A Goodness of Fit Test For Exponentiality Based on Lin-Wong Information

M. Abbasnejad, N. R. Arghami, M. Tavakoli

Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Iran.

Abstract. In this paper, we introduce a goodness of fit test for exponentiality based on Lin-Wong divergence measure. In order to estimate the divergence, we use a method similar to Vasicek’s method for estimating the Shannon entropy. The critical values and the powers of the test are computed by Monte Carlo simulation. It is shown that the proposed test are competitive with other tests of exponentiality based on entropy.

Keywords. Divergence measure, entropy, exponentiality test, order statistics, Vasicek’s sample entropy.

MSC: Primary: 62G10; Secondary: 94A17.

1 Introduction

In social studies, engineering, medical sciences, reliability studies and management science, it is very important to know whether the underlying data follow a particular distribution. So many authors were interested in goodness of fit tests.

Let $X$ be a continuous random variable with distribution function $F(x)$ and probability density function $f(x)$. Consider the following hy-
Potheses

\[
\begin{align*}
H_0 : & \quad f(x) = f_0(x) \\
H_1 : & \quad f(x) \neq f_0(x),
\end{align*}
\]

where \( f_0(x) = \theta e^{-\theta x}, \quad x > 0, \quad \theta > 0, \) and \( \theta \) is unknown.


The entropy of \( X \) is defined by Shannon (1948) as

\[
H(f) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx. \quad (1)
\]


Many researchers presented the goodness of fit tests based on various entropy estimators. Among these various entropy estimators, Vasicek’s sample entropy has been most widely used in goodness of fit tests.

Let \( X_1, X_2, \ldots, X_n \) be a random sample from a continuous distribution \( F \). Using \( F(x) = p \), Vasicek expressed equation (1) as

\[
H(f) = \int_{0}^{1} \log \left\{ \frac{d}{dp} F^{-1}(p) \right\} dp,
\]

and by replacing the distribution function \( F \) by the empirical distribution function \( F_n \) and using a difference operator instead of the differential operator, the derivative of \( F^{-1}(p) \) was estimated by

\[
\frac{X_{(i+m)} - X_{(i-m)}}{2m/n}.
\]

Therefore \( H(f) \) was estimated as

\[
HV_{n,m} = \frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\},
\]

where \( X_{(1)} \leq \ldots \leq X_{(n)} \) are the order statistics and \( m \) is a positive integer smaller than \( n/2 \). For \( i < 1 \), \( X_{(i)} = X_{(1)} \) and for \( i > n \), \( X_{(i)} = \)
The asymmetric Kullback-Leibler distance of $f$ from $f_0$ is:

$$D(f, f_0) = \int_0^{+\infty} f(x) \log \frac{f(x)}{f_0(x)} \, dx = -H(f) - \ln \theta + \theta E(X).$$

It is well known that $D(f, f_0) \geq 0$ and the equality holds if and only if $f(x) = f_0(x)$ almost everywhere. Ebrahimi et al. (1992) introduced an exponentiality test based on Kullback-Leibler information and estimated the test statistic by Vasicek entropy estimator. So, they introduced the test statistic as

$$TV_{n,m} = \exp(HV_{n,m}) \exp(\log(X)) + 1.$$  


Another distance of $f$ from $f_0$ is introduced by Renyi (1961) as

$$D_r(f, f_0) = \frac{1}{r - 1} \log \int_0^{+\infty} \left( \frac{f(x)}{f_0(x)} \right)^{r-1} f(x) \, dx, \quad r > 0 \neq 1.$$  

$D(f, f_0) \geq 0$ and the equality holds if and only if $f = f_0$. Abbasnejad (2011) introduced a test based on Renyi information for normality and exponentiality.

The rest of the paper is organized as follows. In Section 2, we proposed a test statistic for exponentiality based on Lin-Wong information. In Section 3, a simulation study is performed to analyze the behavior of the test statistic. We compare the proposed test with the other tests of exponentiality based on information measures.
2 Test statistics

Lin and Wong (1990) introduced a new divergence distance of two density functions \( f(x) \) and \( g(x) \) as

\[
D_{LW}(f, g) = \int_{-\infty}^{\infty} f(x) \log \frac{2f(x)}{f(x) + g(x)} \, dx.
\]

Since Lin-Wong information belongs to Csiszer family, we have \( D_{LW}(f, g) \geq 0 \) and the equality holds if and only if \( f(x) = g(x) \) (See Kapur and Kesavan, 1992). So, it motivates us to use Lin-Wong information as a test statistic for exponentiality.

Lin-Wong information in favor of \( f(x) \) against \( f_0(x) \) is

\[
D_{LW}(f, f_0) = \int_{0}^{\infty} f(x) \log \frac{2f(x)}{f(x) + \theta e^{-\theta x}} \, dx.
\] (2)

Under the null hypothesis \( D_{LW}(f, f_0) = 0 \) and large values of \( D_{LW}(f, f_0) \) favor \( H_1 \).

To estimate \( D_{LW}(f, f_0) \), we use two following methods.

In the first method, using \( F(x) = p \), similar to Vasicek’s method we express equation (2) as

\[
\int_{0}^{1} \log \frac{2(\frac{dF^{-1}(p)}{dp})^{-1}}{(\frac{dF^{-1}(p)}{dp})^{-1} + \theta e^{-(\theta F^{-1}(p))}} \, dp.
\]

Now, replacing \( F \) by \( F_n \) and using difference operator in place of the differential operator, we get an estimator \( L_V \) of \( D_{LW}(f, f_0) \) as

\[
L_V = -\frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{1}{2} + \frac{n}{4mX} (X_{(i+m)} - X_{(i-m)}) e^{-\frac{X_{(i)}}{X}} \right\},
\] (3)

where \( X_{(i)} = X_{(1)} \) for \( i < 1 \) and \( X_{(i)} = X_{(n)} \) for \( i > n \).

In equation (3), we used maximum likelihood estimator, \( 1/X \) instead of \( \theta \). It is obvious that \( L_V \) is invariant with respect to scale transformation.

It must be noted that to estimating \( D_{LW}(f, f_0) \) in the second method we use a method similar to Bowman (1992) for estimating the Shannon entropy by Kernel density function estimation. However, we do not take it into consideration here since its performance is poor in terms of powers.

Now, similar to the proof of Theorem 2 of Alizadeh and Arghami (2011b), we prove that the test based on \( L_V \) is consistent.
Theorem 2.1. Let \( F \) be an unknown continuous distribution with a positive support and \( F_0 \) be the exponential distribution with unspecified parameter. Then under \( H_1 \) the test based on \( L_V \) is consistent.

Proof. As \( n, m \to \infty \) and \( m/n \to 0 \), we have

\[
\frac{2m}{n} = \frac{F_n(X_{(i+m)}) - F_n(X_{(i-m)})}{2} \approx \frac{f(X_{(i+m)}) + f(X_{(i-m)})}{2}(X_{(i+m)} - X_{(i-m)}),
\]

where \( F_n(a) = (\# x_i \leq a)/n = (1/n) \sum I_{(-\infty,X_i]}(a) \), and \( I \) is the indicator function. Therefore noting that \( \frac{1}{X} \) is the MLE of \( \theta \) and it is consistent, we have

\[
L_V = -\frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{1}{2} + \frac{1}{X} e^{-X} \cdot \frac{1}{2} \cdot \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\}
\]

\[
\approx -\frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{1}{2} + \theta e^{-\theta X_{(i)}}, \frac{1}{2} \cdot \frac{f(X_{(i+m)}) + f(X_{(i-m)})}{f(X_{(i)})} \right\}
\]

\[
\approx -\frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{1}{2} + \theta e^{-\theta X_{(i)}}, \frac{1}{2} \cdot \frac{f(X_{(i)})}{f(X_{(i)})} \right\}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{2f(X_i)}{f(X_{(i)}) + \theta e^{-\theta X_{(i)}}} \right\} \to E(\log \left\{ \frac{2f(X)}{f(X) + \theta e^{-\theta X}} \right\})
\]

\[
= \int_{0}^{\infty} f(x) \log \left\{ \frac{2f(x)}{f(x) + f_{0}(x)} \right\} dx = D(f, f_0),
\]

where the last limit holds by the law of large numbers. So, the test based on \( L_V \) is consistent.

Remark 2.1. It may be noted that Lin-Wong information can be used to constructing general goodness of fit tests (not just for exponentiality). One can consider any known density function (with known or unknown parameters) under the null hypothesis and put it instead of the function \( g(x) \) in the definition of Lin-Wong distance to obtain the test statistic. For example, a test of normality had been considered by the authors, however, it had a poor performance.

3 Simulation study

A simulation study is performed to analyze the behavior of the proposed test statistic.
We determine the critical points using Monte Carlo simulation with 10000 replicates. For choice of $m$ we use the formula, $m = [\sqrt{n} + 0.5]$, which was used by Wieczorkowski and Grzegorzewski (1999). Table 1 gives the critical values of $L_V$ for various sample sizes.

We compute the powers of the test based on $L_V$ statistic by Monte Carlo simulation. To facilitate comparison of the power of the proposed test with powers of the tests published, we selected the same three alternatives listed in Ebrahimi et al. (1992), Gurevich and Davidson (2008) and Alizadeh and Arghami (2011a) and their choices of parameters:

(a) the Weibull distribution with density function

$$f(x; \lambda, \beta) = \beta \lambda^\beta x^{\beta-1} \exp\{-\lambda x^\beta\}, \quad x > 0, \ \beta > 0, \ \lambda > 1,$$

(b) The gamma distribution with density function

$$f(x; \lambda, \beta) = \frac{\lambda^\beta x^{\beta-1} \exp\{-\lambda x\}}{\Gamma(\beta)}, \quad x > 0, \ \beta > 0, \ \lambda > 1,$$

(c) The Log-Normal distribution with density function

$$f(x; \nu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\ln(x) - \nu)^2\right), \quad x > 0, -\infty < \nu < \infty, \ \sigma^2 > 0.$$

We also chose the parameters so that $E(X) = 1$, i.e. $\lambda = \Gamma(1 + (1/\beta))$ for the Weibull, $\lambda = \beta$ for the gamma and $\nu = -\sigma^2/2$ for the log-normal family of distributions.
The test statistics of competitor tests are as follows:

(1) Ebrahimi et al. (1992)

\[ TV_{n,m} = \frac{\exp(HV_{n,m})}{\exp(\log(X)) + 1} \]

(2) Abbasnejad (2011)

\[ ED_r^V = \log X + \frac{1}{r-1} \log \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{2m/n}{X(i+m) - X(i-m)} e^{X(i)} \right]^{r-1} \right\} \]

(3) Gurevich and Davidson (2008)

\[ MKL^1_n = \max_{1 \leq m < n/2} \left\{ \frac{n}{2meX} \left[ \prod_{j=1}^{n} (X(j+m) - X(j-m)) \right]^{1/n} \right\} \]

(4) Alizadeh and Arghami (2011a)

\[ TA_{n,m} = \frac{\exp(HA_{n,m})}{\exp(\log(X) + 1)} \]

where

\[ HA_{n,m} = -\frac{1}{n} \sum_{i=1}^{n} \log \left( \frac{\hat{f}(X(i+m)) + \hat{f}(X(i+m))}{2} \right) \]

where \( \hat{f}(X) = \frac{1}{nh} \sum_{i=1}^{n} k \left( \frac{X - X(i)}{h} \right) \), and the Kernel function is chosen to be the standard normal density function and the bandwidth \( h \) is chosen to be the normal optimal smoothing formula, \( h = 1.06sn^{-1/5} \), where \( s \) is the sample standard deviation.

The goodness of fit test based on entropy involves choosing the best integer parameter \( m \). Unfortunately, there is no choice criterion of \( m \), and in general it depends on the alternative. Ebrahimi et al. (1992) tabulated the values of \( m \), which maximize the power of the test. Similar table is given by Abbasnejad (2011). Gurevich and Davidson (2008) obtained their test statistic by maximizing the test statistic of Ebrahimi et al. (1992) over the various values of \( m \) and so they did not need to choose the best values of \( m \). It is shown that by Alizadeh and Arghami (2011a), there is no \( m \) that is optimal for all alternatives. We suggest the value of \( m \) similar to Ebrahimi et al. (1992) based on simulation results. Tables 2-4 show the estimated power of the test \( LV \) and those of the competing tests, at the significance level \( \alpha = 0.05 \) and \( \alpha = 0.01 \) based on the result of 10000 simulation (of sample size 10,20). For the
Table 2. Power comparisons against the gamma distribution at the significance levels $\alpha = 0.01, 0.05$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>$ED_r^V$</th>
<th>$TV_{n,m}$</th>
<th>$MKL_n^+$</th>
<th>$TA_{n,m}(m)$</th>
<th>$LV$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>0.01</td>
<td>0.086</td>
<td>0.136</td>
<td>0.118</td>
<td>0.122(3)</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.295</td>
<td>0.355</td>
<td>0.324</td>
<td>0.362(3)</td>
<td>0.365</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td>0.05</td>
<td>0.584</td>
<td>0.637</td>
<td>0.601</td>
<td>0.692(3)</td>
<td>0.698</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01</td>
<td>0.434</td>
<td>0.577</td>
<td>0.501</td>
<td>0.563(3)</td>
<td>0.590</td>
</tr>
<tr>
<td>0.05</td>
<td></td>
<td>0.05</td>
<td>0.788</td>
<td>0.859</td>
<td>0.790</td>
<td>0.882(3)</td>
<td>0.885</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0.01</td>
<td>0.150</td>
<td>0.244</td>
<td>0.281</td>
<td><strong>0.360(5)</strong></td>
<td>0.342</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.421</td>
<td>0.485</td>
<td>0.550</td>
<td><strong>0.646(5)</strong></td>
<td>0.629</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td>0.05</td>
<td>0.817</td>
<td>0.873</td>
<td>0.902</td>
<td><strong>0.961(5)</strong></td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.01</td>
<td>0.787</td>
<td>0.924</td>
<td>0.921</td>
<td><strong>0.972(5)</strong></td>
<td>0.958</td>
</tr>
<tr>
<td>0.05</td>
<td></td>
<td>0.05</td>
<td>0.962</td>
<td>0.986</td>
<td>0.988</td>
<td><strong>0.998(6)</strong></td>
<td>0.995</td>
</tr>
</tbody>
</table>

Table 3. Power comparisons against the Weibull distribution at the significance levels $\alpha = 0.01, 0.05$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>$ED_r^V$</th>
<th>$TV_{n,m}$</th>
<th>$MKL_n^+$</th>
<th>$TA_{n,m}(m)$</th>
<th>$LV$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>0.01</td>
<td>0.320</td>
<td>0.425</td>
<td>0.364</td>
<td>0.421(3)</td>
<td>0.425</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.668</td>
<td>0.702</td>
<td>0.662</td>
<td>0.759(1)</td>
<td>0.759</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td>0.05</td>
<td>0.832</td>
<td>0.900</td>
<td>0.831</td>
<td>0.904(3)</td>
<td>0.906</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.982</td>
<td>0.986</td>
<td>0.962</td>
<td>0.992(2)</td>
<td>0.993</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
<td>0.05</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000(3)</td>
<td>1.000</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td>0.05</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000(1)</td>
<td>1.000</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0.01</td>
<td>0.650</td>
<td>0.783</td>
<td>0.789</td>
<td><strong>0.860(5)</strong></td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.901</td>
<td>0.929</td>
<td>0.941</td>
<td><strong>0.977(1)</strong></td>
<td>0.969</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td>0.05</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000(2)</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000(1)</td>
<td>1.000</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
<td>0.05</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000(1)</td>
<td>1.000</td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td>0.05</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000(1)</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Table 4. Power comparisons against the log-normal distribution at the significance levels $\alpha = 0.01, 0.05$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>$ED^r$</th>
<th>$TV_{n,m}$</th>
<th>$MKL_n^1$</th>
<th>$TA_{n,m}(m)$</th>
<th>$L_V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>-0.3</td>
<td>0.01</td>
<td>0.073</td>
<td>0.109</td>
<td>0.100</td>
<td>0.097(3)</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.272</td>
<td>0.308</td>
<td>0.298</td>
<td>0.302(4)</td>
<td>0.317</td>
</tr>
<tr>
<td></td>
<td>-0.2</td>
<td>0.01</td>
<td>0.186</td>
<td><strong>0.302</strong></td>
<td>0.254</td>
<td>0.262(3)</td>
<td>0.280</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.511</td>
<td>0.602</td>
<td>0.554</td>
<td>0.593(3)</td>
<td>0.606</td>
</tr>
<tr>
<td></td>
<td>-0.1</td>
<td>0.01</td>
<td>0.637</td>
<td>0.784</td>
<td>0.704</td>
<td>0.775(3)</td>
<td><strong>0.796</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.921</td>
<td>0.954</td>
<td>0.918</td>
<td>0.968(3)</td>
<td><strong>0.961</strong></td>
</tr>
<tr>
<td>20</td>
<td>-0.3</td>
<td>0.01</td>
<td>0.128</td>
<td>0.223</td>
<td>0.236</td>
<td>0.276(7)</td>
<td><strong>0.278</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.397</td>
<td>0.478</td>
<td>0.499</td>
<td><strong>0.569</strong>(8)</td>
<td>0.534</td>
</tr>
<tr>
<td></td>
<td>-0.2</td>
<td>0.01</td>
<td>0.378</td>
<td>0.591</td>
<td>0.621</td>
<td><strong>0.727</strong>(5)</td>
<td>0.679</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.735</td>
<td>0.827</td>
<td>0.848</td>
<td><strong>0.921</strong>(8)</td>
<td>0.872</td>
</tr>
<tr>
<td></td>
<td>-0.1</td>
<td>0.01</td>
<td>0.938</td>
<td>0.991</td>
<td>0.991</td>
<td>0.992(6)</td>
<td><strong>0.993</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.05</td>
<td>0.997</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000(6)</td>
<td><strong>1.000</strong></td>
</tr>
</tbody>
</table>

test statistic $TA_{n,m}$, the value of $m$ which maximizes the power of the test for each alternative is given in parentheses.

According to Tables 2, 3 and 4, the $L_V$ test behaves better than other tests for $n = 10$ and for all of the alternatives (except log-normal(-0.2)). However, for $n = 20$, the $TA_{n,m}$ test is better than the other tests and the $L_V$ test has greater or equal powers for some alternatives. So we can suggest the $L_V$ test statistic for small sample sizes. Also, for large sample sizes, the $L_V$ test has the advantage of having fixed $m$, in comparison with $TA_{n,m}$ and one may prefer the proposed test.

### 4 Conclusion

In this paper, we introduced a goodness of fit test for exponentiality based on Lin-Wong divergence measure. To construct the test statistic we estimated the Lin-Wong distance similar to Vasicek’s method for estimating of the Shannon entropy. By a simulation study the powers of the proposed test were computed under several alternatives and different sample sizes. It is shown that, $L_V$ test compares favorably with the leading competitors specially for small sample sizes.
Acknowledgements

The authors would like to thank the referees for their attention to the topic and useful comments and suggestions. Partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is acknowledged.

References


