SENRA Academic Publishers, British Columbia
Vol. 8, No. 2, pp. 2905-2912, June 2014
Online ISSN: 1920-3853; Print ISSN: 1715-9997

Canadian Journal of pure&applied sciences

NEW MEASURES OF INFORMATION AND THEIR APPLICATIONS IN CODING THEORY

*Om Parkash and Priyanka Kakkar
[1]Department of Mathematics, Guru Nanak Dev University, Amritsar- 143005, India

ABSTRACT

New measures of information including entropy, directed divergence and inaccuracy along with their generalizations have been introduced and their essential and desirable properties are studied. The relations between newly developed measures of directed divergence and the well-known standard measure of divergence existing in the literature of distance measures usually known as Kullback-Leibler's measure have been established. Applications of these measures are provided to the field of coding theory for the study of source coding theorems.

Keywords: Entropy, directed divergence, inaccuracy, mean codeword length, uniquely decipherable code.

INTRODUCTION

Shannon (1948) founded the subject of information theory which is closely related to thermodynamics and physics through the similarity of Shannon's uncertainty measure to the entropy function. It was then realized that entropy is a property of any stochastic system and the concept is now used widely in different disciplines. The tendency of the systems to become more disordered over time is described by the second law of thermodynamics, which states that the entropy of the system cannot spontaneously decrease. Today, information theory is still principally concerned with communications systems, but there are widespread applications in statistics, information processing and computing. Shannon (1948) entropy, also known as measure of uncertainty for a probability distribution $P = (p_1, p_2, ..., p_n)$ is given by

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i \tag{1.1}$$

with the convention that $0\log 0 := 0$. It is to be noted that the base of logarithm is assumed to be 2, unless until specified.

Besides entropy, another basic and fundamental concept usually applied in information theory is that of divergence. The most important and desirable measure of divergence associated with the probability distributions $P = (p_1, p_2, ..., p_n)$ and $Q = (q_1, q_2, ..., q_n)$ is due to Kullback and Leibler (1951) and is given by

$$K(P:Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{1.2}$$

Another basic concept in information theory which connects the above two measures mathematically, that is, entropy and divergence, is that of inaccuracy. This concept is basically associated with two probability distributions $P = (p_1, p_2, ..., p_n), Q = (q_1, q_2, ..., q_n)$ where Q is predicted and P is true probability distribution. This fundamental concept was proposed by Kerridge (1961) and is given by

$$H(P:Q) = \sum_{i=1}^{n} p_i \log q_i \tag{1.3}$$

The above mentioned information measures find their applications in a variety of disciplines such as genetics, finance, economics, political science, biology, analysis of contingency of tables, statistics, signal processing and pattern recognition. It is to emphasis here that the above mentioned non-parametric measures are not sufficient towards their applications in variety of disciplines. For instance, Shannon's measure of entropy always leads to exponential families of distributions but in actual practice, there are many families and distributions which are not exponential in nature. So, restricting to Shannon's entropy means restricting to exponential family only and thus leaving the system to be least flexible. An alternative to this is to use generalized parametric measures of information where the term 'generalized' does not mean superior or more useful but it simply means to be more flexible.

Csiszar (1977) critically investigated Shannon's measure and summarized the significance of this measure and its generalizations along with their scope of applications in the field of coding theory. Some other parametric generalizations of Shannon's entropy have been investigated and studied by Renyi (1961), Havrda and Charvat (1967), Tsallis (1988) etc. Different types of information measures and their mutual relationships have been studied by Garrido (2011). Dahl and Osteras (2010) applied Shannon entropy as a measure of information content in survey data and defined information efficiency as the empirical entropy divided by the maximum attainable entropy. Mathai and Haubold (2007) introduced

*Corresponding author email: omparkash777@yahoo.co.in

generalized entropies, studied some of their properties and examined situations where generalized entropy of order α finds its applications in a variety of mathematical models.

Jain and Mathur (2011) proposed a new symmetric divergence measure and studied its properties and obtained its bounds in terms of some well known divergence measures. Furuichi and Mitroi (2012) introduced some parametric divergence measures combining existing measures of information leading to new inequalities. Taneja (2005) studied some interesting inequalities among symmetric divergence measures whereas some other pioneers who worked towards the deep study of information measures are Csiszar (2008) and Chen *et al*. (2012).

The objective of the present paper is to introduce new measures of information and to extend their applications in the field of coding theory. The organization of this paper is as follows: In section 2 and 3, we have proposed new measures of entropy and studied their essential and desirable properties whereas section 4 and 5 deal with the proposal of new measures of directed divergence and inaccuracy respectively and the study of their properties for validation. In section 6, we have provided the applications of proposed measures to the discipline of coding theory.

2 New non-parametric measure of entropy

In this section, we propose a new measure of entropy to be called M entropy for a probability distribution $P=\left\{(p_1,p_2,...,p_n),\ p_i\geq 0,\sum_{i=1}^{n}p_i=1\right\}$ and study its essential and desirable properties. This new entropy measure is given by the following mathematical expression:

$$M(P)=\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i}-1 \tag{2.1}$$

Here, we take the convention that $0^0:=1$.

To prove that the measure (2.1) is a valid measure of entropy, we study its essential properties as follows:

1. Obviously, $M(P)$ is non-negative.
2. $M(P)$ is permutationally symmetric as it does not change if $p_1,p_2,...,p_n$ are reordered among themselves.
3. $M(P)$ is a continuous function of p_i for all p_i 's.
4. **<u>Concavity</u>**: $M(P)$ is a concave function of p_i for all p_i 's

To prove concavity property, we proceed as follows:

We have

$$\frac{\partial^2 M(P)}{\partial {p_1}^2}=\left(\left(1+\log p_1\right)^2-\frac{1}{p_1}\right)\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i} \tag{2.2}$$

Now, we know that for all $i=1,2,...,n$, we have $0\leq p_i\leq 1$,

that is, $\left(1+\log p_i\right)^2-\dfrac{1}{p_i}\leq 0$. (2.3)

Thus, using (2.3), we have $\dfrac{\partial^2 M(P)}{\partial {p_1}^2}\leq 0$.

So, $M(P)$ is a concave function of p_1. Similarly, it can be proved that $M(P)$ is a concave function of all p_i's.

Under the above conditions, the function $M(P)$ is a correct measure of entropy. Next, we study most desirable properties of $M(P)$.

1. **<u>Expansibility:</u>** We have $M(p_1,p_2,...,p_n,0)=M(p_1,p_2,...,p_n)$. That is, the entropy does not change by the inclusion of an impossible event.
2. For n degenerate distributions, we have $M(P)=0$. This indicates that for certain outcomes, the uncertainty should be zero.
3. **<u>Maximization of entropy</u>**: We use Lagrange's method to maximize the entropy measure (2.1) subject to the natural constraint $\sum_{i=1}^{n}p_i=1$. In this case, the corresponding Lagrangian is

$$L=\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i}-1-\lambda\left(\sum_{i=1}^{n}p_i-1\right) \tag{2.4}$$

Differentiating equation (2.4) with respect to $p_1,p_2,...,p_n$ and equating the derivatives to zero, we get $p_1=p_2=...=p_n$. This further gives $p_i=\frac{1}{n}\ \forall i$. Thus, we observe that the maximum value of $M(P)$ arises for the uniform distribution and this result is most desirable.

4. **<u>Maximum value:</u>** The maximum value of the entropy is given by $M\left(\frac{1}{n},\frac{1}{n},...,\frac{1}{n}\right)=n-1$.

Again, $M'\left(\frac{1}{n},\frac{1}{n},...,\frac{1}{n}\right)=1>0$. Thus, $M\left(\frac{1}{n},\frac{1}{n},...,\frac{1}{n}\right)$ is an increasing function of n, which is again a desirable result as the maximum value of entropy should always increase.

5. **<u>Non-additivity:</u>** Let $P=(p_1,p_2,...,p_n)$ and $Q=(q_1,q_2,...,q_m)$ be two independent probability distributions of two random variables X and Y, so that

$P(X=x_i)=p_i$, $P(Y=y_j)=q_j$ and

$P(X=x_i, Y=y_j)=P(X=x_i)\, P(Y=y_j)=p_i q_j$.

For the joint distributions of X and Y, there are nm possible outcomes with probabilities $p_i q_j$; $i=1,2,...,n$ and $j=1,2,...,m$ so that the entropy of the joint probability distribution, denoted by $M(P*Q)$, is given by

$$M(P*Q)=\prod_{i=1}^{n}\prod_{j=1}^{m}\left(p_i q_j\right)^{-p_i q_j}-1$$

$$=\left(\begin{array}{l} p_1^{-p_1(q_1+q_2+...+q_m)}\, p_2^{-p_2(q_1+q_2+...+q_m)}..\,p_n^{-p_n(q_1+q_2+...+q_m)} \\ q_1^{-q_1(p_1+p_2+...+p_n)}\, q_2^{-q_2(p_1+p_2+...+p_n)}..\,q_m^{-q_m(p_1+p_2+...+p_n)}-1\end{array}\right)$$

$$=\prod_{i=1}^{n} p_i^{-p_i}\prod_{j=1}^{m} q_j^{-q_j}-1 \tag{2.5}$$

Also, we have $M(P)+M(Q)+M(P)M(Q)$

$$=\prod_{i=1}^{n} p_i^{-p_i}\prod_{j=1}^{m} q_j^{-q_j}-1 \tag{2.6}$$

From equation (2.5) and (2.6), we have $M(P*Q)=M(P)+M(Q)+M(P)M(Q)$

Thus, we claim that the new measure of entropy $M(P)$ introduced in (2.1) satisfies all the essential as well as desirable properties of being an entropy measure, it is a valid measure of entropy.

3 Generalized parametric measure of entropy

In this section, we propose a new generalized measure of entropy to be called parametric M-entropy for a probability distribution

$P=\left\{(p_1,p_2,...,p_n),\ p_i\ge 0, \sum_{i=1}^{n} p_i=1\right\}$, given by the following mathematical expression:

$$M_\alpha(P)=\frac{1}{1-\alpha}\left(\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i(1-\alpha)}-1\right),\ \alpha>0\ ,\ \alpha\ne 1 \tag{3.1}$$

with the convention that $0^0:=1$.

We observe that for $\alpha\to 1$, measure (3.1) reduces to Shannon's (1948) entropy as shown below:

$$\underset{\alpha\to 1}{lt}\, M_\alpha(P)=\underset{\alpha\to 1}{lt}\,\frac{1}{1-\alpha}\left(\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i(1-\alpha)}-1\right)$$

$$=-\sum_{i=1}^{n} p_i\log p_i$$

Hence, this measure is a generalization of Shannon's measure and in particular reduces to measure (2.1) for $\alpha=0$.

Next, we study some essential properties of the generalized measure.

1. $M_\alpha(P)$ is non-negative, that is, $M_\alpha(P)\ge 0$.

Proof: Case-I: When $0<\alpha<1$

$$\frac{1}{1-\alpha}\left(\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i(1-\alpha)}-1\right)\ge 0$$

that is, iff $\log\left(\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i(1-\alpha)}\right)\ge 0$

that is, iff $-\sum_{i=1}^{n} p_i\log p_i\ge 0$ which is true.

Case-II: When $\alpha>1$, we have

$$\frac{1}{1-\alpha}\left(\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i(1-\alpha)}-1\right)\ge 0$$

that is, iff $\log\left(\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i(1-\alpha)}\right)\le 0$

that is, iff $-\sum_{i=1}^{n} p_i\log p_i\ge 0$ which is true.

2. $M_\alpha(P)$ is permutationally symmetric as it does not change if $p_1,p_2,...,p_n$ are re-ordered among themselves.

3. $M_\alpha(P)$ is a continuous function of p_i for all p_i's.

4. **Concavity:** $M_\alpha(P)$ is a concave function of p_i for all p_i's.

To prove concavity property, we proceed as follows: We have

$$\frac{\partial^2 M_\alpha(P)}{\partial p_1^2}=\left(\left(1+\log p_1\right)^2(1-\alpha)-\frac{1}{p_1}\right)\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i(1-\alpha)} \tag{3.2}$$

Now, using (2.3) and for $\alpha>0$, we have

$$\left(1+\log p_i\right)^2(1-\alpha)-\frac{1}{p_i}\le 0,\quad i=1,2,....n\ . \tag{3.3}$$

that is $\frac{\partial^2 M_\alpha(P)}{\partial p_1^2}\le 0$.

So, $M_\alpha(P)$ is a concave function of p_1. Similarly, it can be proved that $M_\alpha(P)$ is a concave function of all $p_i's$. Hence, under the above conditions, the function $M_\alpha(P)$ is a correct measure of entropy. Next, we study the most desirable properties of $M_\alpha(P)$.

1. **Expansibility**: We have

$M_\alpha(p_1,p_2,...,p_n,0)=M_\alpha(p_1,p_2,...,p_n)$.

2. For n degenerate distributions, we have $M_\alpha(P)=0$.
3. **Maximization of entropy:** Using Lagrange's method, we observe that the maximum value of $M_\alpha(P)$ arises for the uniform distribution.
4. **Maximum value:** The maximum value of the entropy is given by $M_\alpha\left(\frac{1}{n},\frac{1}{n},...,\frac{1}{n}\right)=\frac{n^{1-\alpha}-1}{1-\alpha}$ which is an increasing function of n, and is again a desirable result as the maximum value of entropy should always increase.

5. **Non-additivity:**
The entropy of the joint probability distribution, denoted by $M_\alpha(P*Q)$, is given by

$$M_\alpha(P*Q)=\frac{1}{1-\alpha}\left(\prod_{i=1}^{n}\prod_{j=1}^{m}\left(p_iq_j\right)^{-p_iq_j(1-\alpha)}-1\right)$$

$$=\frac{1}{1-\alpha}\left(\begin{matrix}p_1^{-p_1(1-\alpha)}p_2^{-p_2(1-\alpha)}\cdots p_n^{-p_n(1-\alpha)}q_1^{-q_1(1-\alpha)}\\ q_2^{-q_2(1-\alpha)}\ldots q_m^{-q_m(1-\alpha)}-1\end{matrix}\right)$$

$$=\frac{1}{1-\alpha}\left(\prod_{i=1}^{n}p_i^{-p_i(1-\alpha)}\prod_{j=1}^{m}q_j^{-q_j(1-\alpha)}-1\right)\qquad(3.4)$$

Also, we have

$$M_\alpha(P)+M_\alpha(Q)+(1-\alpha)M_\alpha(P)M_\alpha(Q)$$

$$=\frac{1}{1-\alpha}\left(\prod_{i=1}^{n}p_i^{-p_i(1-\alpha)}\prod_{j=1}^{m}q_j^{-q_j(1-\alpha)}-1\right)\qquad(3.5)$$

From equation (3.4) and (3.5), we have

$$M_\alpha(P*Q)=M_\alpha(P)+M_\alpha(Q)+(1-\alpha)M_\alpha(P)M_\alpha(Q)$$

Thus, we claim that the new measure of entropy $M_\alpha(P)$ introduced in (3.1) satisfies all the essential as well as desirable properties of being an entropy measure, it is a valid measure of entropy.

4 Some new measures of directed divergence

First measure of directed divergence

We propose a new non-parametric measure of divergence of probability distribution $P=(p_1,p_2,...,p_n)$ from another probability distribution $Q=(q_1,q_2,...,q_n)$ given by

$$D(P:Q)=\prod_{i=1}^{n}q_i^{-p_i}-\prod_{i=1}^{n}p_i^{-p_i}\qquad(4.1)$$

Measure (4.1) is a correct measure of directed divergence since it satisfies the following properties

1. $D(P:Q)\geq 0$

Proof: We have

$$\prod_{i=1}^{n}q_i^{-p_i}-\prod_{i=1}^{n}p_i^{-p_i}\geq 0$$

iff $-\sum_{i=1}^{n}p_i\log q_i\geq-\sum_{i=1}^{n}p_i\log p_i$ which is true.

2. $D(P:Q)=0$ iff $P=Q$.
3. $D(P:Q)$ is a convex function of P and Q.

Proof: We have

$$\frac{\partial^2 D(P:Q)}{\partial p_1^2}=$$

$$\left(\log q_1\right)^2\prod_{i=1}^{n}q_i^{-p_i}+\left(\frac{1}{p_1}-\left(1+\log p_1\right)^2\right)\prod_{i=1}^{n}p_i^{-p_i}>0\qquad(4.2)$$

$$\frac{\partial^2 D(P:Q)}{\partial q_1^2}=\frac{p_1\left(p_1+1\right)\prod_{i=1}^{n}q_i^{-p_i}}{q_1^2}>0\qquad(4.3)$$

From (4.2) and (4.3), it can be seen that $D(P:Q)$ is a convex function of p_1 and q_1. Similarly, it can be proved that $D(P:Q)$ is convex for each p_i and q_i for $i=1,2,...,n.$

Second measure of directed divergence

We propose a new generalized parametric measure of directed divergence of probability distribution $P=(p_1,p_2,...,p_n)$ from another probability distribution $Q=(q_1,q_2,...,q_n)$, given by

$$D_\alpha(P:Q)=\frac{1}{\alpha-1}\left(\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{-p_i(1-\alpha)}-1\right),\ \alpha>1,\ \alpha\neq 1\qquad(4.4)$$

It is observed that for $\alpha\to 1$ in (4.4), we get Kullback-Leibler's (1951) measure of directed divergence.

Thus, we claim that the measure (4.4) is a correct measure of directed divergence as it satisfies all the requisite properties.

Third measure of directed divergence

We introduce another parametric measure of directed divergence corresponding to measure of entropy (3.1), given by

$$D^\alpha(P:Q)=\frac{1}{1-\alpha}\left(\prod_{i=1}^{n}q_i^{-p_i(1-\alpha)}-\prod_{i=1}^{n}p_i^{-p_i(1-\alpha)}\right)$$

for $0\leq\alpha<1,\ \alpha\neq 1.$ (4.5)

Measure (4.5) is a correct measure of directed divergence as it also satisfies the requisite properties of a measure of directed divergence and reduces to Kullback -Leibler

(1951) measure as $\alpha \to 1$ and hence is a generalized measure. In particular, for $\alpha = 0$, it becomes measure (4.1).

Relation between Kullback-Leibler measure $K(P:Q)$ and the measure $D(P:Q)$

The following relationship can be established between $K(P:Q)$ and $D(P:Q)$.

Theorem 4.1. The divergence measure $K(P:Q)$ is no greater than divergence measure $D(P:Q)$, that is,

$$K(P:Q) \le D(P:Q) \qquad (4.6)$$

Proof. We know that

$$K(P:Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$

$$\le \frac{\prod_{i=1}^{n} q_i^{-p_i} - \prod_{i=1}^{n} p_i^{-p_i}}{\prod_{i=1}^{n} p_i^{-p_i}}$$

$$\le \prod_{i=1}^{n} q_i^{-p_i} - \prod_{i=1}^{n} p_i^{-p_i} = D(P:Q)$$

The above relationship can also be shown with the help of figure 1 in which we assume $P = (x, 1-x)$ and $Q = (1-x, x)$, $0 \le x \le 1$. It is to be noted that natural log is taken for calculating the numerical values for the plot of figure 1

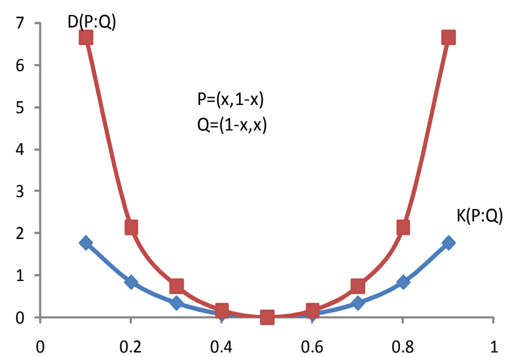

Fig.1. Comparison of the $K(P:Q)$ and $D(P:Q)$ divergence measures for $n = 2$.

Relation between Kullback Leibler measure $K(P:Q)$ and the measure $D_\alpha(P:Q)$:

Theorem 4.2. The divergence measure $K(P:Q)$ is no greater than divergence measure $D_\alpha(P:Q)$, that is,

$$K(P:Q) \le D_\alpha(P:Q), \quad \alpha > 1 \qquad (4.7)$$

Proof. We know that

$$K(P:Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$

$$= \frac{1}{\alpha - 1} \log \left(\frac{\prod_{i=1}^{n} q_i^{-p_i(1-\alpha)}}{\prod_{i=1}^{n} p_i^{-p_i(1-\alpha)}} \right)$$

$$\le \frac{1}{\alpha - 1} \left(\prod_{i=1}^{n} \left(\frac{p_i}{q_i} \right)^{-p_i(1-\alpha)} - 1 \right) = D_\alpha(P:Q)$$

Relation between Kullback Leibler measure $K(P:Q)$ and the measure $D^\alpha(P:Q)$(4.8)

Theorem 4.3. The divergence measure $K(P:Q)$ is no greater than divergence measure $D^\alpha(P:Q)$, that is,

$$K(P:Q) \le D^\alpha(P:Q), \quad 0 \le \alpha < 1 \qquad (4.8)$$

Proof. We know that

$$K(P:Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$

$$\le \frac{1}{1-\alpha} \left(\frac{\prod_{i=1}^{n} q_i^{-p_i(1-\alpha)} - \prod_{i=1}^{n} p_i^{-p_i(1-\alpha)}}{\prod_{i=1}^{n} p_i^{-p_i(1-\alpha)}} \right)$$

$$\le \frac{1}{1-\alpha} \left(\prod_{i=1}^{n} q_i^{-p_i(1-\alpha)} - \prod_{i=1}^{n} p_i^{-p_i(1-\alpha)} \right) = D^\alpha(P:Q)$$

5 New measures of Inaccuracy

First Measure of Inaccuracy

I. We first propose the new non-parametric measure of inaccuracy given by the following mathematical expression:

$$I(P:Q) = \prod_{i=1}^{n} \left(\frac{1}{q_i} \right)^{p_i} - 1 \qquad (5.1)$$

The measure (5.1) represents sum of two uncertainties:
(i) Uncertainty due to our not knowing $P = (p_1, p_2, ..., p_n)$, but knowing only $Q = (q_1, q_2, ..., q_n)$.
(ii) Uncertainty of P even when P is known.

The result (i) can be measured by measure of directed divergence, given by $D(P:Q) = \prod_{i=1}^{n} q_i^{-p_i} - \prod_{i=1}^{n} p_i^{-p_i}$ as defined in (4.1).

The result (ii) is measured by measure of entropy, given by $M(P) = \prod_{i=1}^{n} \left(\frac{1}{p_i} \right)^{p_i} - 1$ as defined in (2.1).

Thus, we have

$$I(P:Q) = D(P:Q) + M(P) \qquad (5.2)$$

Also,

$$I(P:P) = D(P:P) + M(P) = M(P) \tag{5.3}$$

and

$$I(P:Q) \geq M(P)$$

and equality sign holds if and only if $Q = P$.

So, measure (5.1) is an appropriate measure of inaccuracy as it satisfies the following properties:

(i) $I(P:Q) \geq 0$

(ii) $I(P:P)$ is an appropriate measure of entropy.

(iii) $I(P:Q) \geq I(P:P)$ and $I(P:Q)$ reduces to $I(P:P)$ only when $Q = P$.

Second Measure of Inaccuracy

II. We now propose another new parametric measure of inaccuracy given by

$$I_\alpha(P:Q) = \frac{1}{1-\alpha} \frac{\left(\prod_{i=1}^{n} p_i^{-p_i(1-\alpha)} \left(\prod_{i=1}^{n} q_i^{-p_i(1-\alpha)} - 1 \right) \right)}{\prod_{i=1}^{n} q_i^{-p_i(1-\alpha)}}, \ \alpha > 1 \tag{5.4}$$

For $\alpha \to 1$, $I_\alpha(P:Q)$ reduces to Kerridge's (1961) measure of inaccuracy and satisfies all the requisite properties of inaccuracy measure.

Third Measure of Inaccuracy

III. Next, we investigate and propose another parametric measure of inaccuracy, given by the following mathematical expression:

$$I^\alpha(P:Q) = \frac{1}{1-\alpha} \left(\prod_{i=1}^{n} \left(\frac{1}{q_i} \right)^{p_i(1-\alpha)} - 1 \right), \ 0 \leq \alpha < 1, \ \alpha \neq 1 \tag{5.5}$$

Again, for $\alpha \to 1$, we have

$$\underset{\alpha \to 1}{lt}\, I^\alpha(P:Q) = -\sum_{i=1}^{n} p_i \log q_i$$

which is Kerridge's measure of inaccuracy and we claim that measure (5.5) is an appropriate as it satisfies the requisite properties of inaccuracy measure.

In the next section, we provide the applications of the measures of information developed in the above sections.

6 Some new source coding theorems

6.1 Source coding with generalized measure of entropy

Source coding aims to encode the source that produces symbols x_i from X with probabilities p_i where $\sum_{i=1}^{n} p_i = 1$ using an alphabet of size D , that is, to map each symbol x_i to a codeword c_i of length l_i expressed using the D letters of the alphabet. If the set of lengths l_i satisfies the *Kraft's (1949) inequality

$$\sum_{i=1}^{n} D^{-l_i} \leq 1 \tag{6.1}$$

then there exists a uniquely decodable code with these lengths, which means that any sequence $c_{i1}c_{i2}...c_{in}$ can be decoded unambiguously into a sequence of symbols $x_{i1}x_{i2}...x_{in}$ Furthermore, any uniquely decodable code satisfies the Kraft's inequality (6.1). The Kraft's inequality is a basic result in information theory which gives a necessary condition for a code to be uniquely decipherable. Nagaraj (2009) provided a new proof of this inequality and its converse for prefix-free codes by a dynamical systems approach. Parkash and Priyanka (2011) developed some new results which are closely related with the Kraft's inequality.

The Shannon's (1948) source coding theorem indicates that the mean codeword length

$$L = \sum_{i=1}^{n} p_i l_i \tag{6.2}$$

is bounded below by the entropy of the source, that is, Shannon's entropy $H(P)$ and that the best uniquely decodable code satisfies

$$H(P) \leq L < H(P) + 1 \tag{6.3}$$

where the logarithm in the definition of the Shannon entropy is taken in base D . This result indicates that the Shannon entropy $H(P)$ is the fundamental limit on the minimum average length for any code constructed for the source. The lengths of the individual codewords, are given by

$$l_i = -\log_D p_i \tag{6.4}$$

Later, Campbell (1965) introduced the generalized mean codeword length, defined as

$$L_\alpha = \frac{\alpha}{1-\alpha} \log_D \left(\sum_{i=1}^{n} p_i D^{\frac{l_i(1-\alpha)}{\alpha}} \right) \tag{6.5}$$

and proved that Renyi's entropy $H_\alpha(P)$ forms a lower bound to it subject to Kraft's inequality. Sharma and Raina (1980) proved coding theorems for partially received information. Parkash and Kakkar (2012) proposed two new mean codeword lengths, investigated that these lengths satisfy desirable properties as a measure of typical codeword lengths and proved new noiseless coding theorems subject to Kraft's inequality.

Also, we have the following relation between the Shannon's entropy and the generalized entropy (3.1)

$$M_\alpha(P) = \log_\alpha \left(D^{H(P)} \right) \tag{6.6}$$

where $\log_\alpha(.)$ is the $\alpha-$deformed logarithm defined as $\log_\alpha x = \frac{x^{1-\alpha} - 1}{1-\alpha}$.

Now, we consider the following two cases:

Case-I When $0 < \alpha < 1$, we have

$$L \geq H(P)$$

$$\Rightarrow \frac{D^{L(1-\alpha)}-1}{1-\alpha} \geq \frac{D^{H(P)(1-\alpha)}-1}{1-\alpha}$$

$$\Rightarrow K_\alpha = \log_\alpha\left(D^L\right) \geq \log_\alpha\left(D^{H(P)}\right) = M_\alpha(P) \tag{6.7}$$

Case-II When $\alpha > 1$, we have

$L \geq H(P)$

$$\Rightarrow K_\alpha = \log_\alpha\left(D^L\right) \geq \log_\alpha\left(D^{H(P)}\right) = M_\alpha(P) \tag{6.8}$$

Here comes out the new generalized length K_α from (6.7) and (6.8) to which the generalized entropy $M_\alpha(P)$ forms a lower bound. It is a monotonic increasing function of mean codeword length L and it reduces to L when $\alpha \to 1$. The optimal codeword lengths are given by $l_i = -\log_D p_i$ which is similar to equation (6.4) as in case of Shannon's source coding theorem. K_α is not an average of the type $\phi^{-1}\left(\sum_{i=1}^{n} p_i \phi(l_i)\right)$ as introduced by Kolmogorov (1930) and Nagumo (1930) but is a simple expression of the $\alpha-$ deformed logarithm.

Note: When $l_1 = l_2 = ... = l_n = l$, then $K_\alpha \neq l$. Instead, it reduces to $\frac{D^{l(1-\alpha)}-1}{1-\alpha}$ which further reduces to l when $\alpha \to 1$. The above results (6.7) and (6.8) can also be stated in the form of following theorem:

Theorem 6.1. If $l_1, l_2, ..., l_n$ denote the lengths of the uniquely decipherable code for the random variable X, then $K_\alpha \geq M_\alpha(P)$ with equality if and only if $l_i = -\log_D p_i$.

Proof. We have to minimize the following codeword length:

$$K_\alpha = \frac{D^{(1-\alpha)\sum_{i=1}^{n} p_i l_i}-1}{1-\alpha} \tag{6.9}$$

subject to the Kraft's (1949) inequality

$$\sum_{i=1}^{n} D^{-l_i} \leq 1$$

The corresponding Lagrangian is given by

$$J = \frac{D^{(1-\alpha)\sum_{i=1}^{n} p_i l_i}-1}{1-\alpha} + \lambda\left(\sum_{i=1}^{n} D^{-l_i} - 1\right) \tag{6.10}$$

Differentiating (6.10) with respect to $l_i; i = 1, 2, ..., n$ and equating to zero, we get

$$p_i = \lambda D^{-l_i} \tag{6.11}$$

Using $\sum_{i=1}^{n} D^{-l_i} = 1$ and $\sum_{i=1}^{n} p_i = 1$, equation (6.11) gives $\lambda = 1$ and hence $p_i = D^{-l_i}$, that is, $l_i = -\log_D p_i$

Substituting l_i in (6.9), we get the minimum value of K_α as

$$\left[K_\alpha\right]_{\min} = \frac{1}{1-\alpha}\left(\prod_{i=1}^{n}\left(\frac{1}{p_i}\right)^{p_i(1-\alpha)} - 1\right) = M_\alpha(P)$$

6.2 Shannon's source coding via new measures of directed divergence

We know that the measure of directed divergence $D_\alpha(P:Q)$ as given by (4.4) is non-negative, that is,

$$D_\alpha(P:Q) = \frac{1}{\alpha-1}\left(\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{-p_i(1-\alpha)} - 1\right) \geq 0,\ \alpha > 1 \tag{6.12}$$

Substituting $q_i = \frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}},\ i = 1, 2, ..., n$ in (6.12), we get

$$\Rightarrow \prod_{i=1}^{n}\left(\frac{p_i}{D^{-l_i}}\sum_{i=1}^{n} D^{-l_i}\right)^{-p_i(1-\alpha)} \geq 1 \tag{6.13}$$

$$\Rightarrow -\sum_{i=1}^{n} p_i \log_D p_i - \log_D \sum_{i=1}^{n} D^{-l_i} \leq \sum_{i=1}^{n} p_i l_i = L$$

Now, since $\sum_{i=1}^{n} D^{-l_i}$ lies between D^{-1} and 1, therefore the lower bound for L lies between $H(P)$ and $H(P)+1$. Hence, we obtain the following result

$H(P) \leq L < H(P)+1$

which is the Shannon's source coding theorem for uniquely decipherable codes.

Note: On substituting $q_i = \frac{1}{n},\ i = 1, 2, ..., n$ in (6.12) and then taking logarithms both sides, we get

$$H(P) \leq \log_D n \tag{6.14}$$

which is a known result in information theory that shows that maximum value of Shannon's entropy can never be greater than $\log_D n$.

Similarly, if we take the non-negativity of the measure of directed divergence given in equation (4.1) and substitute $q_i = \frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}},\ i = 1, 2, ..., n$ in it, we get

$$\prod_{i=1}^{n}\left(\frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}\right)^{-p_i} \geq \prod_{i=1}^{n} p_l^{-p_i} \tag{6.15}$$

Taking logarithms on both sides of (6.14) and simplifying, we again arrive at inequality (6.3).

Proceeding on similar lines and using the non-negativity of the measure of directed divergence given by (4.5), we again get the Shannon's source coding theorem.

Concluding Remarks: The measures of entropy find tremendous applications in a variety of disciplines viz, biological, economical, physical sciences. Similarly, the measures of divergence have proved to be very useful in a various disciplines of engineering sciences. Since a single measure of information cannot be adequate for each discipline, we need a variety of generalized information measure to extend the scope of their applications. Further, these generalized measures induce flexibility into the system and hence preferred towards optimization problems. Keeping this idea in mind, we have generated various information measures for the discrete probability distribution and provided their applications in field of coding theory. With similar arguments, a variety of information theoretic measures can be developed for continuous probability distributions.

ACKNOWLEDGEMENT

The authors are thankful to Council of Scientific and Industrial Research, New Delhi for providing the financial assistance for the preparation of the manuscript.

REFERENCES

Campbell, LL. 1965. A Coding theorem and Renyi's entropy. Information and control. 8:423-429.

Chen, X., Kar, S. and Ralescu, DA. 2012. Cross-entropy measure of uncertain variables. Information Sciences. 201:53-60.

Csiszar. I. 1977. Information measures: a critical survey. In: Transactions of the 7th Conference on Information Theory, Statistical Decision Functions, Random Processes. Eds. Kozesnik, J. Academia, Prague. B:73-86.

Csiszar, I. 2008. Axiomatic characterizations of information measures. Entropy. 10: 261-273.

Dahl, FA. and Osteras, N. 2010. Quantifying information content in survey data by entropy. Entropy. 12:161-163.

Furuichi, S. and Mitroi, FC. 2012. Mathematical inequalities for some divergences. Physica A. 391:388-400.

Garrido, A. 2011. Classifying entropy measures. Symmetry. 3: 487-502.

Havrda, J. and Charvat, F. 1967. Quantification method of classification processes: concept of structural $\alpha -$ entropy. Kybernetika. 3:30-35.

Jain, KC. and Mathur, R. 2011. A symmetric divergence measure and its bounds. Tamkang Journal of Mathematics .42(4):493-503.

Kerridge, DF. 1961. Inaccuracy and inference. J. R. Statist. Soc. B 23:184-194.

Kolmogorov, A. 1930. Sur la notion de la moyenne. Atti Accad. Naz. Lincei. 12:388-391.

Kullback, S. and Leibler, RA. 1951. On information and sufficiency. Ann. Math. Statistics. 22:79-86.

Mathai, AM. and Haubold, HJ. 2007. On generalized entropy measures and pathways. Physica A. 385:493-500.

Nagaraj, N. 2009. A dynamical systems proof of Kraft–McMillan inequality and its converse for prefix-free codes. Chaos .19(1): 013136 -013140.

Nagumo, M. 1930. Uber eine Klasse der Mittelwerte. Japan. J. Math. 7:71-79.

Parkash, O. and Kakkar, P. 2012. Development of two new mean codeword lengths. Information Sciences. 207:90-97.

Parkash, O. and Priyanka. 2011. Contribution of Kraft's inequality to coding theory. Pacific-Asian Journal of Mathematics. 5(1):45-53.

Renyi, A. 1961. On measures of entropy and information. In: Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1:547-561.

Shannon, CE. 1948. A mathematical theory of communication. Bell System Tech. J. 27:379-423.

Sharma, BD. and Raina, N. 1980. Coding theorem for partial received information. Information Sciences. 20: 181-189.

Taneja, IJ. 2005. Refinement inequalities among symmetric divergence measures. The Australian Journal of Mathematical Analysis and Applications. 2(1):1-23.

Tsallis, C. 1988. Possible generalization of Boltzmann Gibbs statistics. J. Stat. Phys. 52:479-487.

Received: Aug 17, 2013; Revised: Nov 22, 2013; Accepted: Jan 2, 2014