Jordan Journal of Mathematics and Statistics (JJMS) 6(2), 2013, pp. 157-167

SOME WELL KNOWN INEQUALITIES AND ITS APPLICATIONS IN INFORMATION THEORY

K. C. JAIN [(1)] AND RAM NARESH SARASWAT[(2)]

ABSTRACT: Inequalities are very useful to information theory for new results. Inequalities are widely used for some bounds of information divergence measure of in Information Theory. There are many information and divergence measures exists in the literature on information theory and statistics. In this paper we establish some bounds of information and divergence measures using some inequalities and Csiszar's f-divergence measure.

1. INTRODUCTION

Let

$$\Gamma_n = \left\{ P = (p_{1,} p_{2,} \dots\dots p_n) \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}, n \geq 2 \tag{1.1}$$

be the set of all complete finite discrete probability distributions. There are many information and divergence measures exists in the literature of information theory and statistics. Csiszar [2] & [3] introduced a generalized measure of information using f-divergence measure given by

$$I_f(P,Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \tag{1.2}$$

2000 Mathematics Subject Classification. 62B-10, 94A-17,26D15

Keywords: Csiszar's F-Divergence Measure, Inequalities, Triangular Discrimination etc.

Received: July 13. 2012 Accepted: Feb. 5. 2013

Where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a convex function and $P, Q \in \Gamma_n$ **.** As in Csiszar,[2]. We interpret undefined expressions by

$$f(0) = \lim_{t \to 0^+} f(t), 0 f\left(\frac{0}{0}\right) = 0, 0 f\left(\frac{a}{0}\right) = \lim_{\varepsilon \to 0^+} \varepsilon f\left(\frac{a}{\varepsilon}\right) = a \lim_{t \to \infty} \frac{f(t)}{t}, a > 0$$

The Csiszar's f-divergence is a general class of divergence measures that 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, ∞). An important property of this divergence is that many known divergences can be obtained from this measure by appropriately defining the convex function f.

As to the divergence and inaccuracy of information, Kullback and Leibler (1951) [7] studied a measure of information from statistical aspects of view involving two probability distributions associated with the same experiment, calling discrimination function, later different authors named as cross entropy, relative information etc. It is a non-symmetric measure of the difference between two probability distributions P and Q.

It is well known that $I_f(P,Q)$ is a versatile functional form. Most common choices of satisfy $f(1) = 0$, so that $I_f(P,Q) = 0$. Convexity ensures that the divergence measure $I_f(P,Q)$ is always non-negative. Some examples are

- $f(u) = t\log t, \left(f^*(t) = -\log t\right)$ provides the Kullback-Leibler's measure [6, 7].
- $f(t) = (t-1)^2 \left(f^*(t) = \frac{(t-1)^2}{t} \right)$ yields the $\chi^2 -$ divergence [9] and many more.

The basic general properties of $f-$Divergences including their axiomatic properties and some important classes are given [8].

Here we shall give some examples of divergence measures in the category of Csiszar's f-divergence measure.

Kullback-Leibler divergence measure [7]:

If $f(t) = t\log t$ then Kullback and Leibler divergence measure is given by

$$I_f(P,Q) = D(P,Q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right) \tag{1.3}$$

Chi-square divergence [9]**:**

If $f(t) = (t-1)^2$ then χ^2 -Divergence measure given by

$$I_f(P,Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i} = \left[\sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1\right] = \chi^2(P,Q) \tag{1.4}$$

Hellinger Discrimination [5]**:**

If $f(\mathrm{t}) = (\sqrt{t}-1)^2$ then the Hellinger discrimination $h(P,Q)$ is given by

$$I_f(P,Q) = h(P,Q) = [1 - B(P,Q)] = \sum_{i=1}^{n} \left(\sqrt{p_i} - \sqrt{q_i}\right)^2 \tag{1.5}$$

where $B(P,Q) = \sum_{i=1}^{n} \sqrt{p_i q_i}$ is known as Bhattacharya divergence measure [1]

Relative Jensen-Shannon divergence [10]:

If $f(t) = t\log\left(\frac{2t}{t+1}\right)$ then Relative Jensen-Shannon divergence is given by

$$I_f(P,Q) = \sum_{i=1}^{n} p_i \log\left(\frac{2p_i}{p_i + q_i}\right) = F(P,Q) \tag{1.6}$$

Relative arithmetic-geometric divergence [11]:

If $f(t) = \left(\frac{t+1}{2}\right)\log\left(\frac{t+1}{2t}\right)$ then Relative arithmetic-geometric divergence is given by

$$F(P,Q) = \sum_{i=1}^{n}\left(\frac{p_i + q_i}{2}\right)\log\left(\frac{p_i + q_i}{2p_i}\right) = G(P,Q) \tag{1.7}$$

Triangular Discrimination [4]:

If $f(t) = \frac{(t-1)^2}{t+1}$ then Triangular discrimination is given by

$$I_f(P,Q) = \Delta(P,Q) = 2\left[1 - W(P,Q)\right] = \sum_{i=1}^{n}\frac{(p_i - q_i)^2}{p_i + q_i} \tag{1.8}$$

where $W(P,Q) = \sum_{i=1}^{n}\frac{2p_iq_i}{p_i + q_i}$ is known as harmonic mean divergence measure.

2. Some Well Known Inequalities:

Information and divergence measure are very useful and play an important role in many areas like as sensor networks, testing the order in a Markov chain, risk for binary experiments, region segmentation and estimation etc. Most of the achievable limits are thus stated in the form of inequalities involving fundamental measures of information such as entropy and mutual information. Such inequalities form a major tool chain to prove many results in information Theory.

In a sense, these inequalities separate the possibilities from impossibilities in Information Theory. The study of information expressions and inequalities thus are of paramount importance in solving key results in information Theory. The information measures are the usual entropy (single, joint, or conditional) and mutual information (including conditional and those involving multiple random variables). Even though it is not impossible to find a non linear expression involving these measures, they are not much of interest in Information Theory.

The following inequalities are well known in literature of pure and applied mathematics.

$$\frac{x}{1+x} \le \log(1+x) \le x, \quad x>0 \tag{2.1}$$

$$x-\frac{x^2}{2} \le \log(1+x) \le x-\frac{x^2}{2(1+x)}, \quad x>0 \tag{2.2}$$

3. Bounds Of Information Divergence Measure Using Inequalities:

In this section we shall discus some bounds and relations among well known information divergence measure which are may be interested in information theory and statistics using inequalities of (2.1) & (2.2).

Proposition 3.1: Let $P,Q\in\Gamma_n$ be two probability distributions then we have the following inequalities

$$W(P,Q)+2F(P,Q)\le 2 \text{ and } 0\le F(Q,P) \tag{3.1}$$

$$F(Q,P)\le \frac{1}{4}\left[2+\Delta(P,Q)\right] \tag{3.2}$$

Proof:

Using the inequality (2.1)

$$\frac{x}{1+x} \le \log(1+x) \le x, \quad x>0$$

Taking $x=\frac{p_i}{q_i}$ then we get

$$\frac{p_i}{p_i+q_i} \le \log\frac{p_i+q_i}{q_i} \le \frac{p_i}{q_i} \tag{3.3}$$

Multiplying by $2q_i$ of equation (3.3) and Taking summation both side, then we get

$$\sum_{i=1}^{n} \frac{2p_i q_i}{p_i + q_i} \leq 2\log 2 \sum_{i=1}^{n} q_i + 2\sum_{i=1}^{n} q_i \log \frac{p_i + q_i}{2q_i} \leq 2\sum_{i=1}^{n} p_i$$

But we know $W(P,Q) = \sum_{i=1}^{n} \frac{2p_i q_i}{p_i + q_i}$ and $F(Q,P) = \sum_{i=1}^{n} q_i \log\left(\frac{2q_i}{p_i + q_i}\right)$ then we get

$$W(P,Q) \leq 2\log 2 - 2F(Q,P) \leq 2$$

$$W(P,Q) + 2F(Q,P) \leq 2\log_2 2 \leq 2 + 2F(Q,P)$$

$$W(P,Q) + 2F(Q,P) \leq 2 \leq 2 + 2F(Q,P) \tag{3.4}$$

We get the following relations using equation (3.4)

$$W(P,Q) + 2F(Q,P) \leq 2 \quad \text{and} \quad 0 \leq F(Q,P)$$

Hence proved of the relations (3.1)

But

$$\Delta(P,Q) = 2\left[1 - W(P,Q)\right] = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i}$$

$$W(P,Q) = 1 - \frac{1}{2}\Delta(P,Q) \quad \text{put in (3.1)}$$

$$1 - \frac{1}{2}\Delta(P,Q) + 2F(Q,P) \leq 2\log 2$$

$$F(Q,P) \leq \frac{1}{4}\left[2 + \Delta(P,Q)\right], \qquad \therefore\ 2\log_2 2 = 2$$

Hence proved of the result (3.3).

Proposition 3.2:

Let $P, Q \in \Gamma_n$ be two probability distributions then we have the following inequalities

$$\left|F(Q,P) - \chi^2(P,Q)\right| \leq \frac{3}{2} \tag{3.5}$$

Proof:

Using the first and second relation of inequality (2.2)

$$x - \frac{x^2}{2} \le \log(1+x) \quad x > 0$$

Now we take $x = \frac{p_i}{q_i}$ and taking summation both side

$$\sum_{i=1}^{n} \frac{p_i(2p_i - q_i)}{2q_i} \ge \sum_{i=1}^{n} q_i \log\left(\frac{2q_i}{p_i+q_i}\right) - \log 2$$

$$\sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1 + \frac{1}{2} \ge \sum_{i=1}^{n} q_i \log\left(\frac{2q_i}{p_i+q_i}\right) - \sum_{i=1}^{n} \log 2$$

$$\chi^2(P,Q) + \frac{1}{2} \ge F(Q,P) - \log 2$$

$$\because\ \chi^2(P,Q) = \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1, \quad F(Q,P) = \sum_{i=1}^{n} q_i \log\left(\frac{2q_i}{p_i+q_i}\right)$$

$\left|F(Q,P) - \chi^2(P,Q)\right| \le \frac{3}{2}$, hence proved of the result (3.5).

Proposition 3.3:

Let $P,Q \in \Gamma_n$ be two probability distributions then we have the following inequalities

$$0 \le G(Q,P) + A(P,Q) \tag{3.6}$$

$$G(P,Q) + \frac{1}{2}\left[D(P,Q) - D(Q,P)\right] \ge \log \frac{1}{2}\left[h(P,Q) + B(P,Q)\right] \tag{3.7}$$

Proof:

Using the first and last relation of inequality (2.2)

$$x - \frac{x^2}{2} \le x - \frac{x^2}{2(1+x)}, \quad x > 0 \tag{3.8}$$

Now we take $x=\frac{p_i}{q_i}$ then we get

$$\frac{p_i}{q_i}-\frac{p_i^2}{2q_i^2}\le\frac{p_i}{q_i}-\frac{p_i^2}{2q_i(p_i+q_i)}$$

$$\frac{(p_i+q_i)}{2q_i}\ge\frac{1}{2}$$

Taking log both side, multiply by $\frac{p_i+q_i}{2}$ and taking summation both side

$$\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\left(\frac{p_i+q_i}{2q_i}\right)\ge\log\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i+q_i)}{2} \tag{3.9}$$

$$G(Q,P)\ge\log\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i+q_i)}{2},\quad G(Q,P)+A(P,Q)\ge 0$$

Hence proved of the result (3.6)

But we know

$$h(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{p_i}-\sqrt{q_i}\right)^2=\sum_{i=1}^{n}\frac{p_i+q_i}{2}-\sum_{i=1}^{n}\sqrt{p_iq_i}=A(P,Q)-B(P,Q)$$

$$h(P,Q)+B(P,Q)=\sum_{i=1}^{n}\frac{p_i+q_i}{2}=A(P,Q) \tag{3.10}$$

Where $h(P,Q)$ & $B(P,Q)$ are the Hellinger discrimination & Bhattacharya divergence.

From equation (3.9) & (3.10)

$$\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\left(\frac{p_i+q_i}{2p_i}\right)+\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\frac{p_i}{q_i}\geq\log\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i+q_i)}{2}$$

$$G(P,Q)+\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\frac{p_i}{q_i}\geq\log\frac{1}{2}\left[h(P,Q)+B(P,Q)\right]$$

$$G(P,Q)+\frac{1}{2}\left[\sum_{i=1}^{n}p_i\log\frac{p_i}{q_i}-\sum_{i=1}^{n}q_i\log\frac{q_i}{p_i}\right]\geq\log\frac{1}{2}\left[h(P,Q)+B(P,Q)\right]$$

$G(P,Q)+\frac{1}{2}\left[D(P,Q)-D(Q,P)\right]\geq\log\frac{1}{2}\left[h(P,Q)+B(P,Q)\right]$, hence proved of the result (3.7)

Proposition 3.4:

Let $P,Q\in\Gamma_n$ be two probability distributions then we have the following inequalities

$$\left|2G(P,Q)-\chi^2(P,Q)\right|\leq\frac{1}{2}\sum_{i=1}^{n}p_i(6-p_i)-2\log 2 \tag{3.11}$$

Proof:

Using the second and last relation of inequality (2.2)

$$\log(1+x)\leq x-\frac{x^2}{2(1+x)},\quad x>0 \tag{3.12}$$

Now we take $x=\frac{p_i}{q_i}$ and taking summation both sides, then we get

$$\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\left(\frac{p_i+q_i}{2q_i}\right)+\log 2\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\leq\sum_{i=1}^{n}\frac{p_i^2}{2q_i}-\sum_{i=1}^{n}\frac{p_i}{2}+\sum_{i=1}^{n}\left(\frac{3p_i}{2}-\frac{p_i^2}{4}\right)$$

$$G(P,Q)+\log 2\sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\leq\frac{1}{2}\sum_{i=1}^{n}\left(\frac{p_i^2}{q_i}-1\right)+\sum_{i=1}^{n}\left(\frac{3p_i}{2}-\frac{p_i^2}{4}\right)$$

$$G(P,Q) + \log 2 \le \frac{1}{2}\chi^2(P,Q) + \sum_{i=1}^{n}\left(\frac{3p_i}{2} - \frac{p_i^2}{4}\right) \tag{3.13}$$

$$\because \quad G(P,Q) = \sum_{i=1}^{n}\left(\frac{p_i+q_i}{2}\right)\log\left(\frac{p_i+q_i}{2q_i}\right), \quad \chi^2(P,Q) = \sum_{i=1}^{n}\left(\frac{p_i^2}{q_i} - 1\right)$$

$$\left|2G(P,Q) - \chi^2(P,Q)\right| \le \frac{1}{2}\sum_{i=1}^{n} p_i(6-p_i) - 2\log 2$$, hence proved of the result (3.11)

REFERENCES

[1]. Bhattacharya A., “Some analogues to amount of information and their uses in statistical estimation” Sankhya 8 (1946), 1-14.

[2]. Csiszar I. Information measure,“A critical servey.Trans.7th prague conf. on info. Th. Statist. Decius. Funct, Random Processes” and 8th European meeting of statist Volume B. Acadmia Prague, (1978), PP-73-86.

[3]. Csiszar I., “Information-type measures of difference of probability functions and indirect observations” studia Sci.Math.hunger.2 (1961).299-318.

[4]. Dacunha-Castella D., “Ecole d’Ete de probabilities de Saint”-Flour VII-1977 Berline, Heidelberg, New-York:Springer, (1978).

[5]. Hellinger E., “Neue Begrundung der Theorie der quadratischen Foemen von unenlichen vielen Veranderlichen” J. Reine Aug. Math., 136 (1909), 210-271.

[6]. Kulback, S"Information theory & Statistics", Dover Publications, . (1968), New York.

[7]. Kullback S. and R.A Leibler: “On information and sufficiency”. Ann. Nath. Statistics 22(1951), 79-86.

[8]. Österreicher F. “Csiszár’s f-divergences – Basic Properties”, Res. Report Collection (2002), http://rgmia.vu.edu.au/monographs/Csiszár.htm.

[9]. Pearson K., “On the criterion that a give system of deviations from the probable in the case of correlated system of variables in such that it can be reasonable supposed to have arisen from random sampling”, Phil. Mag., 50 (1900),157-172.

[10]. Sibson R., “Information Radius”, Z, Wahrs. undverw.geb. (14) (1969), 149-160.

[11]. Taneja I.J., “New Developments in generalized information measures”, Chapter in: Advances in imaging and Electron Physics, Ed. P. W. Hawkes 91 (1995), 37-135.

Department of Mathematics, Malaviya National Institute of Technology, Jaipur (Rajasthan) - 302017, India**.**

(1) *E-mail address*: jainkc.2003@gmail.com

(2) *E-mail address*: sarswatrn@gmail.com