Journal of Inequalities in Pure and Applied Mathematics

http://jipam.vu.edu.au/

Volume 6, Issue 3, Article 65, 2005

ON A SYMMETRIC DIVERGENCE MEASURE AND INFORMATION INEQUALITIES

PRANESH KUMAR AND ANDREW JOHNSON

MATHEMATICS DEPARTMENT
COLLEGE OF SCIENCE AND MANAGEMENT
UNIVERSITY OF NORTHERN BRITISH COLUMBIA
PRINCE GEORGE BC V2N4Z9, CANADA.
kumarp@unbc.ca

johnsona@unbc.ca

Received 15 October, 2004; accepted 23 May, 2005
Communicated by S.S. Dragomir

ABSTRACT. A non-parametric symmetric measure of divergence which belongs to the family of Csiszár's f-divergences is proposed. Its properties are studied and bounds in terms of some well known divergence measures obtained. An application to the mutual information is considered. A parametric measure of information is also derived from the suggested non-parametric measure. A numerical illustration to compare this measure with some known divergence measures is carried out.

Key words and phrases: Divergence measure, Csiszár's f-divergence, Parametric measure, Non-parametric measure, Mutual information, Information inequalities.

2000 *Mathematics Subject Classification.* 94A17; 26D15.

1. INTRODUCTION

Several measures of information proposed in literature have various properties which lead to their wide applications. A convenient classification to differentiate these measures is to categorize them as: parametric, non-parametric and entropy-type measures of information [9]. 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 [10]. Non-parametric 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 [12]. Measures of entropy express the amount of information contained in a distribution, that is, the amount of uncertainty associated with the outcome of an

ISSN (electronic): 1443-5756

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

195-04

experiment. The classical measures of this type are Shannon's and Rényi's measures [15, 16]. Ferentimos and Papaioannou [9] have suggested methods for deriving parametric measures of information from the non-parametric measures and have studied their properties.

In this paper, we present a non-parametric symmetric divergence measure which belongs to the class of Csiszár's f-divergences ([2, 3, 4]) and information inequalities. In Section 2, we discuss the Csiszár's f-divergences and inequalities. A symmetric divergence measure and its bounds are obtained in Section 3. The parametric measure of information obtained from the suggested non-parametric divergence measure is given in Section 4. Application to the mutual information is considered in Section 5. The suggested measure is compared with other measures in Section 6.

2. CSISZÁR'S $f-$DIVERGENCES AND INEQUALITIES

Let $\Omega = \{x_1, x_2, \dots\}$ be a set with at least two elements and $\mathbb{P}$ the set of all probability distributions $P = (p(x) : x \in \Omega)$ on Ω. For a convex function $f : [0, \infty) \to \mathbb{R}$, the f-*divergence* of the probability distributions P and Q by Csiszár, [4] and Ali & Silvey, [1] is defined as

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

Henceforth, for brevity 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.

Österreicher [13] has discussed basic general properties of f-divergences including their axiomatic properties and some important classes. During the recent past, there has been a considerable amount of work providing different kinds of bounds on the distance, information and divergence measures ([5] – [7], [18]). Taneja and Kumar [17] unified and generalized three theorems studied by Dragomir [5] – [7] which provide bounds on $C(P,Q)$. The main result in [17] is the following theorem:

Theorem 2.1. *Let $f : I \subset \mathbb{R}_+ \to \mathbb{R}$ be a mapping which is normalized, i.e., $f(1) = 0$ and suppose that*

(i) *f is twice differentiable on (r,R), $0 \le r \le 1 \le R < \infty$, (f' and f'' denote the first and second derivatives of f),*
(ii) *there exist real constants m, M such that $m < M$ and $m \le x^{2-s} f''(x) \le M$, $\forall x \in (r,R)$, $s \in \mathbb{R}$.*

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

$$m\,\Phi_s(P,Q) \le C(P,Q) \le M\,\Phi_s(P,Q), \tag{2.2}$$

and

$$m\left(\eta_s(P,Q) - \Phi_s(P,Q)\right) \le C_\rho(P,Q) - C(P,Q) \le M\left(\eta_s(P,Q) - \Phi_s(P,Q)\right), \tag{2.3}$$

where

$$\Phi_s(P,Q) = \begin{cases} {}^2K_s(P,Q), & s \ne 0,1 \\ K(Q,P), & s = 0 \\ K(P,Q), & s = 1 \end{cases} \tag{2.4}$$

(2.5) $$ {}^{2}K_s(P,Q) = [s(s-1)]^{-1}\left[\sum p^s q^{1-s} - 1\right], \quad s \neq 0,1, $$

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

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

and

(2.8) $$ \eta_s(P,Q) = C_{\phi_s'}\left(\frac{P^2}{Q}, P\right) - C_{\phi_s'}(P,Q) = \begin{cases} (s-1)^{-1}\sum (p-q)\left(\frac{p}{q}\right)^{s-1}, & s \neq 1 \\ \sum (p-q)\ln\left(\frac{p}{q}\right), & s = 1 \end{cases}. $$

The following information inequalities which are interesting from the *information-theoretic* point of view, are obtained from Theorem 2.1 and discussed in [17]:

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

(2.9) $$ \frac{m}{2}\chi^2(P,Q) \le C(P,Q) \le \frac{M}{2}\chi^2(P,Q), $$

and

(2.10) $$ \frac{m}{2}\chi^2(P,Q) \le C_\rho(P,Q) - C(P,Q) \le \frac{M}{2}\chi^2(P,Q), $$

where

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

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

(2.12) $$ mK(P,Q) \le C(P,Q) \le MK(P,Q), $$

and

(2.13) $$ mK(Q,P) \le C_\rho(P,Q) - C(P,Q) \le MK(Q,P). $$

(iii) The case $s = \frac{1}{2}$ provides the information bounds in terms of the Hellinger's discrimination, $h(P,Q)$:

(2.14) $$ 4mh(P,Q) \le C(P,Q) \le 4Mh(P,Q), $$

and

(2.15) $$ 4\,m\left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right) \le C_\rho(P,Q) - C(P,Q) \le 4M\left(\frac{1}{4}\eta_{1/2}(P,Q) - h(P,Q)\right), $$

where

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

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

$$mK(P,Q) \leq C(P,Q) \leq MK(P,Q), \tag{2.17}$$

and

$$m\left(\chi^2(Q,P) - K(Q,P)\right) \leq C_\rho(P,Q) - C(P,Q) \leq M\left(\chi^2(Q,P) - K(Q,P)\right). \tag{2.18}$$

3. A Symmetric Divergence Measure of the Csiszár's f−Divergence Family

We consider the function $f : (0, \infty) \to \mathbb{R}$ given by

$$f(u) = \frac{\left(u^2 - 1\right)^2}{2u^{3/2}}, \tag{3.1}$$

and thus the divergence measure:

$$\Psi M(P,Q) := C_f(P,Q) = \sum \frac{\left(p^2 - q^2\right)^2}{2\left(pq\right)^{3/2}}. \tag{3.2}$$

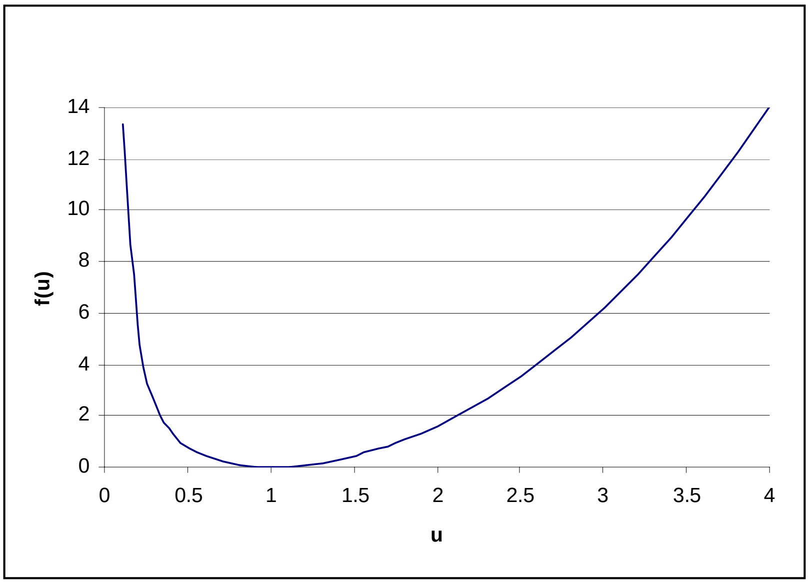
Since

$$f'(u) = \frac{\left(5u^2 + 3\right)\left(u^2 - 1\right)}{4u^{5/2}} \tag{3.3}$$

and

$$f''(u) = \frac{15u^4 + 2u^2 + 15}{8u^{7/2}}, \tag{3.4}$$

it follows that $f''(u) > 0$ for all $u > 0$. Hence $f(u)$ is convex for all $u > 0$ (Figure 3.1).



,

Figure 3.1: Graph of the convex function $f(u)$.

Further $f(1) = 0$. Thus we can say that the measure is *nonnegative* and *convex* in the pair of probability distributions $(P,Q) \in \Omega$.

Noticing that $\Psi M(P,Q)$ can be expressed as

$$\Psi M(P,Q) = \sum \left[\frac{(p+q)(p-q)^2}{pq}\right] \left[\frac{(p+q)}{2}\right] \left[\frac{1}{\sqrt{pq}}\right], \tag{3.5}$$

this measure is made up of the *symmetric chi-square, arithmetic* and *geometric mean* divergence measures.

Next we prove bounds for $\Psi M(P,Q)$ in terms of the well known divergence measures in the following propositions:

Proposition 3.1. *Let $\Psi M(P,Q)$ be as in (3.2) and the symmetric χ^2-divergence*

$$\Psi(P,Q) = \chi^2(P,Q) + \chi^2(Q,P) = \sum \frac{(p+q)(p-q)^2}{pq}. \tag{3.6}$$

Then inequality

$$\Psi M(P,Q) \geq \Psi(P,Q), \tag{3.7}$$

holds and equality, iff $P = Q$.

Proof. From the *arithmetic (AM), geometric (GM)* and *harmonic mean (HM)* inequality, that is, $HM \leq GM \leq AM$, we have

$$HM \leq GM,$$

$$\text{or,} \quad \frac{2pq}{p+q} \leq \sqrt{pq},$$

$$\text{or,} \quad \left(\frac{p+q}{2\sqrt{pq}}\right)^2 \geq \frac{p+q}{2\sqrt{pq}}. \tag{3.8}$$

Multiplying both sides of (3.8) by $\frac{2(p-q)^2}{\sqrt{pq}}$ and summing over all $x \in \Omega$, we prove (3.7). □

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

Proposition 3.2. *Let $\chi^2(P,Q)$ and $\Psi M(P,Q)$ be defined as (2.11) and (3.2), respectively. For $P, Q \in \mathbb{P}^2$ and $0 < r \leq \frac{p}{q} \leq R < \infty$, we have*

$$\frac{15R^4 + 2R^2 + 15}{16R^{7/2}} \chi^2(P,Q) \leq \Psi M(P,Q) \leq \frac{15r^4 + 2r^2 + 15}{16r^{7/2}} \chi^2(P,Q), \tag{3.9}$$

and

$$\begin{aligned} \frac{15R^4 + 2R^2 + 15}{16R^{7/2}} \chi^2(P,Q) &\leq \Psi M_\rho(P,Q) - \Psi M(P,Q) \\ &\leq \frac{15r^4 + 2r^2 + 15}{16r^{7/2}} \chi^2(P,Q), \end{aligned} \tag{3.10}$$

where

$$\Psi M_\rho(P,Q) = \sum \frac{(p-q)(p^2-q^2)(5p^2+3q^2)}{4p^{5/2}q^{3/2}}. \tag{3.11}$$

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

$$f'(u) = \frac{(u^2-1)(3+5u^2)}{4u^{5/2}}, \tag{3.12}$$

and, thus

$$\Psi M_\rho(P,Q) = \sum (p-q) f'\left(\frac{p}{q}\right) = \sum \frac{(p-q)(p^2-q^2)(5p^2+3q^2)}{4p^{5/2}q^{3/2}}. \tag{3.13}$$

Further,

$$f''(u) = \frac{15(u^4+1)+2u^2}{8u^{7/2}}. \tag{3.14}$$

Now if $u \in [a,b] \subset (0,\infty)$, then

$$\frac{15(b^4+1)+2b^2}{8b^{7/2}} \le f''(u) \le \frac{15(a^4+1)+2a^2}{8a^{7/2}}, \tag{3.15}$$

or, accordingly

$$\frac{15R^4+2R^2+15}{8R^{7/2}} \le f''(u) \le \frac{15r^4+2r^2+15}{8r^{7/2}}, \tag{3.16}$$

where r and R are defined above. Thus, in view of (2.9) and (2.10), we get inequalities (3.9) and (3.10), respectively. □

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

Proposition 3.3. *Let $K(P,Q)$, $\Psi M(P,Q)$ and $\Psi M_\rho(P,Q)$ be defined as (2.6), (3.2) and (3.13), respectively. If $P,Q \in \mathbb{P}^2$ and $0 < r \le \frac{p}{q} \le R < \infty$, then*

$$\frac{15R^4+2R^2+15}{8R^{5/2}} K(P,Q) \le \Psi M(P,Q) \le \frac{15r^4+2r^2+15}{8r^{5/2}} K(P,Q), \tag{3.17}$$

and

$$\frac{15R^4+2R^2+15}{8R^{5/2}} K(Q,P) \le \Psi M_\rho(P,Q) - \Psi M(P,Q) \le \frac{15r^4+2r^2+15}{8r^{5/2}} K(Q,P). \tag{3.18}$$

Proof. From (3.4), $f''(u) = \frac{15(u^4+1)+2u^2}{8u^{7/2}}$. Let the function $g:[r,R] \to \mathbb{R}$ be such that

$$g(u) = uf''(u) = \frac{15(u^4+1)+2u^2}{8u^{5/2}}. \tag{3.19}$$

Then

$$\inf_{u\in[r,R]} g(u) = \frac{15R^4+2R^2+15}{8R^{5/2}} \tag{3.20}$$

and

$$\sup_{u\in[r,R]} g(u) = \frac{15r^4+2r^2+15}{8r^{5/2}}. \tag{3.21}$$

The inequalities (3.17) and (3.18) follow from (2.12), (2.13) using (3.20) and (3.21). □

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 $\eta_{1/2}(P,Q)$, $h(P,Q)$, $\Psi M(P,Q)$ and $\Psi M_\rho(P,Q)$ be defined as in (2.7), (2.15), (3.2) and (3.13), respectively. For $P,Q \in \mathbb{P}^2$ and $0 < r \le \frac{p}{q} \le R < \infty$,*

$$\frac{15r^4+2r^2+15}{2r^2}h(P,Q) \le \Psi M(P,Q) \le \frac{15R^4+2R^2+15}{2R^2}h(P,Q), \tag{3.22}$$

and

$$\begin{aligned}\frac{15r^4+2r^2+15}{2r^2}&\left(\frac{1}{4}\eta_{1/2}(P,Q)-h(P,Q)\right)\\ &\le \Psi M_\rho(P,Q)-\Psi M(P,Q)\\ &\le \frac{15R^4+2R^2+15}{2R^2}\left(\frac{1}{4}\eta_{1/2}(P,Q)-h(P,Q)\right).\end{aligned} \tag{3.23}$$

Proof. We have $f''(u) = \frac{15(u^4+1)+2u^2}{8u^{7/2}}$ from (3.4). Let the function $g:[r,R]\to\mathbb{R}$ be such that

$$g(u) = u^{3/2}f''(u) = \frac{15(u^4+1)+2u^2}{8u^2}. \tag{3.24}$$

Then

$$\inf_{u\in[r,R]} g(u) = \frac{15r^4+2r^2+15}{8r^2} \tag{3.25}$$

and

$$\sup_{u\in[r,R]} g(u) = \frac{15R^4+2R^2+15}{8R^2}. \tag{3.26}$$

Thus, the inequalities (3.22) and (3.23) are established using (2.14), (2.15), (3.25) and (3.26). □

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

Proposition 3.5. *Let $K(P,Q)$, $\chi^2(P,Q)$, $\Psi M(P,Q)$ and $\Psi M_\rho(P,Q)$ be defined as in (2.5), (2.10), (3.2) and (3.13), respectively. If $P,Q \in \mathbb{P}^2$ and $0 < r \le \frac{p}{q} \le R < \infty$, then*

$$\frac{15r^4+2r^2+15}{8r^{3/2}}K(P,Q) \le \Psi M(P,Q) \le \frac{15R^4+2R^2+15}{8R^{3/2}}K(P,Q), \tag{3.27}$$

and

$$\begin{aligned}\frac{15r^4+2r^2+15}{8r^{3/2}}&(\chi^2(Q,P)-K(Q,P)\\ &\le \Psi M_\rho(P,Q)-\Psi M(P,Q)\\ &\le \frac{15R^4+2R^2+15}{8R^{3/2}}\left(\chi^2(Q,P)-K(Q,P)\right).\end{aligned} \tag{3.28}$$

Proof. From (3.4), $f''(u) = \frac{15(u^4+1)+2u^2}{8u^{7/2}}$. Let the function $g:[r,R]\to\mathbb{R}$ be such that

$$g(u) = u^2 f''(u) = \frac{15(u^4+1)+2u^2}{8u^{3/2}}. \tag{3.29}$$

Then

$$\inf_{u\in[r,R]} g(u) = \frac{15r^4+2r^2+15}{8r^{3/2}} \tag{3.30}$$

and

$$\sup_{u\in[r,R]} g(u) = \frac{15R^4+2R^2+15}{8R^{3/2}}. \tag{3.31}$$

Thus, (3.27) and (3.28) follow from (2.17), (2.18) using (3.30) and (3.31). □

4. PARAMETRIC MEASURE OF INFORMATION $\Psi M^c(P,Q)$

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 $\theta = (\theta_1, \dots \theta_k)$, the Fisher [10] information matrix be

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

where $||\cdot||_{k\times k}$ 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 $i = 1, \dots, k$;
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 and Papaioannou [9] 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

$$I_x[\theta, k(\theta)] = I_x[f(x,\theta), f(x,k(\theta))], \tag{4.2}$$

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

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

Now we have the following proposition for providing the parametric measure of information from $\Psi M(P,Q)$:

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

$$\phi(u) = \frac{(u^2-1)^2}{2u^{3/2}}, \tag{4.4}$$

and corresponding non-parametric divergence measure

$$\Psi M(P,Q) = \sum \frac{(p^2 - q^2)^2}{2\,(pq)^{3/2}}.$$

Then the parametric measure $\Psi M^C(P,Q)$

$$\Psi M^C(P,Q) := I_x^C[\theta, k(\theta)] = \sum \frac{(p^2 - q^2)^2}{2\,(pq)^{3/2}}. \tag{4.5}$$

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

$$I_x^C[\theta, k(\theta)] = \sum_{x\in\Omega} p(x)\phi\left(\frac{q(x)}{p(x)}\right). \tag{4.6}$$

From (4.4), we have

$$\phi\left(\frac{q(x)}{p(x)}\right) = \frac{\left(p^2 - q^2\right)^2}{2p^{5/2}q^{3/2}}, \tag{4.7}$$

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

Thus, $\Psi M^C(P,Q)$ in (4.5) follows from (4.6) and (4.7). □

Note that the parametric measure $\Psi M^C(P,Q)$ is the same as the non-parametric measure $\Psi M(P,Q)$. Further, since the properties of $\Psi M(P,Q)$ do not require any regularity conditions, $\Psi M(P,Q)$ is applicable to the broad families of probability distributions including the non-regular ones.

5. Applications to the Mutual Information

Mutual information is the reduction in uncertainty of a random variable caused by the knowledge about another. It is a measure of the amount of information one variable provides about another. For two discrete random variables X and Y with a joint probability mass function $p(x,y)$ and marginal probability mass functions $p(x)$, $x \in \mathbb{X}$ and $p(y)$, $y \in \mathbb{Y}$, mutual information $I(X;Y)$ for random variables X and Y is defined by

$$I(X;Y) = \sum_{(x,y)\in\mathbb{X}\times\mathbb{Y}} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}, \tag{5.1}$$

that is,

$$I(X;Y) = K\left(p(x,y), p(x)p(y)\right), \tag{5.2}$$

where $K(\cdot,\cdot)$ denotes the Kullback-Leibler distance. Thus, $I(X;Y)$ is the relative entropy between the joint distribution and the product of marginal distributions and is a measure of how far a joint distribution is from independence.

The chain rule for mutual information is

$$I(X_1,\dots,X_n;Y) = \sum_{i=1}^{n} I(X_i;Y|X_1,\dots,X_{i-1}). \tag{5.3}$$

The conditional mutual information is defined by

$$I(X;Y \mid Z) = ((X;Y)\,|Z) = H(X|Z) - H(X|Y,Z), \tag{5.4}$$

where $H(v|u)$, the conditional entropy of random variable v given u, is given by

$$H(v \mid u) = \sum\sum p(u,v) \ln p(v|u). \tag{5.5}$$

In what follows now, we will assume that

$$t \le \frac{p(x,y)}{p(x)p(y)} \le T, \text{ for all } (x,y) \in \mathbb{X}\times\mathbb{Y}. \tag{5.6}$$

It follows from (5.6) that $t \le 1 \le T$.

Dragomir, Gluščević and Pearce [8] proved the following inequalities for the measure $C_f(P,Q)$:

Theorem 5.1. *Let $f : [0,\infty) \to \mathbb{R}$ be such that $f' : [r,R] \to \mathbb{R}$ is absolutely continuous on $[r,R]$ and $f'' \in L_\infty[r,R]$. Define $f^* : [r,R] \to \mathbb{R}$ by*

$$f^*(u) = f(1) + (u-1)f'\left(\frac{1+u}{2}\right). \tag{5.7}$$

Suppose that $0 < r \leq \frac{p}{q} \leq R < \infty$. *Then*

$$\begin{aligned} |C_f(P,Q) - C_{f^*}(P,Q)| &\leq \frac{1}{4}\chi^2(P,Q)||f''||_\infty \\ &\leq \frac{1}{4}(R-1)(1-r)||f''||_\infty \\ &\leq \frac{1}{16}(R-r)^2||f''||_\infty, \end{aligned} \tag{5.8}$$

where $C_{f^*}(P,Q)$ *is the Csiszár's* f*-divergence (2.1) with* f *taken as* f^* *and* $\chi^2(P,Q)$ *is defined in (2.11).*

We define the mutual information:

$$\text{In } \chi^2\text{-sense: } I_{\chi^2}(X;Y) = \sum_{(x,y)\in\mathbb{X}\times\mathbb{Y}} \frac{p^2(x,y)}{p(x)q(y)} - 1. \tag{5.9}$$

$$\text{In } \Psi M\text{-sense: } I_{\Psi M}(X;Y) = \sum_{(x,y)\in\mathbb{X}\times\mathbb{Y}} \frac{[p^2(x,y) - p^2(x)q^2(y)]}{2[p(x)q(y)]^{3/2}}. \tag{5.10}$$

Now we have the following proposition:

Proposition 5.2. *Let* $p(x,y)$, $p(x)$ *and* $p(y)$ *be such that* $t \leq \frac{p(x,y)}{p(x)p(y)} \leq T$, *for all* $(x,y) \in \mathbb{X}\times\mathbb{Y}$ *and the assumptions of Theorem 5.1 hold good. Then*

$$\begin{aligned} &\left| I(X;Y) - \sum_{(x,y)\in\mathbb{X}\times\mathbb{Y}} [p(x,y) - p(x)q(y)] \ln\left[\frac{p(x,y)+p(x)q(y)}{2p(x)q(y)}\right] \right| \\ &\qquad \leq \frac{I_{\chi^2}(X;Y)}{4t} \leq \frac{4T^{7/2}}{t(15T^4+2T^2+15)} I_{\Psi M}(X;Y). \end{aligned} \tag{5.11}$$

Proof. Replacing $p(x)$ by $p(x,y)$ and $q(x)$ by $p(x)q(y)$ in (2.1), the measure $C_f(P,Q) \equiv I(X;Y)$. Similarly, for $f(u) = u \ln u$, and

$$f^*(u) = f(1) + (u-1)f'\left(\frac{1+u}{2}\right),$$

we have

$$\begin{aligned} I^*(X;Y) &:= C_{f^*}(P,Q) \\ &= \sum_{x\in\Omega}[p(x) - q(x)]\left[\ln\left(\frac{p(x)+q(x)}{2q(x)}\right)\right] \\ &= \sum_{x\in\Omega}[p(x,y) - p(x)q(y)]\left[\ln\left(\frac{p(x,y)+p(x)q(y)}{2p(x)q(y)}\right)\right]. \end{aligned} \tag{5.12}$$

Since $||f''||_\infty = \sup ||f''(u)|| = \frac{1}{t}$, the first part of inequality (5.11) follows from (5.8) and (5.12).

For the second part, consider Proposition 3.2. From inequality (3.9),

$$\frac{15T^4 + 2T^2 + 15}{16T^{7/2}}\chi^2(P,Q) \leq \Psi M(P,Q). \tag{5.13}$$

Under the assumptions of Proposition 5.2, inequality (5.13) yields

$$\frac{I_{\chi^2}(X;Y)}{4t} \leq \frac{4T^{7/2}}{t(15T^4+2T^2+15)} I_{\Psi M}(X;Y), \tag{5.14}$$

and hence the desired inequality (5.11). □

6. NUMERICAL ILLUSTRATION

We consider two examples of symmetrical and asymmetrical probability distributions. We calculate measures $\Psi M(P,Q)$, $\Psi(P,Q)$, $\chi^2(P,Q)$, $J(P,Q)$ and compare bounds. Here, $J(P,Q)$ is the Kullback-Leibler symmetric divergence:

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

Example 6.1 (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

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 $\Psi M(P,Q)$, $\Psi(P,Q)$, $\chi^2(P,Q)$ and $J(P,Q)$ are:

$$\Psi M(P,Q) = 0.00306097, \quad \Psi(P,Q) = 0.00305063,$$
$$\chi^2(P,Q) = 0.00145837, \quad J(P,Q) = 0.00151848.$$

It is noted that

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

The lower and upper bounds for $\Psi M(P,Q)$ from (3.9):

$$\text{Lower Bound} = \frac{15R^4+2R^2+15}{16R^{7/2}} \chi^2(P,Q) = 0.002721899$$
$$\text{Upper Bound} = \frac{15r^4+2r^2+15}{8r^{7/2}} \chi^2(P,Q) = 0.004819452$$

and, thus, $0.002721899 < \Psi M(P,Q) = 0.003060972 < 0.004819452$. The width of the interval is 0.002097553.

Example 6.2 (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

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

From the above data, measures $\Psi M(P,Q)$, $\Psi(P,Q)$, $\chi^2(P,Q)$ and $J(P,Q)$ are calculated:

$$\Psi M(P,Q) = 0.00658200, \quad \Psi(P,Q) = 0.00657063,$$
$$\chi^2(P,Q) = 0.00333883, \quad J(P,Q) = 0.00327778.$$

Note that

$$r\ (= 0.849782156) \le \frac{p}{q} \le R\ (= 1.401219652),$$

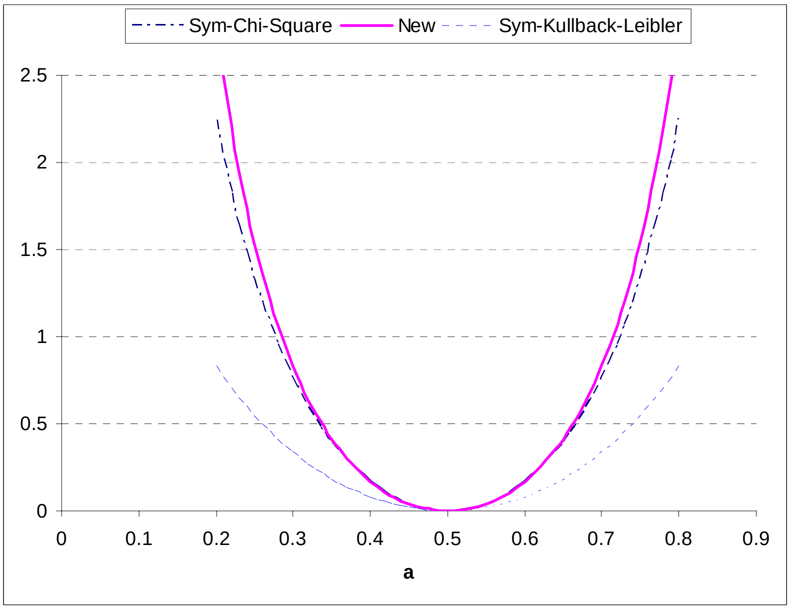
and the lower and upper bounds for $\Psi M(P,Q)$ from (4.5):

$$\text{Lower Bound} = \frac{15R^4 + 2R^2 + 15}{16R^{7/2}}\chi^2(P,Q) = 0.004918045$$
$$\text{Upper Bound} = \frac{15r^4 + 2r^2 + 15}{16r^{7/2}}\chi^2(P,Q) = 0.00895164.$$

Thus, $0.004918045 < \Psi M(P,Q) = 0.006582002 < 0.00895164$. The width of the interval is 0.004033595.

It may be noted that the magnitude and width of the interval for measure $\Psi M(P,Q)$ increase as the probability distribution deviates from symmetry.

Figure 6.1 shows the behavior of $\Psi M(P,Q)$-[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 3.1 that measures $\Psi M(P,Q)$ and $\Psi(P,Q)$ have a steeper slope than $J(P,Q)$.



,

Figure 6.1: New $\Psi M(P,Q)$, Sym-Chi-Square $\Psi(P,Q)$, and Sym-Kullback-Leibler $J(P,Q)$.

REFERENCES

[1] 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** (1966), 131–142.

[2] I. CSISZÁR, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.*, **2** (1967), 299–318.

[3] I. CSISZÁR, Information measures: A critical survey. *Trans. 7th Prague Conf. on Information Theory*, 1974, A, 73–86, Academia, Prague.

[4] I. CSISZÁR AND J. FISCHER, Informationsentfernungen in raum der wahrscheinlichkeist-verteilungen, *Magyar Tud. Akad. Mat. Kutató Int. Kösl*, **7** (1962), 159–180.

[5] 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*; S.S. Dragomir, Ed.; 2000. (http://rgmia.vu.edu.au/monographs/csiszar.htm)

[6] 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*, S.S. Dragomir, Ed.; 2000. (http://rgmia.vu.edu.au/monographs/csiszar.htm)

[7] 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*; S.S. Dragomir, Ed.; 2000. (http://rgmia.vu.edu.au/monographs/csiszar.htm)

[8] S.S. DRAGOMIR, V. GLUŠČEVIĆ AND C.E.M. PEARCE, Approximations for the Csiszár's f-divergence via mid point inequalities, in *Inequality Theory and Applications*, 1; Y.J. Cho, J.K. Kim and S.S. Dragomir, Eds.; Nova Science Publishers: Huntington, New York, 2001, 139–154.

[9] K. FERENTIMOS AND T. PAPAIOPANNOU, New parametric measures of information, *Information and Control,* **51** (1981), 193–208.

[10] R.A. FISHER, Theory of statistical estimation, *Proc. Cambridge Philos. Soc.*, **22** (1925), 700–725.

[11] E. HELLINGER, Neue begründung der theorie quadratischen formen von unendlichen vielen veränderlichen, *Jour. Reine Ang. Math.,* **136** (1909), 210–271.

[12] S. KULLBACK AND A. LEIBLER, On information and sufficiency, *Ann. Math. Statist.*, **22** (1951), 79–86.

[13] F. ÖSTERREICHER, Csiszár's f-divergences-Basic properties, *RGMIA Res. Rep. Coll.*, 2002. (http://rgmia.vu.edu.au/newstuff.htm)

[14] F. ÖSTERREICHER AND I. VAJDA, A new class of metric divergences on probability spaces and its statistical applicability, *Ann. Inst. Statist. Math*. (submitted).

[15] A. RÉNYI, On measures of entropy and information, *Proc. 4th Berkeley Symp. on Math. Statist. and Prob.*, **1** (1961), 547–561, Univ. Calif. Press, Berkeley.

[16] C.E. SHANNON, A mathematical theory of communications, *Bell Syst. Tech. Jour.*, **27** (1958), 623–659.

[17] I.J. TANEJA AND P. KUMAR, Relative information of type-s, Csiszar's f-divergence and information inequalities, *Information Sciences*, 2003.

[18] F. TOPSØE, Some inequalities for information divergence and related measures of discrimination, *RGMIA Res. Rep. Coll.*, **2**(1) (1999), 85–98.