.0017	β=1.0	0.4162	0.1223	β=1.0	1.5412	0.0017
	θ=0.3	0.6121	0.2504	θ=0.3	0.531	0.0036	θ=0.6	0.2971	0.0322	θ=0.6	1.6582	0.0021
300	a=0.4	1.0431	0.9051	a=0.4	0.218	0.3612	a=0.1	0.5798	0.1598	a=0.1	0.4531	0.0301
	b=1.2	1.5612	0.5152	b=1.2	1.842	0.0183	b=0.8	0.1639	0.3219	b=0.8	0.1523	0.0018
	β=1.6	0.2931	0.4321	β=1.6	1.734	0.0294	β=1.0	0.2035	0.431	β=1.0	0.4331	0.0035
	θ=0.3	0.1821	0.2021	θ=0.3	0.426	0.1652	θ=0.6	1.8236	0.052	θ=0.6	0.7021	0.0027
500	a=0.4	0.3531	0.9051	a=0.4	0.362	0.0412	a=0.1	1.6421	0.2134	a=0.1	0.3281	0.0521
	b=1.2	1.5612	0.5152	b=1.2	0.536	0.0012	b=0.8	0.7421	0.3591	b=0.8	0.1872	0.0017
	β=1.6	0.1931	0.4321	β=1.6	1.632	0.0032	β=1.0	1.5281	0.3231	β=1.0	0.3201	0.0023
	θ=0.3	0.8821	0.3013	θ=0.3	0.472	0.0015	θ=0.6	0.4321	0.6013	θ=0.6	0.1982	0.0018

5. Applications of the NEK distribution

The NEKD's improved parameterisation offers a way to more kinds of applications than the previous model. We look at several scenarios where the NEKD can be used, including dependability modelling, financial data analysis, and environmental investigations. The distribution's improved flexibility makes it possible to represent varied datasets more accurately.

In this section, two applications of actual data sets are given in this subsection to highlight the significance of the NEK distribution. Each data set undergoes the Kolmogorov-Smirnov goodness of fit test to determine the validity of the fitted model, and in each case, the p values show that the model fits the data extremely well. We provide some applications of the NEK distribution on two real-life phenomena.

Data Set 1: The first data set is extracted from [14] and refers to 20 observations of the Susquehanna River's maximum flood level located at Harrisburg, Pennsylvania. The unit is a million cubic feet per second. The precise values are: 0.26, 0.27, 0.30, 0.32, 0.32, 0.34, 0.38, 0.38, 0.39, 0.40, 0.41, 0.42, 0.42, 0.42, 0.45, 0.48, 0.49, 0.61, 0.65 and 0. 74. We compute to evaluate the Maximum likelihood estimation (MLEs) and the standard error. Lower values of all these statistics, except greater p-values for the Kolmogorov-Smirnov measures (K-S), imply good fits, including the Akaike Information Criterion (AIC), Anderson-Darling (A*), Cramer-von Mises (W*) and P-value (p). The R programming language is used to do the necessary computations. SAAN(S) is a simulated annealing method, a global optimisation algorithm. It is a starting value, also called the initial kick. Whereas (B) is the optimal value.

Table 3:
Estimated Parameters by MLE with their S.E. and Goodness of Fit.

Distribution	A	B	β	θ
NEK(MLE)	1.15896203	0.06918685	0.27259758	10.70685211
SE	4.5978663	0.2266792	0.5912665	11.7982520
EK(MLE)	0.4794043	4.7252347	-	100.5368155
SE	0.4018444	1.3755054		220.3525677
K(MLE)	3.377843	12.007601	-	-
SE	0.6043048	5.4751547	-	-

Comparing the NEK distribution to the EK and K distributions for datasets 1 & 2, Table 3 & 6 demonstrates that the standard errors are relatively lower and more consistent across all parameters. The NEK distribution provides the best-balanced estimates and standard errors for all parameters.

Table 4:
Statistical measures related to the distributions for data set 1.

Distribution	ℓ	AIC	BIC	CAIC	HQIC
NEK	-16.47043	-24.94086	-20.95793	-22.27419	-24.16335
EK	-15.82544	-25.65088	-22.66368	-24.15088	-25.06775
K	-12.97327	-21.94654	-19.95508	-21.24066	-21.55779

This graph shows that the NEK distribution, which is applied in the data set 1, the data is a good fit compared to other distributions, which are EK & K. The high level of accuracy in the general quality of the fit of the NEK distribution is shown here.

Table 5:
The value of W* A* and P value of data set 1.

Distribution	P-value	A*	W*	Distribution
NEK	0.9105	0.1523	0.0203	NEK
EK	0.5210	0.2261	0.0489	EK
K	0.230	1.3271	.0921	K
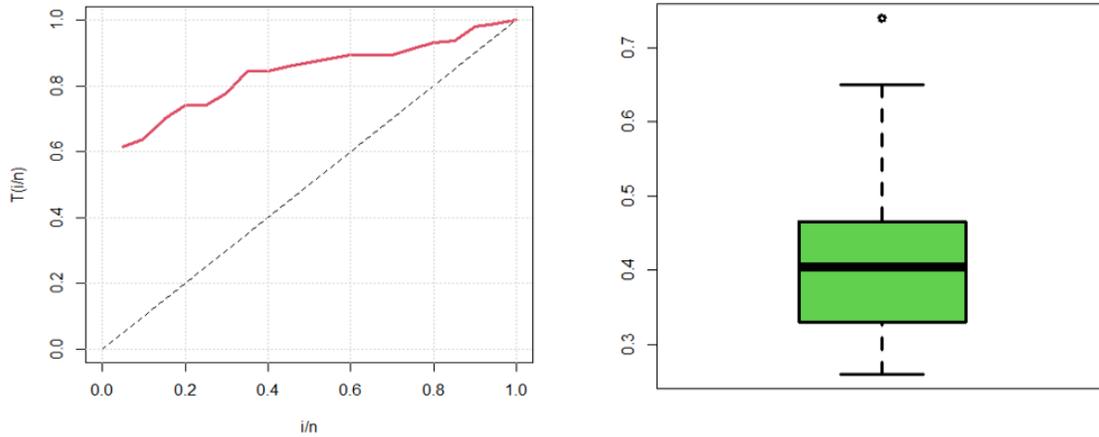
Figure 4: Shows the TTT and Box Plot of NEK from Data set 1.

Figures 4 & 6 show that the dashed diagonal line represents the theoretical expectation assuming perfect model fit, whereas the red curve shows the empirical cumulative distribution of the transformed data under the fitted model. Overall, the red curve's alignment with the line indicates that the model fits the data rather well, with no significant systematic discrepancies, although it is slightly above the diagonal, showing modest deviations. For the majority of the distribution, the empirical probabilities are marginally higher than the theoretical ones, as seen by the upward trend in the TTT plot and also show the box plot for data set I&II.
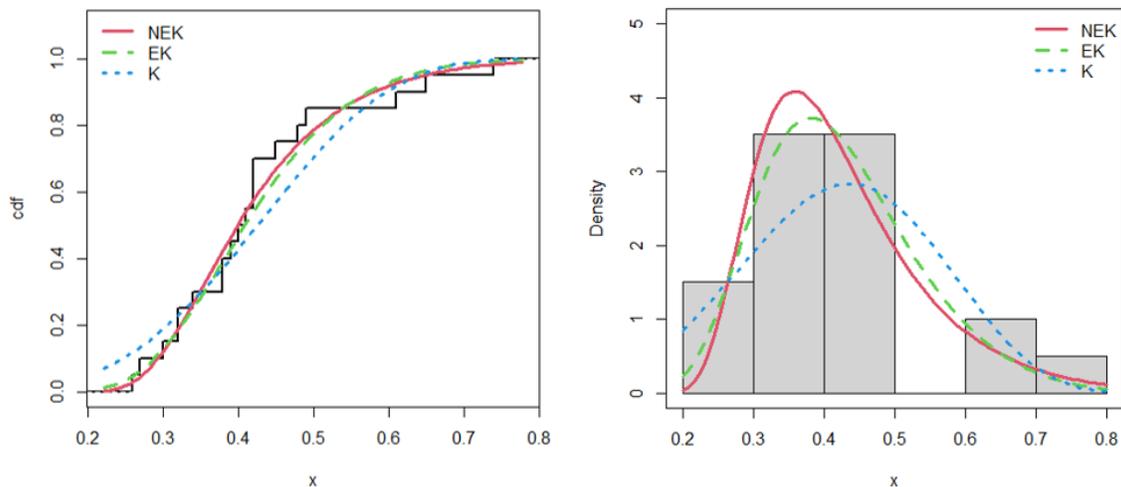
Figure 5: Shows the pdf and cdf graphs of the competitor's distributions for data set 1.

Data set 2: The second data set is on shape measurements of 48 rock samples from a petroleum reservoir. This data was extracted from BP research, image analysis by Ronit Katz, Oxford. This data is given as follows: 0.0903296, 0.189651, 0.228595, 0.200071, 0.280887, 0.311646, 0.176969, 0.464125, 0.148622, 0.164127, 0.231623, 0.144810, 0.179455, 0.276016, 0.438712, 0.420477, 0.183312, 0.203654, 0.172567, 0.113852, 0.191802, 0.197530, 0.163586, 0.200744,0.117063, 0.162394, 0.153481, 0.291029, 0.133083, 0.326635,0.253832, 0.262651, 0.122417, 0.150944, 0.204314, 0.240077, 0.225214, 0.154192, 0.328641, 0.182453, 0.167045, 0.148141, 0.262727, 0.161865, 0.341273,0.276016, 0.230081, 0.200447.

Table 6:
Estimated Parameters by MLE with their S.E. and Goodness of Fit.

Distributions	A	B	β	θ
NEK(MLE)	0.2217281	1.0908688	2.9631321	4.8297660
SE	2.424452	5.011362	18.036655	15.829481
EK(MLE)	0.4621476	7.6199210	-	99.2219316
SE	0.2899777	2.2736480	-	191.0576193
K(MLE)	2.715723	44.464601	-	-
SE	17.4588118	17.4588118	-	-

Table 7:
Statistical measures related to the distributions for data set 2.

Distribution	ℓ	AIC	BIC	CAIC	HQIC
NEK	-58.40772	-108.8154	-101.3306	-107.8852	-105.9869
EK	-58.21548	-110.431	-104.8173	-109.8855	-108.3096
K	-52.49147	-100.9829	-97.24054	-100.7163	-99.56868

Table 8:
The value of W* A* and P value of Data set 2.

Distribution	P-value	A*	W*
NEK	0.8906	0.1957879	0.02944783
EK	0.7379	0.265887	0.02944783
K	0.2101	1.278951	0.2081727
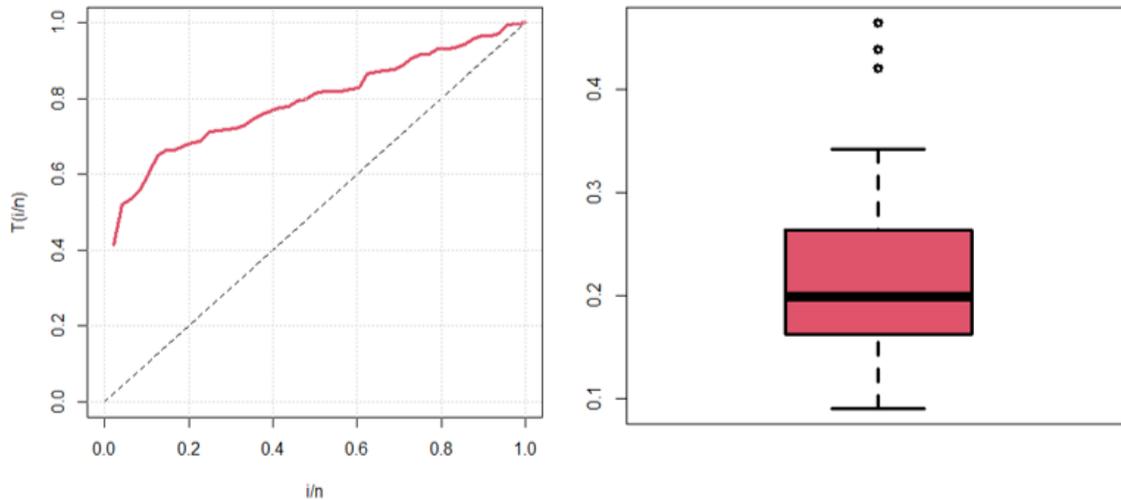
Figure 6: Shows the TTT and Box Plot of NEK from Data set 2.
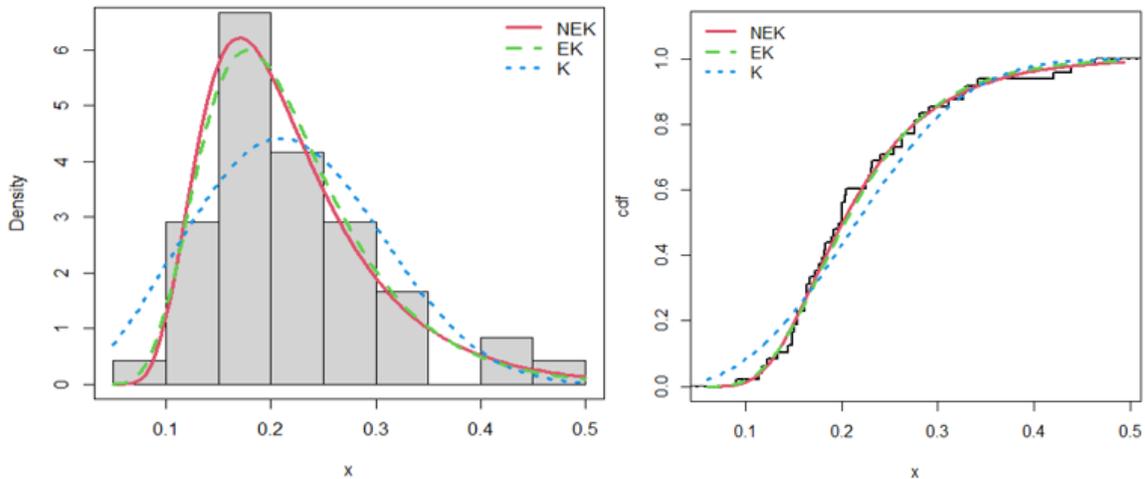
Figure 7: Pdf and Cdf graphs for dataset 2.

The BIC, CAIC, HQIC, KS statistics, its p-value (KS p-value), A∗, and W∗ statistics are some of the metrics used to compare how well these distributions fit. Tables 4, 5, 7, and 8 present the results of these measures for both data sets. The NEKD has a higher P-value and the lowest values for all goodness-of-fit measures. Compared to its rival models, the EK, TEK, and K distributions, the NEKD provides a better fit. The graphical analyses presented in Figures 5 and 7 further support this conclusion. The NEK distribution's excellent precision and general goodness of fit are demonstrated by the tight correspondence between it and the empirical data.

6. Conclusion

The Shako Kumaraswamy distribution, a versatile and all-encompassing extension of the traditional Kumaraswamy distribution, was presented in this work. A simulation is used to thoroughly examine parameter estimation, demonstrating the method's effectiveness and resilience in a range of parameter configurations and sample sizes. Additionally, the Shako Kumaraswamy distribution's improved data fitting capability over other models was demonstrated by its effective application on two empirical data sets from various sectors, demonstrating its real-world applicability. A useful addition to the family of flexible distributions, the Shako Kumaraswamy distribution has both theoretical depth and real-world applications. Future research could examine its regression frameworks, multivariate extensions, or applications in fields including engineering reliability, finance, and environmental studies.

Declaration

References

1. Jamal, F., Kanwal, S., Shafiq, S., Hashim, M., Kayid, M., Muhammad, M., ... & Shawki, A. W. (2025). The New Extended Exponentiated Burr XII distribution: Properties and applications. Journal of Radiation Research and Applied Sciences, 18(1), 101200.
2. Mudholkar, G. S., & Hutson, A. D. (1996). The exponentiated Weibull family: some properties and a flood data application. Communications in Statistics--Theory and Methods, 25(12), 3059-3083.
3. Gupta, R. D., & Kundu, D. (2007). Generalized exponential distribution: Existing results and some recent developments. Journal of Statistical planning and inference, 137(11), 3537-3547.
4. Nadarajah, S., & Kotz, S. (2006). The exponentiated type distributions. Acta Applicandae Mathematica, 92, 97-111.
5. Barreto-Souza, W., & Cribari-Neto, F. (2009). A generalization of the exponential-Poisson distribution. Statistics & Probability Letters, 79(24), 2493-2500.
6. Tahir, M. H., & Cordeiro, G. M. (2016). The exponentiated exponential geometric distribution: Properties and applications. Communications in Statistics - Theory and Methods, 45(6), 1671–1685. https://doi.org/10.1080/03610926.2014.920276.
7. Lemonte, A. J., & Cordeiro, G. M. (2011). The exponentiated generalised inverse Gaussian distribution. Statistics & probability letters, 81(4), 506-517.
8. Nadarajah, S., & Kotz, S. (2003). On the exponentiated Fréchet distribution. Journal of Probability and Statistical Science, 1(2), 177–183.
9. Nadarajah, S. (2006). The exponentiated Gumbel distribution. Journal of Applied Statistics, 33(3), 277–297.
10. Rosaiah, K., Rao, G. S., & Rao, G. S. (2006). The exponentiated log-logistic distribution: Properties and applications. Journal of Statistical Planning and Inference, 136(3), 1003–1012.

11. Abd-Elfattah, A. M., & Omima, A. A. (2009). On the exponentiated generalized exponential distribution and its applications. Applied Mathematical Sciences, 3(47), 2347–2360.

12. Rényi, A. (1961). On measures of entropy and information. In J. Neyman (Ed.), Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 547–561). University of California Press.

13. Shannon, C. E. (1948). A mathematical theory of communication. The Bell system technical journal, 27(3), 379-423.

14. Dumonceaux, R., & Antle, C. E. (1973). Discrimination between the log-normal and the Weibull distributions. Technometrics, 15(4), 923-926.

Author(s) Bio

Shoaib Iqbal: S.I is a dedicated researcher in the field of Statistics. He completed his MSc and MPhil in Statistics from Quaid-i-Azam University, Islamabad. Currently, he is pursuing a PhD in Statistics at The Islamia University of Bahawalpur. His research interests lie in Randomized Response Techniques, Probability Distributions.

Muhammad Hashim: M.H is a dedicated researcher in the field of Statistics. He completed his MSc and MPhil in Statistics from the Islamia University of Bahawalpur. His research interests lie in Probability Distributions.

Farrukh Jamal: F.J is a committed and accomplished researcher in the field of Statistics. He earned his MSc, MPhil, and PhD degrees in Statistics from The Islamia University of Bahawalpur. He currently serves as an Assistant Professor and is actively involved in teaching, research, and academic leadership, Department of Statistics, Faculty of science, University of Tabuk, Tabuk, Saudi Arabia. He has supervised 10 MPhil and 8 PhD students and total publications are 288.

Syeda Raheen Ayesha Saeed: S.R.A.S is a dedicated researcher in the field of Statistics. She completed her BS and MPhil in Statistics from The Islamia University of Bahawalpur. Her research interests lie in Probability Distributions.

Muzna Sarwar: M.S. is a dedicated researcher in the field of Statistics. She completed her MSc in Statistics from BZU and MPhil in Statistics from The Islamia University of Bahawalpur. Currently, she is pursuing a PhD in Statistics at The Islamia University of Bahawalpur. Her research interests lie in Time series, Distribution Theory, Sampling techniques and Machine Learning.

Muhammad Ali Ibrahim: A.I completed his all degrees from, The Islamia University of Bahawalpur, Pakistan, currently he is perusing a PhD in IBMAS at The Islamia University of Bahawalpur, Pakistan.

Author Contribution:

1) **Shoaib Iqbal**

S.I played a key role in deriving the mathematical properties of the proposed distribution. He also contributed to the writing process by helping structure the content and present the technical details in the manuscript.

2) **Muhammad Hashim**
He assisted in writing and running the codes for practical applications of the proposed model to real data.

3) **Farrukh Jamal**
F.J provided the original idea for the project and offered valuable guidance throughout its development and contributed significantly to the development and refinement of the core idea for the paper. He also played a key role in writing and rewriting the content to enhance clarity and improve the overall quality of expression. Additionally, he was responsible for proofreading and polishing the final version to ensure accuracy, coherence, and a professional standard of presentation.

4) **Syeda Raheen Ayesha Saeed**
R.A contributed to the paper by conducting a comprehensive literature review to establish the context and relevance of the study. She also played an active role in drafting sections of the

manuscript and performed thorough proofreading to ensure clarity, coherence, and academic quality.

5) Muzna Sarwar

M.S provided valuable support to the student during the writing process. She assisted in organizing the manuscript and ensuring that the research was clearly and effectively communicated, and polishing the final version to ensure accuracy, coherence, and a professional standard of presentation.

6) Muhammad Ali Ibrahim

A.I provided valuable support to the student during the writing process.