Article

Not peer-reviewed version

Goodness-of-Fit Test for the Kumaraswamy Distribution Via Energy Distance Approach with Applications to Real Data

Joseph Njuki * and Thomas Gilbert

Posted Date: 27 March 2026

doi: 10.20944/preprints202603.2205.v1

Keywords: goodness-of-fit; Kumaraswamy distribution; energy statistics; Monte Carlo simulation; empirical distribution function (EDF) tests

Article

Goodness-of-Fit Test for the Kumaraswamy Distribution via Energy Distance Approach with Applications to Real Data

Joseph Njuki * and Thomas Gilbert

Department of Mathematics and Statistics, Coastal Carolina University, Conway, SC 29526, USA
* Correspondence: jnjuki@coastal.edu

Abstract

In this article, we develop a goodness-of-fit test for the Kumaraswamy distribution based on energy statistics. Due to the availability of its quantile (inverse) function, Kumaraswamy distribution has been shown to be the preferred alternative to the beta distribution, since both have bounded support in the $(0,1)$ interval. The proposed test procedure is simple and more powerful against general alternatives. Under different settings, simulations show that the proposed test is capable of being well controlled for any given significance (nominal) levels. In terms of power comparisons, the proposed test outperforms other existing methods in different settings. We then apply the proposed test to real datasets (underground economy index, food expenditure, and Shasta water reservoir) to demonstrate its competitiveness and usefulness.

Keywords: goodness-of-fit; Kumaraswamy distribution; energy statistics; Monte Carlo simulation; empirical distribution function (EDF) tests

1. Introduction

There is a growing need to model real-life phenomena that are bounded or constrained within a given range such as proportions, ratios, probabilities, etc. These cases arise naturally and play an important role in applied sciences such as hydrology, engineering, economics, and environmental studies. In recent years, bounded distributions such as Kumaraswamy (KW) and beta distributions have been widely used to model such phenomena with supports within the $(0,1)$ interval. The Kumaraswamy (KW) distribution is a continuous probability distribution with a bounded support in the (0,1) interval. Introduced by Kumaraswamy [1], it is a flexible alternative to the beta distribution with similar support in the $(0,1)$ interval. Kumaraswamy [1] applied this distribution to model hydrological events such as daily rainfall and daily stream flow. In such phenomena, Kumaraswamy [1,2] noted that classical probability distribution functions such as Gaussian, log-normal, beta, and empirical distributions (such as polynomial-transformed normal), do not graciously model hydrological random processes.

The probability density function (pdf) of the Kumaraswamy distribution is given as

$$f(x) = abx^{a-1}(1-x^a)^{b-1};\ \ 0 < x < 1, a, b > 0, \tag{1}$$

and its corresponding cumulative distribution function (cdf) is given as

$$F(x) = 1-(1-x^a)^b;\ \ 0 < x < 1, a, b > 0, \tag{2}$$

where a and b are both nonnegative shape parameters. The mathematical short-hand notation for the Kumaraswamy distribution is given as $X \sim KW(a,b)$. Similar to the beta distribution, Beta(a,b), the Kumaraswamy density function $f(x)$ in Eq. (1) is flexible such that $f(x)$ is unimodal if $a > 1$ and $b > 1$ and uniantimodal if $a < 1$ and $b < 1$. In addition, the density $f(x)$ is an increasing function

 doi:10.20944/preprints202603.2205.v1

in x if $a > 1$ and $b \leq 1$, a decreasing function in x if $a \leq 1$ and $b > 1$ and is constant (uniform) if $a = b = 1$, see for example [3] and [4]. The Figure 1 above shows different densities of $KW(a, b)$ for selected values of a and b. Although the Kumaraswamy and beta distributions agree in many cases, the $KW(a, b)$ distribution has some extra advantages over the beta distribution such as its simple explicit formulas for the cumulative distribution function (cdf) and the quantile function, which do not involve any complex or special functions. The quantile function of the Kumaraswamy (KW) distribution is defined as

$$F^{-1}(y) = [1 - (1 - y)^{\frac{1}{b}}]^{\frac{1}{a}}, \quad 0 < y < 1, a, b > 0. \tag{3}$$

This simplicity of the quantile distribution function in (3) has made it easier to have a simple function for the generation and simulation of random variables. In addition, the closed-form quantile formula in (3) provides the median of the distribution as follows.

$$M = \left(1 - 0.5^{\frac{1}{b}}\right)^{\frac{1}{a}}, \; a, b > 0$$

In quantile-based modeling, beta regression is based on the closed-form mean of its distribution. In this case, a robust approach to perform similar regression based on the median is impractical. See more details and discussion on regression based on mean and median in [5], [6] and [7].

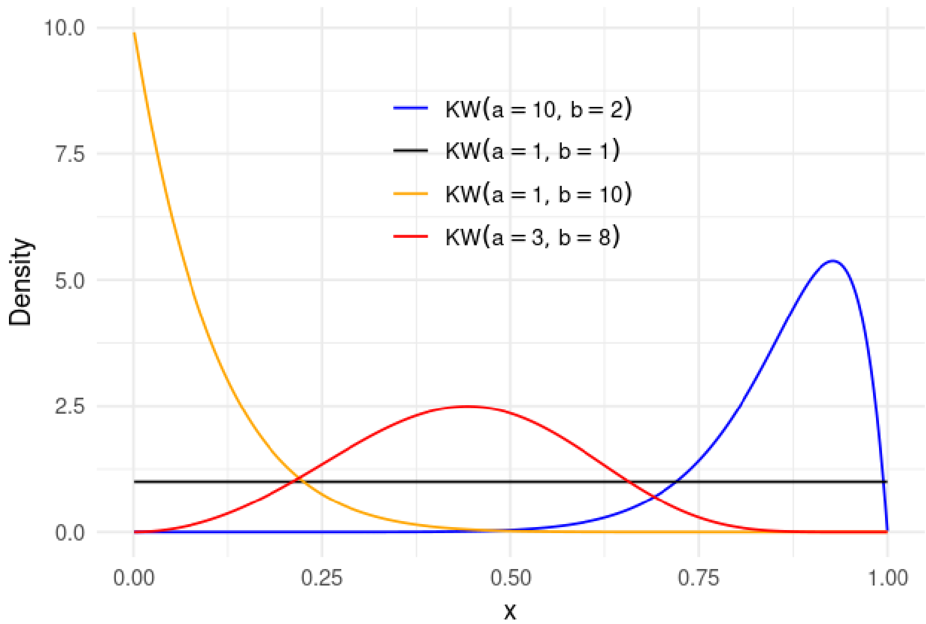


Figure 1. Kumaraswamy densities for selected values of a and b.

Jones [8] provided the r^{th} moment of the Kumaraswamy distribution for $r > -a$ as follows

$$E(X^r) = bB(1 + \frac{r}{a}, b), \tag{4}$$

where $B(.,.)$ is the complete beta function computed at shape parameters $1 + \frac{r}{a}$ and b. In addition, Jones [8] extensively studied the basic statistical properties of the Kumaraswamy (KW) distribution and provided parameter estimates using the maximum likelihood estimation (MLE) method. Nadarajah [9] noted that the Kumaraswamy (KW) distribution is actually a special case of the generalized beta distribution and showed that it is more effective than the beta distribution. Nadar et al. [10] conducted a statistical correlation analysis of the Kumaraswamy distribution for the recorded values.

We thus desire to develop the goodness-of-fit test for the Kumaraswamy distribution with well-established properties to achieve desirable powers for any given sample size based on energy statistics.

 doi:10.20944/preprints202603.2205.v1

In the literature, several goodness-of-fit tests have been developed for both simple and composite hypotheses, see for example [3,11–15] and [16]. For instance, Giles [3] came up with a testing procedure for goodness-of-fit for the Kumaraswamy distribution based on the usual classical EDF tests. Moreover, goodness-of-fit tests involving energy statistics have been proposed for some few distributions, see for example, [11,17,18], and [12]. So far, energy-based tests have been shown to be competitive when compared to other EDF-based tests such as Kolmogorov-Smirnov (KS) and Cramer-Von-Mises, see for example, [19].

In this article, we propose a one-sample (univariate) goodness-of-fit test based on energy statistics ([20]). More recently, similar work has been done by Njuki and Hasan [17], Njuki and Avallone [18], Ofosuhene [12], and Opperman and Ning [11] for the goodness-of-fit tests for the Skew-t, Lindley, Inverse gaussian, and Skew-normal, respectively, using energy statistics. In addition, there have been several studies involving energy statistics such as testing for multivariate normality ([20]), testing for equality of distributions ([21,22]), one-sample goodness-of-fit tests ([11,12,23]), change point analysis ([24–27]), among many others. For a given sequence of independent random variables of size n and with a cdf $G(x)$, the test statistic based on energy statistics will reject the null hypothesis that $F(x) = G(x)$ for large values of the test statistic. If the null distribution, $F(x)$, and the given data come from the same underlying distribution, then the values of the test statistic are expected to be very small.

The energy distance is defined as a statistical distance between the distributions of random vectors which characterizes the equality of distributions, see for example [19,28] and [29]. The concept of energy statistics initially described by [30] is based on the notion of Newton's gravitational potential energy, which is a function of the distance between two bodies; for details, see [31]. The idea of energy statistics therefore is to consider statistical observations as heavenly bodies governed by a statistical potential energy, which is zero if and only if an underlying statistical null hypothesis is true, see for example [19,31]. Śzekely and Rizzo [20,28] thus defined the energy distance between distributions of two independent (univariate) random samples X and Y with finite expectations as

$$\mathcal{E}(\mathbf{X}, \mathbf{Y}) = 2E|X - Y| - E|X - X'| - E|Y - Y'| \geq 0, \tag{5}$$

where $X \stackrel{d}{=} X'$, $Y \stackrel{d}{=} Y'$ and equality holds if and only if X and Y are identically distributed.

Many existing goodness-of-fit tests depend on the distribution function of random variables. Unlike these tests, the energy distance statistic uses the statistical distance between two observations. This gives energy statistic-based tests an invariance property with respect to any distance-preserving transformation of the dataset, see for example [19,26,31]. Energy-type tests have been shown to be typically more powerful against general alternatives than corresponding tests based on non-energy-type statistics (Kolmogorov-Smirnov, Anderson-Darling, Cramers-Von-Mises, etc). Research on goodness-of-fit tests for the Kumaraswamy distribution (KW) distribution is limited. Thus, we aim to expand the statistical research on goodness-of-fit tests for the Kumaraswamy (KW) distribution through an energy statistic-based testing procedure.

The rest of the article is organized as follows: In Section 2, we propose a one-sample (univariate) goodness-of-fit test for the Kumaraswamy (KW) distribution based on energy statistics and establish its theoretical results. We perform a simulation study in Section 3 to compare the results of our testing procedure with other existing methods in terms of Type I error and power. In Section 5, we apply our testing procedure to real-life datasets. The conclusion is provided in Section 6. The proofs of results and supplemental materials are given in the Appendix.

2. Energy Goodness-of-Fit Test for the Kumaraswamy Distribution

We propose a one-sample energy goodness-of-fit test for the Kumaraswamy distribution. The goodness-of-fit problem based on the energy distance in Eq. (5) between distributions is defined so that distributions are compared between the null (hypothesized) distribution, F_0, and the sampled distribution , F, for a given set of observations $x_1, \cdots, x_n$.

Definition 1. *Let $X_1, ...X_n$ be a random sample from a univariate population with distribution F and let $\mathbf{x} = \{x_1, ...x_n\}$ be the observed values of the random variables in the sample. Then, a one-sample energy statistic-based goodness-of-fit test for testing the hypotheses $H_0 : F = F_0$ versus $H_1 : F \neq F_0$ is given as*

$$n\mathcal{E}_n(\mathbf{x}, X) = n\left\{\frac{2}{n}\sum_{i=1}^{n} E|x_i - X| - E|X - X'| - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|\right\}, \tag{6}$$

where X and X' are independent and identically distributed random variables with distribution F_0 and the expectations are taken with respect to the null distribution F_0.

Large values of the test statistic, $n\mathcal{E}_n$, will cause us to reject the null hypothesis, $F = F_0$. Under the null hypothesis, the limiting distribution of $n\mathcal{E}_n$ is a quadratic quantity of the form $\sum_{j=1}^{\infty} \lambda_j Z_j^2$ such that Z_j, $j = 1, 2, \cdots$, are i.i.d. standard normal random variables and λ_j are nonnegative constants that depend on the null distribution F_0. Thus, the energy statistic-based goodness-of-fit test can be implemented by finding the constants λ_j. In practice, this would be difficult and we therefore resort to the use of the Monte Carlo approach to obtain empirical critical values of $n\mathcal{E}_n$ so that $P(n\mathcal{E}_n > C_\alpha) = \alpha$. This fact is guaranteed since the test based on $n\mathcal{E}_n$ is a consistent goodness-of-fit test, see for example, [20]. In order to utilize the goodness-of-fit test statistic defined in (6), we need to derive the expected distances $E|x_i - X|$ and $E|X - X'|$. The derivations of these expected values are taken with respect to the null distribution , F_0, and given in the following propositions.

Proposition 1. *Let $X \sim KW(a, b)$, then for any fixed $x \in \mathbb{R}$*

$$E|x - X| = x - 2x(1 - x^a)^b + bB(1 + \frac{1}{a}, b)(1 - 2G(x^a)), \tag{7}$$

where $B(.,.)$ is the beta function and $G(.)$ is the CDF of a Beta$(1 + \frac{1}{a}, b)$ distribution evaluated at x^a.

The proof of Proposition 1 is deferred to the Appendix.

Proposition 2. *Let X and X' be independent and identically distributed random variables with a $KW(a, b)$ distribution. Then,*

$$E|X - X'| = 2b\big(B(1 + \frac{1}{a}, b) - B(1 + \frac{1}{a}, 2b)\big) - 2ab^2 B(1 + \frac{1}{a}, b)\int_0^1 x^{a-1}(1 - x^a)^{b-1} G(x^a)dx, \tag{8}$$

where $B(.,.)$ is the beta function and $G(.)$ is the CDF of a Beta$(1 + \frac{1}{a}, b)$ distribution at x^a.

The integral part in Eq. (8) involves complex Gaussian hypergeometric functions and can be evaluated using the numerical integration approach for specified values of a and b in R software with the CDF function **pbeta()**.

Corollary 1. *For a special case in Proposition 2, we consider $a = b = 1$, then $X, X' \overset{iid}{\sim} U(0, 1)$. Thus, $E|X - X'| = \frac{1}{3}$.*

See additional details in Proposition 5. The proofs for Proposition 2 and Corollary 1 can be found in the Appendix. The above expectation can also be approximated using the following proposition suggested by [11].

Proposition 3. *The expectation of $|X - X'|$ in Proposition (2) can be approximated as follows. Let X and X' be independent and identically distributed random variables with a well-defined cumulative distribution function, $F(x)$. Given that the quantile or inverse CDF function of X exists, we have*

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2205.v1

$$E|X - X'| = \frac{4}{m}\sum_{i=1}^{m} x_i F^{-1}(x_i) - \frac{2}{m}\sum_{i=1}^{m} F^{-1}(x_i), \tag{9}$$

where m is the number of equally sized subintervals of $[0,1]$, x_i is chosen from the i^{th} subinterval and $F^{-1}(.)$ is defined in Eq. (3).

Proof. The proof of Proposition 3 is provided in Proposition 2.4 of Opperman and Ning [11]. □

To illustrate this through simulation, we compared the exact and simulated values of $E|X - X'|$ for selected values of nonnegative shape parameters a and b, and increasing values of subintervals, m from 10 to 10^5. Table A1 (in the Appendix) provides a comparison between formulas in Propositions 2 and 3, and we observe that exact values are in agreement with the simulated values of $E|X - X'|$ for a large number of subintervals (when $m = 10^5$). Rizzo [32] performed a linearization of the third term of $n\mathcal{E}_n$ in Eq. (6) to improve the computation speed of the test during intensive simulations and applications as given in Proposition 4.

Proposition 4. *Let $X_1, \cdots, X_n$ be a random sample from the distribution F and $X_{(1)}, \cdots, X_{(n)}$ be the ordered sample. Then,*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} |X_i - X_j| = 2\sum_{k=1}^{n}\big((2k-1) - n\big)X_{(k)}.$$

The proof of Proposition 4 is provided in Proposition 3.3 of Ofosuhene [12] and Rizzo [32]. We thus define the one-sample energy statistic-based goodness-of-fit test for Kumaraswamy (KW) distribution stated in Eq. (6) combined with Propositions 1, 2 and 4 as follows.

$$\begin{aligned}\mathcal{K}_n =& n\mathcal{E}_n(\mathbf{x}, X) = n\Bigg\{\frac{2}{n}\sum_{i=1}^{n}\Big[x_i - 2x_i(1-x_i^a)^b + bB(1+\frac{1}{a}, b)(1 - 2G(x_i^a))\Big] \\ &- \Big[2b\big(B(1+\frac{1}{a}, b) - B(1+\frac{1}{a}, 2b)\big) - 2ab^2B(1+\frac{1}{a}, b)\int_0^1 x^{a-1}(1-x^a)^{b-1}G(x^a)dx\Big] \\ &- \frac{2}{n^2}\sum_{k=1}^{n}\big((2k-1) - n\big)x_{(k)}\Bigg\},\end{aligned} \tag{10}$$

where $\mathbf{x} = \{x_1, \cdots, x_n\}$ is a sample of observed values, $B(.,.)$ is the beta function and $G(.)$ is the complete beta distribution function (CDF) with shape parameters $1 + \frac{1}{a}$ and b evaluated at x^a, $0 < x < 1$.

Proposition 5. *We consider the following special cases for the Kumaraswamy distribution when the shape parameters are varied.*

1. *Let $X \sim KW(1,1)$. Then X is a standard uniform random variable, i.e, $X \sim U(0,1)$.*
2. *Let $X \sim KW(1,b)$, $b > 0$. Then X follows the Beta$(1,b)$ distribution.*
3. *Let $X \sim KW(a,1)$. Then X is a beta random variable with shape parameters $\alpha = a > 0$ and $\beta = 1$.*

Proof of Proposition 5.

1. Let $X \sim KW(1,1)$. Then $f(x) = 1$, $0 < x < 1$. Thus, $X \sim U(0,1)$.
2. Let $X \sim KW(1,b)$. Then $f(x) = b(1-x)^{b-1}$, $0 < x < 1$. Thus, $X \sim$ Beta$(1,b)$.
3. Let $X \sim KW(a,1)$. Then $f(x) = ax^{a-1}$, $0 < x < 1$. Thus, $X \sim$ Beta$(a,1)$.

□

3. Simulation Study

In this section, we perform extensive Monte Carlo simulations to investigate the finite sample performance of the proposed test procedure based on energy statistics. First, we determine the ability of the test to control the Type I error for the Kumaraswamy distribution when different values of the shape parameters a and b are considered. Then we compare the power of our proposed test to other

 doi:10.20944/preprints202603.2205.v1

existing well-known EDF tests such as the Kolmogorov-Smirnov and Cramer-von-Mises tests against different alternatives at various chosen parameters and sample sizes. Throughout this paper, we take $B = 10,000$ number of repetitions.

3.1. Empirical Critical Values and Type I Errors

In order to find Type I error rates and powers, we need to obtain empirical critical values by computing a 95% quantile of energy goodness-of-fit test statistics using Eq. (10). To do this, we consider a $KW(3,8)$ distribution with nominal levels at $\alpha = 0.01, 0.05, 0.10$ and sample sizes $n = 10, 25, 50, 75, 100, 150, 200, 300, 500$. The resulting empirical critical values are reported in Table 1. To investigate the capability of our proposed testing procedure in controlling Type I error under the assumption of $KW(a,b)$ distribution, we will consider four (4) different types of Kumaraswamy distribution as follows: $KW(3,8), KW(1,1), KW(1,10)$, and $KW(10,2)$ and their estimated density curves are presented in Figure 1.

We narrowed our consideration to sample sizes $n = 10, 25, 50, 100, 200$ at the levels of significance $\alpha = 0.01, 0.05, 0.10$. Data samples are generated based on each combination of distribution and sample size, and the energy goodness-of-fit test statistic is calculated using the formula in Eq. (10). This process is repeated for B simulations. The empirical Type I error, defined as the probability of rejecting a null hypothesis that is in reality true, can then be found as the proportion of times that the computed energy goodness-of-fit test statistic exceeds the critical value. These values are reported in Table 2 and one can observe that the Type I error rate is controlled at the chosen level of significance for different choices of shape parameters a and b, and sample sizes.

Table 1. Simulated Critical Values for the Kumaraswamy Distribution

$\alpha \backslash n$	10	25	50	75	100	150	200	300	500
0.01	0.7126	0.7143	0.6741	0.6771	0.7011	0.6874	0.7030	0.6791	0.6927
0.05	0.4343	0.4510	0.4448	0.4395	0.4542	0.4531	0.4427	0.4487	0.4430
0.10	0.3334	0.3473	0.3367	0.3395	0.3410	0.3424	0.3404	0.3430	0.3394

Table 2. Simulated Type I Errors of the test for the Kumaraswamy Distribution, $KW(a,b)$

	$a = 3, b = 8$			$a = 1, b = 1$		
n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
10	0.0110	0.0541	0.1031	0.0099	0.0476	0.0995
25	0.0111	0.0529	0.1028	0.0105	0.0505	0.0980
50	0.0093	0.0502	0.1003	0.0104	0.0467	0.0989
100	0.0104	0.0483	0.0995	0.0100	0.0519	0.1007
200	0.0106	0.0507	0.1000	0.0093	0.0473	0.1008
	$a = 1, b = 10$			$a = 10, b = 2$		
n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
10	0.0116	0.0534	0.1019	0.0087	0.0513	0.1001
25	0.0099	0.0507	0.0971	0.0091	0.0495	0.1047
50	0.0097	0.0502	0.1004	0.0097	0.0503	0.0967
100	0.0102	0.0483	0.0998	0.0102	0.0500	0.0917
200	0.0106	0.0526	0.1017	0.0106	0.0463	0.0978

3.2. Power Comparisons

To assess the performance and effectiveness of the proposed energy goodness-of-fit test, we compared our test with other similar existing tests based on empirical distribution functions (EDF). These five well-known EDF test statistics are the Kolmogorov-Smirnov (KS) statistic, the Kuiper (V) statistic ([33]), the Cramér-von Mises (W^2) statistic, the Watson (U^2) statistic ([34]), and the Anderson-Darling (A^2) statistic ([35]). Stephens [36] and D'Agostino and Stephens [37] provided detailed and thorough descriptions of and applications of these different EDF tests. Seven (7) different distributions have been chosen, and samples of five different sizes will be repeatedly run through each test in order to obtain their empirical powers. The procedure for obtaining empirical powers is described below.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 March 2026 doi:10.20944/preprints202603.2205.v1

1. Calculate the critical value by computing a 95% quantile of the energy goodness-of-fit test statistic given in Eq. (10) while assuming that the null distribution (KW) is true.
2. Generate a set of data $x_1, \cdots, x_n$ from one of the specified alternative distributions.
3. Using the *mlkumar()* function of the *univariateML* package in R, we obtain maximum likelihood estimates $\hat{a}$ and $\hat{b}$ of the shape parameters a and b by treating the simulated data as if they were from a $KW(a, b)$ distribution .
4. Using Eq. (10), compute the energy goodness-of-fit statistic for the simulated data.
5. Compare the resulting energy goodness-of-fit statistic and the critical value in step 1, and determine whether or not the energy goodness-of-fit statistic exceeds the critical value.
6. Repeat this process for B times and record the results.

Under the above simulation, the empirical power of the test can then be obtained as the proportion of the number of times the test statistic is greater than the critical value. The empirical powers of the EDF tests considered in this study are calculated in a similar way. The results are reported in Tables 3 and 4 and nominal levels taken as $\alpha = 0.05$ and 0.10. We note that all considered methods are consistent and their powers increase with the increase in sample sizes. For both Tables 3 and 4, our proposed test has outperformed other existing tests against the entire set of alternative distributions in consideration. The Kuiper (V) test statistic consistently gave the lowest power. The Kolmogorov-Sminorv (KS) test statistic is not much better, even falling behind the Kuiper test statistic occasionally. The Cramér-von-Mises and Watson test statistics show noticeably competitive results. The Anderson-Darling statistic appears to be the most effective of the standard EDF tests, showing higher power than the other four nearly all times. However, we can see that for large samples, the energy goodness-of-fit and Anderson-Darling tests are equally competitive.

Table 3. Simulated powers, $\alpha = 5\%$

Distribution	Sample size n	$\mathcal{K}_n$	KS	V	W^2	U^2	A^2
Beta (5,5)	10	0.0646	0.0536	0.0565	0.0557	0.0590	0.0525
	25	0.0682	0.0618	0.0645	0.0637	0.0609	0.0633
	50	0.0752	0.0676	0.0627	0.0686	0.0646	0.0672
	100	0.1122	0.0872	0.0754	0.0955	0.0878	0.1030
	200	0.1792	0.1329	0.1104	0.1631	0.1324	0.1706
Triangular (a = 0, b = 1) (Mode = 1/3)	10	0.2026	0.0731	0.0614	0.0772	0.0678	0.0715
	25	0.2449	0.0957	0.0801	0.1077	0.0917	0.0998
	50	0.3175	0.1481	0.1161	0.1591	0.1382	0.1489
	100	0.4731	0.2407	0.2077	0.2767	0.2401	0.2701
	200	0.6942	0.4171	0.3716	0.5026	0.4471	0.4847
Truncated Normal (a = 0, b = 1) (Mean = 0.2, SD = 5)	10	0.3823	0.0722	0.0522	0.0721	0.0590	0.0587
	25	0.3729	0.0758	0.0528	0.0729	0.0572	0.0613
	50	0.3736	0.0710	0.0488	0.0676	0.0549	0.0587
	100	0.3910	0.0771	0.0559	0.0713	0.0548	0.0662
	200	0.3926	0.0698	0.0461	0.0647	0.0474	0.0571
Trapezoid ($m_1 = 1/4, m_2 = 3/4$) ($n_1 = n_3 = 3$)	10	0.1664	0.0598	0.0640	0.0620	0.0636	0.0569
	25	0.2067	0.0705	0.0869	0.0790	0.0795	0.0737
	50	0.2767	0.0892	0.1181	0.1189	0.1260	0.1124
	100	0.4221	0.1416	0.1951	0.2047	0.2223	0.2166
	200	0.6451	0.2895	0.3891	0.4076	0.4278	0.4296
Truncated Log-Normal (meanlog=0.5, sdlog=0.5)	10	0.0890	0.0775	0.0610	0.0775	0.0647	0.0651
	25	0.1023	0.0872	0.0603	0.0910	0.0684	0.0809
	50	0.1340	0.1100	0.0764	0.1189	0.0866	0.1093
	100	0.2165	0.1640	0.1055	0.1860	0.1271	0.1799
	200	0.3684	0.2714	0.1661	0.3274	0.2185	0.3274
Truncated Gamma ($\alpha = 2, \theta = 6$)	10	0.3290	0.1612	0.1231	0.1812	0.1480	0.1741
	25	0.5564	0.3312	0.2775	0.3815	0.3184	0.3978
	50	0.7652	0.5353	0.4795	0.6145	0.5296	0.6445
	100	0.9321	0.7844	0.7556	0.8602	0.7980	0.8799
	200	0.9953	0.9731	0.9636	0.9880	0.9771	0.9908
Truncated Weibull ($\lambda = 2$, k = 1)	10	0.2561	0.0756	0.0554	0.0692	0.0576	0.0640
	25	0.2903	0.0854	0.0554	0.0879	0.0647	0.0784
	50	0.3090	0.0897	0.0608	0.0924	0.0706	0.0834
	100	0.3664	0.1218	0.0816	0.1326	0.1021	0.1212
	200	0.4664	0.1757	0.1129	0.2003	0.1362	0.1889

Table 4. Simulated powers, $\alpha = 10\%$

Distribution	Sample size n	$\mathcal{K}_n$	KS	V	W^2	U^2	A^2
Beta(5,5)	10	0.1207	0.1055	0.1113	0.1077	0.1056	0.1043
	25	0.1258	0.1163	0.1113	0.1167	0.1168	0.1184
	50	0.1394	0.1254	0.1211	0.1287	0.1225	0.1310
	100	0.1816	0.1513	0.1293	0.1642	0.1469	0.1680
	200	0.2679	0.2144	0.1925	0.2424	0.2160	0.2548
Triangular (a=0, b=1) (Mode=1/3)	10	0.2924	0.1311	0.1139	0.1309	0.1230	0.1244
	25	0.3557	0.1692	0.1465	0.1776	0.1637	0.1653
	50	0.4383	0.2343	0.2019	0.2487	0.2180	0.2371
	100	0.5838	0.3401	0.3021	0.3888	0.3509	0.3686
	200	0.7747	0.5370	0.4926	0.6090	0.5581	0.5921
Truncated Normal (a = 0, b = 1) (Mean = 0.2, SD = 0.5)	10	0.5150	0.1309	0.1040	0.1307	0.1105	0.1191
	25	0.5114	0.1352	0.0996	0.1281	0.1104	0.1168
	50	0.5031	0.1246	0.0990	0.1320	0.1094	0.1127
	100	0.5194	0.1353	0.1078	0.1365	0.1117	0.1166
	200	0.5152	0.1307	0.0931	0.1258	0.1037	0.1131
Trapezoid ($m_1 = 1/4, m_2 = 3/4$) ($n_1 = n_3 = 3$)	10	0.2687	0.1198	0.1216	0.1247	0.1245	0.1128
	25	0.3233	0.1436	0.1528	0.1494	0.1570	0.1388
	50	0.4020	0.1691	0.2058	0.2069	0.2151	0.2090
	100	0.5541	0.2550	0.3070	0.3214	0.3341	0.3323
	200	0.7551	0.4364	0.5194	0.5553	0.5650	0.5793
Truncated Log-Normal (meanlog=0.5, sdlog=0.5)	10	0.1490	0.1419	0.1120	0.1319	0.1204	0.1211
	25	0.1740	0.1547	0.1201	0.1567	0.1326	0.1418
	50	0.2210	0.1924	0.1355	0.2006	0.1580	0.1909
	100	0.3092	0.2524	0.1791	0.2830	0.2143	0.2684
	200	0.4804	0.3942	0.2643	0.4494	0.3343	0.4360
Truncated Gamma ($\alpha = 2, \theta = 6$)	10	0.4121	0.2367	0.1975	0.2523	0.2157	0.2453
	25	0.6325	0.4262	0.3692	0.4718	0.4144	0.4887
	50	0.8140	0.6290	0.5807	0.6968	0.6297	0.7184
	100	0.9507	0.8486	0.8240	0.9009	0.8570	0.9142
	200	0.9967	0.9860	0.9802	0.9929	0.9855	0.9950
Truncated Weibull ($\lambda = 2$, k = 1)	10	0.3879	0.1347	0.1116	0.1328	0.1168	0.1233
	25	0.4089	0.1474	0.1134	0.1484	0.1225	0.1337
	50	0.4297	0.1546	0.1172	0.1580	0.1272	0.1471
	100	0.4869	0.1960	0.1414	0.2037	0.1655	0.1988
	200	0.5769	0.2746	0.1961	0.2939	0.2249	0.2787

4. Discussion

5. Applications

In this section, we demonstrate the effectiveness of the proposed energy goodness-of-fit test using three real-life datasets. The first dataset consists of proportions of income spent on food for a random sample of 38 households in a large US city. This dataset is available in the "*betareg*" package in R software, see for example Cribari-Neto and Zeileis [38]. The second data set is comprised of proportions of GDPs of different countries that are considered to be part of the "hidden economy" and was analyzed by Medina and Schneider [39]. The last dataset consists of the monthly water capacity from the Shasta reservoir in California, USA. The dataset is recorded for the month of February from 1991 to 2010 and can be found in both Tian et al. [4] and Sultana et al. [40]. Since the underlying distribution of these datasets is not known, we resulted to the use of a bootstrap algorithm to approximate the p-value corresponding to our proposed test and that of other tests considered in the study. The procedure is described below.

1. Estimate the parameters a and b using the *mlkumar()* function in the *univariateML* package in R by fitting the real data $y_1, \cdots, y_n$ to a Kumaraswamy distribution, where n is the size of the dataset.
2. Compute the energy goodness-of-fit statistic for the data using the formula in Eq. (10) and denote this value as $\mathcal{K}_n^1$.
3. Using the parameter estimates $\hat{a}$ and $\hat{b}$ obtained in Step 1, simulate a $KW(\hat{a}, \hat{b})$ dataset.
4. Compute the energy goodness-of-fit statistic for the simulated data using Eq. (10), and denote this value as $\mathcal{K}_n^{*1}$.
5. Repeat Steps 3 and 4 for B times, and obtain $\mathcal{K}_n^{*1}, \mathcal{K}_n^{*2}, \cdots, \mathcal{K}_n^{*B}$

6. The bootstrap p-value can then be approximated as

$$\hat{p} = \frac{1}{B} \sum_{b=1}^{B} I(\mathcal{K}_n^{*b} > \mathcal{K}_n^1) \tag{11}$$

where $I(.)$ is an indicator function such that it is 1 when $\mathcal{K}_n^{*b} > \mathcal{K}_n^1$ and 0 otherwise.

A similar procedure is conducted for the EDF tests. The results are reported in Tables A2–A4 in the Appendix. For the food expenditure dataset, our test statistic $\mathcal{K}_n = 0.0996$, which results in a p-value of 0.4466. This leads to not rejecting the null hypothesis that the data can be modeled by a Kumaraswamy distribution. The EDF tests also fail to reject the null hypothesis and thus the data follow a Kumaraswamy distribution. The estimated shape parameters are given as $\hat{a} = 2.9545$ and $\hat{b} = 26.9653$, see Figure A1a.

For the hidden economy dataset, we randomly sampled 250 data observations from the population of 2529 values using the R function command *Sample()* and the *seed* set to 123 for reproducibility purposes. Our test statistic $\mathcal{K}_n = 0.0922$, which results in a p-value of 0.6014. We thus do not reject the null hypothesis, and the data follow a Kumaraswamy distribution. The EDF tests also fail to reject the null hypothesis that the data can be modeled by a Kumaraswamy distribution. The estimated shape parameters are given as $\hat{a} = 2.8833$ and $\hat{b} = 15.4090$, see Figure A2a.

For the Shasta reservoir dataset, our test statistic $\mathcal{K}_n = 0.2147$, which results in a p-value of 0.1576. We therefore conclude that this dataset can be modeled by a Kumaraswamy distribution. The estimated shape parameters for the Shasta dataset are given as $\hat{a} = 6.3475$ and $\hat{b} = 4.9893$, see Figure A3a. In a similar manner, EDF tests result in p-values that conclude the Shasta data follow a Kumaraswamy distribution at the 5% level of significance. We can also observe that the Kumaraswamy density estimates and empirical distribution functions for the three real datasets considered in our study fit sufficiently well, see Figures A1, A2 and A3 available in the Appendix.

6. Conclusion

In this paper, we have proposed a new goodness-of-fit test based on energy statistics for the Kumaraswamy distribution. Through rigorous simulations, our proposed test demonstrates clear effectiveness in terms of its ability to control Type I error rate and power. In power comparisons, our proposed test outperforms other similar existing classical tests (non-energy) such as Anderson-Darling and Watson test statistics, especially when sample sizes are considerably small. These properties make it useful for goodness-of-fit testing for the Kumaraswamy distribution against a wide range of possible alternatives.

When dealing with a similar distribution such as the beta distribution as an alternative, it should be noted that results should be taken with caution due to the inherent similarities between the two distributions. The proposed test is then applied to real datasets to show its usefulness and applicability.

Author Contributions: Conceptualization, J.N.; methodology, J.N. and T.G.; software, J.N. and T.G.; validation, J.N. and T.G.; formal analysis, J.N. and T.G.; investigation, J.N. and T.G.; resources, J.N. and T.G.; data curation, J.N. and T.G.; writing—original draft preparation, J.N. and T.G.; writing—review and editing, J.N. and T.G.; visualization, J.N. and T.G.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: The source of each dataset is provided in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

We present proofs of new results stated in the paper and supplementary materials from our simulations and data applications.

 doi:10.20944/preprints202603.2205.v1

Table A1. Comparisons of results from the Propositions 2 and 3

	$a=3,b=8$		$a=1,b=1$		$a=1,b=10$		$a=10,b=2$	
m	Exact	Sim	Exact	Sim	Exact	Sim	Exact	Sim
25	0.1702	0.1434	0.3333	0.2944	0.0866	0.0710	0.0967	0.0799
50	0.1702	0.1558	0.3333	0.3136	0.0866	0.0779	0.0967	0.0874
10^2	0.1702	0.1626	0.3333	0.3234	0.0866	0.0819	0.0967	0.0917
10^3	0.1702	0.1693	0.3333	0.3323	0.0866	0.0860	0.0967	0.0961
10^4	0.1702	0.1701	0.3333	0.3332	0.0866	0.0865	0.0967	0.0967
10^5	0.1702	0.1702	0.3333	0.3333	0.0866	0.0866	0.0967	0.0967

Table A2. Computed test statistics and p-values of Food Expenditure data

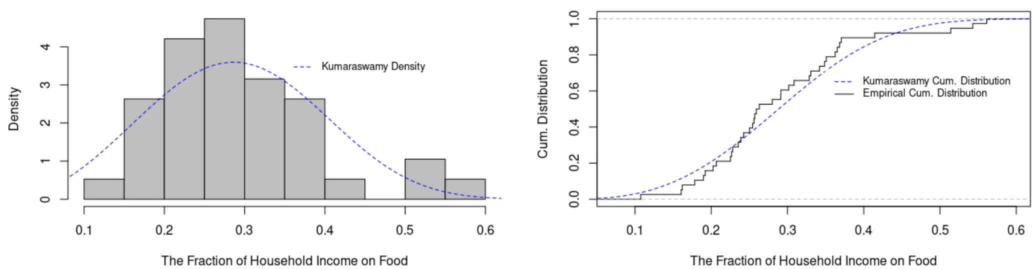
Test Statistic	$n\mathcal{E}_n$	KS	V	W^2	U^2	A^2
Statistic Value	0.0996	0.7781	1.3436	0.1267	0.1129	0.8997
p-value	0.4466	0.5536	0.3124	0.4726	0.2192	0.4224

Table A3. Computed test statistics and p-values of Gini Index for Hidden Economies data

Test Statistic	$n\mathcal{E}_n$	KS	V	W^2	U^2	A^2
Statistic Value	0.0922	0.9782	1.4832	0.1178	0.1161	0.6301
p-value	0.6014	0.2996	0.1898	0.5302	0.2124	0.6152

Table A4. Computed test statistics and p-values of Shasta Water Reservoir data

Test Statistic	$n\mathcal{E}_n$	KS	V	W^2	U^2	A^2
Statistic Value	0.2147	0.2209	1.6915	0.2568	0.1845	1.5988
p-value	0.1576	0.2338	0.0646	0.1758	0.0716	0.1604

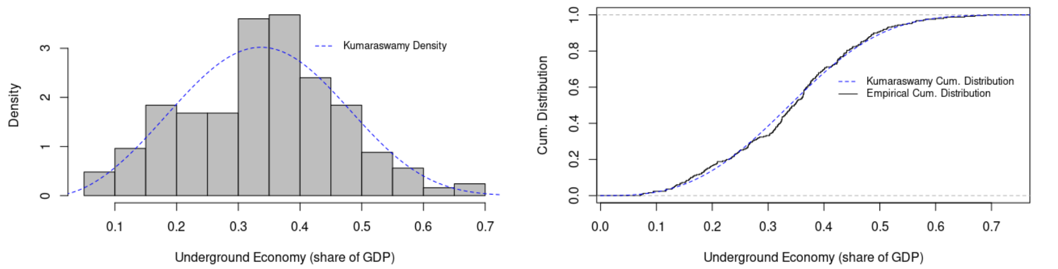


(**a**) Histogram of waiting time

(**b**) Empirical CDF and estimated Kumaraswamy CDF.

Figure A1. Food Expenditure

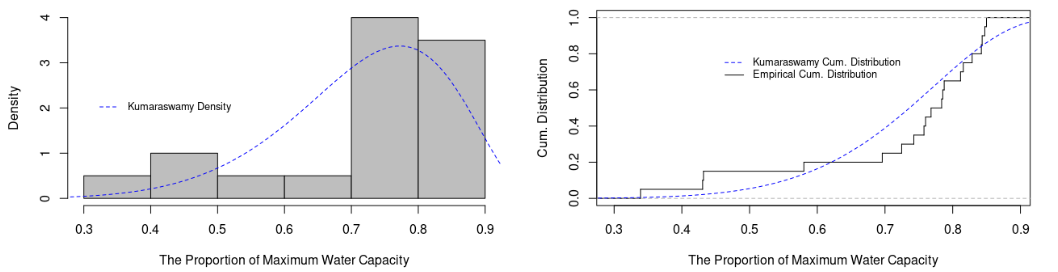
 doi:10.20944/preprints202603.2205.v1



(**a**) Histogram of hidden/underground economy data, with estimated Kumaraswamy density.

(**b**) Empirical CDF and estimated Kumaraswamy CDF.

Figure A2. Hidden Economy



(**a**) Histogram of Shasta reservoir economy data, with estimated Kumaraswamy density.

(**b**) Empirical CDF and estimated Kumaraswamy CDF.

Figure A3. Shasta Reservoir

Proof of Proposition 1. Let $X \sim KW(a,b)$. Then for any fixed $x \in \mathbb{R}$,

$$\begin{aligned} E|x - X| &= \int_0^1 |x-y| f_X(y)dy \\ &= \int_0^x (x-y)f_X(y)dy + \int_x^1 (y-x)f_X(y)dy \\ &= xF_X(x) - \int_0^x yf_X(y)dy + \left[\int_0^1 yf_X(y)dy - \int_0^x yf_X(y)dy\right] - x[1-F_X(x)] \\ &= 2xF_X(x) - x + E(X) - 2\int_0^x yf_X(y)dy \end{aligned}$$

We evaluate the last integral term as follows.

$$\begin{aligned} \int_0^x yf_X(y)dy &= \int_0^x yaby^{a-1}(1-y^a)^{b-1}dy \\ & \quad \text{Apply the U-substitution by letting } u = 1-y^a. \\ &= b\int_{1-x^a}^1 (1-u)^{\frac{1}{a}} u^{b-1}du \\ &= b\int_0^{x^a} w^{\frac{1}{a}}(1-w)^{b-1}dw, \quad \text{letting } \; w = 1-u \\ &= bB(1+\frac{1}{a},b)\int_0^{x^a} \frac{1}{B(1+\frac{1}{a},b)} w^{\frac{1}{a}}(1-w)^{b-1}dw \\ &= bB(1+\frac{1}{a},b)G(x^a) \end{aligned}$$

where $B(.,.)$ is the beta function and $G(.)$ is the complete beta distribution function with parameters $1+\frac{1}{a}$ and b at x^a. Then, with $F(x)$ as given in Eq. 2 and $E(X) = bB(1+\frac{1}{a}, b)$ when $r = 1$ by the formula in Eq. (4), we get:

$$\begin{aligned} E|x - X| =& 2x(1-(1-x^a)^b) - x + bB(1+\frac{1}{a}, b) - 2bB(1+\frac{1}{a}, b)G(x^a) \\ =& x - 2x(1-x^a)^b + bB(1+\frac{1}{a}, b)(1-2G(x^a)) \end{aligned}$$

where $B(.,.)$ is the beta function and $G(x^a)$ is the CDF of a Beta$(1+\frac{1}{a}, b)$ distribution evaluated at x^a. □

Proof of Proposition 2. Let X and X' be independent and identically distributed random variables from a $KW(a,b)$ distribution. Then,

$$\begin{aligned} E|X - X'| =& \int_0^1 \int_0^1 |x-y| f(x) f(y) dy dx \\ =& 2\int_0^1 \int_0^x (x-y) f(x) f(y) dy dx, \text{ by the symmetry.} \\ =& 2\int_0^1 \int_0^x x f(x) f(y) dy dx - 2\int_0^1 \int_0^x y f(x) f(y) dy dx \\ =& 2A - 2B \end{aligned}$$

We proceed to evaluate these two integrals (A & B) separately as follows.

$$\begin{aligned} A =& \int_0^1 \int_0^x x f(x) f(y) dy dx \\ =& \int_0^1 x f(x) \int_0^x ab y^{a-1}(1-y^a)^{b-1} dy dx \\ =& \int_0^1 x f(x)(1-(1-x^a)^b) dx, \text{ by the defition of } F(x) \text{ in Eq. 2} \\ =& ab\int_0^1 x^a(1-x^a)^{b-1}(1-(1-x^a)^b) dx \\ =& ab\int_0^1 x^a(1-x^a)^{b-1} dx - ab\int_0^1 x^a(1-x^a)^{2b-1} dx \\ =& C - D. \end{aligned}$$

We use the U-substitution method to evaluate integrals for C and D.

$$\begin{aligned} C =& ab\int_0^1 x^a(1-x^a)^{b-1} dx, \text{ let } u = 1-y^a \\ =& b\int_0^1 (1-u)^{\frac{1}{a}} u^{b-1} du \\ =& bB(b, 1+\frac{1}{a}) \underbrace{\int_0^1 \frac{1}{B(b,1+\frac{1}{a})} u^{b-1}(1-u)^{\frac{1}{a}} du}_{Beta(b,1+\frac{1}{a}) \text{ integrates to } 1} \\ =& bB(b, 1+\frac{1}{a}) \\ =& bB(1+\frac{1}{a}, b), \text{by the symmetry of the beta distribution.} \end{aligned}$$

Similarly, $D = ab\int_0^1 x^a(1-x^a)^{2b-1}dx = bB(1+\frac{1}{a}, 2b)$ by letting $u = 1 - x^a$ and apply the above solution. Thus, the integral in A becomes

$$A = C - D = bB(1+\frac{1}{a}, b) - bB(1+\frac{1}{a}, 2b) = b\big(B(1+\frac{1}{a}, b) - B(1+\frac{1}{a}, 2b)\big).$$

We proceed to find the integral for B. First, notice that

$$\begin{aligned}\int_0^x yf(y)dy =& ab\int_0^x y^a(1-y^a)^{b-1}dy, \quad \text{let } u = 1-y^a \text{ and apply U-substitution}\\ =& b\int_{1-x^a}^1 (1-u)^{\frac{1}{a}}u^{b-1}du\end{aligned}$$

By letting $z = 1 - u$ and substitute this in the last integral, we obtain

$$\begin{aligned}b\int_{1-x^a}^1 (1-u)^{\frac{1}{a}}u^{b-1}du =& b\int_0^{x^a} z^{\frac{1}{a}}(1-z)^{b-1}dz\\ =& bB(1+\frac{1}{a}, b)\int_0^{x^a} \frac{1}{B(1+\frac{1}{a}, b)}z^{\frac{1}{a}}(1-z)^{b-1}dz\\ =& bB(1+\frac{1}{a}, b)G(x^a),\end{aligned}$$

where $G(.)$ is the CDF of the beta distribution with parameters $1+\frac{1}{a}$ and b. The final integral to evaluate for B simplifies to the following expression.

$$\begin{aligned}B =& \int_0^1\int_0^x yf(x)f(y)dydx = \int_0^1 f(x)\int_0^x yf(y)dy\\ =& \int_0^1 abx^{a-1}(1-x^a)^{b-1}bB(1+\frac{1}{a}, b)G(x^a)dx\\ =& ab^2B(1+\frac{1}{a}, b)\int_0^1 x^{a-1}(1-x^a)^{b-1}G(x^a)dx\end{aligned}$$

Finally, we plug in the expressions for A and B to obtain

$$E|X - X'| = 2A - 2B \tag{A1}$$
$$= 2b\big(B(1+\frac{1}{a}, b) - B(1+\frac{1}{a}, 2b)\big) - 2ab^2B(1+\frac{1}{a}, b)\int_0^1 x^{a-1}(1-x^a)^{b-1}G(x^a)dx, \tag{A2}$$

where $G(.)$ is the CDF of a Beta$(1+\frac{1}{a}, b)$ at x^1 □

Proof of Corollary 1. Let $X, X' \overset{iid}{\sim} KW(1,1) \equiv U(0,1)$. This follows immediately from the proof of Proposition 2 when $a = b = 1$ and thus

$$\begin{aligned}E|X - X'| =& 2(B(2,1) - B(2,2) - 2B(2,1)\int_0^1 G(x)dx, \quad \text{where } G(x) \equiv \text{Beta}(2,1)\\ =& 2(\frac{1}{2} - \frac{1}{6}) - \frac{2}{2}\int_0^1\{\int_0^x 2tdt\}dx\\ =& \frac{2}{3} - \int_0^1 x^2dx\\ =& \frac{2}{3} - \frac{1}{3}\\ =& \frac{1}{3}.\end{aligned}$$

□

 doi:10.20944/preprints202603.2205.v1

References

1. Kumaraswamy, P. A generalized probability density function for double-bounded random processes. *Journal of Hydrology* **1980**, *46*, 79–88.
2. Kumaraswamy, P. Snepower probability density function. *Journal of Hydrology* **1976**, *31*, 181–184.
3. Giles, D.E. New Goodness-of-Fit Tests for the Kumaraswamy Distribution. *Stats* **2024**, *7*, 373–388.
4. Tian, W.; Panf, L.; Tian, C.; Ning, W. Change Point Analysis for Kumaraswamy Distribution. *Mathematics* **2023**, *11*, 553.
5. Hamedi-Shahraki, S.; Rasekhi, A.; Yekaninejad, M.S.; Eshraghian, M.R.; Pakpour, A.H. Kumaraswamy regression modeling for Bounded Outcome Scores. *Pakistan Journal of Statistics and Operation Research* **2021**, *17(1)*, 79–88.
6. Mitnik, P.A.; Baek, S. The Kumaraswamy distribution: Median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Stat Papers* **2013**, *54*, 177–192.
7. Ferrari, S.; Cribari-Neto, F. Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics* **2004**, *31*, 799–815.
8. Jones, M. Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology* **2009**, *6*, 70–81.
9. Nadarajah, S. On the distribution of Kumaraswamy. *Journal of Hydrology* **2008**, *348*, 568–569.
10. Nadar, M.; papadopoulos, A.; Kizilaslan, F. Statistical analysis for Kumaraswamy's distribution based on record data. *Stat Papers* **2013**, *54*, 355–369.
11. Opperman, L.; Ning, W. Goodness-of-fit test for skew normality based on energy statistics. *Random Operators and Stochastic Equations* **2020**, *28*, 227–236.
12. Ofosuhene, P. The energy goodness-of-fit Test for the Inverse gaussian distribution. *Ph.D Thesis: Bowling Green State University* **2020**.
13. Rizzo, M.L.; Śzékely, G.J. Energy Distance. *WIREs Comput Stat* **2016**, *8*, 27–38.
14. Maghami, M.; Bahrami, M. Goodness of Fit Test for the Skew-T Distribution. *Journal of Mathematics and Computer Science* **2015**, *14*, 274–283.
15. Ning, W.; Ngunkeng, G. An empirical likelihood ratio based goodness-of-fit test for skew normality. *Stat Methods Appl* **2013**, *22*, 209–226.
16. Vexler, A.; Shan, G.G.; Kim, S.G.; Tsai, W.M.; Tian, L.L.; Hutson, A.D. An empirical likelihood ratio based goodness-of-fit test for Inverse gausian distributions. *Journal of Statistical Planning and Inference* **2011**, *141*, 2128–2140.
17. Njuki, J.; Hasan, A. A New Goodness-of-Fit Test for Azzalini's Skew-t Distribution Based on the Energy Distance Framework with Applications. *Mathematics* **2025**, *13*. https://doi.org/doi.org/10.3390/math13233833.
18. Njuki, J.; Avallone, R. Energy Statistic-Based Goodness-of-Fit Test for the Lindley Distribution with Application to Lifetime Data. *Stats* **2025**, *8*. https://doi.org/10.3390/stats8040087.
19. Śzékely, G.J.; Rizzo, M.L. *The Energy of Data and Distance Correlation*, 1st ed.; Chapman and Hall: London, UK, 2023.
20. Śzékely, G.J.; Rizzo, M. A new test for multivariate normality. *Journal of Multivariate Analysis* **2005**, *93*, 58–80.
21. Śzékely, G.J.; Rizzo, M.L. Testing for Equal Distributions in high Dimension. *InterStat* **2004**, *11*.
22. Rizzo, M.L. *A test of homogeneity for two multivariate populations, Physical and Engineering Sciences section.*; In: 2002 Proceedings of American Statistical Association: American Statistical Association, Alexandria, VA, 2003.
23. Rizzo, M.L. New goodness-of-fit tests for Pareto distributions. *ASTIN Bulletin: The Journal of the IAA* **2009**, *39*, 691–715.
24. Njuki, J.; Ning, W. Energy statistic-based modified information criterion for detecting the change in distribution. *Journal of Applied Statistics* **2025**, pp. 1–23.
25. Njuki, J.M. Nonparametric Sequential tests for Change Point Analysis Using Energy Statistics. *Ph.D Thesis: Bowling Green State University* **2022**.
26. Matterson, D.S.; James, N.A. A nonparametric Approach for Multiple Change Point Analysis of Multivariate Data. *Journal of the American Statistical Association* **2014**, *109*, 334–345.
27. Kim, A.Y.; Marzban, C.; Percival, D.B.; .; Stuetzle, W. Using labeled data to evaluate change detectors in a multivariate streaming environment. *Signal Processing* **2009**, *89(12)*, 2529–2536.
28. Śzékely, G.J.; Rizzo, M. A Class of Statistical Based on Distances. *Journal of Statistical Planning and Inference* **2013**, *143*, 1249–1272.

29. Rizzo, M.L. *Statistical Computing With R*, 2nd ed.; CRC Press, Taylor & Francis Group: Boca Raton, FL, 2019.
30. Szekely, G.J. E-statistics: Energy of statistical samples. *Technical Report 03-05, BGSU, Department of Mathematics and Statistics* **2000**.
31. Szekely, G.J.; Rizzo, M.L. The Energy of Data. *Annual Review of Statistics and Its Application* **2017**, *4*, 447–479.
32. Rizzo, M. A new rotation invariant goodness-of-fit test. *Ph.D Thesis: Bowling Green State University* **2002**.
33. Kuiper, N.H. Tests concerning random points on a circle. *Proceedings of the Nederlandse Akademie Van Wetenschapen, Series A* **1960**, *63*, 38–47.
34. Watson, G.S. Goodness-of-fit tests on a circle. *Biometrika* **1961**, *48*, 109–114.
35. Anderson, T.W.; Darling, D.A. A test of goodness of fit. *Journal of the American Statistical Association* **1954**, *49*, 765–769.
36. Stephens, M.A. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association* **1974**, *69*, 730–737.
37. D'Agostino, R.B.; Stephens, M.A. Tests Based on EDF Statistics. In *Goodness-of-Fit Techniques*; D'Agostino, R.B.; Stephens, M.A., Eds.; Marcel Dekker: New York, 1986; pp. 97–193.
38. Cribari-Neto, F.; Zeileis, A. Beta Regression in R. *Journal of Statistical Software* **2010**, *34*, 1–24. https://doi.org/10.18637/jss.v034.i02.
39. Medina, L.; Schneider, F.G. Shedding Light on the Shadow Economy: A Global Database and the Interaction with the Official One. *CESifo Working Paper No. 7981* **2019**. https://doi.org/http://dx.doi.org/10.2139/ssrn.3502028.
40. Sultana, F.; Tripathi, Y.M.; Wu, S.J.; Sen, T. Inference for kumaraswamy distribution based on type I progressive hybrid censoring. *Ann. Data. Sci.* **2022**, *9*, 1283–1307.