Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER Computers and Mathematics with Applications 49 (2005) 575–588

An International Journal
computers & mathematics with applications

www.elsevier.com/locate/camwa

A Symmetric Information Divergence Measure of the Csiszár's f-Divergence Class and Its Bounds

P. KUMAR AND S. CHHINA
Mathematics Department
College of Science and Management
University of Northern British Columbia
Prince George BC V2N4Z9, Canada
<kumarp><chhinas@unbc.ca>

(Received February 2004; accepted July 2004)

Abstract—A symmetric measure of information divergence is proposed. This measure belongs to the class of Csiszár's f-divergences. Its properties are studied and bounds in terms of well known divergence measures obtained. A numerical illustration is carried out to compare this measure with some known divergence measures.

Keywords—Parametric measure, Nonparametric measure, Csiszár's f-divergence, Information measure.

1. INTRODUCTION

There are several types of information divergence measures studied in literature which compare two probability distributions and have applications in information theory, statistics and engineering. A convenient classification to differentiate these measures is to categorize them as parametric, nonparametric and entropy-type measures of information [1]. Parametric measures of information measure the amount of information about an unknown parameter θ supplied by the data and are functions of θ. The best known measure of this type is Fisher's measure of information [2]. Nonparametric measures give the amount of information supplied by the data for discriminating in favor of a probability distribution f_1 against another f_2, or for measuring the distance or affinity between f_1 and f_2. The Kullback-Leibler measure is the best known in this class [3]. Measures of entropy express the amount of information contained in a distribution, that is, the amount of uncertainty associated with the outcome of an experiment. The classical measures of this type are Shannon's [4] and Rényi's measures [5]. The construction of measures of information divergence is not an easy task. Methods for deriving parametric measures of information from the nonparametric ones and from the information matrices are suggested in [1].

This research was partially supported by the first author's Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

0898-1221/05/$ - see front matter Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
doi:10.1016/j.camwa.2004.07.017

In this paper, we present a new symmetric nonparametric information divergence measure which belongs to the class of Csiszár's f-divergences [6,7]. In Section 2, we discuss the Csiszár's f-divergences and their properties. Information inequalities are presented in Section 3. New symmetric divergence measure and its bounds are obtained in Section 4. In Section 5, it is shown that the suggested measure can be applied to the parametric family of distributions. A numerical illustration for studying the behavior of new measure is shown in Section 6. Section 7 concludes the paper.

2. CSISZÁR'S F-DIVERGENCES AND PROPERTIES

Let $\Omega = \{x_1, x_2, \ldots\}$ be a set with at least two elements, $\beta(\Omega)$, the set of all subsets of Ω and $\mathbb{P}$, the set of all probability distributions $P = (p(x) : x \in \Omega)$ on Ω. A pair $(P,Q) \in \mathbb{P}^2$ of probability distributions is called a *simple versus simple testing problem*. Two probability distributions P and Q are called *orthogonal* ($P \perp Q$) if there exists an element $A \in \mathbb{P}(\Omega)$ such that $P(A) = Q(A^c) = 0$, where $A^c = \Omega/A$. A testing problem $(P,Q) \in \mathbb{P}^2$ is called *least informative* if $P = Q$ and *most informative* if $P \perp Q$. Further, let $\mathbb{F}$ be a set of convex functions $f : [0,\infty) \longmapsto (-\infty,\infty)]$ continuous at 0, that is, $f(0) = \lim_{u \downarrow 0} f(u)$, $\mathbb{F}_0 = \{f \in \mathbb{F} : f(1) = 0\}$ and let $D_- f$ and $D_+ f$ denote the *left-hand side* and *right-hand side derivatives* of f, respectively. Define $f^* \in \mathbb{F}$, the **-conjugate* (convex) function of f, by $f^*(u) = uf(1/u)$, $u \in (0,\infty)$ and $\tilde{f} = f + f^*$. For a convex function $f : [0,\infty) \to \mathbb{R}$, the *$f$-divergence* of the probability distributions P and Q is defined [6–8],

$$C_f(P,Q) = \sum_{x \in \Omega} q(x) f\left(\frac{p(x)}{q(x)}\right). \tag{2.1}$$

It is well known that $C_f(P,Q)$ is a versatile functional form which results in a number of popular divergence measures [9,10]. Most common choices of f satisfy $f(1) = 0$, so that $C_f(P,P) = 0$. Convexity ensures that divergence measure $C_f(P,Q)$ is nonnegative. Some examples are $f(u) = u \ln u$ provides the Kullback-Leibler measure [3], $f(u) = |u-1|$ results in the variational distance [11,12], $f(u) = (u-1)^2$ yields the χ^2-divergence [13].

The basic general properties of f-divergences including their axiomatic properties and some important classes are given in [9]. For f, f^*, $f_1 \in \mathbb{F}$, $\forall (P,Q) \in \mathbb{P}^2$, $u \in (0,\infty)$,

(i) $C_f(P,Q) = C_{f^*}(Q,P)$.

(ii) UNIQUENESS THEOREM. (See [14].)

$$I_{f_1}(P,Q) = I_f(P,Q), \qquad \text{iff } \exists c \in R : f_1(u) - f(u) = c(u-1).$$

(iii) Let $c \in [D_- f(1), D_+ f(1)]$. Then, $f_1(u) = f(u) - c(u-1)$ satisfies $f_1(u) \geq f(1)$, $\forall u \in [0,\infty)$ while not changing the f-divergence. Hence, without loss of generality $f_1(u) \geq f(1)$, $\forall u \in [0,\infty)$.

(iv) SYMMETRY THEOREM. (See [14].)

$$I_{f^*}(P,Q) = I_f(P,Q), \qquad \text{iff } \exists c \in R : f^*(u) - f(u) = c(u-1).$$

(v) RANGE OF VALUES THEOREM. (See [15].)

$$f(1) \leq I_f(P,Q) \leq f(0) + f^*(0).$$

In the first inequality, equality holds iff $P = Q$. The latter provides f is strictly convex at 1. The difference $I_f(P,Q) - f(1)$ is a quantity that compares the given testing problem $(P,Q) \in \mathbb{P}^2$ with the *least* informative testing problem. Given $f \in \mathbb{F}$, by setting $f(u) := f(u) - f(1)$, we can have $f(1) = 0$ and hence, without loss of generality, $f \in \mathbb{F}_0$. Thus, $I_f(P,Q)$ serves as an appropriate *measure of similarity* between two distributions.

In the second inequality, equality holds iff $P \perp Q$. The latter provides $\tilde{f}(0) \leq f(0)+f^*(0) < \infty$. The difference $I_g(P,Q) := \tilde{f}(0) - I_f(P,Q)$ is a quantity that compares the given testing problem $(P,Q) \in \mathbb{P}^2$ with the *most* informative testing problem. Thus, $I_g(P,Q)$ serves as an appropriate *measure of orthogonality* for the two distributions where the concave function $g : [0,\infty) \to \mathbb{R}$ is given by $g(u) = f(0) + uf^*(0) - f(u)$.

(vi) CHARACTERIZATION THEOREM. (See [7].) Given a mapping $I : \mathbb{P}^2 \longmapsto (-\infty,\infty)$,

(a) I is an f-divergence, that is, there exists an $f \in \mathbb{F}$, such that

$$I(P,Q) = C_f(P,Q), \qquad \forall\, (P,Q) \in \mathbb{P}^2.$$

(b) $C_f(P,Q)$ is invariant under permutation of Ω.

(c) Let $\mathbb{A} = (A_i,\ i \geq 1)$ be a partition of Ω, and $P_{\mathbb{A}} = (P(A_i),\ i \geq 1)$ and $Q_{\mathbb{A}} = (Q(A_i),\ i \geq 1)$ be the restrictions of the probability distributions P and Q to $\mathbb{A}$.

Then, $I(P,Q) \geq I(P_{\mathbb{A}},Q_{\mathbb{A}})$ with equality if $P(A_i) \times p(x) = Q(A_i) \times p(x)$, $\forall x \in A_i$, $i \geq 1$.

(d) Let P_1, P_2 and Q_1, Q_2 be probability distributions on Ω. Then,

$$I(\alpha P_1 + (1-\alpha)P_2, \alpha Q_1 + (1-\alpha)Q_2) \leq \alpha I(P_1,Q_1) + (1-\alpha) I(P_2,Q_2).$$

By characterization theorem, the *-conjugate of a convex function f is $f^*(u) \equiv uf(1/u)$.

For brevity, in what follows now, we will denote $C_f(P,Q)$, $p(x)$, $q(x)$, and $\sum_{x\in\Omega}$ by $C(P,Q)$, p, q, and $\sum$, respectively.

Some popularly practised information divergence measures are as follows.

χ^α-Divergences.

VARIATIONAL DISTANCE. (See [11,12].)

$$V(P,Q) = \sum |p-q|. \tag{2.2}$$

χ^2-DIVERGENCE. (See [13].)

$$\chi^2(P,Q) = \sum \frac{(p-q)^2}{q} = \sum \frac{p^2}{q} - 1. \tag{2.3}$$

SYMMETRIC χ^2-DIVERGENCE.

$$\Psi(P,Q) = \chi^2(P,Q) + \chi^2(Q,P) = \sum \frac{(p+q)(p-q)^2}{pq}. \tag{2.4}$$

KULLBACK AND LEIBLER. (See [3].)

$$K(P,Q) = \sum p \ln\left(\frac{p}{q}\right). \tag{2.5}$$

KULLBACK-LEIBLER SYMMETRIC DIVERGENCE.

$$J(P,Q) = K(P,Q) + K(Q,P) = \sum (p-q)\ln\left(\frac{p}{q}\right). \tag{2.6}$$

TRIANGULAR DISCRIMINATION. (See [16,17].)

$$\Delta(P,Q) = \sum \frac{|p-q|^2}{p+q}. \tag{2.7}$$

SIBSON INFORMATION RADIUS. (See [18–21].)

$$I_r(P,Q) = \begin{cases} (r-1)^{-1}\left[\sum\left(\frac{p^r+q^r}{2}\right)\left(\frac{p+q}{2}\right)^{1-r} - 1\right], & r \neq 1, r > 0, \\ \sum \frac{p\ln p + q\ln q}{2} - \left(\frac{p+q}{2}\right)\ln\left(\frac{p+q}{2}\right), & r = 1. \end{cases} \tag{2.8}$$

TANEJA DIVERGENCE MEASURE. (See [22].)

$$T_r(P,Q) = \begin{cases} (r-1)^{-1}\left[\sum\left(\frac{p^{1-r}+q^{1-r}}{2}\right)\left(\frac{p+q}{2}\right)^{r} - 1\right], & r \neq 1, r > 0, \\ \sum\left(\frac{p+q}{2}\right)\ln\left(\frac{p+q}{2\sqrt{pq}}\right), & r = 1. \end{cases} \tag{2.9}$$

The following divergences are famous divergence measures. It may be noted that they are not members of the family of Csiszár's f-divergences (since the functions $g^\alpha(u) = u^\alpha$, $\alpha \in (0,1)$ are concave).

BHATTACHARYYA DISTANCE. (See [23].)

$$B(P,Q) = \sum \sqrt{pq}. \tag{2.10}$$

HELLINGER DISCRIMINATION. (See [24].)

$$h(P,Q) = \sum \frac{\left(\sqrt{p}-\sqrt{q}\right)^2}{2}. \tag{2.11}$$

RÉNYI MEASURE. (See [5].)

$$R_r(P,Q) = \begin{cases} (r-1)^{-1}\ln\left(\sum p^r q^{1-r}\right), & r \in (0,\infty)\backslash\{1\}, \\ \sum p \ln \frac{p}{q}, & r = 1. \end{cases} \tag{2.12}$$

We have the following inequalities relating to $V(P,Q)$, $\Delta(P,Q)$, $K(P,Q)$, and $h(P,Q)$.

CSISZÁR. (See [6].)

$$K(P,Q) \geq \frac{V^2(P,Q)}{2}. \tag{2.13}$$

CSISZÁR. (See [6,7].)

$$K(P,Q) \geq \frac{V^2(P,Q)}{2} + \frac{V^4(P,Q)}{36}. \tag{2.14}$$

TOPSϕE. (See [17].)

$$K(P,Q) \geq \frac{V^2(P,Q)}{2} + \frac{V^4(P,Q)}{36} + \frac{V^6(P,Q)}{270} + \frac{V^8(P,Q)}{340200}. \tag{2.15}$$

VAJDA AND TOUSSAINT. (See [15] and [25], respectively.)

$$K(P,Q) \geq \max\{L_1(V), L_2(V)\}, \tag{2.16}$$

where, from [15],

$$L_1(V) = \ln\left(\frac{2+V}{2-V}\right) - \frac{2V}{2+V}, \qquad 0 \leq V \leq 2,$$

and from [26],

$$L_2(V) = \frac{V^2}{2} + \frac{V^4}{36} + \frac{V^8}{288}, \qquad 0 \leq V \leq 2.$$

TOPSØE. (See [18].)

$$\frac{1}{2}V^2(P,Q) \le \Delta(P,Q) \le V(P,Q). \tag{2.17}$$

LECAM AND DACUNHA-CASTELLE (SEE [16] AND [17], RESPECTIVELY.).

$$2h(P,Q) \le \Delta(P,Q) \le 4h(P,Q). \tag{2.18}$$

KRAFT. (See [27].)

$$\frac{1}{8}V^2(P,Q) \le h(P,Q)\left(1 - \frac{1}{2}h(P,Q)\right). \tag{2.19}$$

TOPSϕE. (See [18].)

$$\frac{1}{8}V^2(P,Q) \le h(P,Q) \le \frac{1}{2}V(P,Q), \tag{2.20}$$

and

$$K(P,Q) \le (\log 2)\, V(P,Q) + \log c, \tag{2.21}$$

where $c = \max(p_i/q_i)$, $\forall\, i = 1, \dots, n$.

3. INFORMATION BOUNDS

Different kinds of bounds on the distance, information and divergence measures have been studied recently [28–34]. In [28], authors unified and generalized information bounds for $C(P,Q)$ studied in [29–34] given in the following theorem.

THEOREM 3. *Let $f : I \subset R_+ \to \mathbb{R}$ be a mapping which is normalized, i.e., $f(1) = 0$ and suppose that*

(i) *f is twice differentiable on (r,R), $0 \leqslant r \leqslant 1 \leqslant R < \infty$, ($f'$ and f'' denote the first and second derivatives of f),*

(ii) *there exists real constants m, M, such that $m < M$ and $m \leqslant x^{2-s}f''(x) \leqslant M$, $\forall x \in (r,R)$, $s \in \mathbb{R}$.*

If $P, Q \in \mathbb{P}^2$ are discrete probability distributions with $0 < r \leqslant p/q \leqslant R < \infty$,

$$m\,\Phi_s(P,Q) \leqslant C(P,Q) \leqslant M\,\Phi_s(P,Q), \tag{3.1}$$

and

$$m\left(\eta_s(P,Q) - \Phi_s(P,Q)\right) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant M\left(\eta_s(P,Q) - \Phi_s(P,Q)\right), \tag{3.2}$$

where

$$\Phi_s(P,Q) = \begin{cases} {}^2K_s(P,Q), & s \neq 0,1, \\ K(Q,P), & s = 0, \\ K(P,Q), & s = 1, \end{cases} \tag{3.3}$$

$${}^2K_s(P,Q) = [s(s-1)]^{-1}\left[\sum p^s q^{1-s} - 1\right], \qquad s \neq 0,1, \tag{3.4}$$

$$K(P,Q) = \sum p \ln\left(\frac{p}{q}\right),$$

$$C_\rho(P,Q) = C_{f'}\left(\frac{P^2}{Q}, P\right) - C_{f'}(P,Q) = \sum (p-q)\, f'\left(\frac{p}{q}\right), \tag{3.5}$$

$$\eta_s(P,Q) = C_{\phi_s'}\left(\frac{P^2}{Q}, P\right) - C_{\phi_s'}(P,Q),$$

$$= \begin{cases} (s-1)^{-1}\sum (p-q)\left(\frac{p}{q}\right)^{s-1}, & s \neq 1, \\ \sum (p-q)\ln\left(\frac{p}{q}\right), & s = 1 \end{cases} \tag{3.6}$$

As a consequence of this theorem, following information inequalities which are interesting from the *information-theoretic* point of view, are also obtained in [28].

(i) The case $s = 2$ provides the information bounds in terms of the χ^2-divergence, $\chi^2(P,Q)$,

$$\frac{m}{2}\chi^2(P,Q) \leqslant C(P,Q) \leqslant \frac{M}{2}\chi^2(P,Q), \tag{3.7}$$

and

$$\frac{m}{2}\chi^2(P,Q) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant \frac{M}{2}\chi^2(P,Q). \tag{3.8}$$

(ii) For $s = 1$, the information bounds in terms of the Kullback-Leibler divergence, $K(P,Q)$,

$$mK(P,Q) \leqslant C(P,Q) \leqslant MK(P,Q) \tag{3.9}$$

and

$$mK(Q,P) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant MK(Q,P). \tag{3.10}$$

(iii) The case $s = 1/2$ yields the information bounds in terms of the Hellinger's discrimination, $h(P,Q)$,

$$4mh(P,Q) \leqslant C(P,Q) \leqslant 4Mh(P,Q) \tag{3.11}$$

and

$$4m\left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant 4M\left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right). \tag{3.12}$$

(iv) For $s = 0$, the information bounds in terms of the Kullback-Leibler and χ^2-divergences,

$$mK(P,Q) \leqslant C(P,Q) \leqslant MK(P,Q) \tag{3.13}$$

and

$$m\left(\chi^2(Q,P) - K(Q,P)\right) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant M\left(\chi^2(Q,P) - K(Q,P)\right). \tag{3.14}$$

4. NEW INFORMATION DIVERGENCE MEASURE

Let the convex function $f : (0,\infty) \to \mathbb{R}$ be

$$f(u) = \frac{(u+1)(u-1)^2}{u}\ln\frac{u+1}{2\sqrt{u}}. \tag{4.1}$$

Then, we have the following new divergence measure belonging to the Csiszár's f-divergence family,

$$S(P,Q) = \sum \frac{(p+q)(p-q)^2}{pq}\ln\frac{p+q}{2\sqrt{pq}}. \tag{4.2}$$

Since we can express $S(P,Q)$ as

$$S(P,Q) = \sum\left[\frac{(p+q)(p-q)^2}{pq}\right]\ln\left[\frac{p+q}{2}\cdot\frac{1}{\sqrt{pq}}\right],$$

this measure can be termed as the *Symmetric Chi-Square, Arithmetic*, and *Geometric Mean* divergence measure.

It may be noted that, $f(u)$ in (4.1) satisfies $f(1) = 0$, so that $S(P,P) = 0$. Convexity of $f(u)$ ensures that divergence measure $S(P,Q)$ is nonnegative. Thus, we have

(a) $S(P,Q) \geq 0$ and $S(P,Q) = 0$, iff $P = Q$.
(b) $S(P,Q)$ is symmetric with respect to probability distribution.
(c) Since $f^*(u) \equiv uf(1/u) = f(u)$, function $f(u)$ is the *-self conjugate. Therefore, Properties (i)–(vi) of Section 2 hold good for $f(u)$.

We now derive information divergence inequalities providing bounds for $S(P,Q)$ in terms of the well known divergence measures in the following propositions.

Proposition 4.1. *Let $S(P,Q)$ and $\Delta(P,Q)$ be defined as (4.2) and (2.7), respectively. Then, inequality*

$$SM(P,Q) \le 4\sum \frac{(p-q)^2}{\sqrt{pq}} - 2\Delta(P,Q). \tag{4.3}$$

Proof. Consider the *Arithmetic (AM), Geometric(GM)* and *Harmonic mean(HM)* inequality, i.e., HM $\le$ GM $\le$ AM. Then,

$$\begin{aligned} &\mathrm{HM} \le \mathrm{AM} \\ \text{or,}\quad &\frac{2pq}{p+q} \le \frac{p+q}{2} \\ \text{or,}\quad &\ln\frac{p+q}{2\sqrt{pq}} \ge \ln\frac{2\sqrt{pq}}{p+q}. \end{aligned} \tag{4.4}$$

Multiplying both sides of (4.4) by $(p+q)(p-q)^2/pq$, we have

$$\frac{(p+q)(p-q)^2}{pq}\ln\frac{p+q}{2\sqrt{pq}} \ge \frac{(p+q)(p-q)^2}{pq}\ln\frac{2\sqrt{pq}}{p+q}. \tag{4.5}$$

From HM $\le$ GM, we have $2\sqrt{pq}/p+q \le 1$, and thus,

$$\ln\frac{2\sqrt{pq}}{p+q} = \ln\left(1+\left(\frac{2\sqrt{pq}}{p+q}-1\right)\right) \approx \frac{4\sqrt{pq}}{p+q} - \frac{2pq}{(p+q)^2} - \frac{3}{2}. \tag{4.6}$$

Now, from (4.5), (4.6), and summing over all $x \in \Omega$, we get

$$\begin{aligned} \sum \frac{(p+q)(p-q)^2}{pq}\ln\frac{p+q}{2\sqrt{pq}} &\le 4\sum\frac{(p-q)^2}{\sqrt{pq}} - 2\sum\frac{(p-q)^2}{p+q}, \\ SM(P,Q) &\le 4\sum\frac{(p-q)^2}{\sqrt{pq}} - 2\Delta(P,Q), \end{aligned}$$

and hence, the proof.

Next, we derive the information bounds in terms of the χ^2-divergence, that is, $\chi^2(P,Q)$.

Proposition 4.2. *Let $\chi^2(P,Q)$ and $S(P,Q)$ be defined as (2.3) and (4.2), respectively. For $P,Q \in \mathbb{P}^2$ and $0 < r \leqslant p/q \leqslant R < \infty$, we have,*

$$\begin{aligned} &0 \le S(P,Q) \\ &\leqslant \left(\frac{1}{4r^3(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2\ln\frac{r+1}{2\sqrt{r}} + \left(3r^2+4r+3\right)(r-1)^2\right]\chi^2(P,Q) \end{aligned} \tag{4.7}$$

and

$$\begin{aligned} &0 \le S_\rho(P,Q) - S(P,Q) \\ &\leqslant \left(\frac{1}{4r^3(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2\ln\frac{r+1}{2\sqrt{r}} + \left(3r^2+4r+3\right)(r-1)^2\right]\chi^2(P,Q), \end{aligned} \tag{4.8}$$

where

$$S_\rho(P,Q) = \sum\frac{(p-q)^2}{2p^2q}\left[2\left(2p^2+pq+q^2\right)\ln\frac{p+q}{2\sqrt{pq}} + p^2 - 2pq + q\right]. \tag{4.9}$$

Proof. From the expression of $f(u)$ in (4.1), we have

$$f'(u) = \left(\frac{u-1}{2u^2}\right)\left[2\left(2u^2+u+1\right)\ln\frac{u+1}{2\sqrt{u}} + (u-1)^2\right], \tag{4.10}$$

and thus,

$$S_\rho(P,Q) = \sum (p-q) f'\left(\frac{p}{q}\right) = \sum \left(\frac{(p-q)^2}{2p^2 q}\right)\left[2\left(2p^2+pq+q^2\right)\ln\frac{p+q}{2\sqrt{pq}} + p^2 - 2pq + q\right].$$

Further,

$$f''(u) = \left(\frac{1}{2u^3(u+1)}\right)\left[4\left(u^2-u+1\right)(u+1)^2 \ln\frac{u+1}{2\sqrt{u}} + \left(3u^2+4u+3\right)(u-1)^2\right]. \quad (4.11)$$

Now, if $u \in [r,R] \subset (0,\infty)$, then

$$0 \le f''(u) \le \left(\frac{1}{2r^3(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2 \ln\frac{r+1}{2\sqrt{r}} + \left(3r^2+4r+3\right)(r-1)^2\right], \quad (4.12)$$

where r and R are defined above. In view of (3.7), (3.8), and (4.12), we get inequalities (4.7) and (4.8), respectively.

Now, the information bounds in terms of the Kullback-Leibler divergence, $K(P,Q)$ follows.

PROPOSITION 4.3. *Let $K(P,Q)$, $S(P,Q)$, and $S_\rho(P,Q)$ be defined as (2.6), (4.2), and (4.9), respectively. If $P,Q \in \mathbb{P}^2$ and $0 < r \leqslant p/q \leqslant R < \infty$, then,*

$$\begin{aligned} 0 &\le S(P,Q) \\ &\leqslant \left(\frac{1}{2r^2(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2 \ln\frac{r+1}{2\sqrt{r}} + \left(3r^2+4r+3\right)(r-1)^2\right] K(P,Q), \end{aligned} \quad (4.13)$$

and

$$\begin{aligned} 0 &\le S_\rho(P,Q) - S(P,Q) \\ &\leqslant \left(\frac{1}{2r^2(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2 \ln\frac{r+1}{2\sqrt{r}} + \left(3r^2+4r+3\right)(r-1)^2\right] K(Q,P). \end{aligned} \quad (4.14)$$

PROOF. Consider $f''(u)$ as given in (4.11) and let the function $g : [r,R] \to \mathbb{R}$ be such that,

$$g(u) = uf''(u) = \left(\frac{1}{2u^2(u+1)}\right)\left[4\left(u^2-u+1\right)(u+1)^2 \ln\frac{u+1}{2\sqrt{u}} + \left(3u^2+4u+3\right)(u-1)^2\right]. \quad (4.15)$$

Then,

$$0 \le g(u) \le \left(\frac{1}{2r^2(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2 \ln\frac{r+1}{2\sqrt{r}} + \left(3r^2+4r+3\right)(r-1)^2\right]. \quad (4.16)$$

The inequalities (4.13) and (4.14) follow from (3.9), (3.10), and (4.16).

The following proposition provides the information bounds in terms of the Hellinger's discrimination, $h(P,Q)$ and $\eta_{1/2}(P,Q)$.

PROPOSITION 4.4. *Let $h(P,Q)$, $\eta_{1/2}(P,Q)$, $S(P,Q)$, and $S_\rho(P,Q)$ be defined as (2.26), (3.6), (4.2), and (4.11), respectively. For $P,Q \in \mathbb{P}^2$ and $0 < r \leqslant p/q \leqslant R < \infty$,*

$$\begin{aligned} 0 &\le S(P,Q) \\ &\leqslant \left(\frac{2}{R^{3/2}(u+1)}\right)\left[4\left(R^2-R+1\right)(R+1)^2 \ln\frac{R+1}{2\sqrt{R}} + \left(3R^2+4R+3\right)(R-1)^2\right] h(P,Q) \end{aligned} \quad (4.17)$$

and

$$\begin{aligned}0 &\leq S_\rho(P,Q) - S(P,Q)\\ &\leqslant \left(\frac{2}{R^{3/2}(u+1)}\right)\left[4\left(R^2-R+1\right)(R+1)^2\ln\frac{R+1}{2\sqrt{R}} + \left(3R^2+4R+3\right)(R-1)^2\right]\\ &\quad \times \left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right).\end{aligned} \tag{4.18}$$

PROOF. For $f(u)$ in (4.1), we have $f''(u)$ given by (4.11). Let the function $g : [r, R] \to \mathbb{R}$ be such that

$$\begin{aligned}g(u) &= u^{3/2} f''(u)\\ &= \left(\frac{1}{2u^{3/2}(u+1)}\right)\left[4\left(u^2-u+1\right)(u+1)^2\ln\frac{u+1}{2\sqrt{u}} + \left(3u^2+4u+3\right)(u-1)^2\right],\end{aligned} \tag{4.19}$$

then,

$$\begin{aligned}0 &\leq g(u)\\ &\leq \left(\frac{1}{2R^{3/2}(u+1)}\right)\left[4\left(R^2-R+1\right)(R+1)^2\ln\frac{R+1}{2\sqrt{R}} + \left(3R^2+4R+3\right)(R-1)^2\right].\end{aligned} \tag{4.20}$$

Thus, inequalities (4.17) and (4.18) are established using (3.11), (3.12), and (4.20).

Next, follows the information bounds in terms of the Kullback-Leibler and χ^2-divergences.

PROPOSITION 4.5. *Let $\chi^2(P,Q), K(P,Q)$, $S(P,Q)$, and $S_\rho(P,Q)$ be defined as (2.3), (2.6), (4.2), and (4.11), respectively. If $P,Q \in \mathbb{P}^2$ and $0 < r \leqslant p/q \leqslant R < \infty$, then,*

$$\begin{aligned}0 &\leqslant S(P,Q)\\ &\leqslant \left(\frac{1}{2R(R+1)}\right)\left[4\left(R^2-R+1\right)(R+1)^2\ln\frac{R+1}{2\sqrt{R}} + \left(3R^2+4R+3\right)(R-1)^2\right] K(P,Q)\end{aligned} \tag{4.21}$$

and

$$\begin{aligned}0 &\leqslant S_\rho(P,Q) - S(P,Q)\\ &\leqslant \left(\frac{1}{2R(R+1)}\right)\left[4\left(R^2-R+1\right)(R+1)^2\ln\frac{R+1}{2\sqrt{R}} + \left(3R^2+4R+3\right)(R-1)^2\right]\\ &\quad \times \left(\chi^2(Q,P) - K(Q,P)\right).\end{aligned} \tag{4.22}$$

PROOF. From the expression (4.1), we have $f''(u)$ as given in (4.11). Let the function $g : [r, R] \to \mathbb{R}$ be such that,

$$\begin{aligned}g(u) &= u^2 f''(u)\\ &= \left(\frac{1}{2u(u+1)}\right)\left[4\left(u^2-u+1\right)(u+1)^2\ln\frac{u+1}{2\sqrt{u}} + \left(3u^2+4u+3\right)(u-1)^2\right],\end{aligned} \tag{4.23}$$

then,

$$\begin{aligned}0 &\leq g(u)\\ &\leq \left(\frac{1}{2R(R+1)}\right)\left[4\left(R^2-R+1\right)(R+1)^2\ln\frac{R+1}{2\sqrt{R}} + \left(3R^2+4R+3\right)(R-1)^2\right].\end{aligned} \tag{4.24}$$

Thus, (4.21) and (4.22) follow from (3.13), (3.14), and (4.24).

5. PARAMETRIC MEASURE OF INFORMATION

The parametric measures of information are applicable to regular families of probability distributions, that is, to the families for which the following regularity conditions are assumed to be satisfied. Let for $\theta = (\theta_1, \ldots \theta_k)$, the Fisher information matrix [2] be

$$I_x(\theta) = \begin{cases} E_\theta\left[\frac{\partial}{\partial\theta}\log f(X,\theta)\right]^2, & \text{if } \theta \text{ is univariate,} \\ \left\|E_\theta\left[\frac{\partial}{\partial\theta_i}\log f(X,\theta)\frac{\partial}{\partial\theta_j}\log f(X,\theta)\right]\right\|_{k\times k}, & \text{if } \theta \text{ is } k\text{-variate,} \end{cases} \tag{5.1}$$

where $\|\ \|_{k\times k}$ denotes a $k \times k$ matrix.

The regularity conditions are,

(R1) $f(x,\theta) > 0$, for all $x \in \Omega$ and $\theta \in \Theta$,

(R2) $\frac{\partial}{\partial\theta_i} f(X,\theta)$ exists, for all $x \in \Omega$ and $\theta \in \Theta$ and all $i = 1, \ldots, k$,

(R3) $\frac{d}{d\theta_i}\int_A f(x,\theta)\,d\mu = \int_A \frac{d}{d\theta_i} f(x,\theta)\,d\mu$, for any $A \in \mathbb{A}$ (measurable space (X, A) in respect of a finite or σ-finite measure μ), all $\theta \in \Theta$, and all i.

In [1], authors suggested the following method to construct the parametric measure from the nonparametric measure.

Let $k(\theta)$ be a one-to-one transformation of the parameter space Θ onto itself, with $k(\theta) \neq \theta$. The quantity

$$I_x[\theta, k(\theta)] = I_x[f(x,\theta), f(x,k(\theta))], \tag{5.2}$$

can be considered as a parametric measure of information based on $k(\theta)$.

This method is employed to construct the modified Csiszár's measure of information about univariate θ contained in X and based on $k(\theta)$ as,

$$I_x^C[\theta, k(\theta)] = \int f(x,\theta)\,\phi\left(\frac{f(x,k(\theta))}{f(x,\theta)}\right) d\mu. \tag{5.3}$$

Now, we have the following proposition for the parametric measure of information from $S(P,Q)$.

PROPOSITION 5.1. *Let the convex function $\phi : (0,\infty) \to \mathbb{R}$ be*

$$\phi(u) = \frac{(u+1)(u-1)^2}{u}\ln\frac{u+1}{2\sqrt{u}}, \tag{5.4}$$

and the corresponding nonparametric divergence measure,

$$S(P,Q) = \sum \frac{(p+q)(p-q)^2}{pq}\ln\frac{p+q}{2\sqrt{pq}}.$$

Then, the parametric measure $S^C(P,Q)$ is the same as $S(P,Q)$.

PROOF. For discrete random variables X, the expression (5.3) can be written as

$$I_x^C[\theta, k(\theta)] = \sum_{x\in\Omega} p(x)\,\phi\left(\frac{q(x)}{p(x)}\right). \tag{5.5}$$

From (5.4), we have

$$\phi\left(\frac{q(x)}{p(x)}\right) = \frac{(p+q)(p-q)^2}{p^2 q}\ln\frac{p+q}{2\sqrt{pq}}, \tag{5.6}$$

where we denote $p(x)$, and $q(x)$ by p and q, respectively.

Then, $S^C(P,Q)$ follows from (5.5) and (5.6) as

$$S^C(P,Q) := I_x^C[\theta, k(\theta)] = \sum_{x\in\Omega} \frac{(p+q)(p-q)^2}{pq}\ln\frac{p+q}{2\sqrt{pq}} = S(P,Q), \tag{5.7}$$

and hence, the proposition.

6. NUMERICAL ILLUSTRATION

We consider two examples of symmetrical and asymmetrical probability distributions. We calculate measures $S(P,Q)$, $\Psi(P,Q), \chi^2(P,Q)$, $J(P,Q)$, and verify bounds derived above for $S(P,Q)$.

EXAMPLE 1 SYMMETRICAL. Let P be the binomial probability distribution for the random variable X with parameters ($n = 8$, $p = 0.5$) and Q its approximated normal probability distribution. Then, we have Table 1.

Table 1. Binomial Probability Distribution ($n = 8,\ p = 0.5$).

x	0	1	2	3	4	5	6	7	8
$p(x)$	0.004	0.031	0.109	0.219	0.274	0.219	0.109	0.031	0.004
$q(x)$	0.005	0.030	0.104	0.220	0.282	0.220	0.104	0.030	0.005
$p(x)/q(x)$	0.774	1.042	1.0503	0.997	0.968	0.997	1.0503	1.042	0.774

The measures $S(P,Q)$, $\Psi(P,Q), \chi^2(P,Q)$, and $J(P,Q)$ are,

$$S(P,Q) = 0.00001030, \qquad \Psi(P,Q) = 0.00305063,$$
$$\chi^2(P,Q) = 0.00145837, \qquad J(P,Q) = 0.00151848.$$

It is noted that,

$$r\,(= 0.774179933) \leqslant \frac{p}{q} \leqslant R\,(= 1.050330018)\,.$$

The upper bound for $S(P,Q)$ based on $\chi^2(P,Q)$ divergence from (4.7),

$$\text{Upper Bound} = \left(\frac{1}{4r^3(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2\ln\frac{r+1}{2\sqrt{r}} + \left(3r^2+4r+3\right)(r-1)^2\right]\chi^2(P,Q)$$
$$= 0.00158357,$$

and, thus, $0 < S(P,Q) = 0.00001134 < 0.001575814$. The length of the interval is 0.001575814.

The upper bound for $S(P,Q)$ based on $K(P,Q)$ from (4.13),

$$\text{Upper Bound} = \left(\frac{1}{2r^2(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2\ln\frac{r+1}{2\sqrt{r}} + \left(3r^2+4r+3\right)(r-1)^2\right]K(P,Q)$$
$$= 0.002850183,$$

and therefore, $0 < S(P,Q) = 0.00001134 < 0.002850183$. The length of the interval is 0.002850183.

EXAMPLE 2 ASYMMETRICAL. Let P be the binomial probability distribution for the random variable X, with parameters ($n = 8$, $p = 0.4$) and Q its approximated normal probability distribution. Then, we have Table 2.

Table 2. Binomial Probability Distribution ($n = 8, p = 0.4$).

x	0	1	2	3	4	5	6	7	8
$p(x)$	0.017	0.090	0.209	0.279	0.232	0.124	0.041	0.008	0.001
$q(x)$	0.020	0.082	0.198	0.285	0.244	0.124	0.037	0.007	0.0007
$p(x)/q(x)$	0.850	1.102	1.056	0.979	0.952	1.001	1.097	1.194	1.401

The measures $S(P,Q)$, $\Psi(P,Q)$, $\chi^2(P,Q)$, and $J(P,Q)$ are,

$$\Psi M(P,Q) = 0.00001134, \qquad \Psi(P,Q) = 0.00305063,$$
$$\chi^2(P,Q) = 0.00145837, \qquad J(P,Q) = 0.00151848.$$

It is noted that,

$$r\,(=0.774179933) \leqslant \frac{p}{q} \leqslant R\,(=1.050330018)\,.$$

The upper bound for $S(P,Q)$ based on $\chi^2(P,Q)$ divergence from (4.7),

$$\begin{aligned}\text{Upper Bound} &= \left(\frac{1}{4r^3\,(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2\ln\frac{r+1}{2\sqrt{r}}+\left(3r^2+4r+3\right)(r-1)^2\right]\chi^2\,(P,Q)\\ &= 0.001575814,\end{aligned}$$

and thus, $0 < S(P,Q) = 0.00001134 < 0.001575814$. The length of the interval is 0.001575814.

The upper bound for $S(P,Q)$ based on $K(P,Q)$ from (4.13),

$$\begin{aligned}\text{Upper Bound} &= \left(\frac{1}{2r^2\,(r+1)}\right)\left[4\left(r^2-r+1\right)(r+1)^2\ln\frac{r+1}{2\sqrt{r}}+\left(3r^2+4r+3\right)(r-1)^2\right]K\,(P,Q)\\ &= 0.007996539,\end{aligned}$$

and $0 < S(P,Q) = 0.00001134 < 0.007996539$. The length of the interval is 0.007996539.

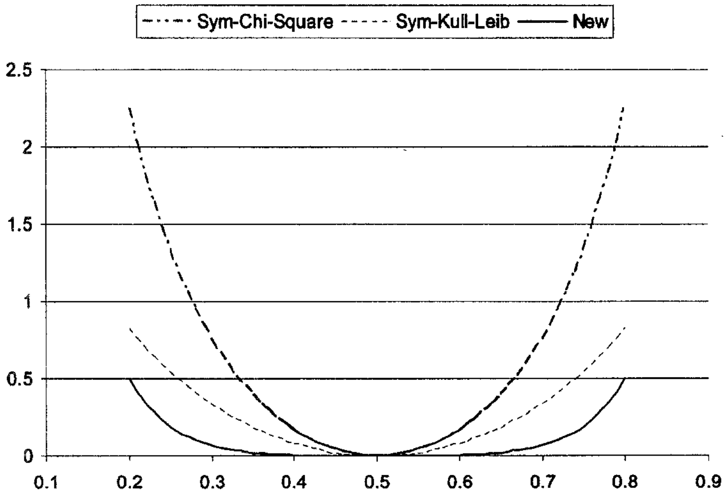

Figure Figure 1. Measures $S(P,Q)$-New, $\Psi(P,Q)$-Sym Chi Square, and $J(P,Q)$- Sym Kullback Leibler.

Figure 1 shows the behavior of $S(P,Q)$-[New], $\Psi(P,Q)$-[Sym-Chi-square] and $J(P,Q)$-[Sym-Kull-Leib]. We have considered $p = (a, 1-a)$ and $q = (1-a, a)$, $a \in [0,1]$. It is clear from Figure 1 that measures $\Psi(P,Q)$ and $J(P,Q)$ have a steeper slope than $S(P,Q)$.

7. CONCLUDING REMARKS

The Csiszár's f-divergence is a general class of divergence measures which includes several divergences used in measuring the distance or affinity between two probability distributions. This class is introduced by using a convex function f, defined on $(0,\infty)$. An important property of this divergence is that many known divergences can be obtained from this measure by appropriately defining the convex function f. Nonparametric measures for the Csiszár's f-divergences are also available. For this class of divergences, its properties, bounds, and relations among well known divergences have been of interest to the researchers. We have introduced a new symmetric divergence measure by considering a convex function and have investigated its properties. Further, we have established its bounds in terms of known divergence measures. Work on one parametric generalization of this measure is in progress and will be reported elsewhere.

REFERENCES

1. K. Ferentimos and T. Papaiopannou, New parametric measures of information, *Information and Control* **51**, 193–208, (1981).
2. R. A. Fisher, Theory of statistical estimation, *Proc. Cambridge Philos. Soc.* **22**, 700–725, (1925).
3. S. Kullback and A.S. Leibler, On information and sufficiency **22**, 79–86, (1951).
4. C.E. Shannon, A mathematical theory of communications, *Bell Syst. Tech. Jour.* **27**, 623–659, (1958).
5. A. Rényi, On measures of entropy and information, In *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, 1 pp. 547–561, University of California Press, Berkeley, CA U.S.A., (1961).
6. I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.* **2**, 299–318, (1967).
7. I. Csiszár, Information measures: A critical survey, In *Trans. Seventh Prague Conf. on Information Theory*, A, pp. 73–86, Academia, Prague, (1974).
8. S.M. Ali and S.D. Silvey, A general class of coefficients of divergence of one distribution from another, *Jour. Roy. Statist. Soc.* **B** (28), 131–142, (1966).
9. F. Österreicher, Csiszár's f-divergences-Basic properties, Res. Report Collection `http://rgmia.vu.edu.au/monographs/csiszar.htm`, (2002).
10. P. Kumar and A. Johnson, On information inequalities for the Csiszár's f-divergences and a new symmetric divergence measure, (submitted).
11. A.N. Kolmogorov, A new invariant for transitive dynamical system, *Dokl. Akad. Nauk USSR* **119**, 861–869, (1957).
12. A.N. Kolmogorov, Entropy per unit time as a metric invariant of automorphisms, *Dokl. Akad. Nauk USSR* **120**, 754–758, (1958).
13. K. Pearson, On the criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonable supposed to have arisen from random sampling, *Phil. Mag.* **50**, 157–172, (1900).
14. F. Liese and I. Vajda, *Teubner-Texte zur Mathematik* 95, Leipzig, (1987).
15. I. Vajda, On f-divergence and singularity of probability measures, *Period. Math. Hungar.* **2**, 223–234, (1972).
16. L. Le Cam, *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York, (1986).
17. F. Topsϕe, 85–98, *Res. Rep. Collection: RGMIA* **2** (1), (1999).
18. R. Sibson, Information radius, *Z. Wahrs. und Verw Geb.* **14**, 149–160, (1969).
19. J. Burbea and C.R. Rao, Entropy differential metric, distance and divergence measures in probability spaces: A unified approach, *Jour. Multi. Analysis* **12**, 575–596, (1982).
20. J. Burbea and C.R. Rao, On the convexity of some divergence measures based on entropy functions, *IEEE Trans. on Information Theory* **IT-28**, 489–495, (1982).
21. M. Salicru and I.J. Taneja, Connections of generalized divergence measures with Fisher information matrix, *Information Sciences* **72**, 251–269, (1993).
22. I.J. Taneja, New developments in generalized information measures, *Advances in Electronics and Electron Physics* (Edited by P.W. Hawkes) **91**, 37–135, (1995).
23. A. Bhattacharyya, On some analogues to the amount of information and their uses in statistical estimation, *Sankhȳa* **8**, 1–14, (1946).
24. E. Hellinger, Neue begründung der theorie quadratischen formen von unendlichen vielen veränderlichen, *Jour. Reine Ang. Math.* **136**, 210–271, (1909).
25. Toussaint, Probability of error, expected divergence, and the affinity of several distributions, *IEEE Trans. on Syst. Man. and Cybern.* **SMC-8**, 482–485, (1978).
26. D. Dacunha-Castelle, *Ecole d'été de probabilités de saint-Fleur*, VII-1977, Springer, (1978).
27. C. Kraft, Some conditions of consistency and uniform consistency of statistical procedure, Univ. of California Publ. in Statistics, Ph.D. Thesis 2, (1955).
28. I.J. Taneja and P. Kumar, Relative information of type-s, Csiszar's f-divergence and information inequalities, *Information Sciences*, (2003).
29. S.S. Dragomir, Some inequalities for the Csiszár's ϕ-divergence, Inequalities for the Csiszár's f-divergence in Information Theory, (Edited by S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.htm`, (2000).
30. S.S. Dragomir, A converse inequality for the Csiszár's ϕ-divergence,Inequalities for the Csiszár's f-divergence in Information Theory, (Edited by S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.htm`, (2000).
31. S.S. Dragomir, Some inequalities for (m, M)-convex mappings and applications for the Csiszár's ϕ-divergence in information theory, Inequalities for the Csiszár's f-divergence in Information Theory, (Edited by S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.htm`, (2000).
32. S.S. Dragomir, Upper and lower bounds for Csiszár's f-divergence in terms of the Kullback-Leibler distance and applications, Inequalities for the Csiszár's f-divergence in Information Theory, (Edited by S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.htm`, (2000).
33. S.S. Dragomir, Upper and lower bounds for Csiszár's f-divergence in terms of the Hellinger discrimination and applications, Inequalities for the Csiszár's f-divergence in Information Theory, (Edited by S.S. Dragomir), `http://rgmia.vu.edu.au/monographs/csiszar.htm`, (2000).

34. S.S. Dragomir, V. Gluščević and C.E.M. Pearce, *Inequality Theory and Applications*, (Edited by Y.J. Cho, J.K. Kim and S.S. Dragomir), Nova Science Publishers, Huntington, New York, (2001).