Journal of Statistical Research *https://doi.org/10.3329/jsr.v56i1.63943*
2022, Vol. 56, No. 1, pp. 1-10 *ISSN 0256 - 422 X*

DENSITY DIVERGENCE AND DENSITY CONVERGENCE

MALAY GHOSH

Department of Statistics, University of Florida, Gainesville, Florida, USA
Email: ghoshm@stat.ufl.edu

PARTHA SARKAR*

Department of Statistics, University of Florida, Gainesville, Florida, USA
Email: sarkarpartha@ufl.edu

SUMMARY

Divergence between two distributions has been of statistical interest for more than a century, beginning with Karl Pearson with his famous chisquare test. The paper revisits some of the well-known density divergence measures, and studies their interrelationship. In addition, it is demonstrated how Scheffe's pointwise density convergence implies convergence of distributions, based on different divergence measures.

Keywords and phrases: α-divergence, Chisquare, Hellinger, Kullback-Leibler, Scheffe's density convergence.

1 Introduction

Measuring the divergence or distance between two distributions has a long statistical history. Its importance was felt many years ago, especially for testing goodness of fit for a given set of data to an assumed model. Beginning with Karl Pearson's classical chisquare goodness of fit test, there exists a multitude of other tests addressing similar problems. Two of the more popular tests, namely the Kolmogorov-Smirnov test and the Cramer-von Mises test measure the closeness of the empirical distribution function, and an assumed null distribution.

While the above two tests and similar other tests are based on measuring the divergence between two distribution functions, of equal importance is to measure the divergence between two density functions. This is very much reflected in the Total Variation (TV) distance between two distributions, which simplifies into an L_1 distance between densities. But there are other equally popular measures of divergence, in particular, the Kullback-Leibler (KL) (Kullback and Leibler, 1951) and the Bhattacharyya-Hellinger (H) (Hellinger, 1909; Bhattacharyya, 1946) measures. However, the last two measures are special cases of a general α-divergence measure, originally introduced by Rényi (1961), followed up later by several others, notably by Amari (1982) and Cressie and Read (1984). Another important example belonging to the α-divergence class is the chisquare divergence.

* Corresponding author

All these divergence measures belong to a more general family, known as f-divergence, studied by Csiszar (1967) and others.

One objective of this note is to set up a relationship between different measures of divergence between two densities using certain elementary inequalities. Many of the results as presented here are well-known, but, at least to our knowledge, not at the level of generality as given here. Also, we have included the proofs, mostly for the sake of completeness, while giving reference to some of the readily available sources. A second objective is to show that the pointwise density convergence as given in Scheffé (1947) is enough to justify density convergence under most of the divergence measures considered, with the exception of the KL divergence. Also, we will state a very strong and useful result related to f-divergence from Gilardoni (2006) which may and possibly has led to a common misconception that TV is the weakest form of convergence between densities. However, in this paper we will explain why this is not always true. An important consequence of our results is that for verifying the convergence of a sequence of pdf's to another pdf either under TV or any α-divergence measure $(0 < \alpha < 1)$, it suffices to show convergence under the H divergence. Finally, we provide two important examples involving the t and F distributions, where pointwise convergence of a sequence of densities to another density yields KL convergence. If the latter holds, so does convergence under the TV or any α-divergence measure $(0 < \alpha < 1)$.

Section 2 of this paper introduces the different divergence measures, and points out the relationship among these. Section 3 shows how Scheffe's pointwise density convergence is enough to guarantee convergence according to the different divergence measures that we have considered.

2 Divergence Measures and Their Relationship

We begin with the Total Variation (TV) divergence between two densities and prove the following result. The result is available in Billingsley (1968, p. 224) who outlined the proof. Here we provide more details.

Theorem 1. *Let p and q be two densities, each absolutely continuous with respect to some σ-finite measure μ. Also, let $P(A) = \int_A p d\mu$ and $Q(A) = \int_A q d\mu$ for every Borel set A. Then*

$$TV(p,q) = \sup_A |P(A) - Q(A)| = (1/2) \int |p - q| d\mu.$$

Proof. First, for every Borel set A, using $\int p d\mu = \int q d\mu = 1$,

$$|P(A) - Q(A)| = |\int_A [p(x) - q(x)] d\mu| = |-\int_{A^c} [p(x) - q(x)] d\mu| = |\int_{A^c} [p(x) - q(x)] d\mu|. \quad (2.1)$$

Hence,

$$
\begin{aligned}
2|P(A) - Q(A)| &= |\int_A [p(x) - q(x)]d\mu| + |\int_{A^c} [p(x) - q(x)]d\mu| \\
&\leq \int_A |p(x) - q(x)|d\mu + \int_{A^c} |p(x) - q(x)|d\mu \\
&= \int |p(x) - q(x)|d\mu. \qquad (2.2)
\end{aligned}
$$

This leads to $|P(A) - Q(A)| \leq (1/2) \int |p(x) - q(x)|d\mu$ for every Borel set A. Since the right hand side in (2.2) does not depend on A, one gets the inequality

$$\sup_A |P(A) - Q(A)| \leq (1/2) \int |p - q|d\mu. \qquad (2.3)$$

Conversely, let $A = \{x : p(x) \geq q(x)\}$. Then, using once again $\int pd\mu = \int qd\mu = 1$,

$$
\begin{aligned}
\int |p(x) - q(x)|d\mu &= \int_A [p(x) - q(x)]d\mu + \int_{A^c} [q(x) - p(x)]d\mu \\
&= 2 \int_A [p(x) - q(x)]d\mu = 2[P(A) - Q(A)] \\
&\leq 2 \sup_A |P(A) - Q(A)|. \qquad (2.4)
\end{aligned}
$$

Combine (2.3) and (2.4) to get the result.

Next we introduce α-divergence by Rényi (1961) between two distributions. This is given by

$$R_\alpha(p, q) = \frac{1}{\alpha - 1} \log[\int p^\alpha q^{1-\alpha} d\mu]]. \qquad (2.5)$$

Similarly, one can define another class of D_α divergence (or information divergence of type $(1 - \alpha)$) as

$$D_\alpha(p, q) = \frac{1}{\alpha(1 - \alpha)} [1 - \int p^\alpha q^{1-\alpha} d\mu]. \qquad (2.6)$$

It is straight forward to see

$$R_\alpha(p, q) = \frac{1}{\alpha - 1} \log[1 - \alpha(1 - \alpha)D_\alpha(p, q)], \qquad (2.7)$$

which is increasing in $D_\alpha(p, q)$. Also, as a special case of the latter, we have

$$D_{1/2}(p, q) = 4[1 - \int p^{1/2} q^{1/2} d\mu] = 2 \int (p^{1/2} - q^{1/2})^2 d\mu = 2H^2(p, q),$$

where $H(p, q) = \{\int (p^{1/2} - q^{1/2})^2 d\mu\}^{1/2}$ is the Bhattacharyya-Hellinger distance between two densities p and q. Further, using L'hospital's rule,

$$\lim_{\alpha \to 1} D_\alpha(p, q) = \int \log(p/q) p d\mu = \text{KL}(p, q);$$

$$\lim_{\alpha\to 0} D_\alpha(p,q) = \int \log(q/p) q d\mu = \mathrm{KL}(q,p),$$

the two KL divergence measures, thus arising as limiting cases of D_α divergence. One can go with $\alpha < 0$ as well. We point out two interesting examples $2D_{-1}(p,q) = \int (p-q)^2/p d\mu$, which is known as the Neyman χ^2 divergence, and $2D_2(p,q) = \int (p-q)^2/q d\mu$, which is known as Pearson χ^2 divergence. From the identity given in (2.7), one immediately gets for all $0 < \alpha < 1$,

$$R_\alpha(p,q) \to 0 \equiv D_\alpha(p,q) \to 0. \tag{2.8}$$

We can also see all of these convergences (except Renyi's divergence) as special cases of f-divergences introduced by Csiszar (1967). The general definition of f-divergence with respect to a convex fuction $f : [0,\infty) \to (-\infty,\infty)$ such that $f(x)$ is finite for all $x > 0$, $f(1) = 0$, and $f(0) = \lim_{x\to 0^+} f(x)$ given by

$$D_f(p,q) = \int f\left(\frac{p}{q}\right) q d\mu. \tag{2.9}$$

It is easy to show if we choose $f(x)$ as $(1/2)|x-1|$, $(x^\alpha - \alpha x)/[\alpha(\alpha-1)]$, $x\log x$, $-\log(x)$, $(x-1)^2$, $(1/x)-1$, then $D_f(p,q)$ will be equal to TV distance, $D_\alpha(p,q)$, $KL(p,q)$, $KL(q,p)$, Pearson χ^2 and Neyman χ^2 divergence respectively. Gilardoni (2006) proved the following result.

Lemma 2.1. *Suppose that the convex function f is differentiable up to order 3 at $x = 1$ with $f''(1) > 0$, then $D_f(p,q) \geq 2f''(1)(TV(p,q))^2$ and the constant $2f''(1)$ is best possible.*

Now, for all of our previous choices of f, one gets $f''(1) = 1$. This implies that $D_\alpha(p,q)$, $KL(p,q)$, $KL(q,p)$, Pearson χ^2 and Neyman χ^2 divergences are bounded below by $2(TV(p,q))^2$. This may lead to a misconception that $TV(p,q)$ is weakest among all distances, while one talks about convergence of densities. However, in the next few paragraphs we will show that that may not always be a valid conclusion.

We now establish an inequality between $TV(p,q)$ and $D_\alpha(p,q)$.

Theorem 2. *$\alpha(1-\alpha)D_\alpha(p,q) \leq TV(p,q)$, $0 \leq \alpha \leq 1$.*

Proof. For $0 \leq \alpha \leq 1$,

$$\begin{aligned}
\alpha(1-\alpha)D_\alpha(p,q) &= 1 - \int p^\alpha q^{1-\alpha} d\mu \\
&\leq 1 - \int (\min(p,q))^\alpha (\min(p,q))^{1-\alpha} d\mu = 1 - \int \min(p,q) d\mu \\
&= (1/2)(2 - 2\int \min(p,q) d\mu) = (1/2)\int [p + q - 2\min(p,q)] d\mu \\
&= (1/2)\int |p-q| d\mu = \mathrm{TV}(p,q).
\end{aligned}$$

The result shows that if $\mathrm{TV}(p,q) \to 0$, so does $D_\alpha(p,q)$ for all $\alpha \in (0,1)$. This does not imply however that $D_\alpha(p,q) \to 0$ when $\alpha \to 0$ or $\alpha \to 1$, i.e., the two KL measures are left out in general. Also, for the H divergence mesure, it gives the inequality $H^2(p,q) \leq 2\text{TV}(p,q)$. There exists however, a second result which provides an upper bound for TV(p,q) in terms of $H(p,q)$. This is attributed to Le Cam, and is given in Wainwright (2019) as an Exercise. We prove the result below.

Theorem 3. *$[TV(p,q)]^2 \leq H^2(p,q)[1 - \frac{1}{4}H^2(p,q)] \leq H^2(p,q)$.*

Proof. The last inequality is obvious since $H^2(p,q) \leq 2$. To prove the first inquality, we proceed as

$$\begin{aligned}
[TV(p,q)]^2 &= (1/4)(\int |p-q|d\mu)^2 = (1/4)[\int (p^{1/2}+q^{1/2})(p^{1/2}-q^{1/2})d\mu]^2 \\
&\leq (1/4)[\int (p^{1/2}+q^{1/2})^2 d\mu][\int (p^{1/2}-q^{1/2})^2 d\mu] \\
&= (1/4)[2 + 2\int (p^{1/2}q^{1/2})d\mu]H^2(p,q) \\
&= (1/4)(4 - H^2(p,q))H^2(p,q) = H^2(p,q)[1 - \frac{1}{4}H^2(p,q)].
\end{aligned}$$

An important consequence of Theorems 2 and 3 is that TV$(p,q) \leq H(p,q) \leq 2[\text{TV}(p,q)]^{1/2}$. Thus convergence of TV$(p,q)$ to zero is equivalent to convergence of $H(p,q)$ to zero. This yields the important equivalence result

$$H(p,q) \to 0 \equiv TV(p,q) \to 0 \equiv D_\alpha(p,q) \to 0 \equiv R_\alpha(p,q) \to 0 \text{ for all } 0 < \alpha < 1.$$

The above TV and Hellinger equivalence is mentioned also in Gibbs and Su (2002). The general Renyi or α divergence, however, was not mentioned there.

The next inequality in this section provides a relationship between KL(p,q) and $D_\alpha(p,q)$.

Theorem 4. *KL$(p,q) \geq (1-\alpha)D_\alpha(q,p)$.*

Proof. With the result $\log(x) \leq x - 1$ for $x \geq 0$, one obtains

$$\begin{aligned}
KL(p,q) &= \int \log(p/q)p d\mu = -\int \log(q/p)p d\mu \\
&= (-1/\alpha)\int \log(q^\alpha/p^\alpha)p d\mu \\
&\geq (-1/\alpha)\int [q^\alpha/p^\alpha - 1]p d\mu \\
&= (1/\alpha)[1 - \int q^\alpha p^{1-\alpha} d\mu] = (1-\alpha)D_\alpha(q,p).
\end{aligned}$$

It follows as an immediate consequence of the above theorem and $D_{1/2}(p,q) = D_{1/2}(q,p)$, KL$(p,q) \geq (1/2)D_{1/2}(q,p) = (1/2)D_{1/2}(p,q) = H^2(p,q) \geq (\text{TV}(p,q))^2$. A sharper inequality KL$(p,q) \geq 2[\text{TV}(p,q)]^2$, attributed to Pinsker, Csiszar and Kullback, is proved in Wainwright (2019). However, this is just a special case of lemma 2.1 by Gilardoni (2006). Also, this inequality may be vacaous when $KL(p,q) > 2$, since total variation distance is atmost 1. For such cases, an alternative bound can be used due to Bretagnolle and Huber (1978) is given by $TV(p,q) \leq \sqrt{1 - exp(-KL(p,q))}$.

The final inequality in this section provides a relationship between the KL divergence and the chisquare divergences.

Theorem 5. *KL*$(p,q) \leq \int [(p-q)^2/q] d\mu = 2D_2(p,q)$.

Proof. Using the elementary inequality $\log z \leq z - 1$ for $z \geq 0$, one gets

$$\begin{aligned} KL(p,q) = \int \log(p/q) p d\mu = & \int \log(p/q)(p/q) q d\mu \\ \leq & \int (p/q - 1)(p/q) q d\mu = \int [(p-q)^2/q^2 + (p-q)/q] q d\mu \\ = & \int [(p-q)^2/q + (p-q)] d\mu = \int [(p-q)^2/q] d\mu = 2D_2(p,q). \end{aligned}$$

Similarly, one can check that $KL(q,p)$ is bounded above by Neyman χ^2 divergence.

It is clear from our discussion that all the divergence measures considered above can be used for detecting the proximity between distributions. For large n if we define efficiency of these divergence measures in a limiting sense it is clear that total variation distance, Hellinger distance, α divergence, Renyi's divergence for $\alpha \in (0,1)$ are equally efficient for showing whether two distributions are close to each other or not. Also for small or large n, the perception that the TV distance is much weaker than the others mentioned above is no more a truth. Thus a researcher will now have more liberty to choose one among these measures which will make her/his calculations simpler. However, it appears that the KL distance is stronger than the above divergence measures, and the χ^2 distance is even stronger than the KL distance. Hence, the χ^2 distance seems to be the most efficient one in terms on checking proximity between two distributions, since if this distance tends to 0, all of the above mentioned distances will tend to 0.

3 Scheffe's Theorem and Its Applications

Consider a sequence of pdf's $\{p_n, n \geq 1\}$. We have found already from Theorems 2 and 3 that $H(p_n,p) \to 0 \equiv TV(p_n,p) \to 0 \equiv D_\alpha(p_n,p) \to 0 \equiv R_\alpha(p_n,p) \to 0$ for all $0 < \alpha < 1$.

The original density convergence theorem of Scheffe implies that if a sequence of pdf's p_n converges to a pdf p pointwise, then TV$(p_n,p) \to 0$. While this is equivalent to $D_\alpha(p_n,p) \to 0$ and $R_\alpha(p_n,p) \to 0$ for all $0 < \alpha < 1$, it may be interesting to see a direct proof of these results. The result follows from an elementary inequality and the dominated convergence theorem.

Theorem 6. *Suppose a sequence of pdf's p_n converges to a pdf p pointwise. Then $D_\alpha(p_n,p) \to 0$. for every $0 < \alpha < 1$.*

Proof. First use the elementary inequality $p_n^\alpha p^{1-\alpha} \geq \min(p_n.p)$. This leads to the inequality

$$D_\alpha(p_n,p) \leq [\alpha(1-\alpha)]^{-1}[1 - \int \min(p_n,p) d\mu] = [2\alpha(1-\alpha)]^{-1} \int [p_n + p - 2\min(p_n,p)] d\mu. \quad (3.1)$$

Since $p_n \to p$ pointwise, $\min(p_n, p) \to p$ pointwise. Further, $\int \min(p_n, p)d\mu \leq \int p d\mu = 1$. Now by the dominated convergence theorem, $\int \min(p_n, p)d\mu \to 1$. Hence, by (3.1), $D_\alpha(p_n, p) \to 0$. This proves the theorem.

One can get a similar direct proof for $R_\alpha(p_n, p) \to 0$ when pdf's p_n converges to a pdf p pointwise. The only issue that is left now is to examine whether p_n converges to p pointwise implies that KL(p, p_n) converges to zero. I am not aware of any general result in this direction. However, it is obvious that if $\log(p/p_n)$ is bounded above by a function integrable with respect to p, then the dominated convergence theorem continues to apply, and the pointwise convergence will imply the KL convergence. We demonstrate this with two examples, one involving Student's t distribution, and the other involving the F distribution.

To this end, I first prove a lemma involving Gamma functions which may be of independent interest.

Lemma 3.1. $((n-1)/2])^{1/2} \leq \Gamma((n+1)/2)/\Gamma(n/2) \leq (n/2)^{1/2}$.

Proof. $\Gamma((n+1)/2) = \int_0^\infty \exp(-z)z^{(n-1)/2}dz = \int_0^\infty \exp(-z)z^{(n-2)/4}z^{n/4}dz$. Now applying the Cauchy-Schwarz inequality, one gets

$$\begin{aligned}\Gamma((n+1)/2) &\leq \left(\int_0^\infty \exp(-z)z^{(n-2)/2}dz\right)^{1/2}\left(\int_0^\infty \exp(-z)z^{n/2}dz\right)^{1/2}\\ &= \Gamma^{1/2}(n/2)\Gamma^{1/2}((n+2)/2) = \Gamma(n/2)(n/2)^{1/2}.\end{aligned}$$

This establishes the upper bound. To obtain the lower bound, once again applying the Cauchy-Schwarz inequality, one gets

$$\begin{aligned}\Gamma(n/2) &= \int_0^\infty \exp(-z)z^{(n-2)/2}dz = \int_0^\infty \exp(-z)z^{(n-1)/4}z^{(n-3)/4}dz\\ &\leq \left(\int_0^\infty \exp(-z)z^{(n-1)/2}dz\right)^{1/2}\left(\int_0^\infty \exp(-z)z^{(n-3)/2}dz\right)^{1/2}\\ &= \Gamma^{1/2}((n+1)/2)\Gamma^{1/2}((n-1)/2) = \Gamma((n+1)/2)/((n-1)/2).\end{aligned}$$

This establishes the lower bound, and thus completes proof of the lemma.

Now, in view of this lemma and the fact that $(1+t^2/n)^{-(n+1)/2}$ converges to $\exp(-t^2/2)$ as $n \to \infty$, one gets $p_n(t) = \Gamma((n+1)/2)/[\Gamma(n/2)(n\pi)^{1/2}])(1+t^2/n)^{-(n+1)/2}$ converges to $\phi(t) = (2\pi)^{-1/2}\exp(-t^2/2)$, the standard normal pdf. It remains to prove that $|\log(\phi(t)/p_n(t)|$ is bounded above by a function not depending on n, integrable with respect to $\phi(t)$.

To this end, by virtue of lemma 3.1, one gets the inequalities $\log(p_n(t)) \leq \log(2\pi)^{-1/2}$ and

$$\begin{aligned}\log(p_n(t)) &\geq (1/2)\log(1-n^{-1}) + \log(2\pi)^{-1/2} - ((n+1)/2)\log(1+t^2/n)\\ &\geq (1/2)\log(1-n^{-1}) + \log(2\pi)^{-1/2} - (t^2/2)(1+n^{-1}).\end{aligned}$$

Noting that $\log\phi(t) = -t^2/2 - \log(2\pi)^{-1/2}$, one has the inequality

$$-t^2/2 \leq \log\phi(t) - \log(p_n(t)) \leq t^2/(2n) - (1/2)\log(1-n^{-1}).$$

Thus $|\log\phi(t) - \log(p_n(t))| \leq t^2/2 + (1/2)\log(n/(n-1)) \leq t^2/2 + (1/2)\log 2$ for $n \geq 2$, an integrable function with respect to $\phi(t)$. This shows that $\mathrm{KL}(\phi(t), p_n(t)) \to 0$ as $n \to \infty$.

Next consider the F statistic. It is well-known that an F statistic with m, n degrees of freedom converges in distribution to χ^2_m/m as $n \to \infty$, while m remains fixed. I prove now the pointwise convergence of a $F_{m,n}$ density to a χ^2_m/m density, and subsequently the convergence of the KL divergence between the two to zero as $n \to \infty$, but m is held fixed.

To this end, we first write the $F_{m,n}$ density as

$$p_{m,n}(u) = [\Gamma((m+n)/2)/(\Gamma(m/2)\Gamma(n/2))](m/n)^{m/2}u^{m/2-1}(1+\frac{m}{n}u)^{-(m+n)/2}.$$

Next, by repeated application of lemma 3.1, one gets the inequality

$$\prod_{j=0}^{m-1}(n+j-1)^{1/2}2^{-m/2} \leq \Gamma((m+n)/2)/\Gamma(n/2) \leq \prod_{j=0}^{m-1}(n+j)^{1/2}2^{-m/2}. \tag{3.2}$$

Hence, $[\Gamma((m+n)/2)/(n^{m/2}\Gamma(n/2))] \to 2^{-m/2}$ as $n \to \infty$. Further, $(1+\frac{m}{n}u)^{-(m+n)/2} \to \exp(-mu/2)$ as $n \to \infty$, but m is held fixed. Thus as $n \to \infty$, but m is held fixed,

$$p_{m,n}(u) \to \exp(-mu/2)u^{m/2-1}m^{m/2}/[2^{m/2}\Gamma(m/2)],$$

which is the pdf of a χ^2_m/m random variable. From now on, we will denote this random variable by U_m with its density given by $g_m(u)$.

Next to show the boundedness of $|\log(g_m(u)/p_{m,n}(u))|$, by virtue of (3.2), one gets the upper bound

$$\begin{aligned}\log g_m(u) - \log p_{m,n}(u) &\leq -(mu/2) - (1/2)\sum_{j=0}^{m-1}\log\frac{n+j-1}{n} + (1/2)(m+n)\log(1+mu/n)\\ &\leq -(mu)/2 + (1/2)(m+n)(mu/n) = m^2u/(2n) \leq m^2u/2. \end{aligned} \tag{3.3}$$

By (3.2) again, and the fact that $(1 + \frac{m}{n}u)^{-(m+n)/2} \leq 1$, one gets the lower bound

$$\begin{aligned}\log g_m(u) - \log p_{m,n}(u) &\geq -(mu/2) - (1/2)\sum_{j=0}^{m-1}\log(n+j-1) + (m/2)\log n \\ &= -(mu/2) - (1/2)\sum_{j=0}^{m-1}\log(1+j/n) \\ &\geq -(mu/2) - (1/2)\sum_{j=0}^{m-1}(j/n) = -(mu/2) - (1/2)m(m-1).\end{aligned} \tag{3.4}$$

Combining (3.3) and (3.4), one gets $|\log g_m(u) - \log p_{m,n}(u)| \leq m^2u/2 + m(m-1)/2$, which is integrable with respect to $g_m(u)$. An application of the dominated convergence theorem once again yields KL$(g_m, p_{m,n}) \to 0$ as $n \to \infty$, while m is held fixed.

4 Acknowledgements

Thanks are due to Debashis Ghosh for pointing out the reference to Gibbs and Su (2002).

References

Amari, S. (1982), "Differential Geometry of Curved Exponential Families-Curvatures and Information Loss," *The Annals of Statistics*, 10, 357 – 385.

Bhattacharyya, A. (1946), "On a Measure of Divergence between Two Multinomial Populations," *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 7, 401–406.

Bretagnolle, J. and Huber, C. (1978), "Estimation des densités : risque minimax," *Séminaire de probabilités de Strasbourg*, 12, 342–363.

Cressie, N. and Read, T. R. C. (1984), "Multinomial Goodness-of-Fit Tests," *Journal of the Royal Statistical Society. Series B (Methodological)*, 46, 440–464.

Csiszar, I. (1967), "Information-type measures of difference of probability distributions and indirect observation," *Studia Scientiarum Mathematicarum Hungarica*, 2, 229–318.

Gibbs, A. L. and Su, F. E. (2002), "On Choosing and Bounding Probability Metrics," *International Statistical Review / Revue Internationale de Statistique*, 70, 419–435.

Gilardoni, G. L. (2006), "On Pinsker's Type Inequalities and Csiszar's f-divergences. Part I: Second and Fourth-Order Inequalities," *CoRR*, abs/cs/0603097.

Hellinger, E. (1909), “Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen.” *Journal für die reine und angewandte Mathematik*, 1909, 210–271.

Kullback, S. and Leibler, R. A. (1951), “On Information and Sufficiency,” *The Annals of Mathematical Statistics*, 22, 79 – 86.

Rényi, A. (1961), “On measures of entropy and information,” in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Berkeley, California, USA, vol. 1.

Scheffé, H. (1947), “A useful convergence theorem for probability distributions,” *The Annals of Mathematical Statistics*, 18, 434–438.

Wainwright, M. J. (2019), *High-dimensional statistics: A non-asymptotic viewpoint*, vol. 48, Cambridge University Press.

Received: *18 October 2022*

Accepted: *28 November 2022*