

A New Estimator of Entropy

Hadi Alizadeh Noughabi, Naser Reza Arghami

Department of Statistics, Ferdowsi University of Mashhad, Iran.
(alizadehhadi@gmail.com)

Abstract. In this paper we propose an estimator of the entropy of a continuous random variable. The estimator is obtained by modifying the estimator proposed by Vasicek (1976). Consistency of estimator is proved, and comparisons are made with Vasicek's estimator (1976), van Es's estimator (1992), Ebrahimi et al.'s estimator (1994) and Correa's estimator (1995). The results indicate that the proposed estimator has smaller mean squared error than above estimators.

1 Introduction

Entropy is a useful measure of uncertainty and dispersion, and has been widely employed in many pattern analysis applications. The entropy of a distribution function F with a probability density function f is defined by Shannon (1948) as:

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

There is an extensive literature on estimating the Shannon entropy nonparametrically. Vasicek (1976), van Es (1992), Correa (1995),

Key words and phrases: Entropy estimator, exponential, information theory, normal, uniform.

Ebrahimi et al. (1994) and Alizadeh Noughabi (2010) have proposed estimates for the entropy of absolutely continuous random variables. Among the various entropy estimators discussed in the literature, Vasicek's estimator has gained prominence in developing entropy-based statistical procedures due to its simplicity. To motivate the estimator, express $H(f)$ in the form of

$$H(f) = \int_0^1 \log \left\{ \frac{d}{dp} F^{-1}(p) \right\} dp, \quad (2)$$

by using the fact that the slope $\frac{d}{dp} F^{-1}(p)$ is simply the reciprocal of the density function at the p th population quantile, i.e.,

$$\frac{d}{dp} F^{-1}(p) = \frac{1}{f(F^{-1}(p))}.$$

So an intuitive idea of estimating the slope would be to estimate F by the empirical distribution function F_n and replace the differential operator by a difference operator. This motivation yields a very simple estimator of the slope which is $\frac{n}{2m}$ times the difference between two sample quantiles whose indexes are $2m$ apart, one on the upper side of the p th sample quantile and the other on the lower side. The entropy estimator is then given by

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

Here, the window size m is a positive integer smaller than $n/2$, $X_{(i)} = X_{(1)}$ if $i < 1$, $X_{(i)} = X_{(n)}$ if $i > n$ and $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ are order statistics based on a random sample of size n . Vasicek proved that $HV_{mn} \rightarrow H(f)$ as $n \rightarrow \infty$, $m \rightarrow \infty$, $\frac{m}{n} \rightarrow 0$.

Van Es (1992) proposed another estimator of entropy based on spacings and proved the consistency and asymptotic normality of this estimator under some conditions. Van Es' estimator is given by

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

Ebrahimi et al. (1994), adjusted the weights of Vasicek's estimator, in order to take into account the fact that the differences are truncated around the smallest and the largest data points. (i.e. $X_{(i+m)} - X_{(i-m)}$ is replaced by $X_{(i+m)} - X_{(1)}$ when $i \leq m$ and

$X_{(i+m)} - X_{(i-m)}$ is replaced by $X_{(n)} - X_{(i-m)}$ when $i \geq n - m + 1$). Their estimator is given by

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

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 proved that $HE_{mn} \rightarrow H(f)$ as $n \rightarrow \infty, m \rightarrow \infty, \frac{m}{n} \rightarrow 0$. They compared their estimator with Vasicek's estimator and Dudewicz and Van der Meulen estimator (1987), and by simulation, showed that their estimator has smaller bias and mean squared error. Also they mentioned that their estimator is better, in terms of bias and MSE, than Mack's estimator, kernel entropy estimator and Theil (1980)'s estimator.

Correa (1995) proposed a modification of Vasicek estimator which produces a smaller MSE; considering the sample information represented as

$$(F_n(X_{(1)}), X_{(1)}), (F_n(X_{(2)}), X_{(2)}), \dots, (F_n(X_{(n)}), X_{(n)}),$$

rewriting Eq. (2) as

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

and noting that the argument of log is the equation of the slope of the straight line that joins the points $(F_n(X_{(i+m)}), X_{(i+m)})$ and $(F_n(X_{(i-m)}), X_{(i-m)})$, Correa (1995) used a local linear model based on $2m + 1$ points to estimate the density of $F(x)$ in the interval $(X_{(i+m)}, X_{(i-m)})$,

$$F(x_{(j)}) = \alpha + \beta x_{(j)} + \varepsilon \quad j = m - i, \dots, m + i.$$

Instead of taking only two points to estimate the slope β , as Vasicek does, he uses all the sample points between $X_{(j-m)}$ and $X_{(j+m)}$, via least square method. The consequent estimator of entropy proposed by Correa (1995) is given by

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

where

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

He compared his estimator with Vasicek's estimator and van Es's estimator. The mean square error (MSE) of his estimator is consistently smaller than the MSE of Vasicek's estimator. Also for some m his estimator behaves better than van Es's estimator. No comparison has been made with Ebrahimi et al.'s estimator. Correa's estimator can be generalized to the two-dimensional case.

Many researchers have used the estimators of entropy for developing entropy-based statistical procedure. See for example Esteban et al. (2001), Park (2003), Choi et al. (2004), Gorla et al. (2005), Choi (2008) and Alizadeh Noughabi and Arghami (2010).

It is clear that

$$s_i(m, n) = \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \quad (4)$$

is not a correct formula for the slope when $i \leq m$ or $i \geq n - m + 1$. In order to correctly estimate the slopes at these points the denominator and/or the numerator should be modified for $i \leq m$ or $i \geq n - m + 1$. Our goal in this paper is, therefore, to remedy this situation, in a way different from that of Ebrahimi et al.

In section 2, we introduce an estimator of entropy and show that it is consistent. Scale invariance of variance and mean squared error of the proposed estimator is established. In section 3 we report results of a comparison of our estimator with the competing estimators by a simulation study.

2 The New Estimator

We see in Ebrahimi et al. (1994) that, for the small sample sizes ($n = 10, 20, 30$), almost everywhere, their estimator underestimates the entropy, in almost all cases. This includes the cases of uniform and normal distributions. Therefore we modify the coefficients of Ebrahimi et al. estimator as

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

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

$$HA_{mn} = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{a_i m} (X_{(i+m)} - X_{(i-m)}) \right\}, \quad (5)$$

where

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

and $X_{(i-m)} = X_{(1)}$ for $i \leq m$ and $X_{(i+m)} = X_{(n)}$ for $i \geq n - m$. Comparing (5) and (3) we obtain

$$\begin{aligned} HA_{mn} &= \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{a_i m} (X_{(i+m)} - X_{(i-m)}) \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{2n}{2a_i m} (X_{(i+m)} - X_{(i-m)}) \right\} \\ &= HV_{mn} + \frac{1}{n} \sum_{i=1}^n \log \frac{2}{a_i} = HV_{mn} + \frac{1}{n} \left(\sum_{i=1}^m \log 2 + \sum_{i=n-m+1}^n \log 2 \right) \\ &= HV_{mn} + \frac{2}{n} (m \log(2)). \end{aligned} \quad (6)$$

Also from Ebrahimi et al. (1994) we have

$$HE_{mn} = HV_{mn} + \frac{2}{n} \left\{ m \log(2m) + \log \frac{(m-1)!}{(2m-1)!} \right\}. \quad (7)$$

Therefore we obtain from (6) and (7)

$$HA_{1n} = HE_{1n}.$$

Remark. Theil (1980) computed the entropy $H(f_n^{ME})$ of an empirical maximum entropy density f_n^{ME} , which is related to HV_{1n} , HE_{1n} and HA_{1n} , as follows.

$$\begin{aligned} H(f_n^{ME}) &= HV_{1n} + \frac{2 - 2 \log 2}{n} \\ &= HA_{1n} - \frac{2}{n} \log 2 + \frac{2 - 2 \log 2}{n} \\ &= HA_{1n} + \frac{2 - 4 \log 2}{n} = HE_{1n} + \frac{2 - 4 \log 2}{n}. \end{aligned}$$

Theorem 2.1. Let X_1, \dots, X_n be a random sample from distribution $F(x)$. Then

- i) $HA_{mn} > HV_{mn}$
- ii) $HA_{mn} \geq HE_{mn}$

Proof. i). From (6) we have

$$HA_{mn} = HV_{mn} + \frac{2}{n} \{m \log 2\}.$$

Since $\frac{2}{n} \{m \log 2\} > 0$ then (i) holds.

ii). From (6) and (7) it is enough to show

$$\frac{2}{n} \{m \log(2)\} \geq \frac{2}{n} \left\{ m \log(2m) + \log \frac{(m-1)!}{(2m-1)!} \right\},$$

or equivalently

$$m^m \leq m(m+1)\dots(2m-1).$$

But the above inequality is true for all $m \geq 1$. For $m = 1$ inequality convert to equality and for $m > 1$ we have strict inequality.

The next theorem states that the scale of the random variable X has no effect on the accuracy of HA_{mn} in estimating $H(f)$. Similar results have been obtained for HV_{mn} and HE_{mn} by Mack (1988) and Ebrahimi (1994), respectively.

Theorem 2.2. *Let X_1, \dots, X_n be a sequence of i.i.d. random variables with entropy $H^X(f)$ and let $Y_i = kX_i, i = 1, \dots, n$, where $k > 0$. Let HA_{mn}^X and HA_{mn}^Y be entropy estimators for $H^X(f)$ and $H^Y(g)$ respectively. (here g is pdf of $Y = kX$). Then the following properties hold.*

- i) $E(HA_{mn}^Y) = E(HA_{mn}^X) + \log k,$
- ii) $Var(HA_{mn}^Y) = Var(HA_{mn}^X),$
- iii) $MSE(HA_{mn}^Y) = MSE(HA_{mn}^X).$

Proof. Since

$$HV_{mn}^{kX} = HV_{mn}^X + \log(k),$$

then from (6) we have

$$\begin{aligned} E(HA_{mn}^{kX}) &= E(HV_{mn}^{kX}) + \frac{2}{n} (m \log(2)) \\ &= E(HV_{mn}^X) + \log(k) + \frac{2}{n} (m \log(2)) \\ &= E(HA_{mn}^X) + \log(k). \end{aligned}$$

Also

$$\text{Var}(HA_{mn}^{kX}) = \text{Var}(HV_{mn}^{kX}) = \text{Var}(HV_{mn}^X) = \text{Var}(HA_{mn}^X),$$

and

$$\begin{aligned} \text{MSE}(HA_{mn}^{kX}) &= \text{Var}(HA_{mn}^{kX}) + \left\{ E(HA_{mn}^{kX}) - H^{kX}(f) \right\}^2 \\ &= \text{Var}(HA_{mn}^X) + \left\{ E(HA_{mn}^X) + \log(k) - H^X(f) - \log(k) \right\}^2 \\ &= \text{Var}(HA_{mn}^X) + \left\{ E(HA_{mn}^X) - H^X(f) \right\}^2 = \text{MSE}(HA_{mn}^X). \end{aligned}$$

Therefore the proof of this theorem is complete.

The following theorem establishes the consistency of HA_{mn} .

Theorem 2.3. *Let C be the class of continuous densities with finite entropies and let X_1, \dots, X_n be a random sample from $f \in C$. If $n \rightarrow \infty$, $m \rightarrow \infty$ and $\frac{m}{n} \rightarrow 0$, then*

$$HA_{mn} \longrightarrow H(f).$$

Proof. It is obvious by (6) and consistency of HV_{mn} .

3 Simulation study

A simulation study was performed to analyze the behavior of the proposed estimator of entropy, HA_{mn} . Some comparisons among Vasicek’s estimator, van Es’s estimator, Correa’s estimator, Ebrahimi et al.’s estimator and our estimator were done. For each sample size 10000 samples were generated and the RMSEs of the estimators were computed. We considered normal, exponential and uniform distributions which are the same three distributions considered in Correa (1995).

Still an open problem in entropy estimation is the optimal choice of m for given n . We choose to use the following heuristic formula (see Grzegorzewski and Wieczorkowski (1999)):

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

In addition, we considered the other values of m .

Table 1: proposed values of m for different values of n

Sample size n	Windows size m
$n \leq 7$	2
$8 \leq n \leq 14$	3
$15 \leq n \leq 24$	4
$25 \leq n \leq 35$	5
$36 \leq n \leq 45$	6
$46 \leq n \leq 60$	7
$61 \leq n \leq 80$	8

In order to compute our estimator for a given data set, one needs to specify the order of spacings m . Since n is always known, it is obvious that m may be taken as a function of n .

In practice, of course, a general guide for the choice of m for a fixed n would be valuable to the users. However, simulations show that the optimal m (in terms of RMSE) also depends on the distribution that one may have in mind. This is shown in Tables 2-4. From these tables we see that there is no m that is optimal for all distribution. Thus if one has a particular distribution in mind, one can choose the optimal m from Tables 2-4. Otherwise, that is if one wants to guard against all distributions a compromise should be made.

Generally, with increasing n , an optimal choice of m also increases, while the ratio m/n tends to zero.

We suggest the values of m which the proposed estimator obtains reasonably good (not best) RMSE. These values of m are tabulated in Table 1.

Tables 2-4 contain the RMSE values (and standard deviation) of the five estimators at different sample size for each of the three considered distributions. In the last four columns of each Table we have shown the quantity

$$R_i = \frac{H_i - HA_{mn}}{H_i} \times 100, \quad i = 1, 2, 3, 4$$

which shows the RMSE-performance of the HA_{mn} with respect to the others four, where $H_1 = HV_{mn}$, $H_2 = HVE_{mn}$, $H_3 = HC_{mn}$ and $H_4 = HE_{mn}$.

Table 2: Root of mean square error (and standard deviation) of estimators in estimate of entropy $H(f)$ for standard normal distribution.

n	m	HV_{mn}	HVE_{mn}	HC_{mn}	HE_{mn}	HA_{mn}	R_1	R_2	R_3	R_4
5	2	0.987 (0.414)	0.506 (0.451)	0.782 (0.409)	0.652 (0.414)	0.537 (0.414)	45.59	-6.13	31.33	17.64
	3	1.064 (0.401)	0.490 (0.424)	0.823 (0.398)	0.648 (0.401)	0.496 (0.401)	53.38	-1.22	39.73	23.46
10	3	0.619 (0.266)	0.366 (0.282)	0.465 (0.268)	0.403 (0.266)	0.302 (0.266)	51.21	17.49	35.05	25.06
	4	0.666 (0.259)	0.376 (0.279)	0.498 (0.259)	0.394 (0.259)	0.266 (0.259)	60.06	29.26	46.59	32.49
20	5	0.719 (0.255)	0.380 (0.276)	0.537 (0.253)	0.389 (0.255)	0.256 (0.255)	64.39	32.63	52.33	34.19
	4	0.374 (0.175)	0.276 (0.182)	0.263 (0.177)	0.245 (0.175)	0.183 (0.175)	51.07	33.70	30.42	25.31
	5	0.396 (0.181)	0.295 (0.186)	0.282 (0.184)	0.243 (0.181)	0.181 (0.181)	54.29	38.64	35.82	25.51
30	6	0.418 (0.178)	0.306 (0.184)	0.297 (0.182)	0.239 (0.178)	0.182 (0.178)	56.46	40.52	38.72	23.85
	5	0.281 (0.142)	0.243 (0.146)	0.192 (0.144)	0.184 (0.142)	0.142 (0.142)	49.47	41.56	26.04	22.83
	6	0.292 (0.144)	0.258 (0.147)	0.202 (0.147)	0.180 (0.144)	0.146 (0.144)	50.00	43.41	27.72	18.89
50	7	0.305 (0.144)	0.269 (0.145)	0.209 (0.147)	0.177 (0.144)	0.154 (0.144)	49.51	42.75	26.32	12.99
	8	0.199 (0.110)	0.212 (0.110)	0.134 (0.112)	0.128 (0.110)	0.114 (0.110)	42.71	46.23	14.93	10.94
	9	0.204 (0.110)	0.223 (0.111)	0.137 (0.112)	0.125 (0.110)	0.121 (0.110)	40.69	45.74	11.68	3.20
		0.209 (0.111)	0.233 (0.110)	0.140 (0.114)	0.123 (0.111)	0.133 (0.111)	36.36	42.92	5.00	-8.13

Table 3: Root of mean square error (and standard deviation) of estimators in estimate of entropy $H(f)$ for exponential distribution with mean one.

n	m	HV_{mn}	HVE_{mn}	HC_{mn}	HE_{mn}	HA_{mn}	R_1	R_2	R_3	R_4
5	2	0.930 (0.557)	0.590 (0.581)	0.750 (0.561)	0.659 (0.557)	0.589 (0.557)	36.67	0.17	21.47	10.62
	3	0.964 (0.567)	0.574 (0.570)	0.758 (0.582)	0.644 (0.567)	0.574 (0.567)	40.46	0.00	24.27	10.87
10	3	0.564 (0.358)	0.389 (0.372)	0.435 (0.362)	0.400 (0.358)	0.358 (0.358)	36.52	7.97	17.70	10.50
	4	0.581 (0.368)	0.397 (0.381)	0.443 (0.375)	0.392 (0.368)	0.383 (0.368)	34.08	3.53	13.54	2.30
20	5	0.595 (0.372)	0.388 (0.377)	0.446 (0.382)	0.381 (0.372)	0.436 (0.372)	26.72	-12.37	2.24	-14.44
	4	0.352 (0.241)	0.272 (0.249)	0.269 (0.244)	0.260 (0.241)	0.242 (0.241)	31.25	11.03	10.04	6.92
	5	0.364 (0.247)	0.282 (0.253)	0.277 (0.251)	0.259 (0.247)	0.259 (0.247)	28.85	8.16	6.50	0.00
30	6	0.364 (0.247)	0.279 (0.251)	0.274 (0.252)	0.251 (0.247)	0.288 (0.247)	20.88	-3.23	-5.11	-14.74
	5	0.270 (0.193)	0.224 (0.197)	0.206 (0.194)	0.203 (0.193)	0.197 (0.193)	27.04	12.05	4.37	2.96
	6	0.271 (0.195)	0.227 (0.199)	0.208 (0.198)	0.200 (0.195)	0.215 (0.195)	20.66	5.29	-3.37	-7.50
50	7	0.275 (0.199)	0.232 (0.200)	0.212 (0.203)	0.200 (0.199)	0.239 (0.199)	13.09	-3.02	-12.74	-19.50
	7	0.198 (0.150)	0.181 (0.151)	0.156 (0.152)	0.153 (0.150)	0.164 (0.150)	17.17	99.91	-5.13	-7.19
	8	0.198 (0.150)	0.186 (0.152)	0.155 (0.151)	0.150 (0.150)	0.176 (0.150)	11.11	5.38	-13.55	-17.33
	9	0.197 (0.151)	0.187 (0.151)	0.156 (0.154)	0.151 (0.151)	0.195 (0.151)	1.02	-4.28	-25.00	-29.14

Table 4: Root of mean square error (and standard deviation) of estimators in estimate of entropy $H(f)$ for uniform distribution on $(0,1)$.

n	m	HV_{mn}	HVE_{mn}	HC_{mn}	HE_{mn}	HA_{mn}	R_1	R_2	R_3	R_4
5	2	0.778 (0.346)	0.405 (0.405)	0.575 (0.343)	0.461 (0.346)	0.374 (0.346)	51.93	7.65	34.96	18.87
	3	0.839 (0.332)	0.379 (0.379)	0.598 (0.338)	0.444 (0.332)	0.341 (0.332)	59.36	10.03	42.98	23.20
10	3	0.453 (0.166)	0.218 (0.218)	0.294 (0.166)	0.235 (0.166)	0.166 (0.166)	63.36	23.85	43.54	29.36
	4	0.485 (0.160)	0.221 (0.221)	0.312 (0.162)	0.213 (0.160)	0.186 (0.160)	61.65	15.84	40.38	12.68
	5	0.533 (0.166)	0.228 (0.228)	0.343 (0.171)	0.210 (0.166)	0.249 (0.166)	53.28	-9.21	27.41	-18.57
20	4	0.276 (0.087)	0.120 (0.120)	0.158 (0.089)	0.135 (0.087)	0.088 (0.087)	68.12	26.67	44.30	34.81
	5	0.291 (0.085)	0.126 (0.126)	0.168 (0.088)	0.123 (0.085)	0.109 (0.085)	62.54	13.49	35.12	11.38
	6	0.308 (0.081)	0.127 (0.127)	0.179 (0.084)	0.112 (0.081)	0.143 (0.081)	53.57	-12.60	20.11	-27.68
30	5	0.211 (0.060)	0.087 (0.087)	0.112 (0.061)	0.097 (0.060)	0.066 (0.060)	68.72	24.14	41.07	31.96
	6	0.221 (0.058)	0.088 (0.088)	0.120 (0.060)	0.089 (0.058)	0.086 (0.058)	61.09	2.27	28.33	3.37
	7	0.235 (0.199)	0.092 (0.200)	0.130 (0.203)	0.084 (0.199)	0.112 (0.199)	52.34	-21.74	13.85	-33.33
50	7	0.155 (0.037)	0.057 (0.057)	0.075 (0.038)	0.062 (0.037)	0.057 (0.037)	63.23	00.00	24.00	8.06
	8	0.163 (0.037)	0.060 (0.060)	0.082 (0.039)	0.059 (0.037)	0.073 (0.037)	55.21	-21.67	10.98	-23.73
	9	0.171 (0.036)	0.061 (0.061)	0.087 (0.038)	0.055 (0.036)	0.090 (0.036)	47.37	-47.54	-3.45	-63.64

We observe that the proposed estimator performs well as compared with other estimators under the normal distribution. Generally, for small sample sizes the new estimator behaves better than other estimators. Also we observe that the estimators HV_{mn} , HE_{mn} and HA_{mn} have the same standard deviation, because we have

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

$$HA_{mn} = HV_{mn} + \frac{2}{n} \{m \log 2\}.$$

Since the new estimator HA_{mn} and the estimators HV_{mn} , HE_{mn} have the same variances and by Theorem 2.1 the new estimator is larger than old ones, with probability 1, then it seem the new estimator reduces the bias.

Acknowledgements

The authors thank two anonymous referees for their helpful comments. Partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is acknowledged.

References

- Alizadeh Noughabi, H. (2010), A new estimator of entropy and its application in testing normality. *Journal of Statistical Computation and Simulation*, **80**(10), 1151-1162.
- Alizadeh Noughabi, H. and Arghami, N. R. (2010), Testing exponentiality based on characterizations of the exponential distribution. to appear in *Journal of Statistical Computation and Simulation*.
- Choi, B. (2008), Improvement of goodness of fit test for normal distribution based on entropy and power comparison. *Journal of Statistical Computation and Simulation*, **78**(9), 781-788.
- Choi, B., Kim, K., and Song, S. H. (2004), Goodness of fit test for exponentiality based on Kullback-Leibler information. *Communications in Statistics-Simulation and Computation*, **33**(2), 525-536.
- Correa, J. C. (1995), A new estimator of entropy. *Communications in Statistics-Theory and Methods*, **24**, 2439-2449.
- Dobrushin, R. L. (1958), Simplified method of experimental estimate of entropy of stationary sequence. *Theory Probability and its Applications*, **3**, 462-464.
- Dudewicz, E. S. and Van der Meulen E. C. (1987), The empiric entropy, a new approach to nonparametric entropy estimation. in: M. L. Puir, J. P. Vilaplana and W. Wertz, eds., *New Perspectives in Theoretical and Applied Statistics*, New York: Wiley, 207-227.
- Ebrahimi, N., Pflughoeft, K., and Soofi, E. (1994), Two measures of sample entropy. *Statistics and Probability Letters*, **20**, 225-234.
- Esteban, M. D., Castellanos, M. E., Morales, D. and Vajda I. (2001), Monte Carlo comparison of four normality tests using different entropy estimates. *Communications in Statistics-Simulation and computation*, **30**(4), 761-785.
- Goria, M. N., Leonenko, N. N., Mergel, V. V., and Novi Inverardi, P. L. (2005), A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *Nonparametric Statistics*, **17**(3), 277-297.

- Grzegorzewski, P. and Wieczorkowski, R. (1999), Entropy-based goodness-of-fit test for exponentiality. *Communications in Statistics-Theory and Methods*, **28**, 1183-1202.
- Hall, P. and Morton, S. C. (1993), On the estimation of entropy. *Annals Institute of Statistical and Mathematics*, **45**, 69-88.
- Hutcheson, K. and Shenton, L. R. (1974), Some moments of an estimate of Shannon's measure of information. *Communications in Statistics*, **3**, 89-94.
- Joe, H. (1989), Estimation of entropy and other functionals of a multivariate density. *Annals Institute of Statistical and Mathematics*, **41**, 683-697.
- Mack, S. P. (1988), A comparative study of entropy estimators and entropy based goodness-of-fit tests. Ph.D. Dissertation, University of California, Riverside.
- Park, S. and Park, D. (2003), Correcting moments for goodness of fit tests based on two entropy estimates. *Journal of Statistical Computation and Simulation*, **73**(9), 685-694.
- Shannon, C. E. (1948), A mathematical theory of communications. *Bell System Technical Journal*, **27**, 379-423; 623-656.
- Theil, J. (1980), The entropy of maximum entropy distribution, *Econom. Lett.*, **5**, 145-148.
- van Es, B. (1992), Estimating functionals related to a density by class of statistics based on spacings. *Scandinavian Journal of Statistics*, **19**, 61-72.
- Vasicek, O. (1976), A test for normality based on sample entropy. *Journal of the Royal Statistical Society, Ser. B*, **38**, 54-59.
- Vatutin, V. A. and Michailov, V. G. (1995), Statistical estimation of entropy of discrete random variables with large numbers of results. *Russian Mathematical Surveys*, **50**, 963-976.