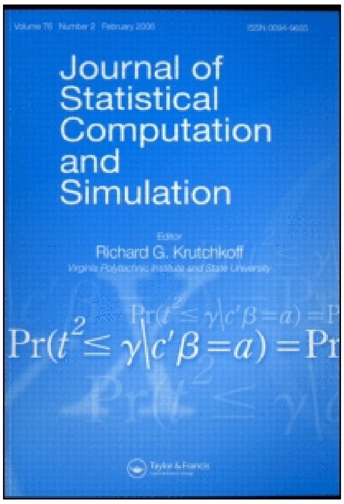

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713650378

Testing exponentiality using transformed data

Hadi Alizadeh Noughabi[a]; Naser Reza Arghami[a]
[a] Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

First published on: 13 April 2010

To cite this Article Alizadeh Noughabi, Hadi and Arghami, Naser Reza(2011) 'Testing exponentiality using transformed data', Journal of Statistical Computation and Simulation, 81: 4, 511 — 516, First published on: 13 April 2010 (iFirst)
To link to this Article: DOI: 10.1080/00949650903348171
URL: http://dx.doi.org/10.1080/00949650903348171

Journal of Statistical Computation and Simulation
Vol. 81, No. 4, April 2011, 511–516

Testing exponentiality using transformed data

Hadi Alizadeh Noughabi* and Naser Reza Arghami

Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

(*Received 14 April 2009; final version received 18 September 2009*)

In this paper, we introduce a test for uniformity and use it as the second stage of an exact goodness-of-fit test of exponentiality. By simulation, the powers of the proposed test under various alternatives are compared with exponentiality test based on Kullback–Leibler information proposed by Ebrahimi *et al*. [N. Ebrahimi, M. Habibullah, and E.S. Soofi, *Testing exponentiality based on Kullback–Leiber information*, J. R. Statist. Soc. Ser. B 54 (1992), pp. 739–748]. The results are impressive, i.e. the proposed test has higher power than the test based on entropy.

Keywords: nonparametric kernel density estimation; test for uniformity; testing exponentiality

1. Introduction

In numerous applications in reliability studies and engineering and management sciences, it is very important to test whether the underlying distribution has a particular form. Most statistical methods assume an underlying distribution in the derivation of their results. However, when we assume that our data follow a specific distribution, we take a serious risk. If our assumption is wrong, then the results obtained may be invalid. For example, the confidence levels of the confidence intervals or error probabilities of tests of hypotheses implemented may be completely off. The consequences of mis-specifying the distribution may prove very costly. One way to deal with this problem is to check the distribution assumptions carefully.

Many researchers have been interested in testing exponentiality. See, for example, [1–4]. Therefore different exponentiality tests have been developed.

In Section 2, we propose a new test statistic for testing uniformity and compare the performance of the proposed test with the existing tests. In Section 3, we introduce a test for exponentiality based on the test introduced in Section 2 and show, by simulation, that it has higher power than Ebrahimi *et al*. [4] test.

*Corresponding author. Email: alizadehhadi@ymail.com

ISSN 0094-9655 print/ISSN 1563-5163 online

DOI: 10.1080/00949650903348171
http://www.informaworld.com

2. Testing uniformity

2.1. *Test statistic and critical points*

Given a random sample $X_1, \ldots, X_n$ from a population with absolutely continuous density function $f(x)$ concentrated on the interval [0,1], distribution function $F(x)$; consider the problem of testing the hypothesis H_0 that the X_i are uniformly distributed, denoted by $U(0, 1)$. For this test we suggest the following test statistic:

$$T = \frac{1}{n}\sum_{i=1}^{n} |x_i \cdot \hat{f}(x_i) - F_0(x_i)|,$$

where $F_0(x)$ is the uniform distribution function and $|\cdot|$ is an absolute function. Also,

$$\hat{f}(x_i) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x_i - x_j}{h}\right),$$

where the kernel function is chosen to be the standard normal density function and the bandwidth h is obtained from the normal optimal smoothing formula, $h = 1.06sn^{-1/5}$, where s is the sample standard deviation (Silverman [9]).

Since the density function and the distribution function of a uniform random variable are equal to 1 and x, respectively, it is obvious that large values of T indicate that the sample is from a non-uniform distribution. Therefore, we reject H_0 at the significance level α if $T \geq C(\alpha)$, where the critical point $C(\alpha)$ is determined by the α quantile of the distribution of the T-statistic under the hypothesis H_0. We determine the critical points $C(\alpha)$ of the T-statistic by means of Monte Carlo simulations for α equal to 0.01, 0.05 and 0.1. For $n \leq 50$, we generated 10,000 uniform samples of size n and computed T. The upper α quantile of the empirical distribution of T was used to determine $C(\alpha)$. Table 1 gives the critical values $C(\alpha)$ for several sample sizes.

2.2. *Power comparison*

In this subsection, we compare the proposed test with the existing tests, namely, Anderson–Darling A^2, Cramer–von Mises W^2, Watson U^2, Kolmogrov–Smirnov D, Kuiper V and the test based on the entropy measure K_{mn} [5], in terms of power.

Table 1. Critical values of T-statistic.

	α		
n	0.01	0.05	0.10
5	1.002	0.518	0.339
10	0.407	0.264	0.202
15	0.295	0.206	0.164
20	0.256	0.176	0.150
25	0.228	0.164	0.141
30	0.202	0.152	0.132
50	0.159	0.130	0.117

The (continuous) alternatives studied consist of those specified by the following distribution functions F:

$$A_k : F(x) = 1 - (1-x)^k, \quad 0 \le x \le 1 \ \ (for\ k = 1.5, 2),$$

$$B_k : F(x) = \begin{cases} 2^{k-1}x^k, & 0 \le x \le 0.5 \\ 1 - 2^{k-1}(1-x)^k, & 0.5 \le x \le 1 \end{cases} \ \ (for\ k = 1.5, 2, 3),$$

$$C_k : F(x) = \begin{cases} 0.5 - 2^{k-1}(0.5-x)^k, & 0 \le x \le 0.5 \\ 0.5 + 2^{k-1}(x-0.5)^k, & 0.5 \le x \le 1 \end{cases} \ \ (for\ k = 1.5, 2).$$

Alternatives A, B and C were used by Stephens [6] in his study of power comparisons of several tests for uniformity. According to Stephens, alternative A gives points closer to zero than expected under the hypothesis of uniformity, whereas B gives points near 0.5 and C give two clusters (close to 0 and 1). The same alternatives were used by Dudewicz and Van der Meulen [5].

The power estimates resulting from our Monte Carlo study against the seven alternatives considered are given in Tables 2 and 3 for $\alpha = 0.05, 0.10$ and $n = 10, 20$. The values in these tables are based on 10,000 simulated samples of size n.

Table 2. Power comparisons of tests with size 0.05.

n	Alternative	T	K_{mn}	D	W^2	V	U^2	A^2
10	$A_{1.5}$	0.041	0.142	0.159	0.169	0.101	0.103	0.163
	A_2	0.050	0.301	0.400	0.435	0.232	0.224	0.417
	$B_{1.5}$	0.200	0.217	0.040	0.027	0.130	0.137	0.015
	B_2	0.418	0.492	0.048	0.023	0.313	0.339	0.010
	B_3	0.821	0.878	0.095	0.053	0.713	0.760	0.021
	$C_{1.5}$	0.026	0.017	0.112	0.099	0.128	0.141	0.127
	C_2	0.025	0.014	0.206	0.158	0.311	0.335	0.235
20	$A_{1.5}$	0.057	0.239	0.281	0.316	0.167	0.164	0.318
	A_2	0.091	0.617	0.699	0.770	0.468	0.440	0.761
	$B_{1.5}$	0.307	0.366	0.056	0.039	0.224	0.246	0.035
	B_2	0.682	0.778	0.122	0.101	0.591	0.651	0.103
	B_3	0.981	0.994	0.411	0.508	0.969	0.984	0.561
	$C_{1.5}$	0.019	0.018	0.149	0.122	0.225	0.243	0.162
	C_2	0.018	0.017	0.310	0.248	0.593	0.652	0.378

Table 3. Power comparisons of tests with size 0.10.

n	Alternative	T	K_{mn}	D	W^2	V	U^2	A^2
10	$A_{1.5}$	0.105	0.229	0.245	0.266	0.178	0.177	0.254
	A_2	0.137	0.451	0.526	0.579	0.346	0.339	0.560
	$B_{1.5}$	0.330	0.332	0.090	0.074	0.221	0.229	0.049
	B_2	0.610	0.621	0.121	0.088	0.453	0.474	0.060
	B_3	0.924	0.934	0.242	0.231	0.821	0.863	0.185
	$C_{1.5}$	0.044	0.045	0.198	0.174	0.218	0.228	0.208
	C_2	0.045	0.041	0.308	0.258	0.454	0.480	0.365
20	$A_{1.5}$	0.148	0.386	0.403	0.446	0.268	0.271	0.437
	A_2	0.211	0.770	0.813	0.862	0.612	0.595	0.860
	$B_{1.5}$	0.447	0.545	0.128	0.109	0.351	0.373	0.100
	B_2	0.798	0.895	0.262	0.273	0.733	0.783	0.289
	B_3	0.995	0.999	0.671	0.805	0.990	0.995	0.837
	$C_{1.5}$	0.050	0.037	0.250	0.210	0.345	0.377	0.263
	C_2	0.062	0.041	0.469	0.422	0.731	0.782	0.549

We see from Tables 2 and 3 that the power of the proposed test is generally less than that of the entropy test introduced by Dudewicz and Van der Meulen [5], while it is greater than the powers of other tests for B alternatives while the reverse holds for other alternatives.

Since, in this paper, our aim is to introduce a more efficient test of exponentiality, the prevalent low powers of the above test, which is going to be used as a tool in the test introduced in the next section, is not of any importance.

3. Testing exponentiality using transformed data

Given a random sample $X_1, \ldots, X_n$ from a continuous probability distribution F with a density $f(x)$ over a non-negative support and with mean $\mu < \infty$, the hypothesis of interest is

$$H_0 : f(x) = f_0(x) = \lambda \exp(-\lambda x),$$

where $\lambda = 1/\mu$ is unspecified. The alternative to H_0 is

$$H_1 : f(x) \neq f_0(x).$$

In order to obtain a test statistic, we use the following theorem, which is proved in Alzaid and Al-Osh [7] and is also mentioned in Balakrishnan and Basu [8].

THEOREM 3.1 *Let X_1 and X_2 be two independent observations from a distribution F. Then $X_1/(X_1 + X_2)$ is distributed as $U(0, 1)$ if and only if F is exponential.*

Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the order statistics of a random sample of size n. First, we transform the sample data to

$$Y_{ij} = \frac{X_{(i)}}{X_{(i)} + X_{(j)}}, \quad i \neq j, i, j = 1, 2, \ldots, n.$$

By the above theorem, under the null hypothesis, each Y_i has a uniform distribution, and it seems to be appropriate to use our proposed test for uniformity (introduced in Section 2) to test the uniformity of the distribution of Y_i's and thus the exponentiality of the distribution of X_i's. Therefore, the summary of the test is as

$$X_1, \ldots, X_n \longrightarrow Y_{ij} = \frac{X_{(i)}}{X_{(i)} + X_{(j)}}, \quad i \neq j \Longrightarrow T = \frac{1}{n'} \sum_{i=1}^{n'} |y_i \hat{f}(y_i) - F_0(y_i)|,$$

where $n' = n(n-1)$.

Large values of T indicate that the sample is from a non-exponential distribution.

For small to moderate sample sizes 5, 10, 15, 20, 25, 30 and 50, we used Monte Carlo methods with 10,000 replicates from the exponential distribution with mean one to obtain critical values of our procedure. These values are reported in Table 4.

To facilitate comparisons of the power of the present test with the powers of the tests published, we selected the same three alternatives listed in Ebrahimi *et al*. [4] and their choices of parameters are:

(1) the Weibull distribution with density function

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

Table 4. Critical values of T-statistic.

	α		
n	0.01	0.05	0.10
5	0.7365	0.3350	0.2120
10	0.3320	0.1965	0.1422
15	0.2401	0.1588	0.1221
20	0.2123	0.1369	0.1094
25	0.1743	0.1195	0.0993
30	0.1594	0.1167	0.0983
50	0.1281	0.0915	0.0837

Table 5. Monte Carlo power estimates of the KL_{mn} and T tests against the gamma distribution.

n	β	α	KL_{mn}	T
10	2	0.01	0.101	0.115
		0.05	0.315	0.334
	3	0.01	0.284	0.317
		0.05	0.627	0.678
	4	0.01	0.485	0.531
		0.05	0.822	0.860
20	2	0.01	0.228	0.313
		0.05	0.502	0.629
	3	0.01	0.658	0.790
		0.05	0.889	0.953
	4	0.01	0.898	0.960
		0.05	0.982	0.997

(2) the gamma distribution with density function

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

(3) the log-normal distribution with density function

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

Table 6. Monte Carlo power estimates of the KL_{mn} and T tests against the Weibull distribution.

n	β	α	KL_{mn}	T
10	2	0.01	0.345	0.366
		0.05	0.681	0.724
	3	0.01	0.855	0.852
		0.05	0.981	0.985
	4	0.01	0.986	0.987
		0.05	1.000	1.000
20	2	0.01	0.734	0.821
		0.05	0.933	0.961
	3	0.01	0.999	1.000
		0.05	1.000	1.000
	4	0.01	1.000	1.000
		0.05	1.000	1.000

Table 7. Monte Carlo power estimates of the KL_{mn} and T tests against the log-normal distribution.

n	v	α	KL_{mn}	T
10	−0.3	0.01	0.083	0.094
		0.05	0.285	0.334
	−0.2	0.01	0.228	0.259
		0.05	0.560	0.636
	−0.1	0.01	0.690	0.764
		0.05	0.938	0.980
20	−0.3	0.01	0.198	0.304
		0.05	0.475	0.600
	−0.2	0.01	0.560	0.713
		0.05	0.835	0.941
	−0.1	0.01	0.985	0.998
		0.05	1.000	1.000

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 $v = -\sigma^2/2$ for the log-normal family of distributions.

We estimated the powers of the Ebrahimi *et al*. test with the powers the present test based on 10,000 samples of size n equal to 10 and 20. Tables 5–7 show the estimated powers at significance levels $\alpha = 0.01$ and $\alpha = 0.05$. The powers reported for the test based on KL_{mn}-statistic are based on the window sizes reported in Ebrahimi *et al*. [4], which give the maximum power for their test.

We observe that the proposed test performs very well compared with the test based on KL_{mn}-statistic for the Weibull, gamma and log-normal alternatives. Also, it can be seen that the relative superiority of the proposed test over Ebrahimi *et al*. test increases with sample size.

Acknowledgements

Partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is acknowledged.

References

[1] H.W. Lilliefors, *On the Kolmogorov–Smirnov test for the exponential distribution with mean unknown*. J. Am. Statist. Assoc. 64 (1969), pp. 387–389.
[2] J. Van-Soest, *Some goodness of fit tests for the exponential distribution.* Statist. Neerlandica 23 (1969), pp. 41–51.
[3] J. Finkelstein and R.E. Schafer, *Important goodness of fit tests*, Biometrika 58 (1971), pp. 641–645.
[4] N. Ebrahimi, M. Habibullah, and E.S. Soofi, *Testing exponentiality based on Kullback–Leiber information*, J. R. Statist. Soc. Ser. B 54 (1992), pp. 739–748.
[5] E. Dudewicz and E. Van der Meulen, *Entropy Based Tests of Uniformity*, J. Am. Statist. Assoc. 76 (1981), pp. 967–974.
[6] M.A. Stephens, *EDF statistics for goodness of fit and some comparisons*, J. Am. Statist. Assoc. 69 (1974), pp. 730–737.
[7] A.A. Alzaid and M.A. Al-Osh, *Characterization of probability distributions based on the relation* $X = U(X_1 + X_2)$, Sankhya Ser. B 53 (1992), pp. 188–190.
[8] N. Balakrishnan and A.P. Basu, *The Exponential Distribution: Theory, Methods and Applications*, Gordon and Breach Publishers, Amsterdam, 1995.
[9] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.