STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, February 2025, pp 891–899.
Published online in International Academic Press (www.IAPress.org)

A New Estimator for Shannon Entropy

Hadi Alizadeh Noughabi [1,*] and Mohammad Shafaei Noughabi [2]

[1] *Department of Statistics, University of Birjand, Birjand, Iran*
[2] *Department of Mathematics and Statistics, University of Gonabad, Gonabad, Iran*

Abstract In this paper we propose a new estimator of the entropy of a continuous random variable. The estimator is obtained by modifying the estimator proposed by Vasicek (1976). Consistency of the proposed estimator is proved, and comparisons are made with Vasicek's estimator (1976), Ebrahimi et al.'s estimator (1994) and Correa's estimator (1995). The results indicate that the proposed estimator has smaller mean squared error than considered alternative estimators. The proposed estimator is applied to a real data set for illustration.

Keywords Information theory, Entropy estimator, Exponential, Normal, Uniform.

AMS 2010 subject classifications 62B10; 94A15

DOI: 10.19139/soic-2310-5070-1844

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:

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

There is an extensive literature on estimating the Shannon entropy nonparametrically. For example, Vasicek (1976), Ebrahimi et al. (1994) and Correa (1995) 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

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

by using the fact that the slope $\frac{d}{dp}F^{-1}(p)$ is simply the reciprocal of the density function at the pth 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

[*] Correspondence to: Hadi Alizadeh Noughabi (Email: alizadehhadi@birjand.ac.ir). Department of Statistics, University of Birjand, Birjand, Iran.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

slope which is $n/2m$ times the difference between two sample quantiles whose indexes are $2m$ apart, one on the upper side of the th sample quantile and the other on $p$the lower side. The entropy estimator is then given by

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

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 $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ are order statistics based on a random sample of size n. Vasicek proved that $HV_{mn} \to H(f)$ as $n \to \infty, m \to \infty, m/n \to 0$.

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 $X_{(i+m)} - X_{(1)}$ is replaced by $X_{(n)} - X_{(1)}$ 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 $HE_{mn} \to H(f)$ as $n \to \infty$, $m \to \infty$, $m/n \to 0$. They compared their estimator with Vasicek's estimator and Dudewicz and Van der Meulen (1987) estimator, 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's (1980) estimator.

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

$$(F_n(X_{(1)}), X_{(1)}), (F_n(X_{(2)}), X_{(2)}), ..., (F_n(X_{(n)}), X_{(n)}),$$

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 $(X_{(i+m)}, X_{(i-m)})$,

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

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. The mean square error (MSE) of his estimator is consistently smaller than the MSE of Vasicek'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), Goria et al. (2005), Choi (2008), Jarrahiferiz and Alizadeh (2017), and Alizadeh and Jarrahiferiz (2020).
It is clear that

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

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. In Section 4, we apply the proposed estimator to a real data example. Some conclusions are presented in Section 5.

2. The New Estimator

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

$$HA_{mn} = \frac{1}{n}\sum_{i=1}^{n} \log\left\{ \frac{n}{2ma_n}(X_{(i+m)} - X_{(i-m)}) \right\}, \tag{5}$$

where

$$a_n = 1 - \frac{1}{\sqrt{n}},$$

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

$$\begin{aligned} HA_{mn} &= \tfrac{1}{n}\sum_{i=1}^{n} \log\left\{ \tfrac{n}{2ma_n}(X_{(i+m)} - X_{(i-m)}) \right\} = \tfrac{1}{n}\sum_{i=1}^{n} \log\left\{ \tfrac{n}{2m}(X_{(i+m)} - X_{(i-m)}) \right\} - \tfrac{1}{n}\sum_{i=1}^{n} \log a_n \\ &= HV_{mn} - \tfrac{1}{n}\sum_{i=1}^{n} \log a_n = HV_{mn} - \log\left(1 - \tfrac{1}{\sqrt{n}}\right). \end{aligned} \tag{6}$$

Also, from Ebrahimi et al. (1994), we have

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

Therefore, we obtain from (6) and (7)

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

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.

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

Theorem 1. Let $X_1, \ldots, X_n$ be a random sample from distribution $F(x)$. Then

$$HA_{mn} \geq HV_{mn} \,.$$

Proof. From (6) we have

$$HA_{mn} = HV_{mn} - \log\left(1 - \frac{1}{\sqrt{n}}\right).$$

Since $\log\left(1 - \frac{1}{\sqrt{n}}\right) < 0$ the proof is complete.

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

$i\,)$ $E\left(HA^Y_{mn}\right) = E\left(HA^X_{mn}\right) + \log\, k$,

$ii\,)$ $Var\left(HA^Y_{mn}\right) = Var\left(HA^X_{mn}\right)$,

$iii\,)$ $MSE\left(HA^Y_{mn}\right) = MSE\left(HA^X_{mn}\right)$.

Proof. Since

$$HV^{kX}_{mn} = HV^X_{mn} + \log(k),$$

then from (6) we have

$$\begin{aligned} E(HA^{kX}_{mn}) &= E(HV^{kX}_{mn}) - \log\left(1 - \tfrac{1}{\sqrt{n}}\right) \\ &= E(HV^X_{mn}) + \log(k) - \log\left(1 - \tfrac{1}{\sqrt{n}}\right) \\ &= E(HA^X_{mn}) + \log(k)\,. \end{aligned}$$

Also

$$Var(HA^{kX}_{mn}) = Var(HV^{kX}_{mn}) = Var(HV^X_{mn}) = Var(HA^X_{mn})\,,$$

and

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

Therefore, the proof of this theorem is complete.

Theorem 3. Let C be the class of continuous densities with finite entropies and let $X_1, \ldots, X_n$ be a random sample from $f \in C$. If $n \to \infty$, $m \to \infty$ and $m/n \to 0$, then

$$HA_{mn} \to H(f),$$

in probability.

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, Correa's estimator, Ebrahimi et al.'s estimator and our estimator were done. For each sample size 100000 samples were generated and the bias and RMSEs of the estimators were

Table 1. Root of mean square error and absolute bias of estimators in estimate of entropy $H(f)$ for standard normal distribution.

	RMSE(AB)						
n	HV_{mn}	HC_{mn}	HE_{mn}	HA_{mn}	R_1	R_2	R_3
5	0.994 (0.902)	0.789 (0.671)	0.658 (0.509)	**0.521 (0.308)**	47.58	33.97	20.82
10	0.621 (0.560)	0.467 (0.382)	0.405 (0.305)	**0.322 (0.181)**	48.15	31.05	20.49
20	0.375 (0.329)	0.266 (0.195)	0.249 (0.172)	**0.194 (0.076)**	48.27	27.07	22.09
30	0.282 (0.243)	0.194 (0.128)	0.186 (0.118)	**0.149 (0.041)**	47.16	23.20	19.89
50	0.198 (0.165)	0.133 (0.074)	0.127 (0.065)	**0.110 (0.013)**	44.44	17.29	13.39

Table 2. Root of mean square error and absolute bias of estimators in estimate of entropy $H(f)$ for exponential distribution with mean one.

	RMSE(AB)						
n	HV_{mn}	HC_{mn}	HE_{mn}	HA_{mn}	R_1	R_2	R_3
5	0.931 (0.747)	0.744 (0.491)	0.660 (0.352)	**0.578 (0.154)**	37.92	22.31	12.42
10	0.564 (0.436)	0.434 (0.238)	0.399 (0.181)	**0.363 (0.058)**	35.64	16.36	9.02
20	0.353 (0.256)	0.269 (0.113)	0.264 (0.101)	**0.244 (0.005)**	30.88	9.29	7.58
30	0.273 (0.190)	0.208 (0.068)	0.206 (0.064)	**0.195 (0.013)**	28.57	6.25	5.34
50	0.197 (0.129)	0.155 (0.033)	0.151 (0.029)	**0.149 (0.024)**	24.37	3.87	1.32

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)):

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

Generally, with increasing n, an optimal choice of m also increases, while the ratio m/n tends to zero. Tables 1-3 contain the absolute bias (AB) and root of mean square error (RMSE) values of the four 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

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

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

Moreover, Figures 1-3 display root of mean square error (RMSE) values of the four estimators at different sample size for each of the three considered distributions.

We observe that the proposed estimator performs well as compared with other estimators. The proposed estimator

Table 3. Root of mean square error and absolute bias of estimators in estimate of entropy $H(f)$ for uniform distribution on (0,1).

	RMSE(AB)						
n	HV_{mn}	HC_{mn}	HE_{mn}	HA_{mn}	R_1	R_2	R_3
5	0.773 (0.693)	0.569 (0.457)	0.457 (0.303)	**0.357 (0.103)**	53.82	37.26	21.88
10	0.454 (0.422)	0.292 (0.241)	0.234 (0.166)	**0.170 (0.041)**	62.55	41.78	27.35
20	0.275 (0.261)	0.157 (0.130)	0.134 (0.102)	**0.086 (0.008)**	68.73	45.22	35.82
30	0.210 (0.201)	0.111 (0.093)	0.096 (0.076)	**0.059 (0.0004)**	71.90	46.85	38.54
50	0.155 (0.151)	0.076 (0.065)	0.063 (0.051)	**0.037 (0.0017)**	76.13	51.32	41.27

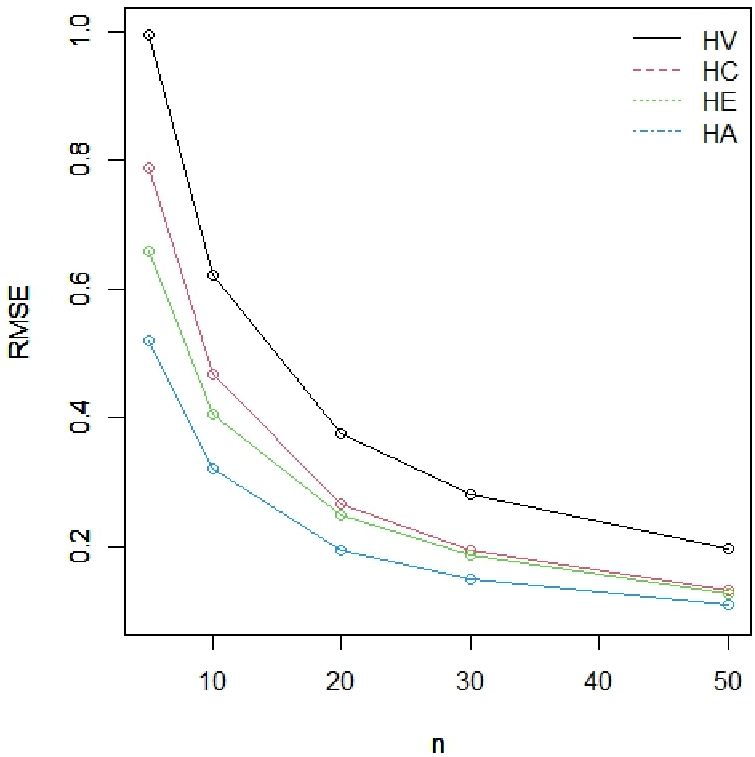

Figure 1. RMSE of the estimators in estimate of entropy for the normal distribution.

has smaller bias and mean squared error than others. For all sample sizes and under different distributions we can see that the new estimator behaves better than the other estimators. Therefore, the proposed estimator can be confidently recommended in practice.

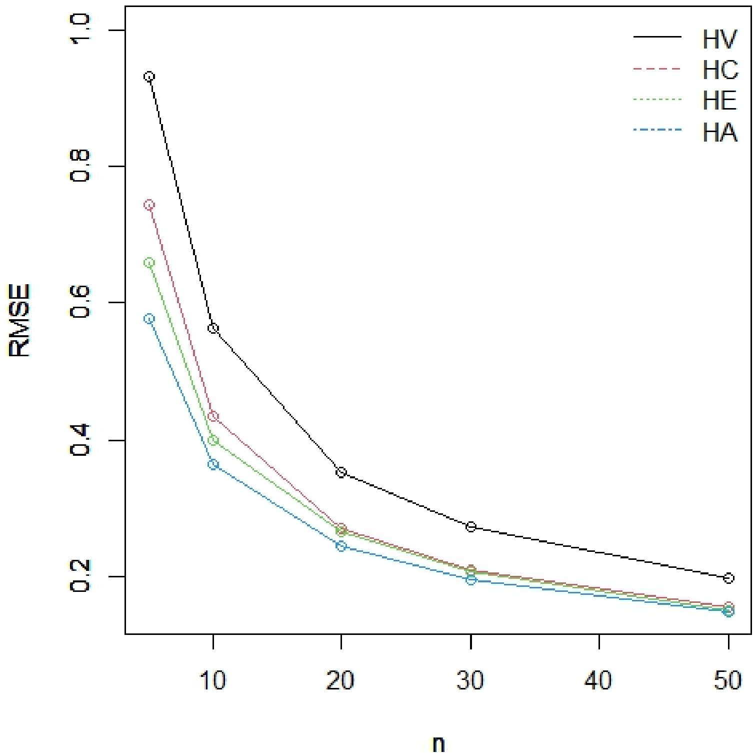

Figure 2. RMSE of the estimators in estimate of entropy for the exponential distribution.

Table 4. Entropy values and absolute bias of the estimators in estimate of entropy $H(f)$ of data.

	HV_{mn}	HC_{mn}	HE_{mn}	HA_{mn}
Entropy value	-0.13146	-0.06679	-0.04054	0.01337
Absolute bias	0.13146	0.06679	0.04054	0.01337

4. Applications to real data

In this section, the newly proposed estimator is applied to a real data set for illustration.
Example 1. We consider the data set discussed in Illowsky and Dean (2018) in Page 317, Table 5.1. The data set consist of smiling times of 55 babies measured in seconds. The data are as follows.
10.4, 19.6, 18.8, 13.9, 17.8, 16.8, 21.6, 17.9, 12.5, 11.1, 4.9, 12.8, 14.8 ,22.8 ,20.0, 15.9, 16.3, 13.4, 17.1 ,14.5, 19.0, 22.8, 1.3, 0.7, 8.9, 11.9 ,10.9 ,7.3 ,5.9, 3.7, 17.9, 19.2, 9.8, 5.8, 6.9, 2.6, 5.8, 21.7 ,11.8, 3.4 ,2.1 ,4.5, 6.3, 10.7, 8.9, 9.4, 9.4 ,7.6, 10.0, 3.3, 6.7 ,7.8, 11.6, 13.8, 18.6.
The data originally follows a uniform distribution U(0,23). We standardize the data to U(0,1). For this transformed data the values and the absolute bias of the considered estimators are obtained and presented in Table 4.
From Table 4, we observe that the proposed estimator HA_{mn} performs well as compared with other estimators. Therefore, in many practical applications, we expect that the proposed estimators are preferable to the competing estimators.

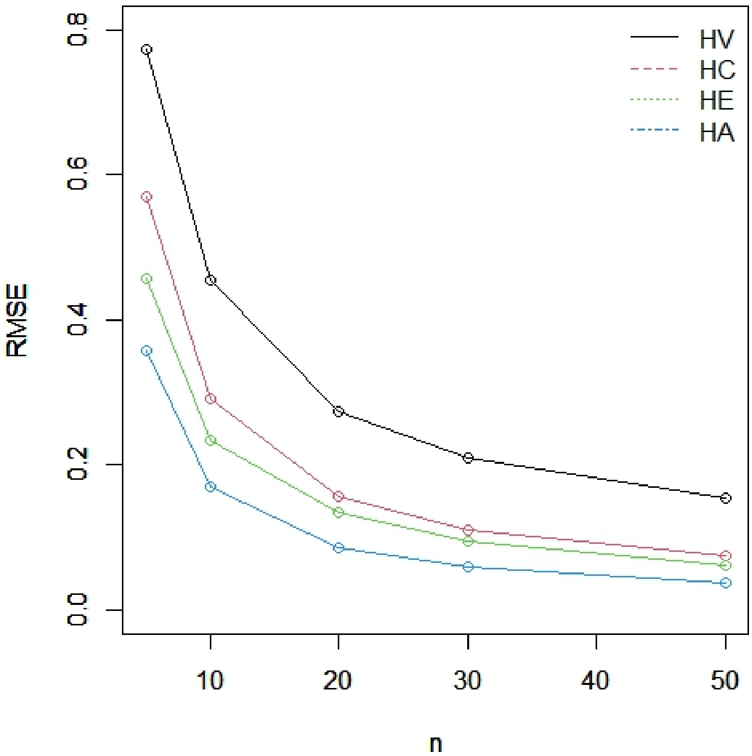

Figure 3. RMSE of the estimators in estimate of entropy for the uniform distribution

5. Conclusions

In this paper, we have first described some prominent methods for entropy estimation and then introduced a new entropy estimator of a continuous random variable. The proposed estimator has constructed based on modification of Vasicek entropy estimator. We have presented the properties of the proposed estimator. We finally have compared the proposed estimator with some prominent existing estimators. We have shown that for different sample sizes the new estimator behaves better than the competitors. Generally, the proposed estimator has a good performance and it can be easily applied in practice.

Acknowledgement

The authors are grateful to anonymous referees and the associate editor for providing some useful comments on an earlier version of this manuscript.

REFERENCES

1. Alizadeh, H.N. and Jarrahiferiz, J. (2020), Tests of fit for the Gumbel distribution: EDF-based tests against entropy-based tests, Journal of Applied Statistics, 47(10), 1885-1900.
2. 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.
3. 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.
4. Correa, J.C. (1995), A new estimator of entropy, Communications in Statistics - Theory and Methods, 24, 2439-2449.
5. Dobrushin, R.L. (1958), Simplified method of experimental estimate of entropy of stationary sequence, Theory Probability and its Applications, 3, 462-464.
6. 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 (Wiley, New York) pp. 207-227.
7. Ebrahimi, N., Pflughoeft, K. and Soofi, E. (1994), Two measures of sample entropy. Statistics and Probability Letters, 20, 225-234.
8. 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.
9. 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.
10. Grzegorzewski, P. and Wieczorkowski, R. (1999), Entropy-based goodness-of-fit test for exponentiality, Communications in Statistics - Theory and Methods, 28, 1183-1202.
11. Hall, P. and Morton, S.C. (1993), On the estimation of entropy, Annals Institute of Statistical and Mathematics, 45, 69-88.
12. Hutcheson, K. and Shenton, L.R. (1974), Some moments of an estimate of Shannon's measure of information, Communications in Statistics, 3, 89-94.
13. Illowsky, B. and Dean, S. (2018), Introductory Statistics, OpenStax, Houston.
14. Jarrahiferiz, J. and Alizadeh, H.N. (2017), Testing exponentiality using different entropy estimates based on type II censored data: a Monte Carlo power comparison, International Journal of Industrial Engineering, 24(5), 556-571.
15. Joe, H. (1989), Estimation of entropy and other functionals of a multivariate density, Annals Institute of Statistical and Mathematics, 41, 683-697.
16. Mack, S.P. (1988), A comparative study of entropy estimators and entropy-based goodness-of-fit tests. Ph.D. Dissertation, University of California, Riverside.
17. 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.
18. Shannon, C.E. (1948), A mathematical theory of communications, Bell System Technical Journal, 27, 379-423; 623-656.
19. Theil, J. (1980), The entropy of maximum entropy distribution, Economic Letters, 5, 145-148.
20. Vasicek, O. (1976), A test for normality based on sample entropy, Journal of the Royal Statistical Society, Ser. B, 38, 54-59.
21. 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.