Series of Information Divergence Measures Using New F-Divergences, Convex Properties and Inequalities

K. C. Jain*, Ram Naresh Saraswat

Department of Mathematics,Malaviya National Institute of Technology,Jaipur (Rajasthan) -302017, INDIA

ABSTRACT: *There are several types of information divergence measures studied in literature of information theory which compare two probability distributions and have applications in information theory, statistics and engineering. In this paper, we derive some families of divergence measures using properties of convex functions and new f-divergence measure. Bounds of new divergences in terms of well-known divergence measure are also considered.*

Additional Key words and Phrases: *Triangular discrimination, symmetric divergence measure, Csiszar,s f-divergence etc.*

AMS Classification 62B-10, 94A-17, 26D15

I. INTRODUCTION

Let

$$\Gamma_n = \left\{ P = (p_1, p_2, \dots\dots p_n) \middle| p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}, n \geq 2 \quad (1.1)$$

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 [1] & [2] introduced a generalized measure of information using f-divergence measure is given by

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

where $f : \mathbf{R}_+ \to \mathbf{R}_+$ is a convex function and $P, Q \in \Gamma_n$.

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 $(0, \infty)$. An important property of this divergence is that many known divergences can be obtained from this measure by appropriately defining the convex function f . These measures have been applied in a variety of fields such as anthropology, genetics, finance, economics, analysis of contingency tables, approximations of probability distributions, signal processing & pattern recognition.

II. NEW F-DIVERGENCE MEASURE OF INFORMATION

Now, we shall consider some properties of a new f-divergence measure and its particular cases which are may be interesting in areas of information theory. Jain and Saraswat, [5, 6] introduced a new f-divergence measure is given by

$$S_f(P,Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i + q_i}{2q_i}\right) \quad (2.1)$$

Where $f : \mathbf{R}_+ \to \mathbf{R}_+$ is a convex function and $P, Q \in \Gamma_n$.

It is shown that using new f-divergence measure we derive some well known divergence measures such as Chi-square divergence, Triangular discrimination and variational distance in this section. Other well-known divergence measures are also derived in [6] using new f-divergence measure. We derived some new series of divergence measures using properties of convex functions in section 3. Using additional properties of convex functions we derived some divergence measure of information in section 4. Some bounds of other new divergence measure in terms of various well-known divergence measures like as Triangular discrimination, Hellinger discrimination and variational distance, in section 5 are also considered.

The following results are also presented by Jain and Saraswat [6].

Proposition 2.1 Let $f : [0, \infty) \to \mathbf{R}$ be convex and $P, Q \in \Gamma_n$ then we have the following inequality

$$S_f(P,Q) \geq f(1) \quad (2.1)$$

Equality holds in (2.1) iff

$$p_i = q_i \ \forall i = 1, 2, .., n \quad (2.2)$$

Corollary 2.1.1 (Non-negativity of new f-divergence measure) Let $f:[0,\infty)\to\mathbf{R}$ be convex and normalized, i.e.

$$f(1)=0 \tag{2.3}$$

Then for any $P,Q\in\Gamma_n$ from (2.1) of proposition 2.1 and (2.3), we have the inequality

$$S_f(P,Q)\geq 0 \tag{2.4}$$

If f is strictly convex, equality holds in (2.5) iff

$$p_i=q_i \quad \forall\, i\in[i,2,\ldots\ldots\ldots\ldots..n] \tag{2.5}$$

and

$$S_f(P,Q)\geq 0 \;\text{ and }\; S_f(P,Q)=0 \;\text{ iff }\; P=Q \tag{2.6}$$

Proposition 2.2 Let $f_1\, \&\, f_2$ are two convex functions and $g=a\,f_1+b\,f_2$ then $S_g(P,Q)=a\,S_{f_1}(P,Q)+b\,S_{f_2}(P,Q)$, where $a\,\&\,b$ are constants and $P,Q\in\Gamma_n$

We now give some examples of well known information divergence measures which are obtained from new f-divergence measure.

- **Chi-square divergence measure** [9]: - If $f(t)=4(t-1)^2$ then Chi-square divergence measure is given by

$$S_f(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{q_i}=\left[\sum_{i=1}^{n}\frac{p_i^2}{q_i}-1\right]=\chi^2(P,Q) \tag{2.7}$$

- **Triangular discrimination** [4]: - If $f(t)=\frac{2(t-1)^2}{t},\ \forall t>0$ then Triangular discrimination is given by

$$S_f(P,Q)=2\left[1-W(P,Q)\right]=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{(p_i+q_i)}=\Delta(P,Q) \tag{2.8}$$

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

- **Variational Distance** [4, 8]:- Let if $f(t)=2\left|t-1\right|$ then Chi-square divergence measure is given by

$$S_f(P,Q)=\sum_{i=1}^{n}\left|p_i-q_i\right|=V(P,Q) \tag{2.9}$$

III. NEW INFORMATION DIVERGENCE MEASURES

In this section we shall find out the new information measures with help of the following convex function. Now we consider the function $f:(0,\infty)\to\mathbf{R}$ given by

$$f_k(t)=t\left(1-\frac{1}{t}\right)^{2k},\qquad k=1,2,3,\ldots\ldots\ldots\ldots\ldots\ldots.. \tag{3.1}$$

Since

$$f_k{}'(t)=\left(1-\frac{1}{t}\right)^{2k-1}\left[\frac{(2k-1)+t}{t}\right] \tag{3.2}$$

and $$f_k{}''(t)=\frac{(t-1)^{2k-2}\left[2k(2k-1)\right]}{t^{2k+1}} \tag{3.3}$$

Function $f_k(t)$ is always convex if k = 1, 2, 3,..positive integers, $\forall t>0$.

$$S_f(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^{2k}}{(p_i+q_i)^{2k-1}}=N_k^c(P,Q)$$

Also we get the following

$$F=\sum_{i=1}^{n} f_k(t)=t\sum_{i=1}^{\infty}\left(1-\frac{1}{t}\right)^{2k} =t\left(1-\frac{1}{t}\right)^{2}\left[1+\left(1-\frac{1}{t}\right)^{2}+--\infty\right]=\frac{(t-1)^2}{t}\left[1+\left(1-\frac{1}{t}\right)^{2}+--\infty\right]$$

Sum always convex for infinity if $\forall t>\frac{1}{2}$

Now we have the following series of convex functions if k=1, 2, 3, 4…of (3.1)

$$\frac{(t-1)^2}{t},\ \frac{(t-1)^4}{t^3},\ \frac{(t-1)^6}{t^5},\ \frac{(t-1)^8}{t^7},\ldots\ldots\ldots\ldots$$

Further we know that if $f_1(t), f_2(t), f_3(t), f_4(t)$................. are convex functions then the function $c_1 f_1(t)+c_2 f_2(t)+c_3 f_3(t)+c_4 f_4(t)+\ldots$ is also convex functions. Where c_1,c_2,c_3,c_4.......... are positive constants such that at least one c_i is not equal to zero. Now taking

$$c_1=1,\ c_2=1,\ c_3=\frac{1}{2!},\ c_4=\frac{1}{3!}\ldots\ldots\ldots$$

and $f_1(t)=\dfrac{(t-1)^2}{t},\quad f_2(t)=\dfrac{(t-1)^4}{t^3},\quad f_3(t)=\dfrac{(t-1)^6}{t^5},\ f_4(t)=\dfrac{(t-1)^8}{t^7}\ldots\ldots\ldots$

We have the following series of convex functions

$$\frac{(t-1)^2}{t}+\frac{(t-1)^4}{t^3}+\frac{(t-1)^6}{2!t^5}+\frac{(t-1)^8}{3!t^7}\ldots\ldots\ldots$$

$$=\frac{(t-1)^2}{t}\left[1+\frac{(t-1)^2}{t^2}+\frac{(t-1)^4}{2!t^4}+\frac{(t-1)^6}{3!t^6}\ldots\ldots\ldots\right]$$

$$g(t)=\frac{(t-1)^2}{t}\exp\left\{\frac{(t-1)^2}{t^2}\right\} \tag{3.4}$$

where exp{.} denotes the exponential function and applying in following function new f-divergence property of (2.2)

$$S_f(P,Q)=g(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i}\exp\left\{\frac{(p_i-q_i)^2}{(p_i+q_i)^2}\right\} \tag{3.5}$$

$$g(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{p_i+q_i}\exp\left\{\frac{(p_i-q_i)^2}{2(p_i+q_i)}.\frac{2}{(p_i+q_i)}\right\} \tag{3.6}$$

Divergence measure $g(P,Q)$ is the combination of triangular and arithmetic divergence measures, similarly, we get

$$c_1=1,c_2=1,c_3=\frac{1}{2!},c_4=\frac{1}{3!}\ldots\ldots\ldots$$

and $f_1(t)=\dfrac{(t-1)^4}{t^3},\ f_2(t)=\dfrac{(t-1)^6}{t^5},\ f_3(t)=\dfrac{(t-1)^8}{t^7},\ f_4(t)=\dfrac{(t-1)^{10}}{t^9}\ldots\ldots\ldots$

then we obtain the following divergence measure of Csiszar's f-divergence class

$$=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^4}{(p_i+q_i)^3}\exp\left\{\frac{(p_i-q_i)^2}{(p_i+q_i)^2}\right\} \tag{3.7}$$

Similarly an appropriate selection of constants and convex functions will result in following series of convex functions

$$f_k^*(t)=\frac{(t-1)^{2k}}{t^{2k-1}}\exp\left\{\frac{(t-1)^2}{t^2}\right\},\quad k=1,2,3,4,\ldots\ldots\ldots \tag{3.8}$$

and the following series of divergence measures of Csiszar's f-divergence class

$$N_k^*(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^{2k}}{(p_i+q_i)^{2k-1}}\exp\left\{\frac{(p_i-q_i)^2}{(p_i+q_i)^2}\right\} \tag{3.9}$$

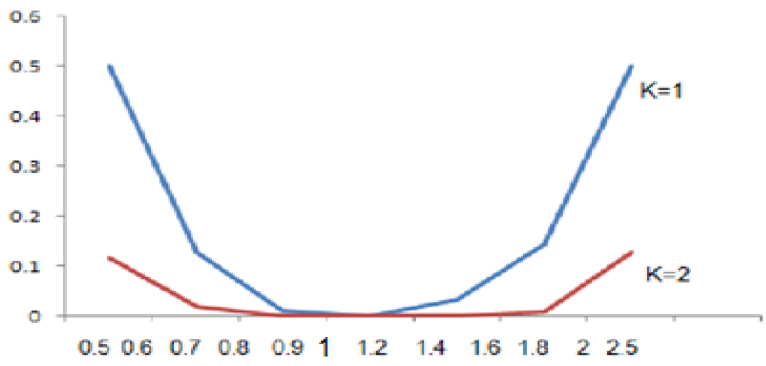

Fig.3.1 graph of the convex function $f_k(t)$

It is clear that from the above graph that the convex functions $f_k^*(t)$ gives a steeper slope with increase in value of k. Further $f_k(1)=0$,so that $N_k^*(P,P)=0$ and the convexity of the function $f_k(t)$ ensure that the measure (3.3) is non-negative. Thus we can say that the measure (3.3) non-negative and convex in the pair of probability distributions $(P,Q)\in\Gamma_n$. Since

$$N_1^*(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{(p_i+q_i)}\exp\left\{\frac{(p_i-q_i)^2}{(p_i+q_i)^2}\right\} \quad (3.10)$$

$$N_2^*(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^4}{(p_i+q_i)^3}\exp\left\{\frac{(p_i-q_i)^2}{(p_i+q_i)^2}\right\} \quad (3.11)$$

$$N_3^*(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^6}{(p_i+q_i)^5}\exp\left\{\frac{(p_i-q_i)^2}{(p_i+q_i)^2}\right\} \quad (3.12)$$

….and so on.

Therefore we can say that the measure $N_k^*(P,Q)$ is made up generalized series of combinations of triangular and arithmetic divergence measure.

IV. SOME NEW OTHERS INFORMATION DIVERGENCE MEASURES

In this section, we will derive some new information divergence measures using the convexity property of divergence measures. We proceed as follows.

Since the sum of two convex functions is again a convex function, therefore we have the following convex functions.

$$\frac{(t-1)^2}{t}+\frac{(t-1)^4}{t^3}=\frac{(t-1)^2(2t^2-2t+1)}{t^3} \quad (4.1)$$

$$\frac{(t-1)^4}{t^3}+\frac{(t-1)^6}{t^5}=\frac{(t-1)^4(2t^2-2t+1)}{t^5}. \quad (4.2)$$

$$\frac{(t-1)^6}{t^5}+\frac{(t-1)^8}{t^7}=\frac{(t-1)^6(2t^2-2t+1)}{t^5} \quad (4.3)$$

..and so on.

The following divergence measures are obtained using new f-divergence measure is given by

$$N_1(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^2(p_i^2+q_i^2)}{(p_i+q_i)^3} \quad (4.4)$$

$$N_2(P,Q)=\frac{1}{2}\sum\frac{(p_i-q_i)^4(p_i^2+q_i^2)}{(p_i+q_i)^5} \quad (4.5)$$

$$N_3(P,Q)=\frac{1}{2}\sum_{i=1}^{n}\frac{(p_i-q_i)^6(p_i^2+q_i^2)}{(p_i+q_i)^7} \quad (4.6)$$

….and so on.

Similarly we can generate various other series of divergence measures using the properties of convex functions. Further results about these divergence measures will be discussed elsewhere.

V. RELATIONS AMONG NEW INFORMATION DIVERGENCE AND OTHER WELL KNOWN DIVERGENCE MEASURES

In this section we will drive inequalities relating $N_k^*(P,Q)$ (for the case k=1, k=2…) with the above divergence measures.

To start with, we will derive inequalities for the divergence measures given by $\Delta_1(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2}{(p_i+q_i)}$,

$$\Delta_2(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^4}{(p_i+q_i)^3},....,\Delta_k(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^{k+1}}{(p_i+q_i)^k} \tag{5.1}$$

The above divergence measures (5.1) may be say the generalized triangular divergence measure of type k. Then we will use these inequalities for relating $N_k^*(P,Q)$ with other divergence measure.

Now again consider the inequality

$$x+1\le\exp\{x\}$$

Replacing x by $\frac{(t-1)^2}{t^2}$

$$\frac{(t-1)^2}{t^2}+1\le\exp\left\{\frac{(t-1)^2}{t^2}\right\}$$

$$\frac{2t^2-2t+1}{t^2}\le\exp\left\{\frac{(t-1)^2}{t^2}\right\}$$

$$\frac{(t-1)^2(2t^2-2t+1)}{t^3}=\frac{(t-1)^2}{t}+\frac{(t-1)^4}{t^3}\le\frac{(t-1)^2}{t}\exp\left\{\frac{(t-1)^2}{t^2}\right\}$$

$$\frac{(t-1)^2}{t}+\frac{(t-1)^4}{t^3}\le\frac{(t-1)^2}{t}\exp\left\{\frac{(t-1)^2}{t^2}\right\}$$

$$N_1(P,Q)=\sum_{i=1}^{n}\frac{(p_i-q_i)^2(p_i^2+q_i^2)}{(p_i+q_i)^3}\le N_1^*(P,Q) \tag{5.2}$$

Now replacing t by $\frac{p+q}{2q}$, multiplying q and finally summing over all t in the above inequality, we obtain

$$N_1(P,Q)\le N_1^*(P,Q) \tag{5.3}$$

$$\Delta(P,Q)\le N_1(P,Q)\le N_1^*(P,Q) \tag{5.4}$$

$\because\quad \Delta(P,Q)\le\chi_1^2(Q,P)$ [Jain & Saraswat, 5]

$$1/2V^2(P,Q)\le\Delta(P,Q)\le N_1(P,Q)\le N_1^*(P,Q) \tag{5.5}$$

$\because\quad \frac{1}{2}V^2(P,Q)\le\Delta(P,Q)$ [Topse, 9]

$$2h(P,Q)\le\Delta(P,Q)\le N_1(P,Q)\le N_1^*(P,Q) \tag{5.6}$$

$\because\quad 2h(P,Q)\le\Delta(P,Q)$ [Dacunha-Castelle, 3] where $h(P,Q)$ is Hellinger discrimination [4]

REFERENCES

[1] **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

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

[3] **Dacunha D.-Castelle,** Ecolee d'Ete de Probabilites de Saint-Flour VII-1977, Berlin, Heidlberg, New York: Springer, 1978

[4] **Hellinger E.,** Neue Begrundung der Theorie der quadratischen Formen Von unendlichen vielen veranderlichen,J.Rein.Aug.Math,136(1909),210-271.

[5] **Jain K. C. and R. N. Saraswat** "A New Information Inequality and its Application in Establishing Relation among various f-Divergence Measures", Journal of Applied Mathematics, Statistics and Informatics (JAMSI), 8(1) (2012), 17-32.

[6] **Jain K. C. and R. N. Saraswat** "Some bounds of information divergence measure in term of Relative arithmetic-geometric divergence measure" International Journal of Applied Mathematics and Statistics, Volume 32, Number 2 processing (2013), pp. 48-58.

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

[8] **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.

[9] **Topse F.** Some inequalitiesfor information divergence and related measure of discrimination. Res. Rep. Collection, RGMIA,2,1,85-98