EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 12, No. 3, 2019, 790-820
ISSN 1307-5543 – www.ejpam.com
Published by New York Business Global

# Divergence Measures Estimation and Its Asymptotic Normality Theory in the discrete case

Amadou Diadie Ba[1,*], Gane Samb Lo[1,2,3]

[1] *U.F.R. S.A.T., LERSTAD, Gaston Berger University, Saint Louis, Sénégal*
[2] *LASTA, Pierre et Marie University, Paris, France*
[3] *African University of Sciences and Technology, Abuja, Nigeria*

**Abstract.** In this paper we provide the asymptotic theory of the general of $\phi$-divergences measures, which include the most common divergence measures : Rényi and Tsallis families and the Kullback-Leibler measure. We are interested in divergence measures in the discrete case. One sided and two-sided statistical tests are derived as well as symmetrized estimators. Almost sure rates of convergence and asymptotic normality theorem are obtained in the general case, and next particularized for the Rényi and Tsallis families and for the Kullback-Leibler measure as well. Our theoretical results are validated by simulations.

**2010 Mathematics Subject Classifications**: 62G05, 62G07, 62F05
**Key Words and Phrases**: $\phi$-divergence measure estimation, Rényi, Tsallis, Kullback-Leibler divergence measures, symmetrized divergence measures

## 1. Introduction

### 1.1. Motivations

In this paper, we study the convergence of $\phi-$divergence measure estimator for empirical discrete probability distributions supported on a finite set.

Let throughout the following $\mathcal{X} = \{c_1, c_2, \cdots, c_r\}$ $(r \geq 2)$ be a finite countable space. The probability distributions on $\mathcal{X}$ are finite dimensional vectors $\mathbf{p}$ in

$$\mathcal{P}(\mathcal{X}) = \left\{ \mathbf{p} = (p_c)_{c\in\mathcal{X}} : p_c \geq 0,\ \forall c \in \mathcal{X} \ \text{ and } \sum_{c\in\mathcal{X}} p_c = 1 \right\}.$$

A divergence measure on $\mathcal{P}(\mathcal{X})$ is a function

$$\begin{array}{rccc} \mathcal{D}: & (\mathcal{P}(\mathcal{X}))^2 & \longrightarrow & \overline{\mathbb{R}} \\ & (\mathbf{p}, \mathbf{q}) & \longmapsto & \mathcal{D}(\mathbf{p}, \mathbf{q}) \end{array} \tag{1}$$

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v12i3.3437

*Email addresses:* ba.amadou-diadie@ugb.edu.sn (A.D.Ba), gane-samb.lo@ugb.edu.sn (G.S. Lo)

 

such that $\mathcal{D}(\mathbf{p},\mathbf{p})=0$ for any $\mathbf{p}$ such that $(\mathbf{p},\mathbf{p})$ in the domain of application of $\mathcal{D}$.

The function $\mathcal{D}$ is not necessarily a mapping. And if it is, it is not always symmetric and it does neither have to be a metric. In lack of symmetry, the following more general notation is more appropriate :

$$\begin{array}{rccc}\mathcal{D}: & \mathcal{P}_1(\mathcal{X})\times\mathcal{P}_2(\mathcal{X}) & \longrightarrow & \overline{\mathbb{R}} \\ & (\mathbf{p},\mathbf{q}) & \longmapsto & \mathcal{D}(\mathbf{p},\mathbf{q}),\end{array} \tag{2}$$

where $\mathcal{P}_1(\mathcal{X})$ and $\mathcal{P}_2(\mathcal{X})$ are two families of probability distributions on $\mathcal{X}$, not necessarily the same. To better explain our concern, let us introduce some of the most celebrated divergence measures.

Let $(\mathbf{p},\mathbf{q})\in\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{X})$ with $\mathcal{X}=\{c_1,c_2,\cdots,c_r\}$, and let $X$ and $Y$ two randoms variables such that

$$\mathbb{P}(X=c_j)=p_j,\quad \text{and}\quad \mathbb{P}(Y=c_j)=q_j,\quad j\in\{1,\cdots,r\},$$

and set $\mathbf{p}=(p_1,\cdots,p_r)^t$ and $\mathbf{q}=(q_1,\cdots,q_r)^t$.

The four most popular divergence measures are :

(1) The $L_2^2$-divergence measure :

$$\mathcal{D}_{L_2}(\mathbf{p},\mathbf{q})=\sum_{j=1}^{r}(p_j-q_j)^2. \tag{3}$$

(2) The family of Rényi's divergence measures indexed by $\alpha\neq 1$, $\alpha>0$, known under the name of Rényi-$\alpha$ :

$$\mathcal{D}_{R,\alpha}(\mathbf{p},\mathbf{q})=\frac{1}{\alpha-1}\log\left(\sum_{j=1}^{r}p_j^{\alpha}q_j^{1-\alpha}\right). \tag{4}$$

(3) The family of Tsallis divergence measures indexed by $\alpha\neq 1$, $\alpha>0$, also known under the name of Tsallis-$\alpha$ :

$$\mathcal{D}_{T,\alpha}(\mathbf{p},\mathbf{q})=\frac{1}{\alpha-1}\left(\sum_{j=1}^{r}p_j^{\alpha}q_j^{1-\alpha}-1\right). \tag{5}$$

(4) The Kullback-Leibler divergence measure

$$\mathcal{D}_{KL}(\mathbf{p},\mathbf{q})=\sum_{j=1}^{r}p_j\log(p_j/q_j). \tag{6}$$

The latter, the Kullback-Leibler divergence measure, may be interpreted as a limit case of both the Rényi's family and the Tsallis' one by letting $\alpha \to 1$. As well, for $\alpha$ near 1, the Tsallis family $\mathcal{D}_{T,\alpha}(\mathbf{p},\mathbf{q})$ may be seen as derived from $\mathcal{D}_{R,\alpha}(\mathbf{p},\mathbf{q})$ based on the first order expansion of the logarithm function in the neighborhood of the unity. Here for ease of notation we refer the notation log as the natural logarithm.

From this small sample of divergence measures, we may give the following remarks :

For both the Rényi and the Tsallis families, we may have computation problems. So without loss of generality, suppose

$$p_j > 0 \ \text{ and } \ q_j > 0, \ \ \forall j \in D = \{1, 2, \cdots, r\} \qquad (\mathbf{BD}) \tag{7}$$

If Assumption (7) holds, we do not have to worry about summation problems, especially for Tsallis, Rényi and Kulback-Leibler measures, in the computations arising in estimation theories. This explains why Assumption (7) is systematically used in a great number of works in that topic, for example, in [20], [10], [8], and recently in [1] to cite a few.

It is clear from the very form of these divergence measures that we do not have symmetry, unless for the special case where $\alpha = 1/2$. So we define the following symmetric version of divergence measures

$$\mathcal{D}^{(s)}(\mathbf{p},\mathbf{q}) = \frac{\mathcal{D}(\mathbf{p},\mathbf{q}) + \mathcal{D}(\mathbf{q},\mathbf{p})}{2},$$

provided that $\mathcal{D}(\mathbf{p},\mathbf{q})$ and $\mathcal{D}(\mathbf{q},\mathbf{p})$ are finite.

Both families are build on the following summation

$$\mathcal{S}_{\alpha}(\mathbf{p},\mathbf{q}) = \sum_{j \in D} p_j^{\alpha} q_j^{1-\alpha}, \ \ \text{with} \ \ \alpha \neq 1, \ \ \alpha > 0.$$

Although we are focusing on the aforementioned divergence measures in this paper, it is worth mentioning that there exist quite a few number of them. Let us cite for example the ones named after : Ali-Silvey or $f$-divergence (see [6]), Cauchy-Schwarz, Jeffrey divergence (see [5]), Chernoff (see [5]) , Jensen-Shannon (see [5]). According to [3], there is more than a dozen of different divergence measures in the literature.

Before coming back to our divergence measures estimation of interest, we want to highlight some important applications of them. Indeed, divergence has proven to be useful in applications. Let us cite a few of them :

(a) They heavily intervene in information theory and recently in machine learning.

(b) They have been used as similarity measures in image registration or multimedia classification (see [17]).

(c) They are also used as loss functions in evaluating and optimizing the performance of density estimation methods (see [8]).

(d) Divergence estimates can also be used to determine sample sizes required to achieve given performance levels in hypothesis testing.

(e) There has been a growing interest in applying divergence to various fields of science and engineering for the purpose of estimation, classification, etc. (See [2] and [13]).

(f) Divergence also plays a central role in the frame of large deviations results including the asymptotic rate of decrease of error probability in binary hypothesis testing problems.

(g) The estimation of divergence between the samples drawn from unknown distributions gauges the distance between those distributions. Divergence estimates can then be used in clustering and in particular for deciding whether the samples come from the same distribution by comparing the estimate to a threshold.

(h) Divergence gauges how differently two random variables are distributed and it provides a useful measure of discrepancy between distributions.

The reader may find more applications and descriptions in the following papers : [11], [7], [18], [9], [17], and [15].

In the next subsection, we describe the frame in which we place the estimation problems we deal in this paper.

## 1.2. Statistical Estimations

The divergence measures may be applied to two statistical problems among others.

**(A)** First, it may be used as a fitting problem as described here. Let $X_1, X_2, \cdots$ a sample of replications of $X$ with an unknown probability distribution $\mathbf{p}$ and we want to test the hypothesis that $\mathbf{p}$ is equal to a known and fixed probability $\mathbf{p}_0$. Theoretically, we can answer this question by estimating a divergence measure $\mathcal{D}(\mathbf{p}, \mathbf{p}_0)$ by a plug-in estimator $\mathcal{D}(\widehat{\mathbf{p}}_n, \mathbf{p})$ where, for each $n \geq 1$, $\mathbf{p}$ is replaced by an estimator $\widehat{\mathbf{p}}_n$ of the probability law, which is based on sample $X_1$, $X_2$, ..., $X_n$, to be precised.
From there establishing an asymptotic theory of $\Delta_n = \mathcal{D}(\widehat{\mathbf{p}}_n, \mathbf{p}_0) - \mathcal{D}(\mathbf{p}, \mathbf{p}_0)$ is thought to be necessary to conclude.

**(B)** Next, it may be used as tool of comparing for two distributions. We may have two samples and wonder whether they come from the same probability distribution. Here, we also may two different cases.

**(B1)** In the first, we have two independent samples $X_1, X_2, ....$ and $Y_1, Y_2, ....$ respectively from a random variable $X$ and $Y$ according the probability distributions $\mathbf{p}$ and $\mathbf{q}$. Here the estimated divergence $\mathcal{D}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$, where $n$ and $m$ are the sizes of the available samples, is the natural estimator of $\mathcal{D}(\mathbf{p}, \mathbf{q})$ on which depends the statistical test of the hypothesis : $\mathbf{p} = \mathbf{q}$.

**(B2)** But the data may also be paired $(X, Y)$, $(X_1, Y_1), (X_2, Y_2), ...$, that is $X_i$ and $Y_i$ are measurements of the same case $i = 1, 2, ...$ In such a situation, testing the equality of the margins $\mathbf{p}_X = \mathbf{p}_Y$ should be based on an estimator $\widehat{\mathbf{p}}^{(n)}_{X,Y}$ of the joint probability law of the couple $(X, Y)$ based of the paired observations $(X_i, Y_i)$, $i = 1, 2, \ldots, n$.

We did not encounter the approach (B2) in the literature. In the (B1) approach, almost all the papers used the same sample size, at the exception of [19], for the double-size estimation problem. In our view, the study case should rely on the available data so that using the same sample size may lead to a loss of information. To apply their method, one should take the minimum of the two sizes and then loose information. We suggest to come back to a general case and then study the asymptotic theory of $\mathcal{D}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$ based on samples $X_1, X_2, .., X_n$. and $Y_1, Y_2, ..., Y_m$. In this paper, we will systematically use arbitrary samples sizes.

## 1.3. Previous work

In the context of the situation (B1), there are several papers dealing with the estimation of the divergence measures. As we are concerned in this paper by the weak laws of the estimators, our review on that problematic did not return significant things. Instead, the literature presented us many kinds of results on almost-sure efficiency of the estimation, with rates of convergences and laws of the iterated logarithm, $L^p$ $(p = 1, 2)$ convergences, etc. To be precise, [4] used recent techniques based on functional empirical process to provide a series of interesting rates of convergence of the estimators in the case of one-sided approach for the class de Rényi, Tsallis, Kullback-Leibler to cite a few. Unfortunately, the authors did not address the problem of integrability, taking that the divergence measures are finite. Although the results should be correct under the boundedness assumption (7) (**BD**) we described earlier, a new formulation in that frame would be welcome. In the context of the situation (B1), we may cite first the works of [20] and [10]. They both used divergence measures based on probability density functions and concentrated of Rényi-$\alpha$, Tsallis-$\alpha$ and Kullback-Leibler.

Specifically, [10] defined Reyni and Tsallis estimators by correcting the plug-in estimator and established that, as long as $\mathcal{D}_{R,\alpha}(\mathbf{p},\mathbf{q}) \geq c$ and $\mathcal{D}_{T,\alpha}(\mathbf{p},\mathbf{q}) \geq c$, for some constant $c > 0$, then

$$\mathbb{E}\left|\mathcal{D}_{R,\alpha}(\widehat{\mathbf{p}}_n,\widehat{\mathbf{q}}_n) - \mathcal{D}_{R,\alpha}(\mathbf{p},\mathbf{q})\right| \leq c\left(n^{-1/2} + n^{-\frac{3s}{2s+d}}\right)$$

and

$$\mathbb{E}\left|\mathcal{D}_{T,\alpha}(\widehat{\mathbf{p}}_n,\widehat{\mathbf{q}}_n) - \mathcal{D}_{T,\alpha}(\mathbf{p},\mathbf{q})\right| \leq c\left(n^{-1/2} + n^{-\frac{3s}{2s+d}}\right).$$

[19] used a $k-$nearest-neighbor approach to prove that if $|\alpha - 1| < k$, $(\alpha \neq 1)$ then

$$\lim_{n,m\to\infty} \mathbb{E}\left[\mathcal{D}_{T,\alpha}(\widehat{\mathbf{p}}_n,\widehat{\mathbf{q}}_m) - \mathcal{D}_{T,\alpha}(\mathbf{p},\mathbf{q})\right]^2 = 0$$

and

$$\lim_{n,m\to\infty} \mathbb{E}\left(\mathcal{D}_{R,\alpha}(\widehat{\mathbf{p}}_n,\widehat{\mathbf{q}}_m)\right) = \mathcal{D}_{R,\alpha}(\mathbf{p},\mathbf{q}).$$

There has been recent interest in deriving convergence rates for divergence estimators (see [16] and [10]). The rates are typically derived in terms of smoothness $s$ of the densities.

Similarly, [21] showed that when $s > d$ a $k$-nearest-neighbor style estimator achieves rate $n^{-2/d}$ (in absolute error) ignoring logarithmic factors. In a follow up work, the authors improved this result to $O(n^{-1/2})$ using an ensemble of weak estimators, but they require $s > d$ orders of smoothness.

[20] provided an estimator for Rényi$-\alpha$ divergences as well as general density functionals that uses a *mirror image* kernel density estimator. They obtained exponential inequalities for the deviation of the estimators from the true value.

[12] studied an $\varepsilon-$nearest neighbor estimator for the $L_2-$divergence that enjoys the same rate of convergence as the projection-based estimator of [10].

### 1.4. Main contributions

Our main contribution may be summarized as follows : for data sampled from one or two unknown random variables, we derive almost sure convergency and central limit theorems for empirical $\phi-$ divergences. We will focus on divergence measures between discrete probability distribution. As well, our results applied to the approaches (A) and (B1) defined above. As a consequence, we estimate divergence measures by their plug-in counterparts, meaning that we replace the probability mass function (*p.m.f.*) in the expression of the divergence measure by a nonparametric estimator of the *p.m.f.*
We also wish to get first general laws for an arbitrary functional of the form

$$J(\mathbf{p},\mathbf{q}) = \sum_{j\in D} \phi(p_j, q_j), \tag{8}$$

where $\phi : (0,1)^2 \to \mathbb{R}$ is a twice continously differentiable function. The results on the functional $J(\mathbf{p},\mathbf{q})$, which is also known under the name of $\phi$-divergence, will lead to those on the particular cases of the Tsallis, Rényi, and Kullback-Leibler measures.

### 1.5. Overview of the paper

The rest of the paper is organized as follows. In SECTION 2, we define estimators of the *p.m.f.*'s $p_j$ and $q_j$ based on two i.i.d. samples according respectively to $\mathbf{p}$ and to $\mathbf{q}$. These ones allow us to define the empirical $\phi-$divergences. In SECTION 3, we will give our full results for functional $J(\mathbf{p},\mathbf{q})$ both one sided and two-sided approaches. In Section 4, we will particularize the results for specific measures we already described. In SECTION 5 we present some simulations confirming our results. Finally in SECTION 6, we conclude.

## 2. Empirical $\phi-$ divergence

### 2.1. Notations and main results

Before we state the main results we need a few definitions. Let $X$ and $Y$ two randoms variables defined on the probability distributions $(\mathcal{X},\mathcal{A},\mathbb{P})$ with $\mathcal{X}=\{c_1,c_2,\cdots,c_r\}$ and $\mathbf{p}=(p_j)_{1\leq j\leq r}$ and $\mathbf{q}=(q_j)_{1\leq j\leq r}$ two discrete probability distributions on $\mathcal{X}$ such that, for any $j\in D$

$$p_j=\mathbb{P}(X=c_j) \text{ and } q_j=\mathbb{P}(Y=c_j).$$

We suppose that (7) is satisfied that is $\forall\, j\in D,\ \ p_j>0 \text{ and } q_j>0.$
Define the empirical probability distribution generated by i.i.d. random variables $X_1,\cdots,X_n$ from the probability distribution $\mathbf{p}$ as

$$\widehat{\mathbf{p}}_n=(\widehat{p}_n^{c})_{c\in\mathcal{X}}, \quad \text{where } \widehat{p}_n^{c_j} \ = \ \frac{1}{n}\sum_{i=1}^{n}1_{c_j}(X_i) \tag{9}$$

where $1_{c_j}(X_i)=\begin{cases}1 & \text{if } X_i=c_j\\ 0 & \text{otherwise}\end{cases}$ for any $j\in D$.

$\widehat{\mathbf{q}}_m$ is defined in the same way by $Y_1,\cdots,Y_m \overset{i.i.d.}{\sim} \mathbf{q}$ that is

$$\widehat{\mathbf{q}}_m=(\widehat{q}_m^{c})_{c\in\mathcal{X}}, \quad \text{where } \widehat{q}_m^{c_j}=\frac{1}{m}\sum_{i=1}^{m}1_{c_j}(Y_i). \tag{10}$$

For a given $j\in D$, this empirical estimator $\widehat{p}_n^{c_j}$ of $p_j$ is strongly consistent and asymptotically normal. Precisely, when $n$ tends to infinity,

$$\widehat{p}_n^{c_j}-p_j \xrightarrow{a.s.} 0 \tag{11}$$

$$\sqrt{n}(\widehat{p}_n^{c_j}-p_j) \overset{\mathcal{D}}{\rightsquigarrow} Z_{p_j}, \tag{12}$$

where $Z_{p_j} \overset{d}{\sim} \mathcal{N}(0, p_j(1-p_j))$.

We denote by $\xrightarrow{a.s.}$ the *almost sure convergence* and $\overset{\mathcal{D}}{\rightsquigarrow}$ the *convergence in distribution*. The notation $\overset{d}{\sim}$ denote the *equality in distribution*.
These asymptotic properties derive from the law of large numbers and central limit theorem.

For sake of simplicity, we introduce the two following notations :

$$\Delta_{p_n}^{c_j} = \widehat{p}_n^{c_j} - p_j, \quad \Delta_{q_m}^{c_j} = \widehat{q}_m^{c_j} - q_j, \quad \forall j \in D,$$

$$a_n = \sup_{j \in D} |\Delta_{p_n}^{c_j}|, \qquad b_m = \sup_{j \in D} |\Delta_{q_m}^{c_j}|, \quad \text{and} \quad c_{n,m} = \max(a_n, b_m). \tag{13}$$

For any $j \in D$, set

$$\delta_n(p_j) = \sqrt{n/p_j}\Delta_{p_n}^{c_j} \quad \text{and} \quad \delta_m(q_j) = \sqrt{m/q_j}\Delta_{q_m}^{c_j}.$$

We recall that, since for a fixed $j \in D$, $n\widehat{p}_n^{c_j}$ has a binomial distribution with parameters $n$ and success probability $p_j$, we have

$$\mathbb{E}\left(\widehat{p}_n^{c_j}\right) = p_j \quad \text{and} \quad \mathbb{V}(\widehat{p}_n^{c_j}) = \frac{p_j(1-p_j)}{n}.$$

Furthermore, by the strong law of large numbers, we know that

$$\Delta_{p_n}^{c_j} \xrightarrow{a.s.} 0, \quad \text{as} \quad n \to +\infty,$$

for a fixed $j \in D$.

And finally, by the asymptotic Gaussian limit of the multinomial law (see for example Chapter 1, Section 4 in [14]), we have

$$\left(\delta_n(p_j),\ j \in D\right) \overset{\mathcal{D}}{\rightsquigarrow} Z(\mathbf{p}) \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \Sigma_{\mathbf{p}}), \quad \text{as} \quad n \to +\infty, \tag{14}$$

$$\text{and} \qquad \left(\delta_m(q_j),\ j \in D)\right) \overset{\mathcal{D}}{\rightsquigarrow} Z(\mathbf{q}) \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \Sigma_{\mathbf{q}}), \quad \text{as} \quad m \to +\infty, \tag{15}$$

where $Z(\mathbf{p}) = (Z_{p_j}, j \in D)^t$ and $Z(\mathbf{q}) = (Z_{q_j}, j \in D)^t$ are two independent centered Gaussian random vectors of dimension $r$ having respectively the following elements :

$$(\Sigma_{\mathbf{p}})_{(i,j)} = (1-p_j)\delta_{ij} - \sqrt{p_i p_j}(1-\delta_{ij}), \quad (i,j) \in D^2 \tag{16}$$

$$(\Sigma_{\mathbf{q}})_{(i,j)} = (1-q_j)\delta_{ij} - \sqrt{q_i q_j}(1-\delta_{ij}), \quad (i,j) \in D^2, \tag{17}$$

where $\delta_{ij} = \begin{cases} 1 & \text{for} \quad i = j \\ 0 & \text{for} \quad i \neq j \end{cases}$.

For a fixed $j \in D$, we have also

$$\mathbb{E}\left(\widehat{q}_m^{c_j}\right) = q_j, \qquad \mathbb{V}(\widehat{q}_m^{c_j}) = \frac{q_j(1-q_j)}{m}, \quad \text{and} \quad \Delta_{q_m}^{c_j} \overset{a.s.}{\longrightarrow} 0, \text{ as } m \to +\infty.$$

## 2.2. $\phi$-divergence measure

**Definition 1.** *The $\phi$-divergence between the two probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is given by*

$$J(\boldsymbol{p}, \boldsymbol{q}) = \sum_{j \in D} \phi(p_j, q_j), \tag{18}$$

*where $\phi : [0,1]^2 \to \mathbb{R}$ is a measurable function having continuous second order partial derivatives.*

The results on the functional $J(\mathbf{p}, \mathbf{q})$ will lead to those on the particular cases of the Tsallis, Rényi, and Kullback-Leibler measures.

Based on (9) and (10), we will use the following empirical $\phi$-divergences :

$$\begin{aligned} J(\widehat{\mathbf{p}}_n, \mathbf{q}) &= \sum_{j \in D} \phi(\widehat{p}_n^{c_j}, q_j), \qquad J(\mathbf{p}, \widehat{\mathbf{q}}_m) = \sum_{j \in D} \phi(p_j, \widehat{q}_m^{c_j}), \\ \text{and} \qquad & J(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m) = \sum_{j \in D} \phi(\widehat{p}_n^{c_j}, \widehat{q}_m^{c_j}). \end{aligned}$$

Denote

$$\begin{aligned} \phi_1^{(1)}(s,t) &= \frac{\partial \phi}{\partial s}(s,t), \ \phi_2^{(1)}(s,t) = \frac{\partial \phi}{\partial t}(s,t), \quad \phi_1^{(2)}(s,t) = \frac{\partial^2 \phi}{\partial s^2}(s,t), \\ \text{and} \quad \phi_2^{(2)}(s,t) &= \frac{\partial^2 \phi}{\partial t^2}(s,t), \qquad \phi_{1,2}^{(2)}(s,t) = \phi_{2,1}^{(2)}(s,t) = \frac{\partial^2 \phi}{\partial s \partial t}(s,t). \end{aligned}$$

Set

$$A_{1,p} = \sum_{j \in D} |\phi_1^{(1)}(p_j, q_j)|, \quad A_{2,q} = \sum_{j \in D} |\phi_2^{(1)}(p_j, q_j)|, \tag{19}$$

$$A_{3,q} = \sum_{j \in D} |\phi_1^{(1)}(q_j, p_j)|, \quad \text{and} \quad A_{4,p} = \sum_{j \in D} |\phi_2^{(1)}(q_j, p_j)|. \tag{20}$$

# 3. Statements of the main results

## 3.1. Main results

Here are our main results. The first concerns the almost sure efficiency of the estimators.

**Theorem 1.** *Let $\boldsymbol{p}$ and $\boldsymbol{q}$ two probability distributions and $\widehat{\boldsymbol{p}}_n$ and $\widehat{\boldsymbol{q}}_m$ be generated by i.i.d. samples $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ according respectively to $\boldsymbol{p}$ and $\boldsymbol{q}$ and given by* (9) *and* (10)*, the assumption* (7) *be satisfied. Then the following asymptotic results hold.*

*(a) One sample*

$$\limsup_{n\to+\infty} \frac{|J(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - J(\boldsymbol{p}, \boldsymbol{q})|}{a_n} \leq A_{1,p}, \quad a.s., \tag{21}$$

$$\limsup_{m\to+\infty} \frac{|J(\boldsymbol{p}, \widehat{\boldsymbol{q}}_m) - J(\boldsymbol{p}, \boldsymbol{q})|}{b_m} \leq A_{2,q}, \quad a.s., \tag{22}$$

*(b) Two samples :*

$$\limsup_{(n,m)\to(+\infty,+\infty)} \frac{|J(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - J(\boldsymbol{p}, \boldsymbol{q})|}{c_{n,m}} \leq A_{1,p} + A_{2,q} \quad a.s., \tag{23}$$

*where $a_n$, $b_m$ and $c_{n,m}$ are as in* (13) *and $A_{1,p}$ and $A_{2,q}$ as in* (19).

*Proof.* In the proof, we will systematically use the mean values theorem. In the multivariate handling, we prefer to use the Taylor-Lagrange-Cauchy as stated in [22], page 230.

For a fixed $j \in D$, we have

$$\begin{aligned} \phi(\widehat{p}_n^{c_j}, q_j) &= \phi(p_j + \Delta_{p_n}^{c_j}, q_j) \\ &= \phi(p_j, q_j) + \Delta_{p_n}^{c_j} \phi_1^{(1)}(p_j + \theta_{1,j}\Delta_{p_n}^{c_j}, q_j), \end{aligned} \tag{24}$$

by applying the mean value theorem to the function $(.) \mapsto \phi((.), q_j)$ and where $\theta_{1,j}$ is some number lying between 0 and 1. In the sequel, any $\theta_{i,j}$, $i = 1, 2, \cdots$ satisfies $|\theta_{i,j}| < 1$.

By applying again the mean values theorem to the function $(.) \mapsto \phi_1^{(1)}((.), q_j)$, we have

$$\phi_1^{(1)}(p_j + \theta_{1,j}\Delta_{p_n}^{c_j}, q_j) = \phi_1^{(1)}(p_j, q_j) + \theta_{1,j}\Delta_{p_n}^{c_j}\phi_1^{(2)}(p_j + \theta_{2,j}\Delta_{p_n}^{c_j}, q_j).$$

We can write (24) as

$$\phi(\widehat{p}_n^{c_j}, q_j) = \phi(p_j, q_j) + \Delta_{p_n}^{c_j}\phi_1^{(1)}(p_j, q_j) + \theta_{1,j}(\Delta_{p_n}^{c_j})^2\phi_1^{(2)}(p_j + \theta_{2,j}\Delta_{p_n}^{c_j}, q_j).$$

Now we have, by summation over $j \in D$,

$$J(\widehat{\mathbf{p}}_n, \mathbf{q}) - J(\mathbf{p}, \mathbf{q}) = \sum_{j\in D} \Delta_{p_n}^{c_j}\phi_1^{(1)}(p_j, q_j)$$

$$+ \sum_{j\in D} \theta_{1,j}(\Delta_{p_n}^{c_j})^2 \phi_1^{(2)}(p_j + \theta_{2,j}\Delta_{p_n}^{c_j}, q_j), \tag{25}$$

hence

$$|J(\widehat{\mathbf{p}}_n, \mathbf{q}) - J(\mathbf{p}, \mathbf{q})| \leq a_n \sum_{j\in D} |\phi_1^{(1)}(p_j, q_j)| + a_n^2 \sum_{j\in D} |\phi_1^{(2)}(p_j + \theta_{2,j}\Delta_{p_n}^{c_j}, q_j)|.$$

Therefore

$$\frac{|J(\widehat{\mathbf{p}}_n, \mathbf{q}) - J(\mathbf{p}, \mathbf{q})}{a_n} \leq A_{1,p} + a_n \sum_{j\in D} |\phi_1^{(2)}(p_j + \theta_{2,j}\Delta_{p_n}^{c_j}, q_j)|,$$

and then

$$\limsup_{n\to\infty} \frac{|J(\widehat{\mathbf{p}}_n, \mathbf{q}) - J(\mathbf{p}, \mathbf{q})}{a_n} \leq A_{1,p},$$

since $A_{1,p} < \infty$, $a_n \xrightarrow{a.s.} 0$, and

$$\sum_{j\in D} |\phi_1^{(2)}(p_j + \theta_{2,j}\Delta_{p_n}^{c_j}, q_j)| \to \sum_{j\in D} |\phi_1^{(2)}(p_j, q_j)| < \infty \quad \text{as} \quad n\to\infty.$$

This proves (21).

Formula (22) is obtained in a similar way. We only need to adapt the result concerning the first coordinate to the second.

The proof of (23) comes by splitting $\sum_{j\in D}\left(\phi(\widehat{p}_n^{c_j}, \widehat{q}_m^{c_j}) - \phi(p_j, q_j)\right)$, into the following two terms

$$\begin{aligned}\sum_{j\in D}\left(\phi(\widehat{p}_n^{c_j}, \widehat{q}_m^{c_j}) - \phi(p_j, q_j)\right) &= \sum_{j\in D}\left(\phi(\widehat{p}_n^{c_j}, \widehat{q}_m^{c_j}) - \phi(p_j, \widehat{q}_m^{c_j})\right) \\ &+ \sum_{j\in D}\left(\phi(p_j, \widehat{q}_m^{c_j}) - \phi(p_j, q_j)\right) \\ &\equiv I_{n,1} + I_{n,2}.\end{aligned}$$

We already know how to handle $I_{n,2}$. As to $I_{n,1}$, we may still use the Taylor-Lagrange-Cauchy formula since we have, for a fixed $j \in D$,

$$\left\|(\widehat{p}_n^{c_j}, \widehat{q}_m^{c_j}) - (p_j, \widehat{q}_m^{c_j})\right\|_\infty = \left\|(\widehat{p}_n^{c_j} - p_j), 0)\right\|_\infty = a_n \to 0.$$

By the Taylor-Lagrange-Cauchy (see [22], page 230), we have

$$\begin{aligned} I_{n,1} &= \sum_{j\in D} \Delta_{p_n}^{c_j} \phi(\widehat{p}_n^{c_j} + \theta_j \Delta_{p_n}^{c_j}, \widehat{q}_m^{c_j}) \\ &\leq a_n \sum_{j\in D} |\phi(\widehat{p}_n^{c_j} + \theta_j \Delta_{p_n}^{c_j}, \widehat{q}_m^{c_j})| \\ &= a_n(A_1 + o(1)). \end{aligned}$$

From there, the combination of these remarks directs to the result.

The second main result concerns the asymptotic normality of the estimators.

Let

$$V_1(\mathbf{p}, \mathbf{q}) = \sum_{j\in D} p_j(1-p_j)(\phi_1^{(1)}(p_j, q_j))^2 - 2 \sum_{(i,j)\in D^2,\, i\neq j} p_i p_j \phi_1^{(1)}(p_i, q_i)\phi_1^{(1)}(p_j, q_j) \quad (26)$$

and

$$V_2(\mathbf{p}, \mathbf{q}) = \sum_{j\in D} q_j(1-q_j)(\phi_2^{(1)}(p_j, q_j))^2 - 2 \sum_{(i,j)\in D^2,\ i\neq j} q_i q_j \phi_2^{(1)}(p_i, q_i)\phi_2^{(1)}(p_j, q_j) \quad (27)$$

**Theorem 2.** *Under the same assumptions as in Theorem 1, the following central limit theorems hold.*

*(a) One sample : as* $n \to +\infty$,

$$\sqrt{n}(J(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - J(\boldsymbol{p}, \boldsymbol{q})) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}\left(0, V_1(\boldsymbol{p}, \boldsymbol{q})\right), \quad (28)$$

$$\sqrt{m}(J(\boldsymbol{p}, \widehat{\boldsymbol{q}}_m) - J(\boldsymbol{p}, \boldsymbol{q})) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}\left(0, V_2(\boldsymbol{p}, \boldsymbol{q})\right). \quad (29)$$

*(b) Two samples : for* $(n, m) \to (+\infty, +\infty)$ *and* $nm/(n+m) \to \gamma \in (0,1)$,

$$\left(\frac{nm}{mV_1(\boldsymbol{p}, \boldsymbol{q}) + nV_2(\boldsymbol{p}, \boldsymbol{q})}\right)^{1/2} (J(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - J(\boldsymbol{p}, \boldsymbol{q})) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}\left(0, 1\right). \quad (30)$$

*Proof.* Let us prove (28). By going back to (25), we have

$$\sqrt{n}(J(\widehat{\mathbf{p}}_n, \mathbf{q}) - J(\mathbf{p}, \mathbf{q})) = \sum_{j\in D} \sqrt{p_j}\delta_n(p_j)\phi_1^{(1)}(p_j, q_j) + \sqrt{n}R_{1,n}$$

where

$$R_{1,n} = \sum_{j\in D} \theta_{1,j}(\Delta_{p_n}^{c_j})^2 \phi_1^{(2)}(p_j + \theta_{2,j}\Delta_{p_n}^{c_j}, q_j).$$

Now using Formula (14) above, we get,

$$\sum_{j\in D}\sqrt{p_j}\delta_n(p_j)\phi_1^{(1)}(p_j,q_j)\overset{\mathcal{D}}{\rightsquigarrow}\sum_{j\in D}\phi_1^{(1)}(p_j,q_j)\sqrt{p_j}Z_{p_j},\quad \text{as}\quad n\rightarrow+\infty$$

which follows a centered normal law of variance $V_1(\mathbf{p},\mathbf{q})$ :

$$V_1(\mathbf{p},\mathbf{q})=\sum_{j\in D}(1-p_j)(\phi_1^{(1)}(p_j,q_j))^2-2\sum_{(i,j)\in D^2,\,i\neq j}\sqrt{p_ip_j}\phi_1^{(1)}(p_i,q_i)\phi_1^{(1)}(p_j,q_j)$$

since

$$\begin{aligned}\mathbb{V}\left(\sum_{j\in D}\phi_1^{(1)}(p_j,q_j)\sqrt{p_j}Z_{p_j}\right) &= \sum_{j\in D}\mathbb{V}(\phi_1^{(1)}(p_j,q_j)\sqrt{p_j}Z_{p_j})\\ &\quad+\ 2\sum_{(i,j)\in D^2,\,i\neq j}\mathbb{C}\text{ov}(\phi_1^{(1)}(p_i,q_i)\sqrt{p_i}Z_{p_i},\phi_1^{(1)}(p_j,q_j)\sqrt{p_j}Z_{p_j})\\ &= \sum_{j\in D}p_j(\phi_1^{(1)}(p_j,q_j))^2\mathbb{V}(Z_{p_j})\\ &\quad+\ 2\sum_{(i,j)\in D^2,\,i\neq j}\sqrt{p_ip_j}\phi_1^{(1)}(p_i,q_i)\phi_1^{(1)}(p_j,q_j)\mathbb{C}\text{ov}(Z_{p_i},Z_{p_j})\\ &= \sum_{j\in D}p_j(1-p_j)(\phi_1^{(1)}(p_j,q_j))^2\\ &\quad-\ 2\sum_{(i,j)\in D^2,\,i\neq j}p_ip_j\phi_1^{(1)}(p_i,q_i)\phi_1^{(1)}(p_j,q_j).\end{aligned}$$

Finally, the proof will be complete if we show that $\sqrt{n}R_{1,n}$ converges, in probability, to zero, as $n$ tends to infinity. We have

$$\left|\sqrt{n}R_{1,n}\right|\leq\sqrt{n}a_n^2\sum_{j\in D}|\phi_1^{(2)}(p_j+\theta_{2,j}\Delta_{p_n}^{c_j},q_j)|. \tag{31}$$

Let show that

$$\sqrt{n}a_n^2=o_{\mathbb{P}}(1).$$

By the Bienaymé-Tchebychev inequality, we have, for any $\epsilon>0$ and for $j\in D$,

$$\mathbb{P}(\sqrt{n}(\widehat{p}_n^{c_j}-p_j)^2\geq\epsilon)=\mathbb{P}\left(|\widehat{p}_n^{c_j}-p_j|\geq\frac{\sqrt{\epsilon}}{n^{1/4}}\right)\leq\frac{p_j(1-p_j)}{\epsilon n^{1/2}},$$

which implies that $\sqrt{n}a_n^2$ converges in probability to 0 as $n\rightarrow+\infty$.

Finally from (31) we have $\sqrt{n}R_{1,n} \stackrel{\mathbb{P}}{\to} 0$ as $n \to +\infty$ which implies

$$\sqrt{n}(J(\widehat{\mathbf{p}}_n, \mathbf{q}) - J(\mathbf{p}, \mathbf{q})) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}\left(0, V_1(\mathbf{p}, \mathbf{q})\right), \quad \text{as} \quad n \to +\infty.$$

This ends the proof of (28).

The result (29) is obtained by a symmetry argument by swapping the roles of $\mathbf{p}$ and $\mathbf{q}$.

Now, it remains to prove Formula (30) of the theorem. Let us use bi-variate Taylor-Lagrange-Cauchy formula to get,

$$\begin{aligned} J(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m) - J(\mathbf{p}, \mathbf{q}) &= \sum_{j \in D} \Delta_{p_n}^{c_j} \phi_1^{(1)}(p_j, q_j) + \sum_{j \in D} \Delta_{q_m}^{c_j} \phi_2^{(1)}(p_j, q_j) \\ & \frac{1}{2} \sum_{j \in D} \biggl( (\Delta_{p_n}^{c_j})^2 \phi_1^{(2)} + \Delta_{p_n}^{c_j} \Delta_{q_m}^{c_j} \phi_{1,2}^{(2)} + (\Delta_{q_m}^{c_j})^2 \phi_2^{(2)} \biggr) \biggl( u_n^{c_j}, v_m^{c_j} \biggr), \end{aligned}$$

where

$$(u_n^{c_j}, v_n^{c_j}) = (p_j + \theta \Delta_{p_n}^{c_j},\ q_j + \theta_j \Delta_{q_m}^{c_j}).$$

Thus we get

$$J(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m) - J(\mathbf{p}, \mathbf{q}) = \frac{1}{\sqrt{n}} N_n(\mathbf{p}) + \frac{1}{\sqrt{m}} N_m(\mathbf{q}) + R_{n,m},$$

where

$$N_n(\mathbf{p}) = \sum_{j \in D} \sqrt{p_j} \delta_n(p_j) \phi_1^{(1)}(p_j, q_j) \stackrel{\mathcal{D}}{\rightsquigarrow} \sum_{j \in D} \phi_1^{(1)}(p_j, q_j) Z_{p_j}, \quad \text{as} \quad n \to +\infty,$$

$$N_m(\mathbf{q}) = \sum_{j \in D} \sqrt{q_j} \delta_m(q_j) \phi_2^{(1)}(p_j, q_j) \stackrel{\mathcal{D}}{\rightsquigarrow} \sum_{j \in D} \phi_2^{(1)}(p_j, q_j) Z_{q_j}, \quad \text{as} \quad m \to +\infty,$$

and $R_{n,m}$ is given by

$$\frac{1}{2} \sum_{j \in D} \biggl( (\Delta_{p_n}^{c_j})^2 \phi_1^{(2)} + \Delta_{p_n}^{c_j} \Delta_{q_m}^{c_j} \phi_{1,2}^{(2)} + (\Delta_{q_m}^{c_j})^2 \phi_2^{(2)} \biggr) \biggl( u_n^{c_j}, v_m^{c_j} \biggr).$$

First, we have that $N_n(\mathbf{p})$ and $N_m(\mathbf{p})$ are independents and hence

$$N_n(\mathbf{p}) \stackrel{\mathcal{L}}{\sim} \mathcal{N}\left(0, V_1(\mathbf{p}, \mathbf{q})\right) \quad \text{and} \quad N_m(\mathbf{q}) \stackrel{\mathcal{L}}{\sim} \mathcal{N}\left(0, V_2(\mathbf{p}, \mathbf{q})\right).$$

Therefore

$$\frac{1}{\sqrt{n}} \sum_{j \in D} \sqrt{p_j} \delta_n(p_j) \phi_1^{(1)}(p_j, q_j) + \frac{1}{\sqrt{m}} \sum_{j \in D} \sqrt{q_j} \delta_m(q_j) \phi_2^{(1)}(p_j, q_j) = \mathcal{N}\left(0, \frac{V_1(\mathbf{p}, \mathbf{q})}{n} + \frac{V_2(\mathbf{p}, \mathbf{q})}{m}\right)$$

$$+ \ o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right).$$

Thus

$$J(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m) - J(\mathbf{p}, \mathbf{q}) = \mathcal{N}\left(0, \frac{V_1(\mathbf{p}, \mathbf{q})}{n} + \frac{V_2(\mathbf{p}, \mathbf{q})}{m}\right) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right) + R_{n,m}.$$

Next, we have

$$\begin{aligned}\frac{1}{\sqrt{\frac{V_1(\mathbf{p},\mathbf{q})}{n} + \frac{V_2(\mathbf{p},\mathbf{q})}{m}}} \left(J(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m) - J(\mathbf{p}, \mathbf{q})\right) = \mathcal{N}(0,1) &+ o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}} \frac{1}{\sqrt{\frac{V_1(\mathbf{p},\mathbf{q})}{n} + \frac{V_2(\mathbf{p},\mathbf{q})}{m}}}\right) \\ &+ o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}} \frac{1}{\sqrt{\frac{V_1(\mathbf{p},\mathbf{q})}{n} + \frac{V_2(\mathbf{p},\mathbf{q})}{m}}}\right) \\ &+ \frac{1}{\sqrt{\frac{V_1(\mathbf{p},\mathbf{q})}{n} + \frac{V_2(\mathbf{p},\mathbf{q})}{m}}} R_{n,m}.\end{aligned}$$

That leads to

$$\sqrt{\frac{nm}{mV_1(\mathbf{p}, \mathbf{q}) + nV_2(\mathbf{p}, \mathbf{q})}} \left(J(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m) - J(\mathbf{p}, \mathbf{q})\right) = \mathcal{N}(0,1) + o_{\mathbb{P}}(1) + \sqrt{\frac{nm}{mV_1(\mathbf{p}, \mathbf{q}) + nV_2(\mathbf{p}, \mathbf{q})}} R_{n,m},$$

since $m/(mV_1(\mathbf{p}, \mathbf{q}) + nV_2(\mathbf{p}, \mathbf{q}))$ and $m/(nV_1(\mathbf{p}, \mathbf{q}) + nV_2(\mathbf{p}, \mathbf{q}))$ are bounded, and then

$$o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}} \frac{1}{\sqrt{\frac{V_1(\mathbf{p},\mathbf{q})}{n} + \frac{V_2(\mathbf{p},\mathbf{q})}{m}}}\right) = o_{\mathbb{P}}\left(\sqrt{\frac{m}{mV_1(\mathbf{p}, \mathbf{q}) + nV_2(\mathbf{p}, \mathbf{q})}}\right) = o_{\mathbb{P}}(1),$$

and

$$o_{\mathbb{P}}\left(\frac{1}{\sqrt{m}} \frac{1}{\sqrt{\frac{V_1(\mathbf{p},\mathbf{q})}{n} + \frac{V_2(\mathbf{p},\mathbf{q})}{m}}}\right) = o_{\mathbb{P}}\left(\sqrt{\frac{n}{mV_1(\mathbf{p}, \mathbf{q}) + nV_2(\mathbf{p}, \mathbf{q})}}\right) = o_{\mathbb{P}}(1).$$

It remains to prove that $\left|\sqrt{\frac{nm}{mV_1(\mathbf{p},\mathbf{q})+nV_2(\mathbf{p},\mathbf{q})}} R_{n,m}\right| = o_{\mathbb{P}}(1)$. But we have, by the continuity assumptions on $\phi$ and on its partial derivatives and by the uniform converges of $\Delta_{p_n}^{c_j}$ and $\Delta_{q_m}^{c_j}$ to zero, that

$$\left|\sqrt{\frac{nm}{mV_1(\mathbf{p}, \mathbf{q}) + nV_2(\mathbf{p}, \mathbf{q})}} R_{n,m}\right| \leq$$

$$\frac{1}{2}\left(\sqrt{n}a_n^2(\sum_{j\in D}\phi_1^{(2)}(p_j,q_j)+o(1))\right)\left(\sqrt{\frac{m}{mV_1(\mathbf{p},\mathbf{q})+nV_2(\mathbf{p},\mathbf{q})}}\right)$$
$$+\ \frac{1}{2}\left(\sqrt{m}b_m^2(\sum_{j\in D}\phi_2^{(2)}(p_j,q_j)+o(1))\right)\left(\sqrt{\frac{n}{mV_1(\mathbf{p},\mathbf{q})+nV_2(\mathbf{p},\mathbf{q})}}\right)$$
$$+\ \frac{1}{2}\left(\sqrt{n}a_m b_m(\sum_{j\in D}\phi_2^{(2)}(p_j,q_j)+o(1))\right)\left(\sqrt{\frac{n}{mV_1(\mathbf{p},\mathbf{q})+nV_2(\mathbf{p},\mathbf{q})}}\right).$$

As previously, we have $\sqrt{n}a_n^2 = o_{\mathbb{P}}(1)$, $\sqrt{m}b_m^2 = o_{\mathbb{P}}(1)$ and $\sqrt{n}a_m b_m = o_{\mathbb{P}}(1)$.

From there, the conclusion is immediate.

## 3.2. Direct extensions

Quite a few number of divergence measures are not symmetric. Among these non-symmetric measures are some of the most interesting ones. For such measures, estimators of the form $J(\widehat{\mathbf{p}}_n,\mathbf{q})$, $J(\mathbf{p},\widehat{\mathbf{q}}_m)$ and $J(\widehat{\mathbf{p}}_n,\widehat{\mathbf{q}}_m)$ are not equal to $J(\mathbf{q},\widehat{\mathbf{p}}_n)$, $J(\widehat{\mathbf{q}}_m,\mathbf{p})$ and $J(\widehat{\mathbf{q}}_m,\widehat{\mathbf{p}}_n)$ respectively.

In one-sided tests, we have to decide whether the hypothesis $\mathbf{p}=\mathbf{q}$, for $\mathbf{q}$ known and fixed, is true based on data from $\mathbf{p}$. In such a case, we may use one of the statistics $J(\widehat{\mathbf{p}}_n,\mathbf{q})$ or $J(\mathbf{q},\widehat{\mathbf{p}}_n)$ to perform the tests. We may have information that allows us to prefer one of them. If not, it is better to use both of them, upon the finiteness of both $J(\mathbf{p},\mathbf{q})$ and $J(\mathbf{q},\mathbf{p})$, in a symmetrized form as

$$J^{(s)}(\mathbf{p},\mathbf{q}) = \frac{J(\mathbf{p},\mathbf{q})+J(\mathbf{q},\mathbf{p})}{2}. \tag{32}$$

The same situation applies when we face double-side tests, i.e., testing $\mathbf{p}=\mathbf{q}$ from data generated by $\mathbf{p}$ et $\mathbf{q}$.

**Asymptotic a.e. efficiency**.

**Theorem 3.** *Under the same assumptions as in Theorem 1, the following hold.*

*(a) One sample :*

$$\limsup_{n\to+\infty}\frac{\left|J^{(s)}(\widehat{\boldsymbol{p}}_n,\boldsymbol{q})-J^{(s)}(\boldsymbol{p},\boldsymbol{q})\right|}{a_n}\leq\frac{1}{2}\left(A_{1,p}+A_{4,p}\right)\quad a.e., \tag{33}$$

$$\limsup_{n\to+\infty}\frac{\left|J^{(s)}(\boldsymbol{p},\widehat{\boldsymbol{q}}_m)-J^{(s)}(\boldsymbol{p},\boldsymbol{q})\right|}{b_n}\leq\frac{1}{2}\left(A_{2,q}+A_{3,q}\right)\quad a.e., \tag{34}$$

*(b) Two samples :*

$$\limsup_{(n,m)\to(+\infty,+\infty)} \frac{\left|J^{(s)}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - J^{(s)}(\boldsymbol{p}, \boldsymbol{q})\right|}{c_{n,m}} \leq \frac{1}{2}\left(A_{1,p} + A_{2,q} + A_{3,q} + A_{4,p}\right), \ \ a.e. \tag{35}$$

*Proof.* The proof is established by considering the following equalities and by using Theorem 1

$$J^{(s)}(\widehat{\mathbf{p}}_n, \mathbf{q}) - J^{(s)}(\mathbf{p}, \mathbf{q}) = \frac{1}{2}\left(J(\widehat{\mathbf{p}}_n, \mathbf{q}) - J(\mathbf{p}, \mathbf{q})\right) + \frac{1}{2}\left(J(\mathbf{q}, \widehat{\mathbf{p}}_n) - J(\mathbf{q}, \mathbf{p})\right), \tag{36}$$

$$J^{(s)}(\mathbf{p}, \widehat{\mathbf{q}}_m) - J^{(s)}(\mathbf{p}, \mathbf{q})) = \frac{1}{2}\left(J(\mathbf{p}, \widehat{\mathbf{q}}_m) - J(\mathbf{p}, \mathbf{q})\right) + \frac{1}{2}\left(J(\widehat{\mathbf{q}}_m, \mathbf{p}) - J(\mathbf{q}, \mathbf{p})\right), \tag{37}$$

$$J^{(s)}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m) - J^{(s)}(\mathbf{p}, \mathbf{q})) = \frac{1}{2}\left(J(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m) - J(\mathbf{p}, \mathbf{q})\right) + \frac{1}{2}\left(J(\widehat{\mathbf{q}}_m, \widehat{\mathbf{p}}_n) - J(\mathbf{q}, \mathbf{p})\right). \tag{38}$$

**Asymptotic Normality**.

In addition to $V_1(\mathbf{p}, \mathbf{q})$ and $V_2(\mathbf{p}, \mathbf{q})$ defined in (26) and (27), denote

$$V_3(\mathbf{p}, \mathbf{q}) = \sum_{j\in D} q_j(1-q_j)(\phi_1^{(1)}(q_j, p_j))^2 - 2\sum_{(i,j)\in D^2,\ i\neq j} q_i q_j \phi_1^{(1)}(q_i, p_i)\phi_1^{(1)}(q_j, p_j), \tag{39}$$

$$V_4(\mathbf{p}, \mathbf{q}) = \sum_{j\in D} p_j(1-p_j)(\phi_2^{(1)}(q_j, p_j))^2 - 2\sum_{(i,j)\in D^2,\ i\neq j} p_i p_j \phi_2^{(1)}(q_i, p_i)\phi_2^{(1)}(q_j, p_j), \tag{40}$$

and finally

$$V_{1:4}(\mathbf{p}, \mathbf{q}) = \frac{1}{4}(V_1(\mathbf{p}, \mathbf{q}) + V_4(\mathbf{p}, \mathbf{q})) \ \text{ and } \ V_{2:3}(\mathbf{p}, \mathbf{q}) = \frac{1}{4}(V_2(\mathbf{p}, \mathbf{q}) + V_3(\mathbf{p}, \mathbf{q})).$$

We have

**Theorem 4.** *Under the same assumptions as in Theorem 1, the following hold.*

*(a) One sample : as $n \to +\infty$,*

$$\sqrt{n}\left(J^{(s)}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - J^{(s)}(\boldsymbol{p}, \boldsymbol{q})\right) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{1:4}(\boldsymbol{p}, \boldsymbol{q})), \tag{41}$$

$$\sqrt{n}\left(J^{(s)}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - J^{(s)}(\boldsymbol{p}, \boldsymbol{q})\right) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{2:3}(\boldsymbol{p}, \boldsymbol{q})). \tag{42}$$

*(b) Two samples : for $(n,m) \to (+\infty, +\infty)$ and $nm/(n+m) \to \gamma \in (0,1)$,*

$$\left(\frac{nm}{mV_{1:4}(\boldsymbol{p}, \boldsymbol{q}) + nV_{2:3}(\boldsymbol{p}, \boldsymbol{q})}\right)^{1/2}\left(J^{(s)}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - J^{(s)}(\boldsymbol{p}, \boldsymbol{q})\right) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0,1). \tag{43}$$

*Proof.* The proof follows directly by using equalities (36), (37), and (38) above and Theorem 2.

**Remark** : We have just seen that Theorems 3 and 4 are direct consequences of the main Theorems 1 and 2. The same will be true for the following Corollaries concerning the particular cases which proofs will be omitted.

## 4. Particular Cases

### 4.1. Rényi and Tsallis families

These two families are expressed through the summation

$$\mathcal{S}_\alpha(\mathbf{p},\mathbf{q}) = \sum_{j\in D} p_j^\alpha q_j^{1-\alpha}, \quad \alpha > 0, \quad \alpha \neq 1, \tag{44}$$

which is of the form of $\phi-$divergence measure with

$$\phi(x,y) = x^\alpha y^{1-\alpha}, \quad (x,y) \in \{(p_j,q_j),\ j \in D\}.$$

**A-(a)- The asymptotic behavior of the Tsallis divergence measure.**

Denote

$$A_{T,\alpha,1} = \frac{\alpha}{|\alpha-1|} \sum_{j\in D} (p_j/q_j)^{\alpha-1} \quad \text{and} \quad A_{T,\alpha,2} = \sum_{j\in D} (p_j/q_j)^{\alpha}.$$

We have

**Corollary 1.** *Under the same assumptions as in Theorem 1, and for any $\alpha > 0,\ \alpha \neq 1$, the following hold.*

*(a) One sample :*

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}_{T,\alpha}(\widehat{\boldsymbol{p}}_n,\boldsymbol{q}) - \mathcal{D}_{T,\alpha}(\boldsymbol{p},\boldsymbol{q})|}{a_n} \leq A_{T,\alpha,1} \quad a.s.,$$

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}_{T,\alpha}(\boldsymbol{p},\widehat{\boldsymbol{q}}_n) - \mathcal{D}_{T,\alpha}(\boldsymbol{p},\boldsymbol{q})|}{b_n} \leq A_{T,\alpha,2} \quad a.s.$$

*(b) Two samples :*

$$\limsup_{(n,m)\to(+\infty,+\infty)} \frac{|\mathcal{D}_{T,\alpha}(\widehat{\boldsymbol{p}}_n,\widehat{\boldsymbol{q}}_m) - \mathcal{D}_{T,\alpha}(\boldsymbol{p},\boldsymbol{q})|}{c_{n,m}} \leq A_{T,\alpha,1} + A_{T,\alpha,2} \quad a.s.$$

Denote

$$V_{T,\alpha,1}(\mathbf{p},\mathbf{q}) = \left(\frac{\alpha}{\alpha-1}\right)^2 \Bigg( \sum_{j\in D} p_j(1-p_j)(p_j/q_j)^{2\alpha-2} - 2 \sum_{(i,j)\in D^2,\, i\neq j} (p_i p_j)^{\alpha}(q_i q_j)^{\alpha-1} \Bigg)$$

$$\text{and} \quad V_{T,\alpha,2}(\mathbf{p},\mathbf{q}) = \sum_{j\in D} q_j(1-q_j)(p_j/q_j)^{2\alpha} - 2 \sum_{(i,j)\in D^2,\ i\neq j} (p_i p_j)^{\alpha}(q_i q_j)^{1-\alpha}.$$

We have

**Corollary 2.** *Under the same assumptions as in Theorem 1, and for any $\alpha > 0,\ \alpha \neq 1$, the following hold.*

*(a) One sample : as $n \to +\infty$,*

$$\sqrt{n}\,(\mathcal{D}_{T,\alpha}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - \mathcal{D}_{T,\alpha}(\boldsymbol{p}, \boldsymbol{q})) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{T,\alpha,1}(\boldsymbol{p}, \boldsymbol{q})),$$

$$\sqrt{n}\,(\mathcal{D}_{T,\alpha}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - \mathcal{D}_{T,\alpha}(\boldsymbol{p}, \boldsymbol{q})) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{T,\alpha,2}(\boldsymbol{p}, \boldsymbol{q})).$$

*(b) Two samples : for $(n,m) \to (+\infty, +\infty)$ and $nm/(n+m) \to \gamma \in (0,1)$,*

$$\left(\frac{nm}{nV_{T,\alpha,2}(\boldsymbol{p},\boldsymbol{q}) + mV_{T,\alpha,1}(\boldsymbol{p},\boldsymbol{q})}\right)^{1/2} (\mathcal{D}_{T,\alpha}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}_{T,\alpha}(\boldsymbol{p}, \boldsymbol{q})) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0,1).$$

As to the symmetrized form

$$\mathcal{D}^{(s)}_{T,\alpha}(\mathbf{p},\mathbf{q}) = \frac{\mathcal{D}_{T,\alpha}(\mathbf{p},\mathbf{q}) + \mathcal{D}_{T,\alpha}(\mathbf{q},\mathbf{p})}{2},$$

we need the supplementaries notations :

$$A_{T,\alpha,3} = \frac{\alpha}{|\alpha-1|} \sum_{j\in D} (q_j/p_j)^{\alpha-1}, \quad A_{T,\alpha,4} = \sum_{j\in D} (q_j/p_j)^{\alpha},$$

$$V_{T,\alpha,3}(\mathbf{p},\mathbf{q}) = \left(\frac{\alpha}{\alpha-1}\right)^2 \left( \sum_{j\in D} q_j(1-q_j)(q_j/p_j)^{2-2\alpha} - 2 \sum_{(i,j)\in D^2,\ i\neq j} (q_i q_j)^{2-\alpha}(p_i p_j)^{\alpha-1} \right),$$

$$\text{and} \quad V_{T,\alpha,4}(\mathbf{p},\mathbf{q}) = \sum_{j\in D} p_j(1-p_j)(q_j/p_j)^{2\alpha} - 2 \sum_{(i,j)\in D^2,\, i\neq j} (p_i p_j)^{1-\alpha}(q_i q_j)^{\alpha}.$$

We have

**Corollary 3.** *Under the same assumptions as in Theorem 1, and for any* $\alpha > 0$, $\alpha \neq 1$, *the following hold.*

*(a) One sample :*

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}^{(s)}_{T,\alpha}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - \mathcal{D}^{(s)}_{T,\alpha}(\boldsymbol{p}, \boldsymbol{q})|}{a_n} \leq (A_{T,\alpha,1} + A_{T,\alpha,4})/2 =: A^{(s)}_{T,\alpha,1} \quad a.s,$$

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}^{(s)}_{T,\alpha}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - \mathcal{D}^{(s)}_{T,\alpha}(\boldsymbol{p}, \boldsymbol{q})|}{b_n} \leq (A_{T,\alpha,2} + A_{T,\alpha,3})/2 =: A^{(s)}_{T,\alpha,2} \quad a.s.$$

*(b) Two samples :*

$$\limsup_{(n,m)\to(+\infty,+\infty)} \frac{|\mathcal{D}^{(s)}_{T,\alpha}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}^{(s)}_{T,\alpha}(\boldsymbol{p}, \boldsymbol{q})|}{c_{n,m}} \leq A^{(s)}_{T,\alpha,1} + A^{(s)}_{T,\alpha,2} \quad a.s.$$

Denote

$$\begin{aligned} V_{T,\alpha,1:4}(\mathbf{p}, \mathbf{q}) &= V_{T,\alpha,1}(\mathbf{p}, \mathbf{q}) + V_{T,\alpha,4}(\mathbf{p}, \mathbf{q}) \\ \text{and} \quad V_{T,\alpha,2:3}(\mathbf{p}, \mathbf{q}) &= V_{T,\alpha,2}(\mathbf{p}, \mathbf{q}) + V_{T,\alpha,3}(\mathbf{p}, \mathbf{q}). \end{aligned}$$

We have

**Corollary 4.** *Under the same assumptions as in Theorem 1, and for any* $\alpha > 0$, $\alpha \neq 1$, *the following hold.*

*(a) One sample : as* $n \to +\infty$,

$$\sqrt{n}\left(\mathcal{D}^{(s)}_{T,\alpha}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - \mathcal{D}^{(s)}_{T,\alpha}(\boldsymbol{p}, \boldsymbol{q})\right) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{T,\alpha,1:4}(\boldsymbol{p}, \boldsymbol{q})),$$

$$\sqrt{n}\left(\mathcal{D}^{(s)}_{T,\alpha}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - \mathcal{D}^{(s)}_{T,\alpha}(\boldsymbol{p}, \boldsymbol{q})\right) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{T,\alpha,2:3}(\boldsymbol{p}, \boldsymbol{q})).$$

*(b) Two samples : for* $(n, m) \to (+\infty, +\infty)$ *and* $nm/(n+m) \to \gamma \in (0, 1)$,

$$\left(\frac{nm}{mV_{T,\alpha,1:4}(\boldsymbol{p}, \boldsymbol{q}) + nV_{T,\alpha,2:3}(\boldsymbol{p}, \boldsymbol{q})}\right)^{1/2} \left(\mathcal{D}^{(s)}_{T,\alpha}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}^{(s)}_{T,\alpha}(\boldsymbol{p}, \boldsymbol{q})\right) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, 1).$$

### A-(b)- The asymptotic behavior of the Rényi-$\alpha$ divergence measure.

The treatment of the asymptotic behavior of the Rényi-$\alpha$, $\alpha > 0$, $\alpha \neq 1$ is obtained from Part **(A)-(a)** by expansions and by the application of the delta method.

We first remark that

$$\mathcal{D}_{R,\alpha}(\mathbf{p}, \mathbf{q}) = \frac{1}{\alpha - 1} \log\left(\mathcal{S}_\alpha(\mathbf{p}, \mathbf{q})\right).$$

**Corollary 5.** *Under the same assumptions as in Theorem 1, and for any $\alpha > 0$, $\alpha \neq 1$, the following hold.*

*(a) One sample :*

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}_{R,\alpha}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - \mathcal{D}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})|}{a_n} \leq \frac{A_{T,\alpha,1}}{\mathcal{S}_\alpha(\boldsymbol{p}, \boldsymbol{q})} =: A_{R,\alpha,1} \quad a.s.,$$

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}_{R,\alpha}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - \mathcal{D}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})|}{b_n} \leq \frac{A_{T,\alpha,2}}{\mathcal{S}_\alpha(\boldsymbol{p}, \boldsymbol{q})} =: A_{R,\alpha,2} \quad a.s.$$

*(b) Two samples :*

$$\limsup_{(n,m)\to(+\infty,+\infty)} \frac{|\mathcal{D}_{R,\alpha}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})|}{c_{n,m}} \leq A_{R,\alpha,1} + A_{R,\alpha,2} \quad a.s.$$

Denote

$$V_{R,\alpha,1}(\mathbf{p}, \mathbf{q}) = \frac{V_{T,\alpha,1}(\mathbf{p}, \mathbf{q})}{\mathcal{S}^2_\alpha(\mathbf{p}, \mathbf{q})} \text{ and } V_{R,\alpha,2}(\mathbf{p}, \mathbf{q}) = \frac{V_{T,\alpha,2}(\mathbf{p}, \mathbf{q})}{\mathcal{S}^2_\alpha(\mathbf{p}, \mathbf{q})}.$$

We have

**Corollary 6.** *Under the same assumptions as in Theorem 1, and for any $\alpha > 0$, $\alpha \neq 1$, the following hold*

*(a) One sample : as $n \to +\infty$,*

$$\sqrt{n}\Big(\mathcal{D}_{R,\alpha}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - \mathcal{D}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})\Big) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{R,\alpha,1}(\boldsymbol{p}, \boldsymbol{q})),$$

$$\sqrt{n}\Big(\mathcal{D}_{R,\alpha}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - \mathcal{D}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})\Big) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{R,\alpha,2}(\boldsymbol{p}, \boldsymbol{q})).$$

*(b) Two samples : for $(n, m) \to (+\infty, +\infty)$ and $nm/(n + m) \to \gamma \in (0, 1)$,*

$$\left(\frac{nm}{nV_{R,\alpha,2}(\boldsymbol{p}, \boldsymbol{q}) + mV_{R,\alpha,1}(\boldsymbol{p}, \boldsymbol{q})}\right)^{1/2}\Big(\mathcal{D}_{R,\alpha}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})\Big) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, 1).$$

As to the symmetrized form

$$\mathcal{D}^{(s)}_{R,\alpha}(\mathbf{p}, \mathbf{q}) = \frac{\mathcal{D}_{R,\alpha}(\mathbf{p}, \mathbf{q}) - \mathcal{D}_{R,\alpha}(\mathbf{q}, \mathbf{p})}{2},$$

we need the supplementary notations :

$$A_{R,\alpha,3} = \frac{A_{T,\alpha,3}}{\mathcal{S}_\alpha(\mathbf{p}, \mathbf{q})}, \quad A_{R,\alpha,4} = \frac{A_{T,\alpha,4}}{\mathcal{S}_\alpha(\mathbf{p}, \mathbf{q})}$$

$$V_{R,\alpha,3}(\mathbf{p}, \mathbf{q}) = \frac{V_{T,\alpha,3}(\mathbf{p}, \mathbf{q})}{\mathcal{S}^2_\alpha(\mathbf{p}, \mathbf{q})}, \quad \text{and } V_{R,\alpha,4}(\mathbf{p}, \mathbf{q}) = \frac{V_{T,\alpha,4}(\mathbf{p}, \mathbf{q})}{\mathcal{S}^2_\alpha(\mathbf{p}, \mathbf{q})}.$$

**Corollary 7.** *Under the same assumptions as in Theorem 1, and for any $\alpha > 0,\ \alpha \neq 1$, the following hold.*

*(a) One sample :*

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}^{(s)}_{R,\alpha}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - \mathcal{D}^{(s)}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})|}{a_n} \leq (A_{R,\alpha,1} + A_{R,\alpha,4})/2 =: A^{(s)}_{R,\alpha,1}, \quad a.s.,$$

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}^{(s)}_{R,\alpha}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - \mathcal{D}^{(s)}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})|}{a_n} \leq (A_{R,\alpha,2} + A_{R,\alpha,3})/2 =: A^{(s)}_{R,\alpha,2}, \quad a.s.$$

*(b) Two samples :*

$$\limsup_{(n,m)\to(+\infty,+\infty)} \frac{|\mathcal{D}^{(s)}_{R,\alpha}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}^{(s)}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})|}{c_{n,m}} \leq A^{(s)}_{R,\alpha,1} + A^{(s)}_{R,\alpha,2}, \quad a.s.$$

Denote

$$\begin{aligned} V_{R,\alpha,1:4}(\mathbf{p},\mathbf{q}) &= V_{R,\alpha,1}(\mathbf{p},\mathbf{q}) + V_{R,\alpha,4}(\mathbf{p},\mathbf{q}) \\ \text{and } V_{R,\alpha,2:3}(\mathbf{p},\mathbf{q}) &= V_{R,\alpha,2}(\mathbf{p},\mathbf{q}) + V_{R,\alpha,3}(\mathbf{p},\mathbf{q}). \end{aligned}$$

We have

**Corollary 8.** *Under the same assumptions as in Theorem 1, and for any $\alpha > 0,\ \alpha \neq 1$, the following hold.*

*(a) One sample : as $n \to +\infty$,*

$$\sqrt{n}\Big(\mathcal{D}^{(s)}_{R,\alpha}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - \mathcal{D}^{(s)}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})\Big) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{R,\alpha,1:4}(\boldsymbol{p}, \boldsymbol{q})),$$

$$\sqrt{n}\Big(\mathcal{D}^{(s)}_{R,\alpha}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - \mathcal{D}^{(s)}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})\Big) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{R,\alpha,2:3}(\boldsymbol{p}, \boldsymbol{q})).$$

*(b) Two samples : as $(n,m) \to (+\infty, +\infty)$ and $nm/(n+m) \to \gamma \in (0,1)$,*

$$\left(\frac{nm}{nV_{R,\alpha,2:3}(\boldsymbol{p},\boldsymbol{q}) + mV_{R,\alpha,1:4}(\boldsymbol{p},\boldsymbol{q})}\right)^{1/2} \Big(\mathcal{D}^{(s)}_{R,\alpha}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}^{(s)}_{R,\alpha}(\boldsymbol{p}, \boldsymbol{q})\Big) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0,1).$$

### B - Kulback-Leibler Measure

Here we have

$$\mathcal{D}_{KL}(\mathbf{p},\mathbf{q}) = \sum_{j\in D} \phi(p_j, q_j),$$

where

$$\phi(x,y) = x\log(x/y), \quad (x,y) \in \{(p_j,q_j),\ j \in D\}.$$

We have

**Corollary 9.** *Under the same assumptions as in Theorem 1, the following hold.*

*(a) One sample :*

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}_{KL}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - \mathcal{D}_{KL}(\boldsymbol{p}, \boldsymbol{q})|}{a_n} \leq \sum_{j\in D} |1+\log(p_j/q_j)| =: A_{KL,1}(\boldsymbol{p}, \boldsymbol{q}), \quad a.s.,$$

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}_{KL}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - \mathcal{D}_{KL}(\boldsymbol{p}, \boldsymbol{q})|}{b_n} \leq \sum_{j\in D} (p_j/q_j) =: A_{KL,2}(\boldsymbol{p}, \boldsymbol{q}), \quad a.s.$$

*(b) Two samples :*

$$\limsup_{(n,m)\to(+\infty,+\infty)} \frac{|\mathcal{D}_{KL}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}_{KL}(\boldsymbol{p}, \boldsymbol{q})|}{c_{n,m}} \leq A_{KL,1}(\boldsymbol{p}, \boldsymbol{q}) + A_{KL,2}(\boldsymbol{p}, \boldsymbol{q}), \quad a.s.$$

Denote

$$\begin{aligned} V_{KL,1}(\mathbf{p},\mathbf{q}) &= \sum_{j\in D} p_j(1-p_j)\left(1+\log(p_j/q_j)\right)^2 \\ &\quad - 2 \sum_{(i,j)\in D^2,\ i\neq j} p_i p_j (1+\log(p_i/q_i))(1+\log(p_j/q_j)) \\ \text{and} \quad V_{KL,2}(\mathbf{p},\mathbf{q}) &= \sum_{j\in D} q_j(1-q_j)(p_j/q_j)^2 - 2 \sum_{(i,j)\in D^2,\ i\neq j} p_i p_j. \end{aligned}$$

We have

**Corollary 10.** *Under the same assumptions as in Theorem 1, the following hold.*

*(a) One sample : as $n\to+\infty$*

$$\sqrt{n}\left(\mathcal{D}_{KL}(\widehat{\boldsymbol{p}}_n, \boldsymbol{q}) - \mathcal{D}_{KL}(\boldsymbol{p}, \boldsymbol{q})\right) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{KL,1}(\boldsymbol{p}, \boldsymbol{q})),$$

$$\sqrt{n}\left(\mathcal{D}_{KL}(\boldsymbol{p}, \widehat{\boldsymbol{q}}_n) - \mathcal{D}_{KL}(\boldsymbol{p}, \boldsymbol{q})\right) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{KL,2}(\boldsymbol{p}, \boldsymbol{q})).$$

*(b) Two samples : as $(n,m)\to(+\infty,+\infty)$ and $nm/(n+m)\to\gamma\in(0,1)$,*

$$\left(\frac{nm}{nV_{KL,2}(\boldsymbol{p}, \boldsymbol{q}) + mV_{KL,1}(\boldsymbol{p}, \boldsymbol{q})}\right)^{1/2} \left(\mathcal{D}_{KL}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}_{KL}(\boldsymbol{p}, \boldsymbol{q})\right) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0,1).$$

As to the symmetrized form

$$\mathcal{D}_{KL}^{(s)}(\mathbf{p},\mathbf{q}) = \frac{\mathcal{D}_{KL}(\mathbf{p},\mathbf{q}) + \mathcal{D}_{KL}(\mathbf{q},\mathbf{p})}{2},$$

we need the supplementary notations :

$$\begin{aligned}
A_{KL,3}(\mathbf{p},\mathbf{q}) &= \sum_{j\in D} |1+\log(q_j/p_j)|\,, \quad A_{KL,4}(\mathbf{p},\mathbf{q}) = \sum_{j\in D} q_j/p_j \\
V_{KL,3}(\mathbf{p},\mathbf{q}) &= \sum_{j\in D} q_j(1-q_j)(1+\log(q_j/p_j))^2 \\
&\quad -2 \sum_{(i,j)\in D^2,\ i\neq j} q_i q_j (1+\log(q_i/p_i))(1+\log(q_j/p_j)), \\
\text{and} \ \ V_{KL,4}(\mathbf{p},\mathbf{q}) &= \sum_{j\in D} p_j(1-p_j)(q_j/p_j)^2 - 2\sum_{(i,j)\in D^2,\ i\neq j} q_i q_j.
\end{aligned}$$

We have

**Corollary 11.** *Under the same assumptions as in Theorem 1, the following hold.*

*(a) One sample :*

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}^{(s)}_{KL}(\widehat{\boldsymbol{p}}_n,\boldsymbol{q}) - \mathcal{D}^{(s)}_{KL}(\boldsymbol{p},\boldsymbol{q})|}{a_n} \leq (A_{KL,1}(\boldsymbol{p},\boldsymbol{q}) + A_{KL,4}(\boldsymbol{p},\boldsymbol{q}))/2 =: A^{(s)}_{KL,1}(\boldsymbol{p},\boldsymbol{q}), \quad a.s.,$$

$$\limsup_{n\to+\infty} \frac{|\mathcal{D}^{(s)}_{KL}(\boldsymbol{p},\widehat{\boldsymbol{q}}_n) - \mathcal{D}^{(s)}_{KL}(\boldsymbol{p},\boldsymbol{q})|}{b_n} \leq (A_{KL,2}(\boldsymbol{p},\boldsymbol{q}) + A_{KL,3}(\boldsymbol{p},\boldsymbol{q}))/2 =: A^{(s)}_{KL,2}(\boldsymbol{p},\boldsymbol{q}), \quad a.s.$$

*(b) Two samples :*

$$\limsup_{(n,m)\to(+\infty,+\infty)} \frac{|\mathcal{D}^{(s)}_{KL}(\widehat{\boldsymbol{p}}_n,\widehat{\boldsymbol{q}}_m) - \mathcal{D}^{(s)}_{KL}(\boldsymbol{p},\boldsymbol{q})|}{c_{n,m}} \leq A^{(s)}_{KL,1}(\boldsymbol{p},\boldsymbol{q}) + A^{(s)}_{KL,2}(\boldsymbol{p},\boldsymbol{q}), \quad a.s.$$

Denote

$$\begin{aligned}
V_{KL,1:4}(\mathbf{p},\mathbf{q}) &= V_{KL,1}(\mathbf{p},\mathbf{q}) + V_{KL,4}(\mathbf{p},\mathbf{q}), \\
V_{KL,2:3}(\mathbf{p},\mathbf{q}) &= V_{KL,2}(\mathbf{p},\mathbf{q}) + V_{KL,3}(\mathbf{p},\mathbf{q}).
\end{aligned}$$

We have

**Corollary 12.** *Under the same assumptions as in Theorem 1, the following hold.*

*(a) One sample : as $n \to +\infty$,*

$$\sqrt{n}\Big(\mathcal{D}^{(s)}_{KL}(\widehat{\boldsymbol{p}}_n,\boldsymbol{q}) - \mathcal{D}^{(s)}_{KL}(\boldsymbol{p},\boldsymbol{q})\Big) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{KL,1:4}(\boldsymbol{p},\boldsymbol{q})),$$

$$\sqrt{n}\Big(\mathcal{D}^{(s)}_{KL}(\boldsymbol{p},\widehat{\boldsymbol{q}}_n) - \mathcal{D}^{(s)}_{KL}(\boldsymbol{p},\boldsymbol{q})\Big) \stackrel{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0, V_{KL,2:3}(\boldsymbol{p},\boldsymbol{q})).$$

*(b) Two samples : for* $(n, m) \to (+\infty, +\infty)$ *and* $nm/(n+m) \to \gamma \in (0,1)$,

$$\left(\frac{nm}{mV_{KL,1:4}(\boldsymbol{p},\boldsymbol{q}) + nV_{KL,2:3}(\boldsymbol{p},\boldsymbol{q})}\right)^{1/2}\left(\mathcal{D}^{(s)}_{KL}(\widehat{\boldsymbol{p}}_n, \widehat{\boldsymbol{q}}_m) - \mathcal{D}^{(s)}_{KL}(\boldsymbol{p},\boldsymbol{q})\right) \overset{\mathcal{D}}{\rightsquigarrow} \mathcal{N}(0,1).$$

## 5. Simulations

To assess the performance of ours estimators, we present a simulation study on a finite sample. For simplicity, in our experiments we consider tree outcomes for the randoms variables $X$ and $Y$, $c_1, c_2, c_3$ with respectives *p.m.f.'s* $p_1,\ p_2,\ p_3$ and $q_1,\ q_2,\ q_3$.
Our aim is to compare the performance of the divergences measures estimators as well as their symmetrized forms with one or two samples when sample sizes increase.
Suppose that $p_1 = 0.4,\ p_2 = 0.25,\ p_3 = 0.35$ and $q_1 = 0.27,\ q_2 = 0.32,\ q_3 = 0.41$.
We set first $\alpha = 0.99$ since

$$\lim_{\alpha \to 1} \mathcal{D}_{T,\alpha}(\mathbf{p},\mathbf{q}) = \lim_{\alpha \to 1} \mathcal{D}_{R,\alpha}(\mathbf{p},\mathbf{q}) = \mathcal{D}_{KL}(\mathbf{p},\mathbf{q})$$

and second $\alpha = 0.5$ since

$$\mathcal{D}_{T,0.5}(\mathbf{p},\mathbf{q}) = \mathcal{D}_{T,0.5}(\mathbf{q},\mathbf{p}) \quad \text{and} \quad \mathcal{D}_{R,0.5}(\mathbf{p},\mathbf{q}) = \mathcal{D}_{R,0.5}(\mathbf{q},\mathbf{p}).$$

True and the symmetrized form of our interest divergence measures are then

$$\begin{aligned}\mathcal{D}_{T,0.99}(\mathbf{p},\mathbf{q}) &\approx 0.03969,\ \ \mathcal{D}^{(s)}_{T,0.99}(\mathbf{p},\mathbf{q}) \approx 0.03854, \quad \mathcal{D}_{R,0.99}(\mathbf{p},\mathbf{q}) \approx 0.03970,\\ \mathcal{D}^{(s)}_{R,0.99}(\mathbf{p},\mathbf{q}) &\approx 0.03854, \quad \mathcal{D}_{KL}(\mathbf{p},\mathbf{q}) \approx 0.04012, \ \text{ and } \ \mathcal{D}^{(s)}_{KL}(\mathbf{p},\mathbf{q}) \approx 0.03893.\end{aligned}$$

and

$$\mathcal{D}_{R,0.5}(\mathbf{p},\mathbf{q}) \approx 0.01951452.$$

In each FIGURES 1, 2, 3, 4, 5, 6, 7, and 8, left panels represent the plots of divergence measure estimator, built from sample sizes of $n = 100, 200, \cdots, 30000$, and the true divergence measure (represented by horizontal black line). The middle panels show the histograms of the data and where the red line represents the plots of the theoretical normal distribution calculated from the same mean and the same standard deviation of the data. The right panels concern the Q-Q plot of the data which display the observed values against normally distributed data (represented by the red line). We see that the underlying distribution of the data is normal since the points fall along a straight line.

## 6. Conclusion

This paper joins a growing body of literature on estimating divergence measures in the discrete case and on finite sets. We adopted the plug-in method and we derived almost sure rates of convergence and asymptotic normality of the most common divergence measures in one sample, two samples as well as symmetrical form of divergence measures, all this, by means of the functional $\phi-$divergence measure.

Figure 1 displays the performances of the Stallis divergence estimators of order $\alpha = 0.99$, $\mathcal{D}_{T,\alpha}(\widehat{\mathbf{p}}_n, \mathbf{q})$, $\mathcal{D}_{T,\alpha}(\mathbf{p}, \widehat{\mathbf{q}}_m)$, and $\mathcal{D}_{T,\alpha}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$.

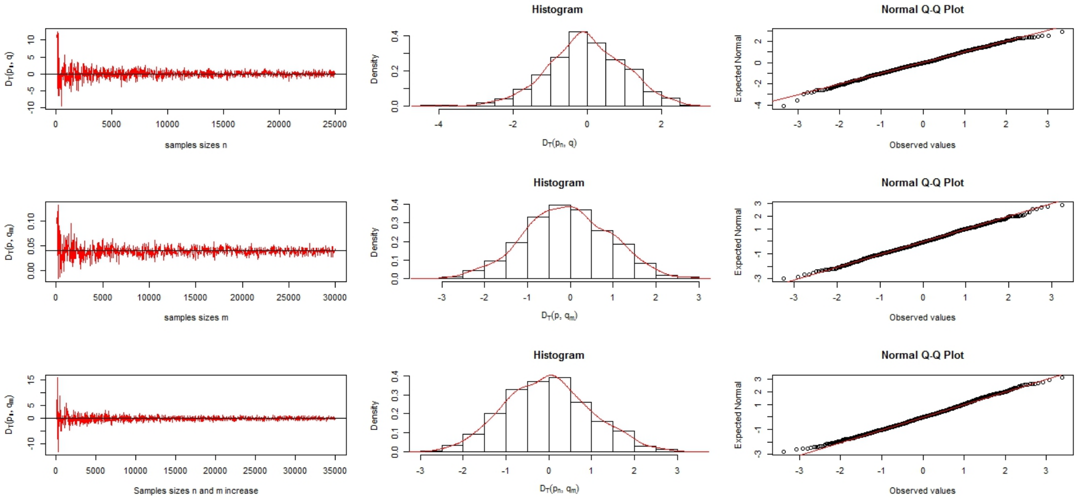


Figure 1: Plots when samples sizes increase, histograms and normal Q-Q plots of $\mathcal{D}_{T,\alpha}(p_n, q)$, $\mathcal{D}_{T,\alpha}(p, q_m)$, and $\mathcal{D}_{T,\alpha}(p_n, q_m)$ ($\alpha = 0.99$) versus $\mathcal{N}(0, 1)$.

Figure 2 concerns the performances of the symmetrized form of Stallis divergence estimators $\mathcal{D}^{(s)}_{T,\alpha}(\widehat{\mathbf{p}}_n, \mathbf{q})$, $\mathcal{D}^{(s)}_{T,\alpha}(\mathbf{p}, \widehat{\mathbf{q}}_m)$ and $\mathcal{D}^{(s)}_{T,\alpha}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$.

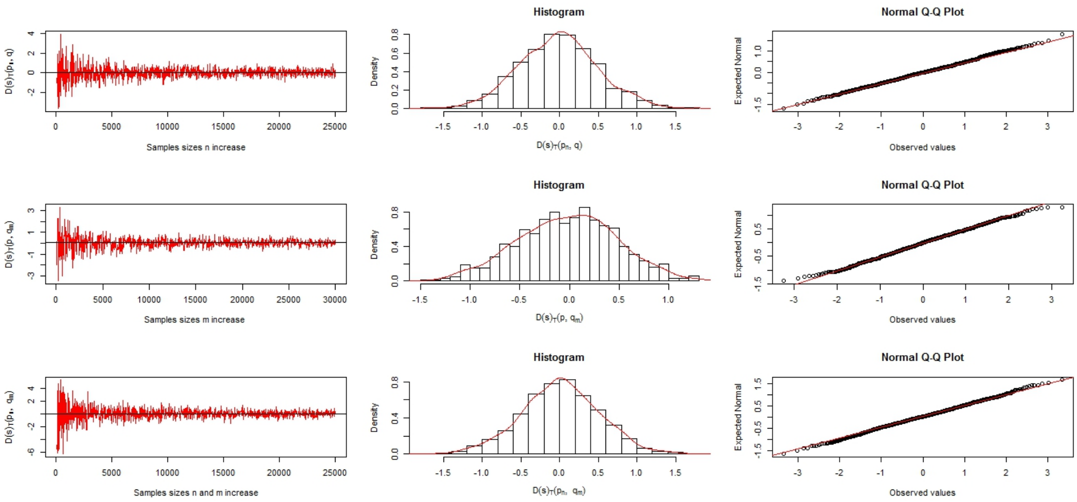


Figure 2: Plots when samples sizes increase, histograms and normal Q-Q plots of $\mathcal{D}^{(s)}_{T,\alpha}(p_n, q)$, $\mathcal{D}^{(s)}_{T,\alpha}(p, q_m)$, and $\mathcal{D}^{(s)}_{T,\alpha}(p_n, q_m)$ ($\alpha = 0.99$) versus $\mathcal{N}(0, 1)$.

Figure 3 displays the performances of the Rényi divergence estimators of order $\alpha = 0.99$, $\mathcal{D}_{R,\alpha}(\widehat{\mathbf{p}}_n, \mathbf{q})$, $\mathcal{D}_{R,\alpha}(\mathbf{p}, \widehat{\mathbf{q}}_m)$, and $\mathcal{D}_{R,\alpha}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$.

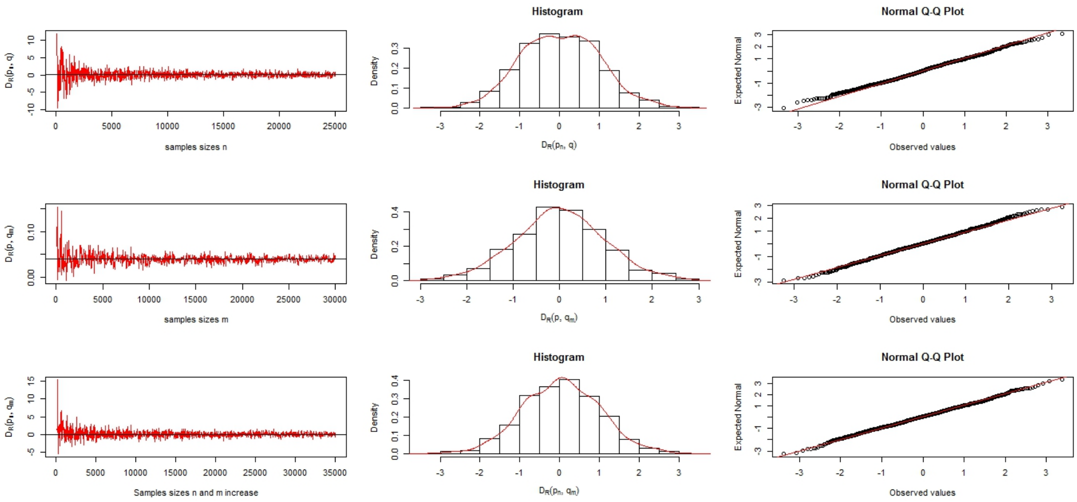


Figure 3: Plots when samples sizes increase, histograms and normal Q-Q plots of $\mathcal{D}_{R,\alpha}(p_n, q)$, $\mathcal{D}_{R,\alpha}(p, q_m)$, and $\mathcal{D}_{R,\alpha}(p_n, q_m)$ ($\alpha = 0.99$) versus $\mathcal{N}(0, 1)$.

Figure 4 concerns the performances of the symmetrized form of Rényi divergence estimators $\mathcal{D}^{(s)}_{R,\alpha}(\widehat{\mathbf{p}}_n, \mathbf{q})$, $\mathcal{D}^{(s)}_{R,\alpha}(\mathbf{p}, \widehat{\mathbf{q}}_m)$, and $\mathcal{D}^{(s)}_{R,\alpha}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$.

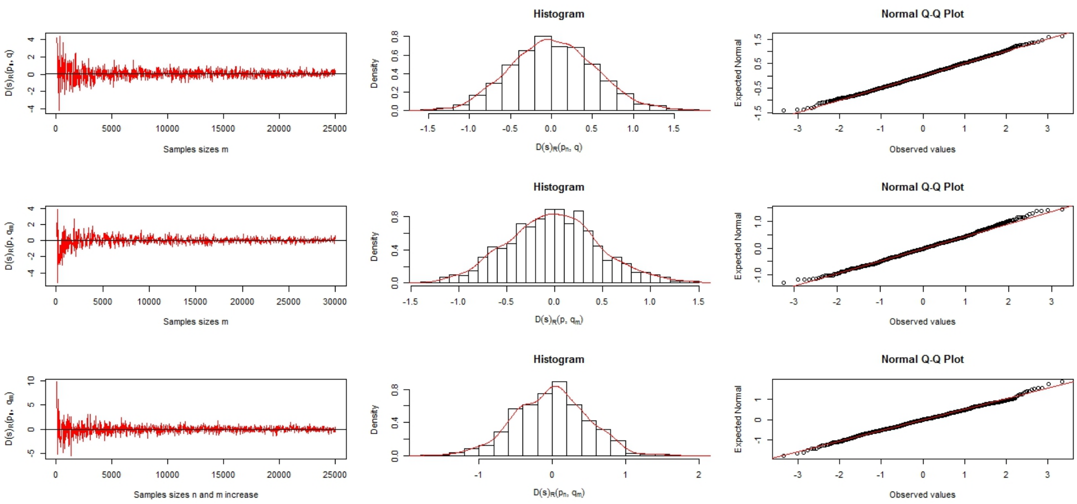


Figure 4: Plots when samples sizes increase, histograms and normal Q-Q plots of $\mathcal{D}^{(s)}_{R,\alpha}(p_n, q)$, $\mathcal{D}^{(s)}_{R,\alpha}(p, q_m)$, and $\mathcal{D}^{(s)}_{R,\alpha}(p_n, q_m)$ ($\alpha = 0.99$) versus $\mathcal{N}(0, 1)$.

Figure 5 displays the performances of the Rényi divergence estimators of order $\alpha = 0.5$, $\mathcal{D}_{R,\alpha}(\widehat{\mathbf{p}}_n, \mathbf{q})$, $\mathcal{D}_{R,\alpha}(\mathbf{p}, \widehat{\mathbf{q}}_m)$, and $\mathcal{D}_{R,\alpha}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$.

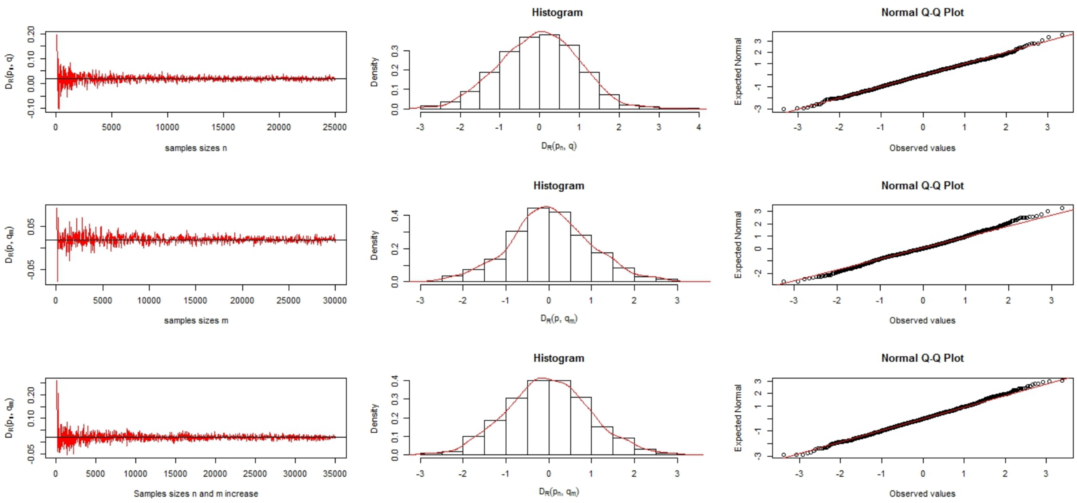


Figure 5: Plots when samples sizes increase, histogram and normal Q-Q plots of $\mathcal{D}_{R,\alpha}(p_n, q)$, $\mathcal{D}_{R,\alpha}(p, q_m)$, and $\mathcal{D}_{R,\alpha}(p_n, q_m)$ $(\alpha = 0.5)$ versus $\mathcal{N}(0, 1)$.

Figure 6 concerns the performances of the symmetrized form of Rényi divergence estimators $\mathcal{D}^{(s)}_{R,\alpha}(\widehat{\mathbf{p}}_n, \mathbf{q})$, $\mathcal{D}^{(s)}_{R,\alpha}(\mathbf{p}, \widehat{\mathbf{q}}_m)$, and $\mathcal{D}^{(s)}_{R,\alpha}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$.

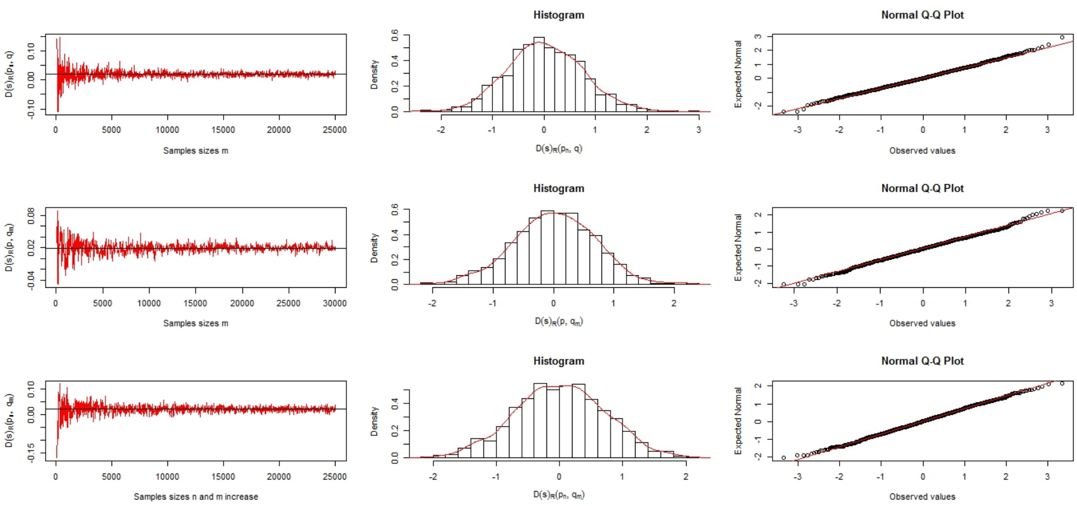


Figure 6: Plots when samples sizes increase, histograms and normal Q-Q plots of $\mathcal{D}^{(s)}_{R,\alpha}(p_n, q)$, $\mathcal{D}^{(s)}_{R,\alpha}(p, q_m)$, and $\mathcal{D}^{(s)}_{R,\alpha}(p_n, q_m)$ $(\alpha = 0.5)$ versus $\mathcal{N}(0, 1)$.

Figure 7 displays the performances of the Kullback-Leibler estimators, $\mathcal{D}_{KL}(\widehat{\mathbf{p}}_n, \mathbf{q})$, $\mathcal{D}_{KL}(\mathbf{p}, \widehat{\mathbf{q}}_m)$, and $\mathcal{D}_{KL}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$.

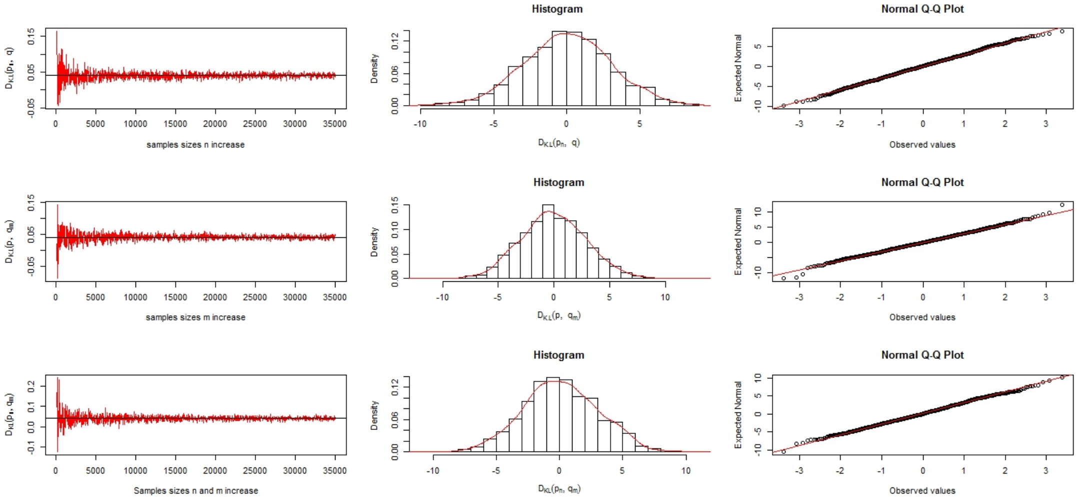


Figure 7: Plots when samples sizes increase, histograms and normal Q-Q plots of $\mathcal{D}_{KL}(p_n, q)$, $\mathcal{D}_{KL}(p, q_m)$, and $\mathcal{D}_{KL}(p_n, q_m)$ versus $\mathcal{N}(0, 1)$.

Figure 8 concerns the performances of the symmetrized form of Kullback-Leibler estimators, $\mathcal{D}^{(s)}_{KL}(\widehat{\mathbf{p}}_n, \mathbf{q})$, $\mathcal{D}^{(s)}_{KL}(\mathbf{p}, \widehat{\mathbf{q}}_m)$, and $\mathcal{D}^{(s)}_{KL}(\widehat{\mathbf{p}}_n, \widehat{\mathbf{q}}_m)$.

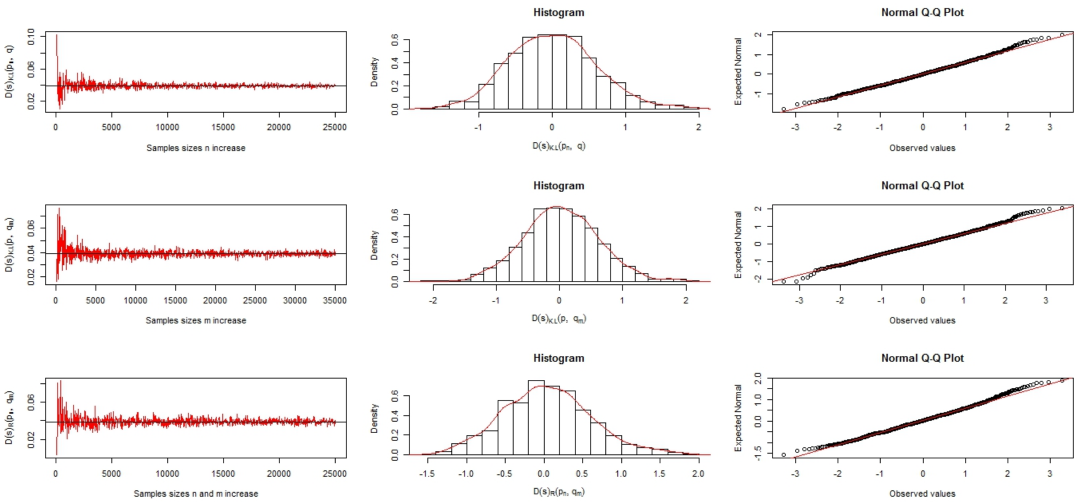


Figure 8: Plots when samples sizes increase, histograms and normal Q-Q plots of $\mathcal{D}^{(s)}_{KL}(p_n, q)$, $\mathcal{D}^{(s)}_{KL}(p, q_m)$, and $\mathcal{D}^{(s)}_{KL}(p_n, q_m)$ versus $\mathcal{N}(0, 1)$.

## References


[1] Lô G.S. Bâ A.D. and Gaston Berger University Bâ, D. Divergence measures estimation and its asymptotic normality theory using wavelets empirical processes i. *Journal of Statistical Theory and Applications*, 17:158–171, 2018.

[2] P. K. Bhattacharya. Efficient estimation of a shift parameter from grouped data. *The Annals of Mathematical Statistics*, 38:1770–1787, 1967.

[3] Andrzej Cichocki and Shun-ichi Amari. Families of alpha- beta- and gamma- divergences: Flexible and robust measures of similarities. *Entropy*, 12:1532–1568, 2010.

[4] Hamza Dhaker. Kernel-type estimators of divergence measures and its strong uniform consistency. *American Journal of Theoretical and Applied Statistics*, 5:13, 2016.

[5] Atif Ahmet Evren. Some applications of kullback-leibler and jeffrey's divergences in multinomial population. *Journal of Selcuk University Natural and Applied Science*, 1, 2012.

[6] Topsoe F. Some inequalities for information divergences and related measures of discrimination. *IEEE Transactions on Information Theory*, 46, 2000.

[7] K. Fukunaga and R.R. Hayes. The reduced parzen classifier. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:423–425, 1989.

[8] Peter Hall. On kullback-leibler loss and density estimation. *The Annals of Statistics*, 15:1491–1519, 1987.

[9] Trevor Hastie and Robert Tibshirani. Classification by pairwise coupling. *The Annals of Statistics*, 26:451–471, 1998.

[10] Akshay Krishnamurthy, Kirthevasan Kandasamy, Barnabas Poczos, and Larry Wasserman. Nonparametric estimation of renyi divergence and friends. *PMLR*, pages 919–927, 2014.

[11] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.

[12] David Kllberg and Oleg Seleznjev. Estimation of entropy-type integral functionals. *Communications in Statistics - Theory and Methods*, 45:887–905, 2016.

[13] Ce Liu, Ce Liu, and Hueng-Yeung Shum. Kullback-leibler boosting, 2015.

[14] Gane Samb Lo. *Weak Convergence (IA). Sequences of Random Vectors.* Statistics and Probability African Society, 2016.

[15] David Mackay. Information theory, inference, and learning algorithms, 1995.

[16] Kevin R. Moon and Alfred O. Hero. Ensemble estimation of multivariate f-divergence. *2014 IEEE International Symposium on Information Theory*, 2014.

[17] Pedro Moreno, Purdy Ho, and Nuno Vasconcelos. A kullback-leibler divergence based kernel for svm classification in multimedia applications, 2004.

[18] Timo Ojala, Matti Pietikinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29:51–59, 1996.

[19] Barnabas Poczos. On the estimation of alpha-divergences, 2015.

[20] Shashank Singh and Barnabas Poczos. Generalized exponential concentration inequality for renyi divergence estimation. *PMLR*, pages 333–341, 2014.

[21] Kumar Sricharan, Dennis Wei, and Alfred O. Hero. Ensemble estimators for multivariate entropy estimation. *IEEE Transactions on Information Theory*, 59:4374–4388, 2013.

[22] Georges Valiron. *Cours d'analyse mathématique*. Masson et Cie, 1966.