A CLASS OF RÉNYI INFORMATION ESTIMATORS FOR MULTIDIMENSIONAL DENSITIES

By Nikolai Leonenko, Luc Pronzato and Vippal Savani

Cardiff University, UK and CNRS/Université de Nice–Sophia Antipolis, France

A class of estimators of the Rényi and Tsallis entropies of an unknown distribution f in $\mathbb{R}^m$ is presented, based on k-th nearest-neighbor distances in a sample of N i.i.d. vectors distributed with f. We show that entropies of any order q can be estimated consistently with minimal assumptions on f. The method can be extended straightforwardly to the estimation of the statistical distance between two distributions using one i.i.d. sample from each.

1. Introduction. We consider the problem of estimating the Rényi (1961) entropy

$$H_q^* = \frac{1}{1-q} \log \int_{\mathbb{R}^m} f^q(x)\, dx\,,\ q \neq 1\,, \tag{1.1}$$

or the Havrda-Charvát (1967) entropy (also called Tsallis (1988) entropy)

$$H_q = \frac{1}{q-1} \left(1 - \int_{\mathbb{R}^m} f^q(x)\, dx \right)\,,\ q \neq 1\,, \tag{1.2}$$

of a random vector $X \in \mathbb{R}^m$ with probability measure μ having a density f with respect to the Lebesgue measure, from N independent and identically distributed (i.i.d.) samples $X_1, \ldots, X_N$, $N \geq 2$. When q tends to 1, both H_q and H_q^* tend to the (Boltzmann-Gibbs-) Shannon entropy

$$H_1 = -\int_{\mathbb{R}^m} f(x)\, \log f(x)\, dx\,. \tag{1.3}$$

We consider a new class of estimators of H_q and H_q^* based on the approach proposed by Kozachenko and Leonenko (1987) for the estimation of H_1, see also Goria et al. (2005). Within the classification made in (Beirlant et al., 1997), which also contains an outstanding overview of nonparametric Shannon entropy estimation, the method

AMS 2000 subject classifications. 94A15, 62G20.

Key words and phrases. entropy estimation, estimation of statistical distance, estimation of divergence, nearest-neighbor distances, Rényi entropy, Havrda-Charvát entropy, Tsallis entropy.

falls in the category of nearest-neighbor distances. When $m = 1$, it is related to sample-spacing methods, see, e.g., Vasicek (1976) for an early reference concerning Shannon entropy. It also has some connections with the more recent random-graph approach of Redmond and Yukich (1996), see also Hero and Michel (1999). For $q \neq 1$ the method relies on the estimation of the integral

$$I_q = \int_{\mathbb{R}^m} f^q(x)\,dx \tag{1.4}$$

through the computation of conditional moments of nearest-neighbor distances, and thus possesses some similarities with Evans et al. (2002). The method is, however, original by several aspects that will be enhanced below, see Section 5. Note that $I_1 = 1$ since f is a p.d.f. and that I_q is finite when $q < 0$ only if f is of bounded support. Indeed, $I_q = \int_{\{x:f(x)\geq 1\}} f^q(x)dx + \int_{\{x:f(x)<1\}} f^q(x)dx > \int_{\{x:f(x)<1\}} f^q(x)dx > \mu_{\mathcal{L}}\{x : f(x) < 1\}$, with $\mu_{\mathcal{L}}$ the Lebesgue measure. Also, when f is bounded, I_q tends to the (Lebesgue) measure of its support $\mu_{\mathcal{L}}\{x : f(x) > 0\}$ when $q \to 0^+$. Some other properties of I_q are summarized in Lemma 1 of Section 6. For the sake of completeness, we also consider the case $q = 1$, that is, the estimation of Shannon entropy, with results obtained as corollaries of those for $q \neq 1$ (at the expense of requiring slightly stronger conditions than in (Kozachenko and Leonenko, 1987)). Notice that H_q^* can be expressed as a function of H_q, $H_q^* = \log[1-(q-1)H_q]/(1-q)$, with, for any q, $d(H_q^*)/d(H_q) = 1/I_q$ and $d^2(H_q^*)/d(H_q)^2 = (q-1)/I_q^2$. H_q^* is thus a strictly increasing concave (respectively convex) function of H_q for $q < 1$ (respectively $q > 1$) and the maximizations of H_q^* and H_q are equivalent. Hence in what follows we shall speak indifferently of q-entropy maximizing distributions.

The entropy (1.2) is of interest in the study of nonlinear Fokker-Planck equations, with $q < 1$ for the case of subdiffusion and $q > 1$ for superdiffusion, see Tsallis and Bukman (1996). Values of $q \in [1, 3]$ are used in (Alemany and Zanette, 1992) to study the behavior of fractal random walks. Applications for quantizer design, characterization of time-frequency distributions, image registration and indexing, texture classification, image matching etc., are indicated in (Hero and Michel, 1999; Hero et al., 2002; Neemuchwala et al., 2005). Entropy minimization is used in (Pronzato et al., 2004; Wolsztynski et al., 2005) for parameter estimation in semi-parametric models. Entropy estimation is a basic tool for independent component analysis in signal processing, see, e.g., (Miller and Fisher, 2003). The entropy H_q is a concave function of the density for $q > 0$ (and convex for $q < 0$). Hence, q-entropy maximizing distributions, under some specific constraints, are uniquely defined for

$q > 0$. For instance, when the constraint is that the distribution is finitely supported, then the q-entropy maximizing distribution is uniform. More interestingly, for any dimension $m \geq 1$ the q-entropy maximizing distribution with a given covariance matrix is of the multidimensional Student-t type if $m/(m+2) < q < 1$, see Vignat et al. (2004). This generalizes the well-known property that Shannon entropy H_1 is maximized for the normal distribution. Such entropy-maximization properties can be used to derive nonparametric statistical tests, following the same approach as in (Vasicek, 1976) where normality is tested with H_1; see also Goria et al. (2005).

Section 2 develops some of the motivations and applications just mentioned (see also Section 3.3 for signal and image processing applications). The main results of the paper are presented in Section 3. The paper is focussed on entropy estimation, but in Section 3.3 we show how a slight modification of the method also permits to estimate statistical distances and divergences between two distributions. Section 4 gives some examples and Section 5 indicates some related results and possible developments. The proofs of the results of Section 3 are collected in Section 6.

2. Properties, motivation and applications.

2.1. *Nonlinear Fokker-Planck equation and entropy.* Consider a family of time-dependent p.d.f. f_t. The p.d.f. that maximizes Réyi entropy (1.1) (and Tsallis entropy (1.2)) subject to the constraints $\int_{\mathbb{R}} f_t(x)\,dx = 1$, $\int_{\mathbb{R}} [x - \bar{x}(t)]\, f_t^q(x)\,dx = 0$, $\int_{\mathbb{R}} [x - \bar{x}(t)]^2\, f_t^q(x)\,dx = \sigma_q^2(t)$, for fixed $q > 1$, is

$$f_t(x) = \frac{1}{Z(t)}\, \frac{1}{\{1 + \beta(t)(q-1)[x - \bar{x}(t)]^2\}^{1/(q-1)}}\,, \quad x \in \mathbb{R}\,, \tag{2.5}$$

where

$$Z(t) = \frac{B\left(\frac{1}{2}, \frac{1}{q-1} - \frac{1}{2}\right)}{\sqrt{(q-1)\beta(t)}}\,, \quad \beta(t) = \frac{1}{2\sigma_q^2(t) Z^{q-1}(t)}$$

where $B(x, y)$ denotes the beta function, see Tsallis et al. (1995); Tsallis and Bukman (1996). Its variance is

$$\sigma_1^2(t) = \begin{cases} \frac{1}{(5-3q)\beta(t)} & \text{if } q < 5/3\,, \\ \infty & \text{if } q \geq 5/3\,; \end{cases}$$

hence, for applications with data of finite variance the parameter q satisfies $1 < q < 5/3$.

An important property of f_t given by (2.5) is that it is solution of the nonlinear Fokker-Planck (or Kolmogorov) equation

$$\frac{\partial}{\partial t} f_t(x) = -\frac{\partial}{\partial x}[f_t(x)F(x)] + D\,\frac{\partial^2}{\partial x^2} f_t^{2-q}(x) \tag{2.6}$$

when the driving term is $F(x) = a - bx$, see Tsallis and Bukman (1996). The time-dependent parameters then satisfy

$$\begin{aligned}
&-\frac{1}{3-q}\frac{dZ^{3-q}(t)}{dt} + 2(2-q)D\beta(t_0)Z^2(t_0) - bZ^{3-q}(t) = 0\,,\\
&\frac{\beta(t)}{\beta(t_0)} = \left[\frac{Z(t_0)}{Z(t)}\right]^2,\\
&\frac{d\bar{x}(t)}{dt} = a - b\,\bar{x}(t)\,,
\end{aligned}$$

or

$$\begin{aligned}
\beta^{(q-3)/2}(t) &= \beta^{(q-3)/2}(t_0)\exp[-b(3-q)(t-t_0)]\\
&\quad -2Db^{-1}(2-q)[\beta(t_0)Z^2(t_0)]^{(q-1)/2}\{\exp[-b(3-q)(t-t_0)] - 1\}\,.
\end{aligned}$$

In the limit, when $q \to 1$, the standard linear diffusion equation is recovered and the t-distribution (2.5) becomes Gaussian, while the Tsallis entropy becomes the Shannon entropy (1.3) and the constraints correspond to the classical Gaussian maximum entropy principle. Other entropy maximization results are presented in the next paragraph.

Remark 2.1. *Let X and Y be two independent random vectors respectively in $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$. Define $Z = (X, Y)$ and denote $f(x, y)$ the joint density for Z and $f_1(x)$, $f_2(y)$ the marginal densities for X and Y, so that $f(x, y) = f_1(x)f_2(y)$. Then it is well known that for the Shannon entropy (1.3) and the Rényi entropy (1.1) we have the following additive property:*

$$H_q^*(f) = H_q^*(f_1) + H_q^*(f_2)\,,\ q \in \mathbb{R}\,,$$

while for the Tsallis entropy (1.2) we obtain that

$$H_q(f) = H_q(f_1) + H_q(f_2) + (1-q)H_q(f_1)H_q(f_2)\,.$$

The first property is known in the physical literature as the extensivity property of Shannon and Rényi entropies, while the second is known as nonextensivity (with q the parameter of nonextensivity).

REMARK 2.2. *The paper (Frank and Daffertshofer, 2000) presents a survey of results related to entropies in connection with nonlinear Fokker-Planck equations and normal or anomalous diffusion processes. In particular, the so-called Sharma and Mittal entropy*

$$H_{q,s} = \frac{1}{1-s}\left[1 - (I_q)^{(s-1)/(q-1)}\right], \ q,s > 0, \ q,s \neq 1,$$

with I_q given by (1.4), represents a possible unification of the (nonextensive) Tsallis entropy (1.2) and (extensive) Rényi entropy (1.1). It satisfies the following properties:

$$\lim_{s\to 1} H_{q,s} = H_q^*, \ \lim_{s,q\to 1} H_{q,s} = H_1, \ H_{q,q} = H_q,$$

while

$$\lim_{q\to 1} H_{q,s} = H_s^G = \frac{1}{s-1}\{1 - \exp[-(s-1)H_1]\}, \ s > 0, \ s \neq 1,$$

where H_s^G is known as Gaussian entropy. Notice that a consistent estimator of $H_{q,s}$ can be obtained from the estimator of I_q presented in Section 3.

2.2. *Entropy maximizing distributions.* The m-dimensional random vector $X = ([X]_1, \ldots, [X]_m)^\top$ is said to have a multidimensional Student distribution $T(\nu, \Sigma, \mu)$ with mean $\mu \in \mathbb{R}^m$, scaling (correlation) matrix Σ (and covariance matrix $C = \nu\Sigma/(\nu-2)$) and ν degrees of freedom if its p.d.f. is

$$f_\nu(x) = \frac{1}{(\nu\pi)^{m/2}} \frac{\Gamma(\frac{m+\nu}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{|\Sigma|^{1/2}[1 + (x-\mu)^\top[\nu\Sigma]^{-1}[x-\mu)]^{(m+\nu)/2}}, \tag{2.7}$$

$x \in \mathbb{R}^m$. The characteristic function of the distribution $T(\nu, \Sigma, \mu)$ is

$$\phi(\zeta) = \mathbb{E}\exp(i\langle\zeta, X\rangle) = \exp(i\langle\zeta,\mu\rangle)\, K_{\nu/2}(\sqrt{\nu\zeta^\top\Sigma\zeta})\,(\sqrt{\nu\zeta^\top\Sigma\zeta})^{\nu/2}\,\frac{2^{1-\nu/2}}{\Gamma(\frac{\nu}{2})},$$

$\zeta \in \mathbb{R}^m$, where $K_{\nu/2}$ denotes the modified Bessel function of the second order. If $\nu = 1$, then (2.7) is the m-variate Cauchy distribution. If $(\nu+m)/2 = i$ integer, then (2.7) is the m-variate Pearson type VII distribution. If Y is $\mathcal{N}(0,\Sigma)$ and νS^2 is independent of Y and is $\mathcal{X}^2$-distributed with ν degrees of freedom, then $X = Y/s + \mu$ has the p.d.f. (2.7). The limiting form of (2.7) as $\nu \to \infty$ if the m-variate normal distribution $\mathcal{N}(\mu,\Sigma)$. The Rényi entropy (1.1) of (2.7) is

$$H_q^* = \frac{1}{1-q}\log\frac{B(\frac{q(m+\nu)}{2} - \frac{m}{2}, \frac{m}{2})}{B^q(\frac{\nu}{2}, \frac{m}{2})} + \frac{1}{2}\log[(\pi\nu)^m|\Sigma|] - \log\Gamma\left(\frac{m}{2}\right), \ q > \frac{m}{m+\nu}.$$

It converges as $\nu \to \infty$ to the Rényi entropy

$$H_q^*(\mu,\Sigma) = \log[(2\pi)^{m/2}|\Sigma|^{1/2}] - \frac{m}{2(1-q)}\,\log q = H_1(\mu,\Sigma) - \frac{m}{2}\left(1 + \frac{\log q}{1-q}\right) \tag{2.8}$$

of the multidimensional normal distribution $\mathcal{N}(\mu,\Sigma)$. When $q \to 1$, $H_q^*(\mu,\Sigma)$ tends to $H_1(\mu,\Sigma) = \log[(2\pi e)^{m/2}|\Sigma|^{1/2}]$, the Shannon entropy of $\mathcal{N}(\mu,\Sigma)$. For $m/(m+2) < q < 1$, the q-entropy maximizing distribution under the constraint

$$\mathbb{E}(X-\mu)(X-\mu)^\top = C \tag{2.9}$$

is the Student distribution $T(\nu,(\nu-2)C/\nu,0)$ with $\nu = 2/(1-q) - m > 2$, see Vignat et al. (2004). For $q > 1$, we define $p = m + 2/(q-1)$ and the q-entropy maximizing distribution under the constraint (2.9) has then finite support given by

$$\Omega_q = \{x \in \mathbb{R}^m : (x-\mu)^\top[(p+2)C]^{-1}(x-\mu) \le 1\}\,.$$

Its p.d.f. is

$$f_p(x) = \begin{cases} \frac{\Gamma(\frac{p}{2}+1)}{|C|^{1/2}[\pi(p+2)]^{m/2}\Gamma(\frac{p-m}{2}+1)} \\ \quad\times[1-(x-\mu)^\top[(p+2)C]^{-1}(x-\mu)]^{1/(q-1)} & \text{if } x \in \Omega_q \\ 0 & \text{otherwise}\,. \end{cases} \tag{2.10}$$

The characteristic function of the p.d.f. (2.10) is given by

$$\phi(\zeta) = \exp(i\langle\zeta,\mu\rangle)\,2^{p/2}\Gamma\left(\frac{p}{2}+1\right)|\zeta^\top(p+2)C\zeta|^{-p/2}J_{p/2}(|\zeta^\top(p+2)C\zeta|)\,,$$

$\zeta \in \mathbb{R}^m$, where $J_{\nu/2}$ denotes the Bessel function of the first kind.

Alternatively, f_ν for $q < 1$ or f_p for $q > 1$ also maximizes Shannon entropy (1.3) under a logarithmic constraint, see (Kapur, 1989; Zografos, 1999). Indeed, when $q < 1$ $f_\nu(x)$ given by (2.7) with $\nu = 2/(1-q) - m$ and $\Sigma = (\nu-1)C/\nu$ maximizes H_1 under the constraint

$$\int_{\mathbb{R}^m} \log(1 + x^\top[(\nu-2)C]^{-1}x)\,f(x)\,dx = \Psi\left(\frac{\nu+m}{2}\right) - \Psi\left(\frac{\nu}{2}\right)\,,$$

and when $q > 1$ $f_p(x)$ given by (2.10) with $p = 2/(q-1) + m$ maximizes H_1 under

$$\int_{\mathbb{R}^m} \log(1 - x^\top[(p+2)C]^{-1}x)\,f(x)\,dx = \Psi\left(\frac{p}{2}\right) - \Psi\left(\frac{p+m}{2}\right)\,,$$

where $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function.

2.3. *Information spectrum.* Considered as a function of q, H_q^* (1.1) is known as the spectrum of Rényi information, see Song (2001). The value for $q=2$ corresponds to the negative logarithm of the well-known efficacy parameter $I\!\!E f(X)$ that arises in asymptotic efficiency considerations. Consider now

$$\dot{H}_1 = \lim_{q\to 1} \frac{dH_q^*}{dq}\,. \tag{2.11}$$

It satisfies

$$\begin{aligned}\dot{H}_1 &= \lim_{q\to 1} \frac{\log \int_{\mathbb{R}^m} f^q(x)\,dx}{(1-q)^2} + \frac{\int_{\mathbb{R}^m} f^q(x)\,\log f(x)\,dx}{(1-q)\int_{\mathbb{R}^m} f^q(x)\,dx} \\ &= -\frac{1}{2}\left\{\int_{\mathbb{R}^m} f(x)\,[\log f(x)]^2\,dx - \left[\int_{\mathbb{R}^m} f(x)\,\log f(x)\,dx\right]^2\right\} \\ &= -\frac{1}{2}\mathrm{var}[\log f(X)]\,.\end{aligned}$$

The quantity $S_f = -2\dot{H}_1 = \mathrm{var}[\log f(X)]$ gives a measure of the intrinsic shape of the density f; it is a location and scale invariant positive functional,

$$S_f = S_g \quad \text{when} \quad f(x) = \sigma^{-1} g[(x-\mu)/\sigma]\,.$$

For the multivariate normal distribution $\mathcal{N}(\mu,\Sigma)$ H_q^* is given by (2.8) and $S_f = m/2$. For the one-dimensional Student distribution with ν degrees of freedom (for which $I\!\!E X^{\nu-1}$ exists, but not $I\!\!E X^{\nu}$), with density

$$f_\nu(x) = \frac{1}{(\nu\pi)^{1/2}}\,\frac{\Gamma(\frac{\nu}{2}+\frac{1}{2})}{\Gamma(\frac{\nu}{2})}\,\frac{1}{(1+x^2/\nu)^{(\nu+1)/2}}\,,$$

we obtain

$$H_q^* = \frac{1}{1-q}\,\log \frac{B(\frac{q(\nu+1)}{2}-\frac{1}{2},\frac{1}{2})}{B^q(\frac{\nu}{2},\frac{1}{2})} + \frac{1}{2}\,\log\nu\,,\ q > \frac{1}{\nu+1}\,, \tag{2.12}$$

$$S_{f_\nu} = \begin{cases} \frac{\pi^2}{3} \simeq 3.2899 & \text{for } \nu = 1 \text{ (Cauchy distribution)} \\ 9-\frac{3}{4}\pi^2 \simeq 1.5978 & \text{for } \nu = 2 \\ \frac{4}{3}\pi^2 - 12 \simeq 1.1595 & \text{for } \nu = 3 \\ \frac{775}{36} - \frac{25}{12}\pi^2 \simeq 0.9661 & \text{for } \nu = 4 \\ 3\pi^2 - \frac{115}{4} \simeq 0.8588 & \text{for } \nu = 5 \end{cases}$$

and more generally

$$S_{f_\nu} = \frac{1}{4}(\nu+1)^2\{\dot{\Psi}(\nu/2) - \dot{\Psi}[(\nu+1)/2]\}\,,$$

with $\dot{\Psi}(x)$ the trigamma function, $\dot{\Psi}(x) = d^2\log\Gamma(x)/dx^2$. Note that in terms of characterization of the shape of a distribution S_f is more interesting that, say, the kurtosis: for instance, the kurtosis is not defined for f_ν when $\nu \leq 4$; for $\nu = 6$, f_6 and the bi-exponential (Laplace) distribution f_L have the same kurtosis but different values of S_f, $S_{f_6} = 147931/3600 - (49/12)\pi^2 \simeq 0.7911$ and $S_{f_L} = 1$. For the multivariate Student distribution (2.7), we get

$$S_{f_\nu} = \frac{1}{4}(\nu+m)^2\{\dot\Psi(\nu/2) - \dot\Psi[(\nu+m)/2]\}\,.$$

The q-entropy maximizing property of the Student distribution can be used to test that the observed samples is Student distributed, and the estimation of S_f then provides information about ν, which for instance finds important applications in financial mathematics, see Heyde and Leonenko (2005).

3. Main results. Let $\rho(x,y)$ denote the Euclidian[1] distance between two points x, y of $\mathbb{R}^m$. For a given sample $X_1, \ldots, X_N$, and a given X_i in the sample, from the $N-1$ distances $\rho(X_i, X_j)$, $j = 1, \ldots, N$, $j \neq i$, we form the order statistics

$$\rho^{(i)}_{1,N-1} \leq \rho^{(i)}_{2,N-1} \leq \cdots \leq \rho^{(i)}_{N-1,N-1}\,.$$

Therefore, $\rho^{(i)}_{1,N-1}$ is the nearest-neighbor distance from the observation X_i to some other X_j in the sample, $j \neq i$, and similarly $\rho^{(i)}_{k,N-1}$ is the k-th nearest-neighbor distance from X_i to some other X_j.

3.1. *Rényi and Tsallis entropies.* We shall estimate I_q, $q \neq 1$, by

$$\hat I_{N,k,q} = \frac{1}{N}\sum_{i=1}^{N} (\zeta_{N,i,k})^{1-q}\,, \tag{3.13}$$

with

$$\zeta_{N,i,k} = (N-1)\, C_k\, V_m\, (\rho^{(i)}_{k,N-1})^m\,, \tag{3.14}$$

where $V_m = \pi^{m/2}/\Gamma(m/2+1)$ is the volume of the unit ball $\mathcal{B}(0,1)$ in $\mathbb{R}^m$ and

$$C_k = \left[\frac{\Gamma(k)}{\Gamma(k+1-q)}\right]^{1/(1-q)}\,.$$

Remark 3.1. *When f is known, a Monte-Carlo estimator of I_q based on the sample $X_1, \ldots, X_N$ is*

$$\frac{1}{N}\sum_{i=1}^{N} f^{q-1}(X_i)\,. \tag{3.15}$$

[1]See Section 5 for an extension to other metrics.

The nearest-neighbor estimator $\hat{I}_{N,k,q}$ *given by (3.13) could thus also be considered as a plug-in estimator,*

$$\hat{I}_{N,k,q} = \frac{1}{N} \sum_{i=1}^{N} [\hat{f}_{N,k}(X_i)]^{q-1}$$

where $\hat{f}_{N,k}(x) = 1/\{(N-1)C_k V_m [\rho_{k+1,N}(x)]^m\}$ *with* $\rho_{k+1,N}(x)$ *the* $(k+1)$*-th nearest-neighbor distance from* x *to the sample. One may notice the ressemblance between* $\hat{f}_{N,k}(x)$ *and the density function estimator* $\hat{\hat{f}}_{N,k}(x) = k/\{NV_m[\rho_{k+1,N}(x)]^m\}$ *suggested by Loftsgaarden and Quesenberry (1965); see also Moore and Yackel (1977); Devroye and Wagner (1977).*

We suppose that $X_1, \ldots, X_N$, $N \geq 2$, are i.i.d. with a probability measure μ having a density f with respect to the Lebesgue measure[2]. The main results of the paper are as follows.

THEOREM 3.1 (ASYMPTOTIC UNBIASEDNESS). *The estimator* $\hat{I}_{N,k,q}$ *given by (3.13) satisfies*

$$\mathbb{E}\hat{I}_{N,k,q} \to I_q \,, \; N \to \infty \,, \tag{3.16}$$

for $q < 1$ *provided that* I_q *given by (1.4) exists, and for any* $q \in (1, k+1)$ *if* f *is bounded.*

Under the conditions of Theorem 3.1, $\mathbb{E}(1 - \hat{I}_{N,k,q})/(q-1) \to H_q$ as $N \to \infty$, which provides an asymptotically unbiased estimate of the Tsallis entropy of f.

THEOREM 3.2 (CONSISTENCY). *The estimator* $\hat{I}_{N,k,q}$ *given by (3.13) satisfies*

$$\hat{I}_{N,k,q} \xrightarrow{L_2} I_q \,, \; N \to \infty \,, \tag{3.17}$$

(and thus $\hat{I}_{N,k,q} \xrightarrow{\text{p}} I_q$, $N \to \infty$*) for* $q < 1$ *provided that* I_{2q-1} *exists, and for any* $q \in (1, (k+1)/2)$ *when* $k \geq 2$ *(respectively* $q \in (1, 1 + 1/[2m])$ *when* $k = 1$*) if* f *is bounded.*

---

[2]However, if μ has a finite number of singular components superimposed to the absolutely continuous component f, one can remove all zero distances from the $\rho^{(i)}_{k,N-1}$ in the computation of the estimate (3.13), which then enjoys the same properties as in Theorems 3.1 and 3.2, i.e., yields a consistent estimator of the Rényi and Tsallis entropies of the continuous component f.

Corollary 3.1. *Under the conditions of Theorem 3.2,*

$$\hat{H}_{N,k,q} = (1 - \hat{I}_{N,k,q})/(q-1) \xrightarrow{L_2} H_q \tag{3.18}$$

and

$$\hat{H}^*_{N,k,q} = \log(\hat{I}_{N,k,q})/(1-q) \xrightarrow{\mathrm{p}} H^*_q \tag{3.19}$$

as $N \to \infty$, which provides consistent estimates of the Rényi and Tsallis entropies of f.

Remark 3.2. *We show the following in the proof of Theorem 3.2: when $q < 1$ and $I_{2q-1} < \infty$, or $1 < q < (k+2)/2$ and f is bounded,*

$$\mathbb{E}(\zeta_{N,i,k}^{1-q} - I_q)^2 \to \Delta_{k,q} = I_{2q-1}\frac{\Gamma(k+2-2q)\Gamma(k)}{\Gamma^2(k+1-q)} - I_q^2\,,\ N \to \infty\,.$$

Notice that

$$\lim_{k\to\infty} \Delta_{k,q} = I_{2q-1} - I_q^2 = \mathrm{var}[f^{q-1}(X)] = N \mathrm{var}\left[\frac{1}{N}\sum_{i=1}^N f^{q-1}(X_i)\right],$$

that is, the limit of $\Delta_{k,q}$ for $k \to \infty$ equals N times the variance of the Monte-Carlo estimator (3.15) (which forms a lower bound on the variance of an estimator I_q based on the sample $X_1, \ldots, X_N$).

3.2. *Shannon entropy.* For the estimation of H_1 ($q = 1$) we take the limit of $\hat{H}_{N,k,q}$ as $q \to 1$, which gives

$$\hat{H}_{N,k,1} = \frac{1}{N}\sum_{i=1}^N \log \xi_{N,i,k} \tag{3.20}$$

with

$$\xi_{N,i,k} = (N-1)\exp[-\Psi(k)]\, V_m\, (\rho_{k,N-1}^{(i)})^m \tag{3.21}$$

where $\Psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function. We then have the following.

Corollary 3.2. *Suppose that f is bounded and that I_{q_1} exists for some $q_1 < 1$. Then H_1 exists and the estimator (3.20) satisfies $\mathbb{E}\hat{H}_{N,k,1} \to H_1$ as $N \to \infty$.*

Corollary 3.3. *Suppose that f is bounded and that I_{q_1} exists for some $q_1 < 1$. Then the estimator (3.20) satisfies $\hat{H}_{N,k,1} \xrightarrow{L_2} H_1$ as $N \to \infty$.*

REMARK 3.3. *The factor* $N-1$ *in (3.14) and (3.21) can be replaced by any term* $C(N,k,q)$ *such that* $C(N,k,q) = N + o(N^\alpha)$, $N \to \infty$, $\alpha < 1$, *without modifying the asymptotic properties stated in Theorems 3.1-3.2 and Corollaries 3.1-3.3 (however, the finite sample behavior of the estimators (3.18-3.20) is affected by such a modification).*

REMARK 3.4. *We obtain the following in the proof of Corollary 3.3:*

$$\mathbb{E}(\log \xi_{N,i,k} - H_1)^2 \to \mathrm{var}[\log f(X)] + \dot{\Psi}(k)\,,\ N \to \infty\,,$$

with $\dot{\Psi}(z) = d^2 \log \Gamma(z)/dz^2$, *where* $\mathrm{var}[\log f(X)]$ *forms a lower bound on the variance of a Monte-Carlo estimation of* H_1 *based on* $\log f(X_i)$, $i = 1, \ldots, N$ *(note that* $\dot{\Psi}(k) \to 0$ *as* $k \to \infty$*).*

REMARK 3.5. *Notice that* $\hat{H}^*_{N,k,q}$ *given by (3.19) is a smooth function of* q. *Its derivative at* $q = 1$ *can be used as an estimate of* $\dot{H}_1$ *defined by (2.11). Straightforward calculations give*

$$\begin{aligned}\lim_{q\to 1} \frac{d\hat{H}^*_{N,k,q}}{dq} &= \frac{\dot{\Psi}(k)}{2} - \frac{m^2}{2}\,\frac{1}{N}\sum_{i=1}^{N}\left[\log \rho^{(i)}_{k,N-1} - \frac{1}{N}\sum_{j=1}^{N}\log \rho^{(j)}_{k,N-1}\right]^2 \\ &= \frac{1}{2}\left[\dot{\Psi}(k) - \frac{1}{N}\sum_{i=1}^{N}(\log \xi_{N,i,k} - \hat{H}_{N,k,1})^2\right]\end{aligned}$$

and $S_f = -2\dot{H}_1$ *can be estimated by*

$$\hat{S}_{f_{N,k}} = \frac{1}{N}\sum_{i=1}^{N}(\log \xi_{N,i,k} - \hat{H}_{N,k,1})^2 - \dot{\Psi}(k)\,. \tag{3.22}$$

3.3. *Relative entropy and divergences.* In some situations, the statistical distance between distributions can be estimated through the computation of entropies, so that the method of k-th nearest-neighbor distances presented above can be applied straightforwardly. For instance, the q-Jensen difference

$$J_q^\beta(f,g) = H_q^*[\beta f + (1-\beta)g] - [\beta H_q^*(f) + (1-\beta)H_q^*(g)]\,,\ 0 \le \beta \le 1\,,$$

see Basseville (1996), can be estimated if we have three samples, respectively distributed according to f, g and $\beta f + (1-\beta)g$. Suppose that we have one sample S_i, $i = 1, \ldots, s$, of i.i.d. variables generated from f and one sample T_j, $j = 1, \ldots, t$, of i.i.d. variables generated from g, with s, t increasing at a constant rate as a function of $N = s + t$. Then, $H_q^*(f)$ and $H_q^*(g)$ can be estimated consistently from the two samples when $n \to \infty$, see Corollary 3.1. Also, as $N \to \infty$, the estimator $\hat{H}^*_{N,k,q}$ based on the sample $X_1, \dots, X_N$ with $X_i = S_i$, $i = 1, \dots, s$, and $X_i = T_{i-s}$, $i = s+1, \dots, N$, converges to $H^*_q[\beta f + (1-\beta) g]$, with $\beta = s/N$, and J^β_q can therefore be estimated consistently from the two samples. This situation is encountered for instance in the image matching problem presented in (Neemuchwala et al., 2005), where entropy is estimated through the random graph approach of Redmond and Yukich (1996).

As shown below, some other types of distances or divergences, that are not expressed directly through entropies, can also be estimated by the nearest-neighbor method.

Let $K(f, g)$ denote the Kullback-Leibler relative entropy,

$$K(f,g) = \int_{\mathbb{R}^m} f(x) \log \frac{f(x)}{g(x)} \, dx = \breve{H}_1 - H_1 \,, \tag{3.23}$$

where H_1 is given by (1.3) and

$$\breve{H}_1 = - \int_{\mathbb{R}^m} f(x) \log g(x) \, dx \,.$$

Given N independent observations $X_1, \dots, X_N$ distributed with the density f and M observations $Y_1, \dots, Y_M$ distributed with g we wish to estimate $K(f, g)$. The second term H_1 can be estimated by (3.20), with asymptotic properties given by Corollaries 3.2-3.3. The first term $\breve{H}_1$ can be estimated in a similar manner, as follows: given X_i in the sample, $i \in \{1, \dots, N\}$, consider $\breve{\rho}(X_i, Y_j)$, $j = 1, \dots, M$, and the order statistics

$$\breve{\rho}^{(i)}_{1,M} \le \breve{\rho}^{(i)}_{2,M} \le \cdots \le \breve{\rho}^{(i)}_{M,M} \,,$$

so that $\breve{\rho}^{(i)}_{k,M}$ is the k-th nearest-neighbor distance from X_i to some Y_j, $j \in \{1, \dots, M\}$. Then, one can prove, similarly to Corollaries 3.2 and 3.3, that

$$\breve{H}_{N,M,k} = \frac{1}{N} \sum_{i=1}^{N} \log\{M \exp[-\Psi(k)] V_m (\breve{\rho}^{(i)}_{k,M})^m\} \tag{3.24}$$

is an asymptotically unbiased and consistent estimator of $\breve{H}_1$ (when now both N and M tend to infinity) when g is bounded and

$$J_q = \int_{\mathbb{R}^m} f(x) \, g^{q-1}(x) \, dx \tag{3.25}$$

exists for some $q < 1$. The difference

$$\begin{aligned}\breve{H}_{N,M,k} - \hat{H}_{N,k,1} &= m \log \left[\prod_{i=1}^{N} \breve{\rho}_{k,M}^{(i)}\right]^{1/N} + \log M - \Psi(k) + \log V_m \\ &\quad - m \log \left[\prod_{i=1}^{N} \rho_{k,N}^{(i)}\right]^{1/N} - \log(N-1) + \Psi(k) - \log V_m \\ &= m \log \left[\prod_{i=1}^{N} \frac{\breve{\rho}_{k,M}^{(i)}}{\rho_{k,N}^{(i)}}\right]^{1/N} + \log \frac{M}{N-1} \qquad (3.26)\end{aligned}$$

thus gives an asymptotically unbiased and consistent estimator of $K(f,g)$. Obviously a similar technique can be used to estimate the (symmetric) Kullback-Leibler divergence $K(f,g) + K(g,f)$. Note in particular that when f is unknown and only the sample $X_1, \ldots, X_N$ is available while g is known, then the term $\breve{H}_1$ in $K(f,g)$ can be estimated either by (3.24) with a sample $Y_1, \ldots, Y_M$ generated from g, with M taken arbitrarily large, or more simply by the Monte-Carlo estimator

$$\breve{H}_{1,N}(g) = -\frac{1}{N} \sum_{i=1}^{N} \log g(X_i)\,, \qquad (3.27)$$

the term H_1 being still estimated by (3.20). This forms an alternative to the method in (Broniatowski, 2003). Compared to the method in (Jiménez and Yukich, 2002) based on Voronoi tessellations (see also Miller (2003) for a Voronoi-based method for Shannon entropy estimation), it does not require any computation of multidimensional integral. In some applications one wishes to optimize $K(f,g)$ with respect to g that belongs to some class G (possibly parametric), with f fixed. Note that only the first term $\breve{H}_1$ of (3.23) need then be estimated. (Maximum likelihood estimation, with $g = g_\theta$ in a parametric class, is a most typical example: θ is then estimated by minimizing $\breve{H}_{1,N}(g_\theta)$, see (3.27).)

The Kullback-Leibler relative entropy can be used to construct a measure of mutual information (MI) between statistical distributions, with applications in image (Viola and Wells, 1995; Neemuchwala et al., 2005) and signal processing (Miller and Fisher, 2003). Let a_i and b_i denote the gray levels of pixel i in two images A and B respectively, $i = 1, \ldots, N$. The image matching problem consists in finding an image B in a data base that ressembles a given reference image A. The MI method corresponds to maximizing $K(f, f_x f_y)$, with f the joint density of the pairs (a_i, b_i) and f_x (respectively f_y) the density of gray levels in image A (respectively B). We have $K(f, f_x f_y) = H_1(f_x) + H_1(f_y) - H_1(f)$, where each term can be estimated by

(3.20) from one of the three samples (a_i), (b_i) or (a_i, b_i) (but A being fixed, only the last two terms need be estimated).

Another example of statistical distance between distributions is given by the following non-symmetric Bregman distance

$$(3.28)\ D_q(f,g) = \int_{\mathbb{R}^m} \left[g^q(x) + \frac{1}{q-1} f^q(x) - \frac{q}{q-1} f(x) g^{q-1}(x) \right] dx\,,\ q \neq 1\,,$$

or its symmetrized version

$$K_q(f,g) = \frac{1}{q}[D_q(f,g) + D_q(g,f)] = \frac{1}{q-1} \int_{\mathbb{R}^m} [f(x) - g(x)]\, [f^{q-1}(x) - g^{q-1}(x)]\, dx\,,$$

see, e.g., Basseville (1996). Given N independent observations from f and M from g, the first and second terms in (3.28) can be estimated by using (3.13). In the last term, the integral J_q given by (3.25) can be estimated by

$$\hat{I}_{N,M,k,q} = \frac{1}{N} \sum_{i=1}^{N} \{ M C_k V_m (\breve{\rho}_{k,M}^{(i)})^m \}^{1-q}\,.$$

Similarly to Theorem 3.1, $\hat{I}_{N,M,k,q}$ is asymptotically unbiased, $N, M \to \infty$, for $q < 1$ if J_q exists and for any $q \in (1, k+1)$ if g is bounded. We also obtain a property similar to Theorem 3.2: $\hat{I}_{N,M,k,q}$ is a consistent estimator of J_q, $N, M \to \infty$, for $q < 1$ if J_{2q-1} exists and for any $q \in (1, (k+2)/2)$ if g is bounded. (Notice, however, the difference with Theorem 3.2: when $q > 1$ the cases $k = 1$ and $k \geq 2$ need not be distinguished for the estimation of J_q and the upper bound on the admissible values for q is slightly larger than in Theorem 3.2.)

4. Examples.

Influence of k. Figure 1 presents H_q^* as a function of q (full line) for the normal distribution $\mathcal{N}(0, I_3)$ in $\mathbb{R}^3$. On the same figure are plotted the estimates $\hat{H}_{N,k,q}^*$ for $k = 1, \ldots, 5$ obtained from a sample of size $N = 1000$. Note that $\hat{H}_{N,k,q}^*$ is defined only for $q < k+1$ and quickly separates from the theoretical value H_q^* when $q > (k+1)/2$ or $q < 1$.

Information spectrum, estimation of var[log $f(X)$]. Figure 2 presents $\hat{H}_{N,k,q}^*(t_5)$ (dashed line) as a function of q, estimated from a sample of 10,000 data generated with the Student distribution with 5 degrees of freedom. (We are mainly interested in the behavior of H_q^* in the neighborhood of $q = 1$ and take $k = 1$.) The full lines correspond to $H_q^*(t_\nu)$ given by (2.12), for $\nu = 3, 4, \ldots, 7$ (the slope decreases as

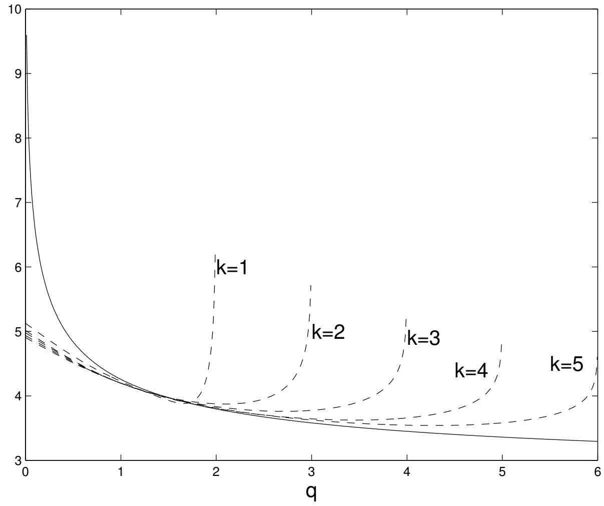

FIG. 1. H_q^* *(full line) and* $\hat{H}_{N,k,q}^*$ *(dashed lines) as functions of* q *for the normal distribution* $\mathcal{N}(0, I_3)$ *in* $\mathbb{R}^3$ *(*$N = 1000$*).*

ν increases, the curves have been shifted in order to coincide with $\hat{H}_{N,k,q}^*(t_5)$ at $q = 1$). The curve closest to $\hat{H}_{N,k,q}^*(t_5)$ around $q = 1$ is $H_q^*(t_5)$. This is confirmed by using the method suggested in Remark 3.5: the estimated value of S_f obtained by (3.22) is $\hat{S}_{f_{N,k}} \simeq 0.8874$, close enough to the value of $S_{f_5} \simeq 0.8588$ to estimate ν correctly (the next closest values are $S_{f_4} \simeq 0.9661$, $S_{f_6} \simeq 0.7911$, see Section 2.3).

Estimation of Kullback-Leibler divergence We use the same sample as above, consisting of 10,000 data generated with f, the Student distribution with 5 degrees of freedom. First, we estimate the Kullback-Leibler relative entropy $K(f, t_\nu)$ given by (3.23), using (3.27) for the estimation of $\breve{H}_1$ and (3.20) for the estimation of H_1, the entropy of f. The estimates obtained for different values of ν between 1 and 11 are given by stars in Figure 3. We then generate M samples from f_ν ($M = 10,000$), $\nu = 1, \ldots, 11$, and estimate $K(f_\nu, f)$ through (3.26). This corresponds to the dots in Figure 3. The sum of the two estimators, which estimates the symmetric Kullback-Leibler divergence, is indicated by circles in Figure 3. The three estimators tend to favor the hypothesis $f = t_5$.

q-entropy maximizing distributions. We generate $N = 500$ i.i.d. samples from the mixture of the three-dimensional Student distribution $T(\nu, (\nu-2)/\nu I_3, 0)$ with $\nu = 5$ and the normal distribution $\mathcal{N}(0, I_3)$, with relative weights β and $1 - \beta$. The

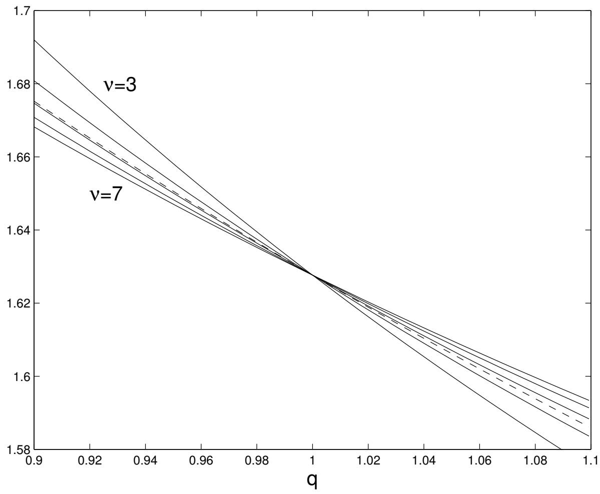

FIG. 2. *$\hat{H}^*_{N,k,q}(t_5)$ (dashed line) and $H^*_q(t_\nu)$, $\nu = 3, 4, \ldots, 7$ (full line) as functions of q for the Student distribution t_ν ($N = 10000$).*

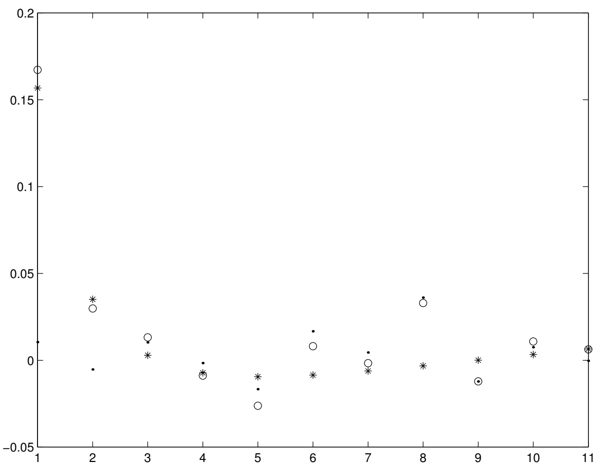

FIG. 3. *Estimation of the Kullback-Leibler relative entropies $K(f, t_\nu)$ (stars) and $K(t_\nu, f)$ (dots) and of the symmetric Kullback-Leibler divergence $K(f, t_\nu) + K(t_\nu, f)$ (circles) as functions of ν following the method of Section 3.3.*

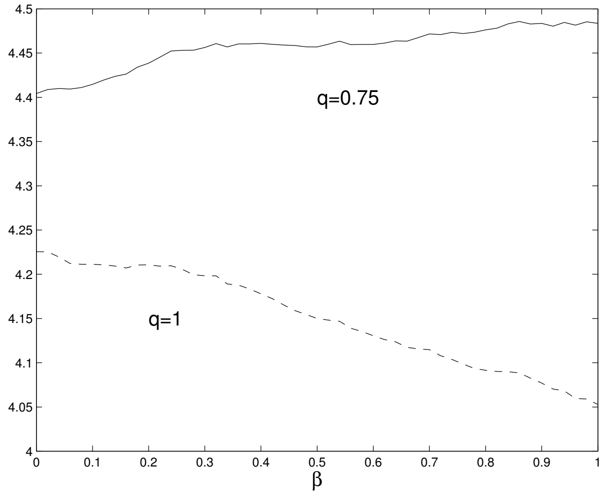

FIG. 4. *$\hat{H}^*_{N,k,q}$ for $q = 0.75$ (full line) and $\hat{H}_{N,k,1}$ (dashed line) in a mixture of Student and normal distributions as functions of the mixture coefficient β ($N = 500$, the estimates are averaged over 100 repetitions).*

variance of both distribution is the identity I_3, the Student distribution is q-entropy maximizing for $q = 1 - 2/(\nu + m) = 0.75$, see Section 2.2, the normal distribution maximizes Shannon entropy ($q = 1$). Figure 4 presents a plot of $\hat{H}^*_{N,k,q}$ for $q = 0.75$ and $\hat{H}_{N,k,1}$ (both averaged over 100 repetitions) as functions of the mixture coefficient β.

5. Related results and further developments. The asymptotic behavior of the moments of k-th nearest-neighbor distances is considered in (Evans et al., 2002) under the conditions that f is continuous, $f > 0$ on a compact convex subset $\mathcal{C}$ of $\mathbb{R}^m$, with f having bounded partial derivatives on $\mathcal{C}$. Only the case $q < 1$ is considered, and convergence in probability $\hat{I}_{N,k,q} \xrightarrow{\text{p}} I_q$, $N \to \infty$, is obtained for $m \geq 2$. The random graph approach of Hero and Michel (1999) supposes that the distribution is supported on $[0, 1]^m$, with some smoothness assumptions on f. Using a result from Redmond and Yukich (1996), a strongly consistent estimator of H^*_q is constructed when $0 < q < 1$ (up to an unknown bias term independent of f, related to the graph properties). Comparatively, our results cover a larger range of values for q and do not rely on regularity or bounded support assumptions for f.

The paper (Jiménez and Yukich, 2002) gives a method for estimating statistical distances between distributions with densities f and g based on Voronoi tessellations. Given an i.i.d. sample from f, it relies of the comparison between the Lebesgue measure (volume) and the measure for g of the Voronoi cells (polyhedra) constructed from the sample. Voronoi tessellations are also used in Miller (2003) to estimate the Shannon entropy of f based on an i.i.d. sample. The method requires the computation of the volumes of the Voronoi cells and no asymptotic results are indicated. Comparatively, the method based on nearest neighbors does not require any computation of (multidimensional) integrals. A possible motivation for using Voronoi tessellations could be the natural adaptation to the shape of the distribution. However, it should be noticed that the metric used to compute nearest-neighbor distances can be adapted to the observed sample: for $X_1, \ldots, X_N$ a sample having a non-spherical distribution, its empirical covariance matrix $\hat{\Sigma}_N$ can be used to define a new metric through $\|x\|^2_{\hat{\Sigma}_N} = x^\top \hat{\Sigma}_N^{-1} x$, the volume V_m of the unit ball in this metric becoming $|\hat{\Sigma}_N|^{1/2}\pi^{m/2}/\Gamma(m/2+1)$.

$\sqrt{N}$-consistency of an estimator of H_1 based on nearest-neighbor distances ($k = 1$) is proved in (Tsybakov and van der Meulen, 1996) for $m = 1$ and sufficiently regular densities f with unbounded support. On the other hand, $\sqrt{N}$-consistency of the estimator $\hat{I}_{N,k,q}$ is still an open issue. As for the case of spacing methods, where the spacing m can be taken as an increasing function of the sample size N, see e.g., Vasicek (1976); van Es (1992), it might be of interest to let $k = k_N$ increase with N. See also Remarks 3.2, 3.4. Properties of nearest-neighbor distances with $k_N \to \infty$ are considered for instance in (Loftsgaarden and Quesenberry, 1965; Moore and Yackel, 1977; Devroye and Wagner, 1977; Liero, 1993).

A central limit theorem for functions $h(\rho)$ of nearest-neighbor distances is obtained by Bickel and Breiman (1983) for $k = 1$ and by Penrose (2000) for $k \geq 1$. However, the results are restricted to the case of bounded functions, which does not cover the situation $h(\rho) = \rho^{m(1-q)}$, see (3.13), or $h(\rho) = \log(\rho)$, see (3.20). Conditions for the asymptotic normality of $\hat{I}_{N,k,q}$ are under current investigation.

6. Proofs.

The following lemma summarizes some properties of I_q.

Lemma 1.

- *(i) If f is bounded, then $I_q < \infty$ for any $q > 1$.*
- *(ii) If $I_q < \infty$ for some $q < 1$, then $I_{q'} < \infty$ for any $q' \in (q, 1)$.*
- *(iii) If f is of finite support, $I_q < \infty$ for any $q \in [0, 1)$.*

Proof.

(i) If $f(x) < \bar{f}$ and $q > 1$, $I_q = \int_{f\le 1} f^q + \int_{f>1} f^q \le \int_{f\le 1} f + \bar{f}^q \int_{f>1} f < \infty$.

(ii) If $q < q' < 1$, $I_{q'} = \int_{f\le 1} f^{q'} + \int_{f>1} f^{q'} \le \int_{f\le 1} f^q + \int_{f>1} f < \infty$ if $I_q < \infty$.

(iii) If $\mu_{\mathcal{S}} = \mu_{\mathcal{L}}\{x : f(x) > 0\} < \infty$ and $0 \le q < 1$, $I_q = \int_{f\le 1} f^q + \int_{f>1} f^q \le \mu_{\mathcal{S}} + \int_{f>1} f < \infty$.

■

The proofs of Theorems 3.1 and 3.2 use the following lemmas.

LEMMA 2 (LEBESGUE, 1910). *If $g \in L_1(\mathbb{R}^m)$, then for any sequence of open balls $\mathcal{B}(x, R_k)$ of radius tending to zero as $k \to \infty$ and for $\mu_{\mathcal{L}}$-almost any $x \in \mathbb{R}^m$,*

$$\lim_{k\to\infty} \frac{1}{V_m R_k^m} \int_{\mathcal{B}(x,R_k)} g(t)dt = g(x)\,.$$

Proof. See Lebesgue (1910). ■

LEMMA 3. *For any $\beta > 0$,*

$$\int_0^\infty x^\beta\, F(dx) = \beta \int_0^\infty x^{\beta-1}\,[1 - F(x)]\,dx \tag{6.29}$$

and

$$\int_0^\infty x^{-\beta}\, F(dx) = \beta \int_0^\infty x^{-\beta-1}\, F(x)\,dx\,, \tag{6.30}$$

in the sense that if one side converges so does the other.

Proof. See Feller (1966), vol. 2, p. 150, for (6.29). The proof is similar for (6.30). Define $\alpha = -\beta < 0$ and $I_{a,b} = \int_a^b z^\alpha\, F(dx)$ for some a, b, with $0 < a < b < \infty$. Integration by parts gives

$$I_{a,b} = [b^\alpha F(b) - a^\alpha F(a)] - \alpha \int_a^b x^{\alpha-1}\, F(x)\,dx$$

and, since $\alpha < 0$,

$$\lim_{b\to\infty} I_{a,b} = I_{a,\infty} = -a^\alpha F(a) - \alpha \int_a^\infty x^{\alpha-1}\, F(x)\,dx < \infty\,.$$

Suppose that $\int_0^\infty x^{-\beta}\, F(dx) = J < \infty$. It implies $\lim_{a\to 0^+} I_{0,a} = 0$, and since $I_{0,a} > a^\alpha F(a)$, $\lim_{a\to 0^+} a^\alpha F(a) = 0$. Therefore, $\lim_{a\to 0^+} -\alpha \int_a^\infty x^{\alpha-1}\, F(x)\,dx = J$.

Conversely, suppose that $\lim_{a\to 0^+} -\alpha \int_a^\infty x^{\alpha-1}\, F(x)\,dx = J < \infty$. Since $I_{a,\infty} < -\alpha \int_a^\infty x^{\alpha-1}\, F(x)\,dx$, $\lim_{a\to 0^+} I_{a,\infty} = J$. ■

6.1. *Proof of Theorem 3.1.* Since the X_i's are i.i.d.,

$$\mathbb{E}\hat{I}_{N,k,q} = \mathbb{E}\zeta_{N,i,k}^{1-q} = \mathbb{E}[\mathbb{E}(\zeta_{N,i,k}^{1-q}|X_i = x)]\,,$$

where the random variable $\zeta_{N,i,k}$ is defined by (3.14). Its distribution function conditional to $X_i = x$ is given by

$$\begin{aligned} F_{N,x,k}(u) &= \Pr(\zeta_{N,i,k} < u | X_i = x) \\ &= \Pr[\rho_{k,N-1}^{(i)} < R_N(u) | X_i = x] \end{aligned}$$

where

$$R_N(u) = \{u/[(N-1)V_m C_k]\}^{1/m}\,. \tag{6.31}$$

Let $\mathcal{B}(x,r)$ be the open ball of centre x an radius r. We have

$$\begin{aligned} F_{N,x,k}(u) &= \Pr\{k \text{ elements of more } \in \mathcal{B}[x, R_N(u)]\} \\ &= \sum_{j=k}^{N-1} \binom{N-1}{j} p_{N,u}^j (1-p_{N,u})^{N-1-j} \\ &= 1 - \sum_{j=0}^{k-1} \binom{N-1}{j} p_{N,u}^j (1-p_{N,u})^{N-1-j} \end{aligned}$$

where

$$p_{N,u} = \int_{\mathcal{B}[x,R_N(u)]} f(t)\,dt\,.$$

Lemma 2 gives

$$F_{N,x,k}(u) \to F_{x,k}(u) = 1 - \exp(-\lambda u) \sum_{j=0}^{k-1} \frac{(\lambda u)^j}{j!}$$

when $N \to \infty$ for μ-almost any x, with $\lambda = f(x)/C_k$, that is, $F_{N,x,k}$ tends to the Erlang distribution $F_{x,k}$, with p.d.f.

$$f_{x,k}(u) = \frac{\lambda^k u^{k-1} \exp(-\lambda u)}{\Gamma(k)}\,.$$

Direct calculation gives

$$\int_0^\infty u^{1-q} f_{x,k}(u)\,du = \frac{\Gamma(k+1-q)}{\lambda^{1-q}\Gamma(k)} = f^{q-1}(x)$$

for any $q < k+1$.

Suppose first that $q < 1$, and consider the random variables (U, X) with joint p.d.f. $f_{N,x,k}(u)f(x)$ on $\mathbb{R} \times \mathbb{R}^m$, where $f_{N,x,k}(u) = dF_{N,x,k}(u)/du$. The function

$u \to u^{1-q}$ is bounded on every bounded interval, and the generalized Helly-Bray Lemma, see Loève (1977, p. 187), implies

$$I\!E\hat{I}_{N,k,q} = \int_{\mathbb{R}^m}\int_0^\infty u^{1-q} f_{N,x,k}(u)\, f(x)\, du\, dx \to \int_{\mathbb{R}^m} f^q(x)\, dx = I_q\,,\ N \to \infty\,,$$

which completes the proof.

Suppose now that $1 < q < k+1$. Note that from Lemma 1 (i) $I_q < \infty$. Consider

$$J_N = \int_0^\infty u^{(1-q)(1+\delta)} F_{N,x,k}(du)\,.$$

We show that $\sup_N J_N < \infty$ for some $\delta > 0$. From Theorem 2.5.1 of Bierens (1994, p. 34), it implies

$$z_{N,k}(x) = \int_0^\infty u^{1-q}\, F_{N,x,k}(du) \to z_k(x) = \int_0^\infty u^{1-q}\, F_{x,k}(du) = f^{q-1}(x)\,,\ N \to \infty \tag{6.32}$$

for μ-almost any x in $\mathbb{R}^m$.

Define $\beta = (1-q)(1+\delta)$, so that $\beta < 0$, and take $\delta < (k+1-q)/(q-1)$ so that $\beta + k > 0$. From (6.30),

$$\begin{aligned}
J_N &= -\beta \int_0^\infty u^{\beta-1}\, F_{N,x,k}(u)\, du \\
&= -\beta \int_0^1 u^{\beta-1}\, F_{N,x,k}(u)\, du \ - \beta \int_1^\infty u^{\beta-1}\, F_{N,x,k}(u)\, du \\
&\le -\beta \int_0^1 u^{\beta-1}\, F_{N,x,k}(u)\, du \ - \beta \int_1^\infty u^{\beta-1}\, du \\
&= 1 - \beta \int_0^1 u^{\beta-1}\, F_{N,x,k}(u)\, du\,. \qquad (6.33)
\end{aligned}$$

Since $f(x)$ is bounded, say by $\bar{f}$, we have

$$\forall x \in \mathbb{R}^m\,,\ \forall u \in \mathbb{R}\,,\ \forall N\,,\ p_{N,u} \le \bar{f} V_m [R_N(u)]^m = \frac{\bar{f} u}{(N-1)C_k}\,.$$

It implies

$$\begin{aligned}
\frac{F_{N,x,k}(u)}{u^k} &\le \sum_{j=k}^{N-1} \binom{N-1}{j} \frac{\bar{f}^j u^{j-k}}{C_k^j (N-1)^j} \\
&\le \sum_{j=k}^{N-1} \frac{\bar{f}^j u^{j-k}}{C_k^j j!} = \frac{\bar{f}^k}{C_k^k k!} + \sum_{j=k+1}^{N-1} \frac{\bar{f}^j u^{j-k}}{C_k^j j!} \\
&\le \frac{\bar{f}^k}{C_k^k k!} + \frac{\bar{f}^k}{C_k^k} \sum_{i=1}^{N-k-1} \frac{\bar{f}^i u^i}{C_k^i i!}
\end{aligned}$$

$$\leq \frac{\bar{f}^k}{C_k^k k!} + \frac{\bar{f}^k}{C_k^k} \sum_{i=1}^{\infty} \frac{\bar{f}^i u^i}{C_k^i i!} = \frac{\bar{f}^k}{C_k^k k!} + \frac{\bar{f}^k}{C_k^k} \left\{ \exp\left[\frac{\bar{f} u}{C_k}\right] - 1 \right\},$$

and thus for $u < 1$,

$$\frac{F_{N,x,k}(u)}{u^k} < U_k = \frac{\bar{f}^k}{C_k^k k!} + \frac{\bar{f}^k}{C_k^k} \left\{ \exp\left[\frac{\bar{f}}{C_k}\right] - 1 \right\}. \tag{6.34}$$

Therefore, from (6.33),

$$J_N \leq 1 - \beta U_k \int_0^1 u^{k+\beta-1} \, du = 1 - \frac{\beta U_k}{k+\beta} < \infty \tag{6.35}$$

which implies (6.32). Now we only need to prove that

$$\int_{\mathbb{R}^m} z_{N,k}(x) \, f(x) \, dx \to \int_{\mathbb{R}^m} z_k(x) \, f(x) \, dx = I_q \,, \; N \to \infty \,.$$

But this follows from Lebesgue's bounded convergence theorem, since $z_{N,k}(x)$ is bounded (take $\delta = 0$ in J_N). ■

6.2. *Proof of Theorem 3.2.* We shall use the same notations as in the proof of Theorem 3.1, and write $\hat{I}_{N,k,q} = (1/N) \sum_{i=1}^{N} \zeta_{N,i,k}^{1-q}$, so that

$$\mathbb{E}(\hat{I}_{N,k,q} - I_q)^2 = \frac{\mathbb{E}(\zeta_{N,i,k}^{1-q} - I_q)^2}{N} + \frac{1}{N^2} \sum_{i \neq j} \mathbb{E}\{(\zeta_{N,i,k}^{1-q} - I_q)(\zeta_{N,j,k}^{1-q} - I_q)\} \,. \tag{6.36}$$

We consider the cases $q < 1$ and $q > 1$ separately.

$q < 1$. Note that $2q - 1 < q < 1$ and Lemma 1 (ii) gives $I_q < \infty$ when $I_{2q-1} < \infty$. Consider the first term on the right-hand side of (6.36). We have

$$\mathbb{E}(\zeta_{N,i,k}^{1-q} - I_q)^2 = \mathbb{E}(\zeta_{N,i,k}^{1-q})^2 + I_q^2 - 2 I_q \mathbb{E}\zeta_{N,i,k}^{1-q} \tag{6.37}$$

where the last term tends to $-2I_q^2$ from Theorem 3.1. Consider the first term,

$$\mathbb{E}(\zeta_{N,i,k}^{1-q})^2 = \int_{\mathbb{R}^m} \int_0^{\infty} u^{2(1-q)} f_{N,x,k}(u) \, f(x) \, du \, dx \,.$$

Since the function $u \to u^{1-q}$ is bounded on every bounded interval, it tends to

$$\int_{\mathbb{R}^m} \int_0^{\infty} u^{2(1-q)} f_{x,k}(u) \, f(x) \, du \, dx = I_{2q-1} \frac{\Gamma(k+2-2q)\Gamma(k)}{\Gamma^2(k+1-q)}$$

for any $q < (k+2)/2$ (generalized Helly-Bray Lemma, Loève (1977, p. 187)). Therefore, $\mathbb{E}(\zeta_{N,i,k}^{1-q} - I_q)^2$ tends to a finite limit, and the first term on the right-hand side of (6.36) tends to zero as $N \to \infty$.

Consider now the second term of (6.36). We show that

$$\mathbb{E}\{(\zeta_{N,i,k}^{1-q} - I_q)(\zeta_{N,j,k}^{1-q} - I_q)\} = \mathbb{E}\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} + I_q^2 - 2I_q\mathbb{E}\zeta_{N,i,k}^{1-q} \to 0\,,$$

$N \to \infty$. Since $\mathbb{E}\zeta_{N,i,k}^{1-q} \to I_q$ from Theorem 3.1, we only need to show that $\mathbb{E}\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} \to I_q^2$.

Define

$$\begin{aligned} F_{N,x,y,k}(u,v) &= \Pr\{\zeta_{N,i,k} < u\,,\ \zeta_{N,j,k} < v | X_i = x, X_j = y\}\,,\\ &= \Pr\{\rho_{k,N-1}^{(i)} < R_N(u)\,,\ \rho_{k,N-1}^{(j)} < R_N(v) | X_i = x, X_j = y\}\,, \end{aligned}$$

so that

$$\mathbb{E}\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} = \int_{\mathbb{R}^m}\int_{\mathbb{R}^m}\int_0^\infty\int_0^\infty u^{1-q}v^{1-q}F_{N,x,y,k}(du,dv)\, f(x)\, f(y)\, dx\, dy\,. \tag{6.38}$$

From the definition of $R_N(u)$, see (6.31), there exist $N_0 = N_0(x,y,u,v)$ such that $\mathcal{B}[x, R_N(u)] \cap \mathcal{B}[y, R_N(v)] = \emptyset$ for $N > N_0$ and thus

$$F_{N,x,y,k}(u,v) = \sum_{j=k}^{N-2}\sum_{l=k}^{N-2}\binom{N-2}{j}\binom{N-2-j}{l} p_{N,u}^j p_{N,v}^l (1 - p_{N,u} - p_{N,v})^{N-2-j-l}$$

with $p_{N,u} = \int_{\mathcal{B}[x,R_N(u)]} f(t)dt$, $p_{N,v} = \int_{\mathcal{B}[y,R_N(v)]} f(t)dt$. Hence, for $N > N_0$

$$\begin{aligned} F_{N,x,y,k}(u,v) &= F_{N-1,x,k}(u) + F_{N-1,y,k}(v) - 1\\ &\quad + \sum_{j=0}^{k-1}\sum_{l=0}^{k-1}\binom{N-2}{j}\binom{N-2-j}{l} p_{N,u}^j p_{N,v}^l (1 - p_{N,u} - p_{N,v})^{N-2-j-l}\,. \end{aligned}$$

Similarly to the proof of Theorem 3.1, we then obtain

$$F_{N,x,y,k}(u,v) \to F_{x,y,k}(u,v) = F_{x,k}(u)F_{y,k}(v)\,,\ N \to \infty \tag{6.39}$$

for $\mu_{\mathcal{L}}$-almost any x and y with

$$\int_0^\infty\int_0^\infty u^{1-q}v^{1-q}F_{x,y,k}(du,dv) = f^{q-1}(x)f^{q-1}(y) \tag{6.40}$$

for any $q < k+1$. Since the function $u \to u^{1-q}$ is bounded on every bounded interval (6.38) gives

$$\mathbb{E}\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} \to \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} f^q(x)\, f^q(y)\, dx\, dy = I_q^2\,,\ N \to \infty$$

(generalized Helly-Bray Lemma, Loève (1977, p. 187)). This completes the proof that $\mathbb{E}(\hat{I}_{N,k,q} - I_q)^2 \to 0$, and thus $\hat{I}_{N,k,q} \xrightarrow{\text{p}} I_q$, when $N \to \infty$.

$q > 1$. Note that from Lemma 1 (i) I_q and I_{2q-1} both exist. Consider the first term on the right-hand side of (6.36). We have again (6.37) where the last term tends to $-2I_q^2$ (the assumptions of the theorem imply $q < k+1$ so that Theorem 3.1 applies). Consider the first term of (6.37). Define

$$J'_N = \int_0^\infty u^{2(1-q)(1+\delta)}\, F_{N,x,k}(du)\,,$$

we show that $\sup_N J'_N < \infty$ for some $\delta > 0$. From the assumptions of the theorem, $2q < k+2$. Let $\beta = 2(1-q)(1+\delta)$, so that $\beta < 0$ and take $\delta < (k+2-2q)/[2(q-1)]$ so that $\beta + k > 0$. Using Lemma 3 and developments similar to the proof of Theorem 3.1 we obtain

$$\begin{aligned} J'_N &= -\beta \int_0^\infty u^{\beta-1}\, F_{N,x,k}(du) \\ &\leq 1 - \beta \int_0^1 u^{\beta-1}\, F_{N,x,k}(du) \\ &\leq 1 - \beta U_k \int_0^1 u^{k+\beta-1}\, du = 1 - \frac{\beta U_k}{k+\beta} < \infty\,, \end{aligned}$$

where U_k is given by (6.34). Theorem 2.5.1 of Bierens (1994) then implies

$$\int_0^\infty u^{2(1-q)}\, F_{N,x,k}(du) \to \int_0^\infty u^{2(1-q)}\, F_{x,k}(du) = \frac{\Gamma(k+2-2q)\Gamma(k)}{\Gamma^2(k+1-q)} f^{2q-2}(x)$$

for μ-almost any x, $q < (k+2)/2$ and Lebesgue's bounded convergence theorem gives $I\!\!E(\zeta_{N,i,k}^{1-q})^2 \to I_{2q-1}\Gamma(k+2-2q)\Gamma(k)/\Gamma^2(k+1-q)$, $N \to \infty$. The first term of (6.36) thus tends to zero.

Consider now the second term. As in the case $q < 1$, we only need to show that $I\!\!E\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} \to I_q^2$. Define

$$J''_N = \int_0^\infty \int_0^\infty u^{(1-q)(1+\delta)} v^{(1-q)(1+\delta)} F_{N,x,y,k}(du, dv)\,.$$

Using (6.39, 6.40), proving that

$$\sup_N J''_N < J(x,y) < \infty$$

for some $\delta > 0$ will then establish that

$$\int_0^\infty \int_0^\infty u^{1-q} v^{1-q} F_{N,x,y,k}(du, dv) \to f^{q-1}(x) f^{q-1}(y)\,,\ N \to \infty\,, \tag{6.41}$$

for μ-almost x and y, see Theorem 2.5.1 of Bierens (1994). Using (6.38), if

$$\int_{\mathbb{R}^m} \int_{\mathbb{R}^m} J(x,y)\, f(x)\, f(y)\, dx\, dy < \infty \tag{6.42}$$

Lebesgue's bounded convergence theorem will then complete the proof.

Integration by parts, as in the proof of Lemma 3, gives

$$J''_N = \beta^2 \int_0^\infty \int_0^\infty u^{\beta-1} v^{\beta-1} F_{N,x,y,k}(u,v)\, du\, dv\,,$$

where $\beta = (1-q)(1+\delta) < 0$. We use different bounds for $F_{N,x,y,k}(u,v)$ on three different parts of the (u,v) plane.

(i) Suppose that $\max[R_N(u), R_N(v)] \leq \|x-y\|$, which is equivalent to $(u,v) \in \mathcal{D}_1 = [0,\Lambda] \times [0,\Lambda]$ with

$$\Lambda = \Lambda(k,N,x,y) = (N-1)V_m C_k \|x-y\|^m\,.$$

This means that the balls $\mathcal{B}[x, R_N(u)]$ and $\mathcal{B}[y, R_N(v)]$ either do not intersect, or, when they do, their intersection does not contain neither x nor y. In that case, we use

$$F_{N,x,y,k}(u,v) < \min[F_{N-1,x,k}(u), F_{N-1,y,k}(v)] < F^{1/2}_{N-1,x,k}(u)\, F^{1/2}_{N-1,y,k}(v)\,,$$

and

$$\begin{aligned} J''^{(1)}_N &= \beta^2 \int_{\mathcal{D}_1} u^{\beta-1} v^{\beta-1} F_{N,x,y,k}(u,v)\, du\, dv \\ &< \beta^2 \left[\int_0^\Lambda u^{\beta-1} F^{1/2}_{N-1,x,k}(u)\, du\right] \left[\int_0^\Lambda v^{\beta-1} F^{1/2}_{N-1,y,k}(v)\, dv\right] \\ &< \beta^2 \left[U_k^{1/2} \int_0^1 u^{\beta-1+k/2}\, du + \int_1^\infty u^{\beta-1}\, du\right]^2 \\ &= \beta^2 \left[U_k^{1/2} \frac{2}{2\beta+k} - \frac{1}{\beta}\right]^2 < \infty \end{aligned}$$

where we used the bound (6.34) for $F_{N-1,x,k}(u)$ when $u<1$, $F_{N-1,x,k}(u) < 1$ for $u \geq 1$ and choose $\delta < (k+2-2q)/[2(q-1)]$ so that $2\beta + k > 0$ (this choice of δ is legitimate since $q < (k+2)/2$).

(ii) Suppose, without any loss of generality, that $u < v$ and consider the domain defined by $R_N(u) \leq \|x-y\| < R_N(v)$, that is, $(u,v) \in \mathcal{D}_2 = [0,\Lambda] \times (\Lambda, \infty)$. The cases $k=1$ and $k \geq 2$ must be treated separately since $\mathcal{B}[y, R_N(v)]$ contains x.

When $k=1$, $F_{N,x,y,1}(u,v) = F_{N-1,x,1}(u)$ and we have

$$\begin{aligned} J''^{(2)}_N &= \beta^2 \int_{\mathcal{D}_2} u^{\beta-1} v^{\beta-1} F_{N,x,y,1}(u,v)\, du\, dv \\ &< \beta^2 \left[\int_0^\Lambda u^{\beta-1} F_{N-1,x,1}(u)\, du\right] \left[\int_\Lambda^\infty v^{\beta-1}\, dv\right] \\ &< \beta^2 \left[U_1 \int_0^1 u^\beta\, du + \int_1^\infty u^{\beta-1}\, du\right] \left(-\frac{\Lambda^\beta}{\beta}\right) \end{aligned}$$

$$= -\beta \left[\frac{U_1}{\beta+1} - \frac{1}{\beta} \right] \Lambda^\beta$$

$$< J^{(2)}(x,y) = -\beta \left[\frac{U_1}{\beta+1} - \frac{1}{\beta} \right] V_m^\beta C_1^\beta \|x-y\|^{m\beta} , \tag{6.43}$$

where we used (6.34) and take $\delta < (2-q)/(q-1)$ so that $\beta > -1$ (this choice of δ is legitimate since $q < 2$).

Suppose now that $k \geq 2$. We have

$$F_{N,x,y,k}(u,v) < F_{N-1,x,k}^{1-\alpha}(u)\, F_{N-1,y,k-1}^{\alpha}(v) \,,\ \forall \alpha \in (0,1) \,.$$

Developments similar to those used for the derivation of (6.34) give for $v < 1$

$$\frac{F_{N-1,y,k-1}(v)}{v^{k-1}} < V_{k-1} = \frac{\bar{f}^{k-1}}{C_k^{k-1}(k-1)!} + \frac{\bar{f}^{k-1}}{C_k^{k-1}} \left\{ \exp\left[\frac{\bar{f}}{C_k} \right] - 1 \right\} . \tag{6.44}$$

We obtain

$$\begin{aligned} J_N''^{(2)} &= \beta^2 \int_{\mathcal{D}_2} u^{\beta-1} v^{\beta-1}\, F_{N,x,y,k}(u,v)\, du\, dv \\ &< \beta^2 \left[\int_0^\Lambda u^{\beta-1} F_{N-1,x,k}^{1-\alpha}(u)\, du \right] \left[\int_\Lambda^\infty v^{\beta-1} F_{N-1,y,k-1}^{\alpha}(v)\, dv \right] \\ &< \beta^2 \left[U_k^{1-\alpha} \int_0^1 u^{\beta-1+(1-\alpha)k}\, du + \int_1^\infty u^{\beta-1}\, du \right] \\ &\quad \times \left[V_{k-1}^\alpha \int_0^1 v^{\beta-1+(k-1)\alpha}\, dv + \int_1^\infty v^{\beta-1}\, dv \right] \\ &= \beta^2 \left[\frac{U_k^{1-\alpha}}{k(1-\alpha)+\beta} - \frac{1}{\beta} \right] \left[\frac{V_{k-1}^\alpha}{(k-1)\alpha+\beta} - \frac{1}{\beta} \right] < \infty , \end{aligned}$$

where we used (6.34, 6.44) and require $\beta + k(1-\alpha) > 0$ and $\beta + (k-1)\alpha > 0$. For that we take $\alpha = \alpha_k = k/(2k-1)$. Indeed, from the assumptions of the theorem, $q < (k+1)/2 < (k^2+k-1)/(2k-1)$ so that we can choose $\delta < [(k^2+k-1) - q(2k-1)]/[(q-1)(2k-1)]$ which ensures that both $\beta + k(1-\alpha_k) > 0$ and $\beta + (k-1)\alpha_k > 0$.

(iii) Suppose finally that $\|x-y\| < \min[R_N(u), R_N(v)]$, that is, $(u,v) \in \mathcal{D}_3 = (\Lambda,\infty)\times(\Lambda,\infty)$. In that case, each of the balls $\mathcal{B}[x, R_N(u)]$ and $\mathcal{B}[y, R_N(v)]$ contains both x and y. Again, the case $k=1$ and $k \geq 2$ must be distinguished.

When $k=1$, $F_{N,x,y,1}(u,v) = 1$ and

$$J_N''^{(3)} = \beta^2 \int_{\mathcal{D}_3} u^{\beta-1} v^{\beta-1}\, F_{N,x,y,1}(u,v)\, du\, dv$$

$$= \beta^2 \left[\int_\Lambda^\infty u^{\beta-1}\, du \right]^2 = \Lambda^{2\beta} < J^{(3)}(x,y) = V_m^{2\beta} C_1^{2\beta} \|x-y\|^{2m\beta} . \tag{6.45}$$

When $k \geq 2$,

$$F_{N,x,y,k}(u,v) < F^{1/2}_{N-1,x,k-1}(u)\, F^{1/2}_{N-1,y,k-1}(v)\,,$$

and

$$\begin{aligned} J_N''^{(3)} &= \beta^2 \int_{\mathcal{D}_3} u^{\beta-1} v^{\beta-1}\, F_{N,x,y,k}(u,v)\, du\, dv \\ &< \beta^2 \left[\int_\Lambda^\infty u^{\beta-1} F^{1/2}_{N-1,x,k-1}(u)\, du\right] \left[\int_\Lambda^\infty v^{\beta-1} F^{1/2}_{N-1,y,k-1}(v)\, dv\right] \\ &< \beta^2 \left[V^{1/2}_{k-1} \frac{2}{2\beta+k-1} - \frac{1}{\beta}\right]^2 < \infty \end{aligned}$$

where we used (6.44) and take $\delta < [(k+1)-2q]/[2(q-1)]$ so that $k-1+2\beta > 0$ (this choice of δ is legitimate since $q < (k+1)/2$).

Summarizing the three cases above, we obtain $J_N'' = J_N''^{(1)} + 2J_N''^{(2)} + J_N''^{(3)}$ with different bounds for $J_N''^{(2)}$ and $J_N''^{(3)}$ depending whether $k=1$ or $k \geq 2$. This proves (6.41).

When $k \geq 2$, the bound on J_N'' does not depend on x, y, and Lebesgue's bounded convergence theorem implies $\mathbb{E}\{\zeta^{1-q}_{N,i,k}\zeta^{1-q}_{N,j,k}\} \to I_q^2$, which completes the proof of the theorem, see (6.42).

When $k = 1$, the condition (6.42) is satisfied if $2m\beta > -1$, see (6.43, 6.45), which is ensured by the choice $\delta < (2m+1-2qm)/[2m(q-1)]$ (legitimate since $q < 1 + 1/(2m)$). Indeed, the convexity of the function $z \to z^\gamma/2$ for $\gamma < 0$ gives

$$\|x-y\|^\gamma < m^{\gamma/2-1} \sum_{i=1}^m |x_i - y_i|^\gamma < \sum_{i=1}^m |x_i - y_i|^\gamma\,,$$

and thus

$$\begin{aligned} \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} \|x-y\|^\gamma f(x)\, f(y)\, dx\, dy &< \sum_{i=1}^m \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} |x_i-y_i|^\gamma f(x)\, f(y)\, dx\, dy \\ &= \sum_{i=1}^m \int_{-\infty}^\infty f_i(t) \left(\int_{-\infty}^\infty |s|^\gamma f_i(t+s)\, ds\right) dt \\ &< \sum_{i=1}^m \int_{-\infty}^\infty f_i(t) \left[2\int_0^1 s^\gamma \bar{f} ds + \int_{-\infty}^\infty f_i(s+t) ds\right] dt \\ &= m\left(\frac{2\bar{f}}{1+\gamma} + 1\right) < \infty \end{aligned}$$

when $\gamma > -1$, where we denoted $f_i(t) = \int_{\mathbb{R}^{m-1}} f(x) \left(\prod_{j\neq i} dx_j\right)_{|x_i=t}$ if $m > 1$ and $f_i(t) = f(t)$ if $m = 1$. When $\delta < (2m+1-2qm)/[2m(q-1)]$ Lebesgue's bounded convergence theorem thus implies $I\!\!E\{\zeta_{N,i,k}^{1-q}\zeta_{N,j,k}^{1-q}\} \to I_q^2$, which completes the proof of the theorem. ■

6.3. *Proof of Corollary 3.2.* The existence of H_1 directly follows from that of I_{q_1} for $q_1 < 1$ and the boundedness of f. We have

$$I\!\!E\hat{H}_{N,k,1} = I\!\!E \log \xi_{N,i,k} = I\!\!E[I\!\!E(\log \xi_{N,i,k}|X_i = x)] \,,$$

where the only difference between the random variables $\zeta_{N,i,k}$ (3.21) and $\xi_{N,i,k}$ (3.14) is the substitution of $\exp[-\Psi(k)]$ for C_k. Similarly to the proof of Theorem 3.1 we define

$$\begin{aligned} F_{N,x,k}(u) &= \Pr(\xi_{N,i,k} < u|X_i = x) \\ &= \Pr[\rho_{k,N-1}^{(i)} < R_N(u)|X_i = x] \end{aligned}$$

with now

$$R_N(u) = (u/\{(N-1)V_m \exp[-\Psi(k)]\})^{1/m} \,.$$

Following the same steps as in the proof of Theorem 3.1 we then obtain

$$F_{N,x,k}(u) \to F_{x,k}(u) = 1 - \exp(-\lambda u) \sum_{j=0}^{k-1} \frac{(\lambda u)^j}{j!} \,,\ N \to \infty \,,$$

for $\mu_{\mathcal{L}}$-almost any x, with $\lambda = f(x)\exp[\Psi(k)]$. Direct calculation gives

$$\int_0^\infty \log(u) F_{x,k}(du) = -\log f(x) \,.$$

We shall use again Theorem 2.5.1 of Bierens (1994), p. 34, and show that

$$J_N = \int_0^\infty |\log(u)|^{1+\delta}\, F_{N,x,k}(du) < \infty \,, \tag{6.46}$$

for some $\delta > 0$, which implies

$$\int_0^\infty \log(u)\, F_{N,x,k}(du) \to \int_0^\infty \log(u) F_{x,k}(du) = -\log f(x) \,,\ N \to \infty \,,$$

for $\mu_{\mathcal{L}}$-almost any x. The convergence

$$\int_{\mathbb{R}^m} \int_0^\infty \log(u)\, F_{N,x,k}(du)\, f(x)\, dx \to H_1 \,,\ N \to \infty \,,$$

then follows from Lebesgue's bounded convergence theorem.

In order to prove (6.46), we write

$$J_N = \int_0^1 |\log(u)|^{1+\delta}\, F_{N,x,k}(du) + \int_1^\infty |\log(u)|^{1+\delta}\, F_{N,x,k}(du) \,. \tag{6.47}$$

Since f is bounded, we can take $q_2 > 1$ (and smaller than $k+1$) such that

$$\int_0^\infty u^{1-q_2}\, F_{N,x,k}(du) < \infty\,,$$

see (6.35). Since $|\log(u)|^{1+\delta}/u^{1-q_2} \to 0$ when $u \to 0$, it implies that the first integral on the right-hand side of (6.47) is finite. Similarly, since by assumption I_{q_1} exists for some $q_1 < 1$, $\int_0^\infty u^{1-q_1}\, F_{N,x,k}(du) < \infty$ and $|\log(u)|^{1+\delta}/u^{1-q_1} \to 0$, $u \to \infty$, implies that the second integral on the right-hand side of (6.47) is finite, which completes the proof. ■

6.4. *Proof of Corollary 3.3.* Similarly to the proof of Corollary 3.2, we only need to replace $\zeta_{N,i,k}$ (3.21) by $\xi_{N,i,k}$ (3.14) and C_k by $\exp[-\Psi(k)]$ in the proof of Theorem 3.2. When we now compute

$$\mathbb{E}(\hat{H}_{N,k,1}-H_1)^2 = \frac{\mathbb{E}(\log \xi_{N,i,k} - H_1)^2}{N} + \frac{1}{N^2}\sum_{i\neq j} \mathbb{E}\{(\log \xi_{N,i,k}-H_1)(\log \xi_{N,j,k}-H_1)\}\,,$$

in the first term, $\mathbb{E}(\log \xi_{N,i,k} - H_1)^2$ tends to

$$\int_{\mathbb{R}^m} \log^2 f(x)\, f(x)\, dx - H_1^2 + \dot{\Psi}(k) = \mathrm{var}[\log f(X)] + \dot{\Psi}(k)\,,$$

with $\dot{\Psi}(z)$ is the trigamma function, $\dot{\Psi}(z) = d^2 \log \Gamma(z)/dz^2$, and for the second term the developments are similar to those in Theorem 3.2. For instance, equation (6.41) now becomes

$$\int_0^\infty \int_0^\infty \log u\, \log v\, F_{N,x,y,k}(du, dv) \to \log f(x) \log f(y)\,,\ N \to \infty\,,$$

for μ-almost x and y. We can then show that $\mathbb{E}\{\log \xi_{N,i,k} \log \xi_{N,j,k}\} \to H_1^2$, so that $\mathbb{E}(\hat{H}_{N,k,1} - H_1)^2 \to 0$, $N \to \infty$. ■

Acknowledgments. The first author gratefully acknowledges financial support from EPSRC grant RCMT091. The work of the second author was partially supported by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors's view. The authors wish to thank Anatoly A. Zhigljavsky from Cardiff School of Mathematics for helpful discussions.

REFERENCES

Alemany, P., Zanette, S., 1992. Fractal random walks from a variational formalism for Tsallis entropies. Physical Review E 49 (2), 956–958.

Basseville, M., 1996. Information: entropies, divergences et moyennes. Research Report IRISA nb. 1020, May 1996, 74 pages.

Beirlant, J., Dudewicz, E., Györfi, L., van der Meulen, E., 1997. Nonparametric entropy estimation; an overview. International Journal of Mathematical and Statistical Sciences 6 (1), 17–39.

Bickel, P., Breiman, L., 1983. Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Annals of Statistics 11 (1), 185–214.

Bierens, H., 1994. Topics in Advanced Econometrics. Cambridge University Press, Cambridge.

Broniatowski, M., 2003. Estimation of the Kullback-Leibler divergence. Mathematical Methods of Statistics 12 (4), 391–409.

Devroye, L., Wagner, T., 1977. The strong uniform consistency of nearest neighbor density estimates. Annals of Statistics 5 (3), 536–540.

Evans, D., Jones, A., Schmidt, W., 2002. Asymptotic moments of near-neighbour distance distributions. Proc. Royal Soc. Lond. A458, 2839–2849.

Feller, W., 1966. An Introduction to Probability Theory and Its Applications. Wiley, New York.

Frank, T., Daffertshofer, A., 2000. Exact time-dependent solutions of the Rényi Fokker-Planck equation and the Fokker-Planck equations related to the entropies proposed by Sharma and Mittal. Physica A285, 351–366.

Goria, M., Leonenko, N., Mergel, V., Inverardi, P. N., 2005. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. Journal of Nonparametric Statistics 17 (3), 277–297.

Havrda, M., Charvát, F., 1967. Quantification method of classification processes: concept of structural α-entropy. Kybernetika 3, 30–35.

Hero, A., Ma, B., Michel, O., Gorman, J., 2002. Applications of entropic spanning graphs. IEEE Signal Proc. Magazine (Special Issue on Mathematics in Imaging) 19 (5), 85–95.

Hero, A., Michel, O., 1999. Asymptotic theory of greedy approximations to minimal k-point random graphs. IEEE Transactions on Information Theory 45, 1921–1939.

Heyde, C., Leonenko, N., 2005. Student processes. Advances in Applied Probability 37 (2), in press.

Jiménez, R., Yukich, J., 2002. Asymptotics for statistical distances based on Voronoi tessellations. Journal of Theoretical Probability 15 (2), 503–541.

Kapur, J. N., 1989. Maximum-Entropy Models in Science and Engineering. John Wiley & Sons Inc., New York.

Kozachenko, L., Leonenko, N., 1987. On statistical estimation of entropy of random vector. Problems Infor. Transmiss. 23 (2), 95–101, (translated from Problemy Peredachi Informatsii, in Russian, vol. 23, No. 2, pp. 9-16, 1987).

Lebesgue, H., 1910. Sur l'intégration des fonctions discontinues. Ann. École Norm. 27, 361–450.

Liero, H., 1993. A note on the asymptotic behaviour of the distance of the k-th nearest neighbour. Statistics 24, 235–243.

Loève, M., 1977. Probability Theory I. Springer (4th edition), Heidelberg.

Loftsgaarden, D., Quesenberry, C., 1965. A nonparametric estimate of a multivariate density fucntion. Annals of Math. Stat. 36, 1049–1051.

Miller, E., 2003. A new class of entropy estimators for multidimendional densities. In: Proc. ICASSP'2003.

Miller, E., Fisher, J., 2003. ICA using spacings estimates of entropy. In: Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation. Nara, Japan, pp. 1047–1052.

Moore, D., Yackel, J., 1977. Consistency properties of nearest neighbor density function estimators. Annals of Statistics 5, 143–154.

Neemuchwala, H., Hero, A., Carson, P., 2005. Image matching using alpha-entropy measures and entropic graphs. Signal Processing 85, 277–296.

Penrose, M., 2000. Central limit theorems for k-nearest neighbour distances. Stochastic Processes and their Applications 85, 295–320.

Pronzato, L., Thierry, E., Wolsztynski, E., June 2004. Minimum entropy estimation in semi parametric models: a candidate for adaptive estimation? In: Di Bucchianico, A., Läuter, H., Wynn, H. (Eds.), mODa'7 – Advances in Model–Oriented Design and Analysis, Proceedings of the 7th Int. Workshop, Heeze (Netehrlands). Physica Verlag, Heidelberg, pp. 125–132.

Redmond, C., Yukich, J., 1996. Asymptotics for Euclidian functionals with power-weighted edges. Stochastic Processes and their Applications 61, 289–304.

Rényi, A., 1961. On measures of entropy and information. In: Proc. 4th Berkeley Symp. on Math. Statist. and Prob. pp. 547–561.

Song, K.-S., 2001. Rényi information, likelihood and an intrinsic distribution measure. Journal of Statistical Planning and Inference 93, 51–69.

Tsallis, C., 1988. Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics 52 (1/2), 479–487.

Tsallis, C., Bukman, D., 1996. Anomalous diffusion in the presence of external forces: exact time-dependent solutions and their thermostatistical basis. Physical Reviews E54 (3), 2197–2200.

Tsallis, C., Levy, S., Souza, A., Maynard, R., 1995. Statistical-mechanical foundation of the ubiquity of Lévy distributions in nature. Physical Review Letters 75 (20), 3589–3593.

Tsybakov, A., van der Meulen, E., 1996. Root-n consistent estimators of entropy for densities with unbounded support. Scandinavian Journal of Statistics 23, 75–83.

van Es, B., 1992. Estimating functionals related to a density by a class of statistics based on spacings. Scandinavian Journal of Statistics 19, 61–72.

Vasicek, O., 1976. A test for normality based on sample entropy. Journal of Royal Statistical Society B38 (1), 54–59.

Vignat, C., Hero, A., Costa, J., 2004. About closedness by convolution of the Tsallis maximizers. Physica A340, 147–152.

Viola, P., Wells, W., 1995. Alignment by maximization of mutual information. In: Proc. 5th IEEE Int. Conf. on Computer Vision. Cambridge, MA, pp. 16–23.

Wolsztynski, E., Thierry, E., Pronzato, L., 2005. Minimum entropy estimation in semi parametric models. Signal Processing 85, 937–949.

Zografos, K., 1999. On maximum entropy characterization of Pearson's type II and VII multivariate distributions. Journal of Multivariate Analysis 71, 67–75.

Nikolai Leonenko
Cardiff University
Cardiff School of Mathematics
Senghennydd Road
Cardiff CF24 4AG, UK
e-mail: leonenkon@Cardiff.ac.uk

Luc Pronzato
Laboratoire I3S
CNRS/Université de Nice–Sophia Antipolis
Les Algorithmes, 2000 route des Lucioles
BP 121, 06903 Sophia-Antipolis Cedex, France
e-mail: pronzato@i3s.unice.fr

Vippal Savani
Cardiff University
Cardiff School of Mathematics
Senghennydd Road
Cardiff CF24 4AG, UK
e-mail: savaniv@Cardiff.ac.uk