stats

MDPI

Article

Energy Statistic-Based Goodness-of-Fit Test for the Lindley Distribution with Application to Lifetime Data

Joseph Njuki * and Ryan Avallone

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

Abstract

In this article, we propose a goodness-of-fit test for a one-parameter Lindley distribution based on energy statistics. The Lindley distribution has been widely used in reliability studies and survival analysis, especially in applied sciences. The proposed test procedure is simple and more powerful against general alternatives. Under different settings, Monte Carlo simulations show that the proposed test is able to be well controlled for any given nominal levels. In terms of power, the proposed test outperforms other existing similar methods in different settings. We then apply the proposed test to real-life datasets to demonstrate its competitiveness and usefulness.

Keywords: Lindley distribution; goodness-of-fit; energy statistics; empirical distribution function (EDF) tests; exponential distribution

check for updates

Academic Editor: Wei Zhu

Received: 25 August 2025
Revised: 20 September 2025
Accepted: 24 September 2025
Published: 26 September 2025

Citation: Njuki, J.; Avallone, R. Energy Statistic-Based Goodness-of-Fit Test for the Lindley Distribution with Application to Lifetime Data. *Stats* **2025**, *8*, 87. https://doi.org/10.3390/stats8040087

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Research on reliability studies cannot be complete without mentioning exponential distributions. This has led many researchers to overlook other more flexible distributions in the analysis of lifetime data such as Gamma, Weibull, Lindley, and Log-normal distributions. The Lindley distribution was proposed by Lindley [1,2] as part of the larger exponential family and as a more superior competitor distribution for models based on the exponential distribution. In fact, Ghitany et al. [3] found that the Lindley distribution is a better alternative distribution for exponential distributions and provided a wide range of properties of the Lindley distribution such as moments, failure rate function, mean residual life function, and maximum likelihood estimation, among many others. The Lindley distribution is an example of an exponential family and has been expressed as a mixture of two well-known distributions, Exponential and Gamma distributions such that its probability distribution function (pdf), $f(x)$ in Equation (1) can be written as

$$f(x) = wf_1(x) + (1-w)f_2(x), \quad x > 0,$$

where $f_1(x) \equiv \text{Exp}(\theta)$, $f_2(x) \equiv \Gamma(2,\theta)$, $w = \frac{\theta}{1+\theta}$, and $\theta > 0$ is the scale parameter.

In recent years, studies involving Lindley distribution have gained momentum and have shown great success in areas such as modeling and analyzing survival, reliability, and failure lifetime-related data. One advantage of the Lindley distribution over the exponential distribution is the fact that it possesses an increasing hazard rate function and a decreasing mean residual life function whereas in the latter both functions are constant; see more details in Ghitany et al. [3]. These studies are very important and common in applied sciences such as construction, engineering, medicine,

insurance, banking, and many others. Several authors have studied possible applications of the Lindley distribution. Such studies stem from distribution properties (Ghitany et al. [3], Lindley [1]), bayesian estimation (Ali et al. [4], Ali et al. [4]), estimation of the stress–strength reliability parameter (Al-Mutairi et al. [5]) and load-sharing parallel system model and its application (Singh and Gupta [6]). Currently, there are several generalizations and extensions of the Lindley distribution such as the double Lindley distribution, the Kumaraswamy quasi Lindley distribution, the Log-Lindley distribution, the exponentiated power Lindley distribution, and the generalized weighted Lindley distribution, among others; see, for example, Satheesh Kumar and Jose [7], Ibrahim et al. [8], Nedjar and Zeghdoudi [9], Ramos and Louzada [10], Gómez-Déniz et al. [11], and Elbatal and Elgarhy [12].

Definition 1. *Let $X \sim Lindley(\theta)$. The pdf and cdf of X are given by*

$$f(x) = \frac{\theta}{1+\theta}(1+x)e^{-\theta x}, \quad x > 0, \quad \theta > 0 \tag{1}$$

and

$$F_X(x) = 1 - \frac{\theta + 1 + \theta x}{1+\theta}e^{-\theta x}, \quad x > 0, \quad \theta > 0, \tag{2}$$

respectively, and θ is the scale parameter.

Its mean and variance are given as

$$\text{Mean:} \quad \mu = \frac{\theta + 2}{\theta(\theta+1)} \tag{3}$$

and

$$\text{Variance:} \quad \sigma^2 = \frac{\theta^2 + 4\theta + 2}{\theta^2(\theta+1)^2} \tag{4}$$

Figure 1 provides some examples of the forms that the density function in Equation (1) can take for selected values of the scale parameter θ.

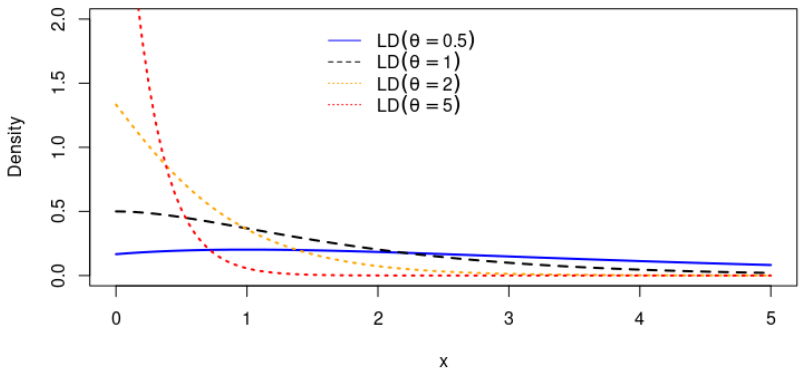


Figure 1. Lindley densities for selected values of θ.

We note that both method of moments (MoM) and maximum likelihood (ML) estimates of the scale parameter θ are the same and defined as follows.

Definition 2. *Let $X_1, X_2, \cdots, X_n$ be a random sample from the Lindley distribution, the maximum likelihood (ML) or method of moments (MoM) estimate of the scale parameter θ is given as*

$$\hat{\theta} = \frac{-(\bar{X}-1) + \sqrt{(\bar{X}-1)^2 + 8\bar{X}}}{2\bar{X}}, \quad \bar{X} > 0, \tag{5}$$

where $\bar{X}$ is the sample mean.

Ghitany et al. [3] proved that in this estimate, $\hat{\theta}$ is positively biased, that is, $E\hat{\theta} - \theta > 0$. In addition, the estimate is consistent and asymptotically normal with mean zero and standard deviation $\frac{1}{\sigma}$, where σ^2 is given in Equation (4); for further details, see, for example, Ghitany et al. [3]. We will use this estimate of the scale parameter θ in the development of the energy goodness-of-fit statistic for the Lindley distribution. These well-established properties of the Lindley distribution to provide a desirable environment for the construction of goodness-of-fit tests. Although we have seen a significant amount of research on the Lindley distribution and its applications, there is limited research in the literature on the goodness-of-fit test for the distribution. Authors such as Noughabi and Noughabi [13], Noughabi and Noughabi [14], and Noughabi [15] have proposed a goodness-of-fit test for the Lindley distribution based on estimates of the Gini index, Empirical likelihood ratio (ELR) and Kullback–Leibler (KL) discrimination, respectively. In the literature, different goodness-of-fit tests for different statistical distributions have been proposed, see, for example, Best and Rayner [16], Shan et al. [17], Vexler et al. [18], and Rizzo [19], Noughabi and Noughabi [20], among many others. For example, Vexler et al. [18] and Vexler and Gurevich [21] applied the empirical likelihood ratio to test goodness-of-fit for inverse gaussian distribution and on sample entropy, respectively. Best and Rayner [16] proposed smooth tests of fit for the Lindley and Poisson–Lindley distributions. Ning and Ngunkeng [22] and Gupta and Chen [23] proposed goodness-of-fit tests for the Skew-normal distribution based on the empirical likelihood ratio and Pearson's χ^2 tests, respectively. Other tests available in the literature are the well-known and common EDF tests, which are based on empirical distribution functions, see, for example, Stephens [24] and Stephens [25]. We thus introduce a more competitive testing procedure based on energy statistics proposed by Sźekely [26] and Rizzo [27]. Sźekely and Rizzo [28] have demonstrated extensively the superiority of energy-based tests over many other tests, such as EDF-based, Gini index and Kullback–Leibler discrimination tests.

In this article, we propose a one-sample (univariate) goodness-of-fit test based on energy statistics (Sźekely [26] and Rizzo [27]). In a more recent work, Logan and Ning [29] proposed a goodness-of-fit test using energy statistics for the Skew-normal distribution and Ofosuhene [30] developed a goodness-of-fit test based on energy statistics and distance correlation for inverse-gaussian distribution. There are a few other research works in the literature involving the energy goodness-of-fit test for some known distributions, see, for example, Sźekely and Rizzo [28] and Rizzo [19]. 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 that $F(x) = G(x)$ for large values of test statistics. If the null distribution $F(x)$ and the given data come from the same underlying distribution, then values of the test statistic are expected to be small. In addition, there are several other studies involving energy distance statistics such as testing for multivariate normality (Sźekely and Rizzo [31]), testing for equality of distributions (Sźekely and Rizzo [32], Rizzo [33]), and change point analysis (Njuki and Ning [34], Njuki [35], Matterson and James [36], Kim et al. [37]), among many others.

The energy distance is defined as a statistical distance between the distributions of random vectors which characterizes equality of distributions, see, for example, Sźekely and Rizzo [28,38,39] and Sźekely [26]. The concept of energy statistics initially described by Sźekely [40] is based on the notion of Newton's gravitational potential energy, which is a function of the distance between two bodies; for further details, see Sźekely and Rizzo [38]. 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, Sźekely and Rizzo [28,38]. In this work, we propose a procedure that is more superior for the goodness-of-fit test for the Lindley

distribution (Lindley [1])-based energy statistics. Unlike the proposed method based on energy statistics, most of the existing methods depend on the (empirical) distribution function of random variables. Energy-type tests have been shown to be typically more powerful against general alternatives than corresponding tests, especially those based on classical statistics (non-energy type) such as Kolmogorov–Smirnov, Anderson Darling (Anderson and Darling [41]), Watson statistic (Watson [42]), and Kuiper statistic (Kuiper [43]). Furthermore, energy statistic-based tests have an additional advantage such that these tests have an invariance property with respect to any distance-preserving transformation of the dataset, see, for example, Sźekely and Rizzo [28,38] and Matterson and James [36]. Tests based on energy statistics can be easily extended to multivariate and high dimension settings, see, for example, Sźekely and Rizzo [31], Rizzo [33] and Sźekely and Rizzo [32].

We organize this article as follows. In Section 2, we propose a testing procedure based on energy statistics for the goodness-of-fit test and establish its theoretical results on the Lindley distribution. We perform different simulations in Section 3 to compare the results of our energy goodness-of-fit test with other existing ones. In Section 4, we demonstrate the application of our method using three real-life datasets. The conclusion is provided in Section 5. The proofs to new results and other supplemental materials are given in the Appendix A.

2. The Energy Goodness-of-Fit Statistic for the Lindley Distribution

We propose a test procedure for the goodness-of-fit test based on the energy statistics for the Lindley distribution. Next, we define the energy goodness-of-fit test based on the characterization of equal distributions.

Definition 3. *(Energy distance). If X and Y are independent random vectors with $E(X) < \infty$ and $E(Y) < \infty$, then the energy distance between X and Y is defined as follows:*

$$\mathcal{E}(\boldsymbol{X}, \boldsymbol{Y}) = 2E|X - Y| - |X - X'| - |Y - Y'| \geq 0, \tag{6}$$

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

The energy distance between two independent distributions is thus given by Equation (6) and hence, under the null hypothesis, it is assumed that the data follow a null (Lindley) distribution against a general alternative hypothesis. Let $\mathbf{x} = \{x_1, \cdots, x_n\}$ be a sample of random observations from the distribution F and $X \sim F_0$, null distribution, then the empirical one-sample goodness-of-fit test based on energy statistics for testing $H_0 : F = F_0$ versus $H_1 : F \neq F_0$ is defined as

$$n\mathcal{E}_n(\mathbf{x}, X) = n\left\{\frac{2}{n}\sum_{i=1}^{n} E|x_i - X| - E|X - X'| - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|\right\}, \tag{7}$$

where X and X' are independently and identically distributed random variables from the null distribution, F_0. The null hypothesis, $F = F_0$, is rejected for large values of the test statistic $n\mathcal{E}_n$, where $F_0 \equiv \text{Lindley}(\theta)$ in our case. Under the null hypothesis, the limiting distribution of $n\mathcal{E}_n(\mathbf{x}, X)$ 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. Thus, the goodness-of-fit test can be implemented by finding the constants λ_j. In practice, this could be difficult, and we therefore resort to the use of the Monte Carlo simulation 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 against all general alternatives, see, for example Sźekely and Rizzo [31] and Móri et al. [44].

The establishment of the energy statistic-based goodness-of-fit test for the Lindley distribution in Equation (7) depends on the expected distances $E|x_i - X|$ and $E|X - X'|$, where X and X' are i.i.d. copies from the Lindley distribution.

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

$$E|x - X| = x + \frac{2xe^{-\theta x}}{1+\theta} + \frac{2e^{-\theta x}}{1+\theta} + \frac{4e^{-\theta x}}{\theta(1+\theta)} + \frac{\theta - 2}{\theta(1+\theta)} - \frac{2}{1+\theta} \tag{8}$$

A full proof of Proposition 1 is provided in Appendix A.

Proposition 2. *Let X and X' be independent and identically distributed random variables from a Lindley distribution with the shape parameter θ. Then,*

$$E|X - X'| = \frac{2\theta^2 + 6\theta + 3}{2\theta(1+\theta)^2} \tag{9}$$

The detailed proof of Proposition 2 involving integration by parts and expansion of the joint expectation is deferred to Appendix A.

Proposition 3. *Let $X_1, X_2, \cdots, X_n$ be a sequence of independent random variables from the null distribution F_0 and let $X_{(1)}, X_{(2)}, \cdots, X(n)$ be the ordered sample. Then, the last term of the test statistic $n\mathcal{E}_n$ in Equation (7) can be linearized as follows;*

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j| = \frac{2}{n^2}\sum_{k=1}^{n}((2k-1) - n)x_{(k)} \tag{10}$$

Proof. The proof of Proposition 3 is given by Proposition 3.3 of Ofosuhene [30] and Rizzo [27]. □

This linearization will reduce the computational complexity of the test from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$, which is useful during extensive operations and simulations. Thus, the one-sample (univariate case) energy goodness-of-fit test for the Lindley distribution based on Equation (7) together with Propositions 1, 2, and 3 is defined as

$$\begin{aligned}\mathbf{Q}_n = n\mathcal{E}_n(\mathbf{x}, X) = n\Bigg\{&\frac{2}{n}\sum_{i}^{n}\left(x_i + \frac{2x_ie^{-\theta x_i}}{1+\theta} + \frac{2e^{-\theta x_i}}{1+\theta} + \frac{4e^{-\theta x_i}}{\theta(1+\theta)} + \frac{\theta-2}{\theta(1+\theta)} - \frac{2}{1+\theta}\right)\\ &- \frac{2\theta^2+6\theta+3}{2\theta(1+\theta)^2} - \frac{2}{n^2}\sum_{k=1}^{n}((2k-1)-n)x_{(k)}\Bigg\},\end{aligned} \tag{11}$$

where $\mathbf{x} = \{x_1, \cdots, x_n\}$ is the observed sample values and $X \sim \text{Lindley}(\theta)$, $\theta > 0$.

3. Simulations and Discussions

In this section, we perform Monte Carlo simulations to investigate the performance of our proposed test in terms of controlling the Type I error and power comparisons for varying values of the scale parameter θ. In practice, the goodness-of-fit tests based on the empirical distribution function (EDF) are most commonly used; we thus compare powers of the proposed test and the EDF tests under various combinations of alternative distributions and sample sizes. These popular and well-known tests are Kolmogorov–Smirnov (KS), Anderson–Darling (A^2), Cramer-von-Mises (W^2), Watson (U^2), and Kuiper (V) statistics. Throughout this article, we consider the number of simulations B to be 10,000.

3.1. Simulated Type I Error (Size of the Test) and Empirical Critical Value

We find the empirical Type I error and critical values of the proposed test under the Lindley(θ) distribution assumption. For these simulations, we consider the nominal levels of $\alpha = 0.01, 0.05, 0.10$ and random samples of sizes $n = 10, 25, 50, 100, 150, 200, 300, 500, 1000$. The algorithm applied to find the empirical critical values and the simulated Type I error is as follows.

1. Generate lifetime dataset of size n from a Lindley (θ) distribution.
2. Compute the energy goodness-of-fit statistic for the data using the formula in Equation (11).
3. Repeat Steps 1 and 2 B times and obtain energy goodness-of-fit statistics Q_n^j, where $j = 1, 2, \cdots, B$.
4. Obtain the empirical critical values by computing a 95% quantile of energy statistics computed in Step 3 and denote it as Q_n^*.
5. Repeat Steps 1 through 4 and compute the empirical Type I error as the number of times the energy goodness-of-fit statistic Q_n^j is greater than the critical value Q_n^*.

Table 1 gives the simulated critical values, and they remain stable across a wide range of sample sizes and nominal levels, indicating consistency in the test's threshold behavior. Empirical Type I errors are reported in Table 2 for selected values of the scale parameter θ. These simulated errors confirm that our proposed test is accurately able to control the Type I error rate near the nominal level $\alpha = 0.05$ for $\theta = 0.5, 1, 2, 5$ and sample sizes $n = 10, 25, 50, 100, 200, 500$. We note that these findings suggest the test is reliable and robust under the null hypothesis across different distributional shapes and sample sizes.

Table 1. Simulated critical values for the Lindley distribution.

$n \backslash \alpha$	**0.01**	**0.05**	**0.10**
10	5.8810	3.6102	2.7356
25	5.8833	3.6165	2.7349
50	5.8531	3.5978	2.7671
100	5.7122	3.5832	2.7429
150	5.7904	3.6832	2.7262
200	5.7634	3.5960	2.7391
300	5.5254	3.5417	2.7782
500	5.7311	3.5207	2.7915
1000	5.8376	3.5765	2.6762

Table 2. Simulated Type I error control of the test for the nominal level $\alpha = 0.05$.

n	$\theta = 0.5$	$\theta = 1$	$\theta = 2$	$\theta = 5$
10	0.0510	0.0500	0.0522	0.0516
25	0.0457	0.0450	0.0486	0.0504
50	0.0501	0.0493	0.0481	0.0502
100	0.0533	0.0499	0.0493	0.0523
200	0.0519	0.0461	0.0507	0.0514
500	0.0496	0.0512	0.0484	0.0524

3.2. Power Comparisons Under Different Alternative Distributions

In this section, we perform simulations for B times under different alternative distributions to investigate the proposed test in terms of power. We consider sample sizes $n = 10, 25, 50, 100, 200$ with the nominal level $\alpha = 0.05$. The single-parameter alternative distributions considered in this study are reported in Table 3 and their respective parameters specified in Table 4. These alternatives cover probability densities with decreasing failure rate,

increasing failure rate, unimodal, and bathtub failure rate functions. To compare its performance, we consider the common and well-known goodness-of-fit tests based on distribution functions, namely, the Kolmogorov–Smirnov test (KS), the Kuiper test (V), the Cramer-von-Mises test (W^2), the Watson's test (U^2), and the Anderson–Darling test (A^2) statistics. The empirical power of our proposed test can be estimated using the following algorithm.

1. Obtain the critical value under the Lindley (θ) distribution assumption as explained in Section 3.1.
2. Generate a dataset $x_1, \ldots, x_n$ from the desired alternative distribution.
3. Process the data as if they were from a Lindley (θ) distribution and use the Maximum Likelihood Estimation (MLE) formula in Equation (5) to find the estimator $\hat{\theta}$ of the scale parameter θ.
4. Compute the energy goodness-of-fit statistic for the dataset $x_1, \ldots, x_n$ using Equation (11).
5. Based on the energy goodness-of-fit statistic and the critical value in Step 1, determine whether or not the null hypothesis is rejected.
6. Repeat Steps 1 through 5 for B times and calculate the empirical power as the fraction of rejections of the test in B times.

The empirical powers for the EDF (non-energy) tests considered in this study are also obtained in a similar manner. The results and comparisons are reported in Table 4. We notice that all considered methods are consistent and their powers increase with the increase in sample sizes. In Table 4, the proposed test has outperformed other existing tests against all choices of alternative distributions considered in this study. More importantly, we noticed that our proposed test has a larger power even for small sample sizes such as $n = 10$ and $n = 25$,, and it is evident in the case of uniform distribution. Also, it is worth to note that for large samples, the Anderson–Darling and Cramer-von-Mises test statistics are competitive in most cases and their powers are more or less the same as that of our proposed test. For small sample sizes, there is a mixture of results for the EDF-based tests, and the Anderson–Darling test has the lowest power when $n = 10$ except for the exponential model.

Table 3. Alternative distributions for the power comparison.

Distribution	Density	Notation
Weibull	$f(x) = \frac{1}{\theta}\exp(-\frac{x}{\theta})$	$W(\theta)$
Gamma	$f(x) = \frac{x^{\theta-1}\exp(-x)}{\Gamma(\theta)}$	$\Gamma(\theta)$
Half-Normal	$\frac{2\theta}{\pi}\exp(\frac{x^2\theta^2}{\pi})$	$HN(\theta)$
Chen	$f(x) = \theta x^{\theta-1}\exp(x^\theta)\exp(1-e^\theta)$	$CH(\theta)$
Uniform	$f(x) = \frac{1}{3}$	$U(0,3)$
Modified Extreme Value	$f(x) = 1 - \exp\left(\frac{1-e^x}{\theta}\right)$	$EV(\theta)$
Exponential	$f(x) = \exp(-x)$	$Exp(\theta)$
Log-Normal	$f(x) = \frac{1}{x\sqrt{2\pi}}\exp(-\frac{(\ln x-\theta)^2}{2})$	$LN(\theta)$
Dhillon	$f(x) = \frac{1+(\theta x)\lambda e^{\theta x}}{\lambda x e^{\theta x}+1}\exp(-\ln(\lambda x e^{\theta x}+1))$	$DL(\theta)$
Nakagami	$f(x) = \frac{2\theta^\theta}{\Gamma(\theta)}x^{2\theta-1}\exp(-\theta x^2)$	$NK(\theta)$
Linear Failure Rate	$f(x) = (1+\theta x)\exp(-x-\frac{\theta x^2}{2})$	$LF(\theta)$

Table 4. Power comparisons at the level of significance $\alpha = 0.05$.

Distribution	Sample Size n	$n\mathcal{E}_n$	KS	V	W^2	U^2	A^2
$W(1.3)$	10	0.1615	0.0762	0.0767	0.0770	0.0804	0.0505
	25	0.2527	0.1318	0.1317	0.1560	0.1415	0.1330
	50	0.3876	0.2175	0.2006	0.2648	0.2232	0.2625
	100	0.6422	0.4328	0.3891	0.5182	0.4337	0.5497
	200	0.8795	0.7208	0.6691	0.8090	0.7113	0.8524
$\Gamma(1.5)$	10	0.3820	0.0581	0.0687	0.0662	0.0727	0.0433
	25	0.4166	0.0738	0.0865	0.0816	0.0871	0.0696
	50	0.5027	0.1128	0.1258	0.1297	0.1365	0.1375
	100	0.6381	0.2048	0.2227	0.2344	0.2424	0.2841
	200	0.8067	0.3803	0.3823	0.4337	0.4275	0.5515
$HN(1)$	10	0.1274	0.0741	0.0797	0.0783	0.0798	0.0534
	25	0.2164	0.1344	0.1281	0.1471	0.1291	0.1231
	50	0.3720	0.2224	0.2004	0.2720	0.2187	0.2429
	100	0.6505	0.4323	0.3900	0.5231	0.4080	0.4976
	200	0.9144	0.7241	0.6940	0.8304	0.6938	0.8214
$CH(0.8)$	10	0.1646	0.0955	0.1075	0.1112	0.1070	0.0787
	25	0.3270	0.2049	0.1949	0.2470	0.1993	0.2065
	50	0.5696	0.3591	0.3503	0.4642	0.3678	0.4327
	100	0.8836	0.6781	0.6731	0.8060	0.6728	0.7912
	200	0.9960	0.9485	0.9548	0.9866	0.9484	0.9855
$U(0,3)$	10	0.6792	0.1657	0.2442	0.2135	0.2257	0.1572
	25	0.9463	0.3989	0.6001	0.5631	0.5298	0.5253
	50	0.9988	0.7488	0.9291	0.9013	0.8545	0.9098
	100	1.0000	0.9852	0.9996	0.9988	0.9943	0.9990
	200	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$EV(1.5)$	10	0.2357	0.1398	0.1412	0.1579	0.1504	0.1259
	25	0.4913	0.2898	0.3045	0.3706	0.3159	0.3231
	50	0.8073	0.5641	0.5830	0.7007	0.5950	0.6706
	100	0.9854	0.8885	0.9074	0.9651	0.9078	0.9583
	200	1.0000	0.9971	0.9990	0.9999	0.9984	1.0000
$Exp(1)$	10	0.1652	0.0597	0.0501	0.0555	0.0485	0.0688
	25	0.2019	0.0658	0.0601	0.0717	0.0620	0.0792
	50	0.2496	0.0866	0.0710	0.0994	0.0761	0.1158
	100	0.3424	0.1401	0.1113	0.1644	0.1202	0.1796
	200	0.4845	0.2465	0.1838	0.2867	0.2051	0.2998
$LN(0.8)$	10	0.4163	0.1166	0.1309	0.1265	0.1394	0.0979
	25	0.6218	0.2098	0.2567	0.2406	0.2846	0.2510
	50	0.8538	0.4356	0.5542	0.4911	0.5974	0.5962
	100	0.9901	0.8044	0.9117	0.8635	0.9340	0.9648
	200	1.0000	0.9950	0.9990	0.9980	0.9997	1.0000
$DL(1.5)$	10	0.6249	0.1546	0.1754	0.1797	0.1904	0.1270
	25	0.8680	0.3988	0.3940	0.4656	0.4367	0.4611
	50	0.9749	0.7180	0.7106	0.7864	0.7730	0.8309
	100	0.9994	0.9727	0.9660	0.9875	0.9829	0.9958
	200	1.0000	0.9999	0.9999	1.0000	1.0000	1.0000
$LF(2)$	10	0.1294	0.1129	0.1108	0.1208	0.1204	0.0825
	25	0.2791	0.2159	0.1981	0.2724	0.2194	0.2193
	50	0.5028	0.3942	0.3576	0.4804	0.3804	0.4493
	100	0.8170	0.6842	0.6528	0.7986	0.6766	0.7797
	200	0.9914	0.9617	0.9499	0.9870	0.9531	0.9848
$NK(0.6)$	10	0.1986	0.1206	0.1237	0.1412	0.1342	0.0971
	25	0.4136	0.2478	0.2357	0.3030	0.2575	0.2695
	50	0.7019	0.5004	0.4699	0.6091	0.4943	0.5902
	100	0.9431	0.8108	0.7878	0.9091	0.8135	0.9087
	200	0.9994	0.9876	0.9871	0.9976	0.9874	0.9985

4. Application to the Real-Life Datasets

In this section, the performance of the proposed test is investigated using three real datasets. We will apply the test to determine whether or not the underlying distribution is the Lindley one. We use the bootstrap algorithm process through simulations to find the approximate p-value associated with the proposed test as described below.

1. Fit the original data $x_1, \ldots, x_n$ with a Lindley distribution, Lindley(θ), and use the formula in Equation (5) to obtain the maximum likelihood estimate (MLE) of θ.
2. Use the formula in Equation (11) to calculate the energy goodness-of-fit statistic of the original data and denote it $\mathcal{Q}_n^1$.
3. Simulate $y_1, \ldots, y_n$, a random sample of size n, from a Lindley $(\hat{\theta})$ distribution, where $\hat{\theta}$ is obtained in step 1.
4. Compute the energy goodness-of-fit statistic using the formula in Equation (11) for the simulated data and denote it $\mathcal{Q}_n^{1*}$.
5. Repeat this simulation procedure for B times and obtain B energy goodness-of-fit statistics and denote them as $\mathcal{Q}_n^{1*}, \ldots, \mathcal{Q}_n^{B*}$.
6. The p-value will then be approximated as

$$\hat{p} = \frac{1}{B} \sum_{b=1}^{B} I\left(\mathcal{Q}_n^{b*} \geq \mathcal{Q}_n^1\right)$$

where $I(\cdot)$ is an indicator function that takes the value of one when $\mathcal{Q}_n^{b*} \geq \mathcal{Q}_n^1$ and zero otherwise.

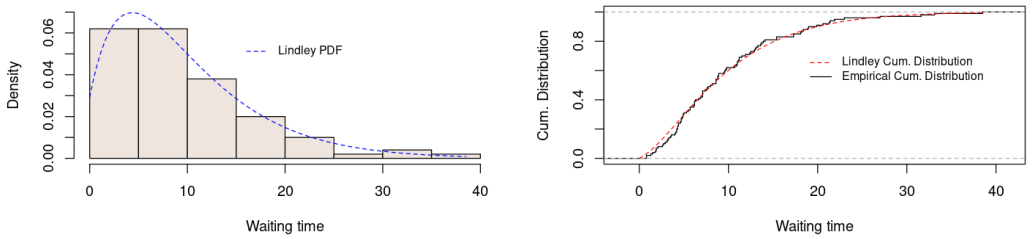
In our applications, the first dataset is listed in Table A1, which contains waiting times (in minutes) before service for 100 bank customers analyzed in [3]. The maximum likelihood estimate (MLE) of the scale parameter is $\hat{\theta} = 0.1866$, and its corresponding value of our test statistic is $n\mathcal{E}_n = 2.3598$. Since the simulated p-value is approximately 0.9243, we do not reject the null that the data follow the Lindley distribution. Other tests except the Cramer-von-Mises statistic suggested that the data follow the Lindley distribution. Their estimated p-values and test statistics are reported in Table A4.

The second dataset provided in Table A2 contains failure times of the electronic components in an accelerated life test. The data are reported and analyzed by Lawless [45]. The MLE of our test statistic is $\hat{\theta} = 0.0702$, and its corresponding value of test statistic is $n\mathcal{E}_n = 6.0811$. Since the simulated p-value is approximately 0.9357, we do not reject the null that the data follow the Lindley distribution. We observe in Table A5 a similar conclusion for other tests except for the Cramer-von-Mises test statistic, which fails to support the underlying distribution assumption as the Lindley one.

Table A3 provides our last dataset, which contains the quality of the yarn by recording the number of stress cycles needed before the yarn breaks. This dataset was originally analyzed by [46], and Shanker and Fesshaye [47] used the same dataset as an illustrative example in their analysis. The MLE of our test statistic is $\hat{\theta} = 0.0090$, and the test statistic value is $n\mathcal{E}_n = 152.3379$. The empirical p-value is approximately 0.3954, and therefore we do not reject the null hypothesis that the data follow the Lindley distribution. In addition, the Kolmogorov–Smirnov test statistic, the Anderson–Darling test statistic, the Kuiper test statistic, and the Watson test statistic provide little to no evidence to reject the null hypothesis that the data do follow the Lindley distribution. The Cramer-von-Mises test statistic again fails to support the fact that the data can be modeled with the Lindley distribution. The approximate p-values and their corresponding test statistics are reported in Table A6.

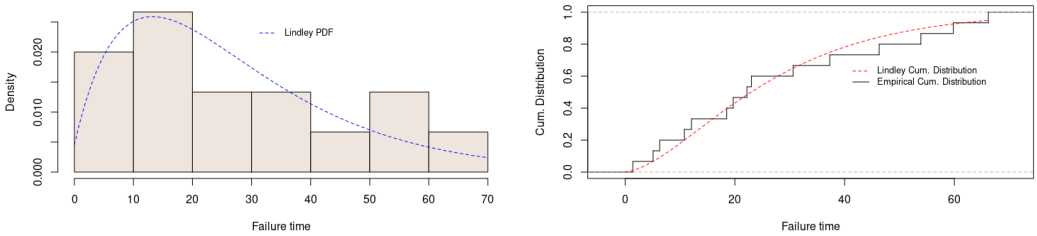
Finally, Figures 2a, 3a, and 4a provide density estimates for each of our three datasets considered in this study. In addition, Figures 2b, 3b, and 4b compare empirical and

theoretical distribution functions for these real datasets. In both situations, the Lindley distribution seems to fit the datasets adequately. Surprisingly, it is worthy to note that the Cramer-von-Mises test rejected the null hypothesis for all of our empirical applications.



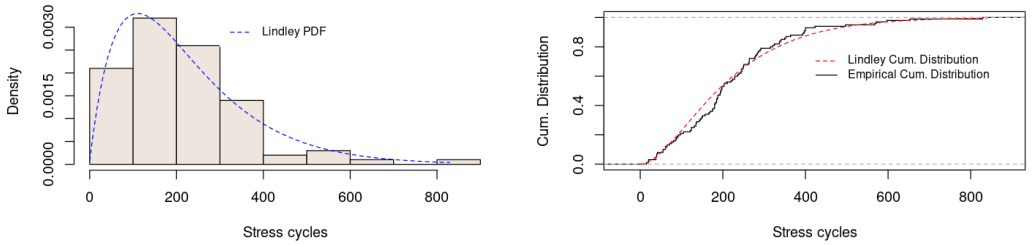
(**a**) Histogram of waiting time (**b**) Empirical CDF and estimated Lindley CDF

Figure 2. Banks' waiting time (in minutes) for customers before service.



(**a**) Histogram of failure times (**b**) Empirical CDF and estimated Lindley CDF

Figure 3. Failure times of the electronic components in an accelerated life test.



(**a**) Histogram of stress cycles (**b**) Empirical CDF and estimated Lindley CDF

Figure 4. Number of stress cycles for yarns' failure.

5. Conclusions

We have proposed a goodness-of-fit test based on energy statistics for the Lindley distribution. Under the null assumption, the proposed test is simple, robust, and computationally efficient to implement for any size of the sample. Monte Carlo simulations are carried out to evaluate the finite sample performance of the proposed test and compare the testing method with other similar existing tests in terms of controlling Type I errors and powers. It is evident that our test outperformed other tests especially for smaller sample sizes.

We successfully applied our proposed procedure to three real-life datasets to demonstrate its applicability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/stats8040087/s1, Proof of Proposition 1; Prooof of Proposition 2.

Author Contributions: J.N.: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Software, Resources, Supervision, Project Administration, Visualization, Writing—original draft preparation, Writing—review and editing; R.A.: Methodology, Software, Validation, Investigation,

Writing—original draft preparation, Visualization, Formal analysis, Data curation, Writing—review and editing. 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 the datasets are provided in the paper.

Acknowledgments: The authors are grateful to three anonymous reviewers for their helpful comments on improving the contents of this article.

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

Appendix A

Appendix A.1

We present proofs of results in the article and supporting materials from our simulations and data applications.

Proof of Proposition 1. Let $X \sim \text{Lindley}(\theta)$ and $x \in \mathbb{R}$. Then;

$$\begin{aligned}
E|x-X| &= \int_0^\infty |x-y| f_X(y)dy \\
&= \int_0^x (x-y) f_X(y)dy + \int_x^\infty (y-x) f_X(y)dy \\
&= x\int_0^x f_X(y)dy - \int_0^x y f_X(y)dy + \int_x^\infty y f_X(y)dy - x\int_x^\infty f_X(y)dy \\
&= xF_X(x) - \int_0^x y f_X(y)dy + \left(\int_0^\infty y f_X(y)dy - \int_0^x y f_X(y)dy\right) - x(1-F_X(x)) \\
&= 2xF(x) - x + E(X) - 2\int_0^x y f_X(y)dy \\
&= 2x\big(1-(\frac{1-\theta-\theta x}{1+\theta})\big)e^{-\theta x} - x + \frac{\theta+2}{\theta(1+\theta)} \\
&- \frac{2\theta}{1+\theta}\left[(-xe^{-\theta x} - x^2e^{-\theta x}) + (\frac{-e^{-\theta x}+1-2xe^{-\theta x}}{\theta}) - \frac{2xe^{-\theta x}-2}{\theta^2}\right] \\
&= x + \frac{2xe^{-\theta x}}{1+\theta} + \frac{2e^{-\theta x}}{1+\theta} + \frac{4e^{-\theta x}}{\theta(1+\theta)} + \frac{\theta-2}{\theta(1+\theta)} - \frac{2}{1+\theta},
\end{aligned}$$

where the CDF $F(x)$ and mean $E(X)$ of the Lindley distribution are explicitly given in Equation (2) and Equation (3), respectively. □

Proof of Proposition 2. Let $X, X' \overset{iid}{\sim} \text{Lindley}(\theta)$.

$$\begin{aligned}
E|X-X'| &= \int_0^\infty \int_0^\infty |x-y| f_X(x) f_{X'}(y)dydx \\
&= 2\int_0^\infty \int_0^x (x-y) f_X(x) f_{X'}(y)dydx \\
&= 2\int_0^\infty \int_0^x (x-y)\frac{\theta^2}{1+\theta}(1+x)e^{-\theta x}\frac{\theta^2}{1+\theta}(1+y)e^{-\theta y}dydx \\
&= 2m^2\int_0^\infty \int_0^x (x+x^2+x^2y-y-y^2-xy^2)e^{-\theta x}e^{-\theta y}dydx, \quad m = \frac{\theta^2}{1+\theta}. \\
&= 2m^2\left(\frac{3}{4\theta^3} + \frac{7}{4\theta^4} - \frac{3}{8\theta^5} - \frac{1}{4\theta^5} + \frac{2}{\theta^5} - \frac{1}{4\theta^3} - \frac{1}{4\theta^4} - \frac{5}{8\theta^5}\right) \\
&= 2\left(\frac{\theta^2}{1+\theta}\right)^2\left(\frac{2}{4\theta^3} + \frac{6}{4\theta^4} + \frac{6}{8\theta^5}\right) \\
&= \frac{2\theta^2+6\theta+3}{2\theta(1+\theta)^2}.
\end{aligned}$$

□

The complete proofs of Propositions 1 and 2 are provided in the attached Supplementary file. Below are results and datasets used in our empirical applications presented.

Table A1. Waiting time (in minutes).

0.8	0.8	1.3	1.5	1.8	1.9	1.9	2.1	2.6	2.7
2.9	3.1	3.2	3.3	3.5	3.6	4.0	4.1	4.2	4.2
4.3	4.3	4.4	4.4	4.6	4.7	4.7	4.8	4.9	4.9
5.0	5.3	5.5	5.7	5.7	6.1	6.2	6.2	6.2	6.3
6.7	6.9	7.1	7.1	7.1	7.1	7.4	7.6	7.7	8.0
8.2	8.6	8.6	8.6	8.8	8.8	8.9	8.9	9.5	9.6
9.7	9.8	10.7	10.9	11.0	11.0	11.1	11.2	11.2	11.5
11.9	12.4	12.5	12.9	13.0	13.1	13.3	13.6	13.7	13.9
14.1	15.4	15.4	17.3	17.3	18.1	18.2	18.4	18.9	19.0
19.9	20.6	21.3	21.4	21.9	23.0	27.0	31.6	33.1	38.5

Table A2. Failure times of electronic components.

1.4	5.1	6.3	10.8	12.1
18.5	19.7	22.2	23.0	30.6
37.3	46.3	53.9	59.8	66.2

Table A3. Number of stress cycles.

86	146	251	653	98	249	400	292	131	169
175	176	76	264	15	364	195	262	88	264
157	220	42	321	180	198	38	20	61	121
282	224	149	180	325	250	196	90	229	166
38	337	65	151	341	40	40	135	597	246
211	180	93	315	353	571	124	279	81	186
497	182	423	185	229	400	338	290	398	71
246	185	188	568	55	55	61	244	20	284
393	396	203	829	239	236	286	194	277	143
198	264	105	203	124	137	135	350	193	188

Table A4. Computed test statistics of waiting time data for different tests.

Test Statistic	Statistic Value	p Value
$n\mathcal{E}_n$	2.3598	0.9243
KS	0.6819	0.7304
V	0.0554	0.9998
W^2	1.1085	0.0017
U^2	0.0585	0.6255
A^2	0.4901	0.7632

Table A5. Computed test statistics of electronic life test data for different tests.

Test Statistic	Statistic Value	p Value
$n\mathcal{E}_n$	6.0811	0.9357
KS	0.4503	0.9843
V	0.0383	1.0000
W^2	0.9029	0.0065
U^2	0.0388	0.8588
A^2	0.3379	0.9210

Table A6. Computed test statistics of stress cycles data for different tests.

Test Statistic	Statistic Value	p Value
$n\mathcal{E}_n$	152.3379	0.3954
KS	0.9907	0.1744
V	0.1615	1.0000
W^2	1.6630	0.0000
U^2	0.1862	0.0527
A^2	0.9120	0.3855

References

1. Lindley, D.V. Fiducial distribution and Bayes' theorem. *J. R. Stat. Soc. Ser. B (Methodol.)* **1958**, *20*, 102–107. [CrossRef]
2. Lindley, D.V. *Introduction Probability and Statistics from a Bayesian Viewpoint*; Part II: Inference; Cambridge University Press: New York, NY, USA, 1965.
3. Ghitany, M.E.; Atieh, B.; Nadarajah, S. Lindley distribution and its application. *Math. Comput. Simul.* **2008**, *78*, 493–506. [CrossRef]
4. Ali, S.; Aslam, M.; Kazmi, S.M.A. A study of the effect of the loss function on Bayes estimate, posterior risk and hazard function for Lindley distribution. *Appl. Math. Model.* **2013**, *37*, 6068–6078. [CrossRef]
5. Al-Mutairi, D.K.; Ghitany, M.E.; Kundu, D. Inferences on the stress-strength reliability from Lindley distributions. *Commun. Stat. -Theory Methods* **2013**, *42*, 1443–1463. [CrossRef]
6. Singh, B.; Gupta, P.K. Load-sharing system model and its application to the real data set. *Math. Comput. Simul.* **2012**, *82*, 1615–1629. [CrossRef]
7. Satheesh Kumar, C.; Jose, R. On Double Lindley Distribution and Some of its Properties. *Am. J. Math. Manag. Sci.* **2018**, *38*, 23–43. [CrossRef]
8. Ibrahim, M.; Yadav, A.S.; Yousof, H.M.; Goual, H.; Hamedani, G. A new extension of Lindley distribution: Modified validation test, characterizations and different methods of estimation. *Commun. Stat. Appl. Methods* **2019**, *26*, 475–493. [CrossRef]
9. Nedjar, S.; Zeghdoudi, H. On gamma Lindley distribution: Properties and simulations. *J. Comput. Appl. Math.* **2016**, *298*, 167–174. [CrossRef]
10. Ramos, P.L.; Louzada, F. The generalized weighted Lindley distribution: Properties, estimation, and applications. *Cogent Math.* **2016**, *3*, 1256022. [CrossRef]
11. Gómez-Déniz, E.; Sordo, M.A.; Calderín-Ojeda, E. The log-Lindley distribution as an alternative to the beta regression model with applications in insurance. *Insur. Math. Econ.* **2014**, *54*, 49–57. [CrossRef]
12. Elbatal, I.; Elgarhy, M. Statistical properties of Kumaraswamy quasi Lindley distribution. *Int. J. Math. Trends Technol.* **2013**, *4*, 237–246.
13. Noughabi, H.A.; Noughabi, M.S. Gini Index based goodness-of-fit test for Lindley distribution. *Commun. Stat.-Simul. Comput.* **2024**, *54*, 2065–2075. [CrossRef]
14. Noughabi, H.A.; Noughabi, M.S. Empirical likelihood ratio-based goodness of fit test for Lindley distribution. *Stat. Optim. Inf. Comput.* **2024**, *12*, 869–881. [CrossRef]
15. Noughabi, H.A. A powerful goodness-of-fit test for Lindley distribution with application to real data. *Pak. J. Stat. Oper. Res.* **2021**, *17*, 761–769. [CrossRef]
16. Best, D.J.; Rayner, J.C.W. *Smooth Tests of Fit for More Flexible Alternatives to the Exponential and Poisson: The Lindley and Poisson-Lindley Distributions*; Working Paper; National Institute for Applied Statistics Research Australia, University of Wollongong: Wollongong, Australia, 2014; Volume 8, pp. 7–14.
17. Shan, G.; Vexler, A.; Wilding, G.E.; Hutson, A.D. Simple and exact empirical likelihood ratio tests for normaluity based on moment relations. *Commun. Stat.-Simul. Comput.* **2010**, *40*, 141–158. [CrossRef]
18. 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. *J. Stat. Plan. Inference* **2011**, *141*, 2128–2140. [CrossRef]
19. Rizzo, M.L. New goodness-of-fit tests for Pareto distributions. *ASTIN Bull. J. IAA* **2009**, *39*, 691–715. [CrossRef]
20. Noughabi, H.A.; Noughabi, M.S. On Testing the Adequacy of the Lindley Model and Power Study. *Stat. Optim. Inf. Comput.* **1970**, *11*, 719–729. [CrossRef]
21. Vexler, A.; Gurevich, G. Empirical likelihood ratios applied to goodness-of-fit tests on sample entropy. *Comput. Stat. Data Anal.* **2010**, *54*, 531–545. [CrossRef]
22. Ning, W.; Ngunkeng, G. An empirical likelihood ratio based goodness-of-fit test for skew normality. *Stat. Methods Appl.* **2013**, *22*, 209–226. [CrossRef]
23. Gupta, A.K.; Chen, T. Goodness-of-fit tests for skew normal distribution. *Commun. Stat.-Simul. Comput.* **2001**, *30*, 907–930.

24. 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, NY, USA, 1986; pp. 97–193.
25. Stephens, M.A. Edf statistics for goodness of fit and some comparisons. *J. Am. Stat. Assoc.* **1974**, *69*, 730–737. [CrossRef]
26. Sźekely, G.J. *E-Statistics: Energy of Statistical Samples*; Technical Report No. 02–16; Bowling Green State University, Department of Mathematics and Statistics: Bowling Green, OH, USA, 2002.
27. Rizzo, M.L. A New Rotation Invariant Goodness-of-Fit Test. Ph.D. Thesis, Bowling Green State University, Bowling Green, OH, USA, 2002.
28. Sźekely, G.J.; Rizzo, M.L. *The Energy of Data and Distance Correlation*, 1st ed.; Chapman and Hall: London, UK, 2023.
29. Logan, O.; Ning, W. Goodness-of-fit test for skew normality based on energy statistics. *Random Oper. Stoch. Equ.* **2020**, *28*, 227–236.
30. Ofosuhene, P. The Energy Goodness-of-Fit Test for the Inverse Gaussian Distribution. Ph.D. Thesis, Bowling Green State University, Bowling Green, OH, USA, 2020.
31. Sźekely, G.J.; Rizzo, M. A new test for multivariate normality. *J. Multivar. Anal.* **2005**, *93*, 58–80. [CrossRef]
32. Sźekely, G.J.; Rizzo, M.L. Testing for Equal Distributions in high Dimension. *InterStat* **2004**, *11*.
33. Rizzo, M.L. *A Test of Homogeneity for Two Multivariate Populations, Physical and Engineering Sciences Section*; American Statistical Association: Alexandria, VA, USA, 2003.
34. Njuki, J.; Ning, W. Energy statistic-based modified information criterion for detecting the change in distribution. *J. Appl. Stat.* **2025**, 1–23. [CrossRef]
35. Njuki, J.M. Nonparametric Sequential Tests for Change Point Analysis Using Energy Statistics. Ph.D. Thesis, Bowling Green State University, Bowling Green, OH, USA, 2022.
36. Matterson, D.S.; James, N.A. A nonparametric Approach for Multiple Change Point Analysis of Multivariate Data. *J. Am. Stat. Assoc.* **2014**, *109*, 334–345. [CrossRef]
37. Kim, A.Y.; Marzban, C.; Percival, D.B.; Stuetzle, W. Using labeled data to evaluate change detectors in a multivariate streaming environment. *Signal Process.* **2009**, *89*, 2529–2536. [CrossRef]
38. Sźekely, G.J.; Rizzo, M.L. The Energy of Data. *Annu. Rev. Stat. Its Appl.* **2017**, *4*, 447–479. [CrossRef]
39. Sźekely, G.J.; Rizzo, M. A Class of Statistical Based on Distances. *J. Stat. Plan. Inference* **2013**, *143*, 1249–1272. [CrossRef]
40. Sźekely, G.J. *E-Statistics: Energy of Statistical Samples*; Technical Report 03-05; BGSU, Department of Mathematics and Statistics: Bowling Green, OH, USA, 2000.
41. Anderson, T.W.; Darling, D.A. A test of goodness of fit. *J. Am. Stat. Assoc.* **1954**, *49*, 765–769. [CrossRef]
42. Watson, G.S. Goodness-of-fit tests on a circle. *Biometrika* **1961**, *48*, 109–114. [CrossRef]
43. Kuiper, N.H. Tests concerning random points on a circle. *Proc. Ned. Akad. Van Wet. Ser. A* **1960**, *63*, 38–47. [CrossRef]
44. Móri, F.T.; Sźekely, G.J.; Rizzo, M.L. On energy tests of normality. *J. Stat. Plan. Inference* **2021**, *213*, 1–15. [CrossRef]
45. Lawless, J.F. *Statistical Models and Methods for Lifetime Data*; John Wiley & Sons: New York, NY, USA, 1982.
46. Picciotto, R. Tensile Fatigue Characteristics of a Sized Polyester/Viscose Yarn and Their Effect on Weaving Performance. Master's Thesis, Department of Textile Technology, North Carolina State University, Raleigh, NC, USA, 1970.
47. Shanker, R.; Fesshaye, H. On modeling of lifetime data using Akash, Shanker, Lindley and exponential distributions. *Biom. Biostat. Int. J.* **2016**, *3*, 214–224. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.