STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **12**, July 2024, pp 869–881.
Published online in International Academic Press (www.IAPress.org)

Empirical Likelihood Ratio-based Goodness of Fit Test for the Lindley Distribution

Hadi Alizadeh Noughabi [1,*], Mohammad Shafaei Noughabi [2]

[1] *Department of Statistics, University of Birjand, Birjand, Iran*
[2] *Department of Mathematics and Statistics, University of Gonabad, Gonabad, Iran*

Abstract The Lindley distribution may serve as a useful reliability model. In this article, we propose a goodness of fit test for the Lindley distribution based on the empirical likelihood (EL) ratio. The properties of the proposed test are stated and the critical values are obtained by Monte Carlo simulation. Power comparisons of the proposed test with some known competing tests are carried out via simulations. Finally, an illustrative example is presented and analyzed.

Keywords Lindley distribution, Empirical likelihood ratio, Goodness of fit test, Monte Carlo simulation, Test power

AMS 2010 subject classifications 62G10; 62P12

DOI: 10.19139/soic-2310-5070-1481

1. Introduction

The modeling and analyzing lifetime data are crucial in many applied sciences including medicine, engineering, insurance and finance, amongst others. It is well known that the Lindley distribution is one of the fundamental models applied for reliability models. The Lindley distribution has been discussed by many authors in different practical cases, such as Bayesian estimation Ali et al. (2013), loading-sharing system mode Singh and Gupta (2012) and stress-strength reliability model Al-Mutairi el al. (2013). It deserves mentioning that the Lindley distribution provides a flexible shape to model the lifetime data. Moreover, Ghitany et al. (2008) presented a comprehensive study about its important mathematical and statistical properties, estimation of parameter and application showing the superiority of Lindley distribution over of the bank customers.

Since the distribution was proposed, it has been overlooked in the literature partly due to the popularity of the exponential distribution in the context of reliability analysis. Nonetheless, it has recently received considerable attention as a lifetime model to analyze survival data in the competing risks analysis and stress-strength reliability studies; see, for example, Ghitany et al. (2008), Mazucheli and Achcar (2011), Gupta and Singh (2013), Al-Mutairi el al. (2013), and Wang (2013), Valiollahi et al. (2017), Altun (2019), Kumar and Jose (2019), and Ibrahim et al. (2019), among others.

Ghitany et al. (2008) provide a nice overview of various statistical properties of the Lindley distribution. Furthermore, they argue that the Lindley distribution could be a better lifetime model than the exponential distribution using a real data set.

Therefore, it is a clear need to check whether the Lindley model is a satisfactory model for the observations.

Goodness of fit (GOF) tests are designed to measure how well the observed sample data fits some proposed model. One class of GOF tests that can be used consists of tests based on the distance between the empirical and

*Correspondence to: Hadi Alizadeh Noughabi (Email: alizadehhadi@birjand.ac.ir). Department of Statistics, University of Birjand, Birjand, Iran.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 2024 International Academic Press

hypothesized distribution functions. Five of the known tests in this class are Cramer-von Mises (W^2), Kolmogorov-Smirnov (D), Kuiper (V), Watson (U^2), and Anderson-Darling (A^2). For more details about these tests, see D'Agostino and Stephens (1986).

In parametric statistics, based on Neyman-Pearson lemma the maximum likelihood ratio test is a uniformly most powerful test. Suppose that $X_1, ..., X_n$ are a random sample and we wish to test the hypothesis

$$H_0 : X_1, ..., X_n \sim f_0,$$

versus

$$H_1 : X_1, ..., X_n \sim f_1.$$

The most powerful test statistic for the above hypothesis is the likelihood ratio

$$\frac{\prod_{i=1}^{n} f_1(X_i)}{\prod_{i=1}^{n} f_0(X_i)}.$$

where $f_0(x)$ and $f_1(x)$ are completely known.

As we know in nonparametric statistics, the alternative distribution is unknown and therefore, for goodness of fit tests based on EL ratio, we need to estimate the likelihood function $\prod_{i=1}^{n} f_1(X_i)$ and then we can use the likelihood ratio statistic. Vexler and Gurevich (2010) estimated the likelihood ratio as

$$\prod_{i=1}^{n} \frac{2m}{n\left(X_{(i+m)} - X_{(i-m)}\right)},$$

and then introduced a test statistic for goodness of fit. Their test statistic is as

$$T_{mn} = \frac{\prod_{i=1}^{n} \frac{2m}{n\left(X_{(i+m)} - X_{(i-m)}\right)}}{\prod_{i=1}^{n} f_0(X_i; \hat{\theta})},$$

where $\hat{\theta}$ is the maximum likelihood estimator (MLE) of θ. Also, $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ are the order statistics obtained from $X_1, ..., X_n$ and $X_{(i)} = X_{(1)}$ if $i < 1$, and $X_{(i)} = X_{(n)}$ if $i > n$.

Since T_{mn} depends on m, they proposed the following test statistic.

$$T_{mn} = \frac{\min_{1 \leq m \leq n^{\delta}} \prod_{i=1}^{n} \frac{2m}{n\left(X_{(i+m)} - X_{(i-m)}\right)}}{\prod_{i=1}^{n} f_0(X_i; \hat{\theta})}.$$

Here, $\delta \in (0, 1)$ and $\hat{\theta}$ is the MLE of θ.

They used their test statistic and proposed tests for the normal and uniform distributions. Moreover, Vexler et al. (2011) applied the above test statistic and introduced a goodness of fit test for the inverse Gaussian distribution.

Recently, the empirical likelihood methods are applied in many statistical problems see for example, Vexler et al. (2011b,c), Vexler and Gurevich (2011), Gurevich and Vexler (2011), Shan et al. (2011), Vexler and Yu (2011), Yu et al. (2011), Vexler et al. (2012a,b), and Vexler et al. (2014), Vexler and Zou (2018), Gurevich and Vexler (2018), Zou el al. (2019), Vexler el al. (2019) and Vexler (2020).

The main contribution of the paper can express as follows. In this paper, we apply ELR-based test for the Lindley distribution. The method of Vexler and Gurevich (2010) is stated and based on this method, we propose a goodness of fit test for the Lindley distribution. Table of critical values and properties of the tests are presented. We show

through extensive simulation studies that the proposed goodness of fit test is more powerful, or at least as good as the classical EDF-tests for different choices of sample sizes and alternatives. We also investigate the behavior of the tests for the Lindley model with real data.
This article is organized as follows. Section 1 describes the Lindley distribution and the procedure for estimating the parameter of this model. Section 3 presents the EL statistic for test of fit for the Lindley distribution. Section 4 gives the results of the power comparison of the proposed test with some known competing tests under various alternatives. Section 5 contains an illustrative example. The following section contains a brief conclusion.

2. The Lindley Distribution

If the density function of the random variable X be as follows, then we say that X has a Lindley distribution.

$$f_0(x;\theta) = \frac{\theta^2}{\theta+1}(1+x)e^{-\theta x}, \qquad x>0,\ \ \theta>0.$$

Lindley distribution was proposed by Lindley (1958) in the context of Bayesian statistics, as a counter example of fiducial statistics. The cumulative distribution function of the Lindley distribution is as

$$F_0(x;\theta) = 1 - \frac{\theta+1+\theta x}{\theta+1}e^{-\theta x}.$$

The mean and variance of the distribution are

$$\mu = E(X) = \frac{\theta+2}{\theta(\theta+1)},$$

and

$$\sigma^2 = Var(X) = \frac{\theta^2+4\theta+2}{\theta^2(\theta+1)^2}.$$

Ghitany et al. (2008) conducted a detailed study about various properties of Lindley distribution including skewness, kurtosis, hazard rate function, mean residual life function, stochastic ordering, stress-strength reliability, among other things; estimation of its parameter and application to model waiting time data in a bank.
In the literature of survival analysis and reliability theory, the exponential distribution is widely used as a model of lifetime data. However, the exponential distribution only provides a reasonable fit for modeling phenomenon with constant failure rates. Distributions like gamma, Weibull and lognormal have become suitable alternatives to the exponential distribution in many practical situations. Ghitany et al. (2008) found that the Lindley distribution can be a better model than one based on the exponential distribution.
The Lindley distribution belongs to an exponential family and it can be written as a mixture of an exponential with parameter and a gamma distribution with parameters $(2,\theta)$.

$$f_0(x;\theta) = pf_1(x) + (1-p)f_2(x) \qquad x>0,$$

where $p=\theta/(1+\theta)$, $f_1(x)=\theta e^{-\theta x}$ and $f_2(x)=\theta^2 x e^{-\theta x}$.

Shanker et al. (2015) discussed a comparative study of Lindley and exponential distributions for modelling various lifetime data sets from biomedical science and engineering, and concluded that there are lifetime data where exponential distribution gives better fit than Lindley distribution and in majority of data sets Lindley distribution gives better fit than exponential distribution.

Since for computing the test statistics, we need to estimate the parameter θ, we apply the MLE approach to estimate the unknown parameter. Suppose $X_1,...,X_n$ is a random sample from the Lindley distribution, the estimator for both MLE and method of moments estimate of the parameter θ is

$$\hat{\theta} = \frac{-\left(\bar{X}-1\right)+\sqrt{\left(\bar{X}-1\right)^2+8\bar{X}}}{2\bar{X}}, \qquad \bar{X}>0.$$

Ghitany et al. (2008) showed that the estimator $\hat{\theta}$ of θ is positively biased: $E(\hat{\theta}) - \theta > 0$, and it is consistent and asymptotically normal $\sqrt{n}\left(\hat{\theta} - \theta\right) \to N(0, 1/\sigma^2)$.
We will use the ML estimator for the proposed statistic to test the goodness of fit for the Lindley distribution.

In complete sample case, Ghitany et al. (2008) developed different distributional properties, reliability characteristics and some inferential procedures for the Lindley distribution. Krishna and Kumar (2011) discussed reliability estimation in Lindley distribution with progressively type II right censored sample. Gupta and Singh (2013) gave parameter estimation of Lindley distribution with hybrid censored data. Also, Al-Mutairi et al. (2013) studied inferences on stress-strength reliability for Lindley distribution with complete sample information. Kumar et al. (2015) discussed estimation of stress-strength reliability using progressively first failure censoring. These studies suggest that in many real-life situations Lindley distribution serves as a better lifetime model than the so far popular distributions like exponential, gamma, Rayleigh, Weibull etc.

3. The EL Goodness of Fit Test

Given a random sample $X_1, ..., X_n$ from a continuous probability distribution F with a density function $f(x)$, the hypothesis of interest is

$$H_0: \ f(x) = f_0(x;\theta) = \frac{\theta^2}{\theta+1}(1+x)e^{-\theta x}\,, \qquad for\ some\ \theta \in \Theta,$$

where θ is specified or unspecified and $\Theta = R^+$. The alternative to H_0 is

$$H_1: \ f(x) \neq f_0(x;\theta)\,, \qquad for\ any\ \theta\,,$$

The likelihood ratio test statistic for the above hypothesis is defined as

$$LR = \frac{\prod_{i=1}^{n} f_{H_1}(X_i)}{\prod_{i=1}^{n} f_{H_0}(X_i;\theta)}.$$

When density function under H_1 is known (f_{H_1}), Neyman-Pearson lemma guarantees that the LR test is the uniformly most powerful (UMP) test. If it is unknown, we will apply the maximum empirical likelihood method to estimate the numerator. Also, the maximum likelihood method will be applied to estimate the parameter θ of a Lindley distribution under the null hypothesis.

Consider

$$L_f = \prod_{i=1}^{n} f_{H_1}(X_i) = \prod_{i=1}^{n} f_{H_1}(X_{(i)}) = \prod_{i=1}^{n} f_i\,,$$

where $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ are the order statistics of the observations and $f(X_{(i)}) = f_i$. We apply the empirical likelihood method to derive the values of f_i that maximize L_f with the constraint $\int f(s)ds = 1$ under the alternative hypothesis. The following lemma, proved by Vexler and Gurevich (2010), express this constraint more explicitly.

Lemma 1. Let $f(x)$ be a density function. Then

$$\sum_{j=1}^{n} \int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx = 2m \int_{X_{(1)}}^{X_{(n)}} f(x)dx - \sum_{k=1}^{m-1} (m-k) \int_{X_{(n-k)}}^{X_{(n-k+1)}} f(x)dx - \sum_{k=1}^{m-1} (m-k) \int_{X_{(k)}}^{X_{(k+1)}} f(x)dx\,,$$

where $X_{(j)} = X_{(1)}$ if $j \leq 1$ and $X_{(j)} = X_{(n)}$ if $j \geq n$.

Let

$$\Delta_m = \frac{1}{2m}\sum_{j=1}^{n}\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx,$$

and since $\int_{X_{(1)}}^{X_{(n)}} f(x)dx \leq \int_{-\infty}^{\infty} f(x)dx = 1$, from Lemma 1,

$$\Delta_m \leq 1.$$

When $m/n \to 0$ as $m, n \to \infty$, we can expect that $\Delta_m \approx 1$.

The integration $\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx$ can be approximated by $\left(X_{(j+m)} - X_{(j-m)}\right) f_j$ and thus

$$\sum_{j=1}^{n}\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx \approx \sum_{j=1}^{n}\left(X_{(j+m)} - X_{(j-m)}\right) f_j.$$

Therefore, Δ_m can be approximated by

$$\hat{\Delta}_m = \frac{1}{2m}\sum_{j=1}^{n}\left(X_{(j+m)} - X_{(j-m)}\right) f_j.$$

Now, we apply the Lagrange multiplier method to maximize $l = \log(L_f) = \sum\limits_{j=1}^{n} \log f_j$, under the constrain $\hat{\Delta}_m \leq 1$. We have

$$l(f_1, f_2, ..., f_n, \eta) = \sum_{j=1}^{n} \log f_j + \eta\left(\frac{1}{2m}\sum_{j=1}^{n}\left(X_{(j+m)} - X_{(j-m)}\right) f_j - 1\right),$$

where η is a Lagrange multiplier. By taking the derivative of the above equation respect to each f_j and η, we obtain the values of $f_1, f_2, ..., f_n$. The form of values is as

$$f_j = \frac{2m}{n\left(X_{(j+m)} - X_{(j-m)}\right)}, \qquad j = 1, ..., n,$$

where $X_{(j)} = X_{(1)}$ if $j \leq 1$ and $X_{(j)} = X_{(n)}$ if $j \geq n$.

We therefore construct the likelihood ratio test statistic to test the goodness of fit for the Lindley distribution based on the maximum empirical likelihood method as

$$T_{mn} = \frac{\prod\limits_{j=1}^{n} \frac{2m}{n\left(X_{(j+m)} - X_{(j-m)}\right)}}{\max\limits_{\theta} \prod\limits_{j=1}^{n} f_{H_0}(X_j; \theta)},$$

where θ is the parameter of a Lindley distribution.

Clearly, the test statistic T_{mn} strongly depends on the value of m and for a given n, the value of m must be determined. It is not possible to have one value of m, for a given n, that would result in a test attaining the maximum power for all alternatives. Therefore, similar to Vexler and Gurevich (2010) and Vexler et al. (2011), we propose the following test statistic.

$$T_n = \frac{\min\limits_{1 \leq m < n^{\delta}} \prod\limits_{j=1}^{n} \frac{2m}{n\left(X_{(j+m)} - X_{(j-m)}\right)}}{\max\limits_{\theta} \prod\limits_{j=1}^{n} f_{H_0}(X_j; \theta)},$$

where $\delta \in (0,1)$. Here, we choose $\delta = 0.5$ for the power study of our test because for this value of δ, we attain good (not best) powers for all alternative distributions.
The following theorems give some asymptotic properties of the proposed test statistic. First, we denote

$$h(x,\theta) = \frac{\partial \log f_{H_0}(x;\theta)}{\partial \theta}.$$

Assume the following conditions are hold.
(C1) $E(\log f(X_1))^2 < \infty$;
(C2) under the null hypothesis, $\left|\theta - \hat{\theta}\right| \to 0$ as $n \to \infty$;
(C3) under the alternative hypothesis, $\theta \to \theta_0$ as $n \to \infty$, where θ_0 is a constant vector with finite components;
(C4) There are open intervals $\Theta_0 \subseteq R^2$ and $\Theta_1 \subseteq R^2$ containing θ and θ_0, respectively. There also exists a function $s(x)$ such that $|h(x,\xi)| \leq s(x)$ for all $x \in R$ and $\xi \in \Theta_0 \cup \Theta_1$.

Theorem 1. Assume that the conditions C1-C4 hold. Then, under H_0,

$$\frac{1}{n}\log(T_n) \to 0,$$

in probability as $n \to \infty$.

Theorem 2. Assume that the conditions C1-C4 hold. Then, under H_1,

$$\frac{1}{n}\log(T_n) \to E\log\left(\frac{f_{H_1}(X_1)}{f_{H_0}(X_1;\theta_0)}\right),$$

in probability as $n \to \infty$. Hence, the test is consistent.

Vexler and Gurevich (2010) showed that the above theorems are satisfied for any null family of distributions. Therefore, when the null hypothesis is the Lindley distribution Theorems 1 and 2 hold.

4. Simulation Study

Since deriving the exact distribution of the proposed test statistic is complicated, we study the null distribution of the test statistic T_n via Monte Carlo simulations using 100,000 runs for each sample size. Upper tail percentiles are obtained for values 0.99, 0.95, and 0.90. These values are presented in Table 1.

We evaluate in Table 2 the estimated type I error control using the 0.05 percentiles of the proposed test ($\alpha = 0.05$). We generated random samples from a spectrum of Lindley populations and then obtained the actual sizes of the test. The results are presented in Table 2. We observe that the empirical percentiles given in Table 1 provides an excellent type I error control.

Through Monte Carlo simulations, the power values of the proposed test against various alternatives are computed. Since the tests of fit based on the empirical distribution function are commonly used in practice, we compare the performance of the EDF-tests and the proposed ELR based goodness of fit test under various alternative distributions. The well-known EDF-tests are Cramer-von Mises (W^2), Kolmogorov-Smirnov (D), Kuiper (V), Watson (U^2), and Anderson-Darling (A^2). The test statistics of these tests are briefly described as follows. For more details about these tests, see D'Agostino and Stephens (1986).
Let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ are the order statistics based on the random sample $X_1, ..., X_n$.

Table 1. Critical values of the $\log{(T_n)}/n$ statistic.

	α		
n	0.01	0.05	0.10
5	1.5671	1.1695	0.9908
10	0.8458	0.6569	0.5722
15	0.6241	0.4944	0.4344
20	0.4887	0.3904	0.3454
25	0.4062	0.3268	0.2900
30	0.3543	0.2866	0.2558
40	0.2839	0.2318	0.2077
50	0.2376	0.1959	0.1763
100	0.1432	0.1196	0.1082

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

n	$\theta = 0.5$	$\theta = 2$	$\theta = 4$	$\theta = 8$
10	0.0539	0.0467	0.0465	0.0456
20	0.0572	0.0461	0.0440	0.0446
30	0.0572	0.0450	0.0439	0.0436
50	0.0558	0.0443	0.0428	0.0405

1. The Cramer-von Mises statistic (1931):

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(\frac{2i-1}{2n} - F_0(X_{(i)};\hat{\theta}) \right)^2 .$$

2. The Watson statistic (1961):

$$U^2 = W^2 - n\big(\bar{P} - 0.5\big)^2,$$

where $\bar{P}$ is the mean of $F_0(X_{(i)};\hat{\theta}),\ \ i = 1, ..., n$.

3. The Kolmogorov-Smirnov statistic (1933):

$$D = \max(D^+, D^-),$$

where

$$D^+ = \max_{1\le i\le n} \left\{ \frac{i}{n} - F_0(X_{(i)};\hat{\theta}) \right\}; \quad D^- = \max_{1\le i\le n} \left\{ F_0(X_{(i)};\hat{\theta}) - \frac{i-1}{n} \right\}.$$

4. The Kuiper statistic (1960):

$$V = D^+ + D^-.$$

5. The Anderson-Darling statistic (1952):

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \left\{ \log F_0(X_{(i)};\hat{\theta}) + \log\left[1 - F_0(X_{(n-i+1)};\hat{\theta}) \right] \right\}.$$

In the above test statistics, $F_0(x)$ is the cumulative distribution function of the Lindley distribution and $\hat{\theta}$ is the maximum likelihood estimate of the parameter θ.

The following alternatives are considered in power comparison.

- the Weibull distribution with density $\theta x^{\theta-1}\exp\left(-x^{\theta}\right)$, denoted by $W(\theta)$;
- the gamma distribution with density $\Gamma(\theta)^{-1}x^{\theta-1}\exp\left(-x\right)$, denoted by $\Gamma(\theta)$;
- the lognormal distribution $LN(\theta)$ with density $(\theta x)^{-1}(2\pi)^{-1/2}\exp\left(-(\log x)^2\Big/(2\theta^2)\right)$;
- the half-normal HN distribution with density $\Gamma(2/\pi)^{1/2}\exp\left(-x^2/2\right)$;
- the uniform distribution U with density $1, \quad 0 \leq x \leq 1$;
- the modified extreme value $EV(\theta)$, with distribution function $1-\exp\left(\theta^{-1}(1-e^{x})\right)$;
- the linear increasing failure rate law $LF(\theta)$ with density $(1+\theta x)\exp\left(-x-\theta x^2/2\right)$;
- Dhillon's (1981) distribution with distribution function $1-\exp\left(-(\log(x+1))^{\theta+1}\right)$;
- Chen's (2000) distribution $CH(\theta)$, with distribution function $1-\exp\left(2\left(1-e^{x^{\theta}}\right)\right)$.

These alternatives include densities f with decreasing failure rates (DFR), increasing failure rates (IFR) as well as models with unimodal failure rate (UFR) functions and bathtub failure rate (BFR) functions.

We compute the power values of the tests under the above alternatives by Monte Carlo simulations as follows. Under each alternative 100,000 samples of size 10, 20, 30 and 50 are generated and the test statistics are calculated. Then power of the corresponding test is computed by the frequency of the event "the statistic is in the critical region". Tables 3 and 4 display and compare the power values of the tests at the significance level $\alpha = 0.05$. For each sample size and alternative, the bold type in these tables indicates the tests achieving the maximal power.

From Table 3, the increasing failure rates alternatives, it is seen that the tests based on T_n and W^2 statistics have the most power. The power differences between the test T_n and the other tests are substantial, especially where $Unif(0,1)$ was the alternative. Also, it is evident from Table 3, for small sample sizes the proposed test has the most power. For the rest alternatives, the Anderson-Darling test and the proposed test have the most power. Therefore, although there is no uniformly most powerful test against all alternatives, the tests based on T_n, W^2 and A^2 statistics can be recommended. The powerful tests against different alternatives are presented in Table 5.
In general, we can conclude that the proposed test T_n has a good performance against IFR alternatives and therefore can be used in practice.

Table 3. Empirical powers of the tests against IFR alternatives at significance level 5%.

$Alternative$	n	W^2	D	V	U^2	A^2	T_n
$W(1.4)$	10	0.1303	0.1174	0.1104	0.1170	0.0894	**0.1580**
	20	**0.2258**	0.1966	0.1761	0.1884	0.1917	0.2196
	30	**0.3237**	0.2691	0.2330	0.2635	0.2967	0.2924
	50	**0.5098**	0.4231	0.3736	0.4167	0.5036	0.4210
$\Gamma(2)$	10	0.1175	0.1028	0.1101	0.1188	0.0810	**0.1679**
	20	0.2011	0.1754	0.1772	0.1935	0.1800	**0.2415**
	30	0.2879	0.2412	0.2369	0.2687	0.2827	**0.3257**
	50	**0.4745**	0.4014	0.3875	0.4408	0.5104	0.4711
HN	10	0.0952	0.0887	0.0844	0.0875	0.0678	**0.1059**
	20	**0.1364**	0.1234	0.1084	0.1149	0.1076	0.1309
	30	**0.1835**	0.1552	0.1340	0.1446	0.1492	0.1647
	50	**0.2839**	0.2321	0.1960	0.2139	0.2445	0.2330
U	10	0.3386	0.2647	0.3088	0.2957	0.2615	**0.4227**
	20	0.6318	0.4888	0.6071	0.5477	0.5793	**0.7899**
	30	0.8309	0.6764	0.8143	0.7416	0.8056	**0.9557**
	50	0.9756	0.9000	0.9777	0.9417	0.9756	**0.9994**
$CH(1)$	10	0.0937	0.0868	0.0772	0.0789	0.0673	**0.0986**
	20	**0.1364**	0.1220	0.0998	0.1061	0.1074	0.1252
	30	**0.1826**	0.1557	0.1230	0.1332	0.1477	0.1613
	50	**0.2796**	0.2301	0.1810	0.1933	0.2379	0.2294
$CH(1.5)$	10	0.4268	0.3505	0.3359	0.3553	0.3348	**0.4311**
	20	**0.7600**	0.6343	0.6239	0.6480	0.7160	0.7156
	30	**0.9200**	0.8205	0.8176	0.8370	0.9071	0.8890
	50	**0.9943**	0.9684	0.9736	0.9763	0.9943	0.9896
$LF(2)$	10	0.1386	0.1235	0.1113	0.1187	0.0972	**0.1438**
	20	**0.2282**	0.1943	0.1706	0.1802	0.1851	0.1964
	30	**0.3292**	0.2723	0.2327	0.2527	0.2828	0.2591
	50	**0.5133**	0.4204	0.3663	0.3955	0.4662	0.3794
$LF(4)$	10	0.2056	0.1790	0.1594	0.1700	0.1469	**0.2065**
	20	**0.3777**	0.3160	0.2752	0.2980	0.3192	0.3061
	30	**0.5308**	0.4386	0.3864	0.4204	0.4758	0.4160
	50	**0.7680**	0.6595	0.6067	0.6401	0.7313	0.6052
$EV(0.5)$	10	0.0923	0.0861	0.0749	0.0782	0.0670	**0.0980**
	20	**0.1384**	0.1221	0.1020	0.1074	0.1068	0.1260
	30	**0.1833**	0.1557	0.1242	0.1345	0.1467	0.11589
	50	**0.2779**	0.2262	0.1803	0.1933	0.2378	0.2303
$EV(1.5)$	10	0.1681	0.1456	0.1420	0.1547	0.1170	**0.1965**
	20	**0.3359**	0.2706	0.2529	0.2634	0.2658	0.3157
	30	**0.4612**	0.3645	0.3618	0.3805	0.4152	0.4478
	50	**0.7218**	0.5906	0.5811	0.5943	0.6880	0.6816

Table 4. Empirical powers of the tests against UFR, DFR and BFR alternatives at significance level 5%.

Alternative	n	W^2	D	V	U^2	A^2	T_n
$LN(0.8)$	10	0.1413	0.1302	0.1279	0.1403	0.1068	**0.1686**
	20	0.2221	0.1968	0.2204	0.2448	0.2110	**0.2847**
	30	0.3180	0.2720	0.3268	0.3652	0.3440	**0.4160**
	50	0.5147	0.4436	0.5541	0.6054	0.6131	**0.6225**
$LN(1.5)$	10	0.5140	0.4823	0.3849	0.4001	**0.5544**	0.2399
	20	0.8027	0.7664	0.6690	0.6869	**0.8197**	0.5861
	30	0.9257	0.9020	0.8342	0.8489	**0.9306**	0.7834
	50	0.9900	0.9842	0.9642	0.9697	**0.9905**	0.9439
$DL(1)$	10	0.0877	0.0813	0.0809	0.0862	0.0629	**0.1074**
	20	0.1185	0.1064	0.1139	0.1236	0.1041	**0.1401**
	30	0.1486	0.1274	0.1445	0.1619	0.1445	**0.1803**
	50	0.2123	0.1771	0.2245	0.2533	0.2394	**0.2427**
$DL(1.5)$	10	0.1999	0.1735	0.1751	0.1937	0.1462	**0.2587**
	20	0.3844	0.3271	0.3228	0.3634	0.3601	**0.4030**
	30	0.5568	0.4783	0.4598	0.5241	**0.5677**	0.5498
	50	0.8123	0.7363	0.7129	0.7832	**0.8509**	0.7505
$W(0.8)$	10	0.1960	0.1750	0.1288	0.1366	**0.2748**	0.0238
	20	0.3570	0.3095	0.2295	0.2438	**0.4417**	0.0833
	30	0.4933	0.4319	0.3201	0.3476	**0.5752**	0.1517
	50	0.7062	0.6330	0.5093	0.5395	**0.7720**	0.2788
$\Gamma(0.4)$	10	0.5137	0.4712	0.3701	0.3914	**0.7163**	0.0731
	20	0.8109	0.7663	0.6579	0.6850	**0.9222**	0.3885
	30	0.9354	0.9074	0.8310	0.8551	**0.9810**	0.6493
	50	0.9943	0.9894	0.9697	0.9762	**0.9990**	0.8992
$CH(0.5)$	10	0.3912	0.3546	0.2711	0.2860	**0.5728**	0.0980
	20	0.6670	0.6127	0.4979	0.5281	**0.8141**	0.2122
	30	0.8331	0.7839	0.6733	0.7102	**0.9251**	0.4089
	50	0.9669	0.9464	0.8924	0.9137	**0.9903**	0.6948

Table 5. Powerful tests against different alternatives

IFR	UFR	DFR-BFR
W^2 & T_n	A^2 & T_n	A^2

5. An Illustrative Example

Through an example, we illustrate how the proposed test can be applied to test the goodness of fit for the Lindley distribution when the observations are available.

Example 1. We consider the data set discussed by Ghitany et al. (2008). The data set consists waiting times (in minutes) before service of 100 bank customers. The waiting times (in minutes) are as follows:

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.
Histogram of these data and a fitted Lindley density function are displayed in Figure 1.

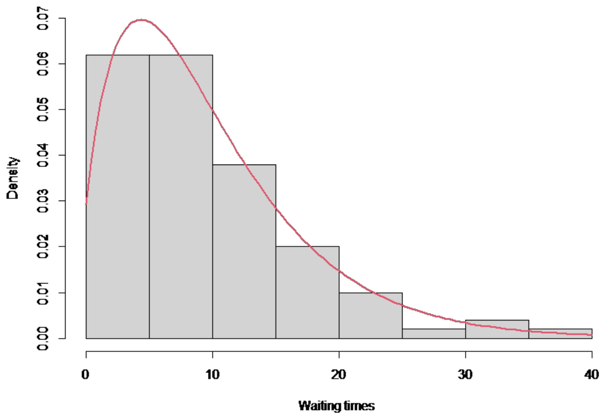

Figure 1. Histogram of data in Example 1 and a fitted Lindley density function.

Krishna and Kumar (2011) considered four reliability models, namely exponential, Lindley, gamma, and lognormal. According to Bayesian information criterion (BIC), they found that the Lindley model is the best fit for these data. Thus, Lindley distribution is fitting the above data quite satisfactorily. The main advantage of using Lindley distribution over gamma and lognormal distributions is that it involves only one parameter. Hence, maximum likelihood and other inferential procedures become simple to deal with, especially from computational point of view. The proposed procedure can be used to investigate whether the data come from a Lindley distribution. The ML estimator of θ is computed as:

$$\hat{\theta} = 0.1866\,.$$

The value of the test statistic is $T_n = 3027.424(\log{(T_n)}/n = 0.0802)$ and the critical value at the 5% is obtained from Table 1 as 0.1196. Since the values of the test statistic is smaller than the critical value, the Lindley hypothesis is accepted for these data at the significance level of 0.05. Therefore, we can conclude that the data come from a Lindley distribution.

6. Conclusions

In this paper, we have proposed a goodness of fit test for the Lindley distribution based on the empirical likelihood ratio, and have shown that the test outperform the EDF-goodness of fit tests which are commonly used in practice. We have carried out an extensive power comparison using Monte Carlo simulations. Through the obtained results, we have shown that the proposed test outperforms in most cases all other competitor tests. Finally, we have presented a real data set and have illustrated how the proposed test can be applied to test the goodness of fit for the Lindley distribution when a sample is available.

Acknowledgement

The authors are grateful to anonymous referees and the associate editor for providing some useful comments on an earlier version of this manuscript.

REFERENCES

1. Ali, S., Aslam, M., Kazmi, S. (2013), A study of the effect of the loss function on Bayes Estimate, posterior risk and hazard function for Lindley distribution, Applied Mathematical Modelling, 37, 6068–6078.
2. Al-Mutairi, D.K., Ghitany, M.E., Kundu, D. (2013), Inferences on the stress-strength reliability from Lindley distributions, Communications in Statistics-Theory and Methods, 42, 1443–1463.
3. Altun, E. (2019), Two-sided Lindley distribution with inference and applications, Journal of the Indian Society for Probability and Statistics, 20, 255–279.
4. Anderson, T.W. and Darling, D.A. (1954), A test of goodness of fit, Journal of American Statistical Association, 49, 765-769.
5. D'Agostino, R.B. and Stephens, M.A. (Eds.) (1986), Goodness of Fit Techniques, New York: Marcel Dekker.
6. Ghitany, M.E., Atieh, B. and Nadarajah, S. (2008), Lindley distribution and its Application, Mathematics Computing and Simulation, 78, 493-506.
7. Gupta, PK, Singh, B. (2013), Parameter estimation of Lindley distribution with hybrid censored data, International Journal of System Assurance Engineering and Management, 4, 378–385.
8. Gurevich, G. and Vexler, A. (2011), A Two-sample empirical likelihood ratio test based on samples entropy, Statistics and Computing, 21, 657-670.
9. Gurevich, G. and Vexler, A. (2018), A density based empirical likelihood approach for testing bivariate normality, Journal of Statistical Computation and Simulation, 88, 2540-2560.
10. Ibrahim M., Yadav, A.S., Yousof, H.M., Goual, H., Hamedani, G. (2019), A new extension of Lindley distribution: modified validation test, characterizations and different methods of estimation, Communications for Statistical Applications and Methods, 26:473-495.
11. Kolmogorov, A.N. (1933), Sulla Determinazione Empirica di une legge di Distribuzione. Giornale dell'Intituto Italiano degli Attuari, 4, 83-91.
12. Krishna, H. and Kumar, K. (2011), Reliability estimation in Lindley distribution with progressively type II right censored sample, Mathematics and Computers in Simulation, 82, 281–294.
13. Kuiper, N.H. (1960), Tests concerning random points on a circle, Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A, 63, 38-47.
14. Kumar, C.S. and Jose, R. (2019), On double Lindley distribution and some of its properties, American Journal of Mathematical and Management Sciences, 38, 23-43.
15. Kumar, K., Krishna, H. and Garg, R. (2015), Estimation of P(Y ¡ X) in Lindley distribution using progressively first failure censoring, International Journal of System Assurance Engineering and Management, 6, 330–341.
16. Lindley, D.V. (1958), Fiducial distribution and Bayes' theorem, Journal of the Royal Statistical Society, 20, 102–107.
17. Mazucheli, J. and Achcar, J.A. (2011), The Lindley distribution applied to competing risks lifetime data, Computer Methods and Programs in Biomedicine, 104, 188- 192.
18. Shan, G., Vexler, A., Wilding, G. E. and Hutson A. D. (2011), Simple and exact empirical likelihood ratio tests for normality based on moment relations, Communications in Statistics-Simulation and Computation, 40, 141–158.
19. Shanker, R., Hagos, F., Sujatha, S. (2015), On modeling of Lifetimes data using exponential and Lindley distributions, Biometrics & Biostatistics International Journal, 2, 1–9.
20. Singh, B. and Gupta, P.K. (2012), Load-sharing system model and its application to the real data set, Mathematics and Computers in Simulation, 82, 1615–1629.
21. Valiollahi, R., Asgharzadeh, A. and Ng, H.K.T. (2017), Prediction for Lindley distribution based on type-II right censored samples, Journal of the Iranian Statistical Society, 16, 1-19.
22. Vexler, A. (2020), Univariate likelihood projections and characterizations of the multivariate normal distribution, Journal of Multivariate Analysis, 179, 104643.
23. Vexler, A. and Gurevich, G. (2010), Empirical likelihood ratios applied to goodness of fit tests based on sample entropy, Computational Statistics & Data Analysis, 54, 531–545.
24. Vexler, A. and Gurevich, G. (2011), A note on optimality of hypothesis testing, Mathematics in Engineering, Science and Aerospace, 2, 243-250.
25. Vexler, A. and Yu, J. (2011), Two-sample density-based empirical likelihood tests for incomplete data in application to a pneumonia study, Biometrical Journal, 53, 628–651.
26. Vexler, A. and Zou, L. (2018), Empirical likelihood ratio tests with power one, Statistics & Probability Letters, 140, 160-166.
27. Vexler, A., Kim, Y.M., Yu, J., Lazar, N.A. and Hutson, A.D. (2014), Computing critical values of exact tests by incorporating Monte Carlo simulations combined with statistical tables, Scandinavian Journal of Statistics, 41, 1013–1030.
28. Vexler, A., Liu, S. and Schisterman, E.F. (2011c), Nonparametric likelihood inference based on cost-effectively-sampled-data, Journal of Applied Statistics, 38, 769–783.
29. Vexler, A., Shan, G., Kim, S., Tsai, W. M., L. Tian and Hutson, A.D. (2011a), An empirical likelihood ratio-based goodness of fit test for Inverse Gaussian distributions, Journal of Statistical Planning and Inference, 141, 6, 2128-2140.
30. Vexler, A., Tsai, W-M. and Malinovsky, Y. (2012a), Estimation and testing based on data subject to measurement errors: from parametric to non-parametric likelihood methods, Statistics in Medicine, 31, 2498-2512.

31. Vexler, A., Tsai, W-M., Gurevich, G. and Yu, J. (2012b), Two-sample density-based empirical likelihood ratio tests based on paired data, with application to a treatment study of Attention-Deficit/Hyperactivity Disorder and Severe Mood Dysregulation, Statistics in Medicine, 31, 1821–1837.
32. Vexler, A., Yu, J. and Hutson, A.D. (2011b), Likelihood testing populations modeled by autoregressive process subject to the limit of detection in applications to longitudinal biomedical data, Journal of Applied Statistics. 38, 1333-1346.
33. Vexler, A., Zou, L., Hutson, A.D. (2019), The empirical likelihood prior applied to bias reduction of general estimating equations, Computational Statistics & Data Analysis, 138, 96-106.
34. von Mises, R. (1931), Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik, Leipzig and Vienna: Deuticke.
35. Wang, M. (2013), A new three-parameter lifetime distribution and associated inference. arXiv:1308.4128 [stat.ME].
36. Watson, G.S. (1961), Goodness of fit tests on a circle, Biometrika, 48, 109-114.
37. Yu, J., Vexler, A. and Tian, L. (2010), Analyzing incomplete data subject to a threshold using empirical likelihood methods: an application to a pneumonia risk study in an ICU setting, Biometrics, 66, 123-130.
38. Yu, J., Vexler, A., Kim, S. and Hutson, A.D. (2011), Two-sample empirical likelihood ratio tests for medians in application to biomarker evaluations, The Canadian Journal of Statistics. 39, 671–689.
39. Zou, L., Vexler, A., Yu, J., Wan, H. (2019), A sequential density-based empirical likelihood ratio test for treatment effects, Statistics in Medicine, 38, 2115-2125.