DISCUSSION
PAPER

9206

ROOT-n CONSISTENT ESTIMATORS OF ENTROPY FOR DENSITIES WITH UNBOUNDED SUPPORT

A.B. TSYBAKOV and E.C. VAN DER MEULEN

INSTITUT DE
STATISTIQUE

UNIVERSITE CATHOLIQUE DE LOUVAIN
VOIE DU ROMAN PAYS, 34
B-1348 LOUVAIN-LA-NEUVE
BELGIUM

ROOT-n CONSISTENT ESTIMATORS OF ENTROPY FOR DENSITIES WITH UNBOUNDED SUPPORT

by

A.B. Tsybakov* **and E.C. van der Meulen****

December 1992

Abstract

We consider the truncated version of the entropy estimator proposed by Kozachenko and Leonenko (1987) and based on spacings of order 1. We prove the mean squared $\sqrt{n}$-consistency of the estimator for a class of densities with unbounded support, including the Gaussian density.

AMS 1991 Subject Classification: 62 G05, 62 G20

Key words: entropy, spacings, $\sqrt{n}$-consistency, densities with unbounded support

---

* CORE and Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium and Institute for Problems of Information Transmission, Russian Academy of Sciences, Moscow, Russia.

** Department of Mathematics, Katholieke Universiteit Leuven, Leuven, Belgium.

1. Introduction

Let $X_1, \ldots, X_n$ be a sample of i.i.d. random variables with common density $f(x)$, $x \in \mathbb{R}^1$. Consider the problem of estimation of the unknown entropy

$$H(f) = -\int f(x) \ln f(x) dx. \tag{1.1}$$

This problem has various applications in hypothesis testing and information theory, and was studied in many papers starting from Dobrushin (1958) and Tarasenko (1968). There exist two main approaches to the construction of entropy estimators. First, the "plug-in" approach which consists in substitution of $f(x)$ in (1.1) by suitable nonparametric density estimators (see Györfi and van der Meulen (1987,1990), Dudewicz and van der Meulen (1987), and the references therein). We shall not pursue this approach here.

The second approach is based on spacings. Let $X_{n,1} \leq X_{n,2} \leq \ldots \leq X_{n,n}$ be the order statistics of $X_1, \ldots, X_n$. Consider the statistic

$$\widehat{H}_{m,n} = \frac{1}{n} \sum_{i=1}^{n-m} \ln \left(\frac{n}{m} (X_{n,i+m} - X_{n,i}) \right), \tag{1.2}$$

where m is a positive integer less than n. It turns out that under appropriate conditions this statistic is a consistent (in probability or a.s.) and asymptotically normal estimator of $H(f)$ as $n \to \infty$. In fact, a simple heuristic argument suggests that $X_{n,i+m} - X_{n,i} \approx \dfrac{m}{nf(X_{n,i})}$ if $m/n \to 0$ as $n \to \infty$. Thus, $\widehat{H}_{m,n}$ is a kind of empirical approximation for $H(f)$.

The asymptotic properties of $\widehat{H}_{m,n}$ (and versions of it based on spacings of order $2m$) have been studied by Cressie (1976), Vasicek (1976), and Dudewicz and van der Meulen (1981). Note that (1.2) is a special example of a more general statistic

$$T_{m,n} = \frac{1}{n} \sum_{i=1}^{n-m} h \left(\frac{n}{m} (X_{n,i+m} - X_{n,i}) \right), \tag{1.3}$$

where h is some measurable real-valued function. Statistics of type (1.3) have been investigated by Bickel and Breiman (1983), Hall (1984,1986), Beirlant and van Zuijlen (1985), Beirlant (1986), Khashimov (1989) and van Es (1992) under various conditions on h, that sometimes include the case $h(x) = \ln x$, and sometimes not.

Let us give here a brief summary of the results concerning the asymptotic behaviour of $\widehat{H}_{m,n}$ as $n \to \infty$. We distinguish between the two cases: 1) $m \to \infty$ as $n \to \infty$, and 2) m is fixed.

1) $m \to \infty$. This case was considered in the papers of Vasicek (1976), Dudewicz and van der Meulen (1981), Beirlant and van Zuijlen (1985), Hall (1986), Beirlant (1986), Khashimov (1989), van Es (1992).

The asymptotic properties of the statistic (1.3) were studied under various assumptions on h and the density f. For our case of $h(x) = \ln x$ it was proved that $\widehat{H}_{m,n}$ (or one of its modifications) is a consistent (in probability or a.s.) estimator of $H(f)$ provided $m, n \to \infty$, $\frac{m}{n} \to 0$, and f satisfies some general conditions (Vasicek (1976), Beirlant and van Zuijlen (1985)). The asymptotic distribution of $\widehat{H}_{m,n}$ was derived under the following assumption on f:

(1.4) *f is a bounded density which is bounded away from zero on its support.*

Hall (1986, Th. 3) and van Es (1992, Th. 4) show that if (1.4) holds and $m = m_n \to \infty$, then under some additional regularity conditions and conditions on the speed with which $m \to \infty$, one has

$$\sqrt{n}(\widehat{H}_{m,n} - H(f)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}\ (\ln f(X_1))), \tag{1.5}$$

as $n \to \infty$. The variance Var $(\ln f(X_1))$ is the smallest possible variance for entropy estimators in the local asymptotical minimax sense (Levit (1978)). Hence $\widehat{H}_{m,n}$ proves to be an asymptotically efficient entropy estimator if $m \to \infty$.

2) *m fixed.* It turns out that for m fixed $\widehat{H}_{m,n}$ is not a consistent estimator of entropy and has asymptotically the *constant bias* $\Psi(m) - \ln m$, since

$$\widehat{H}_{m,n} \to H(f) + \Psi(m) - \ln m, \quad m = 1, 2, \dots, \tag{1.6}$$

in probability, as $n \to \infty$, where Ψ denotes the digamma function, $\Psi(x) = -(\ln \Gamma(x))'$. In particular, $\Psi(1) = -C_E$ where C_E is the Euler constant: $C_E = -\int\limits_0^\infty e^{-t} \ln t \, dt = 0,5772\dots$. The result (1.6) was proved by Hall (1984) for densities satisfying (1.4). For the uniform density f this was known earlier (see Tarasenko (1968) and Cressie (1976) and the references therein). Clearly, (1.6) suggests that for m fixed we can correct our estimator, and consider

$$\widetilde{H}_{m,n} = \widehat{H}_{m,n} - \Psi(m) + \ln m$$

which is a consistent entropy estimator.

For the asymptotic normality of $\widehat{H}_{m,n}$ with m fixed we refer to Cressie (1976), Dudewicz and van der Meulen (1981), Hall (1984) and Beirlant (1986), who prove that under some conditions including (1.4) we have

$$\sqrt{n}(\widetilde{H}_{m,n} - H(f)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (2m^2 - 2m + 1)\Psi'(m) - 2m + 1 + \mathrm{Var}\,[\ln f(X_1)]\right),$$

as $n \to \infty$. In particular, $\Psi'(1) = \pi^2/6$, and thus for $m = 1$

$$\sqrt{n}(\widetilde{H}_{m,n} - H(f)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\pi^2}{6} - 1 + \mathrm{Var}\,[\ln f(X_1)]\right).$$

We note that

$$(2m^2 - 2m + 1)\Psi'(m) - 2m + 1 > 0, \quad m = 2, 3, \ldots$$

Therefore, $\widetilde{H}_{m,n}$ is not an asymptotically efficient entropy estimator for m fixed.

Note that the results on the asymptotic normality of $\widehat{H}_{m,n}$ and $\widetilde{H}_{m,n}$ relate to the densities satisfying (1.4) which is a rather restrictive class. For densities with unbounded support there are for $\widehat{H}_{m,n}$ only the consistency results by Vasicek (1976) and Beirlant and van Zuijlen (1985) available. So far no rates of convergence were given. In this paper we establish the $\sqrt{n}$-convergence result for another entropy estimator based on spacings on a class of densities with *unbounded support*.

We study the estimator which is somewhat different from $\widetilde{H}_{m,n}$ and is defined by

$$H_n^0 = \frac{1}{n}\sum_{i=1}^{n} \ln(n\rho_i) + \ln 2 + C_E \tag{1.7}$$

where C_E is the Euler constant,

$$\rho_i = \min\{a_n, \min_{j \neq i} |X_i - X_j|\},$$

and $a_n \to 0$ is a sequence of positive numbers. In the following we assume that $a_n = 1/\sqrt{n}$. Definition (1.7) gives a truncated version of the estimate of Kozachenko and Leonenko (1987), who considered the case $\rho_i = \min_{j \neq i} |X_i - X_j|$. In the Kozachenko-Leonenko estimator we have, instead of the difference of two order statistics of the type $X_{n,i+1} - X_{n,i}$ or of the type $X_{n,i+1} - X_{n,i-1}$ (as in

$\widetilde{H}_{m,n}$ for the case $m = 1$ or $m = 2$, respectively), the minimum of two differences of two order statistics of the first kind, viz. $\min\{X_{n,i+1} - X_{n,i}, X_{n,i} - X_{n,i-1}\}$. This estimator may therefore be interpreted as intermediate between the versions of $\widetilde{H}_{m,n}$ for $m = 1$ and $m = 2$. In fact, the additive term $\ln 2$ is analogous to the case $m = 2$, which is due to the fact that in each ρ_i the order statistics $X_{n,i-1}$ and $X_{n,i+1}$, which are separated by two steps, are present. However, the Euler constant appears as in the case of $m = 1$ since H_n^0 depends on the spacings of order 1.

For technical reasons it is more convenient to consider the estimator which is the slightly modified version of (1.7):

$$H_n = \frac{1}{n}\sum_{i=1}^{n} \zeta_i \tag{1.8}$$

where $\zeta_i = \ln\{2\rho_i\gamma(n-1)\}, \gamma = \exp\{C_E\}$. Note that, the $\sqrt{n}$-consistency of the estimator (1.8) entails the $\sqrt{n}$-consistency of (1.7).

The advantage of the Kozachenko-Leonenko estimate is that its definition can be easily extended to the case of N-dimensional X_i, in contrast to the usual entropy estimators $\widehat{H}_{1,n}, \widetilde{H}_{1,n}$ based on one-dimensional spacings. Kozachenko and Leonenko (1987) prove the mean squared consistency of their estimator under rather general conditions on the density f in the multivariate case.

In this paper we consider the estimator H_n in the one-dimensional case only. In Theorem 1 we prove that the bias of H_n is of order $O\left(\frac{1}{\sqrt{n}}\right)$, $n \to \infty$. In Theorem 2 we show that the variance of H_n is of order $O\left(\frac{1}{n}\right)$. Thus, our estimator H_n converges to $H(f)$ with the rate $O\left(\frac{1}{\sqrt{n}}\right)$ which is the best possible rate for the estimation of density functionals in the asymptotical minimax sense (Levit (1978)). However, we don't specify the constant factor which multiplies the rate of convergence. Our results hold for densities f *with unbounded support and exponentially decreasing tails*, such as the Gaussian density.

2. The results

Assume the following

(A0) $\int f(x)|\ln f(x)|dx < \infty$.

(A1) f is twice continuously differentiable and strictly positive on $\mathbb{R}^1$.

(A2) $\int f(x)\exp(-bf(x))dx \leq Cb^{-1}$.

Here and later C stands for finite positive constants, possibly different.

The following theorem gives the rate of convergence for the bias of H_n.

Theorem 1. *Assume* (A0)-(A2). *Then for the estimate* (1.8) *we have*

$$E(H_n) - H(f) = O\left(\frac{1}{\sqrt{n}}\right),$$

as $n \to \infty$.

Now we give the convergence result for the variance of H_n. The conditions on f for this result are in some respect less restrictive than in Theorem 1, and we state them separately.

(B1) The density f is Lipschitz continuous and strictly positive on $\mathbb{R}^1$.

(B2) There exists $a > 0$ such that the integrals $\int f(x)\left(\frac{\sup_{|z-x|\leq a} f(z)}{f(x)}\right)^j dx$, $\int f(x)\left(\frac{\sup_{|z-x|\leq a} f(z)}{f(x)}\right)^j \ln^2 f(x)dx$ are finite for $j = 1, 2, 3$.

Theorem 2. *Assume* (A0), (B1), (B2). *Then for the estimate* (1.8) *we have*

$$E(H_n - E(H_n))^2 = O\left(\frac{1}{n}\right),$$

as $n \to \infty$.

From Theorems 1 and 2 we get the following corollary which is the main result of this paper.

Corollary. *Assume* (A0)-(A2), (B2). *Then the estimate* (1.8) *is* $\sqrt{n}$*-consistent in the mean squared sense, i.e.*

$$E\left((\sqrt{n}[H_n - H(f)])^2\right) = O(1),$$

as $n \to \infty$.

Note that the conditions (B1) and (B2) for the variance convergence are rather general, and they hold for densities with polynomially decreasing tails. The main restriction is given by (A2) which, in fact, is almost equivalent to the condition that $f(x)$ has exponentially decreasing tails. One can easily check that if $f(x)$ is bounded and

$$C_1 \exp(-\alpha|x|^\beta) \leq f(x) \leq C_2 \exp(-\alpha|x|^\beta)$$

for some constants $0 < C_1 < C_2 < \infty$ $\alpha > 0$, $\beta > 1$, and $|x|$ large enough, then (A2) is satisfied. For example, (A2) holds for the Gaussian density.

3. Proof of Theorem 1

We introduce some notation. Set

$$b_n = 2\gamma a_n(n-1) = 2\gamma(n-1)/\sqrt{n},$$

and for every $u > 0$ define

$$r_n(u) = \frac{u}{2\gamma(n-1)}.$$

Denote

$$\rho_i' = \min_{j \neq i} |X_i - X_j|, \zeta_i' = \ln(2\rho_i'\gamma(n-1)),$$

and introduce the following conditional distributions

$$F_{n,x}^{(1)}(u) = P\{\exp \zeta_i < u \mid X_i = x\} = P\{\min(a_n, \min_{j \neq i} |X_i - X_j|) < r_n(u)\},$$

$$F_{n,x}(u) = P\{\exp \zeta_i' < u \mid X_i = x\} = P\{\min_{j \neq i} |X_i - X_j| < r_n(u) \mid X_i = x\}.$$

Set

$$F_x(u) = 1 - \exp(-f(x)u/\gamma).$$

Note that $F_{n,x}^{(1)}(u) = F_{n,x}(u)$, $u \leq b_n$, and $F_{n,x}^{(1)}(u) = 1$, $u > b_n$. (In fact, $e^{\zeta_i} \leq 2\rho_i\gamma(n-1) \leq b_n$.) Using this and proceeding as in Kozachenko and Leonenko (1987) we obtain

$$H(f) = \int \left[\int_0^\infty \ln u dF_x(u) \right] f(x)dx,$$

(3.1)

$$E(H_n) = E(\zeta_1) = \int \left[\int_0^{b_n} \ln u dF_{n,x}(u) \right] f(x)dx,$$

and

(3.2)
$$|H(f) - E(H_n)| \leq \int [I_1 + I_2 + I_3] f(x)dx$$

where

$$I_1 = \left|\int\limits_{b_n}^{\infty} \ln u dF_x(u)\right|,$$

$$I_2 = \left|\int\limits_0^1 \ln u dF_{n,x}(u) - \int\limits_0^1 \ln u dF_x(u)\right|,$$

$$I_3 = \left|\int\limits_1^{b_n} \ln u dF_{n,x}(u) - \int\limits_1^{b_n} \ln u dF_x(u)\right|.$$

We estimate these terms separately.

Estimation of I_1. By definition

$$I_1 = \left|\int\limits_{b_n f(x)/\gamma}^{\infty} \ln(t\gamma/f(x))e^{-t}dt\right|.$$

Consider the two cases: 1) $b_n f(x)/\gamma < 1$, and 2) $b_n f(x)/\gamma \geq 1$. In the first case

$$I_1 = \left|\int\limits_{\frac{b_n f(x)}{\gamma}}^{\infty} \ln t e^{-t}dt\right| \leq \int\limits_0^{\infty} |\ln t| e^{-t}dt \leq C.$$

In the second case

$$I_1 = \left|\int\limits_{\frac{b_n f(x)}{\gamma}}^{\infty} \ln t e^{-t}dt\right| \leq \int\limits_{\frac{b_n f(x)}{\gamma}}^{\infty} e^{-t/2}dt = 2\exp(-b_n f(x)/\gamma).$$

Since also $I\{b_n f(x)/\gamma < 1\} \leq \exp(1 - b_n f(x)/\gamma)$ we have

$$\begin{aligned} I_1 &\leq CI\{b_n f(x)/\gamma < 1\} + 2\exp(-b_n f(x)/\gamma) \\ &\leq C\exp(-b_n f(x)/\gamma). \end{aligned} \tag{3.3}$$

Estimation of I_2.

$$\begin{aligned} I_2 &= \left|\int\limits_0^1 \ln u dF_{n,x}(u) - \int\limits_0^1 \ln u dF_x(u)\right| \\ &\leq \int\limits_0^1 |\ln u|\, |f_{n,x}(u) - f_x(u)| du \end{aligned} \tag{3.4}$$

where $f_{n,x}$ and f_x are the densities of $F_{n,x}$ and F_x respectively. We have

$$f_x(u) = \frac{f(x)}{\gamma}\exp(-f(x)u/\gamma),$$
$$f_{n,x}(u) = \frac{d}{du}F_{n,x}(u) = \frac{1}{2\gamma}S_{n,x}^{n-2}(u)\Delta_{n,x}(u)$$

where

$$S_{n,x}(u) = \left(1 - \int\limits_{x-r_n(u)}^{x+r_n(u)} f(y)dy\right), \Delta_{n,x}(u) = f(x+r_n(u)) + f(x-r_n(u)).$$

By (A1) we get

$$|f_{n,x}(u) - \overline{f}_{n,x}(u)| \le C\frac{u^2}{(n-1)^2} \tag{3.5}$$

where

$$\overline{f}_{n,x}(u) = \frac{f(x)}{\gamma}S_{n,x}^{n-2}(u).$$

Now we evaluate the rate of convergence of $S_{n,x}^{n-2}(u)$ to $\exp\left(-\frac{f(x)}{\gamma}u\right)$.

Denote

$$\alpha_n = (n-1)\left(\int\limits_{x-r_n(u)}^{x+r_n(u)} f(y)dy - \frac{f(x)u}{\gamma(n-1)}\right).$$

Since f is twice differentiable

$$|\alpha_n| \le C(n-1)r_n^3(u), \quad \forall\ u. \tag{3.6}$$

Denote $\Delta = \frac{1}{n-1}$, $\quad b = f(x)u/\gamma + \alpha_n$. Using the inequality

$$\left|\exp(\frac{1}{\Delta}\ln(1-b\Delta)) - \exp(-b)\right| \le Cb^2|\Delta|$$

(which is true for $b > 0$ and Δ small enough) we get that for n large

$$\left|S_{n,x}^{n-1}(u) - \exp\left(-\left(\frac{f(x)u}{\gamma} + \alpha_n\right)\right)\right| \le C/n. \tag{3.7}$$

Also, by (3.6)

$$\left|\exp\left(-\left(\frac{f(x)u}{\gamma} + \alpha_n\right)\right) - \exp\left(-\frac{f(x)u}{\gamma}\right)\right| \le \tag{3.8}$$
$$\le \left|\exp(-\alpha_n) - 1\right| \le C(n-1)r_n^3(u).$$

From (3.7) and (3.8) we obtain

$$(3.9) \qquad \left|S^{n-1}_{n,x}(u) - \exp\left(-\frac{f(x)u}{\gamma}\right)\right| \le C\left(nr_n^3(u) + \frac{1}{n}\right)$$

where C does not depend on x. The same is true if we change the exponent $n-1$ in (3.9) into $n-2$. Thus

$$(3.10) \qquad \left|\overline{f}_{n,x}(u) - \exp\left(-\frac{f(x)}{\gamma}u\right)\frac{f(x)}{\gamma}\right| \le C\left(nr_n^3(u) + \frac{1}{n}\right).$$

Combining (3.5) and (3.10) and using the definition of $f_x(u)$ we get

$$\left|f_x(u) - f_{n,x}(u)\right| \le C\left(\frac{u^2}{n^2} + nr_n^3(u) + \frac{1}{n}\right) \le Cn^{-1}$$

for $u \le 1$. Substituting this into (3.4) we find

$$(3.11) \qquad I_2 \le \int_0^1 |\ln u| du \cdot \frac{C}{n} \le \frac{C}{n}.$$

Estimation of I_3. Note that

$$\int_1^{b_n} \ln u dF_{n,x}(u) = \int_1^{b_n} \frac{1}{u}\left(1 - F_{n,x}(u)\right) du,$$

$$\int_1^{b_n} \ln u dF_x(u) = \int_1^{b_n} \frac{1}{u}\left(1 - F_x(u)\right) du.$$

Therefore,

$$(3.12) \qquad I_3 \le \int_1^{b_n} \frac{1}{u}\left|F_{n,x}(u) - F_x(u)\right| du.$$

To estimate the difference $\left|F_{n,x}(u) - F_x(u)\right|$ we use the following simple fact: if $\frac{v}{n-1} \le v_0$ where $v_0 > 0$ is a constant then there exists $C = C(v_0)$ such that

$$(3.13) \qquad \left|\left(1 - \frac{v}{n-1}\right)^{n-1} - \exp(-v)\right| \le C\exp(-v)\frac{v^2}{n-1}.$$

Denote $v = (n-1) \int\limits_{x-r_n(u)}^{x+r_n(u)} f(t)dt$. If $u \le b_n$ then $\frac{v}{n-1} \le Cr_n(u) \le v_0$ for some constant v_0. For this choice of v (3.13) gives

$$\begin{aligned}
&\left|S_{n,x}^{n-1}(u) - \exp(-v)\right| \le \\
&\le C\exp(-v)\frac{\left((n-1)\int\limits_{x-r_n(u)}^{x+r_n(u)} f(t)dt\right)^2}{n-1} \\
&\le C\exp(-v)\frac{1}{n-1}\left(\left(\frac{f(x)u}{\gamma}\right)^2 + |\alpha_n|^2\right) \\
&\le C\exp(-v)\frac{1}{n-1}\left(\left(\frac{f(x)u}{\gamma}\right)^2 + (n-1)r_n^3(u)\right)
\end{aligned} \tag{3.14}$$

where we used (3.6). Also, by (3.6)

$$\begin{aligned}
&\left|\exp(-v) - \exp\left(-\frac{f(x)u}{\gamma}\right)\right| \le \\
&\le \exp\left(-\frac{f(x)u}{\gamma}\right)\left|\exp\left(Cr_n^3(u)(n-1)\right) - 1\right| \\
&\le \frac{C}{\sqrt{n}}\exp\left(-\frac{f(x)u}{\gamma}\right), \quad \text{if } u \le b_n.
\end{aligned} \tag{3.15}$$

From (3.14), (3.15) and the definitions of F_x, $F_{n,x}$ we find

$$\begin{aligned}
&\left|F_{n,x}(u) - F_x(u)\right| \le \\
&\le C\exp\left(-\frac{f(x)u}{\gamma}\right)\left\{\frac{1}{\sqrt{n}} + \frac{1}{n-1}\left[\left(\frac{f(x)u}{\gamma}\right)^2 + (n-1)r_n^3(u)\right]\right\} \\
&\le C\exp\left(-\frac{f(x)u}{\gamma}\right)\left[\frac{1}{\sqrt{n}} + \left(\frac{f(x)u}{\gamma}\right)^2\frac{1}{n-1}\right]
\end{aligned} \tag{3.16}$$

if $u \le b_n$. Substitute (3.16) into (3.12). Then we obtain

$$\begin{aligned}
I_3 &\le \frac{C}{\sqrt{n}}\int\limits_1^{b_n} \exp\left(-\frac{f(x)u}{\gamma}\right)\left(1 + \frac{f(x)u}{\gamma}\right)\frac{du}{u} \\
&\le \frac{C}{\sqrt{n}}\left[\int\limits_{f(x)}^{\infty} \exp(-u'/\gamma)\frac{du'}{u'} + \int\limits_{f(x)}^{\infty} \exp(-u'/\gamma)du'\right] \\
&\le \frac{C}{\sqrt{n}}\left(|\ln f(x)| + 1\right).
\end{aligned} \tag{3.17}$$

Now we use (3.2), (3.3), (3.11), (3.17) and (A0) to obtain the following evaluation of the bias:

$$|H(f) - E(H_n)| \leq C \left[\int f(x) \exp(-b_n f(x)/\gamma) dx + \frac{1}{\sqrt{n}} \right].$$

This proves the Theorem since in view of (A2) the last integral does not exceed Cb_n^{-1}.

4. Proof of Theorem 2.

Consider the joint conditional distributions

$$F^{(1)}_{n,x,y}(u,w) = P\{e^{\zeta_i} < u,\ e^{\zeta_j} < w \mid X_i = x,\ X_j = y\},$$

$$\begin{aligned} F_{n,x,y}(u,w) &= P\{e^{\zeta_i'} < u,\ e^{\zeta_j'} < w \mid X_i = x,\ X_j = y\} \\ &= 1 - P\{\min_{k \neq j} |y - X_k| \geq r_n(w)\} - P\{\min_{k \neq j} |x - X_k| \geq r_n(u)\} + \\ &\quad + P\{\min_{k \neq i} |x - X_k| \geq r_n(u),\ \min_{k \neq j} |y - X_k| \geq r_n(w)\}. \end{aligned}$$

Since $e^{\zeta_i} \leq 2\rho_i \gamma(n-1) \leq b_n$ we have

$$F^{(1)}_{n,x,y}(u,w) = F_{n,x,y}(u,w), \quad u, w \leq b_n, \tag{4.1}$$

and $F^{(1)}_{n,x,y}(u,w) = 1$, when $u > b_n$ or $w > b_n$.

Note that

$$\begin{aligned} &F_{n,x,y}(u,w) = 1 \text{ if } |x-y| < \min(r_n(u), r_n(w)), \\ &F_{n,x,y}(u,w) = \begin{cases} 1 - F_{n,x}(u) \text{ if } |x-y| < r_n(w), \\ 1 - F_{n,y}(w) \text{ if } |x-y| < r_n(u), \end{cases} \end{aligned} \tag{4.2}$$

and, if $|x-y| \geq \max(r_n(u), r_n(w))$ then

$$F_{n,x,y}(u,w) = 1 - F_{n,x}(u) - F_{n,y}(w) + (1 - \int_A f(t)dt)^{n-2} \tag{4.3}$$

where $A = \{t : |x-t| < r_n(u)\} \cup \{t : |y-t| < r_n(w)\}$. Furthermore

$$E(H_n - E(H_n))^2 = \frac{1}{n^2} \sum_{i,j=1}^{n} \left(E(\zeta_i \zeta_j) - E^2(\zeta_i) \right). \tag{4.4}$$

To prove the Theorem it suffices to show

$$E(\zeta_i\zeta_j) - E^2(\zeta_i) = O\left(\frac{1}{n}\right), \tag{4.5}$$

as $n \to \infty$, uniformly in $i, j \in \{1, \dots, n\}$.

Using (4.1) we get

$$E(\zeta_i\zeta_j \mid X_i = x,\ X_j = y) = \int_0^{b_n}\int_0^{b_n} \ln w \ln u dF_{n,x,y}(u,w),$$

and

$$E(\zeta_i\zeta_j) = \int\int[\int_0^{b_n}\int_0^{b_n} \ln w \ln u dF_{n,x,y}(u,w)]f(x)f(y)dxdy. \tag{4.6}$$

Together with (3.1), (4.2), (4.3) this entails

$$\begin{aligned} E(\zeta_i\zeta_j) - E^2(\zeta_i) &= \\ &= -\int\int\left[\int_0^{b_n}\int_0^{b_n} \ln w \ln u dF_{n,x}(u)dF_{n,y}(w)\right. \\ &\quad \left. -\int_0^{b_n}\int_0^{b_n} \ln w \ln u dF_{n,x,y}(u,w)\right] f(x)f(y)dxdy \\ &= Q_1 + Q_2 + Q_3 \end{aligned} \tag{4.7}$$

where

$$Q_1 = \int_0^{b_n}\int_0^{b_n}\int_{\mathcal{D}(u,w)} \ln u \ln w f(x)f(y)dF_{n,x,y}(u,w)dxdy,$$

$$\begin{aligned} Q_2 = -\int_0^{b_n}\int_0^{b_n}\int_{|x-y|>r_n(u)+r_n(w)} & \ln u \ln w f(x)f(y)[dF_{n,x}(u)dF_{n,y}(w) \\ & -dF_{n,x,y}(u,w)]dxdy, \end{aligned}$$

$$Q_3 = -\int_0^{b_n}\int_0^{b_n}\int_{\mathcal{D}(u,w)} \ln u \ln w f(x)f(y)dF_{n,x}(u)dF_{n,y}(w)dxdy,$$

$$\mathcal{D}(u,w) = \{x, y : \max(r_n(u), r_n(w)) \le |x-y| \le r_n(u) + r_n(w)\}.$$

(Note that the inetgration is restricted to the region where $|x-y| > \max(r_n(u), r_n(w))$ in view of (4.2), (4.3)).

Let us estimate the terms $Q_j, \quad j = 1,2,3.$

Estimation of Q_1.

We have that if $x, y \in \mathcal{D}(u,w)$ then

$$F_{n,x,y}(u,w) = 1 - F_{n,x}(u) - F_{n,y}(w) + (1 - \int\limits_a^b f(t)dt)^{n-2}$$

where $a = \min\{y - r_n(w),\ x - r_n(u)\}$, $b = \max\{y + r_n(w),\ x + r_n(u)\}$, and there exist two possibilities: 1) $a = y - r_n(w)$, $b = x + r_n(u)$ or 2) $a = x - r_n(u)$, $b = y + r_n(w)$. Since both cases are symmetric we assume in the following that $a = y - r_n(w)$, $b = x + r_n(u)$, and we estimate only the part of Q_1 corresponding to this choice of a, b. For this part of Q_1 we have $z = y - x \geq 0$, and the density of $F_{n,x,y}$ has the form

$$f_{n,x,y}(u,w) =$$
$$= \frac{(n-3)(n-2)}{(n-1)^2}\frac{1}{(2\gamma)^2}\left(1 - \int\limits_{x-r_n(u)}^{y+r_n(w)} f(t)dt\right)^{n-4} f\left(y + r_n(w)\right) f\left(x - r_n(u)\right).$$

Thus, instead of Q_1 it is sufficient to estimate the integral

$$G = \int\limits_0^{b_n}\int\limits_0^{b_n} G(u,w)\ln u \ln w du dw$$

where

$$G(u,w) = \int\int\limits_{\mathcal{D}(u,w)} f(x)f(y)f(x - r_n(u))f(y + r_n(w))\times$$
$$\times\left(1 - \int\limits_{x-r_n(u)}^{y+r_n(w)} f(t)dt\right)^{n-4} dxdy \leq \int f(x)f(x - r_n(u))L(x)dx$$

where

$$L(x) = \int\limits_{0\leq z\leq r_n(u)+r_n(w)} f(x+z)f(x+z+r_n(w))\left(1 - \int\limits_{x-r_n(u)}^{x+z+r_n(w)} f(t)dt\right)^{n-4} dz.$$

Denote $f_{1n}(x) = \sup\limits_{|z|\le 3a_n} f(x+z)$. By the Lipschitz condition of f

$$(4.8)\qquad \int\limits_{x-r_n(u)}^{x+z+r_n(w)} f(t)dt \ge f(x)(z+r_n(w)+r_n(u)) - C(z+r_n(w)+r_n(u))^2$$

$$\ge f(x)(r_n(w)+r_n(u)) - C(r_n(w)+r_n(u))^2$$

since $0 \le z \le r_n(u)+r_n(w)$. From (4.8) we find (use $1-x \le e^{-x}$) that

$$(4.9)\qquad \begin{aligned} L(x) &\le Cf_{1n}^2(x)\exp(-(r_n(u)+r_n(w))f(x)(n-4))\times \\ &\quad \times \exp(C(n-4)(r_n(u)+r_n(w))^2)(r_n(u)+r_n(w)) \\ &\le C(r_n(u)+r_n(w))f_{1n}^2(x)\exp(-(r_n(u)+r_n(w))f(x)(n-4)) \end{aligned}$$

if $u, w \le b_n$.

This gives

$$|G| \le C\int\Bigg[\int\limits_0^{b_n}\int\limits_0^{b_n} |\ln u|\;|\ln w|(r_n(u)+r_n(w))\times$$

$$\times \exp(-(r_n(u)+r_n(w))f(x)(n-4))dudw\Bigg] f(x)f_{1n}^3(x)dx.$$

Using the definition of r_n and denoting $u' = uf_{2n}(x)$, $w' = wf_{2n}(x)$ we find

$$(4.10)\qquad |G| \le \frac{C}{n}\int\Bigg[\int\limits_0^{\infty}\int\limits_0^{\infty} |\ln(u'f_{2n}^{-1}(x))|\;|\ln(w'f_{2n}^{-1}(x))|(u'+w')\times$$

$$\times \exp(-C(u'+w'))du'dw'\Bigg] f(x)\frac{f_{1n}^3(x)}{f^3(x)}dx.$$

The integral here is finite in view of the following lemma which is a simple consequence of (B2).

Lemma 1. *Under the conditions of Theorem 2 the integrals*

$$\int f(x)\frac{f_{1n}^j(x)}{f^j(x)}dx,\quad \int f(x)\frac{f_{1n}^j(x)}{f^j(x)}|\ln f(x)|dx,$$

$$\int f(x)\frac{f_{1n}^j(x)}{f^j(x)}\ln^2 f(x)dx,\quad j=1,2,3,$$

are finite for n large enough.

It follows from (4.10) and from Lemma 1 that $G = O(1/n)$ as $n \to \infty$, and thus

$$Q_1 = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty. \tag{4.11}$$

Estimation of Q_2.

If $|x-y| > r_n(u) + r_n(w)$ then

$$F_{n,x,y}(u,w) = 1 - F_{n,x}(u) - F_{n,y}(w) + S_{n,x,y}^{n-2}(u,w)$$

where

$$S_{n,x,y}(u,w) = (1 - \int\limits_{x-r_n(u)}^{x+r_n(u)} f(t)dt - \int\limits_{y-r_n(w)}^{y+r_n(w)} f(t)dt),$$

$$dF_{n,x,y}(u,w) = \frac{(n-2)(n-3)}{(2\gamma(n-1))^2}\Delta_{n,y}(w)\Delta_{n,x}(u)S_{n,x,y}^{n-4}(u,w)dudw.$$

Thus,

$$\begin{aligned}
&dF_{n,x}(u)dF_{n,y}(w) - dF_{n,x,y}(u,w) = \\
&= \frac{1}{(2\gamma)^2}\Delta_{n,x}(u)\Delta_{n,y}(w)\left[S_{n,x}^{n-2}(u)S_{n,y}^{n-2}(w) - \frac{(n-2)(n-3)}{(n-1)^2}S_{n,x,y}^{n-4}(u,w)\right]dudw \\
&= \left[G_1(u,w) + G_2(u,w) + G_3(u,w)\right]dudw,
\end{aligned}$$

where

$$\begin{aligned}
G_1(u,w) &= \frac{1}{(2\gamma)^2}\Delta_{n,x}(u)\Delta_{n,y}(w)\left[S_{n,x}^{n-2}(u)S_{n,y}^{n-2}(w) - S_{n,x,y}^{n-2}(u,w)\right], \\
G_2(u,w) &= \frac{1}{(2\gamma)^2}\left(1 - \frac{(n-2)(n-3)}{(n-1)^2}\right)\Delta_{n,x}(u)\Delta_{n,y}(w)S_{n,x,y}^{n-2}(u,w), \\
G_3(u,w) &= \frac{1}{(2\gamma)^2}\,\Delta_{n,x}(u)\Delta_{n,y}(w)\left(S_{n,x,y}^2(u,w) - 1\right)S_{n,x,y}^{n-4}(u,w),
\end{aligned}$$

and

$$\begin{aligned}
Q_2 &= -\sum_{j=1}^{3}\int\limits_0^{b_n}\int\limits_0^{b_n} \ln u \ln w \left[\int\limits_{|x-y|>r_n(u)+r_n(w)} G_j(u,w)f(x)f(y)dxdy\right]dudw \\
&= -\sum_{j=1}^{3} Q_{2j}.
\end{aligned}$$

Let us estimate Q_{21} first. Consider $G_1(u,w)$. Note that since for every $\alpha, \beta \geq 0$

$$\begin{aligned} 0 &\leq (1-(\alpha+\beta)+\alpha\beta)^{n-2} - (1-(\alpha+\beta))^{n-2} \\ &\leq (n-2)\alpha\beta\exp(-(n-1)(\alpha+\beta-\alpha\beta)) \end{aligned}$$

we have

$$\begin{aligned} &0 \leq S^{n-2}_{n,x}(u)S^{n-2}_{n,y}(w) - S^{n-2}_{n,x,y}(u,w) \leq \\ &\leq (n-2)\int\limits_{x-r_n(u)}^{x+r_n(u)} f(t)dt \int\limits_{y-r_n(w)}^{y+r_n(w)} f(t)dt \times \\ &\quad \times \exp\Bigg[-(n-1)\Bigg(\int\limits_{x-r_n(u)}^{x+r_n(u)} f(t)dt + \int\limits_{y-r_n(w)}^{y+r_n(w)} f(t)dt - \\ &\quad - \int\limits_{x-r_n(u)}^{x+r_n(u)} f(t)dt \int\limits_{y-r_n(w)}^{y+r_n(w)} f(t)dt\Bigg)\Bigg] \leq \\ &\leq C\frac{(n-2)uw}{(n-1)^2} f_{1n}(x)f_{1n}(y)\exp\left(-\frac{1}{\gamma}(f(x)u+f(y)w)\right) \end{aligned} \tag{4.12}$$

where we used the fact that $u, w \leq b_n$ and that f is Lipschitz continuous. Also

$$0 \leq \Delta_{n,x}(u) \leq 2f_{1n}(x). \tag{4.13}$$

By (4.12), (4.13) we get

$$0 \leq G_1(u,w) \leq \frac{Cuw}{n} f^2_{1n}(x)f^2_{1n}(y)\exp\left(-\frac{1}{\gamma}(f(x)u+f(y)w)\right).$$

Thus

$$\begin{aligned} |Q_{21}| &\leq \frac{C}{n}\int\Bigg[\int\limits_0^{b_n}\int\limits_0^{b_n} |\ln u|\,|\ln w| uw \times \\ &\quad \times \exp\left(-\frac{1}{\gamma}(f(x)u+f(y)w)\right)dudw\Bigg] f^2_{1n}(x)f^2_{1n}(y)f(x)f(y)dxdy \\ &\leq \frac{C}{n}\int\Bigg[\int\limits_0^{\infty} |\ln u| u\exp\left(-\frac{1}{\gamma}f(x)u\right)du\Bigg] \times \\ &\quad \times \Bigg[\int\limits_0^{\infty} |\ln w| w\exp\left(-\frac{1}{\gamma}f(y)w\right)dw\Bigg] f^2_{1n}(x)f^2_{1n}(y)f(x)f(y)dxdy. \end{aligned} \tag{4.14}$$

Changing variables: $u' = uf_{2n}(x)$, $w' = wf_{2n}(x)$, and using Lemma 1 we obtain

$$Q_{21} = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty. \tag{4.15}$$

Let us estimate Q_{22} now. We have

$$0 \le S^{n-2}_{n,x,y}(u,w) \le \exp\biggl(-\frac{n-2}{\gamma(n-1)}(f(x)u + f(y)w)\biggr). \tag{4.16}$$

These inequalities together with (4.13) give

$$|Q_{22}| \le \frac{C}{n}\int\biggl[\int\limits_0^{b_n}\int\limits_0^{b_n} |\ln u|\, |\ln w| uw \times$$
$$\times \exp\biggl(-\frac{1}{\gamma}(f(x)u + f(y)w)\biggr)dudw\biggr] f_{1n}(x)f_{1n}(y)f(x)f(y)dxdy.$$

Proceeding as in (4.14), (4.15) we prove that

$$Q_{22} = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty. \tag{4.17}$$

To estimate the term Q_{23} we use a similar argument: apply (4.13), (4.16) and the inequality

$$0 \le S^2_{n,x,y}(u,w) - 1 \le \frac{1}{\gamma^2(n-1)^2}(f_{1n}(x)u + f_{1n}(y)w)^2 + \frac{2}{\gamma(n-1)}(f_{1n}(x)u +$$
$$f_{1n}(y)w)) + \frac{2}{\gamma(n-1)}(f_{1n}(x)u + f_{1n}(y)w).$$

Thus, we get

$$Q_{23} = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty. \tag{4.18}$$

Combining (4.15), (4.17) and (4.18) we get

$$Q_2 = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty. \tag{4.19}$$

Estimation of Q_3

We have

$$|Q_3| = \left|\int_0^{b_n}\int_0^{b_n}\int_{\mathcal{D}(u,w)} \ln u \ln w f(x)f(y)f_{n,x}(u)f_{n,y}(w)dxdydudw\right| \le$$

$$\le \frac{1}{(2\gamma)^2}\int f(x)dx \int_{x-r_n(u)-r_n(w)}^{x+r_n(u)+r_n(w)} f(y)dy\times$$

$$\times \int_0^{b_n}\int_0^{b_n} |\ln u|\,|\ln w| S^{n-2}_{n,x}(u)S^{n-2}_{n,y}(w)\Delta_{n,x}(u)\Delta_{n,y}(w)dudw,$$

since on $\mathcal{D}(u,w)$ we have

$$x - r_n(u) - r_n(w) \le y \le x + r_n(u) + r_n(w).$$

Also by this reason on $\mathcal{D}(u,w)$

$$\Delta_{n,y}(w) \le 2\max_{|t|\le a_n} f(y+t) \le 2\max_{|t|\le 3a_n} f(x+t) = 2f_{1n}(x), \tag{4.20}$$

and

$$0 \le S^{n-2}_{n,x,y}(u,w) \le \exp\left(-\frac{n-2}{\gamma(n-1)}(u+w)f(x)\right). \tag{4.21}$$

In fact,

$$\int_{y-r_n(w)}^{y+r_n(w)} f(t)dt \ge 2f(x)r_n(w) - \int_{y-r_n(w)}^{y+r_n(w)} |f(t)-f(x)|dt \ge$$

$$\ge \frac{w}{\gamma(n-1)}f(x) - Ca_n^2.$$

Applying (4.13), (4.16), (4.20), (4.21) we find

$$|Q_3| \le C\int f(x)f_{1n}^2(x)dx \int_{x-r_n(u)-r_n(w)}^{x+r_n(u)+r_n(w)} f(y)dy\times$$

$$\times \int_0^{b_n}\int_0^{b_n} |\ln u|\,|\ln w| \exp(-\frac{1}{\gamma}f(x)(u+w))dudw \le$$

$$\le \frac{C}{n}\int f(x)f_{1n}^3(x)dx \int_0^{b_n}\int_0^{b_n}(u+w)|\ln u|\,|\ln w|\exp(-\frac{1}{\gamma}f(x)(u+w))dudw.$$

Now we act as in (4.14), (4.15) and use Lemma 1 to prove that the integral is finite. Thus,

$$Q_3 = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty. \tag{4.22}$$

In view of (4.7), (4.11), (4.19) and (4.22) the proof of Theorem 2 is complete.

5. Conclusion

In this paper we established the mean-squared $\sqrt{n}$-consistency of entropy estimators for a class of densities including those with unbounded support. The method of the paper can be generalized to the case of multivariate densities as mentioned in the introduction. This will be the subject of our future work.

Acknowledgement

The authors would like to thank Prof. L. Györfi for valuable discussion on the subject of the paper.

References:

Beirlant, J. (1986), Limit theory for spacings of order statistics from general univariate distribution, *Publ. Inst. Stat. Univ. Paris*, XXXI, fasc. 1, 27–57.

Beirlant, J. and van Zuijlen, M.C.A. (1985), The empirical distribution function and strong laws for functions of order statistics of uniform spacings, *Journal of Multivariate Analysis*, 16, 300–317.

Bickel, P.J. and Breiman, L. (1983), Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test, *Annals of Probability*, 11, 185-214.

Cressie, N. (1976), On the logarithms of high order spacings, *Biometrika*, 63, 343–355.

Dobrushin, R.L. (1958), The simplified method of experimental estimation of the entropy of stationary sequence, *Theory Prob. Appl.* 3, 462–464.

Dudewicz, E.J. and van der Meulen, E.C. (1981), Entropy-based tests of uniformity, *Journal of American Stat. Assoc.*, 76, 967–974.

Dudewicz, E.J. and van der Meulen, E.C. (1987), The empiric entropy, a new approach to nonparametric entropy estimation, in Puri M.L. et al. eds., *New Perspectives in Theoretical and Applied Statistics*, Wiley, N.Y., 207–227.

Györfi, L. and van der Meulen, E.C. (1987), Density-free convergence properties of various estimators of entropy, *Computational Statistics and Data Analysis*, 5, 425–436.

Györfi, L. and van der Meulen, E.C. (1990), An entropy estimate based on a kernel density estimation, in *Colloquia Mathematica Societatis János Bolyai*, vol. 57: *Limit Theorems in Probability and Statistics*, 229–240.

Hall, P. (1984), Limit theorems for sums of general functions of m-spacings, *Math. Proc. Comb. Soc.*, 96, 517–532.

Hall, P. (1986), On powerful distributional tests based on sample spacings, *Journal of Multivariate Analysis*, 19, 201–225.

Khashimov Sh.A. (1989), Asymptotic properties of functions of spacings, *Theor. Prob. Appl.*, 34, 298–307.

Kozachenko, L.F. and Leonenko, N.N. (1987), On statistical estimation of entropy of a random vector, *Probl. of Information Transmission*, 23, 95–101.

Levit, B.Ya. (1978), Asymptotically efficient estimation of nonlinear functionals, *Probl. of Information Transmission*, 14, 204–209.

Tarasenko, F.P. (1968), 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.

van Es, B. (1992), Estimating functionals related to a density by a class of statistics based on spacings. *Scand. J. Statist.*, 19, 61–72.

Vasicek, O. (1976), A test for normality based on entropy, *Journal Royal Statist. Soc.* Ser. B, 38, 54–59.