C. R. Acad. Sci. Paris, Ser. I 352 (2014) 75–80

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

New Kernel-type estimator of Shanonn's entropy

Nouvel estimateur à noyau de l'entropie de Shanonn

Salim Bouzebda [a], Issam Elhattab [b]

[a] *Laboratoire de mathématiques appliquées de Compiègne, Université de technologie de Compiègne, BP 529, 60205 Compiègne cedex, France*
[b] *ENCG–Casablanca, Université Hassan-II Mohammedia, Morocco*

ARTICLE INFO

Article history:
Received 16 February 2012
Accepted after revision 19 November 2013
Available online 20 December 2013

Presented by Paul Deheuvels

ABSTRACT

In the present Note, we propose an estimator of Shanonn's entropy based on smooth estimators of quantile density. The consistency and asymptotic normality of the proposed estimates are obtained.

RÉSUMÉ

Dans cette Note, nous proposons un nouvel estimateur de l'entropie de Shanonn basé sur l'estimateur à noyau de la densité de quantile. Nous obtenons la consistance et la normalité de l'estimateur proposé.

Version française abrégée

Soient $X_1, \ldots, X_n$ des variables aléatoires réelles, indépendantes et de même loi que X. Pour tout $x \in \mathbb{R}$, soit $F(x) = \mathbb{P}(X \leqslant x)$ la fonction de répartition de X. Soit $\mathcal{Q}(\cdot) := F^{-1}(\cdot)$ la fonction *inverse*, ou *de quantiles*, de $F(\cdot)$ définie par :

$$\mathcal{Q}(t) := \inf\{x\colon\ F(x) \geqslant t\}, \quad \text{pour } 0 < t < 1.$$

Posons $x_F = \sup\{x\colon\ F(x) = 0\}$, $x^F = \inf\{x\colon\ F(x) = 1\}$, $\infty \leqslant x_F < x^F \leqslant \infty$. Nous supposons, dans la suite, que $F(\cdot)$ admet une densité $f(\cdot)$ par rapport à la mesure de Lebesgue. Soit $q(x) = \mathrm{d}\mathcal{Q}(x)/\mathrm{d}x = 1/f(\mathcal{Q}(x))$, pour $0 < x < 1$, la fonction de densité du quantile. Nous définissons l'entropie associée à la densité $f(\cdot)$, si elle existe, par :

$$\mathcal{H}(X) = -\int_{\mathbb{R}} f(x) \log f(x)\, \mathrm{d}x.$$

Supposons que $|\mathcal{H}(X)| < \infty$. On peut réécrire $\mathcal{H}(X)$ (voir par exemple [23]) sous la forme suivante :

$$\mathcal{H}(X) = \int_{[0,1]} \log\left(\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{Q}(x)\right) \mathrm{d}x = \int_{[0,1]} \log\big(q(x)\big)\, \mathrm{d}x.$$

Rappelons que la fonction de quantile empirique $\mathcal{Q}_n(\cdot)$ est définie par :

$$\mathcal{Q}_n(t) := X_{k;n}, \quad (k-1)/n < t \leqslant k/n, \ \text{pour } k = 1, \ldots, n,$$

E-mail addresses: salim.bouzebda@utc.fr (S. Bouzebda), issam.elhattab@gmail.com (I. Elhattab).

1631-073X/$ – see front matter
http://dx.doi.org/10.1016/j.crma.2013.11.011

où $X_{1,n}, \dots, X_{n,n}$ désignent les statistiques d'ordre associées à $X_1, \dots, X_n$. Une version lissée de $\mathcal{Q}_n(\cdot)$ peut être définie comme suit, voir [9] :

$$\widehat{\mathcal{Q}}_n(t) = \int_0^1 \mathcal{Q}_n(x) K_n(t,x)\, \mathrm{d}\mu_n(x), \quad \text{pour } t \in (0,1),$$

ce qui permet ensuite d'estimer la densité du quantile $q(\cdot)$ par :

$$\widehat{q}_n(t) = \frac{\mathrm{d}}{\mathrm{d}t} \widehat{\mathcal{Q}}_n(t) = \frac{\mathrm{d}}{\mathrm{d}t} \int_0^1 \mathcal{Q}_n(x) K_n(t,x)\, \mathrm{d}\mu_n(x), \quad \text{pour } t \in (0,1), \tag{1}$$

où les suite de noyaux $\{K_n(t,x),\ (t,x) \in [0,1] \times (0,1)\}_{n \geqslant 1}$ et de mesures $\{\mu_n(\cdot)\}_{n \geqslant 1}$ satisfont certaines conditions de régularité. Dans ce travail, nous caractérisons les propriétés asymptotiques de l'estimateur suivant, pour tout $\varepsilon \in]0, 1/2[$ petit,

$$\mathcal{H}_{n,\varepsilon}(X) = \varepsilon \log\big(\widehat{q}_n(\varepsilon)\big) + \varepsilon \log\big(\widehat{q}_n(1-\varepsilon)\big) + \frac{1}{n} \sum_{i=\lceil \varepsilon n \rceil}^{\lfloor (1-\varepsilon) n \rfloor} \log \widehat{q}_n\big(F_n(X_i)\big),$$

où $F_n(\cdot)$ désigne la fonction de répartition empirique, i.e.,

$$F_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{]-\infty, t]}(X_i), \quad \text{pour } t \in \mathbb{R}.$$

1. Introduction and estimation

Let X be a random variable [r.v.] with cumulative distribution function $F(x) = \mathbb{P}(X \leqslant x)$ for $x \in \mathbb{R}$ and a density function $f(\cdot)$ with respect to Lebesgue measure on $\mathbb{R}$. Then its differential (or Shannon) entropy is defined by:

$$\mathcal{H}(X) = -\int_{\mathbb{R}} f(x) \log f(x)\, \mathrm{d}x. \tag{2}$$

We assume that $\mathcal{H}(X)$ is properly defined by the integral (2), in the sense that:

$$\big|\mathcal{H}(X)\big| < \infty. \tag{3}$$

We recall from (cf. [2, p. 237], [4, p. 108]) that the finiteness of $\mathcal{H}(X)$ is guaranteed if both $\mathbb{E}\|\mathbf{X}\|^2 < \infty$, where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$ (in which case $\mathcal{H}(X) < \infty$) and $f(\cdot)$ is bounded (in which case $\mathcal{H}(X) > -\infty$). In [2], Ash gives an example of a density function on $\mathbb{R}$ for which $\mathcal{H}(X) = \infty$ and also one for which $\mathcal{H}(X) = -\infty$. We refer to [15, Section 4] for conditions characterizing (3) in terms of $f(\cdot)$. The concept of differential entropy was introduced in Shannon's original paper [18]. Since this early epoch, the notion of entropy has been the subject of great theoretical and applied interest. We refer to [10] (see their Chapter 8) for a comprehensive overview of differential entropy and their mathematical properties. In the literature, several proposals have been made to estimate entropy. Dmitriev and Tarasenko [12] and Ahmad and Lin [1] proposed estimators of the entropy using kernel-type estimators of the density $f(\cdot)$; we refer to [5,6] for more details. Vasicek [23] proposed an entropy estimator based on spacings. Inspired by the work of [23], some authors [22,24,13,7,16] proposed modified entropy estimators, improving in some respects the properties of Vasicek's estimator. The reader will find in [3] detailed accounts of the theory as well as surveys for entropy estimators.

This Note aims to introduce a new entropy estimator and obtains its asymptotic properties. Let us first set out the basic definitions and notation which will be used throughout the sequel. For each distribution function $F(\cdot)$, we define the quantile function by:

$$\mathcal{Q}(t) := \inf\big\{x\colon\ F(x) \geqslant t\big\}, \quad \text{for } 0 < t < 1.$$

Let:

$$x_F = \sup\big\{x\colon\ F(x) = 0\big\} \quad \text{and} \quad x^F = \inf\big\{x\colon\ F(x) = 1\big\}, \quad -\infty \leqslant x_F < x^F \leqslant \infty.$$

We assume that the distribution function $F(\cdot)$ has a density $f(\cdot)$ (with respect to Lebesgue measure in $\mathbb{R}$), and that $f(x) > 0$ for all $x \in (x_F, x^F)$. Let:

$$q(x) = \mathrm{d}\mathcal{Q}(x)/\mathrm{d}x = 1/f\big(\mathcal{Q}(x)\big), \quad \text{for } 0 < x < 1,$$

be the quantile density function (qdf). The entropy $\mathcal{H}(X)$, defined by (2), can be expressed in the form of a quantile-density functional as:

$$\mathcal{H}(X) = \int_{[0,1]} \log\left(\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{Q}(x)\right)\mathrm{d}x = \int_{[0,1]} \log\bigl(q(x)\bigr)\,\mathrm{d}x. \tag{4}$$

Vasicek's estimator was constructed by replacing the quantile function $\mathcal{Q}(\cdot)$ in (4) by the empirical quantile function and using a difference operator instead of the differential one. The derivative $(\mathrm{d}/\mathrm{d}x)\mathcal{Q}(x)$ is then estimated by a function of spacings. In this work, we construct our estimator of entropy by replacing $q(\cdot)$, in (4), by an appropriate estimator $\widehat{q}_n(\cdot)$ of $q(\cdot)$. We shall consider the kernel-type estimator of $q(\cdot)$ introduced by [14] and studied by [8] and [9]. Our choice is motivated by the well-asymptotic properties of this estimator. First, smoothing reduces the random variation in the data, resulting in a more efficient estimator. Second, smoothing gives a smooth curve for the quantile function that better displays the interesting features of the population distribution; at this point we may refer to [9] for more details and discussion on the subject. It is noteworthy here to point out that many of the quantile functions used in some applications do not have tractable distribution functions. As was pointed out in [21,20], this makes the analytical study of the properties of $\mathcal{H}(X)$ of these distributions by means of (2) difficult. Hence the alternative definition and properties of entropy in terms of quantile functions can be used in this situation. The entropy expressed as a density-quantile functional has advantages. The analytical properties of the entropy function obtained in this approach can be used as alternative tools in modeling data. Sometimes, the quantile-based approach is better in terms of tractability. Finally, we note that the paper by [17] gives an evidence of good properties of the statistical tests for goodness-of-fit that are based on entropy, which is expressed as a quantile-density functional. In addition, in the last reference, the unifying roles of entropy statistics including, among others, those introduced by Moran, Vasicek and Dudewicz and van der Meulen, were discussed. In [9] were established the asymptotic properties of $\widehat{q}_n(\cdot)$ on all compact $U \subset]0,1[$, which avoids boundary problems. Since the entropy is defined as an integral on $]0,1[$ of a functional of $q(\cdot)$, it is not suitable to substitute directly $\widehat{q}_n(\cdot)$ in (4) to estimate $\mathcal{H}(X)$. To circumvent the boundary effects, we will proceed as follows, as in [7]. We set for small $\varepsilon \in]0,1/2[$:

$$\mathcal{H}_\varepsilon(X) := \varepsilon\log\bigl(q(\varepsilon)\bigr) + \varepsilon\log\bigl(q(1-\varepsilon)\bigr) + \int_{\varepsilon}^{1-\varepsilon} \log\bigl(q(x)\bigr)\,\mathrm{d}x.$$

In view of (4), we can see that:

$$\mathcal{H}(X) - \mathcal{H}_\varepsilon(X) = o\bigl(\eta(\varepsilon)\bigr), \tag{5}$$

where

$$-\eta(\varepsilon) \to 0, \quad \text{as } \varepsilon \to 0.$$

The choice of ε close to zero guaranteed the closeness of $\mathcal{H}_\varepsilon(X)$ to $\mathcal{H}(X)$, then the problem of the estimation of $\mathcal{H}(X)$ is reduced to the estimation of $\mathcal{H}_\varepsilon(X)$. Given an independent and identically distributed random [i.i.d.] sample $X_1,\ldots,X_n$, we define, in order to introduce our entropy estimator, in a first step, a kernel quantile density estimator $\widehat{q}_n(\cdot)$ of qdf $q(\cdot)$. Towards this aim, we proceed as follows. Let $X_{1;n} \leqslant \cdots \leqslant X_{n;n}$ denote the order statistics of $X_1,\ldots,X_n$. The empirical quantile function $\mathcal{Q}_n(\cdot)$ based upon these random variables is given by:

$$\mathcal{Q}_n(t) := X_{k;n}, \quad (k-1)/n < t \leqslant k/n, \text{ for } k = 1,\ldots,n. \tag{6}$$

Let $\{K_n(t,x)\colon (t,x) \in [0,1]\times(0,1)\}_{n\geqslant1}$ be a sequence of kernels and $\{\mu_n\}_{n\geqslant1}$ a sequence of σ-finite measure on $[0,1]$. A smoothed version of $\mathcal{Q}_n(\cdot)$ (see, e.g., [9]) can be defined as:

$$\widehat{\mathcal{Q}}_n(t) = \int_0^1 \mathcal{Q}_n(x)K_n(t,x)\,\mathrm{d}\mu_n(x), \quad \text{for } t \in (0,1).$$

Finally, following [8], we estimate $q(\cdot)$ by:

$$\widehat{q}_n(t) = \frac{\mathrm{d}}{\mathrm{d}t}\widehat{\mathcal{Q}}_n(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 \mathcal{Q}_n(x)K_n(t,x)\,\mathrm{d}\mu_n(x), \quad \text{for } t \in (0,1). \tag{7}$$

Clearly, in order to obtain a meaningful qdf estimator in this way, the sequence of kernels $\{K_n(t,x)\colon (t,x) \in [0,1]\times(0,1)\}_{n\geqslant1}$ must satisfy certain differentiability conditions and, together with the sequence of measures $\{\mu_n\}_{n\geqslant1}$, must satisfy certain variational conditions. These conditions will be detailed in the next section.

A familiar example is the convolution-kernel estimator:

$$\widehat{q}_n(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 h_n^{-1}\mathcal{Q}_n(x)K\left(\frac{t-x}{h_n}\right)\mathrm{d}x, \quad \text{for } t \in (0,1),$$

where $K(\cdot)$ denotes a kernel function, namely a measurable function integrating to 1 on $\mathbb{R}$, and has bounded derivative, and $\{h_n\}_{n\geqslant 1}$ is a sequence of positive reals fulfilling $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. In this case, Csörgö et al. [11] define:

$$\bar{q}_n(t) = h_n^{-1} \int_{1/(n+1)}^{n/(n+1)} K\left(\frac{t-x}{h_n}\right) \mathrm{d}\mathcal{Q}_n(x) = h_n^{-1} \sum_{i=1}^{n-1} K\left(\frac{t-i/n}{h_n}\right)(X_{i+1;n} - X_{i;n}), \quad \text{for } t \in (0,1).$$

Calculations using summation by parts show that:

$$\widehat{q}_n(t) = \bar{q}_n(t) + h_n^{-1}\left[K\left(\frac{t-1}{h_n}\right)X_{n;n} - K\left(\frac{t}{h_n}\right)X_{1;n}\right], \quad \text{for } t \in (0,1).$$

The new family of estimators that we propose to estimate the entropy is given by:

$$\mathcal{H}_{n,\varepsilon}(X) = \varepsilon \log\bigl(\widehat{q}_n(\varepsilon)\bigr) + \varepsilon \log\bigl(\widehat{q}_n(1-\varepsilon)\bigr) + \frac{1}{n} \sum_{i=\lceil \varepsilon n \rceil}^{\lfloor (1-\varepsilon)n \rfloor} \log \widehat{q}_n\bigl(F_n(X_i)\bigr), \tag{8}$$

where

$$F_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{]-\infty,t]}(X_i), \quad \text{for } t \in \mathbb{R}, \tag{9}$$

and where $\mathbb{1}_A$ stands for the indicator function of the event A. It is noteworthy here to point out that compared with the estimator given in [7], our estimator, defined in (8), reduces the computational requirement that, from a practical standpoint, can be very significant when n is large.

The remainder of this Note is organized as follows. The consistency and normality of our estimator are discussed in the next section. Our arguments, used to establish the asymptotic normality of our estimator, make use of an original application of the invariance principle for the uniform empirical process.

2. Main results

Throughout the following, $U := U(\varepsilon) := [\varepsilon, 1-\varepsilon]$ for $\varepsilon \in]0, 1/2[$ arbitrarily fixed (free of the sample size n). The following conditions are used to establish the main results. The previously defined notation continues to be used below.

(Q.1) The qdf $q(\cdot)$ is twice differentiable in $(0,1)$. Moreover,

$$0 < \min\bigl\{q(0), q(1)\bigr\} \leqslant \infty.$$

(Q.2) There is a $\varsigma > 0$ such that:

$$\sup_{t\in(0,1)} \bigl\{t(1-t)\bigl|J(t)\bigr|\bigr\} \leqslant \varsigma,$$

where $J(t) = \mathrm{d}\log\{q(t)\}/\mathrm{d}t$ is the score function.

(Q.3) Either $q(0) < \infty$ or $q(\cdot)$ is nonincreasing in some interval $(0, t_*)$, and either $q(1) < \infty$ or $q(\cdot)$ is nondecreasing in some interval $(t^*, 1)$, where $0 < t_* < t^* < 1$.

We will make the following assumptions on the sequence of kernels $K_n(\cdot,\cdot)$.

(K.1) For each n and each $(t,x) \in U \times (0,1)$, $K_n(t,x) \geqslant 0$, and for each $t \in U$,

$$\int_0^1 K_n(t,x)\,\mathrm{d}\mu_n(x) = 1;$$

(K.2) There is a sequence $\delta_n \downarrow 0$ such that:

$$R_n = \sup_{t\in U}\left[1 - \int_{t-\delta_n}^{t+\delta_n} K_n(t,x)\,\mathrm{d}\mu_n(x)\right] \to 0, \quad \text{as } n \to \infty;$$

(K.3) For any function $g(\cdot)$ that is at least three times differentiable in $(0,1)$,

$$\widehat{g}_n(t) := \int_0^1 g(x) K_n(t,x)\, \mathrm{d}\mu_n(x),$$

is differentiable in t on U, and

$$\sup_{t\in U}\big|g(t) - \widehat{g}_n(t)\big| = O\big(n^{-\alpha}\big), \quad \alpha > 0;$$

$$\sup_{t\in U}\left|g'(t) - \frac{\mathrm{d}}{\mathrm{d}t}\widehat{g}_n(t)\right| = O\big(n^{-\beta}\big), \quad \beta > 0;$$

(K.4) For some finite positive constant C and $\varrho > 1$, suppose:

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\widehat{g}_n(s) - \frac{\mathrm{d}}{\mathrm{d}t}\widehat{g}_n(t)\right| \leqslant C|s-t|^{\varrho}.$$

Note that the conditions (Q.1)–(Q.3) and (K.1)–(K.3), which we will use to establish the convergence in probability of $\mathcal{H}_{n,\varepsilon}(X)$, match those used by [9] to establish the asymptotic properties of $\widehat{q}_n(\cdot)$.

We will use the following additional condition:

(Q.4)

$$\mathbb{E}\big\{\log^2\big(q\big(F(X)\big)\big)\big\} < \infty.$$

Let, for all $\varepsilon \in]0, 1/2[$,

$$\mathbb{E}\big\{\log^2\big(q\big(F(X)\big)\big)\big\} - \left\{\varepsilon \log^2\big(q(\varepsilon)\big) + \varepsilon \log^2\big(q(1-\varepsilon)\big) + \int_{\varepsilon}^{1-\varepsilon} \log^2\big(q(x)\big)\,\mathrm{d}x\right\} = o\big(\vartheta(\varepsilon)\big),$$

where $\vartheta(\varepsilon) \to 0$, as $\varepsilon \to 0$. We are now in a position to state our result concerning consistency of $\mathcal{H}_{n,\varepsilon}(X)$ defined in (8), where the qdf estimator is given in (7).

Theorem 2.1. *Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with a quantile density function $q(\cdot)$ fulfilling* (Q.1)–(Q.4)*. Let $K_n(\cdot,\cdot)$ fulfill* (K.1)–(K.3)*. Then, we have*:

$$\big|\mathcal{H}_{n,\varepsilon}(X) - \mathcal{H}_{\varepsilon}(X)\big| = O_{\mathbb{P}}\big(n^{-1/2} M(q; K_n) + n^{-\beta} + n^{-1/2}(\log\log n)^{1/2}\big), \tag{10}$$

where

$$M(q; K_n) = M q \widehat{M}_n(1)\sqrt{\delta_n \log \delta_n^{-1}} + M_{q'} + \sqrt{\widehat{M}_n\big(q^2\big) R_n'(1)} + n^{-1/2} A_{\gamma}(n) M_q \widehat{M}_n(1),$$

$$\widehat{M}_n(g) = \sup_{u\in U}\int_0^1 \big|g(x) K_n(u,x)\big|\,\mathrm{d}\mu_n(x),$$

$$R_n'(g) = \sup_{u\in U}\int_{[0,1]\setminus U^{\delta_n}} \big|g(x) K_n(u,x)\big|\,\mathrm{d}\mu_n(x),$$

$$M_g = \sup_{u\in U}\big|g(u)\big|,$$

where δ_n is given in condition (K.2) *and U^{δ_n} is a δ_n-neighborhood of U.*

In the sequel, by a Brownian bridge $\{B(t)\colon\ 0 \leqslant t \leqslant 1\}$, is meant a centered Gaussian process with continuous sample paths, and covariance function:

$$\mathbb{E}\big(B(s)B(t)\big) = s \wedge t - st, \quad \text{for } 0 \leqslant s, t \leqslant 1.$$

The main result, concerning the normality of $\mathcal{H}_{n,\varepsilon}(X)$, to be proved here may now be stated precisely as follows.

Theorem 2.2. *Assume that the conditions* (Q.1)–(Q.4) *and* (K.1)–(K.4) *hold with* $\alpha > 1/2$ *and* $\beta > 1/2$ *in* (K.3)*. On a probability space rich enough, one can construct a sequence* $\{B_n(u)\colon u \in [0,1]\}$ *of Brownian bridges, such that we have*:

$$\sqrt{n}\big(\mathcal{H}_{n,\varepsilon}(X) - \mathcal{H}_{\varepsilon}(X)\big) - \psi(n) = o_{\mathbb{P}}(1), \tag{11}$$

where

$$\psi(n) := \int_{\varepsilon}^{1-\varepsilon} \left\{\frac{q'(x)}{q(x)}\right\} B_n(x)\,\mathrm{d}x$$

is a centered Gaussian random variable with variance equal to:

$$\mathrm{Var}\big(\psi_n(\varepsilon)\big) = \mathrm{Var}\big\{\log\big(q\big(F(X)\big)\big)\big\} + o\big(\vartheta(\varepsilon) + \big(\eta(\varepsilon)\big)^2\big).$$

Remark 1. Condition (Q.4) is extremely weak and is satisfied by all commonly encountered distributions including many important heavy tailed distributions for which the moments do not exist (see, e.g., [19]).

Remark 2. We mention that condition (K.3) is satisfied when $\alpha, \beta > 1/2$ for the difference kernels $K_n(t,x)\,\mathrm{d}\mu_n(x) = h_n^{-1} K((t-x)/h_n)\,\mathrm{d}x$ with $h_n = O(n^{-\nu})$, where $1/4 < \nu < 1$. A typical example for difference kernels satisfying theses conditions is the Gaussian kernel.

Remark 3. In order to extract methodological recommendations for the use of the entropy estimate proposed in this work, it will be interesting to conduct extensive Monte Carlo experiments to compare, in terms of mean squared error, several estimators as in [7], which would go well beyond the scope of the present Note.

Acknowledgements

The authors would like to acknowledge the referee for his/her constructive criticism and helpful comments, which led to important improvements over an earlier draft.

References

[1] I.A. Ahmad, P.E. Lin, A nonparametric estimation of the entropy for absolutely continuous distributions, IEEE Trans. Inf. Theory IT-22 (3) (1976) 372–375.
[2] R. Ash, Information Theory, Intersci. Tracts Pure Appl. Math., vol. 19, Interscience Publishers, John Wiley & Sons, New York–London–Sydney, 1965.
[3] J. Beirlant, E.J. Dudewicz, L. Györfi, E.C. van der Meulen, Nonparametric entropy estimation: an overview, Int. J. Math. Stat. Sci. 6 (1) (1997) 17–39.
[4] T. Berger, Rate distortion theory, in: A Mathematical Basis for Data Compression, in: Prentice-Hall Ser. Inf. Syst. Sci., Prentice Hall Inc., Englewood Cliffs, NJ, 1971.
[5] S. Bouzebda, I. Elhattab, A strong consistency of a nonparametric estimate of entropy under random censorship, C. R. Acad. Sci. Paris, Ser. I 347 (13–14) (2009) 821–826.
[6] S. Bouzebda, I. Elhattab, Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy, Electron. J. Stat. 5 (2011) 440–459.
[7] S. Bouzebda, I. Elhattab, A. Keziou, L. Lounis, New entropy estimator with an application to test of normality, Commun. Stat., Theory Methods 42 (12) (2013) 2245–2270.
[8] C. Cheng, Uniform consistency of generalized kernel estimators of quantile density, Ann. Stat. 23 (6) (1995) 2285–2291.
[9] C. Cheng, E. Parzen, Unified estimators of smooth quantile and quantile density functions, J. Stat. Plan. Inference 59 (2) (1997) 291–307.
[10] T.M. Cover, J.A. Thomas, Elements of Information Theory, second edition, Wiley–Interscience [John Wiley & Sons], Hoboken, NJ, 2006.
[11] M. Csörgő, L. Horváth, P. Deheuvels, Estimating the quantile-density function, in: Nonparametric Functional Estimation and Related Topics, Spetses, 1990, in: NATO Adv. Stud. Inst. Ser., Ser. C, Math. Phys. Sci., vol. 335, Kluwer Acad. Publ., Dordrecht, 1991, pp. 213–223.
[12] J.G. Dmitriev, F.P. Tarasenko, The estimation of functionals of a probability density and its derivatives, Teor. Verojatnost. i Primenen. 18 (1973) 662–668.
[13] M.D. Esteban, M.E. Castellanos, D. Morales, I. Vajda, Monte Carlo comparison of four normality tests using different entropy estimates, Commun. Stat., Simul. Comput. 30 (4) (2001) 761–785.
[14] M. Falk, On the estimation of the quantile density function, Stat. Probab. Lett. 4 (2) (1986) 69–73.
[15] L. Györfi, E.C. van der Meulen, On the nonparametric estimation of the entropy functional, in: Nonparametric Functional Estimation and Related Topics, Spetses, 1990, in: NATO Adv. Stud. Inst. Ser., Ser. C, Math. Phys. Sci., vol. 335, Kluwer Acad. Publ., Dordrecht, 1991, pp. 81–95.
[16] E. Parzen, Nonparametric statistical data modeling, J. Am. Stat. Assoc. 74 (365) (1979) 105–131, with comments by John W. Tukey, Roy E. Welsch, William F. Eddy, D.V. Lindley, Michael E. Tarter and Edwin L. Crow, and a rejoinder by the author.
[17] E. Parzen, Goodness of fit tests and entropy, J. Comb. Inf. Syst. Sci. 16 (2–3) (1991) 129–136.
[18] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, 623–656.
[19] K.-S. Song, Limit theorems for nonparametric sample entropy estimators, Stat. Probab. Lett. 49 (1) (2000) 9–18.
[20] S.M. Sunoj, P.G. Sankaran, Quantile based entropy function, Stat. Probab. Lett. 82 (6) (2012) 1049–1053.
[21] S.M. Sunoj, P.G. Sankaran, A.K. Nanda, Quantile based entropy function in past lifetime, Stat. Probab. Lett. 83 (1) (2013) 366–372.
[22] B. van Es, Estimating functionals related to a density by a class of statistics based on spacings, Scand. J. Stat. 19 (1) (1992) 61–72.
[23] O. Vasicek, A test for normality based on sample entropy, J. R. Stat. Soc. B 38 (1) (1976) 54–59.
[24] R. Wieczorkowski, P. Grzegorzewski, Entropy estimators—improvements and comparisons, Commun. Stat., Simul. Comput. 28 (2) (1999) 541–567.