Journal of Frame and Matrix Theory (JFMT) 2 (2025) No. 1, 59–69
Online ISSN: 3041-8526
Published by Faculty of Mathematics and Computer Science
Hakim Sabzevari University
http://dx.doi.org/10.22034/jfmt.2025.492456.1020

Testing Uniformity Based on Cumulative Entropy of Comparison Distribution Function

Malihe Abbasnejad[a,*], M. Mahdizadeh[a]

[a] *Department of Mathematics and Computer Science, Hakim Sabzevari University, Sabzevar, Iran*

(Communicated by Ghadir Sadeghi)

Abstract

Testing for uniformity arises in some important problems in statistics. In this paper, we use cumulative entropy of comparison distribution to construct a uniformity test. The asymptotic null distribution of the test statistics is derived. The necessary critical values for small sample sizes are determined numerically. The proposed test is compared with its competitors using Monte Carlo experiments. It is observed that the introduced test has good power properties against some alternative distributions.

Keywords: Goodness-of-fit test, Information theory, Uniformity.
MSC 2020: Primary: 62G10; Secondary: 94A17.

*Corresponding author
Email addresses: m.abbasnejad@hsu.ac.ir (Malihe Abbasnejad), mahdizadeh.m@hsu.ac.ir; mahdizadeh.m@live.com (M. Mahdizadeh)

Received: 5 December 2024 *Accepted:* 3 April 2025

1 Introduction

The validity of parametric inference procedures highly depends on suitability of particular distributions assumed for the data. Testing for model adequacy is thus an important topic in statistics. Numerous procedures have been developed in the literature to determine if a random sample is drawn from a specific distribution. Such methods are broadly classified as goodness-of-fit (GOF) tests, which have been extensively studied by many authors. Testing for uniformity is of particular importance because GOF tests for other distributions may be reduced to that for uniformity. Let X be a nonnegative absolutely continuous random variable. The corresponding probability density function, distribution function, and survival function are denoted by $f(x)$, $F(x)$ and $\bar{F}(x) = 1 - F(x)$, respectively. [15] introduced cumulative residual entropy (CRE) as a measure of uncertainty based on the survival function. For the random variable X, it is defined as

$$\mathcal{E}(X) = -\int_0^\infty \bar{F}(x) \ln \bar{F}(x) dx.$$

This measure has found applications in reliability and image alignment. [8] proposed a similar information measure based on the distribution function, and called it cumulative entropy (CE). It is defined as

$$\mathcal{CE}(X) = -\int_0^\infty F(x) \ln F(x) dx.$$

Information-theoretic measures have been widely used for developing inference procedures. Some examples in the context of GOF tests include [1], [2], [3], [5], [6], [9], [11], [12], [13], [19], and [20].

Let X and Y be two non-negative random variables whose distribution functions are F and G, respectively. To compare F and G, [14] introduced the comparison distribution function as $D(u) = F(G^{-1}(u))$, for $0 < u < 1$. It can be shown that $D(u)$ is the distribution function of the standard uniform distribution if and only if $F = G$. Motivated by this property, we construct a GOF test based on the CE of comparison distribution function.

In Section 2, the test statistics is introduced and its asymptotic null distribution is shown to be normal. The necessary critical values for small sample sizes are determined numerically. In Section 3, a simulation study is conducted to compare this test with the existing tests. The results indicate that the new test has higher power for some alternative distributions.

2 The test Statistic

Let $X_1, \dots, X_n$ be a random sample from a population with absolutely distribution function $F(x)$, defined on $\mathbb{R}^+$. The problem of interest is to test the hypotheses

$$H_0 : F(x) = F_0(x) \text{ for all } x \in \mathbb{R}^+ \quad \text{vs.} \quad H_1 : F(x) \neq F_0(x) \text{ for some } x \in \mathbb{R}^+, \tag{2.1}$$

where $F_0(x)$ is a fully specified distribution function. Suppose that for $i = 1, \dots, n$, X_i is transformed as $Y_i = F_0(X_i)$. From the probability integral transformation theorem, it follows that Y_i has a standard uniform distribution. This is to say that the above hypothesis testing problem can be reduced to that of testing uniformity based on Y_i's. Thus, we consider the following special case in (2.1):

$$H_0 : F(x) = x \text{ for all } x \in (0,1) \quad \text{vs.} \quad H_1 : F(x) \neq x \text{ for some } x \in (0,1), \tag{2.2}$$

To develop a decision rule, we consider the comparison distribution function as $D(u) = F_0(F^{-1}(u))$, for $0 < u < 1$, where $F_0(x) = x$. The CE of $D(u)$ is then given by

$$\begin{aligned} \mathcal{CE}(D) &= -\int_0^1 D(u) \ln D(u) du \\ &= -\int_0^1 F_0(x) \ln F_0(x) dF(x) \\ &= -\int_0^1 x \ln x dF(x) \\ &= -E_F(X \ln X). \end{aligned} \tag{2.3}$$

If $F(x) = x$, then it is easily seen that $\mathcal{CE}(D) = 1/4$, i.e. the CE value of the standard uniform distribution. This motivated us to consider small or large values of $\mathcal{CE}(D) - 1/4$ as an evidence against uniformity.

In practice, $F(x)$ is unknown and $\mathcal{CE}(D)$ in (2.3) cannot be evaluated. An estimate of this quantity is needed to construct a test statistic. This can be simply obtained using

$$U_n = -\frac{1}{n} \sum_{i=1}^{n} X_i \ln X_i.$$

An application of the central limit theorem shows that under the null hypothesis in (2.2) we have

$$Z = \frac{\sqrt{n}\,(U_n - 1/4)}{\sqrt{5/432}} \xrightarrow{d} N(0,1),$$

where $\xrightarrow{d}$ denotes convergence in distribution. Thus, the critical region of an asymptotic test of level α for (2.2) is given by $|Z| > z_{\alpha/2}$, where $z_{\alpha/2}$ is $100(1-\alpha/2)$ percentile of the standard normal distribution. A simulation study was carried out to demonstrate the asymptotic null distribution of Z. To do so, 10,000 samples of sizes $n = 50, 100$ were generated from standard uniform distribution. Then the test statistic was computed from each sample. Figure 1 shows the histogram of Z values overlaid with the standard normal density curve. It is observed that the curve fits the histogram very closely.

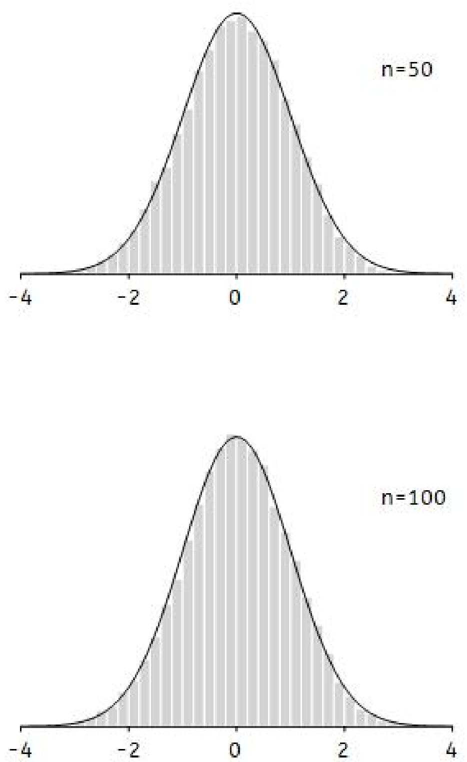

Figure 1: Histogram of the null distribution of Z for $n = 50, 100$ overlaid with the standard normal density curve.

It is difficult to derive the null distribution of Z for small sample sizes analytically. We thus employed a simulation study to approximate $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of this distribution, denoted by $z^*_{1-\alpha/2}$ and $z^*_{\alpha/2}$, respectively. The estimated values based on 10,000 replications are reported in Table 1 for some choices of n and α. The null hypothesis is rejected at the significance level α if $Z < z^*_{1-\alpha/2}$ or $Z > z^*_{\alpha/2}$.

Table 1: The estimated critical values of Z for some choices of n and α.

n	α	$z^*_{1-\alpha/2}$	$z^*_{\alpha/2}$	n	α	$z^*_{1-\alpha/2}$	$z^*_{\alpha/2}$
5	0.1	-1.7242	1.5762	15	0.1	-1.6915	1.6073
	0.05	-2.0922	1.787		0.05	-1.9856	1.9061
	0.01	-2.7455	2.1121		0.01	-2.6359	2.394
6	0.1	-1.7061	1.592	20	0.1	-1.7201	1.5814
	0.05	-2.0171	1.8244		0.05	-2.0189	1.8785
	0.01	-2.7566	2.1659		0.01	-2.64	2.3786
7	0.1	-1.6998	1.5845	25	0.1	-1.6891	1.5911
	0.05	-2.0144	1.8311		0.05	-2.0656	1.8834
	0.01	-2.691	2.267		0.01	-2.7981	2.4242
8	0.1	-1.7127	1.574	30	0.1	-1.6952	1.5772
	0.05	-2.0506	1.836		0.05	-2.0086	1.8706
	0.01	-2.7069	2.2176		0.01	-2.588	2.3782
9	0.1	-1.7055	1.5949	40	0.1	-1.6641	1.6133
	0.05	-2.0829	1.8516		0.05	-1.9762	1.9083
	0.01	-2.6974	2.2718		0.01	-2.5924	2.4593
10	0.1	-1.7146	1.5602	50	0.1	-1.687	1.6165
	0.05	-2.0607	1.8203		0.05	-2.0025	1.8964
	0.01	-2.6708	2.3044		0.01	-2.6432	2.4855

3 Power Comparison

A simulation study was performed to evaluate performance of the proposed test with the existing uniformity tests. In doing so, we considered the some well-known GOF tests: Kolmogrov-Smirnov [10], Cramer-Von Mises [7, 21], Anderson-Darling [4], Smith and Bain [17], and entropy based test [9]. Let $X_{(1)} \leq \cdots \leq X_{(n)}$ be order statistics of the random sample $X_1, \ldots, X_n$. The test statistics for the above-mentioned tests are as follows:

- Kolmogrov-Smirnov Statistic

$$KS = \max\left\{\max_{1\leq i\leq n}\left[\frac{i}{n} - X_{(i)}\right], \max_{1\leq i\leq n}\left[X_{(i)} - \frac{i-1}{n}\right]\right\}.$$

- Cramer-von Mises statistic

$$CvM = \frac{1}{12n} + \sum_{i=1}^{n}\left(\frac{2i-1}{n} - X_{(i)}\right)^2.$$

- Anderson-Darling statistic

$$AD = -n - \frac{2}{n}\sum_{i=1}^{n}\frac{2i-1}{n}\left[\left(i-\frac{1}{2}\right)\ln X_{(i)} + \left(n-i+\frac{1}{2}\right)\ln(1-X_{(i)})\right].$$

- Smith and Bain statistic

$$R = 1 - \hat{\rho}^2,$$

where $\hat{\rho}$ is the correlation coefficient for pairs $(X_{(i)}, k_i)$, with $k_i = i/(n+1)$ $(i = 1, \ldots, n)$.

- Entropy based statistic

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

where $m \le n/2$ is a positive integer called window size, $X_{(i)} = X_{(1)}$ if $i < 1$, and $X_{(i)} = X_{(n)}$ if $i > n$.

In power comparisons, we considered the following alternatives distributions which have been frequently used in the literature [9]:

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

In particular, $A_{1.5}$, A_2, A_3 $B_{1.5}$, B_2, B_3, $C_{1.5}$, C_2 and C_3 were included in our study. As pointed out by [18], alternative A has points closer to zero than expected under the uniformity hypothesis, while alternative B has points near 0.5, and alternative C has two clusters near 0 and 1.

Under each alternative distribution, 10,000 samples of sizes $n = 10, 20, 50, 100$ were generated, and used to estimate power of different tests for the significance levels $\alpha = 0.05, 0.01$. In using the entropy based statistic, $m = \lfloor n/4 \rfloor$ was used. It should be mentioned that for $n = 10, 20$, power of our test was estimated based on the critical values in Table 1. Also, for $n = 50, 100$, its power was estimated based on the asymptotic null distribution of Z.

The estimated powers are reported in Tables 2 and 3. It is observed that there is not a single test outperforming the others against all alternatives. The CvM test has the highest power for alternative A. Also, the entropy based test has the highest power for alternative B. In this case, Z test is a good competitor of the entropy based test, while the other tests generally have low powers. Finally, the proposed test has the best performance for alternative C. It is worth noting that this test has the advantage of known asymptotic null distribution.

Table 2: Estimated powers of the tests for $\alpha = 0.01, 0.05$ and $n = 10, 20$.

n	α	Alt.	KS	CvM	AD	R	$K_{m,n}$	Z
10	0.01	$A_{1.5}$	0.0477	**0.0611**	0.0516	0.0150	0.0343	0.0168
		A_2	0.1644	**0.2119**	0.1784	0.0278	0.0974	0.0257
		A_3	0.5207	**0.6234**	0.5749	0.0627	0.3370	0.0281
		$B_{1.5}$	0.0046	0.0024	0.0010	0.0094	**0.0423**	0.0376
		B_2	0.0047	0.0014	0.0002	0.0106	**0.1274**	0.1087
		B_3	0.0047	0.0003	0.0000	0.0208	**0.4237**	0.3376
		$C_{1.5}$	0.0331	0.0289	0.0344	0.0182	0.0127	**0.0354**
		C_2	0.0372	0.0524	0.0731	0.0341	0.0269	**0.1217**
		C_3	0.1555	0.1068	0.2087	0.1024	0.0800	**0.3770**
	0.05	$A_{1.5}$	0.1588	**0.1825**	0.1758	0.0651	0.1189	0.0742
		A_2	0.3947	**0.4490**	0.4268	0.0999	0.2741	0.1020
		A_3	0.8057	**0.8702**	0.8414	0.1673	0.6419	0.0908
		$B_{1.5}$	0.0370	0.0281	0.0155	0.0497	**0.1618**	0.1372
		B_2	0.0503	0.0255	0.0103	0.0591	**0.3479**	0.3124
		B_3	0.0974	0.0637	0.0255	0.0848	**0.7411**	0.6752
		$C_{1.5}$	0.1144	0.1016	0.1025	0.0719	0.0661	**0.1152**
		C_2	0.1959	0.1560	0.2374	0.1200	0.1162	**0.2844**
		C_3	0.3670	0.2981	0.4942	0.2501	0.2728	**0.6577**
20	0.01	$A_{1.5}$	0.1129	**0.1416**	0.1334	0.0288	0.0782	0.0296
		A_2	0.4266	**0.5330**	0.5117	0.0790	0.3281	0.0541
		A_3	0.9298	**0.9703**	0.9621	0.2295	0.8760	0.0436
		$B_{1.5}$	0.0087	0.0042	0.0018	0.0126	**0.1395**	0.0753
		B_2	0.0203	0.0068	0.0042	0.0182	**0.4562**	0.3136
		B_3	0.0793	0.0577	0.0450	0.0618	**0.9343**	0.7954
		$C_{1.5}$	0.0408	0.0359	0.0519	0.0215	0.0035	**0.0845**
		C_2	0.1231	0.0788	0.1405	0.0721	0.0991	**0.3056**
		C_3	0.3359	0.2406	0.4950	0.2734	0.4348	**0.7876**
	0.05	$A_{1.5}$	0.2841	**0.3232**	0.3196	0.0985	0.2419	0.1434
		A_2	0.7051	**0.7715**	0.7617	0.2048	0.6170	0.2121
		A_3	0.9887	**0.9968**	0.9950	0.4105	0.9758	0.1320
		$B_{1.5}$	0.0615	0.0420	0.0328	0.0607	**0.3499**	0.2740
		B_2	0.1300	0.1077	0.1064	0.0873	**0.7356**	0.6638
		B_3	0.4205	0.5185	0.5630	0.1887	**0.9908**	0.9707
		$C_{1.5}$	0.1474	0.1177	0.1569	0.0911	0.1124	**0.2312**
		C_2	0.3098	0.2484	0.3784	0.2284	0.2671	**0.5565**
		C_3	0.6395	0.6487	0.8241	0.6607	0.7127	**0.9230**

Table 3: Estimated powers of the tests for $\alpha = 0.01, 0.05$ and $n = 50, 100$.

n	α	Alt.	KS	CvM	AD	R	$K_{m,n}$	Z
50	0.01	$A_{1.5}$	0.3543	**0.4425**	0.4336	0.0875	0.2932	0.1345
		A_2	0.9196	**0.9629**	0.9618	0.3488	0.8937	0.2430
		A_3	**1**	**1**	**1**	0.7268	0.9999	0.0887
		$B_{1.5}$	0.0248	0.0126	0.0154	0.0186	**0.3950**	0.3769
		B_2	0.1436	0.1372	0.2222	0.0867	**0.9447**	0.9132
		B_3	0.7747	0.9255	0.9736	0.3755	**1**	**1**
		$C_{1.5}$	0.0877	0.0479	0.0786	0.0598	0.0084	**0.2180**
		C_2	0.3344	0.2604	0.4416	0.4224	0.0691	**0.7408**
		C_3	0.8731	0.9430	0.9884	0.9802	0.6908	**0.9982**
	0.05	$A_{1.5}$	0.6078	**0.6858**	0.6852	0.2197	0.5770	0.3259
		A_2	0.9837	0.9933	**0.9945**	0.5760	0.9811	0.4818
		A_3	**1**	**1**	**1**	0.8753	**1**	0.2858
		$B_{1.5}$	0.1278	0.1130	0.1449	0.0781	**0.7044**	0.6444
		B_2	0.4835	0.6108	0.7383	0.2545	**0.9927**	0.9842
		B_3	0.9796	0.9989	**1**	0.6294	**1**	**1**
		$C_{1.5}$	0.2580	0.2012	0.2720	0.2209	0.0607	**0.4794**
		C_2	0.6567	0.6904	0.7975	0.7731	0.2961	**0.9158**
		C_3	0.9927	0.9992	0.9993	0.9994	0.9434	**0.9997**
100	0.01	$A_{1.5}$	0.7445	0.8228	**0.8419**	0.2479	0.6844	0.3091
		A_2	0.9995	**1**	**1**	0.7766	0.9995	0.5339
		A_3	**1**	**1**	**1**	0.9890	**1**	0.2871
		$B_{1.5}$	0.0966	0.0645	0.1484	0.0482	**0.8223**	0.7404
		B_2	0.6375	0.7737	0.9385	0.3281	**0.9997**	0.9993
		B_3	0.9996	**1**	**1**	0.8668	**1**	**1**
		$C_{1.5}$	0.2107	0.1235	0.2150	0.2357	0.0177	**0.5294**
		C_2	0.7623	0.8264	0.9248	0.9492	0.2604	**0.9796**
		C_3	0.9999	**1**	**1**	**1**	0.9966	**1**
	0.05	$A_{1.5}$	0.9109	0.9452	**0.9532**	0.4719	0.8910	0.5630
		A_2	**1**	**1**	**1**	0.9162	0.9999	0.7733
		A_3	**1**	**1**	**1**	0.9976	**1**	0.5184
		$B_{1.5}$	0.3493	0.3823	0.5613	0.1832	**0.9573**	0.9049
		B_2	0.9357	0.9865	0.9985	0.6335	**1**	**1**
		B_3	**1**	**1**	**1**	0.9663	**1**	**1**
		$C_{1.5}$	0.4827	0.4447	0.5635	0.5656	0.0854	**0.7535**
		C_2	0.9625	0.9878	0.9944	0.9953	0.6166	**0.9958**
		C_3	**1**	**1**	**1**	**1**	0.9998	**1**

To visually assess the consistency of the proposed test, its power against different alternatives was plotted as a function of n. The result is depicted in Figure 2 when $\alpha = 0.05$. It can be seen that for each alternative, the power tends to 1 as n increases from 10 to 100, supporting the consistency property.

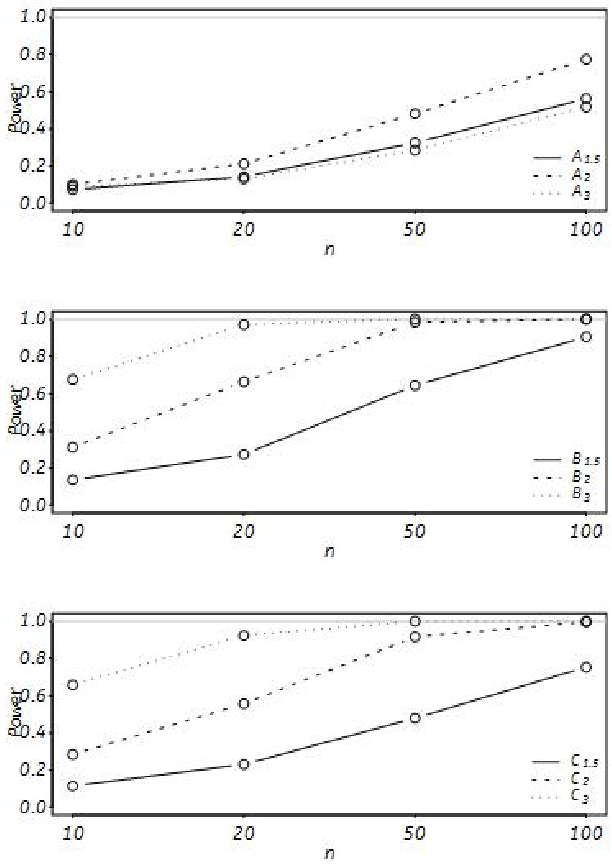

Figure 2: Estimated powers of the proposed test for $\alpha = 0.05$ as a function of n.

4 Conclusion

Many statistical methods require some parametric assumptions for the data at hand. It is essential to ensure the validity of such assumptions before using the statistical methods. This is formally done using GOF tests developed for many classical probability distributions. In particular, testing for uniformity has received considerable attention in the literature because GOF test for other distributions may be done via that for uniformity.

The CE is a measure of uncertainty that has found many applications in statistical inference. In this article, a uniformity test based on the CE of comparison distribution function is developed. The power of the proposed test is compared with its competitors in a simulation study. The results indicate that the new test has the best performance against the alternatives from C family. Motivated by the good power behavior of this test, we plan to study GOF test for the Rayleigh, generalized exponential, and Pareto distributions based on the CE of comparison distribution function.

References

[1] M. Abbasnejad, *Modified estimators of Renyi entropy with application in testing exponentiality based on transformed data*, İstatistik, **9** (2016) 42–55. zbl MR pdf

[2] M. Abbasnejad, *Testing exponentiality based on Renyi entropy of transformed data*, J. Statist. Res. Iran, **8** (2011) 149–162. doi

[3] M. Abbasnejad, N. R. Arghami, and M. Tavakoli, *A goodness of fit test for exponentiality based on Lin-Wong information*, J. Iran. Stat. Soc., **11** (2012) 191–202. zbl MR pdf

[4] T. W. Anderson, D. A. Darling, *A test of goodness-of-fit*, J. Amer. Statist. Assoc., **49** (1954) 765–769. doi

[5] I. Arizono, H. Ohta, *A test for normality based on Kullback-Leibler information*, Amer. Statist., **43** (1989) 20–22. doi

[6] S. Baratpour, A. Habibi Rad, *Exponentiality test based on the progressive type II censoring via cumulative entropy*, Commun. Stat. Simulation Comput., **45** (2016) 2625–2637. zbl MR doi

[7] H. Cramer, *On the composition of elementary errors: II, statistical applications*, Scand. Actuar. J., 1 (1928) 13–74. zbl doi

[8] A. Di Crescenzo, M. Longobardi, *On cumulative entropies*, J. Statist. Plann. Inference, **139** (2009) 4072–4087. zbl MR doi

[9] E. J. Dudewicz, E. C. Van derMeulen, *Entropy-based tests of uniformity*, J. Amer. Statist. Assoc., **76** (1981) 967–974. zbl MR doi

[10] A. N. Kolmogorov, *Sulla determinazione empirica di una legge di distibuziane*, Giornale dellIstituto Italiano degli Attuari, **4** (1933) 83–91. pdf

[11] M. Mahdizadeh, *New indices of sample informational energy with application to testing*

uniformity, Afr. Mat., **31** (2020) 273–295. zbl MR doi

[12] M. Mahdizadeh, *On testing uniformity using an information-theoretic measure*, Commun. Stat., Simulation Comput., **46** (2017) 6173–6196. zbl MR doi

[13] M. Mahdizadeh, N. R. Arghami, *Efficiency of ranked set sampling in entropy estimation and goodness-of-fit testing for the inverse Gaussian law*, J. Stat. Comput. Simulation, **80** (2010) 761–774. zbl MR doi

[14] E. Parzen, *Statistical methods mining, two sample data analysis, comparison distributions, and quantile limit theorems*, Szyszkowicz B, Asymptotic methods in probability and statistics, An international conference at Carleton Univ., Ottawa, Ontario, Canada, Amsterdam: North-Holland/ Elsevier, (1998) 611–617. zbl

[15] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang, *Cumulative residual entropy: a new measure of information*, IEEE Trans. Inform. Theory, **50** (2004) 1220–1228. zbl MR doi

[16] N. V. Smirnov, *Estimate of derivation between empirical distribution functions in two independent samples*, Bull. Mosc. Univ., **2** (1939) 3–16. link

[17] R. M. Smith, L. J. Bain, *Correlation type goodness-of-fit statistics with censored sampling*, Comm. Statist. Theory Methods A5, **5** (1976) 119–132. zbl MR doi

[18] M. A. Stephens, *EDF statistics for goodness of fit and some comparisons*, J. Amer. Statist. Assoc., **69** (1974) 730–737. doi

[19] M. Tavakoli, N. R. Arghami, and M. Abbasnejad, *A goodness of fit test for normality based on Balakrishnan- Sanghvi information*, J. Iran. Stat. Soc., **18** (2019) 177–190. zbl MR doi

[20] O. Vasicek, *A test for normality based on sample entropy*, J. R. Stat. Soc., Ser. B, **38** (1976) 54–59. zbl MR

[21] R. Von Mises, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und Theoretischen Physik*, Leipzig: Deuticke, 1931. link