Kernel-type estimators of divergence measures and its strong uniform consistency

Hamza Dhaker, Papa Ngom, El Hadji Deme, Pierre Mendy

Kernel-type estimators of divergence measures and its strong uniform consistency

Hamza Dhaker[a,b], Papa Ngom[a,*], El Hadji Deme[c], Pierre Mendy[b]

[a] *LMA,Laboratoire de Mathematiques Appliquees*
[b] *LMDAN,Universit Cheikh Anta Diop, Dakar-Fann, BP 5005, Senegal*
[c] *LERSTAD, Universite Gaston Berger, UFR SAT Saint-Louis, BP 234, Senegal*

Abstract

In this paper, we develop a kernel-type estimators of divergence measures for continuous distributions. We use the method based on empirical process techniques for consistence kernel-type function estimators to show a general result for the strong uniform consistency of our proposal divergence estimators.

Keywords: divergence mesures, kernel estimation, strong uniform, consistency
AMS Subject Classification : 62F12, 62E20.

---

1. Introduction

Given a samples from two distributions, one fundamental and classical question to ask is: how close are the two distributions? First, one must specify what it means for two distributions to be close, for which many different measures quantifying the degree of these distributions have been studied in the past. They are frequently called distance measures, although some of them are not strictly metrics. The divergence measures play an important role in statistical theory, especially in large theories of estimation and testing. They have been applied to different areas, such as medical image registration ([28]), classification and retrieval. In machine learning, it is often convenient to view training data as a set of distributions and use divergence measures to estimate dissimilarity between examples. This idea has been used in neuroscience, where the neural response pattern of an individual is modeled as a distribution, and divergence measures is used to compare responses across subjects (see, e.g [22]). Later many papers have appeared in the literature, where divergence or entropy type measures of information have been used in testing statistical hypotheses. For more examples and other possible applications of divergence measures, see the extended technical report ([30, 31]). For these applications and others, it is crucial to accurately estimate divergences.
The class of divergence measures is large; it includes the Rényi-α ([32, 33]), Tsallis-α ([38]), Kullback-Leibler (KL), Hellinger, Bhattacharyya, Euclidean divergences, etc. These divergence measures can be related to the Csiszár-f divergence ([5]). The Kullback-Leibler, Hellinger and Bhattacharyya are special cases of Rényi-α and Tsallis-α divergences. But the Kullback Leibler one is the most popular of these divergence measures. Estimation of divergence and its applications have been many studies using different approaches and specific. For example Pardo [27] presented methods and applications in the context of discrete distributions. By exploring a nonparametric method for estimating the divergence in the continuous case, Poczos and Schneider [30] use a k-nearest-neighbor estimator and show that one does not need a consistent density estimator to consistently estimate Rényi-α and Tsallis-α divergences.

In the nonparametric setting, a number of authors have proposed various estimators which are provably consistent. Krishnamurthy and Kandasamy [23] used an initial plug-in estimator for estimating by estimates of the higher order terms in the von Mises expansion of the divergence functional. In their frameworks, they proposed tree estimators for Rényi-α, Tsallis-α, and Euclidean divergences between two continuous distributions and established the rates of convergence of these estimators.

*Corresponding author
Email address: `papa.ngom@ucad.edu.sn` (Papa Ngom)

The main purpose of this paper is to analyze estimators for divergence measures between two continuous distributions. Our approach is similar on those of Krishnamurthy and Kandasamy [23] and is based on plug-in estimation scheme: first, apply a consistent density estimator for the underlying densities, and then plug them into the desired formulas. Unlike of their frameworks, we study the strong consistence estimators of a general class of divergence measures. We emphasize that the plug-in estimation technique are heavily used by [3, 14] in the case of entropy. Bouzebda [3] propose a method to establish consistency for kernel-type estimators of the differential entropy. We generalize this method for a large class of divergence measures in order to establish the consistency of kernel-type estimators of divergence measure when the bandwidth is allowed to range in a small interval which may decrease in length with the sample size. Our results will be immediately applicable to proving strong consistency for Kenel-type estimation of this class of divergence measures.

The rest of this paper is organized as follows: in Section 2, we introduce divergence measures and we construct their kernel-type estimators. In Section 3, we study the uniform strong consistency of the proposal estimators. Section 4 is devoted on the proofs.

2. Kenel-type estimators of Divergence Measures

Let us begin by standardizing notation and presenting some basic definitions. We will be concerned with two densities, $f, g : \mathbb{R}^d \mapsto [0,1]$ where $d \geq 1$ denotes the dimension. The divergence measures of interest are Rényi-α, Tsallis-α are defined respectively as follows

$$\mathcal{D}_\alpha^R(f,g) = \frac{1}{\alpha-1}\log\int_{\mathbb{R}^d} f^\alpha(x)g^{1-\alpha}(x)dx, \quad \alpha\in\mathbb{R}\setminus\{1\} \tag{1}$$

$$\mathcal{D}_\alpha^T(f,g) = \frac{1}{\alpha-1}\left(\int_{\mathbb{R}^d} f^\alpha(x)g^{1-\alpha}(x)dx - 1\right), \quad \alpha\in\mathbb{R}\setminus\{1\} \tag{2}$$

These quantities are nonnegative, and they are zero iff $f = g$ almost surely (a.s). Remark that In the special cases for $\alpha = 1/2, 1$, we obtain from (1) and (2) the well known Hellinger, Kullback and Bhattacharyya.

$$\begin{aligned}
\mathcal{D}_{1/2}^T(f,g) = 2\mathcal{D}^H(f,g) &= 2\left(1 - \int_{\mathbb{R}^d} f^{1/2}(x)g^{1/2}(x)dx\right),\\
\mathcal{D}_{1/2}^R(f,g) = 2\mathcal{D}^B(f,g) &= -\log\int_{\mathbb{R}^d} f^{1/2}(x)g^{1/2}(x)dx,\\
\lim_{\alpha\longrightarrow 1}\mathcal{D}_\alpha^R(f,g) = \mathcal{D}^{KL}(f,g) &= \int_{\mathbb{R}^d} f(x)\log\frac{f(x)}{g(x)}dx,
\end{aligned}$$

which is related to the Shannon entropy. For some statistical properties for the Shannon entropy, one can refer to [3].

$$H_S(f) = \int_{\mathbb{R}^d} f(x)\log f(x)dx$$

via

$$\mathcal{D}^{KL}(f,g) = V_S(f,g) - H_S(f)$$

where

$$V_S(f,g) = \int_{\mathbb{R}^d} f(x)\log g(x))dx$$

For the following, we focus only on the estimation of $\mathcal{D}_\alpha^T(f,g)$ and $\mathcal{D}_\alpha^R(f,g)$. The Kullback-Leibler, Hellinger and Bhattacharyya can be deducing immediately.
We will next provide consistent estimator for the following quantity

$$\mathcal{D}_\alpha(f,g) = \int_{\mathbb{R}^d} f^\alpha(x)g^{1-\alpha}(x)dx, \tag{3}$$

whenever this integral is meaningful. Plugging it estimates into the appropriate formula immediately leads to consistent estimator for the divergence measures $\mathcal{D}_\alpha^R(f,g)$, $\mathcal{D}_\alpha^T(f,g)$.

Now, assuming that for the rest of the document, the density f is unknown, and the density g known and satisfies : $\int_{\mathbb{R}^d} g^{1-\alpha}(x)dx$ is finite, this implies that $\mathcal{D}_\alpha(f,g)$ is finite. Next, consider $X_1,...,X_n, n \geq 1$ a sequence of independent and identically distributed $\mathbb{R}^d$-valued random vectors, with cumulative distribution function F a density function $f(\cdot)$ with respect to Lebesgue measure on $\mathbb{R}^d$. The following conditions are needed for the following sections. To construct our divergence estimators we define, We start by a kernel density estimator for $f(\cdot)$, and then substituting $f(\cdot)$ by its estimator in the divergence like functional of $f(\cdot)$. For this, we introduce a measurable function $K(\cdot)$ that satisfies the conditions.

(K.1) $K(\cdot)$ is of bounded variation on $\mathbb{R}^d$

(K.2) $K(\cdot)$ is right continuous on $\mathbb{R}^d$

(K.3) $\|K\|_\infty = \sup_{x\in\mathbb{R}^d} \mid K(x) \mid < \infty$

(K.4) $\int_{\mathbb{R}^d} K(t)dt = 1$.

Rosenblatt [34] first proposed an estimator $f(\cdot)$ and Parzen [26] generalizes thereafter eventually leading to the Parzen-Rosenblatt estimator, defined in the following way for any $x \in \mathbb{R}^d$

$$\widehat{f_{n,h_n}}(x) = \frac{1}{nh_n^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \tag{4}$$

where $0 < h_n < 1$ is the bandwidth sequence. Assuming that the density f is continuous, one obtains a strongly consistent estimator $\widehat{f_{n,h_n}}$ of f, that is, one has with probability 1, $\widehat{f_{n,h_n}}(x) \longrightarrow f(x)$, $x \in \mathbb{R}^d$. There are also results concerning uniform convergence and convergence rates. For proving such results one usually writes the difference $\widehat{f_{n,h_n}}(x) - f(x)$ as the sum of a probabilistic term $\widehat{f_{n,h_n}}(x) - \mathbb{E}\widehat{f_{n,h_n}}(x)$ and a deterministic term $\mathbb{E}\widehat{f_{n,h_n}}(x) - f(x)$, the so-called bias. On can refer to [15, 18, 20] , among other authors.

After having estimated $\widehat{f_{n,h_n}}(\cdot)$, we estimate $\mathcal{D}_\alpha(f,g)$ by setting

$$\widehat{\mathcal{D}}_\alpha(\widehat{f_{n,h_n}}, g) = \int_{A_{n,h_n}} \widehat{f}^{\alpha}_{n,h_n}(x) g^{1-\alpha}(x)dx, \quad \alpha \neq 1 \tag{5}$$

where $A_{n,h_n} = \{x \in \mathbb{R}^d, \ \ \widehat{f_{n,h_n}}(x) \geq \gamma_n\}$ and $\gamma_n \downarrow 0$ is a sequence of positive constant. Thus, using 5, the associated divergences $\mathcal{D}_\alpha^R(f,g)$ and $\mathcal{D}_\alpha^T(f,g)$ can be estimated by:

$$\begin{aligned}
\widehat{\mathcal{D}}_\alpha^R(\widehat{f_{n,h_n}}, g) &= \frac{1}{\alpha - 1}\log \widehat{\mathcal{D}}_\alpha(\widehat{f_{n,h_n}}, g),\\
\widehat{\mathcal{D}}_\alpha^T(\widehat{f_{n,h_n}}, g) &= \frac{1}{\alpha - 1}\left(\widehat{\mathcal{D}}_\alpha(\widehat{f_{n,h_n}}, g) - 1\right).
\end{aligned}$$

The approach use to define the plug-in estimators is also developed in [3] in order to introduce a kernel-type estimators of Shannon's entropy. Some statistical properties of these divergences is related on those of the kernel estimator $\widehat{f_{n,h_n}}(\cdot)$ of the continuous density f.

The limiting behavior of $\widehat{f_{n,h_n}}(\cdot)$, for appropriate choices of the bandwidth h_n, has been widely studied in the literature, examples include the work of Deroye [10, 11] Bosq [2] and Prakasa [29]. In particular, under our assumptions, the condition that $h_n \downarrow 0$ together with $nh_n \uparrow \infty$ is necessary and sufficient for the convergence in probability of $\widehat{f_{n,h_n}}(x)$ towards the limit $f(x)$, independently of $x \in \mathbb{R}^d$ and the density $f(\cdot)$. We can find other results of uniform consistency of the estimator $\widehat{f_{n,h_n}}(x)$ in [6, 15, 9] and the references therein. In the next section, we will use their methods to establish convergence results for the estimates $\widehat{\mathcal{D}}_\alpha(\widehat{f_{n,h_n}}, g)$ and deduce the convergence results of $\widehat{\mathcal{D}}_\alpha^R(\widehat{f_{n,h_n}}, g)$ and $\widehat{\mathcal{D}}_\alpha^T(\widehat{f_{n,h_n}}, g)$.

3. Statistical properties of the estimators

We first study the strong consistency of the estimator $\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h_n}, g)$ defined in (5). We shall consider another, but more appropriate and more computationally convenient, centering factor than the expectation $\mathbb{E}\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h_n}, g)$ which is delicate to handle. This is given by

$$\widehat{\mathbb{E}\widehat{\mathcal{D}}}_\alpha(\widehat{f}_{n,h_n}, g) := \int_{A_{n,h_n}} \left(\mathbb{E}\widehat{f}_{n,h_n}(x)\right)^\alpha g^{1-\alpha}(x)dx.$$

Lemma 1. *Let $K(\cdot)$ satisfy (**K.1-2-3-4**) and let $f(\cdot)$ be a continuous bounded density . Then, for each pair of sequence $(h_n^{'})_{n\geq1}$, $(h_n^{''})_{n\geq1}$ such that $0 < h_n^{'} < h_n \leq h_n^{''}$, together with $h_n^{''} \longrightarrow 0$, $nh_n^{'}/\log(n) \longrightarrow \infty$ as $n \longrightarrow \infty$, for any $\alpha \in (0,1)$, one has with probability 1*

$$\sup_{h_n^{'}\leq h\leq h_n^{''}} \left|\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h}, g) - \widehat{\mathbb{E}\widehat{\mathcal{D}}}_\alpha(\widehat{f}_{n,h}, g)\right| = O\left(\left(\frac{\log(1/h_n^{'}) \vee \log\log n}{nh_n^{'}}\right)^{\alpha/2}\right).$$

The proof of Lemma 1 is postponed until Section 5.

Lemma 2. *Let $K(\cdot)$ satisfy (**K.3-4**) and let $f(\cdot)$ be a uniformly Lipschitz and continuous density. Then, for each pair of sequence $(h_n^{'})_{n\geq1}$, $(h_n^{''})_{n\geq1}$ such that $0 < h_n^{'} < h_n \leq h_n^{''}$, together with $h_n^{''} \longrightarrow 0$, as $n \longrightarrow \infty$, for any $\alpha \in (0,1)$, we have*

$$\sup_{h_n^{'}\leq h\leq h_n^{''}} \left|\widehat{\mathbb{E}\widehat{\mathcal{D}}}_\alpha(\widehat{f}_{n,h}, g) - \mathcal{D}_\alpha(f, g)\right| = O\left(\gamma_n^\alpha \vee h_n^{''\alpha/d}\right).$$

The proof of Lemma 2 is postponed until Section 5.

Theorem 1. *Let $K(\cdot)$ satisfy (**K.1-2-3-4**) and let $f(\cdot)$ be a uniformly Lipschitz, bounded and continuous density. Then, for each pair of sequence $(h_n^{'})_{n\geq1}$, $(h_n^{''})_{n\geq1}$ such that $0 < h_n^{'} < h_n \leq h_n^{''}$, together with $h_n^{''} \longrightarrow 0$, $nh_n^{'}/\log(n) \longrightarrow \infty$ as $n \longrightarrow \infty$, for any $\alpha \in (0,1)$, one has with probability 1*

$$\sup_{h_n^{'}\leq h\leq h_n^{''}} \left|\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h}, g) - \mathcal{D}_\alpha(f, g)\right| = O\left(\left(\frac{\log(1/h_n^{'}) \vee \log\log n}{nh_n^{'}}\right)^{\alpha/2} \vee \gamma_n^\alpha \vee h_n^{''\alpha/d}\right).$$

This, in turn, implies that

$$\lim_{n\to\infty} \sup_{h_n^{'}\leq h\leq h_n^{''}} \left|\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h}, g) - \mathcal{D}_\alpha(f, g)\right| = 0 \qquad a.s. \tag{6}$$

The proof of Theorem 1 is postponed until Section 5.
The following corollaries handle respectively the uniform deviation of the estimate $\widehat{\mathcal{D}}_\alpha^T(\widehat{f}_{n,h}, g)$ and $\widehat{\mathcal{D}}_\alpha^R(\widehat{f}_{n,h}, g)$ with respect to $\mathcal{D}_\alpha^T(f, g)$ and $\mathcal{D}_\alpha^R(f, g)$.

Corollary 1. *Assuming that the assumptions of Theorem 1 hold. Then, we have*

$$\sup_{h_n^{'}\leq h\leq h_n^{''}} \left|\widehat{\mathcal{D}}_\alpha^T(\widehat{f}_{n,h}, g) - \mathcal{D}_\alpha^T(f, g)\right| = O\left(\left(\frac{\log(1/h_n^{'}) \vee \log\log n}{nh_n^{'}}\right)^{\alpha/2} \vee \gamma_n^\alpha \vee h_n^{''\alpha/d}\right).$$

This, in turn, implies that

$$\lim_{n\to\infty} \sup_{h_n^{'}\leq h\leq h_n^{''}} \left|\widehat{\mathcal{D}}_\alpha^T(\widehat{f}_{n,h}, g) - \mathcal{D}_\alpha^T(f, g)\right| = 0 \qquad a.s. \tag{7}$$

The proof of Corollary 1 is postponed until Section 5.

Corollary 2. *Assuming that the assumptions of Theorem 1 hold. Then, we have*

$$\sup_{h'_n\leq h\leq h''_n}\left|\widehat{\mathcal{D}}^R_\alpha(\widehat{f}_{n,h},g)-\mathcal{D}^R_\alpha(f,g)\right| = O\left(\left(\frac{\log(1/h'_n)\vee\log\log n}{nh'_n}\right)^{\alpha/2}\vee\gamma_n^\alpha\vee h_n''^{\alpha/d}\right)$$

This, in turn, implies that

$$\lim_{n\rightarrow\infty}\sup_{h'_n\leq h\leq h''_n}\left|\widehat{\mathcal{D}}^R_\alpha(\widehat{f}_{n,h},g)-\mathcal{D}^R_\alpha(f,g)\right| = 0 \qquad a.s. \tag{8}$$

The proof of Corollary 2 is postponed until Section 5.
Note that, the divergence estimation such as (5) also requires the appropriate choice of the smoothing parameter h_n. The result given in (6), (7) and (8) show that any choice of h between h'_n and h''_n ensures the strong consistency of the underlying divergence estimates. In other word, the fluctuation of the bandwidth in a small interval do not affect the consistency of the nonparametric estimator of these divergences.

The work Bouzebda and Elhattab [3] very important for establishing of our results, the authors have created a class of compactly supported densities. We need the following additional conditions.

(**F.1**) $f(\cdot)$ has a compact support say $\mathbb{I}$ and is s-time continuously differentiable, and there exists a constant $0<M<\infty$ such that

$$\sup_{x\in\mathbb{I}}\left|\frac{\partial^s f(x)}{\partial x_1^{j_1}\ldots\partial x_d^{j_d}}\right|\leq M,\quad j_1+\cdots+j_d=s.$$

(**K.5**) $K(\cdot)$ is of order s, i.e., for some constant $\varrho\neq 0$,

$$\int_{\mathbb{R}^d}u_1^{j_1}\ldots u_d^{j_d}K(u)du=0,\quad j_1,\ldots,j_d\geq 0,\quad j_1+\cdots+j_d=1,\ldots,s-1,$$

and

$$\int_{\mathbb{R}^d}|u_1^{j_1}\ldots u_d^{j_d}|K(u)du=\varrho,\quad j_1,\ldots,j_d\geq 0,\quad j_1+\cdots+j_d=s.$$

Under (**F.1**) the expression $\mathcal{D}_\alpha(f,g)$ may be written as follows

$$\mathcal{D}_\alpha(f,g)=\int_{\mathbb{I}}f^\alpha(x)g^{1-\alpha}dx. \tag{9}$$

Theorem 2. *Assuming conditions (**K.1-2-3-4-5**) hold. Let $f(\cdot)$ fulfill (**F.1**). Then for each pair of sequences $0<h_n^{'}<h_n\leq h_n^{''}$ with $h_n^{''}\longrightarrow 0$, $nh'_n/\log n\longrightarrow\infty$ as $n\longrightarrow\infty$, for any $\alpha\in(0,1)$,, we have*

$$\limsup_{n\longrightarrow\infty}\sup_{h'_n\leq h\leq h''_n}\frac{\sqrt{(nh)^\alpha}\left|\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h},g)-\mathcal{D}_\alpha(f,g)\right|}{\sqrt{(\log(1/h)\vee\log\log n)^\alpha}}\leq\zeta(\mathbb{I})\int_{\mathbb{R}^d}g^{1-\alpha}(x)dx\quad a.s,$$

where

$$\zeta(\mathbb{I})=\sup_{x\in\mathbb{I}}\left\{f(x)\int_{\mathbb{R}^d}K^2(u)du\right\}^{\alpha/2}.$$

The proof of Theorem 2 is postponed until Section 5.

Corollary 3. *Assuming that the assumptions of the Theorem 2 hold. Then,*

$$\limsup_{n\longrightarrow\infty}\sup_{h'_n\leq h\leq h''_n}\frac{\sqrt{(nh)^\alpha}\left|\widehat{\mathcal{D}}^T_\alpha(\widehat{f}_{n,h},g)-\mathcal{D}^T_\alpha(f,g)\right|}{\sqrt{(\log(1/h)\vee\log\log n)^\alpha}}\leq\frac{1}{1-\alpha}\zeta(\mathbb{I})\int_{\mathbb{R}^d}g^{1-\alpha}(x)dx\quad a.s,$$

Corollary 4. *Assuming that the assumptions of the Theorem 2 hold. Then, for any $\gamma > 0$ we have*

$$\limsup_{n\longrightarrow\infty} \sup_{h_n'\leq h\leq h_n''} \frac{\sqrt{(nh)^{\alpha}}\left|\widehat{\mathcal{D}}_{\alpha}^{R}(\widehat{f}_{n,h}, g) - \mathcal{D}_{\alpha}^{R}(f, g)\right|}{\sqrt{(\log(1/h) \vee \log\log n)^{\alpha}}} \leq \frac{1}{(1-\alpha)\gamma^{\alpha}}\zeta(\mathbb{I}) \ \ a.s,$$

The proof of Corollaries 3 and 4 are given in Section 5.

Now, assume that there exists a sequence $\{\mathbb{I}_n\}_{n\geq 1}$ of strictly nondecreasing compact subsets of $\mathbb{I}$, such that $\mathbb{I} = \cup_{n\geq 1}\mathbb{I}_n$. For the estimation of the support $\mathbb{I}$ we may refer to ([12]) and the references therein. Throughout, we let $h \in [h_n', h_n'']$, where h_n' and h_n'' are as in Corollaries (3) and (4). We chose an estimator of $\zeta(\mathbb{I})$ in the Corollaries (3) and (4) as the form

$$\zeta_n(\mathbb{I}_n) = \sup_{x\in\mathbb{I}_n}\left\{\widehat{f}_{n,h}(x)\int_{\mathbb{R}^d} K^2(u)du\right\}^{\alpha/2}.$$

Using the techniques developed in [9] , the Corollaries (3) and (4) one can construct a asymptotic 100% certainty intervals for the true divergences $\mathcal{D}_{\alpha}^{T}(f, g)$, $\mathcal{D}_{\alpha}^{R}(f, g)$.

4. Concluding remarks and future works

In this paper we are concerned with the problem of nonparametric estimation of a class of divergence measures. For this cause, many estimators are available. The most recent ones are the estimates developed by Bouzebal [3] . We introduce an estimator that can be seen as generalization to those previously suggested, in that he was interested in the estimation of entropy. We are focusing on the Rényi-α and the Tsallis-α divergence measures. Under our study, one can easily deduced Kullback-Leibler, Hellinger and Bhattacharyya nonparmetric estimators. The results presented in this work are general, since the required conditions are fulfilled by a large class of densities. We mention that the estimator $\widehat{\mathcal{D}}_{\alpha}(\widehat{f}_{n,h_n}, g)$ in (5) can be calculated by using a Monte-Carlo method under the density g. And a pratical choice of γ_n is $\beta(\log n)^{\delta}$ where $\beta > 0$ and $\delta \geq 0$.
It will be interesting to enrich our results presented here by an additional uniformity in term of γ_n in the supremum appearing in all our theorems, which requires non trivial mathematics, this would go well beyond the scope of the present paper. Another direction of research is to obtain results, in the case where the continuous distributions f and g are both unknown. The problems and the methods described here all are inherently univariate. A natural and useful multivariate extension is the use of copula function.

5. Proofs of our results

Proof of Lemma 1. To prove the strong consistency of $\widehat{\mathcal{D}}_{\alpha}(\widehat{f}_{n,h_n}, g)$, we use the following expression

$$\widehat{\mathbb{E}\mathcal{D}}_{\alpha}(\widehat{f}_{n,h_n}, g) := \int_{A_{n,h_n}} \left(\mathbb{E}\widehat{f}_{n,h_n}(x)\right)^{\alpha} g^{1-\alpha}(x)dx,$$

where $A_{n,h_n} = \{x \in \mathbb{R}^d, \ \ \widehat{f}_{n,h_n}(x) \geq \gamma_n\}$ and $\gamma_n \downarrow 0$ is a sequence of positive constant. Define

$$\Delta_{n,1,h_n} := \underbrace{\widehat{\mathcal{D}}_{\alpha}(\widehat{f}_{n,h_n}, g) - \widehat{\mathbb{E}\mathcal{D}}_{\alpha}(\widehat{f}_{n,h_n}, g)}.$$

We have

$$\begin{aligned}
|\Delta_{n,1,h_n}| &= \left|\int_{A_{n,h_n}} \left(\widehat{f}_{n,h_n}^{\alpha}(x) - \left(\mathbb{E}\widehat{f}_{n,h_n}(x)\right)^{\alpha}\right) g^{1-\alpha}(x)dx\right| \\
&\leq \int_{A_{n,h_n}} \left|\widehat{f}_{n,h_n}^{\alpha}(x) - \left(\mathbb{E}\widehat{f}_{n,h_n}(x)\right)^{\alpha}\right| g^{1-\alpha}(x)dx \\
&\leq \sup_{x\in\mathbb{R}^d}\left|\widehat{f}_{n,h_n}^{\alpha}(x) - \left(\mathbb{E}\widehat{f}_{n,h_n}(x)\right)^{\alpha}\right| \int_{A_{n,h_n}} g^{1-\alpha}(x)dx.
\end{aligned}$$

Since $h(x) = x$ is a 1-Lipschitz function, for $0 < \alpha < 1$ then
$| (h(x))^\alpha - (h(y))^\alpha | \leq | h(x) - h(y) |^\alpha$.
Therefore for $0 < \alpha < 1$, we have

$$\left| \widehat{f}^{\alpha}_{n,h_n}(x) - \left(\mathbb{E}\widehat{f}_{n,h_n}(x) \right)^\alpha \right| \leq \left| \widehat{f}_{n,h_n}(x) - \mathbb{E}\widehat{f}_{n,h_n}(x) \right|^\alpha \leq \left\| \widehat{f}_{n,h_n} - \mathbb{E}\widehat{f}_{n,h_n} \right\|^\alpha_\infty,$$

where $\|\cdot\|_\infty$ denotes, as usual, the supremum norm, i.e., $\|\varphi\|_\infty := \sup_{x\in\mathbb{R}} |\varphi(x)|$. Hence,

$$|\Delta_{n,1,h_n}| \leq \left\| \widehat{f}_{n,h_n} - \mathbb{E}\widehat{f}_{n,h_n} \right\|^\alpha_\infty \int_{A_{n,h_n}} g^{1-\alpha}(x)dx. \tag{10}$$

Finaly,

$$|\Delta_{n,1,h_n}| \leq \left\| \widehat{f}_{n,h_n} - \mathbb{E}\widehat{f}_{n,h_n} \right\|^\alpha_\infty \int_{\mathbb{R}^d} g^{1-\alpha}(x)dx. \tag{11}$$

Using the conditions on the kernel $K(\cdot)$, posed by Einmahl [16]. Consider the class of functions

$$\mathcal{K} := \left\{ K\left((x - \cdot)/h\right) : h > 0,\ x \in \mathbb{R}^d \right\}.$$

For $\varepsilon > 0$, set $N(\varepsilon, \mathcal{K}) = \sup_Q N(\kappa\varepsilon, \mathcal{K}, d_Q)$, where the supremum is taken over all probability measures Q on $(\mathbb{R}^d, \mathcal{B})$, where $\mathcal{B}$ represents the σ-field of Borel sets of $\mathbb{R}^d$, i.e is the smallest containing all the open (and/or closed) balls is $\mathbb{R}^d$. Here, d_Q denotes the $L_2(Q)$-metric and $N(\kappa\varepsilon, \mathcal{K}, d_Q)$ is the minimal number of balls $\{\psi : d_Q(\psi, \psi') < \varepsilon\}$ of d_Q-radius ε needed to cover $\mathcal{K}$.
We assume that $\mathcal{K}$ satisfies the following uniform entropy condition.

(**K.6**) for some $C > 0$ and $\nu > 0$, $N(\varepsilon, \mathcal{K}) \leq C\varepsilon^{-\nu}, 0 < \varepsilon < 1$.

(**K.7**) $\mathcal{K}$ is a pointwise measurable class, that is there exists a countable sub-class $\mathcal{K}_0$ of $\mathcal{K}$ such that we can find for any function $\psi \in \mathcal{K}$ a sequence of functions $\{\psi_m : m \geq 1\}$ in $\mathcal{K}_0$ for which

$$\psi_m(z) \longrightarrow \psi(z), \quad z \in \mathbb{R}^d.$$

This condition is discussed in [34] . It is satisfied whenever K is right continuous.

*Remark that condition (**K.6**) is satisfied whenever (**K.1**) holds, i.e., $K(\cdot)$ is of bounded variation on $\mathbb{R}^d$, we refer the reader to Van der Vaart and Wellner [36], for details on conditions of entropy (see also Pakes and Pollard [25] and Nolan and Pollard [24]). Condition (**K.7**) is satisfied whenever (**K.2**) holds, i.e., $K(\cdot)$ is right continuous, condition is discussed in [36], (see also [9] and [16]).*

From Theorem 1 in [16], whenever $K(\cdot)$ is measurable and satisfies (**K.3-4-6-7**), and when $f(\cdot)$ is bounded, we have for each pair of sequence $(h'_n)_{n\geq1}$, $(h''_n)_{n\geq1}$ such that $0 < h'_n < h \leq h''_n \leq 1$, together with $h''_n \to 0$ and $nh'_n/\log(n) \to \infty$ as $n \longrightarrow \infty$, with probability 1

$$\sup_{h'_n \leq h \leq h''_n} \left\| \widehat{f}_{n,h} - \mathbb{E}\widehat{f}_{n,h} \right\|_\infty = O\left(\sqrt{\frac{\log(1/h'_n) \vee \log\log n}{nh'_n}} \right). \tag{12}$$

Since $\int_{\mathbb{R}^d} g^{1-\alpha}(x)dx < \infty$, in view of (11) and (12), we obtain with probability 1

$$\sup_{h'_n \leq h \leq h''_n} \left| \Delta_{n,1,h} \right| = O\left(\left(\frac{\log(1/h'_n) \vee \log\log n}{nh'_n} \right)^{\alpha/2} \right). \tag{13}$$

It concludes the proof of the lemma.
Proof of Lemma 2.
Let A^c_{n,h_n} be the complement of A_{n,h_n} in $\mathbb{R}^d$ ($i.e$, $A^c_{n,h_n} = \{x \in \mathbb{R}^d,\ \widehat{f}_{n,h_n} < \gamma_n\}$). We have

$$\widehat{\mathbb{E}\mathcal{D}}_\alpha(\widehat{f}_{n,h_n}, g) - \mathcal{D}_\alpha(f, g) = \Delta_{n,2,h_n} + \Delta_{n,3,h_n},$$

with

$$\Delta_{n,2,h_n} := \int_{A_{n,h_n}} \left(\left(\mathbb{E}\widehat{f}_{n,h_n}(x)\right)^{\alpha} - f^{\alpha}(x)\right) g^{1-\alpha}(x)dx$$

and

$$\Delta_{n,3,h_n} := \int_{A^c_{n,h_n}} f^{\alpha}(x) g^{1-\alpha}(x)dx.$$

Term $\Delta_{n,2,h_n}$. Repeat the arguments above in the terms $\Delta_{n,1,h_n}$ with the formal change of $\widehat{f}_{n,h_n}$ by f. We show that, for any $n \geq 1$,

$$|\Delta_{n,2,h_n}| \leq \left\|\mathbb{E}\widehat{f}_{n,h_n} - f\right\|_{\infty}^{\alpha} \int_{A_{n,h_n}} g^{1-\alpha}(x)dx, \tag{14}$$

which implies

$$|\Delta_{n,2,h_n}| \leq \left\|\mathbb{E}\widehat{f}_{n,h_n} - f\right\|_{\infty}^{\alpha} \int_{\mathbb{R}^d} g^{1-\alpha}(x)dx. \tag{15}$$

On the other hand, we know (see, e.g,[16]), that since the density $f(\cdot)$ is uniformly Lipschitz and continuous, we have for each sequences $h'_n < h < h''_n < 1$, with $h''_n \to 0$, as $n \to \infty$,

$$\sup_{h'_n \leq h \leq h''_n} \left\|\mathbb{E}\widehat{f}_{n,h_n} - f\right\|_{\infty} = O(h''^{1/d}_n). \tag{16}$$

Thus,

$$\sup_{h'_n \leq h \leq h''_n} \left|\Delta_{n,2,h}\right| = O(h''^{\alpha/d}_n). \tag{17}$$

Term $\Delta_{n,3,h_n}$. It is obsious to see that

$$\begin{aligned}
\left|\Delta_{n,3,h_n}\right| &= \textstyle\int_{A^c_{n,h_n}} |f^{\alpha}(x)| g^{1-\alpha}(x)dx \\
&\leq \textstyle\int_{A^c_{n,h_n}} \left|E\widehat{f}_{n,h_n}(x) - f^{\alpha}(x)\right| g^{1-\alpha}(x)dx + \int_{A^c_{n,h_n}} E\widehat{f}_{n,h_n}(x) g^{1-\alpha}(x)dx \\
&\leq \left\|\mathbb{E}\widehat{f}_{n,h_n} - f\right\|_{\infty}^{\alpha} \textstyle\int_{A^c_{n,h_n}} g^{1-\alpha}(x)dx + \gamma_n^{\alpha} \int_{A^c_{n,h_n}} g^{1-\alpha}(x)dx
\end{aligned}$$

Thus,

$$\left|\Delta_{n,3,h_n}\right| \leq \left(\left\|\mathbb{E}\widehat{f}_{n,h_n} - f\right\|_{\infty}^{\alpha} + \gamma_n^{\alpha}\right) \int_{A^c_{n,h_n}} g^{1-\alpha}(x)dx. \tag{18}$$

Hence,

$$\left|\Delta_{n,3,h_n}\right| \leq \left(\left\|\mathbb{E}\widehat{f}_{n,h_n} - f\right\|_{\infty}^{\alpha} + \gamma_n^{\alpha}\right) \int_{\mathbb{R}^d} g^{1-\alpha}(x)dx. \tag{19}$$

Thus, in view of (16), we get

$$\sup_{h'_n \leq h \leq h''_n} \left|\Delta_{n,3,h_n}\right| = O\left(\gamma_n^{\alpha} \vee h''^{\alpha/d}_n\right) \tag{20}$$

Finaly, in view of (17) and (20), we get

$$\sup_{h'_n \leq h \leq h''_n} \left|\mathbb{E}\widehat{\mathcal{D}}_{\alpha}(\widehat{f}_{n,h}, g) - \mathcal{D}_{\alpha}(f, g)\right| = O\left(\gamma_n^{\alpha} \vee h''^{\alpha/d}_n\right). \tag{21}$$

It concludes the proof of the lemma.

Proof of Theorem 1. We have

$$\left|\widehat{\mathcal{D}}^R_{\alpha}(\widehat{f}_{n,h_n}, g) - \mathcal{D}^R_{\alpha}(f, g)\right| \leq \left|\widehat{\mathcal{D}}_{\alpha}(\widehat{f}_{n,h}, g) - \widehat{E\mathcal{D}}_{\alpha}(\widehat{f}_{n,h}, g)\right| + \left|\widehat{E\mathcal{D}}_{\alpha}(\widehat{f}_{n,h}, g) - \mathcal{D}_{\alpha}(f, g)\right|.$$

Combinating the Lemmas (1) and (2), we obtain

$$\sup_{h_n'\leq h\leq h_n''}\left|\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h_n},g)-\mathcal{D}_\alpha(f,g)\right| = O\left(\left(\frac{\log(1/h_n')\vee\log\log n}{nh_n'}\right)^{\alpha/2}\right)+O\left(\gamma_n^\alpha\vee h_n''^{\alpha/d}\right).$$

It concludes the proof of the Theorem.

Proof of Corollary 1. Remark that

$$\widehat{\mathcal{D}}_\alpha^T(\widehat{f}_{n,h_n},g)-\mathcal{D}_\alpha^T(f,g)=\frac{1}{\alpha-1}\left(\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h_n},g)-\mathcal{D}_\alpha(f,g)\right).$$

Using the Theorem (1), we have

$$\sup_{h_n'\leq h\leq h_n''}\left|\widehat{\mathcal{D}}_\alpha^T(\widehat{f}_{n,h_n},g)-\mathcal{D}_\alpha^T(f,g)\right| = O\left(\left(\frac{\log(1/h_n')\vee\log\log n}{nh_n'}\right)^{\alpha/2}\vee\gamma_n^\alpha\vee h_n''^{\alpha/d}\right),$$

and the Corollary 1 holds

Proof of Corollary 2. A first order taylor expansion of $y\mapsto\log y$ arround $y=y_0>0$ and $y=\widehat{y}>0$ gives

$$\log\widehat{y}=\log y_0+\frac{1}{y_0}(\widehat{y}-y_0)+o(\|\widehat{y}-y_0\|).$$

Remark that from Theorem 1,

$$\sup_{h_n'\leq h\leq h_n''}\left|\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h},g)-\mathcal{D}_\alpha(f,g)\right| = O\left(\left(\frac{\log(1/h_n')\vee\log\log n}{nh_n'}\right)^{\alpha/2}\vee\gamma_n^\alpha\vee h_n''^{\alpha/d}\right),$$

which turn, implies that

$$\lim_{n\to\infty}\sup_{h_n'\leq h\leq h_n''}\left|\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h},g)-\mathcal{D}_\alpha(f,g)\right|=0 \qquad a.s.$$

Thus, for all

$$\begin{aligned}\widehat{\mathcal{D}}_\alpha^R(\widehat{f}_{n,h_n},g)-\mathcal{D}_\alpha^R(f,g) &= \frac{1}{\alpha-1}\left(\log\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h},g)-\log\mathcal{D}_\alpha(f,g)\right)\\ &= \frac{1}{(\alpha-1)\mathcal{D}_\alpha(f,g)}\left(\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h},g)-\mathcal{D}_\alpha(f,g)\right)\\ &\quad +o\left(\left\|\widehat{\mathcal{D}}_\alpha(\widehat{f}_{n,h},g)-\mathcal{D}_\alpha(f,g)\right\|\right).\end{aligned}$$

Consequently

$$\sup_{h_n'\leq h\leq h_n''}\left|\widehat{\mathcal{D}}_\alpha^R(\widehat{f}_{n,h_n},g)-\mathcal{D}_\alpha^R(f,g)\right| = O\left(\left(\frac{\log(1/h_n')\vee\log\log n}{nh_n'}\right)^{\alpha/2}\vee\gamma_n^\alpha\vee h_n''^{\alpha/d}\right),$$

and the Corollary 2 holds.

Proof of Theorem 2. Under conditions (**F.1**), (**K.5**) and using Taylor expansion of order s we get, for $x\in\mathbb{I}$,

$$|\mathbb{E}\widehat{f}_{n,h_n}-f(x)|=\frac{h^{s/d}}{s!}\left|\int\sum_{k_1+\ldots+k_d}t_1^{k_1}\ldots t_d^{k_d}\frac{\partial^s f(x-h\theta t)}{\partial x_1^{k_1}\ldots\partial x_1^{k_d}}K(t)dt\right|$$

where $\theta = (\theta_1, ..., \theta_d)$ and $0 < \theta_i < 1, \quad i = 1, ...; d$ Thus a straightforward application of Lebesgue dominated convergence theorem gives, for n large enough,

$$\sup_{x\in\mathbb{I}} |\mathbb{E}\widehat{f_{n,h}}(x) - f(x)| = O(h_n^{''})$$

Let $\mathbb{J}$ be a nonempty compact subset of the interior of $\mathbb{I}$. First, note that we have from Corollary 3.1.2. p. 62 of Viallon [37] (see also, [3] , statement (4.16)).

$$\limsup_{n\to\infty} \sup_{h_n'\le h\le h_n'} \sup_{x\in\mathbb{J}} \frac{\sqrt{nh}|\widehat{f_{n,h}}(x) - f(x)|}{\sqrt{\log(1/h) \vee \log\log n}} = \sup_{x\in\mathbb{J}} \left(f(x) \int_{\mathbb{R}^d} K^2(t)dt \right)^{1/2} \tag{22}$$

Set, for all $n \ge 1$,

$$\begin{aligned}
\pi_n(\mathbb{J}) &= \left| \int_{\mathbb{J}} \left(\widehat{f^{\alpha}_{n,h}}(x) - f^{\alpha}(x) \right) g^{1-\alpha}(x)dx \right| \\
&\le \int_{\mathbb{J}} |\widehat{f^{\alpha}_{n,h}}(x) - f^{\alpha}(x)| g^{1-\alpha}(x)dx \\
&\le \int_{\mathbb{J}} |\widehat{f_{n,h}}(x) - f(x)|^{\alpha} g^{1-\alpha}(x)dx \quad \textit{since} \ \ \alpha \in]0,1[, \\
&\le \sup_{x\in\mathbb{J}} |\widehat{f_n}(x) - f(x)|^{\alpha} \int_{\mathbb{J}} g^{1-\alpha}(x)dx, \qquad (23)\\
&\le \sup_{x\in\mathbb{J}} |\widehat{f_n}(x) - f(x)|^{\alpha} \int_{\mathbb{R}^d} g^{1-\alpha}(x)dx. \qquad (24)
\end{aligned}$$

One fined, by combining (22) and (24)

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{h_n'\le h\le h_n''} \frac{\sqrt{(nh)^{\alpha}}\pi_n(\mathbb{J})}{\sqrt{(\log(1/h) \vee \log\log n)^{\alpha}}} \\
&\qquad \le \sup_{x\in\mathbb{J}} \left\{ \left(f(x) \int_{\mathbb{R}^d} K^2(t)dt \right)^{\alpha/2} \right\} \int_{\mathbb{R}^d} g^{1-\alpha}(x)dx. \qquad (25)
\end{aligned}$$

Let $\{\mathbb{J}_\ell\}, \quad \ell = 1, 2, ...,$ be a sequence of nondecreasing nonempty compact subsets of the interior of $\mathbb{I}$ such that

$$\bigcup_{\ell\ge1} \mathbb{J}_\ell = \mathbb{I}$$

Now, from (25), it is straightforward to observe that

$$\begin{aligned}
&\limsup_{\ell\to\infty} \limsup_{n\to\infty} \sup_{h_n'\le h\le h_n''} \frac{\sqrt{(nh)^{\alpha}}\pi_n(\mathbb{J}_\ell)}{\sqrt{(\log(1/h) \vee \log\log n)^{\alpha}}} \\
&\qquad \le \limsup_{\ell\to\infty} \sup_{x\in\mathbb{J}_\ell} \left\{ f(x) \int_{\mathbb{R}^d} K^2(t)dt\}^{\alpha/2} \right\} \int_{\mathbb{R}^d} g^{1-\alpha}(x)dx \\
&\qquad \le \sup_{x\in\mathbb{I}} \left\{ \left(f(x) \int_{\mathbb{R}^d} K^2(t)dt \right)^{\alpha/2} \right\} \int_{\mathbb{R}^d} g^{1-\alpha}(x)dx
\end{aligned}$$

The proof of Theorem 2 is completed.

Proof of Corollary 3. A direct application of the Theorem 2 leeds to the Corollary 3.

Proof of Corollary 4. Here again, set, for all $n \ge 1$,

$$\eta_n(\mathbb{J}) = \left|\frac{1}{\alpha-1}\left(\log\int_{\mathbb{J}}\widehat{f}^{\alpha}_{n,h}(x)g^{1-\alpha} - \log\int_{\mathbb{J}} f^{\alpha}(x)g^{1-\alpha}(x)dx\right)\right|.$$

A first order Taylor expansion of $\log(y)$ leads to

$$\begin{aligned}
\eta_n(\mathbb{J}) &\leq \tfrac{1}{1-\alpha}\tfrac{1}{\int_{\mathbb{J}} f^{\alpha}(x)g^{1-\alpha}(x)dx}\left|\int_{\mathbb{J}}\left(\widehat{f}^{\alpha}_{n,h}(x) - f^{\alpha}(x)\right)g^{1-\alpha}(x)dx\right| + o\left(\|\widehat{f}^{\alpha}_{n,h} - f\|^{\alpha}_{\infty}\right),\\
&\leq \tfrac{1}{1-\alpha}\tfrac{1}{\int_{\mathbb{J}} f^{\alpha}(x)g^{1-\alpha}(x)dx}\pi_n(\mathbb{J}) + o\left(\|\widehat{f}^{\alpha}_{n,h} - f\|^{\alpha}_{\infty}\right),
\end{aligned}$$

Using condition $(F.1)$, $f(\cdot)$ is compactly supported), $f(\cdot)$ is bounded away from zero on its support, thus, we have for n enough large, there exists $\gamma > 0$, such that $f(x) > \gamma$, for all x in the support of $f(\cdot)$. From (23), we have

$$\pi_n(\mathbb{J}) \leq \sup_{x\in\mathbb{J}}|\widehat{f}_{n,h}(x) - f(x)|^{\alpha}\int_{\mathbb{J}} g^{1-\alpha}(x)dx.$$

Hence,

$$\begin{aligned}
\eta_n(\mathbb{J}) &\leq \tfrac{1}{1-\alpha}\tfrac{1}{\gamma^{\alpha}}\tfrac{1}{\int_{\mathbb{J}} g^{1-\alpha}(x)dx}\sup_{x\in\mathbb{J}}|\widehat{f}_n(x) - f(x)|^{\alpha}\int_{\mathbb{J}} g^{1-\alpha}(x)dx\\
&\leq \tfrac{1}{1-\alpha}\tfrac{1}{\gamma^{\alpha}}\sup_{x\in\mathbb{J}}|\widehat{f}_n(x) - f(x)|^{\alpha}
\end{aligned}$$

One fined, by combining the last equation with (22)

$$\limsup_{n\to\infty}\sup_{h'_n\leq h\leq h''_n}\frac{\sqrt{(nh)^{\alpha}}\eta_n(\mathbb{J})}{\sqrt{(\log(\frac{1}{h})\vee\log\log n)^{\alpha}}} \leq \frac{1}{1-\alpha}\frac{1}{\gamma^{\alpha}}\sup_{x\in\mathbb{J}}\left\{\left(f(x)\int_{\mathbb{R}^d}K^2(t)dt\right)^{\alpha/2}\right\}$$

$$\limsup_{ell\to\infty}\limsup_{n\to\infty}\sup_{h'_n\leq h\leq h''_n}\frac{\sqrt{(nh)^{\alpha}}\eta_n(\mathbb{J}_{\ell})}{\sqrt{(\log(1/h)\vee\log\log n)^{\alpha}}}$$

$$\begin{aligned}
&\leq \tfrac{1}{1-\alpha}\tfrac{1}{\gamma^{\alpha}}\limsup_{l\to\infty}\sup_{x\in\mathbb{J}_{\ell}}\left\{f(x)\int_{\mathbb{R}^d}K^2(t)dt\}^{\alpha/2}\right\}\\
&\leq \tfrac{1}{1-\alpha}\tfrac{1}{\gamma^{\alpha}}\sup_{x\in\mathbb{I}}\left\{\left(f(x)\int_{\mathbb{R}^d}K^2(t)dt\right)^{\alpha/2}\right\}
\end{aligned}$$

The proof of Corollary is completed.

[1] Akaike, H. (1954). An approximation to the density function.*Ann. Inst. Statist. Math., Tokyo,* 6, 127-132.

[2] Bosq, D. and Lecoutre, J. P. (1987). Théorie de l'estimation fonctionnelle. *Économie et Statistiques Avancées. Economica, Paris.*

[3] Bouzebda, S. and Elhattab, I.(2011) Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy. *Electronic Journal of Statistics.* 5, 440-459.

[4] Clarkson, J. A. and Adams, C. R. (1933). On definitions of bounded variation for functions of two variables. *Trans. Amer. Math. Soc.*, 35 (4), 824-854.

[5] Csiszár, I. (1967). Information-type measures of differences of probability distributions and indirect observations. *Studia Sci. Math. Hungarica,* 2: 299-318.

[6] Deheuvels, P. (2000). Uniform limit laws for kernel density estimators on possibly unbounded intervals. *In Recent advances in reliability theory (Bordeaux, 2000), Stat. Ind. Technol., pages 477-492. BirkhaBoston*,

[7] Deheuvels, P. and Einmahl, J. (1996). On the strong limiting behavior of local functionals of empirical processes based upon censored data. *Ann. Prob*. 24, 504-525.

[8] Deheuvels, P. and Einmahl, J. (2000). Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications. *Ann. Prob.*28 (7), 1301-1335.
[9] Deheuvels, P. and Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.*, 7(3), 225-277.
[10] Deroye, L. and Gyorfi, L. (1985). Nonparametric density estimation. *Wiley Series in Probability and Mathematical Statistics*: Tracts on Probability and Statistics. John Wiley & Sons Inc., New York. The L1 view.
[11] Devroye, L. and Lugosi, G. (2001). Combinatorial methods in density estimation. *Springer Series in Statistics. Springer-Verlag, New York.*
[12] Devroye, L. and Wise, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.*, 38(3),480-488.
[13] Diehl, S. and Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring. *J. Mult. Analy.*. 25, 299-310.
[14] Dmitriev, J. G. and Tarasenko, F. P. (1973). The estimation of functionals of a probability density and its derivatives. *Teor. Verojatnost. i Primenen.*, 18, 662-668.
[15] Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.*, 13 (1), 1-37.
[16] Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, 33(3), 1380-1403.
[17] Foldes A.and Rejto L., (1981). Strong Uniform Consistency for Nonparametric Survival Curve Estimators from Randomly Censored Data. *Annals of Statistics.* Volume 9, 122-129.
[18] Giné, E.and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.* 38 907-921.
[19] Giné, E. and Mason, D. M. (2008) Uniform in bandwidth estimation of integral functionals of the density function. *Scand. J. Statist.*, 35(4), 739-761
[20] Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes (with discussion).*Ann. Probab.* 12 929-998.
[21] Hobson, E. W. (1958). The theory of functions of a real variable and the theory of Fourier's series. Vol. I. *Dover Publications Inc.*, New York, N.Y.
[22] Johnson , D. H., Gruner, B., C. M. K., and Seshagiri.(2001) Information-theoretic analysis of neural coding. *Journal of Computational Neuroscience*.
[23] Krishnamurthy A., Kandasamy K. (2014). Nonparametric Estimation of Rényi Divergence and Friends *http://www.arxiv.org/1402.2966v2*.
[24] Nolan, D. and Pollard, D.(1987): U-processes: rates of convergence. Ann. Statist., 15(2):780799.
[25] Pakes, A. and Pollard, D.(1989): Simulation and the asymptotics of optimization estimators. Econometrica, 57(5):10271057, 1989.
[26] Parzen, E. (1962). On estimation of a probability density function and mode.*Ann. Math. Statist.*, 33, 1065-1076.
[27] Pardo, L.(2005) Statistical inference based on divergence measures. *CRC Press*.
[28] Pluim B M, Safran M. From breakpoint to advantage. *description, treatment, and prevention of all tennis injuries.* Vista: USRSA, 2004
[29] Prakasa Rao, B. L. S. (1983). Nonparametric functional estimation. *Probability and Mathematical Statistics. Academic Press Inc.* [Harcourt Brace Jovanovich Publishers], New York.
[30] Póczos, B. and Schneider, J. On the estimation of alpha-divergences. *CMU, Auton Lab Technical Report,* `http://www.cs.cmu.edu/bapoczos/ articles/poczos11alphaTR.pdf`.
[31] Póczos, B. Xiong L., Sutherland D, J., and Schneider J. (2012). Nonparametric kernel estimators for image classification. *In IEEE Conference on Computer Vision and Pattern Recognition,*
[32] Rényi, A. (1961). On measures of entropy and information. *In Fourth Berkeley Symposium on Mathematical Statistics and Probability.*
[33] Rényi, A. (1970). Probability Theory. *Publishing Company, Amsterdam.*
[34] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27, 832-837.
[35] Tanner, M. A. and Wong, W.H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.* 11, 989-993.
[36] Van der Vaart, A. W. and Wellner, J. A. (1996). Weak Convergence and Empirical Processes: *With Applications to Statistics. Springer, New York*.
[37] Viallon, V. (2006). Processus empiriques, estimation non parametrique et données censurées. Ph.D. thesis, Université Paris 6.
[38] Villmann, T. and Haase, S. (2010). Mathematical aspects of divergence based vector quantization using Frechet-derivatives. *University of Applied SciencesMittweida.*
[39] Vituškin, A. G. (1955). O mnogomernyh variaciyah. Gosudarstv. *Izdat. Tehn.Teor. Lit., Moscow.*
[40] Watson, G.S. and eadbetter, M.R. (1964a). Hazard Analysis I. *Biometrika,* Vol. 51, 1 and 2, pp. 175-184.
[41] Watson, G.S. and Leadbetter,M.R. (1964b). *Hazard Analysis II. Sankhya: The Indian Journal of Statistics, Series A,* Vol. 26, No. 1, pp. 101-116.