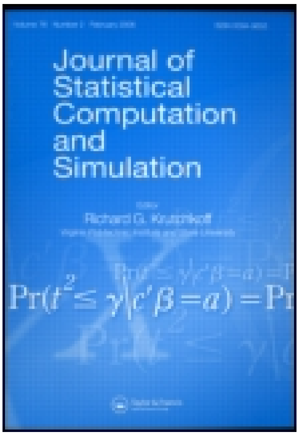

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

Testing normality based on new entropy estimators

Ehsan Zamanzade [a] & Nasser Reza Arghami [a]

[a] Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran

Version of record first published: 08 Jul 2011.

Journal of Statistical Computation and Simulation
Vol. 82, No. 11, November 2012, 1701–1713

Testing normality based on new entropy estimators

Ehsan Zamanzade* and Nasser Reza Arghami

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

(*Received 12 December 2009; final version received 28 May 2011*)

In this paper, we first introduce two new estimators for estimating the entropy of absolutely continuous random variables. We then compare the introduced estimators with the existing entropy estimators, including the first of such estimators proposed by Dimitriev and Tarasenko [*On the estimation functions of the probability density and its derivatives*, Theory Probab. Appl. 18 (1973), pp. 628–633]. We next propose goodness-of-fit tests for normality based on the introduced entropy estimators and compare their powers with the powers of other entropy-based tests for normality. Our simulation results show that the introduced estimators perform well in estimating entropy and testing normality.

Keywords: information theory; entropy estimator; testing normality

1. Introduction

Suppose a random variable X has a distribution function $F(x)$ with an absolutely continuous density function $f(x)$. The entropy $H(f)$ of the random variable is defined by Shannon [1] to be

$$H(f) = -\int_{-\infty}^{\infty} f(x)\log f(x)\,\mathrm{d}x. \tag{1}$$

There are several areas in statistics, the area of goodness-of-fit testing in particular, that use the concept of entropy as a key element. Many researchers have been concerned with the problem of estimating entropy of continuous random variables. Dimitriev and Tarasenko [2] proposed estimator using kernel density estimation, while Vasicek [3], Ebrahimi *et al.* [4] and Correa [5] directly obtained non-parametric entropy estimators. Many researchers have shown interest in developing entropy-based goodness-of-fit tests including Vasicek [3], Arizono and Ohta [6], Dudewicz and van der Muelen [7], Mudholkar and Lin [8], Ebrahimi *et al.* [9], Bowman [10], Lund and Jammalamadaka [11], Esteban *et al.* [12], Park and Park [13], Choi *et al.* [14], Goria *et al.* [15], Yousefzadeh and Arghami [16], Mahdizadeh and Arghami [17], Alizadeh Noughabi and Arghami [18] and Zamanzade and Arghami [19], most of which use Vasicek's [3] version of entropy estimator because of its simplicity and surprisingly relative precision.

*Corresponding author. Email: ehsanzamanzadeh@yahoo.com

ISSN 0094-9655 print/ISSN 1563-5163 online

http://dx.doi.org/10.1080/00949655.2011.592984
http://www.tandfonline.com

Dimitriev and Tarasenko [2] propose to estimate the entropy of an absolutely continuous random variable by

$$HD = -\int_{-\infty}^{\infty} \log(\hat{f}(x))\hat{f}(x)\,\mathrm{d}x,$$

where $\hat{f}(x)$ is the kernel density estimation of $f(x)$, which is given by the formula

$$\hat{f}(X_i) = \frac{1}{nh}\sum_{j=1}^{n} k\left(\frac{X_i - X_j}{h}\right), \tag{2}$$

where h is a bandwidth and k is a kernel function which satisfies

$$\int_{-\infty}^{\infty} k(x)\,\mathrm{d}x = 1$$

and is usually a symmetric probability density function.

Vasicek's estimator is based on the fact that Equation (1) can be expressed as

$$H(f) = \int_0^1 \log\left\{\frac{\mathrm{d}}{\mathrm{d}p}F^{-1}(p)\right\}\,\mathrm{d}p.$$

The estimate was constructed 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 $F^{-1}(p)$ is then estimated by a function of the order statistics.

Assuming that $X_1, \ldots, X_n$ is a random sample, the estimator is given by

$$HV_{mn} = \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\}, \tag{3}$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are order statistics based on a random sample of size n, the window size m is a positive integer smaller than $n/2$ and $X_{(i)} = X_{(1)} i < 1$ or $X_{(i)} = X_{(n)} i > n$.

Ebrahimi *et al.* 's [4] entropy estimator was obtained by modifying the denominator of Equation (3) in order to assign smaller weights to observations in Equation (3) that are replaced by $X_{(1)}$ and $X_{(n)}$. Their proposed entropy estimator is

$$HE_{mn} = \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{n}{c_i m}(X_{(i+m)} - X_{(i-m)})\right\},$$

where

$$c_i = \begin{cases} 1 + \frac{i-1}{m}, & 1 \leq i \leq m, \\ 2, & m+1 \leq i \leq n-m, \\ 1 + \frac{n-i}{m}, & n-m+1 \leq i \leq n. \end{cases}$$

They carried on a simulation study to show that their estimator has smaller bias and smaller mean squared error than Vasicek's [3] and Dudewicz and Van der Meulen's [7] estimators.

Correa [5] proposed another modification of Vasicek's [3] estimator, which produces smaller MSEs, by rewriting Equation (3) as

$$HV_{mn} = -\frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{\hat{F}_n(X_{(i+m)}) - \hat{F}_n(X_{(i-m)})}{X_{(i+m)} - X_{(i-m)}}\right\},$$

and noting that inside the brackets in the above equation is the estimate of the slope of the straight line that joins points $(X_{(i+m)}, \hat{F}_n(X_{(i+m)}))$ and $(X_{(i-m)}, \hat{F}_n(X_{(i-m)}))$. He estimated this slope by local linear regression on $\{X_{(i-m)}, \ldots, X_{(i+m)}\}$, using all of $2m + 1$ points instead of using only the two end points.

He considered the relation:

$$\hat{F}_n(x_j) = \alpha + \beta x_j + \varepsilon, \quad j = m - i, \ldots, m + i,$$

and estimated the slope β by its least square estimator; thus, he introduced his estimator of entropy as

$$HC_{mn} = \frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{\sum_{j=i-m}^{i+m} (X_{(j)} - \bar{X}_{(i)})(j - i)}{n \sum_{j=i-m}^{i+m} (X_{(j)} - \bar{X}_{(i)})^2} \right\},$$

where

$$\bar{X}_{(i)} = \frac{1}{2m + 1} \sum_{j=i-m}^{i+m} X_{(j)}.$$

Van Es [20] introduced the bias-corrected entropy estimator defined by

$$HEs_{mn} = \frac{1}{n - m} \sum_{i=1}^{n-m} \left(\frac{n + 1}{m} \log(X_{(i+m)} - X_{(i)}) \right) + \sum_{k=m}^{n} \frac{1}{k} + \log(m) - \log(n + 1),$$

and established its asymptotic normality.

In Section 2 of this paper, we first introduce two new entropy estimators which are based on another approximation for $F(X_{(i+m)}) - F(X_{(i-m)}), \quad i = 1, \ldots, n$, and then we compare our proposed estimators with entropy estimators proposed by Dimitriev and Tarasenko [2], Vasicek [3], Van Es [20], Ebrahimi *et al.* [4] and Correa [5]. In Section 3, we introduce goodness-of-fit tests for normality based on the proposed entropy estimators and then compare their powers with the powers of other entropy-based tests of normality. Section 4 contains some concluding remarks.

2. New entropy estimators

In this section, we first introduce new entropy estimators and then compare their performances with those of the leading competitors.

2.1. *Introduction of the estimators*

Suppose $X_1, \ldots, X_n$ is a random sample from an unknown absolutely continuous distribution F with a probability density function $f(x)$. To obtain our first entropy estimator, we note that $H(f)$ can be approximated by its sample mean, thus we have

$$\begin{aligned} H(f) &= -\int_{-\infty}^{\infty} f(x) \log f(x) \\ &= -E(\log f(x)) \\ &\approx -\frac{1}{n} \sum_{j=1}^{n} \log f(x_{(i)}), \end{aligned}$$

where '$\approx$' indicates approximate equality. Using Vasicek [3] difference operator, we have

$$f(x_{(i)}) \approx \frac{F(x_{(i+m)}) - F(x_{(i-m)})}{x_{(i+m)} - x_{(i-m)}},$$

where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ are order statistics based on a random sample of size n, the window size m is a positive integer smaller than $n/2$; thus, we have

$$H(f) \approx \frac{1}{n}\sum_{j=1}^{n}\log\left(\frac{x_{(i+m)} - x_{(i-m)}}{F(x_{(i+m)}) - F(x_{(i-m)})}\right).$$

Here, $F(x_{(i+m)}) - F(x_{(i-m)}), \quad i = 1, \ldots, n$, is taken to be the area, under the curve of f, between $x_{(i+m)}$ and $x_{(i-m)}$. We can approximate this area by

$$F\left(x_{(i+m)}\right) - F\left(x_{(i-m)}\right) = \sum_{j=k_1(i)}^{k_2(i)-1}\left(\frac{f(x_{(j+1)}) + f(x_{(j)})}{2}\right)(x_{(j+1)} - x_{(j)}),$$

where

$$k_1(i) = \begin{cases} 1 & \textit{if } \; i \leq m, \\ i - m & \textit{if } \; i > m, \end{cases}$$

$$k_2(i) = \begin{cases} i + m & \textit{if } \; i \leq n - m, \\ i - m & \textit{if } \; i > n - m. \end{cases}$$

Now values of $f(x_{(i)})$ can be estimated by a kernel density estimator (2).

Therefore, we can estimate the entropy $H(f)$ of an unknown continuous probability density function f by

$$HZ1_{mn} = \frac{1}{n}\sum_{i=1}^{n}\log\{b_i\},$$

where

$$b_i = \frac{X_{(i+m)} - X_{(i-m)}}{\sum_{j=k_1(i)}^{k_2(i)-1}\left(\frac{f(X_{(j+1)})+f(X_{(j)})}{2}\right)(X_{(j+1)} - X_{(j)})},$$

$$\hat{f}(X_i) = \frac{1}{nh}\sum_{j=1}^{n} k\left(\frac{X_i - X_j}{h}\right),$$

where k 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 $s(= \sqrt{1/n\sum_{i=1}^{n}(X_i - \bar{X})^2})$ is the sample standard deviation (SD) [21].

To obtain our second estimator, we note that, in the first estimator we do not use equal number of points for computing b_i's. Thus, it seems reasonable to use weights proportional to the number of points used in deriving b_i's. Therefore, $HZ1_{mn}$can be modified to give

$$HZ2_{mn} = \sum_{i=1}^{n} w_i \log\{b_i\},$$

where

$$w_i = \begin{cases} \frac{m+i-1}{\sum_{i=1}^{n} w_i} & \text{if } 1 \leq i \leq m, \\ \frac{2m}{\sum_{i=1}^{n} w_i} & \text{if } m+1 \leq i \leq n-m, \qquad i = 1, \ldots, n. \\ \frac{n-i+m}{\sum_{i=1}^{n} w_i} & \text{if } n-m+1 \leq i \leq n, \end{cases}$$

are weights proportional to the number of points used in computation of b_i's.

The following theorem which states that the scale of the random variable X has no effect on the accuracy of $HZ1_{mn}$ and $HZ2_{mn}$ in estimating $H(f)$, can easily be proved by following the line of argument in Ebrahimi *et al.* [4].

THEOREM 1 *Let $X_1, \dots, X_n$ be a sequence of iid random variables with entropy $H^X(f)$ and let $Y_i = cX_i$, $i = 1, \dots, n$, where $c > 0$. Let HZi^X_{mn} and HZi^Y_{mn}, $i = 1, 2$, be entropy estimators for $H^X(f)$ and $H^Y(g)$, respectively. (here g is the probability distribution function of $Y = cX$). Then the following proprieties hold:*

(*i*) $E(HZi^Y_{mn}) = E(HZi^X_{mn}) + \log c, \ i = 1, 2,$

(*ii*) $\mathrm{Var}(HZi^Y_{mn}) = \mathrm{Var}(HZi^X_{mn}), \ i = 1, 2,$

(*iii*) $\mathrm{MSE}(HZi^Y_{mn}) = \mathrm{MSE}(HZi^X_{mn}), \ i = 1, 2.$

2.2. *RMSE comparisons of entropy estimators*

In this subsection, we report the results of a simulation study which compares the performances of the introduced entropy estimators with the estimators proposed by Dimitriev and Tarasenko [2], Vasicek [3], van Es [20], Correa [5] and Ebrahimi *et al.* [4] in terms of their standard deviation (SDs) and their root mean square errors (RMSEs). For selected values of n, $N = 10,000$ samples of size n were generated from normal, exponential and uniform distributions which are the same three distributions considered by Ebrahimi *et al.* [4] and Correa [5].

Since Dimitriev and Tarasenko [2] make no mention of selecting the kernel function and the value of h in their kernel estimator, we used the standard normal density as the kernel and its corresponding optimal value of h which is equals to $h = 1.06sn^{-1/n}$.

The optimal choice of m (which results in the minimum value of RMSE for given n) is still an open problem. We use a heuristic formula proposed by Grzegorzewski and Wieczorkowski [22] for choosing m subject to n, namely

$$m = [\sqrt{n} + 0.5],$$

for choosing m and computing RMSE and SD of the m-spacing entropy estimators.

Tables 1– 3 contain the RMSE values (and SDs) of the seven estimators at different sample sizes for each of the three considered distributions.

We observe that the proposed estimators compare favorably with the five competitors; in the case of the normal distribution, the proposed estimators perform well in comparison with other m-spacing entropy estimators and in the case of the exponential distribution, the proposed estimators and the entropy estimator proposed by Van Es [20] are quite competitive, but our entropy estimators

Table 1. Root mean square error and bias of the estimators of the entropy of the standard normal distribution.

		RMSE(SD)						
n	m	HD	HV_{mn}	HE_{mn}	HEs_{mn}	HC_{mn}	$HZ1_{mn}$	$HZ2_{mn}$
5	2	0.400 (0.400)	0.994 (0.425)	0.666 (0.425)	0.509 (0.452)	0.793 (0.418)	0.494 (0.407)	0.493 (0.407)
10	3	0.257 (0.247)	0.618 (0.269)	0.408 (0.269)	0.366 (0.283)	0.470 (0.271)	0.303 (0.255)	0.310 (0.255)
15	4	0.213 (0.198)	0.474 (0.211)	0.294 (0.211)	0.318 (0.220)	0.348 (0.213)	0.222 (0.193)	0.232 (0.192)
20	4	0.186 (0.168)	0.373 (0.179)	0.247 (0.179)	0.276 (0.185)	0.265 (0.182)	0.190 (0.170)	0.205 (0.169)
30	5	0.156 (0.134)	0.282 (0.144)	0.186 (0.144)	0.243 (0.148)	0.194 (0.146)	0.148 (0.135)	0.165 (0.135)

Table 2. Root mean square error and bias of the estimators of the entropy of the exponential distribution with mean one.

		RMSE(SD)						
n	m	HD	HV_{mn}	HE_{mn}	HEs_{mn}	HC_{mn}	$HZ1_{mn}$	$HZ2_{mn}$
5	2	0.633 (0.581)	0.930 (0.559)	0.651 (0.559)	0.596 (0.586)	0.743 (0.554)	0.575 (0.573)	0.576 (0.575)
10	3	0.522 (0.394)	0.570 (0.360)	0.404 (0.360)	0.392 (0.373)	0.435 (0.361)	0.391 (0.383)	0.389 (0.383)
15	4	0.482 (0.320)	0.421 (0.284)	0.308 (0.284)	0.310 (0.290)	0.328 (0.290)	0.327 (0.306)	0.321 (0.305)
20	4	0.454 (0.267)	0.356 (0.242)	0.263 (0.242)	0.274 (0.250)	0.272 (0.247)	0.300 (0.261)	0.286 (0.260)
30	5	0.426 (0.219)	0.276 (0.198)	0.201 (0.187)	0.227 (0.201)	0.208 (0.197)	0.266 (0.209)	0.245 (0.208)

Table 3. Root mean square error and bias of the estimators of the entropy of the uniform distribution one (0,1).

		RMSE(SD)						
n	m	HD	HV_{mn}	HE_{mn}	HEs_{mn}	HC_{mn}	$HZ1_{mn}$	$HZ2_{mn}$
5	2	0.404 (0.331)	0.774 (0.346)	0.450 (0.446)	0.407 (0.407)	0.566 (0.336)	0.330 (0.327)	0.330 (0.326)
10	3	0.324 (0.175)	0.455 (0.167)	0.235 (0.167)	0.216 (0.216)	0.295 (0.169)	0.179 (0.176)	0.180 (0.178)
15	4	0.296 (0.130)	0.343 (0.110)	0.159 (0.110)	0.155 (0.155)	0.208 (0.112)	0.137 (0.123)	0.136 (0.127)
20	4	0.283 (0.107)	0.274 (0.087)	0.133 (0.087)	0.126 (0.121)	0.157 (0.088)	0.125 (0.100)	0.121 (0.105)
30	5	0.263 (0.081)	0.210 (0.059)	0.096 (0.059)	0.086 (0.086)	0.110 (0.061)	0.112 (0.073)	0.104 (0.078)

are better in small sample sizes. In the case of the uniform distribution our estimators outperform all competitors in most cases. We also observe that the proposed estimators are approximately equivalent but $HZ1_{mn}$ is slightly better than $HZ2_{mn}$.

3. Testing normality

Normal distribution is the most important distribution in statistics and has a predominant presence in statistical inference. Many statistical techniques are based on the assumption that the data come from this well-known, Bell-shaped distribution. Consequently, the results of these techniques can be completely unreliable if the normality assumption is violated. Thus it becomes very important to check this assumption in an appropriate and efficient way and that is why goodness-of-fit techniques, especially for normal distribution, have attracted the attention of many researchers in statistical inference.

Vasicek [3] showed that his entropy-based normality test achieves higher powers, as compared with (then) best normality tests, namely, Anderson and Darling [23] and Shapiro and Wilk [24] tests.

Many researchers have been interested in testing normality including Vasicek [3], Arizono and Ohta [6], Esteban *et al.* [12], Choi *et al.* [14], Goria *et al.* [15], Farrell and Rogers-Stewart [25], Yazici and Yolacan [26], Meintanis [27] Romão *et al.* [28], Cardoso de Oliveira and Ferreira [29], Paul and Zhang [30] and Alizadeh Noughabi and Arghami [31].

In this section, we first introduce a goodness-of-fit test based on the proposed entropy estimators of Section 2 for testing normality and then we compare the powers of the introduced tests with the tests based on the entropy estimators proposed by Dimitriev and Tarasenko [2] , Vasicek [3] , Van Es [20], Ebrahimi *et al.* [4] and Correa [5].

3.1. *Introduction of the test statistics*

A well-known theorem of information theory [1, p. 55] states that among all continuous distributions that possess a density function f, and have a given variance σ^2, the entropy $H(f)$ is maximized by the normal distribution. Based on this property, following Vasicek [3], we introduce the following statistics for testing normality:

$$TZ1_{mn} = \frac{\exp\{HZ1_{mn}\}}{\hat{\sigma}},$$
$$TZ2_{mn} = \frac{\exp\{HZ2_{mn}\}}{\hat{\sigma}},$$

where

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Small values of $TZ1_{mn}$ and $TZ2_{mn}$ can be regarded as a symptom of non-normality and therefore we reject the hypothesis of normality for small enough values of $TZ1_{mn}$ and $TZ2_{mn}$.

It is worth mentioning that it is obvious that the above test statistics are scale and location invariant.

Unfortunately, the test statistics $TZ1_{mn}$ and $TZ2_{mn}$ are too complicated to allow deriving their exact distributions under normal hypothesis and therefore we obtain their critical values by means of Monte Carlo simulation. Tables 4 and 5 give respectively, the exact critical values (non-asymptotic) of the test statistics $TZ1_{mn}$ and $TZ2_{mn}$, for various sample sizes by Monte Carlo simulations with 10,000 repetitions.

3.2. *Power comparisons*

Ebrahimi *et al*. [4] showed the following relationship between their estimator and the estimator proposed by Vasicek [3]

$$HE_{mn} = HV_{mn} + \frac{2}{n}\left\{m\log(2m) + \log\frac{(m-1)!}{(2m-1)!}\right\}.$$

Thus for a fixed sample size n and fixed m, the test based on Vasicek's [3] and Ebrahimi *et al*.'s [4] entropy estimators have the same power. Therefore, we only compare the tests based on $TZ1_{mn}$

Table 4. Critical values of the $TZ1_{mn}$ statistic for $\alpha = 0.05$.

	m									
n	1	2	3	4	5	6	7	8	9	10
5	3.176	3.290								
6	3.195	3.302	3.392							
7	3.250	3.320	3.389							
8	3.290	3.354	3.425	3.482						
9	3.320	3.374	3.444	3.498						
10	3.348	3.403	3.472	3.516	3.559					
15	3.486	3.523	3.583	3.620	3.638	3.652	3.664			
20	3.577	3.604	3.648	3.690	3.712	3.717	3.717	3.719	3.723	3.728
25	3.641	3.660	3.700	3.739	3.772	3.789	3.788	3.787	3.782	3.775
30	3.703	3.724	3.765	3.792	3.821	3.833	3.845	3.845	3.839	3.834
40	3.768	3.777	3.805	3.835	3.854	3.879	3.892	3.896	3.895	3.893
50	3.810	3.817	3.837	3.867	3.892	3.912	3.928	3.935	3.940	3.947

Table 5. Critical values of the $TZ2_{mn}$ statistic for $\alpha = 0.05$.

	m									
n	1	2	3	4	5	6	7	8	9	10
5	3.143	3.312								
6	3.144	3.288	3.415							
7	3.185	3.292	3.402							
8	3.210	3.301	3.417	3.503						
9	3.233	3.309	3.420	3.510						
10	3.254	3.321	3.427	3.512	3.576					
15	3.388	3.406	3.474	3.546	3.601	3.642	3.672			
20	3.479	3.480	3.520	3.576	3.627	3.666	3.689	3.707	3.721	3.736
25	3.548	3.538	3.561	3.610	3.649	3.691	3.723	3.737	3.750	3.756
30	3.615	3.596	3.617	3.642	3.680	3.717	3.738	3.761	3.772	3.783
40	3.682	3.660	3.675	3.687	3.716	3.740	3.759	3.783	3.797	3.805
50	3.732	3.707	3.712	3.721	3.738	3.760	3.781	3.799	3.817	3.828

and $TZ2_{mn}$ with the tests based on the following test statistics

$$TD = \frac{\exp\{HD\}}{\hat{\sigma}},$$

$$TV_{mn} = \frac{\exp\{HV_{mn}\}}{\hat{\sigma}},$$

$$TEs_{mn} = \frac{\exp\{HEs_{mn}\}}{\hat{\sigma}},$$

$$TC_{mn} = \frac{\exp\{HC_{mn}\}}{\hat{\sigma}},$$

where $\hat{\sigma} = \sqrt{(1/n)\sum_{i=1}^{n}(X_i - \bar{X})^2}$.

It is obvious that all of the above test statistics are location and scale invariant, and we reject the hypothesis of normality for small enough values of TD, TV_{mn}, TEs_{mn} and TC_{mn}.

For power comparisons, we compute the powers of the tests based on statistics $TD, TV_{mn}, TEs_{mn}, TC_{mn}, TZ1_{mn}$ and $TZ2_{mn}$ by means of Monte Carlo simulations under 20 alternatives. The alternatives can be classified into four groups, according to their supports and the shapes of their densities. From the point of view of applied statistics, natural alternatives to normal distribution are in groups I and II. For the sake of completeness, we also consider groups III and IV. This gives additional insight in understanding the behaviour of the new test statistics $TZ1_{mn}$ and $TZ2_{mn}$. Esteban *et al*. [12], in their study of power comparisons of several tests for normality, suggest classifying the alternatives into the following four groups:

Group I: Support $= (-\infty, \infty)$, symmetric.

- Student's t with 1 degree of freedom (i.e. the standard Cauchy);
- Student's t with 3 degrees of freedom;
- Double exponential with parameters $\mu = 0$ (location), $\sigma = 1$ (scale);
- Logistic with parameters $\mu = 0$ (location), $\sigma = 1$ (scale).

Group II: Support $= (-\infty, \infty)$, asymmetric.

- Gumbel with parameters $\alpha = 0$ (location) and $\beta = 1$ (scale);
- Skew normal (SN) with parameters $\mu = 0$ (location), $\sigma = 1$ (scale) and $\alpha = 2$ (shape);

- Skew double exponential (SDE) with parameters $\alpha = 1, \beta = 2$ and $\mu = 0$ (location) (mixture exponential distribution with mean$\beta = 2$, and the negative of an exponential distribution with mean $\alpha = 1$).

Group III: Support $= (0, \infty)$.

- Exponential with mean 1;
- Gamma with parameters $\beta = 1$ (scale) and $\alpha = 2$ (shape);
- Gamma with parameters $\beta = 1$ (scale) and $\alpha = 1/2$ (shape);
- Lognormal (LN) with parameters $\mu = 0$ (scale) and $\sigma = 1$ (shape);
- Lognormal (LN) with parameters $\mu = 0$ (scale) and $\sigma = 2$ (shape);
- Lognormal (LN) with parameters $\mu = 0$ (scale) and $\sigma = \frac{1}{2}$ (shape);
- Weibull with parameters $\beta = 1$ (scale) and $\alpha = \frac{1}{2}$ (shape);
- Weibull with parameters $\beta = 1$ (scale) and $\alpha = 2$ (shape).

Group IV: Support $= (0, 1)$.

- Uniform;
- Beta (2, 2);
- Beta (0.5, 0.5);
- Beta (3, 1.5);
- Beta (2, 1).

Unfortunately, the powers of the proposed tests depend on the window size and the alternative distribution and therefore it is not possible to determine the best value of m, for which the tests attain maximum power under all alternatives. Therefore we used the values of m for which the aforementioned tests attain good (not best) powers for all alternative distributions. These values of m are tabulated in Table 6.

Tables 7– 10 contain the results of 10,000 simulation (of samples size 10, 20 and 40) per case to obtain the power of the proposed tests and those of the competing tests, at a significance level $\alpha = 0.05$. The bold type in these tables indicates the statistic achieving the maximal power.

Table 7 shows that the test based on $TZ2_{mn}$ consistently has the greatest power among entropy-based tests for all alternative distributions in group I and all considered sample sizes.

Table 8 indicates that although the tests based on $TZ2_{mn}$and TD are quite competitive but the test TD is slightly better.

It is evident from Table 9 that the tests based on TV_{mn}, TD and $TZ1_{mn}$ can be the best test for different sample sizes. For small sample size ($n = 10$), the test based on TV_{mn} is the best, for moderate sample size ($n = 20$), the test based onTD has the greatest power and for large sample size ($n = 40$), the test based on $TZ1_{mn}$ performs better than the others.

It is worth mentioning that the differences among the power of tests based on TV_{mn}, TD and $TZ1_{mn}$ in group III are not considerable.

Table 6. Proposed values of m for testing normality for different values of n.

n	m
$n \leq 8$	1
$9 \leq n \leq 15$	2
$16 \leq n \leq 35$	3
$36 \leq n \leq 60$	4
$61 \leq n \leq 80$	5
$81 \leq n \leq 100$	6

Table 7. Power comparisons for normality tests based on $T_D, T_V, T_{Es}, T_C, T_{Z1}$ and T_{Z2} statistics for sample sizes $n = 10, 20$ and 40 under the alternatives in group I with $\alpha = 0.05$.

Sample size	$n = 10$						$n = 20$						$n = 40$					
Alt	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}
$t_{(1)}$	0.583	0.442	0.591	0.409	0.632	**0.638**	0.872	0.737	0.871	0.687	0.885	**0.900**	0.991	0.960	0.987	0.949	0.988	**0.993**
$t_{(3)}$	0.201	0.091	0.167	0.083	0.212	**0.216**	0.371	0.165	0.330	0.138	0.377	**0.402**	0.612	0.289	0.541	0.249	0.561	**0.622**
DE	0.163	0.065	0.140	0.057	0.177	**0.181**	0.304	0.091	0.271	0.070	0.309	**0.344**	0.533	0.197	0.451	0.158	0.476	**0.568**
Logistic	0.087	0.051	0.074	0.047	0.089	**0.091**	0.134	0.051	0.114	0.043	0.133	**0.147**	0.210	0.053	0.160	0.048	0.166	**0.211**
Average	0.258	0.216	0.243	0.149	0.277	**0.281**	0.420	0.261	0.396	0.234	0.426	**0.448**	0.586	0.374	0.534	0.351	0.547	**0.598**

Table 8. Power comparisons for normality tests based on $T_D, T_V, T_{Es}, T_C, T_{Z1}$ and T_{Z2}statistics for sample sizes $n = 10, 20$ and 40 under the alternatives in group II with $\alpha = 0.05$.

Sample size	$n = 10$						$n = 20$						$n = 40$					
Alt	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}
Gumbel(0,1)	**0.154**	0.101	0.113	0.097	0.145	0.144	**0.310**	0.198	0.195	0.185	0.294	0.282	**0.530**	0.399	0.355	0.394	0.528	0.492
SN(0,1,2)	**0.071**	0.058	0.062	0.053	0.068	0.066	**0.102**	0.073	0.073	0.076	0.099	0.096	**0.149**	0.099	0.097	0.102	0.144	0.140
SDE(0,1,2)	0.216	0.117	0.178	0.111	0.220	**0.223**	0.423	0.225	0.353	0.201	0.424	**0.436**	**0.693**	0.420	0.586	0.385	0.659	**0.693**
Average	**0.147**	0.092	0.117	0.087	0.144	**0.147**	**0.278**	0.165	0.207	0.154	0.272	0.271	**0.457**	0.306	0.346	0.293	0.443	0.441

Table 9. Power comparisons for normality tests based on $T_D, T_V, T_{Es}, T_C, T_{Z1}$ and T_{Z2} statistics for sample sizes $n = 10, 20$ and 40 under the alternatives in group III with $\alpha = 0.05$.

Sample size	$n = 10$						$n = 20$						$n = 40$					
Alt	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}
Exp(1)	0.394	**0.416**	0.330	0.404	0.359	0.353	0.790	**0.846**	0.646	0.836	0.745	0.688	0.980	**0.996**	0.943	0.995	0.984	0.960
Gamma(2)	**0.222**	0.179	0.158	0.173	0.209	0.205	**0.507**	0.457	0.322	0.443	0.459	0.416	0.801	**0.832**	0.613	0.825	0.825	0.751
Gamma(1/2)	0.631	0.782	0.679	**0.786**	0.581	0.572	0.959	**0.992**	0.955	0.991	0.947	0.915	0.999	**1**	0.999	**1**	999	0.999
LN(0,1)	**0.565**	0.552	0.485	0.542	0.524	0.516	0.909	**0.927**	0.825	0.919	0.895	0.865	0.997	**0.998**	0.989	**0.998**	0.997	0.993
LN(0,2)	0.869	0.938	0.892	**0.936**	0.846	0.840	0.997	**0.999**	0.997	**0.999**	0.996	0.994	**1**	**1**	**1**	**1**	**1**	**1**
LN(0,1/2)	**0.249**	0.181	0.176	0.171	0.229	0.228	**0.517**	0.404	0.360	0.386	0.470	0.445	0.788	0.740	0.619	0.733	**0.801**	0.757
Weibull(1/2)	0.813	**0.931**	0.876	0.926	0.784	0.776	0.995	**1**	0.997	0.999	0.994	0.987	**1**	**1**	**1**	**1**	**1**	**1**
Weibull(2)	**0.076**	0.075	0.064	0.071	0.074	0.073	**0.148**	0.132	0.089	0.133	0.123	0.110	0.251	**0.267**	0.126	0.276	**0.267**	0.199
Average	0.477	**0.503**	0.457	0.501	0.450	0.445	**0.727**	0.719	0.648	0.713	0.703	0.677	0.852	0.854	0.786	0.853	**0.859**	0.832

Table 10. Power comparisons for normality tests based on $T_D, T_V, T_{Es}, T_C, T_{Z1}$ and T_{Z2} statistics for sample sizes $n = 10, 20$ and 40 under the alternatives in group IV with $\alpha = 0.05$.

Sample size	$n = 10$						$n = 20$						$n = 40$					
Alt	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}	T_D	T_V	T_{Es}	T_C	T_{Z1}	T_{Z2}
Uniform	0.028	0.167	0.061	**0.170**	0.030	0.026	0.084	**0.442**	0.076	0.438	0.099	0.028	0.343	0.846	0.160	**0.854**	0.480	0.126
Beta(2,2)	0.025	0.082	0.037	**0.086**	0.025	0.023	0.028	0.131	0.027	**0.135**	0.028	0.013	0.070	0.257	0.028	**0.288**	0.095	0.015
Beta(1/2,1/2)	0.080	**0.512**	0.238	0.489	0.078	0.060	0.408	**0.914**	0.460	0.902	0.442	0.145	0.912	**0.999**	0.882	**0.999**	0.973	0.726
Beta(3,1/2)	0.065	0.108	0.064	**0.110**	0.061	0.058	0.129	0.224	0.069	**0.225**	0.114	0.079	0.262	0.492	0.118	**0.512**	0.316	0.158
Beta(2,1)	0.093	0.173	0.092	**0.182**	0.081	0.076	0.221	**0.438**	0.131	0.432	0.200	0.130	0.472	0.825	0.289	**0.832**	0.574	0.315
Average	0.058	**0.208**	0.098	0.207	0.055	0.048	0.174	**0.429**	0.152	0.426	0.176	0.079	0.411	0.683	0.295	**0.696**	0.487	0.268

In group IV, the results are not in favour of the proposed tests, TV_{mn} and TC_{mn} give the best powers and the differences between their powers are negligible.

4. Conclusion

In this paper, we showed by simulation, that our two entropy estimators compare favourably with their competitors in terms of RMSE.

We also considered normality tests based on the introduced entropy estimators and compared them with the leading competitors.

Based on these comparisons, the following recommendations can be formulated for the application of the studied tests for testing normality in practice.

1. Use $TZ2_{mn}$ against the alternative distributions that are symmetric and have support on $(-\infty, \infty)$.
2. Use one of the $TZ2_{mn}$ or TD against the alternative distributions that have support on $(-\infty, \infty)$.
3. Use one of the $TZ1_{mn}, TD$ or TV_{mn} against the alterative distributions that have support on $(0, \infty)$.
4. Use TV_{mn} or TC_{mn} against the alterative distributions that have support on (0,1).

Acknowledgment

The authors are very thankful to the referee for his/her valuable comments which substantially improved the earlier version of this paper. Partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is acknowledged.

References

[1] C.E. Shannon, *A mathematical theory of communications*, Bell Syst. Tech. J. 27 (1948), pp. 379–423; 623–656.
[2] Y.G. Dimitriev and F.P. Tarasenko, *On the estimation functions of the probability density and its derivatives,* Theory Probab. Appl. 18 (1973), pp. 628–633.
[3] O. Vasicek, *A test for normality based on sample entropy*, Journal of the Royal Statistical Society, Ser. B, 38 (1976), pp. 54–59.
[4] N. Ebrahimi, K. Pflughoeft, and E. Soofi, *Two measures of sample entropy*, Statist. Probab. Lett. 20 (1994), pp. 225–234.
[5] J.C. Correa, *A new estimator of entropy,* Commun. Statist. Theory Meth. 24 (1995), pp. 2439–2449.
[6] I. Arizono and H. Ohta, *A test for normality based on Kullback-Leibler information*, Am. Statist. 43 (1989), pp. 20–22.
[7] E.J. Dudewicz and E.C. Van der Meulen, *Entropy-based tests of uniformity*, J. Amer. Statist. Assoc. 76 (1981), pp. 967–974.
[8] G.S. Mudholkar and C.T. Lin, *On two applications of characterization theorems of goodness-of-fit,* Colloq. Math. Soc. Janos Bolyai 45 (1984), pp. 395–414.
[9] N. Ebrahimi, M. Habibullah, and E. Soofi, *Testing exponentiality based on Kullback-Leibler information,* J. Roy. Statist. Soc. 54 (1992), pp. 739–748.
[10] A.W. Bowman, *Density based tests for goodness-of-fit*, J. Statist. Comput. Simul. 40 (1992), pp. 1–13.
[11] U. Lund and S.R. Jammalamadaka, *An entropy-base test for goodness of fit of the von Mises distribution,* J. Statist. Comput. Simul. 67 (2000), pp. 319–333.
[12] M.D. Esteban, M.E. Castellanos, D. Morales, and I. Vajda, *Monte Carlo comparison of four normality tests using different entropy estimates,* Commun. Statist. Simul. Comput. 30(4) (2001), pp. 761–285.
[13] S. Park and D. Park, *Correcting moments for goodness of fit tests based on two entropy estimates,* J. Statist. Comput. Simul. 73(9) (2003), pp. 685–694.
[14] B. Choi, K. Kim, and S.H. Song, *Goodness of fit test for exponentiality based on Kullback-Leibler information,* Commun. Statist. Simul. Comput. 33(2) (2004), pp. 525–536.
[15] M.N. Goria, N.N. Leonenko, V.V. Mergel, and P.L. Novi Inverardi, *A new class of random vector entropy estimators and its applications in testing statistical hypotheses*, Nonparametric Statist. 17(3) (2005), pp. 277–297.
[16] F. Yousefzadeh and N.R. Arghami, *Testing exponentiality based on type II censored data and a new cdf estimator*, Commun. Statist. Simul. Comput. 37 (2008), pp. 1479–1499.

[17] M. Mahdizadeh and N.R. Arghami, *Efficiency of ranked set sampling in entropy estimation and goodness-of-fit testing for the inverse Gaussian law*, J. Statist. Comput. Simul. 80(7) (2010), pp. 761–774.
[18] H. Alizadeh Noughabi and N.R. Arghami, *Monte Carlo comparison of five exponentiality tests using different entropy estimates*, J. Statist. Comput. Simul. First published on 5 April 2011 (iFirst).
[19] E. Zamanzade and N.R. Arghami, *Goodness of fit test based on correcting moments of modified entropy estimator*, J. Statist. Comput. Simul. First published on 16 May 2011(iFirst).
[20] B. Van Es, *Estimating functionals related to a density by class of statistics based on spacings*, Scand. J. Statist. 19 (1992), pp. 61–72.
[21] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.
[22] P. Grzegorzewski and R. Wieczorkowski, *Entropy-based goodness-of-fit test for exponentiality,* Commun. Statist. Theory Methods 28 (1999), pp. 1183–1202.
[23] T.W. Anderson and D.A. Darling, *A test of goodness of fit*, J. Amer. Statist. Assoc. 49 (1954), pp. 765–769.
[24] S.S. Shapiro and M.B. Wilk, *Analysis of variance test for normality*, Biometrika 52 (1965), pp. 591–611.
[25] P.G. Farrell and S. Rogers, *Comprehensive study of tests for normality and symmetry: Extending the Spiegelhalter test,* J. Statist. Comput. Simul. 76(9) (2006), pp. 803–816.
[26] B. Yazici and S. Yolacan, *A comparison of various tests of normality*, J. Statist. Comput. Simul. 77(2) (2007), pp. 175–183.
[27] S.G. Meintanis, *Goodness-of-fit testing by transforming to normality: Comparison between classical and characteristic function-based methods,* J. Statist. Comput. Simul. 79(2) (2009), pp. 205–212.
[28] X. Romão, R. Delgado, and A. Costa, *An empirical power comparison of univariate goodness-of-fit tests for normality,* J. Statist. Comput. Simul. 80(5) (2010), pp. 545–591.
[29] I.R. Cardoso de Oliveira and D.F. Ferreira, *Multivariate extension of chi-squared univariate normality test*, J. Statist. Comput. Simul. 80(5) (2010), pp. 513–526.
[30] S.R. Paul and X. Zhang, *Testing for normality in linear regression models*, J. Statist. Comput. Simul. 80(10) (2010), pp. 1101–1113.
[31] H. Alizadeh Noughabi and N.R. Arghami, *Monte Carlo comparison of seven normality tests*, J. Statist. Comput. Simul. 81 (2011), pp. 965–972.