

# Kernel-Type Estimators of Divergence Measures and Its Strong Uniform Consistency

Hamza Dhaker<sup>1,\*</sup>, Papa Ngom<sup>1</sup>, El Hadji Deme<sup>2</sup>, Pierre Mendy<sup>3</sup>

<sup>1</sup>Departement de Mathématiques et Informatique, Faculté des Sciences et Technique, Université Cheikh Anta Diop, Dakar, Sénégal

<sup>2</sup>Sciences Appliquées et Technologie, Unité de Formation et de Recherche, Université Gaston Berger, Saint-Louis, Sénégal

<sup>3</sup>Département de Techniques Quantitatives, Faculté des Sciences Economiques et de Gestion, Université Cheikh Anta Diop, Dakar, Sénégal

## Email address:

hamzaould.yahya@ucad.edu.sn (H. Dhaker), papa.ngom@ucad.edu.sn (P. Ngom), eltinoyaw@yahoo.fr (E. H. Deme),

pierre.mendy@ucad.edu.sn (P. Mendy)

## To cite this article:

Hamza Dhaker, Papa Ngom, El Hadji Deme, Pierre Mendy. Kernel-Type Estimators of Divergence Measures and Its Strong Uniform Consistency. *American Journal of Theoretical and Applied Statistics*. Vol. 5, No. 1, 2016, pp. 13-22. doi: 10.11648/j.ajtas.20160501.13

**Abstract:** Nonparametric density estimation, based on kernel-type estimators, is a very popular method in statistical research, especially when we want to model the probabilistic or stochastic structure of a data set. In this paper, we investigate the asymptotic confidence bands for the distribution with kernel-estimators for some types of divergence measures (Rényi- $\alpha$  and Tsallis- $\alpha$  divergence). Our aim is to use the method based on empirical process techniques, in order to derive some asymptotic results. Under different assumptions, we establish a variety of fundamental and theoretical properties, such as the strong consistency of an uniform-in-bandwidth of the divergence estimators. We further apply the previous results in simulated examples, including the kernel-type estimator for Hellinger, Bhattacharyya and Kullback-Leibler divergence, to illustrate this approach, and we show that that the method performs competitively.

**Keywords:** Divergence Measures, Kernel Estimation, Strong Uniform, Consistency

## 1. Introduction

In this paper, we focus on the similarity between two distributions. Given a sample from one distribution, one of fundamental and classical question to ask is: how to have the similarity between this density with another known density? 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 [21], classification and retrieval. There are several important problems in machine learning and statistics that require the estimation of the distance or divergence between distributions. Divergence between distributions also proves to be useful in neuroscience, For example (see, e.g [14]). employs divergence to quantify the difference between neural response patterns.

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 [23, 24]. The key role of the measure divergence in these various applications, it is necessary to accurately estimate these divergences.

Recently Ngom et al [16] has introduced the Divergence Indicator  $DI$  method by proposing a test for choosing between a random walk and AR(1), using a divergence measure.

The class of divergence measures is large; it includes the Rényi- $\alpha$  [25, 26], Tsallis- $\alpha$  [30], Kullback-Leibler (KL), Hellinger, Bhattacharyya, Euclidean divergences, etc. These divergence measures can be related to the Csiszár- $f$  divergence [3]. The Kullback-Leibler, Hellinger and Bhattacharyya are special cases of Rényi- $\alpha$  and Tsallis- $\alpha$  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. For example Pardo [20] presented methods and applications in the case of discrete distributions. By exploring a nonparametric method for estimating the divergence in the continuous case, Poczos and Schneider [23] proposed a  $k$ -nearest-neighbor estimator and proved the weak

consistency of the estimator Rényi- $\alpha$  and Tsallis- $\alpha$  divergences.

Finding estimators nonparametric of measure divergence, remains an open issue. Krishnamurthy and Kandasamy [15] 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 three estimators for Rényi- $\alpha$ , Tsallis- $\alpha$ , and Euclidean divergences between two continuous distributions and established the rates of convergence of these estimators. The main purpose of this paper is to analyze estimators for divergence measures between two continuous distributions. Our approach is similar to that of Krishnamurthy and Kandasamy [15] and is based on plug-in estimation scheme: first, we apply a consistent density estimator for the underlying densities, and then we plug them into the desired formulas. Unlike of their frameworks, we study the strong consistency estimators for a general class of divergence measures. We emphasize that the plug-in estimation techniques are heavily used by [2, 9] in the case of entropy. Bouzebda [2] proposed 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 proposed estimators. Section 4 is devoted to the proofs. In Section 5, numical examples are proposed in order to illustrate the performance of our method. Finally, in Section 6, we present our conclusion.

## 2. Kernel-Type Estimators of Divergence Measures

In this section we give the notations and then presenting some basic definitions. We are interested 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- $\alpha$ , Tsallis- $\alpha$  are defined respectively as follows

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

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

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

$$D_{1/2}^T(f, g) = 2D^H(f, g) = 2 \left( 1 - \int_{\mathbb{R}^d} f^{1/2}(x) g^{1/2}(x) dx \right),$$

$$D_{1/2}^R(f, g) = 2D^B(f, g) = -\log \int_{\mathbb{R}^d} f^{1/2}(x) g^{1/2}(x) dx,$$

$$\lim_{\alpha \rightarrow 1} D_\alpha^R(f, g) = D^{KL}(f, g) = \int_{\mathbb{R}^d} f(x) \log \frac{f(x)}{g(x)} dx,$$

which is related to the Shannon entropy. For some statistical properties for the Shannon entropy, one can refer to [2].

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

via

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

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  $D_\alpha^T(f, g)$  and  $D_\alpha^R(f, g)$ . The Kullback-Leibler, Hellinger and Bhattacharyya can be deducted immediately.

We will next provide consistent estimator for the following quantity

$$D_\alpha(f, g) = \int_{\mathbb{R}^d} f^\alpha(x) g^{1-\alpha}(x) dx, \quad (3)$$

whenever this integral is meaningful. Plugging its estimates into the appropriate formula immediately leads to a consistent estimator for the divergence measures  $D_\alpha^R(f, g)$  and  $D_\alpha^T(f, g)$ .

Now, assuming that for the rest of the document, the density  $f$  is unknown, and the density  $g$  is known and satisfies:  $\int_{\mathbb{R}^d} g^{1-\alpha}(x) dx$  is finite, this implies that  $D_\alpha(f, g)$  is finite. Next, consider  $X_1, \dots, X_n, n \geq 1$  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 the Lebesgue measure on  $\mathbb{R}^d$ . The following conditions are needed for the remainder of this paper. 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 following 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)  $\|K\|_\infty = \sup_{x \in \mathbb{R}^d} |K(x)| < \infty$

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

Rosenblatt [27] first proposed an estimator for  $f(\cdot)$  and Parzen [19] generalized, it eventually leading to the Parzen-Rosenblatt estimator, defined in the following way for any  $x \in \mathbb{R}^d$

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

where  $0 < h_n < 1$  is the bandwidth sequence. Assuming that the density  $f$  is continuous, one obtains a strongly consistent estimator  $f_{n,h_n}(x)$  of  $f$ , that is, one has with probability 1,

$f_{n,h_n}(x) \rightarrow f(x)$ ,  $x \in \mathbb{R}^d$ . There are also results concerning uniform convergence and convergence rates. For proving such results one usually writes the difference  $f_{n,h_n}(x) - f(x)$  as the sum of a probabilistic term  $f_{n,h_n}(x) - \mathbb{E}f_{n,h_n}(x)$  and a deterministic term  $\mathbb{E}f_{n,h_n}(x) - f(x)$ , also called bias. For further explanation One can refer to [10, 12, 13], among other authors. After having estimated  $f_{n,h_n}(\cdot)$ , we estimate  $D_\alpha(f, g)$  by setting

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

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 constants. Thus, using (5), the associated divergences  $D_\alpha^R(f, g)$  and  $D_\alpha^T(f, g)$  can be estimated by:

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

The approach used to define the plug-in estimators is also developed in [2] in order to introduce a kernel-type estimator of Shannon's entropy. Some statistical properties of these divergences is related to those of the kernel estimator  $f_{n,h_n}(\cdot)$  of the continuous density  $f$ . The limiting behavior of  $f_{n,h_n}(\cdot)$ , for appropriate choices of the bandwidth  $h_n$ , has been widely studied in the literature, examples include the work of Deroye [6, 7] Bosq [1] and Prakasa [22]. 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  $f_{n,h_n}(x)$  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  $f_{n,h_n}(x)$  in [4, 10, 5] and the references therein. In the next section, we will use the methods developed in previous references to establish convergence results for the estimates  $\widehat{D}_\alpha(f_{n,h_n}, g)$  and deduce the convergence results of  $\widehat{D}_\alpha^R(f_{n,h_n}, g)$  and  $\widehat{D}_\alpha^T(f_{n,h_n}, g)$ .

### 3. Statistical Properties of the Estimators

We first study the strong consistency of the estimator  $\widehat{D}_\alpha(f_{n,h_n}, g)$  defined in (5). Throughout the remainder of this paper, we will use the notation for  $\mathbb{E}\widehat{D}_\alpha(f_{n,h_n}, g)$ , which is delicate to handle. This is given by

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

*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  $(h'_n)_{n \geq 1}$ ,  $(h''_n)_{n \geq 1}$  such that  $0 < h'_n < h_n \leq h''_n$ , together with  $h''_n \rightarrow 0$ ,  $nh'_n/\log(n) \rightarrow \infty$  as  $n \rightarrow \infty$  and for any  $\alpha \in (0,1)$ , one has with probability 1

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

The proof of Lemma 1 is postponed until Section 4.

*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  $(h'_n)_{n \geq 1}$ ,  $(h''_n)_{n \geq 1}$  such that  $0 < h'_n < h_n \leq h''_n$ , together with  $h''_n \rightarrow 0$ , as  $n \rightarrow \infty$  and for any  $\alpha \in (0,1)$ , we have

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

The proof of Lemma 2 is postponed until Section 4.

*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  $(h'_n)_{n \geq 1}$ ,  $(h''_n)_{n \geq 1}$  such that  $0 < h'_n < h_n \leq h''_n$ , together with  $h''_n \rightarrow 0$ ,  $nh'_n/\log(n) \rightarrow \infty$  as  $n \rightarrow \infty$  and for any  $\alpha \in (0,1)$ , one has with probability 1

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

This, in turn, implies that

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} |\widehat{D}_\alpha(f_{n,h}, g) - D_\alpha(f, g)| = 0 \text{ a.s.} \quad (6)$$

The proof of Theorem 1 is postponed until Section 4.

The following corollaries handle respectively the uniform deviation of the estimate  $\widehat{D}_\alpha^T(f_{n,h}, g)$  and  $\widehat{D}_\alpha^R(f_{n,h}, g)$  with respect to  $D_\alpha^T(f, g)$  and  $D_\alpha^R(f, g)$ .

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

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

This, in turn, implies that

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} |\widehat{D}_\alpha^T(f_{n,h}, g) - D_\alpha^T(f, g)| = 0 \text{ a.s.} \quad (7)$$

The proof of Corollary 1 is postponed until Section 4.

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

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

This, in turn, implies that

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} |\widehat{D}_\alpha^R(f_{n,h}, g) - D_\alpha^R(f, g)| = 0 \text{ a.s.} \quad (8)$$

The proof of Corollary 2 is postponed until Section 4.

Note that, the divergence estimator such as (5) also requires the appropriate choice of the smoothing parameter  $h_n$ . The results 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 estimators. In other words, the fluctuation of the bandwidth in a small interval do not affect the consistency of the nonparametric estimators of these divergences. The work of Bouzebda and Elhattab [2] is very important for establishing our results, these authors have created a class of compactly supported densities. They used 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

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

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

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 4.

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

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{(nh)^\alpha} |\widehat{D}_\alpha^T(f_{n,h}, g) - D_\alpha^T(f, g)|}{\sqrt{(\log(1/h) \vee \log \log n)^\alpha}} \\ \leq \frac{1}{1-\alpha} \zeta(\mathbb{I}) \int_{\mathbb{R}^d} g^{1-\alpha}(x) dx \text{ a.s.}, \end{aligned}$$

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

$$\limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{(nh)^\alpha} |\widehat{D}_\alpha^R(f_{n,h}, g) - D_\alpha^R(f, g)|}{\sqrt{(\log(1/h) \vee \log \log n)^\alpha}}$$

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

$$\int_{\mathbb{R}^d} u_1^{j_1} \dots u_d^{j_d} K(u) du = 0,$$

$$j_1, \dots, j_d \geq 0, j_1 + \dots + j_d = 1, \dots, s-1,$$

and

$$\int_{\mathbb{R}^d} |u_1^{j_1} \dots u_d^{j_d}| K(u) du = \rho, j_1, \dots, j_d \geq 0,$$

$$j_1 + \dots + j_d = s.$$

Under (F.1) the expression  $D_\alpha(f, g)$  may be written as follows

$$D_\alpha(f, g) = \int_{\mathbb{I}} f^\alpha(x) g^{1-\alpha} dx. \quad (9)$$

*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  $h''_n \rightarrow 0$ ,  $nh'_n/\log n \rightarrow \infty$  as  $n \rightarrow \infty$  and for any  $\alpha \in (0,1)$ , we have

$$\leq \frac{1}{(1-\alpha)\gamma^\alpha} \zeta(\mathbb{I}) \text{ a.s.},$$

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

Now, assume that there exists a sequence  $\{\mathbb{I}_n\}_{n \geq 1}$  of strictly nondecreasing compact subsets of  $\mathbb{I}$ , such that  $\mathbb{I} = \bigcup_{n \geq 1} \mathbb{I}_n$ . For the estimation of the support  $\mathbb{I}$  we may refer to ([8]) 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 choose an estimator of  $\zeta(\mathbb{I})$  in the Corollaries (3) and (4) as the form

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

Using the techniques developed in [5] and the Corollaries (3) and (4) one can construct an asymptotically 100% certainty intervals for the true divergences  $D_\alpha^T(f, g)$  and  $D_\alpha^R(f, g)$ .

## 4. Proofs of Our Results

*Proof of Lemma 1.* to show the strong consistency of

$\widehat{D}_\alpha(f_{n,h_n}, g)$ , we use the following expression

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

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

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

We have

$$|f_{n,h_n}^\alpha(x) - (\mathbb{E}f_{n,h_n}(x))^\alpha| \leq |f_{n,h_n}(x) - \mathbb{E}f_{n,h_n}(x)|^\alpha \leq \|f_{n,h_n} - \mathbb{E}f_{n,h_n}\|_\infty^\alpha,$$

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

$$|\Delta_{n,1,h_n}| \leq \|f_{n,h_n} - \mathbb{E}f_{n,h_n}\|_\infty^\alpha \int_{A_{n,h_n}} g^{1-\alpha}(x) dx. \quad (10)$$

Finally,

$$|\Delta_{n,1,h_n}| \leq \|f_{n,h_n} - \mathbb{E}f_{n,h_n}\|_\infty^\alpha \int_{\mathbb{R}^d} g^{1-\alpha}(x) dx. \quad (11)$$

Using the conditions on the kernel  $K(\cdot)$ , posed by Einmahl [11], consider the class of functions

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

For  $\varepsilon > 0$ , set  $N(\varepsilon, K) = \sup_Q N(K\varepsilon, K, d_Q)$ , where the supremum is taken over all probability measures  $Q$  on  $(\mathbb{R}^d, \mathcal{B})$ , where  $\mathcal{B}$  represents the  $\sigma$ -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  $N(K\varepsilon, K, d_Q)$  is the minimal number of balls  $\{\psi : d_Q(\psi, \psi') < \varepsilon\}$  of  $d_Q$ -radius  $\varepsilon$  needed to cover  $K$ .

We assume that  $K$  satisfies the following uniform entropy condition.

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

(K.7)  $K$  is a pointwise measurable class, that is there exists a countable sub-class  $K_0$  of  $K$  such that we can find for any function  $\psi \in K$  a sequence of functions  $\{\psi_m : m \geq 1\}$  in  $K_0$  for which

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

This condition is discussed in [27]. 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 [28], for details on conditions of entropy (see also Pakes and Pollard [18], and Nolan and Pollard [17]). Condition (K.7) is satisfied whenever (K.2) holds, i.e.,  $K(\cdot)$  is right continuous, this condition is discussed in [28], (see also [5] and [11]).*

From Theorem 1 in [11], whenever  $K(\cdot)$  is measurable and

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

Since  $s(x) = x$  is a 1-Lipschitz function, for  $0 < \alpha < 1$  then

$$|(s(x))^\alpha - (s(y))^\alpha| \leq |s(x) - s(y)|^\alpha.$$

Therefore for  $0 < \alpha < 1$ , we have

satisfies (K.3-4-6-7), and when  $f(\cdot)$  is bounded, we have for each pair of sequence  $(h'_n)_{n \geq 1}$ ,  $(h''_n)_{n \geq 1}$  such that  $0 < h'_n < h \leq h''_n \leq 1$ , together with  $h''_n \rightarrow 0$  and  $nh'_n/\log(n) \rightarrow \infty$  as  $n \rightarrow \infty$ , with probability 1

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

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

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

It concludes the proof of the lemma.

*Proof of Lemma 2.*

Let  $A_{n,h_n}^c$  be the complement of  $A_{n,h_n}$  in  $\mathbb{R}^d$  (i.e.,  $A_{n,h_n}^c = \{x \in \mathbb{R}^d, f_{n,h_n} < \gamma_n\}$ ). We have

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

with

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

and

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

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

$$|\Delta_{n,2,h_n}| \leq \|f_{n,h_n} - \mathbb{E}f_{n,h_n}\|_\infty^\alpha \int_{A_{n,h_n}} g^{1-\alpha}(x) dx, \quad (14)$$

which implies

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

On the other hand, we know (see, e.g., [11]), that since the

density  $f(\cdot)$  is uniformly Lipschitz and continuous, we have for each sequences  $h'_n < h < h''_n < 1$ , with  $h''_n \rightarrow 0$ , as  $n \rightarrow \infty$ ,

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

$$\begin{aligned} |\Delta_{n,3,h_n}| &= \int_{A_{n,h_n}^c} |f^\alpha(x)| g^{1-\alpha}(x) dx \\ &\leq \int_{A_{n,h_n}^c} |\mathbb{E}f_{n,h_n}(x) - f^\alpha(x)| g^{1-\alpha}(x) dx + \int_{A_{n,h_n}^c} \mathbb{E}f_{n,h_n}(x) g^{1-\alpha}(x) dx \\ &\leq \|\mathbb{E}f_{n,h_n} - f\|_\infty^\alpha \int_{A_{n,h_n}^c} g^{1-\alpha}(x) dx + \gamma_n^\alpha \int_{A_{n,h_n}^c} g^{1-\alpha}(x) dx \end{aligned}$$

Thus,

$$|\Delta_{n,3,h_n}| \leq (\|\mathbb{E}f_{n,h_n} - f\|_\infty^\alpha + \gamma_n^\alpha) \int_{A_{n,h_n}^c} g^{1-\alpha}(x) dx. \quad (18)$$

Hence,

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

Thus, in view of (16), we get

$$|\widehat{D}_\alpha^R(f_{n,h_n}, g) - D_\alpha^R(f, g)| \leq |\widehat{D}_\alpha(f_{n,h_n}, g) - \mathbb{E}\widehat{D}_\alpha(f_{n,h_n}, g)| + |\mathbb{E}\widehat{D}_\alpha(f_{n,h_n}, g) - D_\alpha(f, g)|.$$

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

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

It concludes the proof of the Theorem.

*Proof of Corollary 1.* Remark that

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

Using the Theorem (1), we have

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

and the Corollary 1 holds.

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

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

Remark that from Theorem 1,

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

which in turn, implies that

$$\lim_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} |\widehat{D}_\alpha(f_{n,h}, g) - D_\alpha(f, g)| = 0 \text{ a. s.}$$

Thus,

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

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

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

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

$$\sup_{h'_n \leq h \leq h''_n} |\mathbb{E}\widehat{D}_\alpha(f_{n,h}, g) - D_\alpha(f, g)| = O(\gamma_n^\alpha \vee h_n''^{\alpha/d}) \quad (21)$$

is deduced the proof of the lemma.

*Proof of Theorem 1.* We have

Thus, for all

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

Consequently

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

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  $x \in \mathbb{I}$ ,

$$\begin{aligned} |\mathbb{E}f_{n,h} - f(x)| \\ = \frac{h^{s/d}}{s!} \left| \int \sum_{k_1 + \dots + k_d} t_1^{k_1} \dots t_d^{k_d} \frac{\partial^s f(x - h\theta t)}{\partial x_1^{k_1} \dots \partial x_d^{k_d}} K(t) dt \right| \end{aligned}$$

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

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

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 [29] (see also, [2], statement (4.16)).

$$\limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \sup_{x \in \mathbb{J}} \frac{\sqrt{nh} |f_{n,h}(x) - f(x)|}{\sqrt{\log(1/h) \vee \log \log n}} = \sup_{x \in \mathbb{J}} (f(x) \int_{\mathbb{R}^d} K^2(t) dt)^{1/2} \quad (22)$$

Set, for all  $n \geq 1$ ,

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

by combining (22) and (24)

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

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

$$\bigcup_{\ell \geq 1} \mathbb{J}_\ell = \mathbb{I}$$

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

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

The proof of Theorem 2 is completed.

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

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

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

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

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

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

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

Hence,

$$\eta_n(\mathbb{J}) \leq \frac{1}{1-\alpha} \frac{1}{\gamma^\alpha} \frac{1}{\int_{\mathbb{J}} g^{1-\alpha}(x) dx} \sup_{x \in \mathbb{J}} |f_{n,h}(x) - f(x)|^\alpha \int_{\mathbb{J}} g^{1-\alpha}(x) dx \leq \frac{1}{1-\alpha} \frac{1}{\gamma^\alpha} \sup_{x \in \mathbb{J}} |f_{n,h}(x) - f(x)|^\alpha$$

by combining the last equation with (22)

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

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

$$\leq \frac{1}{1-\alpha} \frac{1}{\gamma^\alpha} \limsup_{l \rightarrow \infty} \sup_{x \in \mathbb{J}_\ell} \left\{ f(x) \int_{\mathbb{R}^d} K^2(t) dt \right\}^{\alpha/2}$$

$$\leq \frac{1}{1-\alpha} \frac{1}{\gamma^\alpha} \sup_{x \in \mathbb{I}} \left\{ \left( f(x) \int_{\mathbb{R}^d} K^2(t) dt \right)^{\alpha/2} \right\}$$

The proof of Corollary is completed.

## 5. Simulation Study

Summarizing the ideas and the results given in the previous sections, we propose to study the performance of the kernel-estimators for Hellinger ( $\alpha = \frac{1}{2}$  in Tsallis -  $\alpha$ ), Bhattacharyya ( $\alpha = \frac{1}{2}$  in Rényi -  $\alpha$ ) and Kullback-Leibler ( $\alpha \rightarrow 1$  in Rényi -  $\alpha$ ) measures and their

uniform-in-bandwidth consistency.

Hellinger, Bhattacharyya and Kullback-Leibler divergences are defined respectively as follows:

$$D^H(f, g) = 1 - \int_{\mathbb{R}^d} f^{1/2}(x) g^{1/2}(x) dx,$$

$$D^B(f, g) = -\frac{1}{2} \log \int_{\mathbb{R}^d} f^{\frac{1}{2}}(x) g^{\frac{1}{2}}(x) dx,$$

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

The asymptotic behavior of each bandwidth is performed using the kernel-type estimator of the divergence criteria in corollary 3 and corollary 4 respectively.

We compute, for each chosen value of  $\alpha$ , the expressions

$$|DH| = \limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{(nh)^5} |\hat{D}^H(f_{n,h}, g) - D^H(f, g)|}{\sqrt{(\log(1/h) \vee \log \log n)^5}} \leq 2\zeta(\mathbb{I}) \int_{\mathbb{R}^d} g^{.5}(x) dx$$

$$|DB| = \limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{(nh)^5} |\hat{D}^B(f_{n,h}, g) - D^B(f, g)|}{\sqrt{(\log(1/h) \vee \log \log n)^5}} \leq \frac{2}{\sqrt{\log(n)}} \zeta(\mathbb{I})$$

$$|DK| = \limsup_{n \rightarrow \infty} \sup_{h'_n \leq h \leq h''_n} \frac{\sqrt{nh} |\hat{D}^K(f_{n,h}, g) - D^K(f, g)|}{\sqrt{\log(1/h) \vee \log \log n}} \leq \frac{2}{\log(n)} \zeta(\mathbb{I})$$

where the coresponding bounds ( $BH, BB_n, BK_n$ ) are defined by

$$BH = 2\zeta(\mathbb{I}) \int_{\mathbb{R}^d} g^{.5}(x) dx$$

$$BB_n = \frac{2}{\sqrt{\log(n)}} \zeta(\mathbb{I})$$

$$BK_n = \frac{2}{\log(n)} \zeta(\mathbb{I})$$

We consider an experiment in which the DGP (Data Generating Process) for the true distribution  $f$  are generated from a mixture of two normal distributions,

$$f(x) = 0.5 \times N(0,1) + 0.5 \times N(0,2)$$

and the  $g$  function is supposed to be a normal distribution with mean 1 and variance 2.

The sample size varies from 10 to 1000, and for each size, the statistics  $DH, DB, DK$  and  $BH, BB, BK$  are evaluated.

In order to plot  $DH, DB$  and  $DK$  against sample size, we need to perform three sets experiments.



The results are presented in tables 1-3 and figures 1-3.

**Table 1.**  $|DH|$  and  $BH$  against the sample  $n$ .

| Size $n$ | $ DH $ | $BH$   |
|----------|--------|--------|
| 10       | 0.14   | 0.0627 |
| 20       | 0.25   | 0.0627 |
| 50       | 0.07   | 0.0627 |
| 100      | 0.02   | 0.0627 |
| 300      | 0.01   | 0.0628 |
| 500      | 0.006  | 0.0629 |
| 1000     | 0.003  | 0.0629 |

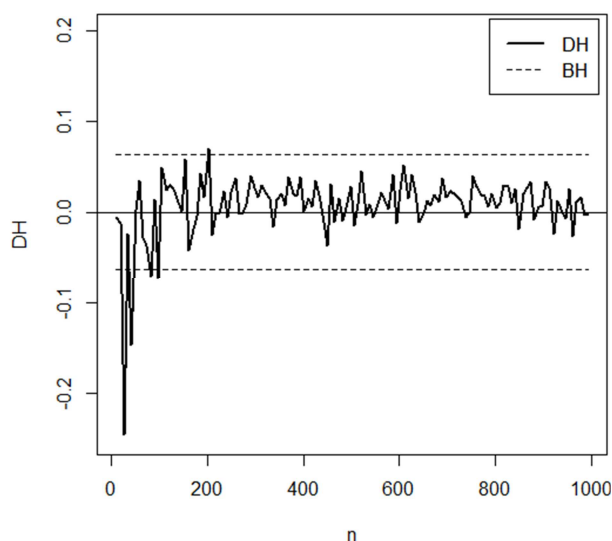
**Table 2.**  $|DB|$  and  $BB_n$  against the sample  $n$ .

| Size $n$ | $ DB $ | $BB_n$ |
|----------|--------|--------|
| 10       | 0.048  | 0.040  |
| 20       | 0.037  | 0.030  |
| 50       | 0.007  | 0.027  |
| 100      | 0.007  | 0.025  |
| 300      | 0.005  | 0.024  |
| 500      | 0.003  | 0.024  |
| 1000     | 0.002  | 0.023  |

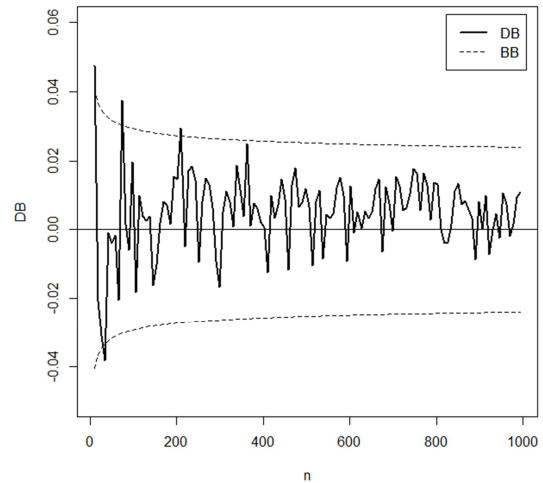
**Table 3.**  $|DK|$  and  $BK_n$  against the sample  $n$ .

| Size $n$ | $ DK $ | $BK_n$ |
|----------|--------|--------|
| 10       | 0.15   | 0.19   |
| 20       | 0.128  | 0.167  |
| 50       | 0.213  | 0.11   |
| 100      | 0.053  | 0.095  |
| 300      | 0.034  | 0.093  |
| 500      | 0.025  | 0.842  |
| 1000     | 0.007  | 0.825  |

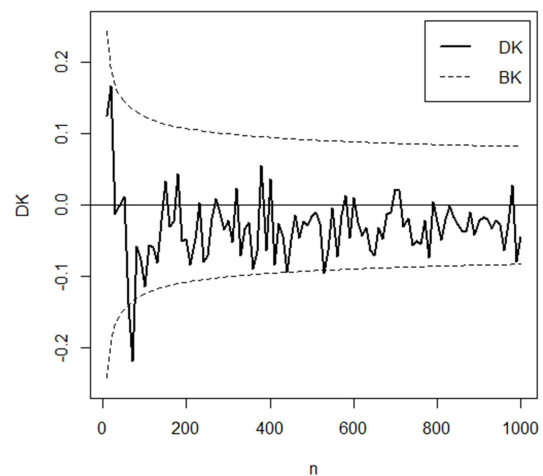
The tables 1-3 show that the kernel-type estimators of the divergence measures converge rapidly to their pseudo-true value, and confirm our asymptotic results. They all show that the discrepancy between the estimated and the true divergence criterion converge rapidly to zero. Similarly, in table 2 and table 3,  $DB$  and  $DK$  converge, as expected, to zero, which is the mean of the asymptotic distribution when the estimated distribution  $f_{n,h}$  is close to  $f$ .



**Fig. 1.**  $DH$ ,  $BH$  and  $-BH$  as a function of  $n$ .



**Fig. 2.**  $DB$ ,  $BB$  and  $-BB$  as a function of  $n$ .



**Fig. 3.**  $DK$ ,  $BK$  and  $-BK$  as a function of  $n$ .

The figures 1-3 show  $DH$ ,  $DB$  and  $DK$  values plots for Hellinger, Bhattacharyya and Kullback-Leibler divergence respectively. The preceding comments from the table 1-3 also apply to the figure 1-3. For dealing with divergence error, it is much revealing to graph  $DH$ ,  $DB$  and  $DK$  against sample size. They also confirm our asymptotic results. We note that, as sample size increases, the value discrepancy plots of divergence error converge, as it should, to zero. These plots provide a great deal of information about how the sample size  $n$  affect the performance of these informational criterions.

## 6. 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 Bouzebda [2]. We introduce an estimator that can be seen as a generalization of those previously suggested, in the sense that Bouzebda was only interested in the case of entropy, while we focus on the Rényi- $\alpha$  and the Tsallis- $\alpha$  divergence measures. Under our study, one can easily deduce Kullback-Leibler, Hellinger and

Bhattacharyya nonparametric 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  $\widehat{D}_\alpha(\hat{f}_{n,h_n}, g)$  in (5) can be calculated by using a Monte-Carlo method under a given distribution  $g$ . And a practical choice of  $\gamma_n$  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  $\gamma_n$  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 are all inherently univariate. A natural and useful multivariate extension appear in the use of copula function.

## References

- [1] Bosq, D. and Lecoutre, J. P. (1987). Théorie de l'estimation fonctionnelle. *Économie et Statistiques Avancées. Economica, Paris*.
- [2] 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.
- [3] Csiszár, I. (1967). Information-type measures of differences of probability distributions and indirect observations. *Studia Sci. Math. Hungarica*, 2: 299-318.
- [4] 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.
- [5] 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.
- [6] Deroye, L. and Györfi, 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.
- [7] Devroye, L. and Lugosi, G. (2001). Combinatorial methods in density estimation. *Springer Series in Statistics*. Springer-Verlag, New York.
- [8] 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.
- [9] Dmitriev, J. G. and Tarasenko, F. P. (1973). The estimation of functionals of a probability density and its derivatives. *Teor. Veroyatnost. i Primenen.*, 18, 662-668.
- [10] 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.
- [11] Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, 33(3), 1380-1403.
- [12] 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.
- [13] Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes (with discussion). *Ann. Probab.* 12 929-998.
- [14] Johnson, D. H., Gruner, B., C. M. K., and Seshagiri.(2001) Information-theoretic analysis of neural coding. *Journal of Computational Neuroscience*.
- [15] Krishnamurthy A., Kandasamy K., Póczos B., and Wasserman L., (2014). Nonparametric Estimation of Rényi Divergence and Friends. <http://www.arxiv.org/1402.2966v2>.
- [16] Ngom, P., Dhaker, H., Mendy, P., Deme, E. Generalized divergence criteria for model selection between random walk and AR(1) model. <https://hal.archives-ouvertes.fr/hal-01207476v1>
- [17] Nolan, D. and Pollard, D. (1987): U-processes: rates of convergence. *Ann. Statist.*, 15(2): 780-799.
- [18] Pakes, A. and Pollard, D.(1989): Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5): 1027-1057, 1989.
- [19] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, 1065-1076.
- [20] Pardo, L.(2005) Statistical inference based on divergence measures. *CRC Press*.
- [21] Pluim B M, Safran M. From breakpoint to advantage. *description, treatment, and prevention of all tennis injuries*. Vista: USRSA, 2004.
- [22] Prakasa Rao, B. L. S. (1983). Nonparametric functional estimation. *Probability and Mathematical Statistics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York.
- [23] Póczos, B. and Schneider, J. On the estimation of alpha-divergences. *CMU, Auton Lab Technical Report*,
- [24] <http://www.cs.cmu.edu/bapoczcos/articles/poczcos11alphaTR.pdf>.
- [25] 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*.
- [26] Rényi, A. (1961). On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*.
- [27] Rényi, A. (1970). Probability Theory. *Publishing Company, Amsterdam*.
- [28] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27, 832-837.
- [29] Van der Vaart, A. W. and Wellner, J. A. (1996). Weak Convergence and Empirical Processes: *With Applications to Statistics*. Springer, New York.
- [30] Viallon, V. (2006). Processus empiriques, estimation non paramétrique et données censurées. Ph. D. thesis, Université Paris 6.
- [31] Villmann, T. and Haase, S. (2010). Mathematical aspects of divergence based vector quantization using Frechet-derivatives. *University of Applied SciencesMittweida*.