On an information divergence measure and information inequalities

Pranesh Kumar* *and* Laura Hunter

Department of Mathematics

College of Science & Management

University of Northern British Columbia

Prince George, BC V2N 4Z9, Canada

Abstract

Information divergence measures are often useful for comparing two probability distributions. An information divergence measure is suggested. This measure is symmetric in respect of the probability distributions and belongs to the well known class of Csiszár's *f*-divergences. Its properties are studied and bounds in terms of often used information divergences are obtained. A numerical illustration based on symmetric and asymmetric probability distributions compares some of these divergence measures.

MSC: 94A17, 26D15

Keywords: Symmetric, Parametric measure, Non-parametric measure, Information divergence, Csiszár's *f*-divergence.

* Corresponding author. Tel.: +250-9606671; Fax: +250-9605544.

(*kumarp@unbc.ca*)

1. Introduction

One problem in Probability Theory which interested researchers has been finding an appropriate measure of divergence (or distance or difference or discrimination or information) between two probability distributions. Non-parametric measures give the amount of information supplied by the data for discriminating in favor of a probability distribution against another, or for measuring the distance or affinity between two probability distributions. The Kullback-Leibler (1951) is the best known information divergence in this class. There are a number of divergence measures being proposed in literature which compare two probability distributions and have been applied in a variety of disciplines like, anthropology, genetics, finance, economics and political science, biology, analysis of contingency tables, approximations of probability distributions, signal processing, pattern recognition [Shannon (1958), Rényi (1961), Csiszár (1967,1974), Ali & Silvey (1966), Vajda (1972), Ferentimos & Papaiopannou (1981), Burbea & Rao (1982a,b), Taneja (1995)]. Several of these measures belong to the class of Csiszár's *f*-divergence which is defined in what follows now.

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 $f : [0, \infty) \mapsto (-\infty, \infty)]$ continuous at 0, that is, $f(0) = \lim_{u \downarrow 0} f(u)$, $\mathbb{F}_0 = \{f \in \mathbb{F} \not\Leftarrow \mho(\nVdash) = \nvdash\}$ 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 $f^*(u) = uf(\frac{1}{u})$, $u \in (0,\infty)$ and $\tilde{f} = f + f^*$.

For a convex function $f : (0,\infty) \to \mathbb{R}$, the Csiszár's f-divergence between two probability distributions P and Q is defined [Csiszár (1967,1974), Ali & Silvey (1966)]:

$$C_f(P,Q) = \sum_{x \in \Omega} \mathrm{q(x)f}\left(\frac{\mathrm{p(x)}}{\mathrm{q(x)}}\right). \quad (1.1)$$

It is well known that $C_f(P,Q)$ results in a number of popular divergence measures [Taneja (1995), Dragomir (2001), Österreicher (2002)]. Some choices of f satisfy $f(1) = 0$, so that $C_f(P,P) = 0$. Convexity ensures that divergence measure $C_f(P,Q)$ is non-negative. Some examples of the well known information divergences who belong to this class are: $f(u) = u \ln u$ provides the Kullback- Leibler measure [Kullback-Leibler (1951)], $f(u) = |u-1|$ results in the variational distance [Kolgomorov (1957,1958)], $f(u) = (u-1)^2$ yields the χ^2-divergence [Pearson (1900)].

The basic general properties of f-divergences including their axiomatic properties and some important classes are given in [Österreicher (2002)]. 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 [Leise & Vajda (1987)]:

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

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 [Leise & Vajda (1987)]:

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

v) Range of Values Theorem [Vajda (1972)]:

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

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 [Csiszár (1974)]: Given a mapping $I : \mathbb{P}^2 \mapsto (-\infty,\infty)$,

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

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

(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

$f^*(u) \equiv uf(\frac{1}{u})$.

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 commonly applied information divergences which belong to the class of $C_f(P,Q)$ are:

Variational Distance [Kolmogorov (1957,1958)]:

$$V(P,Q) = \sum |p - q|. \qquad (1.2)$$

χ^2-divergence [Pearson (1900)]:

$$\chi^2(P,Q) = \sum \frac{(p-q)^2}{q} = \sum \frac{p^2}{q} - 1. \quad (1.3)$$

Symmetric χ^2-divergence:

$$\Psi(P,Q) = \chi^2(P,Q) + \chi^2(Q,P) = \sum \frac{(p+q)(p-q)^2}{pq}. \quad (1.4)$$

Kullback & Leibler (1951):

$$K(P,Q) = \sum p \ln(\frac{p}{q}). \qquad (1.5)$$

Kullback-Leibler Symmetric Divergence:

$$J(P,Q) = K(P,Q) + K(Q,P) = \sum (p-q)\ln(\frac{p}{q}). \qquad (1.6)$$

Triangular Discrimination [Le Cam (1986), Topsϕe (1999)]:

$$\Delta(P,Q) = \sum \frac{|p-q|^2}{p+q}. \qquad (1.7)$$

Sibson Information Radius [Sibson (1969), Burbea & Rao (1982a,b)]:

$$I_r(P,Q) = \begin{cases} (r-1)^{-1}\left[\sum\left(\frac{p^r+q^r}{2}\right)\left(\frac{p+q}{2}\right)^{1-r} - 1\right], & r \neq 1, r > 0 \\ \sum \frac{p\ln p + q\ln q}{2} - \left(\frac{p+q}{2}\right)\ln\left(\frac{p+q}{2}\right), & r = 1 \end{cases}. \qquad (1.8)$$

Taneja Divergence Measure [Taneja (1995)]:

$$T_r(P,Q) = \begin{cases} (r-1)^{-1}\left[\sum\left(\frac{p^{1-r}+q^{1-r}}{2}\right)\left(\frac{p+q}{2}\right)^{r} - 1\right], & r \neq 1, r > 0 \\ \sum\left(\frac{p+q}{2}\right)\ln\left(\frac{p+q}{2\sqrt{pq}}\right), & r = 1 \end{cases}. \qquad (1.9)$$

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 [Bhattacharyya (1946)]:

$$B(P,Q) = \sum \sqrt{pq}. \qquad (1.10)$$

Hellinger Discrimination [Hellinger (1909)]:

$$h(P,Q) = \sum \frac{(\sqrt{p} - \sqrt{q})^2}{2}. \qquad (1.11)$$

Rényi Measure [Rényi (1961)]:

$$R_r(P,Q) = \begin{cases} (r-1)^{-1}\ln\left(\sum p^r q^{1-r}\right), & r \in (0,\infty)\backslash\{1\} \\ \sum p \ln \frac{p}{q}, & r = 1 \end{cases}. \qquad (1.12)$$

The following inequalities provide relationships among $V(P,Q)$, $\Delta(P,Q), K(P,Q)$ and $h(P,Q)$:

Csiszár (1967):

$$K(P,Q) \geq \frac{V^2(P,Q)}{2}. \qquad (1.13)$$

Csiszár (1967,1974):

$$K(P,Q) \geq \frac{V^2(P,Q)}{2} + \frac{V^4(P,Q)}{36}. \qquad (1.14)$$

Topsϕe (1999):

$$K(P,Q) \geq \frac{V^2(P,Q)}{2} + \frac{V^4(P,Q)}{36} + \frac{V^6(P,Q)}{270} + \frac{V^8(P,Q)}{340200}. \qquad (1.15)$$

Vajda (1972) and Toussaint (1978):

$$K(P,Q) \geq \max\{L_1(V), L_2(V)\}, \qquad (1.16)$$

where from Vajda (1972)

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

and from Toussaint (1978)

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

Topsϕe (1999):

$$\frac{1}{2}V^2(P,Q) \leq \Delta(P,Q) \leq V(P,Q). \qquad (1.17)$$

LeCam (1986) and Dacunha-Castelle (1978):

$$2h(P,Q) \leq \Delta(P,Q) \leq 4h(P,Q). \qquad (1.18)$$

Kraft (1955):

$$\frac{1}{8}V^2(P,Q) \leq h(P,Q)\left(1 - \frac{1}{2}h(P,Q)\right). \qquad (1.19)$$

Topsϕe (1999):

$$\frac{1}{8}V^2(P,Q) \leq h(P,Q) \leq \frac{1}{2}V(P,Q). \qquad (1.20)$$

and

$$K(P,Q) \leq (\log 2)V(P,Q) + \log c, \qquad (1.21)$$

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

Now in the next section 2, we discuss information inequalities for the Csiszár's *f*-divergences. Section 3 presents a new information divergence measure which belongs to the class of Csiszár's *f*-divergences and discuss its bounds in terms of some well known information divergences. In section 4, we discuss parametric information divergence which is applicable to the parametric family of distributions. A numerical study to compare new information divergence with some known divergence measures and to evaluate its bounds is done in section 5. Section 6 concludes the paper.

2. Information inequalities

Various inequalities providing bounds on the distance, information and divergence measures have been obtained recently [Dragomir (2001a,b,c,d,e), Dragomir, Gluščević & Pearce (2001), Taneja & Kumar (2004)]. Taneja & Kumar (2004) unified and generalized information bounds for $C(P,Q)$ [Dragomir (2001a,b,c,d,e)] given in the following theorem:

Theorem 2. 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 $m \leqslant x^{2-s} f''(x) \leqslant M, \quad \forall x \in (r, R), \quad s \in \mathbb{R}.$

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

$$m \Phi_s(P,Q) \leqslant C(P,Q) \leqslant M \Phi_s(P,Q), \quad (2.1)$$

and

$$m(\eta_s(P,Q) - \Phi_s(P,Q)) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant M(\eta_s(P,Q) - \Phi_s(P,Q)), \quad (2.2)$$

where

$$\Phi_s(P,Q) = \begin{cases} {}^2K_s(P,Q), & s \neq 0,1 \\ K(Q,P), & s = 0 \\ K(P,Q), & s = 1 \end{cases}, \quad (2.3)$$

$${}^2K_s(P,Q) = [s(s-1)]^{-1}[\sum p^s q^{1-s} - 1], \quad s \neq 0,1, \quad (2.4)$$

$$K(P,Q) = \sum p \ln(\tfrac{p}{q}),$$

$$C_\rho(P,Q) = C_{f'}\left(\tfrac{P^2}{Q},P\right) - C_{f'}(P,Q) = \sum (p-q) f'\left(\tfrac{p}{q}\right), \quad (2.5)$$

$$\eta_s(P,Q) = C_{\phi'_s}\left(\tfrac{P^2}{Q},P\right) - C_{\phi'_s}(P,Q)$$

$$= \begin{cases} (s-1)^{-1}\sum (p-q)\left(\tfrac{p}{q}\right)^{s-1}, & s \neq 1 \\ \sum (p-q)\ln\left(\tfrac{p}{q}\right), & s = 1 \end{cases}. \quad (2.6)$$

As a consequence of this theorem, following information inequalities which are interesting from the *information- theoretic* point of view, are also obtained in [Taneja & Kumar (2004)]:

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

$$\frac{m}{2}\chi^2(P,Q) \leqslant C(P,Q) \leqslant \frac{M}{2}\chi^2(P,Q), \quad (2.7)$$

and

$$\frac{m}{2}\chi^2(P,Q) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant \frac{M}{2}\chi^2(P,Q). \quad (2.8)$$

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

$$mK(P,Q) \leqslant C(P,Q) \leqslant MK(P,Q), \quad (2.9)$$

and

$$mK(Q,P) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant MK(Q,P). \quad (2.10)$$

(iii) The case $s = \frac{1}{2}$ yields the information bounds in terms of the Hellinger's

discrimination, $h(P,Q)$:

$$4mh(P,Q) \leqslant C(P,Q) \leqslant 4Mh(P,Q), \quad (2.11)$$

and

$$4m\left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant 4M\left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right). \quad (2.12)$$

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

χ^2-divergences:

$$mK(P,Q) \leqslant C(P,Q) \leqslant MK(P,Q), \quad (2.13)$$

and

$$m(\chi^2(Q,P) - K(Q,P)) \leqslant C_\rho(P,Q) - C(P,Q) \leqslant M(\chi^2(Q,P) - K(Q,P)). \quad (2.14)$$

In what follows now, we present a new information divergence measure which belongs

to the class of Csiszár's *f*-divergences and discuss its bounds in terms of some well known information divergences.

3. New information divergence

We consider a convex function $f : (0,\infty) \to \mathbb{R}$ be

$$f(u) = \frac{(u-1)^2}{u+1} \ln \frac{u+1}{2\sqrt{u}}. \quad (3.1)$$

Then, a new divergence measure belonging to the Csiszar's *f*-divergence family is defined as:

$$L(P,Q) = \sum \frac{(p - q)^2}{(p+q)} \ln \frac{p+q}{2\sqrt{pq}}. \qquad (3.2)$$

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

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

this measure is based on the well known *Symmetric Chi-Square*, *Arithmetic* and *Geometric Mean* divergence measures.

It may be noted that $f(u)$ in (3.1) satisfies $f(1) = 0$, so that $S(P,P) = 0$. Convexity of $f(u)$ ensures that divergence measure $S(P,Q)$ is non-negative. 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 $f^*(u) \equiv uf(\frac{1}{u}) = f(u)$, function $f(u)$ is the *-self conjugate. Therefore, all the properties (i) to (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 3.1. Let $L(P,Q)$ and $\Delta(P,Q)$ be defined as (3.2) and (1.7), respectively. Then inequality

$$L(P,Q) \leq 4\sum \frac{(p-q)^2\sqrt{pq}}{(p+q)^2} - 1.5\Delta(P,Q). \quad (3.3)$$

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

$$HM \leq AM$$
$$\text{or,} \quad \frac{2pq}{p+q} \leq \frac{p+q}{2}$$
$$\text{or,} \quad \ln \frac{p+q}{2\sqrt{pq}} \geq \ln \frac{2\sqrt{pq}}{p+q}. \quad (3.4)$$

Multiplying both sides of (3.4) by $\frac{(p-q)^2}{(p+q)}$, we have

$$\frac{(p-q)^2}{(p+q)} \ln \frac{p+q}{2\sqrt{pq}} \geq \frac{(\mathrm{p}-\mathrm{q})^2}{(p+q)} \ln \frac{2\sqrt{pq}}{p+q}. \quad 3.5)$$

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

$$\ln \frac{2\sqrt{pq}}{p+q} = \ln\left(1 + \left(\frac{2\sqrt{pq}}{p+q} - 1\right)\right) \approx \frac{4\sqrt{pq}}{p+q} - \frac{2pq}{(p+q)^2} - \frac{3}{2}. \quad (3.6)$$

Now from (3.5), (3.6) and summing over all $x \in \Omega$, we get

$$\sum \frac{(p-q)^2}{(p+q)} \ln \frac{p+q}{2\sqrt{pq}} \leq 4 \sum \frac{(p-q)^2\sqrt{pq}}{} - 1.5 \sum \frac{(p-q)^2}{p+q},$$
$$L(P,Q) \leq 4 \sum \frac{(p-q)^2\sqrt{pq}}{(p+q)^2} - 1.5\Delta(P,Q),$$

and hence the proof.

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

Proposition 3.2. Let $\chi^2(P,Q)$ and $L(P,Q)$ be defined as (2.3) and (4.2), respectively. For $P,Q \in \mathbb{P}^2$ and $0 < r \leqslant \frac{p}{q} \leqslant R < \infty$, we have

$$0 \leq L(P,Q)$$
$$\leqslant \frac{1}{4r^2(r+1)^3}\left[1 + 6r - 14r^2 + 6r^3 + r^4 + 16r^2 \ln \frac{r+1}{2\sqrt{r}}\right]\chi^2(P,Q), \ (3.7)$$

and

$$0 \leq L_\rho(P,Q) - L(P,Q)$$
$$\leqslant \frac{1}{4r^2(r+1)^3}\left[1 + 6r - 14r^2 + 6r^3 + r^4 + 16r^2 \ln \frac{r+1}{2\sqrt{r}}\right]\chi^2(P,Q), \qquad (3.8)$$

where

$$L_\rho(P,Q) = \sum \frac{(p-q)^2}{2p(p+q)^2}\left[(p-q)^2 + 2p(p+3q)\ln\frac{p+q}{2\sqrt{pq}}\right]. \qquad (3.9)$$

Proof. From the expression of $f(u)$ in (3.1), we have

$$f^{'}(u) = \frac{(u-1)}{2u(u+1)^2}\left[(u-1)^2 + 2u(3+u)\ln\frac{u+1}{2\sqrt{u}}\right], \qquad (3.10)$$

and, thus

$$L_\rho(P,Q) = \sum (\mathrm{p}-\mathrm{q})\mathrm{f}'\left(\frac{\mathrm{p}}{\mathrm{q}}\right) = \sum \frac{(p-q)^2}{2p(p+q)^2}\left[(p-q)^2 + 2p(p+3q)\ln\frac{p+q}{2\sqrt{pq}}\right].$$

Further,

$$f^{''}(u) = \frac{1}{2u^2(u+1)^3}\left[1+6u-14u^2+6u^3+u^4+16u^2\ln\frac{u+1}{2\sqrt{u}}\right]. \qquad (3.11)$$

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

$$0 \leq f^{''}(u) \leq \frac{1}{2r^2(r+1)^3}\left[1+6r-14r^2+6r^3+r^4+16r^2\ln\frac{r+1}{2\sqrt{r}}\right], \qquad (3.12)$$

where r and R are defined above. In view of (2.7), (2.8) and (3.12), we get inequalities (3.7) and (3.8), respectively.

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

Proposition 3.3. Let $K(P,Q)$, $L(P,Q)$ and $L_\rho(P,Q)$ be defined as (1.6), (3.2) and (3.9), respectively. If $P,Q \in \mathbb{P}^2$ and $0 < r \leqslant \frac{p}{q} \leqslant R < \infty$, then

$$\begin{aligned} 0 &\leq L(P,Q) \\ &\leqslant \frac{1}{2r(r+1)^3}\left[1+6r-14r^2+6r^3+r^4+16r^2\ln\frac{r+1}{2\sqrt{r}}\right]K(P,Q), \end{aligned} \qquad (3.13)$$

and

$$0 \le L_\rho(P,Q) - L(P,Q)$$
$$\leqslant \frac{1}{2r(r+1)^3}\left[1 + 6r - 14r^2 + 6r^3 + r^4 + 16r^2 \ln\frac{r+1}{2\sqrt{r}}\right]K(P,Q), \qquad (3.14)$$

Proof. Consider $f''(u)$ as given in (3.11) and let the function $g : [r,R] \to \mathbb{R}$ be such that

$$g(u) = uf''(u) = f''(u) = \frac{1}{2u(u+1)^3}\left[1 + 6u - 14u^2 + 6u^3 + u^4 + 16u^2 \ln\frac{u+1}{2\sqrt{u}}\right]. \qquad (3.15)$$

Then

$$0 \le g(u) \le \frac{1}{2r(r+1)^3}\left[1 + 6r - 14r^2 + 6r^3 + r^4 + 16r^2 \ln\frac{r+1}{2\sqrt{r}}\right], \qquad (3.16)$$

The inequalities (3.13) and (3.14) follow from (2.9), (2.10), and (3.16).

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

Proposition 3.4. Let $h(P,Q)$, $\eta_{1/2}(P,Q)$, $L(P,Q)$ and $L_\rho(P,Q)$ be defined as (1.26), (2.6), (3.2) and (3.11), respectively. For $P,Q \in \mathbb{P}^2$ and $0 < r \leqslant \frac{p}{q} \leqslant R < \infty$,

$$0 \le L(P,Q) \leqslant \frac{2}{\sqrt{R}\,(R+1)^3}\left[1 + 6R - 14R^2 + 6R^3 + R^4 + 16R^2 \ln\frac{R+1}{2\sqrt{R}}\right]h(P,Q), \qquad (3.17)$$

and

$$0 \le L_\rho(P,Q) - L(P,Q)$$
$$\leqslant \frac{2}{\sqrt{R}\,(R+1)^3}\left[1 + 6R - 14R^2 + 6R^3 + R^4 + 16R^2 \ln\frac{R+1}{2\sqrt{R}}\right]\left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right). \qquad (3.18)$$

Proof. For $f(u)$ in (3.1), we have $f''(u)$ given by (3.11). Let the function $g : [r,R] \to \mathbb{R}$ be such that

$$g(u) = u^{3/2}f''(u)$$
$$= \frac{1}{2\sqrt{u}\,(u+1)^3}\left[1 + 6u - 14u^2 + 6u^3 + u^4 + 16u^2 \ln \frac{u+1}{2\sqrt{u}}\right]. \tag{3.19}$$

Then

$$0 \leq g(u) \leq \frac{1}{2\sqrt{R}\,(R+1)^3}\left[1 + 6R - 14R^2 + 6R^3 + R^4 + 16R^2 \ln \frac{R+1}{2\sqrt{R}}\right], \tag{3.20}$$

Thus, inequalities (3.17) and (3.18)are established using (2.11),(2.12) and (3.20).

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

Proposition 3.5. Let $\chi^2(P,Q)$,$K(P,Q)$, $L(P,Q)$ and $L_\rho(P,Q)$ be defined as (1.3), (1.6), (3.2) and (3.11), respectively. If $P,Q \in \mathbb{P}^2$ and $0 < r \leqslant \frac{p}{q} \leqslant R < \infty$, then

$$0 \leq L(P,Q) \leqslant \frac{1}{2(R+1)^3}\left[1 + 6R - 14R^2 + 6R^3 + R^4 + 16R^2 \ln \frac{R+1}{2\sqrt{R}}\right]K(P,Q), \tag{3.21}$$

and

$$0 \leq L_\rho(P,Q) - L(P,Q)$$
$$\leqslant \frac{1}{2(R+1)^3}\left[1 + 6R - 14R^2 + 6R^3 + R^4 + 16R^2 \ln \frac{R+1}{2\sqrt{R}}\right]((\chi^2(Q,P) - K(Q,P)). \tag{3.22}$$

Proof. From the expression (3.1), we have $f''(u)$ as given in (3.11). Let the function $g : [r,R] \to \mathbb{R}$ be such that

$$g(u) = u^2 f''(u) = \frac{1}{2(u+1)^3}\left[1 + 6u - 14u^2 + 6u^3 + u^4 + 16u^2 \ln \frac{u+1}{2\sqrt{u}}\right]. \tag{3.23}$$

Then

$$0 \leq g(u) \leq \frac{1}{2(R+1)^3}\left[1 + 6R - 14R^2 + 6R^3 + R^4 + 16R^2 \ln \frac{R+1}{2\sqrt{R}}\right], \tag{3.24}$$

Thus, (3.21) and (3.22) follow from (2.13),(2.14) and (3.24).

In section 4, we discuss parametric information divergence measure which is applicable to the parametric family of distributions.

4. Parametric information divergence

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 .These measures are applicable to the 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 [Fisher (1925)] be

$$I_x(\theta) = \begin{cases} E_\theta\left[\frac{\partial}{\partial\theta}\log f(X,\theta)\right]^2, & \text{if } \theta \text{ is univariate} \\ \|E_\theta\left[\frac{\partial}{\partial\theta_i}\log f(X,\theta)\frac{\partial}{\partial\theta_j}\log f(X,\theta)\right]\|_{k\times k} & \text{if } \theta \text{ is } k\text{-variate} \end{cases}, \quad (4.1)$$

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.

Ferentimos & Papaiopannou (1981) suggested the following method to construct the parametric measure from the non-parametric measure:

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

$$I_x[\theta, k(\theta)] = I_x[f(x,\theta),\ f(x,k(\theta))], \qquad (4.2)$$

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

$$I_x^C[\theta, k(\theta)] = \int f(x,\theta)\phi\left(\frac{f(x,k(\theta))}{f(x,\theta)}\right)d\mu. \qquad (4.3)$$

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

Proposition 4.1. Let the convex function $\phi : (0,\infty) \to \mathbb{R}$ be

$$\phi(u) = \frac{(u-1)^2}{u+1}\ln\frac{u+1}{2\sqrt{u}}, \qquad (4.4)$$

and the corresponding non-parametric divergence measure $L(P,Q) = \sum \frac{(p-q)^2}{p+q}\ln\frac{p+q}{2\sqrt{pq}}$. Then the parametric measure $L^C(P,Q)$ is also the

non-parametric measure $L(P,Q)$.

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

$$I_x^C[\theta, k(\theta)] = \sum_{x\in\Omega} p(x)\phi\left(\frac{q(x)}{p(x)}\right). \qquad (4.5)$$

From (4.4), we have

$$\phi\left(\frac{q(x)}{p(x)}\right) = \frac{(p-q)^2}{p(p+q)}\ln\frac{p+q}{2\sqrt{pq}}, \qquad (4.6)$$

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

Then $L^C(P,Q)$ follows from (4.5) and (4.6) as

$$L^C(P,Q) := I_x^C[\theta, k(\theta)] = \sum_{x\in\Omega} \frac{(p-q)^2}{p+q}\ln\frac{p+q}{2\sqrt{pq}} = L(P,Q), \qquad (4.7)$$

and hence the proposition.

In what follows next, we carry out a numerical study to compare this measure with some known divergence measures and to evaluate its bounds.

5. Numerical illustration

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

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

[Insert Table 1]

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

$$L(P,Q) = 0.00000253,\ \Psi(P,Q) = 0.00305063,\ \chi^2(P,Q) = 0.00145837,\ \ J(P,Q) = 0.00151848.$$

It is noted that

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

The upper bound for $L(P,Q)$ based on $\chi^2(P,Q)$ divergence from (3.7):

$$\begin{aligned}\text{Upper Bound} &= \left(\frac{1}{4r^3(r+1)}\right)\left[4(r^4+r^3+r+1)\ln\frac{r+1}{2\sqrt{r}} + 3r^4 - 2r^3 - 2r^2 - 2r + 3\right]\chi^2(P,Q)\\ &= 0.000051817,\end{aligned}$$

and, thus, $0 < L(P,Q) = 0.00000253 < 0.000051817$. The length of the interval is

0.000051817.

The upper bound for $L(P,Q)$ based on $K(P,Q)$ from (3.13):

$$\text{Upper Bound} = \frac{1}{2r(r+1)^3}\left[1 + 6r - 14r^2 + 6r^3 + r^4 + 16r^2 \ln\frac{r+1}{2\sqrt{r}}\right]K(P,Q)$$
$$= 0.000083538,$$

and therefore, $0 < L(P,Q) = 0.00000253 < 0.000083538$. The length of the interval is 0.000083538.

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

[Insert Table 2]

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

$$L(P,Q) = 0.000002804,\ \Psi(P,Q) = 0.006570635,\ \chi^2(P,Q) = 0.003338836,\ \ J(P,Q) = 0.00327778$$

It is noted that

$$r\ (= 0.849782156) \leqslant \frac{p}{q} \leqslant R\ (= 1.401219652).$$

The upper bound for $L(P,Q)$ based on $\chi^2(P,Q)$ divergence from (3.7):

$$\text{Upper Bound} = \left(\frac{1}{4r^3(r+1)}\right)\left[4(r^4 + r^3 + r + 1)\ln\frac{r+1}{2\sqrt{r}} + 3r^4 - 2r^3 - 2r^2 - 2r + 3\right]\chi^2(P,Q)$$
$$= 0.0000420931,$$

and, thus, $0 < L(P,Q) = 0.000002804 < 0.0000420931$. The length of the interval is0.0000420931.

The upper bound for $L(P,Q)$ based on $K(P,Q)$ from (3.13):

$$\text{Upper Bound} = \frac{1}{2r(r+1)^3}\left[1 + 6r - 14r^2 + 6r^3 + r^4 + 16r^2 \ln\frac{r+1}{2\sqrt{r}}\right]K(P,Q)$$
$$= 0.0000702319,$$

and $0 < L(P,Q) = 0.000002804 < 0.0000702319$. The length of the interval is 0.0000702319.

Figure 1 shows the behavior of $L(P,Q)$ - [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 $L(P,Q)$.

[Insert Figure1]

6. 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*. Non-parametric 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.

Acknowledgment

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

References

Ali, S. M.; Silvey, S. D.,1966. A general class of coefficients of divergence of one distribution from another. *Jour. Roy. Statist. Soc.*, B, 28, 131-142.

Bhattacharyya, A., 1946. On some analogues to the amount of information and their uses in statistical estimation, *Sankhȳa*, 8, 1-14.

Burbea, J.; Rao, C. R.,1982. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Jour. Multi. Analysis*,12, 575-596.

Burbea, J.; Rao, C. R.,1982. On the convexity of some divergence measures based on entropy functions. *IEEE Trans. on Information Theory*, IT-28, 489-495.

Csiszár, I.,1967. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2, 299-318.

Csiszár, I.,1974. Information measures: A critical survey. *Trans. 7th Prague Conf. on Information Theory*, A, 73- 86, Academia, Prague.

Dragomir, S. S.,2001. Some inequalities for the Csiszár′s ϕ-divergence. *Inequalities for the Csiszár′s f-divergence in Information Theory*; S. S. Dragomir, Ed.;

http://rgmia.vu.edu.au/monographs/csiszar.htm

Dragomir, S. S.,2001. A converse inequality for the Csiszár′s ϕ-divergence. *Inequalities for the Csiszár′s f-divergence in Information Theory*; S. S. Dragomir, Ed.;

http://rgmia.vu.edu.au/monographs/csiszar.htm)

Dragomir, S. S., 2001. Some inequalities for (m,M)-convex mappings and applications for the Csiszár's f-divergence in information theory. *Inequalities for the Csiszár's f-divergence in Information Theory*; S; S; Dragomir, Ed.;

http://rgmia.vu.edu.au/momographs/csiszar.htm

Dragomir, S. S.,2001. 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*; S. S. Dragomir, Ed.;

http://rgmia.vu.edu.au/monographs/csiszar.htm

Dragomir, S. S.,2001. 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*; S. S. Dragomir, Ed.;

http://rgmia.vu.edu.au/monographs/csiszar.htm

Dragomir, S. S.; Gluščević, V.; Pearce, C. E. M.,2001. Approximations for the Csiszár′s *f*-divergence via mid point inequalities. *Inequality Theory and Applications*, 1; Y. J. Cho, J. K. Kim and S. S. Dragomir, Eds.; Nova Science Publishers: Huntington, New York, 139-154.

Ferentimos, K.; Papaiopannou, T.,1981. New parametric measures of information, *Information and Control*, 51, 193-208.

Fisher, R. A.,1925. Theory of statistical estimation. *Proc. Cambridge Philos. Soc.*, 22, 700-725.

Hellinger, E.,1909. Neue begründung der theorie quadratischen formen von unendlichen vielen veränderlichen. *Jour. Reine Ang. Math.*, 136, 210-271.

Kolmogorov, A. N.,1957. A new invariant for transitive dynamical system. *Dokl. Akad. Nauk USSR*, 119, 861-869.

Kolmogorov, A. N.,1958. Entropy per unit time as a metric invariant of automorphisms.*Dokl. Akad. Nauk USSR*,120, 754-758.

Kullback, S.; Leibler, A.,1951. On information and sufficiency. *Ann. Math. Statist.*, 22, 79-86.

Le Cam, L.,1986. *Asymptotic Methods in Statistical Decision Theory*. Springer Verlag, New York- Berlin.

Liese, F.; Vajda, I.,1987. Convex statistical distances. *Teubner-Texte zur Mathematik*, 95, Leipzig.

Österreicher, F.,2002. Csiszár′s *f*-divergences-Basic properties. *Res. Report Collection*, http://rgmia.vu.edu.au/monographs/csiszar.htm

Rényi, A.,1961. On measures of entropy and information. *Proc. 4th Berkeley Symp. on Math. Statist. and Prob.*, 1, 547-561, Univ. Calif. Press, Berkeley.

Salicru, M.; Taneja, I. J.,1993. Connections of generalized divergence measures with Fisher information matrix. *Information Sciences*,72, 251-269.

Shannon, C. E.,1958. A mathematical theory of communications. *Bell Syst. Tech. Jour.*, 27, 623-659.

Sibson, R.,1969. Information radius. *Z. Wahrs. und Verw Geb.*, 14, 149-160.

Taneja, I. J.,1995. New developments in generalized information measures. *Advances in Electronics and Electron Physics*; P. W. Hawkes, Ed.; 91, 37-135.

Taneja, I. J; Kumar, P., 2003. Relative information of type-s, Csiszar′s *f*-divergence and information inequalities. *Information Sciences* (In Press).

Topsϕe, F.,1999. Some inequalities for information divergence and related measures of discrimination. *Res. Rep. Collection.: RGMIA*, 2, 1, 85-98.

Vajda, I.,1972. On f-divergence and singularity of probability measures. *Period. Math. Hungar.*, 2, 223-234.

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

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

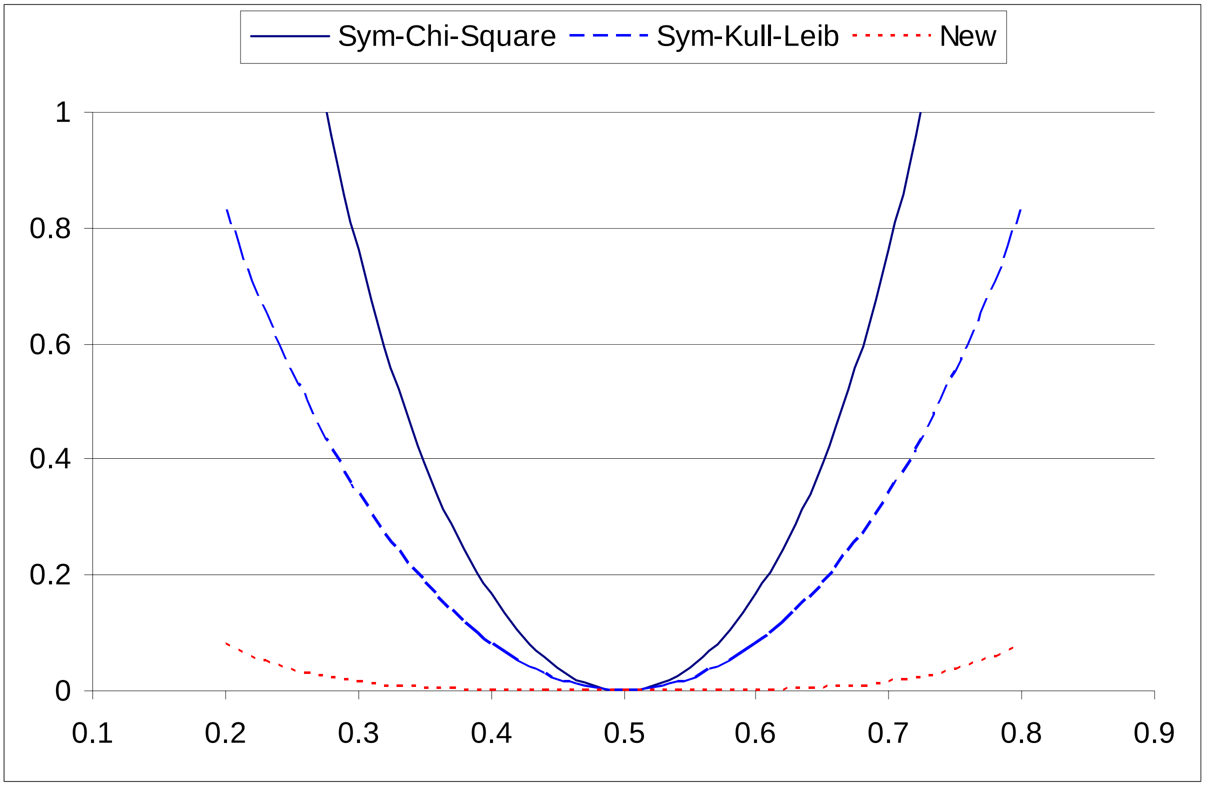

Figure 1. Measures $L(P,Q)$- New, $\Psi(P,Q)$- Sym Chi Square and $J(P,Q)$- Sym Kullback Leibler.