Nonparametric entropy estimation: an overview *

J. Beirlant
Dep. of Mathematics
Kath. Univ. Leuven
Celestijnenlaan 200B
B-3001 Heverlee
Belgium

E. J. Dudewicz
Dep. of Mathematics
Syracuse Univ.
Syracuse
NY 13244-1150
USA

L. Györfi
Dep. of Mathematics
Tech. Univ. of Budapest
Stoczek u.2
H-1521 Budapest
Hungary

E. C. van der Meulen
Dep. of Mathematics
Kath. Univ. Leuven
Celestijnenlaan 200B
B-3001 Heverlee
Belgium

July 17, 2001

Dedicated to

Professor L. L. Campbell, in celebration of his longstanding career in mathematics and statistics and in tribute to his many scholarly contributions to information theory.

Abstract

An overview is given of the several methods in use for the nonparametric estimation of the differential entropy of a continuous random variable. The properties of various methods are compared. Several applications are given such as tests for goodness-of-fit, parameter estimation, quantization theory and spectral estimation.

I Introduction

Let X be a random vector taking values in R^d with probability density function (pdf) $f(x)$, then its differential entropy is defined by

$$H(f) = -\int f(x) \ln f(x) dx. \tag{1}$$

We assume that $H(f)$ is well-defined and is finite.

The concept of differential entropy was introduced in Shannon's original paper ([55]). Since then, entropy has been of great theoretical and applied interest. The basic properties

*This research was supported by the Scientific Exchange Program between the Belgian Academy of Sciences and the Hungarian Academy of Sciences in the field of Mathematical Information Theory, and NATO Research Grant No. CRG 931030.

of differential entropy are described in Chapter 9 of Cover and Thomas [10]. Verdugo Lazo and Rathie [64] provide a useful list containing the explicit expression of $H(f)$ for many common univariate pdf's. Ahmed and Gokhale [1] calculated $H(f)$ for various multivariate pdf's.

The differential entropy has some important extremal properties:
(I) If the density f is concentrated on the unit interval $[0, 1]$ then the differential entropy is maximal iff f is uniform on $[0, 1]$, and then $H(f) = 0$.
(II) If the density is concentrated on the positive half line and has a fixed expectation then the differential entropy takes its maximum for the exponential distribution.
(III) If the density has fixed variance then the differential entropy is maximized by the Gaussian density.

Many distributions in statistics can be characterized as having maximum entropy. For a general characterization theorem see [38].

II Estimates

II.1 Criteria and conditions

If for the i.i.d. sample $X_1, \ldots X_n$ H_n is an estimate of $H(f)$ then we consider the following types of consistencies:
Weak consistency:

$$\lim_{n\to\infty} H_n = H(f) \text{ in probability.} \tag{2}$$

Mean square consistency:

$$\lim_{n\to\infty} E\{(H_n - H(f))^2\} = 0. \tag{3}$$

Strong consistency:

$$\lim_{n\to\infty} H_n = H(f) \text{ a.s.} \tag{4}$$

Root-n consistency results are either of form of **asymptotic normality**, i.e.,

$$\lim_{n\to\infty} n^{1/2}(H_n - H(f)) = N(0, \sigma^2) \tag{5}$$

in distribution, or L_2 **rate of convergence**:

$$\lim_{n\to\infty} nE\{(H_n - H(f))^2\} = \sigma^2. \tag{6}$$

Next we list the usual conditions on the underlying density.
Smoothness conditions:

S1: f is continuous.
S2: f is k times differentiable.
Tail conditions:
T1: $H([X]) < \infty$, where $[X]$ is the integer part of X.
T2: $\inf_{f(x)>0} f(x) > 0$.
Peak conditions:
P1: $\int f(\ln f)^2 < \infty$. (This is also a mild tail condition.)
P2: f is bounded.

II.2 Plug-in estimates

The plug-in estimates of entropy are based on a consistent density estimate f_n of f such that f_n depends on $X_1, \dots X_n$.

(i) We first consider the **integral estimate** of entropy, which is of the form

$$H_n = -\int_{A_n} f_n(x) \ln f_n(x) dx, \tag{7}$$

where, with the set A_n one typically excludes the small or tail values of f_n. The first such estimator was introduced by Dmitriev and Tarasenko [17], who proposed to estimate $H(f)$ by (7) for $d = 1$, where $A_n = [-b_n, b_n]$ and f_n is the kernel density estimator. They showed the strong consistency of H_n defined by (7). See also Prakasa Rao [49]. Mokkadem [44] calculated the expected L_r error of this estimate.

The evaluation of the integral in (7) however requires numerical integration and is not easy if f_n is a kernel density estimator. Joe [37] considers estimating $H(f)$ by (7) when f is a multivariate pdf, but he points out that the calculation of (7) when f_n is a kernel estimator gets more difficult for $d \geq 2$. He therefore excludes the integral estimate from further study.

The integral estimator can however be easily calculated if, for example, f_n is a histogram. This approach is taken by Györfi and van der Meulen [29]. If $\{x \in A_n\} = \{f_n(x) \geq a_n\}$ with $0 < a_n \to 0$, then the strong consistency has been proved under the only condition T1. Carbonez, et al. [7] extended this approach to the estimation of $H(f)$ when the observations are censored.

(ii) The **resubstitution estimate** is of the form

$$H_n = -\frac{1}{n} \sum_{i=1}^{n} \ln f_n(X_i). \tag{8}$$

Ahmad and Lin [2] proposed estimating $H(f)$ by (8), where f_n is a kernel density estimate. They showed the mean square consistency of (8). Joe [37] considered the estimation of $H(f)$

for multivariate pdf's by an entropy estimate of the resubstitution type (8), also based on a kernel density estimate. He obtained asymptotic bias and variance terms, and showed that non-unimodal kernels satisfying certain conditions can reduce the mean square error. His analysis and simulations suggest that the sample size needed for good estimates increases rapidly with the dimension d of the multivariate density. His results rely heavily on conditions T2 and P2. Hall and Morton [35] investigated the properties of an estimator of the type (8) both when f_n is a histogram density estimator and when it is a kernel estimator. For histogram they show (5) under certain tail and smoothness conditions with

$$\sigma^2 = Var(\ln f(X)). \tag{9}$$

They point out that the histogram-based estimator can only be root-n consistent when $d = 1$ or 2, and conclude it is particularly attractive for $d = 1$, since when $d = 2$ any root-n consistent histogram-based estimator of entropy will have significant bias. They suggest an empirical rule for the binwidth, using a penality term. They study the effects of tail behaviour, distribution smoothness and dimensionality on convergence properties, and argue that root-n consistency of entropy estimation requires appropriate assumptions about each of these three features. Their results are valid for a wide class of densities f having unbounded support. They also suggest an application to projection pursuit.

(iii) The next plug-in estimate is the **splitting data estimate**. Here one decomposes the sample into two sub-samples: $X_1, \dots X_l$ and $X_1^*, \dots X_m^*$, $n = l + m$. Based on $X_1, \dots X_l$ one constructs a density estimate f_l, and then, using this density estimate and the second sample, estimates $H(f)$ by

$$H_n = -\frac{1}{m} \sum_{i=1}^{m} I_{[X_i^* \in A_l]} \ln f_l(X_i^*). \tag{10}$$

Györfi and van der Meulen used this approach in [29] for f_l being the histogram density estimate, in [30] for f_l being the kernel density estimate, and in [31] for f_l being any L_1-consistent density estimate such that $[X_i^* \in A_l] = [f_l(X_i^*) \geq a_l]$, $0 < a_l \to 0$. Under some mild tail and smoothness conditions on f the strong consistency is shown for general dimension d.

(iv) The final plug-in estimate is based on a **cross-validation estimate** or leave-one-out density estimate. If $f_{n,i}$ denotes a density estimate based on $X_1, \dots X_n$ leaving X_i out, then the corresponding density estimate is of the form

$$H_n = -\frac{1}{n} \sum_{i=1}^{n} I_{[X_i \in A_n]} \ln f_{n,i}(X_i). \tag{11}$$

Ivanov and Rozhkova [36] proposed such entropy estimate when $f_{n,i}$ is a kernel estimate. They showed strong consistency, and also made an assertion regarding the rate of convergence of the moments $E|H_n - H(f)|^r$, $r \geq 1$. Hall and Morton [35] also studied entropy estimates of the type (11) based on kernel estimator. For $d = 1$, properties of H_n were studied in the context of Kullback-Leibler loss in [34]. Under some conditions the analysis in [35] yields a root-n consistent estimate of the entropy when $1 \leq d \leq 3$. The authors point out that the case $d = 2$ is of practical interest in projection pursuit.

II.3 Estimates of entropy based on sample-spacings.

Since sample-spacings are defined only for $d = 1$, we assume that $X_1, \ldots, X_n$ is a sample of i.i.d. real valued random variables. Let $X_{n,1} \leq X_{n,2} \leq \ldots \leq X_{n,n}$ be the corresponding order statistics. Then $X_{n,i+m} - X_{n,i}$ is called a spacing of order m, or m-spacing ($1 \leq i < i+m \leq n$).

Based on spacings it is possible to construct a density estimate:

$$f_n(x) = \frac{m}{n} \frac{1}{X_{n,im} - X_{n,(i-1)m}}$$

if $x \in [X_{n,(i-1)m}, X_{n,im})$. This density estimate is consistent if, as $n \to \infty$,

$$m_n \to \infty, \, m_n/n \to 0. \tag{12}$$

The estimate of entropy based on sample-spacings can be derived as a plug-in integral estimate using a spacing density estimate. However, surprisingly one can get a consistent spacings based entropy estimate from a non-consistent spacings density estimate, too.
(i) We consider first the m-**spacing estimate** for fixed m:

$$H_{m,n} = \frac{1}{n} \sum_{i=1}^{n-m} \ln\left(\frac{n}{m}(X_{n,i+m} - X_{n,i})\right) - \psi(m) + \ln m, \tag{13}$$

where $\psi(x) = -(\ln \Gamma(x))'$ is the digamma function. Then the corresponding density estimate is not consistent. This implies that in (13) there is an additional term correcting the asymptotic bias. For uniform f the consistency of (13) has been proved by Tarasenko [57] and Beirlant and van Zuijlen [4]. Hall [32] proved the weak consistency of (13) for densities satisfying T2 and P2. Under the conditions T2 and P2 the asymptotic normality of $H_{m,n}$ in form of (5) was studied by Cressie [11], Dudewicz and van der Meulen [19], Hall [32] and Beirlant [3], who all proved (5) under T2 and P2 with

$$\sigma^2 = (2m^2 - 2m + 1)\psi'(m) - 2m + 1 + Var(\ln f(X)). \tag{14}$$

For $m = 1$ (14) yields

$$\sigma^2 = \frac{\pi^2}{6} - 1 + Var(\ln f(X)). \tag{15}$$

Dudewicz and van der Meulen [21] proposed to estimate $H(f)$ by $H(\hat{f})$, where $\hat{f}$ is the empiric pdf, derived from a smoothed version of the empirical distribution function. This led to the introduction of the notion of empiric entropy of order m.

(ii) In order to decrease the asymptotic variance consider next the **m_n-spacing estimate** with $m_n \to \infty$:

$$H_n = \frac{1}{n}\sum_{i=1}^{n-m_n} \ln\left(\frac{n}{m_n}(X_{n,i+m_n} - X_{n,i})\right). \tag{16}$$

This case is considered in the papers of Vasicek [63], Dudewicz and van der Meulen [19], Beirlant and van Zuijlen [4], Beirlant [3], Hall [33] and van Es [62]. In these papers the weak and strong consistencies are proved under (12). Consistencies for densities with unbounded support is proved only in [57] and [4]. Hall [33] and Van Es [62] proved (5) with (9) if f is not uniform but satisfies T2 and P2. Hall [33] showed this result also for the non-consistent choice $m_n/n \to \rho$ if ρ is irrational. This asymptotic variance is the smallest one for an entropy estimator if f is not uniform (cf. Levit [40]). If f is uniform on $[0,1]$ then Dudewicz and van der Meulen [19] and van Es [62] showed, respectively for $m_n = o(n^{1/3-\delta}), \delta > 0$, and for $m_n = o(n^{1/3})$ that

$$\lim_{n\to\infty}(mn)^{1/2}(\hat{H}_n - H(f)) = N(0, 1/3), \tag{17}$$

for slight modifications $\hat{H}_n$ of the m_n-spacing estimate H_n.

II.4 Estimates of entropy based on nearest neighbor distances

The nearest neighbor estimate is defined for general d. Let $\rho_{n,i}$ be the nearest neighbor distance of X_i and the other X_j: $\rho_{n,i} = \min_{j\neq i, j\leq n}\|X_i - X_j\|$. Then the **nearest neighbor estimate** is

$$H_n = \frac{1}{n}\sum_{i=1}^{n} \ln(n\rho_{n,i}) + \ln 2 + C_E, \tag{18}$$

where C_E is the Euler constant: $C_E = -\int_0^\infty e^{-t}\ln t dt$. Under some mild conditions like P1 Kozachenko and Leonenko [39] proved the mean square consistency for general d. Tsybakov and van der Meulen [61] showed root-n rate of convergence for a truncated version of H_n when $d = 1$ for a class of densities with unbounded support and exponentially decreasing tails, such as the Gaussian density. Bickel and Breiman [6] considered estimating a general functional of a density. Under general conditions on f they proved (5). Unfortunately their study excludes the entropy.

III Applications

III.1 Entropy-based tests for goodness-of-fit

Moran [45] was the first to use a test statistic based on 1-spacings for testing the goodness-of-fit hypothesis of uniformity. Moran's statistic is defined by

$$M_n = -\sum_{i=0}^{n} \ln((n+1)(X_{n,i+1} - X_{n,i})), \tag{19}$$

with $X_{n,0} = 0$ and $X_{n,n+1} = 1$. Darling [16] showed that under $H_0 : F(x) = x$, $0 \le x \le 1$, one gets

$$\lim_{n\to\infty} n^{-1/2}(M_n - nC_E) = N(0, \frac{\pi^2}{6} - 1). \tag{20}$$

(This limit law is (5) with (15) for f uniform.)

Cressie [11] generalized the notion of 1-spacings to m_n-spacings, and considered the testing of uniformity. He showed the asymptotic normality of his test statistic

$$L_n = \sum_{i=1}^{n-m_n+1} \ln(X_{n,i+m_n} - X_{n,i}) \tag{21}$$

under the hypothesis of uniformity (see Sec. II.3 for more details) and under the assumption that f is a bounded positive step function on $[0,1]$, and stated this fact, as a generalization, to hold if f is concentrated on $[0,1]$, satisfies conditions T2 and P2, and has a finite number of discontinuities. From this it follows that the test based on rejecting the hypothesis of uniformity for large negative values of L_n will be consistent. Cressie [11] compared his test procedure with Greenwood's ([27]), for which the test statistic is based on the sum of squares of 1-spacings, in terms of Pitman asymptotic relative efficiency, for a sequence of neighboring alternatives which converge to the hypotheses of uniformity as $n \to \infty$. Cressie continued his investigations in [12], [13].

¿From a different point of view, and independently, Vasicek [63] introduced his test statistic

$$\hat{H}_n = \frac{1}{n}\sum_{i=1}^{n} \ln(\frac{n}{2m_n}(X_{n,i+m_n} - X_{n,i-m_n})), \tag{22}$$

where $X_{n,j} = X_{n,n}$ for $j > n$ and $X_{n,j} = X_{n,1}$ for $j < 1$. His test statistic is very close to the $2m_n$-spacing entropy estimate (16), and thus to Cressie's statistic (21). Vasicek [63] proved that $\hat{H}_n$ is a weak consistent estimate of $H(f)$. Making use of the fact that for fixed variance, entropy is maximized by the normal distribution, he proposed an entropy based test for the composite hypothesis of normality, which rejects the null hypothesis if $e^{\hat{H}_n}/S < C = e^{-1.42}$, where S denotes the sample standard deviation. Vasicek [63] provided critical values of his test statistic for various values of m and n based on Monte Carlo simulation and also simulated its power against various alternatives. He found, that compared with other tests for normality, the entropy-based test performs well against various alternatives. Prescott ([50]), however, raised some questions regarding the protection it provides against heavy-tailed alternatives. Mudholkar and Lin [46] extended Vasicek's power simulations for testing normality and concluded that Vasicek's test provides strong protection against light-tailed alternatives.

Using the fact that the uniform distribution maximizes the entropy on $[0,1]$, Dudewicz and van der Meulen [19] extended Vasicek's reasoning and proposed an entropy-based test for uniformity which rejects the null hypothesis if $\hat{H}_n \leq H^*_{\alpha,m,n}$, where $H^*_{\alpha,m,n}$ is set so that the test has level α for given m and n. Dudewicz and van der Meulen [19] give Monte Carlo estimates of $H^*_{\alpha,m,n}$ for specific α, m and n and carry out a Monte Carlo study of the power of their entropy-based test under seven alternatives. Their simulation studies show that the entropy-based test of uniformity performs particularly well for alternatives which are peaked up near 0.5, as compared with other tests of uniformity. Their simulation studies were continued in [22], [23].

Mudholkar and Lin [46] applied the Vasicek logic to develop an entropy-based test for exponentiality, using the fact that among all distributions with given mean and support $(0,\infty)$, the exponential distribution with specified mean has maximum entropy. Their test for exponentiality rejects if $e^{\hat{H}_n}/\bar{X} < e$. They showed consistency of their test and evaluated the test procedure empirically by a Monte Carlo study, for various values of m and n. Their studies conclude that this entropy-based of exponentiality has reasonably good power properties. Following the outline in [19], Gokhale [26] formalized the entropy-based test construction for a broad class of distributions.

As stated in Sec II.3 the asymptotic distribution of m_n-spacing entropy estimate has been shown to be normal under conditions T2 and P2 (including uniformity, excluding the gaussian and the exponential distribution). Since this convergence to normality is very slow Mudholkar and Smethurst [47] developed transformation-based methods to accelerate this convergence. Chaubey, Mudholkar and Smethurst [8] proposed a method for avoiding problems in approximating the null distributions of entropy-based test statistics by constructing a jackknife statistic which is asymptotically normal and may be approximated in moderate sized samples by a scaled t-distribution.

Robinson [53] proposed a test for independence based on an estimate of the Kullback-Leibler divergence, which is closely related to differential entropy. His estimate is of the integral (and resubstitution) type. He shows consistency and focuses on testing serial independence for time series.

Parzen [48] considers entropy-based statistics such as (19) and (22) to test the goodness-of-fit of a parametric model $\{F(x,\theta)\}$.

III.2 Entropy-based parameter estimation

Cheng and Amin [9], Ranneby [51] and Shao and Hahn [56] proposed a consistent parameter estimate, where the maximum likelihood estimate is not consistent. Consider a parametric family of distribution function of real variable

$$\{F(x,\theta)\}.$$

Assume that the distribution function of X is $F(x,\theta_0)$, where θ_0 is the true parameter. The transformed variable

$$X^* = F(X,\theta)$$

is $[0,1]$-valued, and its distribution is uniform on $[0,1]$ iff $\theta = \theta_0$, i.e. its differential entropy is maximal iff $\theta = \theta_0$. Thus it is reasonable to maximize an estimate of the differential entropy. If, for example, the 1-spacings estimate is used then the estimate of the parameter is

$$\theta_n = argmax_\theta \sum_{i=1}^{n-1} \ln(F(X_{n,i+1},\theta) - F(X_{n,i},\theta)).$$

This estimator is referred to as maximum spacing estimator.

III.3 Differential entropy and quantization

The notion of differential entropy is intimately connected with variable-length lossy source coding. Here the actual coding rate is the average codelength of a binary prefix code for the source codewords, but, for the sake of simplicity, it is usually approximated by the (Shannon) entropy of the source coder output. It turns out that for large rates, the entropy of the encoder output can be expressed through the differential entropy of the source.

Let X be a real random variable with density f and finite differential entropy $H(f)$, and assume that the discrete random variable $[X]$ (the integer part of X) has finite entropy. With these very general conditions Rényi [52] proved that the entropy of the sequence $[nX]$ behaves as

$$\lim_{n\to\infty} (H([nX]) - \log n) = H(f) = -\int f(x)\log f(x)dx, \tag{23}$$

where log stands for the logarithm of base 2. This shows that for large n the entropy of the uniform quantizer with step-size $1/n$ is well approximated by $H(f) + \log n$. Using their own (slightly less general) version of the above statement, Gish and Pierce [25] showed that for the squared distortion $D(Q_\Delta)$, and for the entropy $H(Q_\Delta)$, of the Δ step-size uniform scalar quantizer, the following asymptotics holds:

$$\lim_{\Delta\to 0}\left[H(Q_\Delta)+\frac{1}{2}\log(12D(Q_\Delta))\right]=H(f). \tag{24}$$

Here f is assumed to satisfy some smoothness and tail conditions. Again, the differential entropy provides the rule of thumb $D(Q_\Delta)\approx(1/12)2^{-2[H(Q_\Delta)-H(f)]}$ for small Δ. (24) can be proved without any additional smoothness and tail conditions (Györfi, Linder, van der Meulen [28]). The rate and distortion tradeoff depends on the source density only through the differential entropy. A heuristic generalization of (24) was given by Gersho [24]. For k-dimensional vector quantizers and large rates R, he derived the formula

$$D(Q_R)\approx c_k(P)2^{-(2/k)[R-H(f)]}, \tag{25}$$

where Q_R is a vector quantizer of entropy R having quantization cells congruent to a polytope P, and $c_k(P)$ is the normalized second moment of P.

Csiszár [14], [15] investigated the entropy of countable partitions in an arbitrary measure space. He obtained the following strengthening of Rényi's result (23): Let $X=(X_1,\dots,X_k)$ be an $\mathcal{R}^k$ valued random vector with differential entropy $H(f)$, and assume that $[X]=([X_1],\dots,[X_k])$ has finite entropy. Then considering partitions $\mathcal{A}$ of $\mathcal{R}^k$ into Borel measurable sets of equal Lebesgue measure $\epsilon(\mathcal{A})$ and maximum diameter $\delta(\mathcal{A})$, we have

$$\lim_{\delta(\mathcal{A})\to 0}(H_{\mathcal{A}}(X)+\log\epsilon(\mathcal{A}))=H(f),$$

where $H_{\mathcal{A}}(X)$ is the entropy of the partition $\mathcal{A}$ with respect to the probability measure induced by X. Using this result, Linder and Zeger [42] gave a rigorous proof of (25) for sources with a finite differential entropy.

Differential entropy is also the only factor through which the Shannon lower bound to the rate-distortion function depends on the source density. For squared distortion the Shannon lower bound is

$$R_k(D)\geq\frac{1}{k}H(f_k)-\frac{1}{2}\log(2\pi eD),$$

where $R_k(D)$ is the kth order rate-distortion function of the source having density f_k. For stationary sources, when the differential entropy rate $H=\lim_{k\to\infty}(1/k)H(f_k)$ exists, the above lower bound also holds for all $D>0$ when $(1/k)H(f_k)$ is replaced by H and $R_k(D)$ is by $R(D)=\lim_{k\to\infty}R_k(D)$, the rate distortion function of the source. Both bounds are asymptotically (when $D\to 0$) tight, as was shown by Linkov [43] and by Linder and Zamir [41], which makes them extremely useful for relating asymptotic quantizer performances to rate-distortion limits.

III.4 Applications towards econometrics, spectroscopy and statistical simulations

Theil [58] evaluated the entropy (1) of a density estimate which is fitted to the data by a maximum entropy principle. Although he did not so note the entropy of the fitted pdf can be regarded as an estimate of the entropy $H(f)$ of the underlying pdf f. Theil's estimate turns out to be equal to Vasicek's estimate (22) for $m = 1$, apart from a constant. Theil and Laitinen [60] applied the method of maximum entropy estimation of pdf's in econometrics. Their work is at the beginning of quite a few papers on this topic in this field. For an overview of this maximum entropy approach for the estimation of pdf's and the application of it to problems in econometrics see Theil and Fiebig [59].

Rodrigez and Van Ryzin [54] studied the large sample properties of a class of histogram estimators whose derivation is based on the maximum entropy principle.

Entropy principles play also a key role in spectroscopy and image analysis. In practice entropy is not known, but must be estimated from data. For an overview of the available maximum entropy methods in spectroscopy see [18].

In [20] a method is developed to rank and compare random number generators based on their entropy-values. Here the entropy of the random number generator is estimated using (22). A detailed numerical study, ranking 13 random number generators on basis of their entropy estimate, is carried out in [5].

References

[1] Ahmed, N. A. and Ghokale, D. V. Entropy expressions and their estimators for multivariate distribution. *IEEE Trans. Information Theory*, 35, 688-692, 1989.

[2] Ahmad, I. A. and Lin, P. E. A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Trans. Information Theory*, 22, 372-375, 1976.

[3] Beirlant, J. Limit theory for spacing statistics from general univariate distributions. *Pub. Inst. Stat. Univ. Paris XXXI* fasc 1, 27-57, 1986.

[4] Beirlant, J. and Zuijlen, M. C. A. The empirical distribution function and strong laws for functions of order statistics of uniform spacings. *J. Multivar. Anal.* 16, 300-317, 1985.

[5] Bernhofen, L. T., Dudewicz, E.J., Levendovszky, J. and van der Meulen, E.C. Ranking of the best random number generators via entropy-uiformity theory. In *MSI-2000: Multivariate Statistical Analysis in Honor of Professor Minoru Siotani on his 70th Birthday, Vol. II*, Ed. H Hayakawa, M; Aoshima and K. Shimizu, American Sciences Press Inc., Columbus, Ohio, 1996, in press.

[6] Bickel, P. J. and Breiman L. Sums of functions of nearrest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Annals of Statist.*, 11, 185-214, 1983.

[7] Carbonez, A., Györfi, L. and van der Meulen, E.C. Nonparametric entropy estimation based on randomly censored data. *Problems of Control and Inforrmation Theory*, 20, 441-451, 1991.

[8] Chaubey, Y. P., Mudholkar, G. S. and Smethurst, P. A. On entropy-based goodness-of-fit tests: a practical strategy. in *Probability and Statistics*, S. K. Basu and B. K. Sinha, eds., Narosa Publishing House, New Delhi, 116-120, 1993.

[9] Cheng, R. C. H. and Amin, N. A. K. Estimating parameters in continuous univariate distributions with a shifted origin. *J. R. Statist. Soc. B*, 45, 394-403, 1983.

[10] Cover, T. M. and Thomas J. A. *Elements of Information Theory.* Wiley, 1991.

[11] Cressie, N. On the logarithms of higher order spacings. *Biometrika*, 63, 343-355, 1976.

[12] Cressie, N. The minimum of higher order gaps. *Austral. J. Statist.*, 19, 132-143, 1977.

[13] Cressie, N. Power results for tests based on higher order gaps. *Biometrika*, 65, 214-218, 1978.

[14] Csiszár, I. On generalized entropy. *Studia Sci. Math. Hungar*, 4, 401-419, 1969.

[15] Csiszár, I. Generalized entropy and quantization problems. *Transactions of the Sixth Praque Conference on Information Theory, Statistical Decision Functions, Random Processes* (Academia, Prague), 1973.

[16] Darling, D. A. On a class of problems relating to the random division of an interval. *Ann. Math. Stat.*, 24, 239-253, 1953.

[17] Dmitriev, Yu.G. and Tarasenko, F.P. On the estimation functions of the probability density and its derivatives, *Theory Probab. Appl.* 18, 628-633, 1973.

[18] Dudewicz, E.J., Mommaerts, W. and van der Meulen, E.C. Maximum entropy methods in modern spectroscopy: a review and an empiric eentropy approach. In *The Frontiers of Statistical Scirntific Theory and Industrial Applications*, Eds. A. Öztrük and E. C. van der Meulen, American Sciences Press, Inc., Columbus, Ohio, 115-160,1991.

[19] Dudewicz, E.J. and van der Meulen, E.C. Entropy-based tests of uniformity. *J. Amer. Statist. Assoc.* 76, 967-974, 1981.

[20] Dudewicz, E.J. and van der Meulen, E.C. Entropy-based statistical inference, II: Selection-of-the-best/complete ranking for continuous distributions on (0,1), with applications to random number generator. *Statistics and Decisions*, 1, 131-145, 1983.

[21] Dudewicz, E.J. and van der Meulen, E.C. The empiric entropy, a new approach to nonparametric entropy estimation. In *New Perspectives in Theoretical and Applied Statistics* Eds. M. L. Puri, J. Vilaplana and W. Wertz, Wiley, 207-227, 1987.

[22] Dudewicz, E.J. and van der Meulen, E.C. On the power of entropy-based tests against bump-type alternatives. In *Contributions to Stochastics*, Ed. W. Sendler, Physica-Verlag, 84-88, 1987.

[23] Dudewicz, E.J. and van der Meulen, E.C. Entropy-based random number evaluation. *American J. Mathematical and Management Sciences*, 15, 115-153, 1995.

[24] A. Gersho, Asymptotically optimal block quantization. *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.

[25] H. Gish and J. N. Pierce, Asymptotically efficient quantizing. *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, September 1968.

[26] Gokhale, D. V. On entropy-based goodness-of-fit test. *Computational Statistics and Data Analysis*, 1, 157-165, 1983.

[27] Greeenwood M. The statistical study of infectious diseases. *J. Royal Statist. Soc. A.*, 109, 85-110, 1946.

[28] Györfi, L., Linder, T. and van der Meulen, E.C. On the asymptotic optimality of quantizers. *Proceedings of the Eleventh Symposium on Information Theory in the Benelux*, Noordwijkerhout, The Netherlands, October 25-26, 1990, pp. 29-35.

[29] Györfi, L. and van der Meulen, E.C. Density-free convergence properties of various estimators of entropy. *Comput. Statist. Data Anal.*, 5, 425-436, 1987.

[30] Györfi, L. and van der Meulen, E.C. An entropy estimate based on a kernel density estimation. in *Limit Theorems in Probability and Statistics*, Eds. I. Berkes, E. Csáki, P. Révész, North Holland, 229-240, 1989.

[31] Györfi, L. and van der Meulen, E.C. On nonparametric estimation of entropy functionals. in *Nonparametric Functional Estimation and Related Topics* Ed. G. Roussas, Kluwer Academic Publisher, 81-95, 1990.

[32] Hall, P. Limit theorems for sums of general functions of m-spacings. *Math. Proc. Camb. Phil. Soc*, 96, 517-532. 1984.

[33] Hall, P. On powerful distributional tests based on sample spacings. *J. Multivariate Statist.* 19, 201-225, 1986.

[34] Hall, P. On Kullback-Leibler loss and density estimation. *Annals of Statistics*, 15, 1491-1519, 1987.

[35] Hall, P. and Morton S. C. On the estimation of the entropy. *Ann. Inst. Statist. Math.* 45, 69-88, 1993.

[36] Ivanov A. V. and Rozhkova . Properties of the statistical estimate of the entropy of a random vector with a probability density. *Problems of Information Transmission*, 17, 171-178, 1981.

[37] Joe H. On the estimation of entropy and other fuunctionals of a multivariate density. *Ann. Inst. Statist. Math.* 41, 683-697, 1989.

[38] Kagan, A. M., Linnik, Y. V. and Rao, C. R. *Characterization Problems in Mathematical Statistics*. Wiley, 1973.

[39] Kozachenko, L. F. and Leonenko, N. N. Sample estimate of entropy of a random vector. *Problems of Information Transmission*, 23, 95-101, 1987.

[40] Levit B. Ya. Asymptotically optimal estimation of nonlinear functionals *Problems of Information Transmission*, 14, 65-72, 1978.

[41] T. Linder and R. Zamir, On the asymptotic tightness of the Shannon lower bound. *IEEE Trans. Inform. Theory*, vol. 40, pp. 2026–2031, Nov. 1994.

[42] T. Linder and K. Zeger, Asymptotic entropy constrained performance of tessellating and universal randomized lattice quantization. *IEEE Trans. Inform. Theory*, vol. 40, pp. 575–579, March 1994.

[43] Y. N. Linkov, Evaluation of epsilon entropy of random variables for small epsilon. *Problems of Information Transmission*, vol. 1, pp. 12–18, 1965.

[44] Mokkadem, A. Estimation of the entropy and information for absolutely continuous random variables, *IEEE Trans. Information Theory*, 35, 193-196, 1989.

[45] Moran, P. A. P. The random division of an interval. Part I. *J. R. Statist. Soc. B*, 9, 92-98, 1951.

[46] Mudholkar, G. S. and Lin, C. T. On two applications of characterization theorems to goodness-of-fit. *Colloquia Mathematica Soc. János Bolyai*, 395-414, 1984.

[47] Mudholkar, G. S. and Smethurst, P. A. Empirically accelerated convergeence of the spacings statistics to normality. *Austral. J. Statist.*, 32, 129-138, 1990.

[48] Parzen, E. Goodness of fit tests and entropy. *J. Combinatorics, Information and System Sciences*, 16, 129-136, 1991.

[49] Prakasa Rao, B. L. S. *Nonparametric Functional Estimation*, Academic Press, 1983.

[50] Prescott, P. On a test for normality based on sample entropy, *J. Roy. Statist. Soc. Ser. B.* 38, 254-256, 1976.

[51] Ranneby, B. The maximum spacing method. An estimation method related to the maximum likelyhood method. *Scand. J. Statist.*, 11, 93-112, 1984.

[52] A. Rényi, On the dimension and entropy of probability distributions, *Acta Math. Acad. Sci. Hungar.*, vol. 10, pp. 193–215, 1959.

[53] Robinson, P. M. Consistent nonparametric entropy-based testing. *Review of Economic Studies*, 58, 437-453, 1991.

[54] Rodrigez, C. C. and Van Ryzin, J. Large sample properties of maximum entropy histograms. *IEEE Trans. Information Theory*, 32, 751-759, 1986.

[55] Shannon, C. E. A mathematical theory of communication. *Bell System Tech. J.*, 27, 379-423.

[56] Shao, Y. and Hahn, M. G. Limit theorems for the logarithm of sample spacings. *Statistics and Probability Letters*, 24, 121-132, 1995.

[57] Tarasenko, F.P. On the evaluation of an unknown probability density function, the direct estimation of the entropy from independent observations of a continuous random variable, and the distribution-free entropy test of goodness-of-fit. *Proc. IEEE* 56, 2052-2053, 1968.

[58] Theil, H. The entropy of the maximum entropy distribution. *Econom. Letters*, 5, 145-148, 1980.

[59] Theil, H. and Fiebig, D. G. *Exploiting Continuity: Maximum Entropy Estimation of Continuous Distributions.* Harper and Row, 1984.

[60] Theil, H. and Laitinen, K. Singular moment matrices in applied econometrics. in Multivariate Analysis, V, Ed. P. R. Krishnaiah, North-Holland, 629-629, 1980.

[61] Tsybakov, A. B. and van der Meulen, E. C. Root-n consistent estimators of entropy for densities with unbounded support. *Scand. J. Statist.*, 23, 75-83, 1994.

[62] Van Es B. Estimating functionals related to a density by a class of statistics based on spacings. *Scand. J. Statist.* 19, 61-72, 1992.

[63] Vasicek O. A test for normality based on sample entropy. *J. Roy. Statist. Soc. Ser. B* 38, 54-59, 1976.

[64] Verdugo Lazo, A. C. G. and Rathie, P. N. On the entropy of continuous probability distributions. *IEEE Trans. Inform. Theory*, 24, 120-122, 1978.