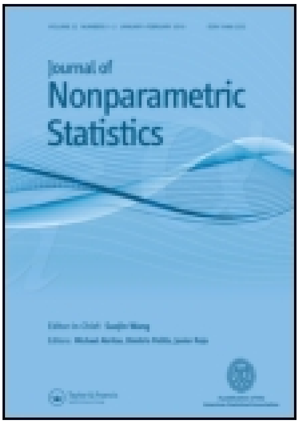

Journal of Nonparametric Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gnst20

A new class of random vector entropy estimators and its applications in testing statistical hypotheses

M. N. Goria [a] , N. N. Leonenko [b] , V. V. Mergel [c] & P. L. Novi Inverardi [a]

[a] Department of Computer and Management Sciences , University of Trento , Trento, 38100, Italy

[b] School of Mathematics , Cardiff University , Senghennydd Road, Cardiff, CF24 4YH, UK

[c] Department of Statistics , University of Florida , Gainesville, FL, 32611-8545, USA

Published online: 16 Aug 2006.

Nonparametric Statistics
Vol. 17, No. 3, April–May 2005, 277–297

A new class of random vector entropy estimators and its applications in testing statistical hypotheses

M. N. GORIA†, N. N. LEONENKO‡, V. V. MERGEL§ and P. L. NOVI INVERARDI*†

†Department of Computer and Management Sciences, University of Trento, Trento 38100, Italy
‡School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF24 4YH, UK
§Department of Statistics, University of Florida, Gainesville, FL 32611-8545, USA

(Received September 2003; in final form June 2004)

This paper proposes a new class of estimators of an unknown entropy of random vector. Its asymptotic unbiasedness and consistency are proved. Further, this class of estimators is used to build both goodness-of-fit and independence tests based on sample entropy. A simulation study indicates that the test involving the proposed entropy estimate has higher power than other well-known competitors under heavy tailed alternatives which are frequently used in many financial applications.

Keywords: Entropy; Multivariate density; Estimator; Goodness-of-fit test; Testing independence; Monte Carlo methods

1. Introduction

We consider the problem of estimation of the Shannon measure of information or (differential) entropy

$$H = -\int_{R^m} f(x) \ln f(x) \, \mathrm{d}x < \infty, \tag{1}$$

of random vector $\xi \in R^m$ with density function $f(x)$, $x \in R^m$, based on independent and identically distributed *(i.i.d.)* sample $X_1, X_2, \ldots, X_N$, $N \geq 2$.

The estimation of entropy of discrete random variables has been considered by Dobrushin [1], Basharin [2], Zubkov [3], Hutcheston and Shenton [4], Stein [5], El-Bassiouny [6], Vatutin and Michailov [7] and others.

In the absolutely continuous case, the estimation of entropy has been considered by Ahmad and Lin [8], Vasicek [9], Joe [10, 11], Hall and Morton [12], Van Es [13], and Tsybakov and van der Meulen [14]. The most elegant results have been obtained by Van Es [13], who proposed a class of statistics based on spacings of increasing order. Van Es also showed that these estimates are almost surely consistent and asymptotically normal under a set of assumptions on the density function. The entropy-based tests of goodness-of-fit exploit the

*Corresponding author. Email: pnoviinv@cs.unitn.it

Nonparametric Statistics
ISSN 1048-5252 print/ISSN 1029-0311 online
http://www.tandf.co.uk/journals
DOI: 10.1080/104852504200026815

so-called maximum entropy principle [15–18] are considered by Vasicek [9], Dudewicz and van der Meulen [19], and others.

The problem of estimation of entropy of random vector with density function from different view point has been considered by Ivanov and Rojkova [20], Joe [10, 11], Gangopadhyay *et al.* [21], and Eggermont and La Riccia [22]. They discussed a class of statistics based on the kernel-type estimators of density [23].

Our paper is motivated by the work of Kozachenko and Leonenko [24], where an estimate of entropy based on the idea of Dobrushin [1] is treated.

We present here the conditions for asymptotic unbiasedness and consistency of a class of simple estimators of entropy. In addition, some entropy-based tests of goodness-of-fit and entropy-based test of independence are presented. At present, we have neither exact nor asymptotic results on the distribution of the proposed entropy estimator. Thus, we will present a simulation study showing that for many families of distribution, widely used in statistics, the hypothesis of asymptotic normal distribution of our estimator seems to be acceptable.

Note that Tsybakov and van der Meulen [14] have been considered a truncated version of the estimate of Kozachenko and Leonenko [24]. They proved the root-n consistentcy of the estimator for a class of one-dimensional densities with unbounded support. However, the approach of Tsybakov and van der Meulen [14] does not appear to be applicable in a multidimensional situation.

2. Estimation of entropy of multivariate density

Let R^m be m-dimensional Euclidean space with Euclidian distance

$$\rho(x, y) = \sqrt{\sum_{j=1}^{m}(x_j - y_j)^2},$$

between $x = (x_1, \ldots, x_m) \in R^m$ and $y = (y_1, \ldots, y_m) \in R^m$. Denote by $v(y, r) = \{x \in R^m\colon \rho(x, y) < r\}$ an open ball of radius $r > 0$ with center $y \in R^m$ and let $|v(y, r)| = r^m c_1(m)$ be its volume, where

$$c_1(m) = \frac{2\pi^{m/2}}{m\Gamma(m/2)}. \tag{2}$$

Now, consider the m-dimensional vector ξ with unknown density function $f(x)$, $x \in R^m$. Denote by $\mathrm{supp}(f) = \{x \in R^m\colon f(x) > 0\}$. The problem is to estimate the entropy H, given in equation (1), based on the independent identically distributed sample $X_1, \ldots, X_N$, $N \geq 2$ of random vector ξ.

For a fixed observation X_i, $i \in \{1, \ldots, N\}$ and fixed $k \in \{1, \ldots, N-1\}$, we define random variables $\rho_{i,k}$ as follows:

$$\begin{aligned}
\rho_{i,1} &:= \min\{\rho(X_i, X_j),\ j \in \{1, \ldots, N\}\backslash\{i\}\} = \rho(X_i, X_{j_1}),\\
\rho_{i,2} &:= \min\{\rho(X_i, X_j),\ j \in \{1, \ldots, N\}\backslash\{i, j_1\}\} = \rho(X_i, X_{j_2}),\\
&\ \vdots
\end{aligned}$$

$$\rho_{i,k} := \min\{\rho(X_i, X_j),\, j \in \{1, \dots, N\}\backslash\{i, j_1, \dots, j_{k-1}\}\} = \rho(X_i, X_{j_k}),$$

$$\vdots$$

$$\rho_{i,N-1} := \max\{\rho(X_i, X_j),\, j \in \{1, \dots, N\}\backslash\{i\}\} = \rho(X_i, X_{j_{N-1}}),$$

where $X_{j_{k-1}}$ is a vector such that with probability 1:

$$\rho(X_i, X_{j_{k-2}}) < \rho(X_i, X_{j_{k-1}}) < \rho(X_i, X_{j_k}).$$

Consequently with probability 1, we have

$$\rho(X_i, X_{j_1}) < \rho(X_i, X_{j_2}) < \cdots < \rho(X_i, X_{j_{N-1}}).$$

We note that $\rho_{i,k}$ is the distance of X_i and its kth nearest neighbor. For a fixed $k \in \{1, \dots, N-1\}$, we let

$$\bar{\rho}_k = \left\{\prod_{i=1}^{N} \rho_{i,k}\right\}^{1/N}$$

denote the geometric mean of random variables $\rho_{1,k}, \dots, \rho_{N,k}$. We define the statistical estimate of equation (1)

$$H_{k,N} = m \ln \bar{\rho}_k + \ln(N-1) - \psi(k) + \ln c_1(m), \tag{3}$$

where $k \in \{1, 2, \dots, N-1\}$, $c_1(m)$ is given in equation (2) and

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} = \frac{\mathrm{d}}{\mathrm{d}z} \ln \Gamma(z) = \int_0^\infty \left(\frac{\mathrm{e}^{-t}}{t} - \frac{\mathrm{e}^{-zt}}{(1-\mathrm{e}^{-t})}\right) \mathrm{d}t \tag{4}$$

is the digamma function. Note that $\psi(1) = -\ln \gamma = -c_2 \sim 0.5772$ is an Euler constant, *i.e.*, $\gamma = \exp\{-\int_0^\infty \mathrm{e}^{-t} \ln t \,\mathrm{d}t\}$.

The estimate $H_{1,N}$ was considered by Kozachenko and Leonenko [24]. We shall prove that the proposed estimates $H_{k,N}$, $k \in \{1, \dots, N-1\}$ are asymptotically unbiased and consistent as $N \to \infty$ under very weak conditions on the density function.

Our main results are the following:

THEOREM 1 *Let $k \in \{1, \dots, N-1\}$ fixed and suppose that there exists an $\varepsilon > 0$ such that*

$$\int_{R^m} |\ln f(x)|^{1+\varepsilon} f(x)\,\mathrm{d}x < \infty, \tag{5}$$

$$\int_{R^m}\int_{R^m} |\ln \rho(x, y)|^{1+\varepsilon} f(x) f(y)\,\mathrm{d}x\,\mathrm{d}y < \infty,$$

and probability density function $f(x)$, $x \in R^m$, is bounded then $\lim_{N\to\infty} EH_{k,N} = H$, *where entropy H is defined in equation* (1)

THEOREM 2 *Let $k, \varepsilon > 0$ as in Theorem* 1 *and*

$$\int_{R^m} |\ln f(x)|^{2+\varepsilon} f(x)\,\mathrm{d}x < \infty, \tag{6}$$

$$\int_{R^m}\int_{R^m} |\ln \rho(x, y)|^{2+\varepsilon} f(x) f(y)\,\mathrm{d}x\,\mathrm{d}y < \infty,$$

and probability density function $f(x)$, $x \in R^m$, is bounded then the entropy estimate $H_{k,N}$ is given in equation (2), *is a weak consistent estimate of H, as $N \to \infty$.*

The proofs of theorems are given in section 6. Theorems 1 and 2 generalize the results of Kozachenko and Leonenko [24], where the case $k = 1$ is considered. Note that our conditions are weaker than those in Kozachenko and Leonenko [24] even in the case $k = 1$.

The above theorems can be compared with the results of van Es [13], where the case $m = 1$ is considered. Our conditions seem weaker than the conditions of Theorem 2 of van Es [13]; in particular, we do not assume that the density function is bounded and bounded away from zero. However, we are unable to present the results on the almost sure convergence and asymptotic distributions of our estimates. Furthermore, the approach of van Es [13] does not appear to be applicable in a multidimensional situation.

3. Test of goodness-of-fit

The main idea for the construction of tests of goodness of fit is based on the maximum entropy principle. Consider a class of densities satisfying certain restrictions. Find a consistent estimator of entropy for the members of the class. Next, using so-called maximum entropy principle (see, *e.g.*, ref. [15] or refs. [17, 18] for more recent developments) determines a member of the class maximizing entropy and find its parametric consistent estimator. Finally, take a function of the above estimators as a test statistic of goodness-of-fit for the member maximizing the entropy.

Tests of goodness-of-fit based on the sample entropy was proposed by Vasicek [9] for one-dimensional normal distribution and by Dudewicz and van der Meulen [19] for uniform distribution, among the others. The essential difference between the tests proposed by these authors and ours lies in the choice of entropy estimator. In our work, we have broaded this list of entropy-based tests. In particular, we consider the goodness-of-fit test for multidimensional normal distribution.

3.1 *Test for normality*

Let $\mathcal{K}$ be a class of m-dimensional density functions $f(x), x \in R^m$ with $\operatorname{supp}\{f\} = R^m$ which satisfy conditions (4) and (5). Note that the density f^* of a nonsingular normal distribution $\mathcal{N}(\alpha, \Sigma)$ belongs to the class $\mathcal{K}$.

It is well known [15, p. 120; 25, p. 533] that among all the distributions with densities $f(x)$, $x \in R^m$ such that $\operatorname{supp}\{f\} = R^m$ and

$$\int_{R^m} x f(x)\,\mathrm{d}x = \alpha, \quad \int_{R^m} [(x-\alpha)(x-\alpha)'] f(x)\,\mathrm{d}x = \Sigma, \tag{7}$$

the entropy $H(f)$ is maximized by the normal distribution, *i.e.*,

$$H(f) \le H(f^*) = \frac{m}{2}\ln(2\pi) + \frac{m}{2} + \left(\frac{1}{2}\right)\ln\det\Sigma. \tag{8}$$

The last equality can be rewritten in the following form:

$$\frac{\exp\{H(f^*)\}}{\{\det\Sigma\}^{1/2}} = \{2\pi\mathrm{e}\}^{m/2}. \tag{9}$$

Let $X_1, X_2, \ldots, X_N$, $N \ge 2$ be an independent sample from a member of $\mathcal{K}$ having finite second order. Consider the sample covariance matrix

$$\Sigma_N = (\sigma_{ijN})_{1\le i,j\le m} = \left(\frac{1}{N-1}\sum_{\nu=1}^{N}(X_\nu^{(i)} - \bar{X}_i)(X_\nu^{(j)} - \bar{X}_j)'\right)_{1\le i,j\le m},$$

as a consistent estimate of Σ, as $N \to \infty$, where $X_\nu = (X_\nu^{(1)}, \ldots, X_\nu^{(m)})'$, $\nu = 1, \ldots, N$, and $\bar{X}_i = (1/N)\sum_{\nu=1}^{N} X_\nu^{(i)}$, $i = 1, \ldots, m$. Then, under the null hypothesis H_0: $X_1, X_2, \ldots, X_N$ are sample from normal distribution $\mathcal{N}_m(\alpha, \Sigma)$, we obtain from equation (9), Theorem 2, and Slutsky's theorem, that

$$\{\det \Sigma_N\}^{-1/2} \exp\{H_{k,N}\} \longrightarrow \{2\pi \mathrm{e}\}^{m/2}$$

in probability as $N \to \infty$, for every fixed $k \in \{1, \ldots, N-1\}$. Under the alternative H_1: $X_1, X_2, \ldots, X_N$ is a sample from any other member of $\mathcal{K}$, we find that for each fixed $k \in \{1, \ldots, N-1\}$

$$\{\det \Sigma_N\}^{-1/2} \exp\{H_{k,N}\} \longrightarrow \{\det \Sigma\}^{-1/2} \exp\{H(f)\} < \{2\pi \mathrm{e}\}^{m/2}.$$

In other words, the test for normality is consistent against such alternatives as mentioned earlier.

3.2 *Test for Laplace distribution*

Now, consider a class $\mathcal{K}_1$ of the densities f with supp$f = R^1$ such that the condition (6) holds and with finite mean. Note that the density of the Laplace distribution

$$f^*(x) = \frac{2}{\alpha}\mathrm{e}^{-|x|/\alpha}, \quad x \in R^1, \ \alpha > 0$$

belongs to this class. It is known [15, p. 56; 26, p. 564] that among all one-dimensional distributions of random variables X with density belonging to $\mathcal{K}_1$ with given $E|X| = g$, the entropy $H(f)$ is maximized by Laplace distribution, *i.e.*,

$$H(f) \leq H(f^*) = 1 + \ln 2\alpha,$$

i.e.,

$$\frac{1}{\alpha}\mathrm{e}^{H(f^*)} = 2\mathrm{e}. \tag{10}$$

One knows that $\alpha_N = (1/N)\sum_{i=1}^{N} |X_i|$ provides a consistent estimate of α based on an independent sample $X_1, X_2, \ldots, X_N$, $N \geq 2$. Consequently, under the null hypothesis H_0: $X_1, X_2, \ldots, X_N$ are sample from Laplace distribution $\mathcal{L}(\alpha)$, we obtain from equation (10), Theorem 2, and Slutsky's theorem that for each $k \in \{1, \ldots, N-1\}$

$$\left(\frac{1}{\alpha_N}\right) \exp\{H_{k,N}\} \longrightarrow 2\mathrm{e}$$

in probability as $N \to \infty$. Under the alternative H_1: $X_1, X_2, \ldots, X_N$ are sample from any other distribution from class $\mathcal{K}_1$, we find that for each fixed $k \in \{1, \ldots, N-1\}$

$$\left(\frac{1}{\alpha_N}\right) \exp\{H_{k,N}\} \longrightarrow \left(\frac{1}{\alpha}\right) \exp\{H(f)\} < 2\mathrm{e},$$

where

$$\alpha = \int_{-\infty}^{\infty} |x| f(x)\, \mathrm{d}x.$$

Hence, the test is consistent against H_1 as mentioned earlier.

3.3 *Test for exponentiality*

Consider the class $\mathcal{K}_2$ of the densities f with supp $f = (0, \infty)$ such that equations (5) and (6) hold. Note that the density function of exponential distribution

$$f^*(x) = \frac{1}{\lambda}\mathrm{e}^{-x/\lambda}, \quad x > 0, \;\; \lambda > 0$$

belongs to this class. It is well known [15, p. 56; 26, p. 564] that among all members of $\mathcal{K}_2$ such that

$$\int_0^\infty x f(x)\,\mathrm{d}x = \lambda, \quad \lambda > 0, \tag{11}$$

the entropy $H(f)$ is maximized by exponential density, *i.e.*,

$$H(f) \le H(f^*) = 1 + \ln\lambda,$$

consequently,

$$\lambda^{-1}\exp\{H(f^*)\} = e. \tag{12}$$

We know that $\lambda_N = (1/N)\sum_{i=1}^N X_i$ provides a consistent estimate of a parameter λ based on the random sample X_i, $i = 1, 2, \ldots, N$. Hence, under H_0: $X_1, X_2, \ldots, X_N$ are sample from exponential distribution, we have from equation (11), Theorem 2, and Slutsky's theorem that for each fixed $k \in \{1, \ldots, N-1\}$

$$(\lambda_N)^{-1}\exp\{H_{k,N}\} \longrightarrow e,$$

in probability as $N \to \infty$.

Under the alternative H_1: $X_1, X_2, \ldots, X_N$ are sample from any other member of $\mathcal{K}_2$, we have that for each fixed $k \in \{1, \ldots, N-1\}$

$$(\lambda_N)^{-1}\exp\{H_{k,N}\} \longrightarrow \lambda^{-1}\exp\{H(f)\} < e,$$

in probability as $N \to \infty$. This means that the above mentioned test of exponentiality is consistent for such alternatives.

3.4 *Test for gamma distribution*

Consider again the class $\mathcal{K}_2$. Clearly, the density function of Gamma distribution

$$f^*(x) = \frac{\alpha^{-p}x^{p-1}\mathrm{e}^{-x/\alpha}}{\Gamma(p)}, \quad x > 0, \;\; \alpha > 0, \;\; p > 1$$

belongs to this class. It is known [15, p. 57] that among all one-dimensional densities $f(x)$ with supp$(f) = (0, \infty)$ such that for given

$$E(X) = g_1 < \infty, \quad E(\ln X) = g_2,$$

the entropy $H(f)$ is maximized by gamma distribution, *i.e.*,

$$H(f) \le H(f^*) = p(1 + \ln\alpha) + \ln\Gamma(p) - (p-1)E(\ln X). \tag{13}$$

The unknown parameters α, p can be consistently estimated by the method of moments, *i.e.*,

$$\alpha_N = \frac{\bar{X}_N}{S_N^2}, \quad p_N = \left(\frac{\bar{X}_N}{S_N}\right)^2,$$

where $\bar{X}_N$, S_N^2 are, respectively, the sample mean and the variance. Further, the parameter $E(\ln X)$ can be consistently estimated [15, p. 270] by the maximum likelihood estimate

$g_N = (\prod_{i=1}^N X_i)^{1/N}$. Consequently under the null hypothesis H_0: $X_1, \ldots, X_N$ is the sample from gamma distribution. We find from equation (12) and Theorem 2 that

$$\zeta_N = p_N(1 + \ln \alpha_N) + \ln \Gamma(p_N) - (p_N - 1)g_N - H_{k,N} \longrightarrow 0,$$

in probability as $N \to \infty$ for every fixed $k \in \{1, \ldots, N-1\}$. Under the alternative H_1: $X_1, \ldots, X_N$ is sample from any other member of class $\mathcal{K}_2$ with finite second order moment, we obtain that for each $k \in \{1, 2, \ldots, N-1\}$ that

$$\zeta_N \longrightarrow \text{const} \neq 0,$$

in probability as $N \to \infty$.

3.5 *Test for β-distribution*

Now, consider a class $\mathcal{K}_3$ of densities f with $\mathrm{supp}(f) = (0, 1)$ which satisfy conditions (5) and (6). Clearly, the density of Beta-distribution

$$f^*(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0, 1), \ \alpha > 0, \ \beta > 0$$

belongs to this class. It is known [15, p. 78] that among all the densities f with $\mathrm{supp}(f) = (0, 1)$ such that given

$$E \ln X = g_1 < \infty, \quad E \ln(1 - X) = g_2,$$

where X is random variable with density f, the entropy is maximized by the density with Beta distribution, *i.e.*,

$$\begin{aligned} H(f) \leq H(f^*) &= \ln B(\alpha, \beta) - (\alpha - 1)E \ln X - (\beta - 1)E \ln(1 - X) \\ &= \ln B(\alpha, \beta) - (\alpha - 1)[\psi(\alpha) - \psi(\alpha + \beta)] - (\beta - 1)[\psi(\beta) - \psi(\alpha + \beta)]. \end{aligned} \tag{14}$$

Let $X_1, X_2, \ldots, X_N$, $N \geq 2$ be an independent sample from a member of $\mathcal{K}_3$ with finite second-order moment. Let $\bar{X}_N$, S_N^2 denote, respectively, the sample mean and the variance, then

$$\begin{aligned} \alpha_N &= \frac{(\bar{X}_N)^2(1 - \bar{X}_N)}{S_N^2} - \bar{X}_N, \\ \beta_N &= \frac{(1 - \bar{X}_N)[\bar{X}_N(1 - \bar{X}_N) - S_N^2]}{S_N^2}, \\ g_{1,N} &= (1/N) \sum_{i=1}^N \ln X_i, \quad g_{2,N} = (1/N) \sum_{i=1}^N \ln(1 - X_i) \end{aligned}$$

are consistent estimators of parameters α, β, g_1, g_2, respectively.

Under the null hypothesis H_0: $X_1, X_2, \ldots, X_N$ are sample from β-distribution $\beta(\alpha, \beta)$, we have from Theorem 2 and equation (13) that for each fixed $k \in \{1, \ldots, N-1\}$

$$\zeta_N = \ln B(\alpha_N, \beta_N) - (\alpha_N - 1)g_{1N} - (\beta_N - 1)g_{2N} - H_{k,N} \longrightarrow 0, \tag{15}$$

in probability as $N \to \infty$. Whereas under the alternative H_1: $X_1, X_2, \ldots, X_N$ are sample from any other distribution from class $\mathcal{K}_2$ with finite second-order moment, we obtain that

$\zeta_N \to \text{const} \neq 0$ as $N \to \infty$ in probability for every fixed $k \in \{1, 2, \dots, N-1\}$. This means that this test is consistent for such alternatives.

Remark 1 Using the same ideas, the entropy-based tests of goodness-of-fit can be proposed for χ^2, Erlang, F, logistic, log normal, uniform, Weibull one-dimensional distributions according to table 3.8 in Kapur [15], p. 84–86.

Remark 2 Kapur [15, chapter 5] presents maximum entropy multivariate log-normal, polynomial exponential-sum, Dirichlet, Logistic, Cauchy, Pareto and some other distributions. An interesting open problem is to describe entropy-based tests of goodness-of-fit for some of these distributions based on the Theorem 2 and Slutsky's theorem.

4. Entropy-based test of independence

The entropy of multivariate distribution can be considered as a measure of interdependence [10, 11, 15, 27–29]. In fact, the components of random vector are independent if and only if the joint entropy is equal to the sum of the marginal entropies; for discrete random vector, the results are given in Kapur [15, p. 215], whereas in continuous case, this characterization follows from Kullback [28]. Here, we present some tests of independence based on the sample entropy in the bivariate case.

Let $X = (X^{(1)}, X^{(2)})$ be a two-dimensional vector with joint density $f(x_1, x_2)$, $(x_1, x_z) \in R^2$ and $f_i(x_i)$, $x_i \in R^1$, $i = 1, 2$ and $f_2(x_2)$, $x_2 \in R^1$ and the conditional density $f(x_i|x_j)$, $i, j = 1, 2, i \neq j$ denote, respectively, the marginals and the conditional densities. Consider, the joint entropy

$$H(X^{(1)}, X^{(2)}) = -\int_{R^2} f(x_1, x_2) \ln f(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2, \tag{16}$$

the marginal entropies

$$H(X^{(1)}) = -\int_{R^1} f_1(x_1) \ln f_1(x_1)\, \mathrm{d}x_1,$$
$$H(X^{(2)}) = -\int_{R^2} f_2(x_2) \ln f_2(x_2)\, \mathrm{d}x_2,$$

and the conditional entropy

$$\begin{aligned} H(X^{(2)}|X^{(1)}) &= -\int_{R^2} f_1(x_1) f(x_2|x_1) \ln f(x_2|x_1)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\ &= -\int_{R^2} f(x_1, x_2) \ln f(x_2|x_1)\, \mathrm{d}x_1\, \mathrm{d}x_2. \end{aligned}$$

The relative entropy or mutual information between $X^{(1)}$ and $X^{(2)}$ is

$$\begin{aligned} I(X^{(1)}, X^{(2)}) &= H(X^{(1)}) + H(X^{(2)}) - H(X^{(1)}, X^{(2)}) \\ &= H(X^{(1)}) - H(X^{(1)}|X^{(2)}) = H(X^{(2)}) - H(X^{(2)}|X^{(1)}). \end{aligned} \tag{17}$$

If $g(x, x_2)$ be some other two-dimensional density given, then Kullback [28] proved that the minimum of a functional

$$J(f, g) = \int_{R^2} f(x_1, x_2) \ln \frac{f(x_1, x_2)}{g(x_1, x_2)}\, \mathrm{d}x_1\, \mathrm{d}x_2,$$

over all bivariate densites $f(x_1, x_2)$ with given marginal $f_i(x_i)$, $i = 1, 2$ is

$$J(f^*, g) = \int_{R^1} f_1(x_1) \ln a(x_1)\, \mathrm{d}x_1 + \int_{R^1} f_2(x_2) \ln b(x_2)\, \mathrm{d}x_2,$$

where the minimizing density is

$$f^*(x_1, x_2) = a(x_1)b(x_2)g(x_1, x_2)$$

and the functions a, b are to be determined by the relation

$$f_1(x_1) = a(x_1) \int_{R^1} b(x_2)g(x_1, x_2)\, \mathrm{d}x_2, \quad f_2(x_2) = b(x_2) \int_{R^1} a(x_1)g(x_1, x_2)\, \mathrm{d}x_1 \tag{18}$$

From Kullback [28], we obtain

$$J(f, g) = J(f^*, g) + J(f, f^*) \geq J(f^*, g) \geq 0$$

with equality if and only if $I(f, f^*) = 0$, *i.e.*, if and only if,

$$f(x_1, x_2) = f^*(x_1, x_2), \text{ a.e.} \tag{19}$$

In particular, if $g(x_1, x_2) = f_1(x_1) f_2(x_2)$, then from equations (18) and (19) we obtain $a(x_1)a(x_2) = 1$, $f^*(x_1, x_2) = f_1(x_1) f_2(x_2)$ and $J(f, f^*) \geq 0$, where the equality holds if and only if $f(x_1, x_2) = f_1(x_1) f_2(x_2)$, a.e. Thus, $X^{(1)}$ and $X^{(2)}$ are independent if and only if $I(X^{(1)}, X^{(2)}) = 0$, or

$$H(X^{(1)}, X^{(2)}) = H(X^{(1)}) + H(X^{(2)}). \tag{20}$$

The characterization of independence given by equation (20) together with Theorem 2 can be used for entropy-based test statistics for this purpose. Let $X_1 = (X_1^{(1)}, X_1^{(2)})', \ldots, X_N = (X_N^{(1)}, X_N^{(2)})'$, $N \geq 2$ be independent observations of random vector $X = (X^{(1)}, X^{(2)})'$ with density function $f(x_1, x_2)$. Consider the entropy estimate $H_{k,N}$, $k \in \{1, \ldots, N-1\}$ defined by the formula (3) with $m = 2$ and entropy estimates $H_{k,N}^1$ and $H_{k,N}^2$, defined via the formula (3) for $m = 1$ for the samples $(X_1^{(1)}, \ldots, X_N^{(1)})$ and $(X_N^{(2)}, \ldots, X_N^{(2)})$, respectively. From Theorems 1 and 2 we obtain that $H_{k,N}$ is asymptotically unbiased and consistent estimate of joint entropy $H(X^{(1)}, X^{(2)})$ and $H_{k,N}^{(1)}$, $H_{k,N}^{(2)}$ are asymptotically unbiased and consistent estimate of the marignals entropies $H(X^{(1)})$ and $H(X^{(2)})$, respectively, under a set of assumptions on the density for every $k \in \{1, \ldots, N-1\}$.

Consider, the class $\mathcal{K}_4$ of two-dimensional distributions with density functions satisfying equations (5) and (6) with $m = 2$ and also its both marignal densities satisfying the same conditions with $m = 1$ for every fixed $k \in \{1, 2, \ldots, N-1\}$. Then, under the null hypothesis H_0: $X^{(1)}$ and $X^{(2)}$ are independent random variables in a class of two-dimensional distributions $\mathcal{K}_4$, we have from equation (20) and Theorem 2 that

$$\zeta_N = H_{k,N} - H_{k,N}^{(1)} - H_{k,N}^{(2)} \longrightarrow 0,$$

in probability as $N \to \infty$ for every fixed $k \in \{1, \ldots, N-1\}$. Whereas under the alternative H_1: $X^{(1)}$ and $X^{(2)}$ are dependent random variables from the class $\mathcal{K}_4$, we have

$$\zeta_N \longrightarrow \text{const} \neq 0,$$

in probability as $N \to \infty$ for every fixed $k \in \{1, 2, \ldots, N-1\}$. This means that this test is consistent for such alternatives.

5. Simulations

5.1 H_{kN} distribution

Here, 5000 samples of different size ($N = 10, 30, 50, 100$) are generated from several alternatives to normalities, namely, Beta, Cauchy, Gamma, Laplace, and Student-t and analyzed the simulated distribution of the statistic $H_{k,N}$ under these alternatives. More explicitly, our aim is twofold: (a) find a model for the distribution of the proposed statistic, which behaves well under different type of alternatives considered and (b) show for a particular choice of alternaltives (Gamma and Student-t) that the proposed model is effective for approximating the simulated distribution of H_{kN}. To this end, for each sample size, starting from the 5000 values of the statistic H_{kN}, we compute summary measures of the simulated distribution of H_{kN} and also examine the Q–Q plot under each alternative, to assert departure from normality. As regards the choice of k-value of spacing order for given N (optimal choice being inconclusive at present), we shall use the following heuristic formula [30] $k = [\sqrt{N} + 0.5]$, besides $k = 1$. This choice, indeed, implies better global performance of $H_{k,N}$ for all distributions and sample sizes considered.

From the simulation study, we found both the summary statistics of H_{kN} and corresponding Q–Q plots exhibited similar behavior under the various alternatives, *i.e.*, the distribution of $H_{k,N}$ is a bit left skewed and moderately leptokurtic; however, as the sample size increases, the related descriptive statistics tend to those of normal. As an illustration, we shall report the results for the Beta-distribution where the summary statistics show higher values and also Q–Q plot in figure 1 shows more pronounced departure from the normality.

In table 1, we report mean, variance, bias, asymmetry and kurtosis measures of $H_{k,N}$ computed from samples of different sample size generated from Beta-distribution with parameters $\alpha = 1.5$ and $\beta = 2.5$.

From the above discussion, it seems convenient to consider a family of distribution, recently introduced by Kotz *et al.* [31] and further discussed by Ayebo and Kozubowski [32] for approximating the distribution of $H_{k,N}$ under the alternatives considered here. This family of distributions depends on four parameters, namely location θ, scale σ, symmetry κ and tail behavior α, and its density has the following form:

$$f(x) = \frac{\alpha}{\sigma\Gamma(1/\alpha)} \frac{\kappa}{1+\kappa^2} \exp\left\{ -\frac{\kappa^\alpha}{\sigma^\alpha}[(x-\theta)^+]^\alpha - \frac{1}{\kappa^\alpha\sigma^\alpha}[(x-\theta)^-]^\alpha \right\}, \quad \alpha, \sigma, \kappa > 0, \quad \theta \in \mathbb{R},$$

where $(x-\theta)^+$ and $(x-\theta)^-$, respectively, are the positive and the negative part of $(x-\theta)$. From table 1, one sees that it seems reasonable to take $\alpha = 2$, light tailed near-normal. We shall use this subfamily of asymmetric exponential power distributions, known as skewed normal distribution, for modeling the distribution of H_{kN} under the various alternatives. For $\kappa = 1$ and $\alpha = 2$, the above family leads to normal distribution.

Matching the first three moments of the skewed normal distribution with those given in the table 1 for $N = 10$, we find $\kappa = 1.7759$, $\sigma^2 = 0.1174$, and $\theta = 0.0086$.

Table 1. Summary measures for Beta (1.5, 2.5).

N	k	Mean ($H_{k,N}$)	Var($H_{k,N}$)	Bias($H_{k,N}$)	$\gamma_1(H_{k,N})$	$\gamma_2(H_{k,N})$
10	3	−0.2259	0.0901	−0.0373	−0.7137	0.8039
30	5	−0.2024	0.0209	−0.0138	−0.4904	0.5334
50	7	−0.2005	0.0102	−0.0119	−0.3690	0.2215
100	10	−0.1978	0.0043	−0.0092	−0.3498	0.3226

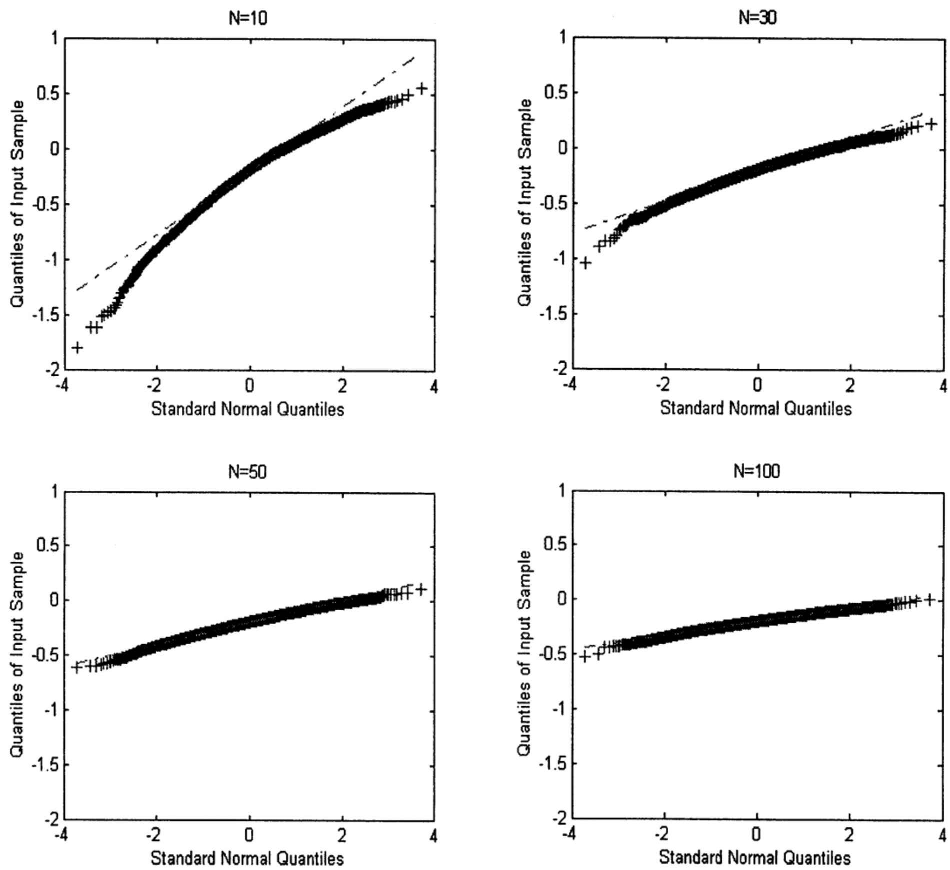

Figure 1. $H_{k,N}$ Monte Carlo Q–Q plot for Beta(1.5, 2.5) distribution.

Note that while the largest value of κ results from Beta(1.5, 2.5) for $N = 10 (\kappa = 1.7759)$, the same is not true for the other parameter values; indeed, the largest values of σ^2 and θ lie with Student-$t_3 (\sigma^2 = 0.2823)$ and Gam(2, 3)$(\theta = 2.7691)$. Obviously, these bounds vary considerably for replications for small sample sizes, but become increasingly stable as the sample size increases. In fact, on repeated simulation study for Beta with same parameter values and $N = 10$, we found $\kappa = 2.1736$, $\sigma^2 = 0.0914$, and $\theta = 0.0541$. Next, to see how closely the above model matches the simulated distribution of $H_{k,N}$, we exhibit below the graphs for Gamma and Student-t distribution alternatives. Here, F_N and F^* represent the simulated empirical distribution and the model (skewed normal) distribution, respectively.

From figure 2, we see that for Gam(2, 3) the two distributions (simulated empirical distribution and skewed normal distribution) match almost perfectly even for small sample size; in fact, $\sup |F_N - F^*| \leq 0.045$. This is not the case for Student-t_3 distribution as shown by figure 3. In this case $(\sup |F_N - F^*| \leq 0.213$ for $N = 10$. This visible mismatch may be attributed to the large variance characterizing the distribution of $H_{k,N}$, when N is small and it disappears for large N as $\sup |F_N - F^*| \leq 0.081$ for $N = 100$ implies.

5.2 *Simulated power comparison*

Here, we compute the simulated power of the goodness-of-fit test for normality based on the $H_{k,N}$ estimator of entropy using Monte Carlo method against the same alternatives considered

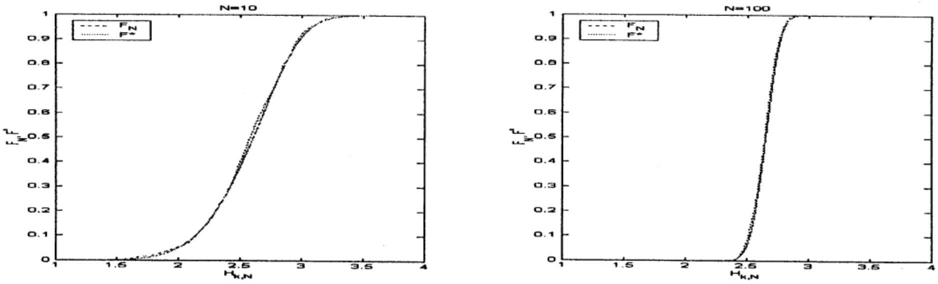

Figure 2. Simulated emprical and skewed normal distribution for Gam(2,3).7

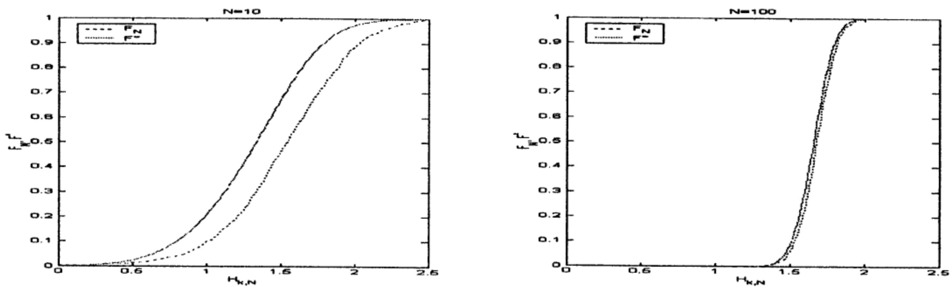

Figure 3. Simulated emprical distribution and skewed normal distribution for Student-t_3.

Table 2. Simulated power of some test statistics of normality.

		Vasicek			van Es			Correa			$H_{N,k}$		
Alternative	N	k	Point	Power	k	Point	Power	k	Point	Power	k	Point	Power
Exp(I)	10	4	2.182	0.4430	3	2.863	0.3400	4	2.662	0.4480	3	2.756	0.2950
	20	5	2.737	0.8620	3	3.093	0.6750	5	3.068	0.8380	6	3.161	0.6120
	30	4	3.079	0.9760	3	3.257	0.8570	4	3.419	0.9760	9	3.364	0.8120
Gam(2)	10	4	2.194	0.2420	2	2.839	0.1650	4	2.673	0.2520	4	2.847	0.1690
	20	5	2.726	0.4890	3	3.112	0.3660	5	3.048	0.4600	7	3.213	0.3380
	30	5	3.060	0.7290	5	3.147	0.4920	5	3.370	0.7190	11	3.353	0.4560
U(0,1)	10	2	2.157	0.1940	2	2.829	0.0810	2	2.560	0.1940	2	2.470	0.1110
	20	5	2.724	0.4690	2	3.104	0.1430	3	3.155	0.4370	2	2.831	0.1390
	30	14	2.573	0.7880	2	3.267	0.2120	13	2.903	0.7400	3	3.140	0.2010
Beta(2,1)	10	3	2.213	0.1780	1	2.483	0.0950	2	2.579	0.1910	3	2.765	0.1080
	20	5	2.736	0.4880	2	3.080	0.1770	5	3.067	0.4580	4	3.116	0.1580
	30	8	2.926	0.7310	2	3.264	0.2550	5	3.377	0.7150	5	3.292	0.2120
Cau(0,1)	10	3	2.215	0.4390	3	2.842	0.6330	1	2.061	0.4160	4	2.882	0.5920
	20	3	2.789	0.7490	9	2.869	0.9260	1	2.557	0.7510	9	3.233	0.9120
	30	3	3.033	0.9170	13	2.904	0.9840	3	3.402	0.9080	13	3.354	0.9830
t(2)	10	2	2.142	0.4140	4	2.772	0.5650	4	2.667	0.3890	4	2.818	0.5290
	20	1	2.247	0.6950	9	2.865	0.8700	1	2.556	0.7000	9	3.209	0.8440
	30	2	2.908	0.8660	12	2.901	0.9630	2	3.309	0.8510	12	3.371	0.9580
Lap(0,1)	10	4	2.190	0.0940	3	2.810	0.1850	4	2.675	0.1070	4	2.863	0.1550
	20	3	2.798	0.1090	8	2.855	0.4030	1	2.545	0.1120	8	3.217	0.3730
	30	2	2.923	0.1770	11	2.925	0.5520	2	3.330	0.1730	11	3.397	0.5200

by Vasicek [9] and compare it with the simulated power of some well-known goodness-of-fit tests. For each $N = 10, 20, 30$, we simulate 5000 samples of size N from the normal distribution to evaluate the critical values for four different tests. The percentage points of the distributions (point, in the following table) were estimated by the corresponding order statistics. For each alternative, 1000 samples of size $N = 10, 20, 30$ were generated, and the simulated power was found by the relative frequency of samples falling in the critical region. The alternatives investigated are exponential, Gamma, Beta, uniform, Student-t, Cauchy, and Laplace. In this simulation study, the value of k was taken as one leading for each sample size N to the highest power.

It is apparent from the table 2 that none of the tests considered performs better than the others against all alternatives. For asymmetric short and long tailed alternatives, the tests proposed by Vasicek and by Correa outperform the other two with the former having a slight superiority on the other for $N = 20, 30$. On the other hand, test based on $H_{k,N}$ which exhibits similar behavior to the test developed by van Es entropy estimator, shows higher power against symmetric and heavy tailed alternatives. In other words, our test is preferable to the others against symmetric heavy tailed alternative distributions. Moreover, it has the additional advantage that it can be applied for testing multinormality. The same is not true for its competitors.

6. Proofs

Before proving Theorems 1 and 2, we mention some known results.

LEMMA 3 [33] *If $g(x) \in L_1(R^m)$ then for any sequences of open balls $v(x, r_k)$ of radius $r_k \to 0$ and for almost all $x \in R^m$*

$$\lim_{k\to\infty} \frac{1}{|v(x, r_k)|} \int_{v(x,r_k)} g(y)\,\mathrm{d}y = g(x).$$

LEMMA 4 [34, p. 862] *Let ξ_n, $n = 1, 2, \ldots$ be a sequence of random variables, which converges in distribution to random variable ξ, and suppose there exists $\varepsilon > 0$ and a constant $c > 0$ such that $E|\xi_n|^{1+\varepsilon} < c$ for all $n \geq 1$. Then, for all $a < 1 + \varepsilon$ and all integers $r < 1 + \varepsilon$,*

$$E|\xi_n|^a \longrightarrow E|\xi|^a, \quad E\xi_n^r \longrightarrow E\xi^r, \;\; n \longrightarrow \infty.$$

LEMMA 5 [24] *Let $F(u)$ be a distribution function. Then, for $\alpha \geq 1$*

$$\int_1^\infty (\ln u)^\alpha\,\mathrm{d}F(u) = \alpha \int_1^\infty (\ln u)^{\alpha-1} u^{-1}(1 - F(u))\,\mathrm{d}u \tag{21}$$

whenever one of the two improper integrals converges.

Proof Integrating by parts, we get

$$\begin{aligned}\int_1^u (\ln(t))^\alpha\,\mathrm{d}F(t) &= -\int_1^u (\ln t)^\alpha d(1 - F(t)) \\ &= -(1 - F(u))(\ln(u))^\alpha + \alpha \int_1^u (\ln(t))^{\alpha-1} t^{-1}(1 - F(t))\,\mathrm{d}t.\end{aligned}$$

To complete the proof, we merely need to show

$$\lim_{u\to\infty}(1-F(u))(\ln u)^{\alpha}=0.$$

This follows from the inequality

$$1-F(u)\le(\ln u)^{-\alpha}\int_u^{\infty}(\ln t)^{\alpha}\,\mathrm{d}F(t).$$

Note that we can likewise show that if the right hand side of equation (21) is finite, then

$$\lim_{u\to\infty}(1-F(u))(\ln u)^{\alpha}=0.$$

■

LEMMA 1 *Let $F(u)$ be a distribution function and $F(0)=0$. Then, for $\alpha\ge 1$*

$$\int_0^1(-\ln u)^{\alpha}\,\mathrm{d}F(u)=\alpha\int_0^1(-\ln u)^{\alpha-1}u^{-1}F(u)\,\mathrm{d}u \tag{22}$$

Proof Integrating by parts, we obtain

$$\int_u^1(-\ln t)^{\alpha}\,\mathrm{d}F(t)=-F(u)(-\ln u)^{\alpha}+\alpha\int_u^1(-\ln t)^{\alpha-1}t^{-1}F(t)\,\mathrm{d}t.$$

To complete the proof it is sufficient to show that

$$\lim_{u\to 0^+}F(u)(-\ln u)^{\alpha}=0,$$

this follows from the inequality

$$F(u)\le(-\ln u)^{-\alpha}\int_0^u(-\ln t)^{\alpha}\,\mathrm{d}F(t).$$

Similarly, we can show that if the right hand side of equation (22) is finite then

$$\lim_{u\to 0^+}F(u)(-\ln u)^{\alpha}=0.$$

■

6.1 *Proof of Theorem 1*

Note that the Theorem 1 for $k=1$ is given in Kozachenko and Leonenko [24]. Now, the sample entropy $H_{N,k}$ defined in equation (3) can be written as follows:

$$H_{N,k}=\left(\frac{1}{N}\right)\sum_{i=1}^{N}\zeta_{i,k},$$

where $\zeta_{i,k}=\ln\{\rho_{i,k}^{m}(N-1)c_1(m)\exp\{-\psi(k)\}\}$ are identically distributed random variables. Consequently,

$$EH_{N,k}=E\zeta_{i,k}$$

for every $i\in\{1,2,\ldots,N\}$ and fixed $k\in\{1,2,\ldots,N-1\}$. Next, if we let

$$R_N(u)=\left[\frac{u}{(N-1)c_1(m)\mathrm{e}^{-\psi(k)}}\right]^{1/m},\quad u\in R^1,$$

then for $y \in R^m$, we have

$$|v(y, R_N(u))| = \frac{u}{(N-1)\mathrm{e}^{-\psi(k)}} \longrightarrow 0$$

as $N \to \infty$. From the property of $\rho_{i,k}$, it follows that the conditional distribution of $\xi_{N,x,k} = \mathrm{e}^{\zeta_{i,k}}$ is

$$F_{N,x,k}(u) = P(\mathrm{e}^{\zeta_{i,k}} < u | X_i = x) = 1 - \sum_{j=0}^{k-1} q_j(N, u),$$

where

$$q_j(N, u) = \binom{N-1}{j} \left(\int_{v(x, R_N(u))} f(y)\,\mathrm{d}y \right)^j \left(1 - \int_{v(x, R_N(u))} f(y)\,\mathrm{d}y \right)^{N-1-j}.$$

Applying the Lemma 3, we have

$$\lim_{N\to\infty} q_j(N, u) = \frac{(f(x)u)^j}{j!(\mathrm{e}^{-\psi(k)})^j} \exp\left\{ -\frac{uf(x)}{\mathrm{e}^{-\psi(k)}} \right\}.$$

Therefore,

$$\begin{aligned} \lim_{N\to\infty} F_{N,x,k}(u) = F_{x,k}(u) &= 1 - \sum_{j=0}^{k-1} \exp\left\{ -\frac{uf(x)}{\mathrm{e}^{-\psi(k)}} \right\} \frac{(f(x)u)^j}{j!(\mathrm{e}^{-\psi(k)})^j} \\ &= 1 - \exp\left\{ -\frac{uf(x)}{\mathrm{e}^{-\psi(k)}} \right\} \sum_{j=0}^{k-1} \frac{(f(x)u)^j}{j!(\mathrm{e}^{-\psi(k)})^j}. \end{aligned} \tag{23}$$

Let $\xi_{x,k}$ be a random variable with distribution function $F_{x,k}$ and having density function

$$f_{x,k}(u) = \left(\frac{f(x)}{\mathrm{e}^{-\psi(k)}(k-1)!} \right) \exp\left\{ -\frac{uf(x)}{\mathrm{e}^{-\psi(k)}} \right\} \frac{(f(x)u)^{k-1}}{(\mathrm{e}^{-\psi(k)})^{k-1}}, \quad u \geq 0. \tag{24}$$

We want to show that for almost all x

$$\lim_{N\to\infty} E\{\zeta_{i,k} | X_i = x\} = -\ln f(x). \tag{25}$$

Note that

$$\begin{aligned} E \ln \xi_{x,k} &= \int_0^\infty \ln u f_{x,k}(u)\,\mathrm{d}u \\ &= \int_0^\infty \ln \frac{\mathrm{e}^{-\psi(k)} t}{f(x)} \left(\frac{t^{k-1}\mathrm{e}^{-t}}{(k-1)!} \right) \mathrm{d}t \\ &= -\psi(k) + \psi(k) - \ln f(x) \end{aligned}$$

and

$$E \ln \xi_{N,x,k} = E\{\zeta_{i,k} | X_i = x\}.$$

Thus, we can rewrite equation (25) in the following form:

$$\lim_{N\to\infty} E \ln \xi_{N,x,k} = E \ln \xi_{x,k} = -\ln(f(x)). \tag{26}$$

Thus it remains to prove equation (26). Note that equation (26) does not follow from equation (25) because the function $\ln(u)$ is not bounded. If we show that there exists $\varepsilon > 0$

and a constant $c > 0$ such that

$$E|\ln \xi_{N,x}|^{1+\varepsilon} < c \tag{27}$$

then equation (26) follows by Lemma 4.

Thus, it remains to prove equation (27). Note that

$$\begin{aligned}E|\ln \xi_{N,x,k}|^{1+\varepsilon} &= \int_0^1 |\ln u|^{1+\varepsilon}\,\mathrm{d}F_{N,x,k}(u) + \int_1^{\sqrt{N-1}} |\ln u|^{1+\varepsilon}\,\mathrm{d}F_{N,x,k}(u)\\ &\quad + \int_{\sqrt{N-1}}^{\infty} |\ln u|^{1+\varepsilon}\,\mathrm{d}F_{N,x,k}(u) =: I_1 + I_2 + I_3.\end{aligned}$$

Using Lemma 5, we obtain

$$\begin{aligned}I_3 &= (1+\varepsilon)\int_{\sqrt{(N-1)}}^{\infty} (\ln u)^{\varepsilon} u^{-1}(1 - F_{N,x,k}(u))\,\mathrm{d}u\\ &= (1+\varepsilon)\int_{\sqrt{(N-1)}}^{\infty} (\ln u)^{\varepsilon} u^{-1} \sum_{j=0}^{k-1} q_j(N,u)\,\mathrm{d}u.\end{aligned}$$

Note that $\binom{N-1}{j} \le (N-1)^j$ and $\int_{v(x,R_N(u))} f(y)\,\mathrm{d}y < 1$, thus

$$\begin{aligned}I_3 &\le (1+\varepsilon)N^k k \left(1 - \int_{v(x,R_N(\sqrt{N-1}))} f(y)\,\mathrm{d}y\right)^{N-k-1}\\ &\quad\times \int_{\sqrt{N-1}}^{\infty} (\ln u)^{\varepsilon} u^{-1}\left(1 - \int_{v(x,R_N(u))} f(y)\,\mathrm{d}y\right)\mathrm{d}u \longrightarrow 0,\end{aligned}$$

as

$$\begin{aligned}&\lim_{N\to\infty}\left(N^k\left(1 - \int_{v(x,R_N(\sqrt{N-1}))} f(y)\,\mathrm{d}y\right)^{N-k-1}\right)\\ &\quad= \lim_{N\to\infty}\left(N^k \exp\left\{-\sqrt{N-1}\,\frac{f(x)}{\exp\{-\psi(k)\}}\right\}\right) = 0\end{aligned}$$

and the tail integral is also negligible [24].

To deal with I_2, we again first apply Lemma 5 thus

$$\begin{aligned}I_2 &\le (1+\varepsilon)\sum_{j=0}^{k-1}\int_1^{\sqrt{(N-1)}} |\ln u|^{\varepsilon} u^{-1}\left(\sup_{x\in\mathrm{supp}(f)} f(x)\right)^j (N-1)^j \left(\frac{u}{(N-1)\exp\{-\psi(k)\}}\right)^j\\ &\quad\times\left(1 - \int_{v(x,R_N(\sqrt{(N-1)}))} f(y)\,\mathrm{d}y\right)^{N-k-1}\mathrm{d}u,\end{aligned}$$

then as $\lim_{N\to\infty}(1 - \int_{v(x,R_N(u))} f(y)\,\mathrm{d}y)^{N-k-1} = \exp\{-f(x)\exp\{\psi(k)\}u\}$

$$I_2 \le (1+\varepsilon)\sum_{j=0}^{k-1}(M^j(f(x)\exp\{\psi(k)\})^{(j+\varepsilon)}\Gamma(j+\varepsilon)) < \infty,$$

where $M = \sup_{x\in\mathrm{supp}(f)} f(x)$.

Consider next I_1. As $F_{N,x,k}(0) = 0$, we may apply Lemma 1 and obtain

$$
\begin{aligned}
I_1 &= \int_0^1 (-\ln u)^{1+\varepsilon}\, \mathrm{d}F_{N,x,k}(u) \\
&= (1+\varepsilon)\int_0^1 (-\ln u)^{\varepsilon} u^{-1} F_{N,x,k}(u)\,\mathrm{d}u \\
&= (1+\varepsilon)\int_0^1 (-\ln u)^{\varepsilon} u^{-1} \left[1 - \sum_{j=0}^{k-1} q_j(N,u)\right] \mathrm{d}u.
\end{aligned}
$$

Thus,

$$
I_1 < (1+\varepsilon)\int_0^1 (-\ln u)^{\varepsilon} u^{-1} \left[1 - \left(1 - \int_{v(x,R_N(u))} f(y)\,\mathrm{d}y\right)^{N-1}\right] \mathrm{d}u
$$

then as $\lim_{N\to\infty}(1 - \int_{v(x,R_N(u))} f(y)\,\mathrm{d}y)^{N-1} = \exp\{-f(x)\exp\{\psi(k)\}u\}$ and $1 - \mathrm{e}^{-x} < x$, for positive x, we have

$$
I_1 < (1+\varepsilon)\int_0^1 (-\ln u)^{\varepsilon} u^{-1} u f(x)\exp\{\psi(k)\}\,\mathrm{d}u = (1+\varepsilon) f(x)\exp\{\psi(k)\}\Gamma(1+\varepsilon) < \infty.
$$

Thus, we have proved equation (27) and thus equation (26).

To complete the proof, we need to show that as $N \to \infty$

$$
E\zeta_{i,k} = \int_{R^m} E(\zeta_{i,k}|X_i = x) f(x)\,\mathrm{d}x \longrightarrow \int_{R^m} (-\ln f(x)) f(x)\,\mathrm{d}x. \tag{28}
$$

From Fatou Lemma, we obtain

$$
\begin{aligned}
\lim_{N\to\infty}\sup \int_{R^m} |E(\zeta_{i,k}|X_i = x)|^{1+\varepsilon} f(x)\,\mathrm{d}x &\le \int_{R^m} \lim_{N\to\infty}\sup |E(\zeta_{i,k}|X_i = x)|^{1+\varepsilon} f(x)\,\mathrm{d}x \\
&= \int_{R^m} |\ln f(x)|^{1+\varepsilon} f(x)\,\mathrm{d}x < \infty.
\end{aligned}
$$

Thus equation (28) follows. This completes the proof of the Theorem 1.

6.2 *Proof of Theorem 2*

Note that passage to the limits in the proof of Theorem 2 can be done similar to that of the Theorem 1. Therefore, we shall omit the details.

Let

$$
\zeta(z) = \sum_{k=0}^{\infty} \frac{1}{k^z}, \quad \mathrm{Re}(z) > 1
$$

be a zeta-function of Riemann and

$$
\zeta(z,v) = \sum_{k=0}^{\infty} \frac{1}{(v+k)^z}, \quad \mathrm{Re}(z) > 1, \quad v \ne 0, -1, -2, \ldots
$$

the generalized Hurwitz zeta-function [35, p. 52].

Then $\zeta(2, k-1) = \zeta(z)$. It is obvious that

$$\mathrm{Var} H_{k,N} = \frac{\mathrm{Var}\zeta_{i,k}}{N} + 2\frac{\sum_{i<j} \mathrm{cov}(\zeta_{i,k}, \zeta_{j,k})}{N^2}. \tag{29}$$

As in the proof of Theorem 1 (for more detail see also the proof of Theorem 2 in ref. [24]), we obtain

$$\begin{aligned}
\lim_{N\to\infty} E(\zeta_{i,k}^2 | X_i = y) &= \int_0^\infty \ln^2 u W_y (u W_y)^{k-1} \frac{\mathrm{e}^{-W_y u}}{(k-1)!}\,\mathrm{d}u \\
&= \int_0^\infty \ln^2\ t t^{k-1} \frac{\mathrm{e}^{-t}}{\Gamma(k)}\,\mathrm{d}t + \ln^2\ W_y - 2\psi(k)\ln W_y \\
&= \zeta(2, k-1) + \ln^2\ f(y),
\end{aligned}$$

where $W_y = f(y)\mathrm{e}^{\psi(k)}$. Using Theorem 1, we obtain

$$\begin{aligned}
\lim_{N\to\infty} \mathrm{Var}\zeta_{i,k} &= \lim_{N\to\infty} (E\zeta_{i,k}^2 - (E\zeta_{i,k})^2) \\
&= \zeta(2, k-1) + I_1 - H^2 \\
&= \zeta(2) + I_1 - H^2,
\end{aligned}$$

where H is entropy defined by equation (1) and

$$I_1 = \int_{R^m} f(x) \ln^2(f(x))\,\mathrm{d}x.$$

Therefore, the first term in equation (29) goes to zero as $N \to \infty$, *i.e.*,

$$\lim_{N\to\infty} \frac{\mathrm{Var}\zeta_{i,k}}{N} = 0.$$

It is interesting to note that in the Gaussian case $I_1 - H^2 = m/2$. Thus, we only need to prove that

$$\mathrm{cov}(\zeta_{i,k}, \zeta_{j,k}) = E\zeta_{i,k}\zeta_{j,k} - (E\zeta_{i,k})^2 \longrightarrow 0, \tag{30}$$

as $N \to \infty$. From separability theorem, we know that there exist N_0 such that $x \neq y$ and $N > N_0$

$$v(x, R_N(u)) \cap v(y, R_N(w)) = \emptyset,$$

where $v(x, R_N(u))$ is defined in Theorem 1. Consequently, we obtain

$$\begin{aligned}
P\{\mathrm{e}^{\zeta_{i,k}} < u, \mathrm{e}^{\zeta_{j,k}} < w | X_i = x, X_j = y) &= 1 - \sum_{m=0}^{k-1} \binom{N-1}{m} I_u^m (1 - I_u)^{N-m-1} \\
&\quad - \sum_{m=0}^{k-1} \binom{N-1}{m} I_w^m (1 - I_w)^{N-m-1} \\
&\quad + \sum_{m=0}^{k-1} \sum_{l=0}^{k-1} \binom{N-1}{m} \binom{N-m-2}{l} \\
&\quad \times I_u^m I_w^l (1 - I_{u,w})^{N-m-l-2} \\
&=: 1 - \Sigma_1 - \Sigma_2 + \Sigma_3,
\end{aligned}$$

where

$$I_u = \int_{v(x,R_N(u))} f(z)\,\mathrm{d}z,$$
$$I_w = \int_{v(y,R_N(w))} f(z)\,\mathrm{d}z,$$
$$I_{u,w} = \int_{v(x,R_N(u))\cup v(y,R_N(w))} f(z)\,\mathrm{d}z = I_u + I_w,$$

for N is large enough ($N > N_0$).

Similar to the proof of Theorem 1, we obtain

$$\lim_{N\to\infty} \Sigma_1 = \mathrm{e}^{-uf(x)/\mathrm{e}^{-\psi(k)}} \sum_{m=0}^{k-1} \frac{(A_x u)^m}{m!},$$

where $A_x = f(x)\mathrm{e}^{\psi(k)}$,

$$\lim_{N\to\infty} \Sigma_2 = \mathrm{e}^{-wf(y)/\mathrm{e}^{-\psi(k)}} \sum_{m=0}^{k-1} \frac{(A_y w)^m}{m!},$$

where $A_y = f(y)\mathrm{e}^{\psi(k)}$ and

$$\begin{aligned}\lim_{N\to\infty} \Sigma_3 &= \lim_{N\to\infty} \sum_{m=0}^{k-1}\sum_{l=0}^{k-1}\left[\frac{(N-2)^{l+m}}{l!m!}\left(\frac{A_x u}{N-1}\right)^m\left(\frac{A_y w}{N-1}\right)^l \mathrm{e}^{-(uA_x+wA_y)(N-2-l-m)/N-1}\right]\\ &= \exp\left\{-\frac{(wf(y)+uf(x))}{\mathrm{e}^{-\psi(k)}}\right\}\sum_{l=0}^{k-1}\frac{(f(x)u)^l\mathrm{e}^{l\psi(k)}}{l!}\sum_{m=0}^{k-1}\frac{(f(y)w)^m\mathrm{e}^{m\psi(k)}}{m!}.\end{aligned}$$

Therefore,

$$\begin{aligned}\lim_{N\to\infty} P(\mathrm{e}^{\zeta_{i,k}} < u, \mathrm{e}^{\zeta_{j,k}} < w | X_i = x, X_j = y) &= 1 - \mathrm{e}^{uA_x}\sum_{m=0}^{k-1}\frac{(A_x u)^m}{m!}\\ &\quad - \mathrm{e}^{-uA_y}\sum_{n=0}^{k-1}\frac{(A_y w)^n}{n!} + \mathrm{e}^{-wA_y-uA_x}\\ &\quad + \sum_{n=0}^{k-1}\frac{(A_x u)^n}{n!}\sum_{m=0}^{k-1}\frac{(A_y w)^m}{m!}.\end{aligned}$$

Note that

$$\frac{\partial^2}{\partial u\partial w}\Sigma_1 = 0, \quad \frac{\partial^2}{\partial u\partial w}\Sigma_2 = 0, \quad \frac{\partial^2}{\partial u\partial w}\Sigma_3 = e^{-(wA_y+uA_x)}A_xA_y\frac{(A_xu)^{k-1}(A_yw)^{k-1}}{((k-1)!)^2}.$$

Now,

$$\begin{aligned}\lim_{N\to\infty} E\{(\zeta_{i,k}\zeta_{j,k}|X_i = x, X_j = y)\} &= \int_{R^m}\int_{R^m}\frac{\ln u \ln w A_xA_y(A_xu)^{k-1}(A_yw)^{k-1}}{\mathrm{e}^{(wA_y-uA_x)}((k-1)!)^2}\,\mathrm{d}u\,\mathrm{d}w\\ &= \ln f(x)\ln f(y).\end{aligned}$$

Thus,

$$\lim_{N\to\infty} E(\zeta_{i,k}\zeta_{j,k}) = \int_{R^m}\int_{R^m} f(x)f(y)\ln f(x)\ln f(y)\,\mathrm{d}x\,\mathrm{d}y = (\lim_{N\to\infty} E\zeta_{i,k})^2.$$

This proves equation (30) and consequently completes the proof of Theorem 2.

Acknowledgements

The comments and suggestions of an anonymous Associated Editor and referee together with those of the Editor have been gratefully appreciated and allowed us to improve the final version of this paper. The second author was partially supported by the ARC grant DP 0345577 and by the EPSRC grant RCMT091.

References

[1] Dobrushin, R.L., 1958, Simplified method of experimental estimate of entropy of stationary sequence. *Theory of Probability and its Application*, **3**, 462–464. (Russian)
[2] Basharin, G.P., 1959, On statistical estimation of entropy of sequence of independent random variables. *Theory of Probability and its Application*, **4**, 333–336.
[3] Zubkov, A.M., 1959, Limit distributions of statistical estimation of entropy. *Theory of Probability and its Application*, **18**, 643–650. (Russian)
[4] Hutcheston, K. and Shenton, D.R., 1974, Some moments of an estimate of Shannon's measure of information. *Communications in Statistics*, **3**, 89–94.
[5] Stein, C., 1977, Lectures on theory of estimation of many parameters. In: *Isseledovanija po statisticheskoi theorii ocenivania*, Vol. 1 (Leningrad: Nauka). (Russian)
[6] El-Bassiouny, A.X., 1985, On convergence of functionals of statistical estimate of entropy in polynomial schemes. *Theoritical Probability in Mathematical Statistics*, **33**, 113–120. (Russian)
[7] Vatutin, V.A. and Michailov, V.G., 1995, Statistical estimation of entropy of discrete random variables with large numbers of results. *Uspeh. Math. Nauk*, **50**, 121–134. (Russian)
[8] Ahmad, I.A. and Lin, P.E., 1976, A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Transaction Information Theory*, **IT-22**, 372–375.
[9] Vasicek, O., 1976, A test of normality based on sample entropy. *Journal of the Royal Statistical Society Series B*, **38**, 254–256.
[10] Joe, H., 1989, Relative entropy measures of multivariate dependence. *Journal of American Statistical Association*, **84**, 1217–1225.
[11] Joe, H., 1989, Estimation of entropy and other functionals of a multivariate density. *Annals of the Institute of Statistical Mathematics*, **41**, 683–697.
[12] Hall, P. and Morton, S.C., 1993, On the estimation of entropy. *Annals of the Institute of Statistical Mathematics*, **45**, 69–88.
[13] van Es, B., 1992, Estimating functionals related to a density by class of statistics based on spacing. *Scandinavian Journal of Statistics*, **19**, 61–72.
[14] Tsybakov, A.B. and van der Meulen, E.C., 1996, Root-n consistent estimators of entropy for densities with unbounded support. *Scandinavian Journal of Statistics*, **23**, 75–83.
[15] Kapur, J.N., 1989, *Maximum Entropy Models in Science and Engineering* (New York: Wiley).
[16] Kozachenko, L.F., 1990, On entropy based tests of goodness of fit for distributions of random variables. *Issledov. Oper. and ASU*, **36**, 7–11. (Russian)
[17] Tagliani, A., 1993, On the application of maximum entropy to the moment problem. *Journal of Mathematical Physics*, **34**, 326–337.
[18] Frontini, M.N. and Tagliani, A., 1997, Entropy-convergence in Stieltjes and Hamburger moment problem. *Applied Mathematics and Computation*, 3, **88**, 39–51.
[19] Dudewicz, E.I. and van der Meulen, E.C., 1981, Entropy based test of uniformity. *Journal of American Statistical Association*, **76**, 974–987.
[20] Ivanov, A.V. and Rojkova, M.N., 1981, On properties of statistical estimation of entropy of random vector having probability density. *Problems Infor. Transmiss.*, **17**, 34–43. (Russian)
[21] Gangopadhyay, A.K., Disario, R. and Dey, D.K., 1997, A nonparametric approach to k-sample inference based on entropy. *Nonparametric Statistics*, **8**, 234–252.
[22] Eggermont, P.P.B. and La Riccia, V.N., 1999, Best asymptotical normality of the kernel density entropy estimator for smooth densities. *IEEE Transaction on Information Theory*, **45**, 1321–1326.
[23] Beirlant, J., Dudewicz, E., Gyorfi, L. and van der Meulen, E.C., 1997, Nonparametric entropy estimation: an overview. *International Journal of Mathematical Statistical Science*, **6**(1), 17–39.

[24] Kozachenko, L.F. and Leonenko N.N., 1987, On statistical estimation of entropy of random vector. *Problems Information Transmission*, **23**, 9–16.
[25] Rao, C.R., 1973, *Linear Statistical Methods and Its Applications* (Nauka: Moskow).
[26] Kagan, A.M., Linnik, Yu.V. and Rao, C.B., 1973, *Characterization Problems in Mathematical Statistics* (New York: Wiley).
[27] Kullback, S., 1959, *Information Theory and Statistics* (New York: Wiley).
[28] Kullback, S., 1968, Probability densties with given marginals. *Annals of Mathematical Statistics*, **39**, 1236–1246.
[29] Joe, H., 1987, Majorization, randomness and dependence for multivariate distributions. *The Annals of Probability*, **15**, 1217–1225.
[30] Grzegorzewski, P. and Wieczorkowski, R., 1999, Entropy-based goodness-of-fit test for exponentiality. *Communications in Statistics. Theory and Methods*, **28**, 1183–1202.
[31] Kotz, S., Kozubowski, T.J. and Podgrski, K., 2002, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering and Finance* (Boston: Birkhauser).
[32] Ayebo, A. and Kozubowski, T.J., 2003, An asymmetric generalization of Gaussian and Laplace laws. *Journal of Probability and Statistical Science*, in press.
[33] Lebesgue, H., 1910, Sur l'integration des functions discontinues. *Ann. Ecole. Norm.*, **27**, 361–450.
[34] Shorack, G.R. and Wellener, J.A., 1986, *Emprical Processes with Application to Statistics* (New York: Wiley).
[35] Margulis, V., 1965, *Handbook of Series* (New York: Academic Press).
[36] Loéve, M., 1962, *Theory Probability* (Moscow).
[37] Shapiro, S.S., Wilk, M.B. and Chen, H.J., 1968, A comparative study of various test of normality. *Journal of American Statistical Association*, **63**, 1343–1372.