SENRA Academic Publishers, British Columbia
Vol. 8, No. 3, pp. 3115-3123, October 2014
Online ISSN: 1920-3853; Print ISSN: 1715-9997

Canadian Journal of pure&applied sciences

NEW INFORMATION THEORETIC MODELS, THEIR DETAILED PROPERTIES AND NEW INEQUALITIES

*Om Parkash[1] and Priyanka Kakkar[2]
[1]Department of Mathematics, Guru Nanak Dev University, Amritsar- 143005
[2]Department of Mathematics, Guru Nanak Dev University, Amritsar- 143005, India

ABSTRACT

New parametric models concerning measures of information including entropy, joint entropy, conditional entropy, directed divergence and inaccuracy have been introduced and their detailed properties have been studied. A new concept to be called alpha logarithm has been introduced and the chain rule for entropy, alpha mutual entropy and directed divergence has been studied. More desirable properties known as subadditivity and strong subadditivity of the entropy function have been studied and certain new inequalities useful in the literature of information theory has been derived.

Keywords: Entropy, directed divergence, inaccuracy, concavity, convexity.

INTRODUCTION

The notion of disorder or chaos, uncertainty or randomness, also known as entropy, was introduced by Clausius in the 19th century in thermodynamics- an integral part of Boltzmann's theory. Subsequently, the probabilistic nature of the concept emerged more clearly with Gibbs work on statistical mechanics. It was one of Shannon's (1948) great insights that entropy could be used as a measure of information content and one's freedom of choice when one selects a message to be communicated over a noisy or noiseless channel. Shannon also stressed the importance of the relative entropy, also known as directed divergence as a measure of redundancy which provides a comparison between two probabilistic systems and typically measures the actual entropy to the maximal possible entropy. This relative entropy played a key role in many of the later discoveries and applications in a various disciplines of mathematical sciences.

Shannon (1948) proposed the first most important and the simplest measure of additive entropy of a probability distribution $P=(p_1, p_2, ..., p_n)$, given by

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

with the convention that $0\log 0 := 0$. Kullback and Leibler (1951) introduced the most important and desirable measure of divergence associated with the probability distributions
$P=(p_1, p_2, ..., p_n)$ and $Q=(q_1, q_2, ..., q_n)$, given by

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

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

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

The above mentioned measures have very nice mathematical properties and have tremendous applications in a variety of disciplines dealing with mathematical sciences. In spite of the fact that these measures are fundamental, we may face problems if we stick to these measures only because of their inadequacy towards applicability in every situation. An alternative is to use a generalized parametric measure of information where the parameter could hopefully be estimated from the data in the same way as ordinary statistical parameters are estimated from the data.

Many authors including Renyi (1961), Havrda Charvat (1967) and Tsallis (2009) introduced various generalized measures of information which are now increasingly used in many fields. Gupta and Bajaj (2013) studied the monotonic behaviour of the conditional Tsallis (2009) and Kapur (1967) entropies. Besenyei and Petz (2013) investigated a kind of partial subadditivity for Shannon and Tsallis entropy. Asgarani (2013) introduced a set of new three-parameter entropies which are expressed in terms of a generalized incomplete Gamma function and shown that for some special values of parameters, some known entropies are recovered. The uniqueness theorem for a two parameter extended relative entropy, that is, directed divergence is proved and its properties are studied by Furuichi (2010). Teixeira and Antunes (2012)

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

described three general definitions of conditional Renyi entropy and studied their properties and values as a function of parameter α. Bercher (2011) discussed the two families of two-parameter entropies and divergences, derived from the standard Renyi and Tsallis entropies and divergences. Furuichi and Mitroi (2012) introduced some parametric extended divergences combining J-divergence and Tsallis entropy defined by generalized logarithmic functions, which lead to new inequalities whereas several other inequalities on generalized entropies have been studied by Furuichi and Mitroi (2012).

The objective of the present study is to develop new information theoretic models and study their properties. The paper is organized as follows: In section 2, new measure of entropy to be called alpha entropy is introduced and its properties are studied. Section 3 deals with the new concept of alpha logarithm and provides the chain rule for alpha entropy and alpha mutual entropy by defining alpha joint entropy and alpha conditional entropy. Subadditivity and strong subadditivity property of entropy has been studied and various new inequalities has been derived. In section 4, new measure of directed divergence, its chain rule and new measure of inaccuracy is introduced.

2. New Measure of Entropy

In this section, we propose a new generalized measure of entropy for a probability distribution

$$P=\left\{\left(p_1,\, p_2,\ldots,\, p_n\right), p_i\geq 0, \sum_{i=1}^{n} p_i=1\right\}$$

of the random variable $X=\left(x_1,x_2,\ldots,x_n\right)$ and studied its essential and desirable properties. This new entropy measure of order α to be called alpha entropy, is given by the following mathematical expression:

$$_{\alpha}A(P)=\frac{\sum_{i=1}^{n} p_i\alpha^{\ln p_i}-1}{1-\alpha},\ \alpha>1, \qquad (2.1)$$

where ln(.) stands for the natural logarithm.

The expression (2.1) can also be written as

$$_{\alpha}A(X)=\frac{\sum_{x} p(x)\alpha^{\ln p(x)}-1}{1-\alpha},\ \alpha>1.$$

The two notations $A_\alpha(X)$ and $A_\alpha(P)$ need not to be confused as they have the same meaning and will be used wherever required for simplicity.

Obviously, we have $\lim_{\alpha\to 1}{}_{\alpha}A(P)=-\sum_{i=1}^{n} p_i\ln p_i$.

Thus, $A_\alpha(P)$ is a generalization of well known Shannon (1948) entropy.

Next, to prove that $_{\alpha}A(P)$ is a valid measure of entropy, we study its essential and desirable properties as follows:

1. Obviously, $_{\alpha}A(P)\geq 0$.
2. $_{\alpha}A(P)$ is permutationally symmetric as it does not change if $p_1,p_2,p_3,\ldots,p_n$ are re-ordered among themselves.
3. $_{\alpha}A(P)$ is a continuous function of p_i for all p_i's.

4. **Concavity:** The Hessian matrix of second order partial derivatives of $_{\alpha}A(P)$ with respect to $p_1,p_2,\ldots,p_n$ is given by

$$\begin{bmatrix} \frac{\alpha^{\ln p_1}(1+\ln\alpha)\ln\alpha}{p_1(1-\alpha)} & 0 & \cdots & 0 \\ 0 & \frac{\alpha^{\ln p_2}(1+\ln\alpha)\ln\alpha}{p_2(1-\alpha)} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \frac{\alpha^{\ln p_n}(1+\ln\alpha)\ln\alpha}{p_n(1-\alpha)} \end{bmatrix},$$

which is negative definite.

Thus, $_{\alpha}A(P)$ is a concave function of p_i for all p_i's.

5. **Expansibility:** We have

$$_{\alpha}A\left(p_1,p_2,p_3,\ldots,p_n,0\right)={}_{\alpha}A\left(p_1,p_2,p_3,\ldots,p_n\right).$$

That is, the entropy does not change by the inclusion of an impossible event.

6. For degenerate distributions, $_{\alpha}A(P)=0$.

This indicates that for certain outcomes, the uncertainty should be zero.

7. **Maximization of entropy:** 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

$$L\equiv\frac{\sum_{i=1}^{n} p_i\alpha^{\ln p_i}-1}{1-\alpha}+\lambda\left(1-\sum_{i=1}^{n} p_i\right). \qquad (2.2)$$

Differentiating equation (2.2) with respect to p_i, $i=1,2,\ldots,n$ and equating the derivatives to zero, we get

$$p_i=e^{\log_\alpha\left(\frac{\lambda(1-\alpha)}{1+\ln\alpha}\right)},\ i=1,2,..,n.$$

Using $\sum_{i=1}^{n} p_i=1$, we get $e^{\log_\alpha\left(\frac{\lambda(1-\alpha)}{1+\ln\alpha}\right)}=\frac{1}{n}$,

that is, $p_i=\frac{1}{n},\ i=1,2,..,n$.

Now $\left(\frac{\partial^2 L}{\partial p_i^2}\right)_{p_i=\frac{1}{n}} = \frac{n\alpha^{-\ln n}(1+\ln\alpha)}{(1-\alpha)} < 0$ for every i.

Thus, we observe that the maximum value of ${}_{\alpha}A(P)$ arises for the uniform distribution and this result is most desirable.

8. The maximum value of the entropy is given by

$${}_{\alpha}A\left(\frac{1}{n},\frac{1}{n},...,\frac{1}{n}\right) = \frac{\alpha^{-\ln n}-1}{1-\alpha}.$$

Also, ${}_{\alpha}A'\left(\frac{1}{n},\frac{1}{n},...,\frac{1}{n}\right) = \frac{\alpha^{-\ln n}\ln\alpha}{n(\alpha-1)} > 0$,

which shows that ${}_{\alpha}A\left(\frac{1}{n},\frac{1}{n},...,\frac{1}{n}\right)$ is an increasing function of n, which is again a desirable result as the maximum value of an entropy should always increase.

9. Non-Additive property:

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

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

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)=p_iq_j.$$

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

$${}_{\alpha}A(P*Q) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} p_iq_j\alpha^{\ln p_iq_j}-1}{1-\alpha}$$

$$= \frac{\sum_{i=1}^{n} p_i\alpha^{\ln p_i}\sum_{j=1}^{m} q_j\alpha^{\ln q_j}-1}{1-\alpha}. \quad (2.3)$$

Also, ${}_{\alpha}A(P)+{}_{\alpha}A(Q)+(1-\alpha)\ {}_{\alpha}A(P)\,{}_{\alpha}A(Q)$

$$= \frac{\sum_{i=1}^{n} p_i\alpha^{\ln p_i}\sum_{j=1}^{m} q_j\alpha^{\ln q_j}-1}{1-\alpha}. \quad (2.4)$$

From (2.3) and (2.4), we have

$${}_{\alpha}A(P*Q) = {}_{\alpha}A(P)+{}_{\alpha}A(Q)+(1-\alpha)\,{}_{\alpha}A(P)\,{}_{\alpha}A(Q).$$

So, measure of entropy ${}_{\alpha}A(P)$ is non-additive.

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

3 Alpha Logarithm-New Concepts

Let us introduce a new function to be called alpha logarithm given by

$$\ln_{\alpha} x \equiv \frac{\alpha^{\ln x}-1}{\alpha-1},\ \alpha\neq 1, \quad (3.1)$$

for any non-negative real number x and α.

For $\alpha\to 1$, alpha logarithm tends to natural logarithm, that is, $\lim_{\alpha\to 1}\ln_{\alpha} x = \ln x$.

Its inverse is the alpha exponential function given by

$$e_{\alpha}^{x} \equiv \left(1+(\alpha-1)x\right)^{\frac{1}{\ln\alpha}}. \quad (3.2)$$

For $\alpha\to 1$, alpha exponential tends to exponential function, that is, $\lim_{\alpha\to 1} e_{\alpha}^{x} = e^{x}$.

Now, let us state some results related to function (3.1) and (3.2)

1. $\ln_{\alpha}(xy) = \ln_{\alpha} x + \ln_{\alpha} y + (\alpha-1)\ln_{\alpha} x \ln_{\alpha} y$ (3.3)
2. $\ln_{\alpha}(xy) = \ln_{\alpha} x + \alpha^{\ln x}\ln_{\alpha} y$. (3.4)
3. $\ln_{\alpha}\left(\frac{1}{x}\right) = -\alpha^{-\ln x}\ln_{\alpha} x$. (3.5)
4. $e_{\alpha}^{x+y+(\alpha-1)xy} = e_{\alpha}^{x}e_{\alpha}^{y}$. (3.6)

Alpha entropy can be written as the alpha logarithm Shannon entropy in the following way:

$${}_{\alpha}A(P) = -\sum_{i=1}^{n} p_i \ln_{\alpha} p_i,\ \alpha>1. \quad (3.7)$$

Using result (3.5), (3.7) can further be written as

$${}_{\alpha}A(P) = \sum_{i=1}^{n} p_i\alpha^{\ln p_i}\ln_{\alpha}\frac{1}{p_i},\ \alpha>1. \quad (3.8)$$

Now, let us introduce the alpha joint entropy and alpha conditional entropy by means of following definitions:

Definition 3.1 For the conditional probability $p(x|y)=p(X=x|Y=y)$, we define alpha conditional entropy as

$${}_{\alpha}A(X|Y) = \sum_{x}\sum_{y} p(x,y)\alpha^{\ln p(x,y)}\ln_{\alpha}\frac{1}{p(x|y)},\ \alpha>1. \quad (3.9)$$

Definition 3.2 For the joint probability $p(x,y)=p(X=x,Y=y)$, we define alpha joint entropy as

$${}_{\alpha}A(X,Y) = \sum_{x}\sum_{y} p(x,y)\alpha^{\ln p(x,y)}\ln_{\alpha}\frac{1}{p(x,y)},\ \alpha>1 \quad (3.10)$$

or

$${}_{\alpha}A(X,Y) = \sum_{x}\sum_{y}\frac{p(x,y)\alpha^{\ln p(x,y)}-1}{1-\alpha},\ \alpha>1 \quad (3.11)$$

or

$$_{\alpha}A(X,Y) = -\sum_x \sum_y p(x,y) \ln_\alpha p(x,y), \ \alpha > 1. \quad (3.12)$$

3.1 Chain rules for alpha entropy

Now, let us study the chain rule for alpha entropy given by means of following theorem:

Theorem 3.1 For the two random variables X and Y, we have

$$_{\alpha}A(X,Y) = {}_{\alpha}A(X) + {}_{\alpha}A(Y|X), \ \alpha > 1. \quad (3.1.1)$$

Proof: Using $p(x,y) = p(y|x)p(x)$ and result (3.4), we have

$$_{\alpha}A(X,Y) = \sum_x \sum_y p(x,y)\alpha^{\ln p(x,y)} \ln_\alpha \frac{1}{p(x,y)}$$

$$= \sum_x \sum_y p(x,y)\alpha^{\ln p(x,y)} \ln_\alpha \frac{1}{p(y|x)} + \sum_x \sum_y p(x,y)\alpha^{\ln p(x)} \ln_\alpha \frac{1}{p(x)}$$

$$= {}_{\alpha}A(Y|X) + {}_{\alpha}A(X).$$

Note: If X and Y are independent, that is, $p(y|x) = p(y)$ for all x and y, then from Theorem 3.1, we have the following pseudo-additivity:

$$_{\alpha}A(X,Y) = {}_{\alpha}A(X) + {}_{\alpha}A(Y) + (1-\alpha)\,{}_{\alpha}A(X)\,{}_{\alpha}A(Y)$$

Theorem 3.2 The following chain rules hold:

(1) $_{\alpha}A(X,Y,Z) = {}_{\alpha}A(X,Y|Z) + {}_{\alpha}A(Z).$ (3.1.2)

(2) $_{\alpha}A(X,Y|Z) = {}_{\alpha}A(X|Z) + {}_{\alpha}A(Y|X,Z)$ (3.1.3)

Proof(1):

$$_{\alpha}A(X,Y|Z) = \sum_x \sum_y \sum_z p(x,y,z)\alpha^{\ln p(x,y,z)} \ln_\alpha \frac{1}{p(x,y|z)}$$

$$= \frac{1}{\alpha-1} \sum_x \sum_y \sum_z p(x,y,z) \left(\alpha^{\ln \frac{p(x,y,z)}{p(x,y|z)}} - \alpha^{\ln p(x,y,z)} \right)$$

$$= \frac{1}{\alpha-1}\left(\sum_x \sum_y \sum_z p(x,y,z)\alpha^{\ln p(z)} - 1 \right) + \frac{1}{1-\alpha}\left(\sum_x \sum_y \sum_z p(x,y,z)\alpha^{\ln p(x,y,z)} - 1 \right)$$

$$= -{}_{\alpha}A(Z) + {}_{\alpha}A(X,Y,Z)$$

that is, $_{\alpha}A(X,Y,Z) = {}_{\alpha}A(X,Y|Z) + {}_{\alpha}A(Z)$.

(2) On similar lines as proved in part (1), we have

$$_{\alpha}A(Y|X,Z) = {}_{\alpha}A(X,Y,Z) - {}_{\alpha}A(X,Z). \quad (3.1.4)$$

Also, from (3.1.1), we have

$$_{\alpha}A(X,Y) = {}_{\alpha}A(X) + {}_{\alpha}A(Y|X). \quad (3.1.5)$$

Therefore, equation (3.1.4) can further be written as

$$_{\alpha}A(Y|X,Z) = {}_{\alpha}A(X,Y|Z) + {}_{\alpha}A(Z) - \left({}_{\alpha}A(Z) + {}_{\alpha}A(X|Z) \right)$$

that is, $_{\alpha}A(Y|X,Z) = {}_{\alpha}A(X,Y|Z) - {}_{\alpha}A(X|Z)$ which proves (3.1.3).

Remark 1. From (3.1.3), we have

$_{\alpha}A(X|Z) \le {}_{\alpha}A(X,Y|Z)$.

2. The part (2) of Theorem 3.2 can further be generalized in the following way:

$$_{\alpha}A(X_1, X_2, ..., X_n|Y) = \sum_{i=1}^{n} {}_{\alpha}A(X_i|X_{i-1}, ..., X_1, Y).$$

Theorem 3.3 Let $X_1, X_2, ..., X_n$ be the random variables .Then we have the following chain rule:

$$_{\alpha}A(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} {}_{\alpha}A(X_i|X_{i-1}, ..., X_1) \quad (3.1.6)$$

Proof: We prove the theorem by induction on n. From Theorem 3.1, we have

$_{\alpha}A(X_1, X_2) = {}_{\alpha}A(X_1) + {}_{\alpha}A(X_2|X_1)$.

Let us assume that the result (3.1.6) is true for some n. From (3.1.1), we have

$$_{\alpha}A(X_1, X_2, ..., X_{n+1}) = {}_{\alpha}A(X_1, X_2, ..., X_n) + {}_{\alpha}A(X_{n+1}|X_n, ..., X_1)$$

$$= \sum_{i=1}^{n} {}_{\alpha}A(X_i|X_{i-1}, ..., X_1) + {}_{\alpha}A(X_{n+1}|X_n, ..., X_1),$$

which shows that (3.1.6) holds for $n+1$.

3.2 Subadditivities for alpha entropy

Theorem 3.4 Let X and Y be two random variables, then we have

$$_{\alpha}A(X|Y) \le {}_{\alpha}A(X), \ \alpha > 1, \quad (3.2.1)$$

with equality if and only if $\alpha = 1$ and $p(x|y) = p(x)$ for all x and y.

Proof: Let $f(x) = \frac{x\alpha^{\ln x} - x}{1-\alpha}, \ \alpha > 1$.

This function is concave function of x. So, by concavity of $f(x)$, we have

$$\sum_y p(y) f(p(x|y)) \le f\left(\sum_y p(y) p(x|y) \right) \quad (3.2.2)$$

Taking summation on x on both sides of (3.2.2), we get

$$\sum_y p(y) \sum_x f(p(x|y)) \le \sum_x f(p(x))$$

$$\Rightarrow \frac{\sum_y\sum_x p(x,y)\alpha^{\ln p(x|y)} - \sum_x\sum_y p(x,y)}{1-\alpha} \le \frac{\sum_x p(x)\alpha^{\ln p(x)} - 1}{1-\alpha},$$

that is,

$$\frac{\sum_y\sum_x p(x,y)\alpha^{\ln p(x|y)} - 1}{1-\alpha} \le \frac{\sum_x p(x)\alpha^{\ln p(x)} - 1}{1-\alpha} \quad (3.2.3)$$

Since $p(y)\alpha^{\ln p(y)} \le p(y)$ and $f(x) \ge 0$ for any $x > 0$, we have

$$p(y)\alpha^{\ln p(y)}\sum_x f\left(p(x|y)\right) \le p(y)\sum_x f\left(p(x|y)\right) \quad (3.2.4)$$

Taking summation on y on both sides of (3.2.4), we get

$$\sum_y p(y)\alpha^{\ln p(y)}\sum_x f\left(p(x|y)\right) \le \sum_y p(y)\sum_x f\left(p(x|y)\right)$$

$$\Rightarrow \sum_y\sum_x p(x,y)\alpha^{\ln p(x,y)} \ln_\alpha \frac{1}{p(x|y)} \le \frac{\sum_y\sum_x p(x,y)\alpha^{\ln p(x|y)} - 1}{1-\alpha} \quad (3.2.5)$$

From (3.2.3) and (3.2.5), we have

$$\sum_y\sum_x p(x,y)\alpha^{\ln p(x,y)} \ln_\alpha \frac{1}{p(x|y)} \le \frac{\sum_x p(x)\alpha^{\ln p(x)} - 1}{1-\alpha}$$

that is, ${}_\alpha A(X|Y) \le {}_\alpha A(X)$.

The equality holds when $\alpha = 1$ and $p(x|y) = p(x)$, that is, when X and Y are independent.

Theorem 3.5 Alpha entropy is sub-additive, that is,

$${}_\alpha A(X,Y) \le {}_\alpha A(X) + {}_\alpha A(Y). \quad (3.2.6)$$

Proof: The proof follows directly from Theorem 3.1 and Theorem 3.4.

Theorem 3.6 For the random variables $X_1, X_2, ..., X_n$, we have

$${}_\alpha A(X_1, X_2, ..., X_n) \le \sum_{i=1}^{n} {}_\alpha A(X_i), \quad (3.2.7)$$

with equality if and only if $\alpha = 1$ and the random variables are independent of each other.

Proof: The proof follows directly by making use of Theorem 3.3 and Theorem 3.4.

Theorem 3.7 For the random variables X, Y, Z with probabilities $p(X = x) = p(x)$, $p(Y = y) = p(y)$, $p(Z = z) = p(z)$ respectively and joint probability $p(x, y, z)$, strong subadditivity

$${}_\alpha A(X,Y,Z) + {}_\alpha A(Z) \le {}_\alpha A(X,Z) + {}_\alpha A(Y,Z) \quad (3.2.8)$$

holds with equality if and only if $\alpha = 1$ and the random variables X and Y are independent for a given random variable Z, that is, when $p(x|y,z) = p(x|z)$.

Proof: By making use of concavity of function $f(x)$ as defined in Theorem 3.4, we have

$$\sum_y p(y|z)\, f\left(p(x|y,z)\right) \le f\left(\sum_y p(y|z)p(x|y,z)\right)$$

$$\Rightarrow \sum_y p(y|z)\, f\left(p(x|y,z)\right) \le f\left(p(x|z)\right).$$

Multiply both sides of above inequality by $p(z)\alpha^{\ln p(z)}$ and taking the summation on z and x, we have

$$\sum_{z,x} p(z)\alpha^{\ln p(z)} \sum_y p(y|z)\, f\left(p(x|y,z)\right) \le \sum_{z,x} p(z)\alpha^{\ln p(z)} f\left(p(x|z)\right)$$

(3.2.9)

Now, $0 \le p(y|z) \le 1$

$$\Rightarrow p(y|z)\alpha^{\ln p(y|z)} \le p(y|z). \quad (3.2.10)$$

So, using the non-negativity of the function $f(x)$ and inequality (3.2.10), we have

$$p(y|z)\alpha^{\ln p(y|z)} \sum_x f\left(p(x|y,z)\right) \le p(y|z)\sum_x f\left(p(x|y,z)\right)$$

for any y, z.

Multiply both sides of above inequality by $p(z)\alpha^{\ln p(z)}$ and then taking the summation on y and z, we have

$$\sum_{y,z} p(z)\alpha^{\ln p(z)} p(y|z)\alpha^{\ln p(y|z)} \sum_x f\left(p(x|y,z)\right) \le \sum_{y,z} p(z)\alpha^{\ln p(z)} p(y|z) \sum_x f\left(p(x|y,z)\right)$$

$$\Rightarrow \sum_{y,z} p(y,z)\alpha^{\ln p(y,z)} \sum_x f\left(p(x|y,z)\right) \le \sum_{y,z} p(z)\alpha^{\ln p(z)} p(y|z) \sum_x f\left(p(x|y,z)\right) \quad (3.2.11)$$

From (3.2.9) and (3.2.11), we have

$$\sum_{y,z} p(y,z)\alpha^{\ln p(y,z)} \sum_x f\left(p(x|y,z)\right) \le \sum_{z,x} p(z)\alpha^{\ln p(z)} f\left(p(x|z)\right)$$

$$\Rightarrow \sum_{x,y,z} p(x,y,z)\alpha^{\ln p(x,y,z)} \ln_\alpha \frac{1}{p(x|y,z)} \le \sum_{x,z} p(x,z)\alpha^{\ln p(x,z)} \ln_\alpha \frac{1}{p(x|z)}$$

that is, ${}_\alpha A(X|Y,Z) \le {}_\alpha A(X|Z)$ (3.2.12)

$\Rightarrow {}_{\alpha}A(X,Y,Z) - {}_{\alpha}A(Y,Z) \le {}_{\alpha}A(X,Z) - {}_{\alpha}A(Z)$.
(Using Theorem 3.1)

The equality holds when $\alpha = 1$ and the random variables X and Y are independent for a given random variable Z, that is, when $p(x|y,z) = p(x|z)$.

Note: Inequality (3.2.6) can be recovered from inequality (3.2.8) by treating the random variable Z as a trivial one which proves that Theorem 3.7 is a generalization of Theorem 3.5.

The more generalized form of Theorem 3.7 is as follows:

Theorem 3.8 For the random variables $X_1, X_2, ..., X_n$, we have

$${}_{\alpha}A(X_{n+1}|X_1, X_2, ..., X_n) \le {}_{\alpha}A(X_{n+1}|X_2, ..., X_n). \tag{3.2.13}$$

Proof: The proof is on similar lines as in Theorem 3.7.

Theorem 3.9 For the random variables X, Y, Z, we have

$${}_{\alpha}A(X,Y|Z) \le {}_{\alpha}A(X|Z) + {}_{\alpha}A(Y|Z). \tag{3.2.14}$$

Proof: Adding $-2\,{}_{\alpha}A(Z)$ to both sides of inequality (3.2.8), we have

$${}_{\alpha}A(X,Y,Z) - {}_{\alpha}A(Z) \le {}_{\alpha}A(X,Z) - {}_{\alpha}A(Z) + {}_{\alpha}A(Y,Z) - {}_{\alpha}A(Z)$$

By making use of Theorem 3.1 in above inequality, we have result (3.2.14).

Above theorem can be generalized in the following way:

Theorem 3.10 For the random variables $X_1, X_2, ..., X_n, Z$, we have

$${}_{\alpha}A(X_1, X_2, ..., X_n|Z) \le {}_{\alpha}A(X_1|Z) + {}_{\alpha}A(X_2|Z) + ... + {}_{\alpha}A(X_n|Z) \tag{3.2.15}$$

Proof: The above result can be proved with the help of mathematical induction.

Theorem 3.11 For the random variables X, Y, Z, we have

$$2\,{}_{\alpha}A(X,Y,Z) \le {}_{\alpha}A(X,Y) + {}_{\alpha}A(Y,Z) + {}_{\alpha}A(Z,X). \tag{3.2.16}$$

Proof: From equation (3.2.12), we have

$${}_{\alpha}A(X|Y,Z) \le {}_{\alpha}A(X|Z). \tag{3.2.17}$$

From equation (3.2.1), we have

$${}_{\alpha}A(X|Z) \le {}_{\alpha}A(X). \tag{3.2.18}$$

So, (3.2.17) and (3.2.18) gives

$${}_{\alpha}A(X|Y,Z) \le {}_{\alpha}A(X). \tag{3.2.19}$$

Again from equation (3.2.12), we have

$${}_{\alpha}A(Z|X,Y) \le {}_{\alpha}A(Z|X). \tag{3.2.20}$$

Adding equation (3.2.19) and (3.2.20), we have

$${}_{\alpha}A(Z|X,Y) + {}_{\alpha}A(X|Y,Z) \le {}_{\alpha}A(Z|X) + {}_{\alpha}A(X) \tag{3.2.21}$$

Applying chain rule in (3.2.21), we have

$${}_{\alpha}A(X,Y,Z) - {}_{\alpha}A(X,Y) + {}_{\alpha}A(X,Y,Z) - {}_{\alpha}A(Y,Z) \le {}_{\alpha}A(Z,X)$$

which proves the theorem.

Theorem 3.12 For the random variables $X_1, X_2, ..., X_n$, we have

$${}_{\alpha}A(X_n|X_1) \le {}_{\alpha}A(X_2|X_1) + {}_{\alpha}A(X_3|X_2) + ... + {}_{\alpha}A(X_n|X_{n-1}) \tag{3.2.22}$$

Proof: From equation (3.1.6), we have

$${}_{\alpha}A(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} {}_{\alpha}A(X_i|X_{i-1}, ..., X_1),$$

that is,

$${}_{\alpha}A(X_1, X_2, ..., X_n) = {}_{\alpha}A(X_1) + {}_{\alpha}A(X_2|X_1) + {}_{\alpha}A(X_3|X_2, X_1) + ... + {}_{\alpha}A(X_n|X_{n-1}, ..., X_1) \tag{3.2.23}$$

Using equation (3.2.12) in (3.2.23), we have

$${}_{\alpha}A(X_1, X_2, ..., X_n) \le {}_{\alpha}A(X_1) + {}_{\alpha}A(X_2|X_1) + {}_{\alpha}A(X_3|X_2) + ... + {}_{\alpha}A(X_n|X_{n-1}) \tag{3.2.24}$$

From the generalization of part (2) of Theorem 3.2, we have

$${}_{\alpha}A(X_n|X_1) \le {}_{\alpha}A(X_2, ..., X_n|X_1)$$
$$= {}_{\alpha}A(X_1, ..., X_n) - {}_{\alpha}A(X_1)$$

(Using Theorem 3.1)

$$\le {}_{\alpha}A(X_2|X_1) + {}_{\alpha}A(X_3|X_2) + ... + {}_{\alpha}A(X_n|X_{n-1}),$$

which proves the theorem.

3.3 Alpha Mutual entropy

Alpha mutual entropy for the two random variables X and Y is defined as the difference between alpha entropy and alpha conditional entropy and is given by

$${}_{\alpha}I(X:Y) = {}_{\alpha}A(X) - {}_{\alpha}A(X|Y). \tag{3.3.1}$$

Also, alpha conditional mutual entropy for the three random variables X, Y, Z is given by

$${}_{\alpha}I(X:Y|Z) = {}_{\alpha}A(X|Z) - {}_{\alpha}A(X|Y,Z). \tag{3.3.2}$$

Theorem 3.13 (1) For the three random variables X, Y, Z, the following chain rule holds:

$${}_{\alpha}I(X:Y,Z) = {}_{\alpha}I(X:Z) + {}_{\alpha}I(X:Y|Z). \tag{3.3.3}$$

(2) For the random variables $X_1, X_2, ..., X_n$ and Y, the following chain rule holds:

$${}_{\alpha}I\left(X_1, X_2,..., X_n : Z\right) = \sum_{i=1}^{n} {}_{\alpha}I\left(X_i : Y \left| X_1, X_2,..., X_{i-1}\right.\right). \tag{3.3.4}$$

Proof: (1) We have

$$\begin{aligned}{}_{\alpha}I\left(X : Y|Z\right) &= {}_{\alpha}A\left(X|Z\right) - {}_{\alpha}A\left(X|Y,Z\right)\\ &= {}_{\alpha}A\left(X|Z\right) - {}_{\alpha}A\left(X\right) + {}_{\alpha}A\left(X\right) - {}_{\alpha}A\left(X|Y,Z\right)\\ &= -{}_{\alpha}I\left(X : Z\right) + {}_{\alpha}I\left(X : Y,Z\right),\end{aligned}$$

which proves the part (1).

(2) We have

$${}_{\alpha}I\left(X_1, X_2,..., X_n : Z\right) = {}_{\alpha}A(X_1, X_2,..., X_n) - {}_{\alpha}A(X_1, X_2,..., X_n|Y)$$

Using remark 2 and equation (3.18) in above equation, we have

$$\begin{aligned}{}_{\alpha}I\left(X_1, X_2,..., X_n : Z\right) &= \sum_{i=1}^{n} {}_{\alpha}A\left(X_i| X_{i-1},..., X_1\right)\\ &\quad - \sum_{i=1}^{n} {}_{\alpha}A\left(X_i| X_{i-1},..., X_1, Y\right)\\ &= \sum_{i=1}^{n} {}_{\alpha}I\left(X_i : Y|X_1, X_2,..., X_{i-1}\right),\end{aligned}$$

which proves part (2).

Note: Applications of inequalities

This is to be remarked that information theoretic inequalities play an important role in the theory of cryptography. To mention the fact, we consider Shannon's (1948) famous secret key cryptosystem model which consists of sender "S", a receiver "R", an eavesdropper "E", and an open channel from "S" to "R". The secret key Z is known to "S" and to "R" only whereas "S" encrypts the text X using Z according to the encryption rule which results in the cryptogram Y that is sent to "R" and can also be received by "E". "R" can recover X with his knowledge of Z. In the present model, a cipher is called perfect if and only if the text X and the cryptogram Y are independent random variables, that is, if and only if ${}_{\alpha}I\left(X,Y\right) = 0$ or equivalently, ${}_{\alpha}A\left(X\right) = {}_{\alpha}A\left(X|Y\right)$ if we apply the parametric entropy introduced in section 2. "R" must be able to recover X uniquely from Y and Z, that is, ${}_{\alpha}A\left(X|Y Z\right) = 0$.

Thus, applying this model, we have

$$\begin{aligned}{}_{\alpha}A\left(X\right) &= {}_{\alpha}A\left(X|Y\right)\\ &\le {}_{\alpha}A\left(XZ|Y\right)\end{aligned}$$

(Using equation (3.1.3))

$$\begin{aligned}&= {}_{\alpha}A\left(Z|Y\right) + {}_{\alpha}A\left(X|Y Z\right)\\ &= {}_{\alpha}A\left(Z|Y\right)\\ &\le {}_{\alpha}A\left(Z\right)\end{aligned}$$

This inequality shows that the entropy of the secret key must be as large as the entropy of the text to be encrypted, and consequently provides a helpful tool for the removal of uncertainty to be removed.

4 New measures of Directed Divergence and Inaccuracy

4.1 Measure of directed divergence

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

$${}_{\alpha}D(P:Q) = \frac{\sum_{i=1}^{n} p_i \alpha^{\ln \frac{p_i}{q_i}} - 1}{\alpha - 1},\ \alpha > 1. \tag{4.1.1}$$

This measure is to be called alpha directed divergence. The measure (4.1.1) can also be written in the following form:

$${}_{\alpha}D(P:Q) = \sum_{i=1}^{n} p_i \ln_{\alpha} \frac{p_i}{q_i},\ \alpha > 1, \tag{4.1.2}$$

or

$${}_{\alpha}D(P:Q) = -\sum_{i=1}^{n} p_i \alpha^{\ln \frac{p_i}{q_i}} \ln_{\alpha} \frac{q_i}{p_i},\ \alpha > 1. \tag{4.1.3}$$

Also, $\lim_{\alpha\to 1} {}_{\alpha}D(P:Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$,

which is Kullback-Leibler's (1951) measure of directed divergence.

Now, to prove the validity of this measure, we will resort to the definition of measure of directed divergence as given by Csiszer (1972) in the form of following theorem:

Theorem 4.1 If $\phi(.)$ is twice differentiable convex function such that $\phi(1) = 0$, then

$$D\left(P:\mathrm{Q}\right) = \sum_{i=1}^{n} q_i \phi\left(\frac{p_i}{q_i}\right), \tag{4.1.4}$$

is a valid measure of directed divergence.

Now, let us take $\phi\left(x\right) = x \ln_{\alpha} x = \frac{x\left(\alpha^{\ln x} - 1\right)}{\alpha - 1},\ \alpha > 1$.

Here $\phi\left(1\right) = 0$ and $\phi'\left(x\right) = \frac{1}{\alpha - 1}\left[\left(1 + \ln\alpha\right)\alpha^{\ln x} - 1\right]$,

$$\phi''\left(x\right) = \frac{\alpha^{\ln x}\left(1 + \ln\alpha\right)\ln\alpha}{x\left(\alpha - 1\right)} > 0,$$

which shows that $\phi\left(x\right)$ is a convex function of x.

So, $\phi\left(x\right)$ satisfies all conditions of Theorem 4.1. Substituting it in equation (4.1.3), we get the measure of directed divergence (4.1.1). Hence, (4.1.1) is a valid measure of directed divergence.

Second Criteria to prove the validity of this measure is to study its properties which are as follows:

1. ${}_{\alpha}D(P:Q) \geq 0$.

Proof: We will find the extremum of ${}_{\alpha}D(P:Q)$ subject to the constraint $\sum_{i=1}^{n} p_i = 1$. Let us consider the Lagrangian given by

$$L \equiv {}_{\alpha}D(P:Q) + \lambda\left(1 - \sum_{i=1}^{n} p_i\right).$$

Now, $\frac{\partial L}{\partial p_i} = \frac{\alpha^{\ln\frac{p_i}{q_i}}(1+\ln\alpha)}{\alpha-1} - \lambda,\ i = 1,2,...,n$ and $\frac{\partial L}{\partial p_i} = 0$, gives

$$p_i = q_i e^{\log_\alpha \frac{\lambda(\alpha-1)}{1+\ln\alpha}}. \tag{4.1.5}$$

Now, using $\sum_{i=1}^{n} p_i = 1$ gives $e^{\log_\alpha \frac{\lambda(\alpha-1)}{1+\ln\alpha}} = 1$.

So, from equation (4.1.5), we get

$p_i = q_i,\ i = 1,2,...,n.$

Also, $\left(\frac{\partial^2 L}{\partial p_i^{\,2}}\right)_{p_i=q_i} = \frac{(1+\ln\alpha)\ln\alpha}{q_i(\alpha-1)} > 0,\ i = 1,2,...,n$.

and $\frac{\partial^2 L}{\partial p_i p_j} = 0,\ i \neq j.$

So, we see that the minimum value of ${}_{\alpha}D(P:Q)$ is obtained when $p_i = q_i,\ i = 1,2,...,n$ and $\left[{}_{\alpha}D(P:Q)\right]_{\min} = 0$. So, ${}_{\alpha}D(P:Q) \geq 0$.

2. ${}_{\alpha}D(P:Q) = 0$ when $P = Q$.

3. ${}_{\alpha}D(P:Q)$ is a convex function of P and Q.

Proof: The Hessian matrix of second order partial derivatives of ${}_{\alpha}D(P:Q)$ with respect to $p_1, p_2, \ldots, p_n$ is given by

$$\begin{bmatrix} \frac{\alpha^{\ln\frac{p_1}{q_1}}(1+\ln\alpha)\ln\alpha}{p_1(\alpha-1)} & 0 & \cdots & 0 \\ 0 & \frac{\alpha^{\ln\frac{p_2}{q_2}}(1+\ln\alpha)\ln\alpha}{p_2(\alpha-1)} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \frac{\alpha^{\ln\frac{p_n}{q_n}}(1+\ln\alpha)\ln\alpha}{p_n(\alpha-1)} \end{bmatrix}$$

which is positive definite. A similar result is also true with respect to $q_1, q_2, \ldots, q_n$. Thus, we conclude that ${}_{\alpha}D(P:Q)$ is a convex function of both $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$.

Hence, ${}_{\alpha}D(P:Q)$ is a valid measure of directed divergence.

Next, we propose the measure of conditional directed divergence by taking in view the definition of alpha entropy and conditional alpha entropy.

Definition 4.1 For the two joint probability distributions $p(x,y)$ and $q(x,y)$ and the two conditional probability distributions $p(y|x)$ and $q(y|x)$, conditional directed divergence is given by the following mathematical expression:

$${}_{\alpha}D\left(p(y|x):q(y|x)\right) = -\sum_{i=1}^{n} p(x,y)\alpha^{\ln\frac{p(x,y)}{q(x,y)}} \ln_\alpha \frac{q(y|x)}{p(y|x)},\ \alpha > 1. \tag{4.1.6}$$

Theorem 4.2 The following chain rule holds for alpha directed divergence for general case, that is, when X and Y are not independent

$${}_{\alpha}D\left(p(x,y):q(x,y)\right) = {}_{\alpha}D\left(p(x):q(x)\right) + {}_{\alpha}D\left(p(y|x):q(y|x)\right) \tag{4.1.7}$$

Proof: The proof follows from the direct calculations:

$${}_{\alpha}D\left(p(x,y):q(x,y)\right) = \sum_x\sum_y p(x,y)\ln_\alpha \frac{p(x,y)}{q(x,y)}$$

$$= \sum_x\sum_y p(x,y)\left(\ln_\alpha \frac{p(x)}{q(x)} + \alpha^{\ln\frac{p(x)}{q(x)}} \ln_\alpha \frac{p(y|x)}{q(y|x)}\right)$$

(Using equation (3.4))

$$= \sum_x\sum_y p(x,y)\ln_\alpha \frac{p(x)}{q(x)} - \sum_x\sum_y p(x,y)\alpha^{\ln\frac{p(x)}{q(x)}}\alpha^{\ln\frac{p(y|x)}{q(y|x)}} \ln_\alpha \frac{q(y|x)}{p(y|x)}$$

(Using equation (3.5))

$$= \sum_x\sum_y p(x,y)\ln_\alpha \frac{p(x)}{q(x)} - \sum_x\sum_y p(x,y)\alpha^{\ln\frac{p(x,y)}{q(x,y)}} \ln_\alpha \frac{q(y|x)}{p(y|x)} \tag{4.1.8}$$

$$= {}_{\alpha}D\left(p(x):q(x)\right) + {}_{\alpha}D\left(p(y|x):q(y|x)\right).$$

Note: When X and Y are independent, that is, when $p(y|x) = p(y)$ and $q(y|x) = q(y)$, alpha directed divergence has a pseudoadditivity as shown below:

From (4.1.8)

$${}_{\alpha}D\left(p(x,y):q(x,y)\right) = \sum_x\sum_y p(x,y)\ln_\alpha \frac{p(x)}{q(x)} - \sum_x\sum_y p(x,y)\alpha^{\ln\frac{p(x,y)}{q(x,y)}} \ln_\alpha \frac{q(y|x)}{p(y|x)}$$

$$= {}_{\alpha}D\left(p(x):q(x)\right) + {}_{\alpha}D\left(p(y):q(y)\right)\sum_x p(x)\alpha^{\ln\frac{p(x)}{p(x)}}$$

$$= {}_{\alpha}D\left(p(x):q(x)\right) + {}_{\alpha}D\left(p(y):q(y)\right) + (\alpha-1)\,{}_{\alpha}D\left(p(x):q(x)\right){}_{\alpha}D\left(p(y):q(y)\right)$$

4.2 Measure of Inaccuracy

We propose the new parametric measure of inaccuracy given by the following mathematical expression:

$$_{\alpha}I(P:Q)=\sum_{i=1}^{n}p_i\alpha^{\ln p_i}\ln_{\alpha}\frac{1}{q_i},\ \alpha>1\ . \qquad (4.2.1)$$

Letting $\alpha \to 1$, measure (4.2.1) reduces to Kerridge's (1961) measure of inaccuracy given by

$$I(P:Q)=-\sum_{i=1}^{n}p_i\ln q_i\ .$$

Measure (4.2.1) is a valid measure of inaccuracy as it satisfies the following properties:

1. $_{\alpha}I(P:Q)=\sum_{i=1}^{n}p_i\alpha^{\ln p_i}\ln_{\alpha}\frac{1}{q_i}\geq 0$. (4.2.2)

Since measure (4.2.1) is a sum of alpha entropy and alpha directed divergence given by (2.1) and (4.1.1) respectively, both of which are non-negative quantities, therefore (4.2.2) holds.

2. $_{\alpha}I(P:P)=\sum_{i=1}^{n}p_i\alpha^{\ln p_i}\ln_{\alpha}\frac{1}{p_i},\ \alpha>1$ is a valid measure of entropy as proved in Section 2.

3. $_{\alpha}I(P:Q)\geq{}_{\alpha}I(P:P)$ and $_{\alpha}I(P:Q)$ reduces to $_{\alpha}I(P:Q)$ only when $Q=P$.

CONCLUDING REMARKS

The non-additive Tsallis entropy which is supposed to be a firm basis of the non-extensive statistical mechanics having applications in diverse disciplines of mathematical sciences motivated us to introduce the non-additive measure of entropy in the manuscript. It is expected that the proposed measure will perform equally well in all the application areas parallel to Tsallis entropy. Also, we have made the detailed study of the proposed entropy measure in the form of inequalities, theorems and developed corresponding measures of information applicable to a variety of mathematical disciplines.

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

Asgarani, S. 2013. A set of new three-parameter entropies in terms of a generalized incomplete Gamma function. Physica A. 392:1972-1976.

Bercher, JF. 2011. Escort entropies and divergences and related canonical distribution. Physics Letters A. 375:2969-2973.

Besenyei, A. and Petz, D. 2013. Partial subadditivity of entropies. Linear Algebra Appl. 439:3297-3305.

Csiszer, I. 1972. A class of measures of informativity of observation channels. Periodic Math. Hungarica. 2:191-213.

Furuichi, S. 2010. An axiomatic characterization of a two-parameter extended relative entropy. Journal of Mathematical Physics. 51:123302-123310.

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

Furuichi, S., Minculete, N. and Mitroi, FC. 2012. Some inequalities on generalized entropies. Journal of Inequalities and Applications. 2012:226-241.

Gupta, N. and Bajaj, RK. 2013. On partial monotonic behaviour of some entropy measures. Statistics and Probability Letters. 83:1330-1338.

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

Kapur, JN. 1967. Generalized entropy of order α and type β . The Maths. Seminar. 4:78-94.

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

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

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.

Tsallis, C. 2009. Non-additive entropy: The concept and its use. Eur. Phys. J. A. 40:257-266.

Teixeira, A., Matos, A. and Antunes, L. 2012. Conditional Renyi entropies. IEEE Transactions on Information Theory. 58 (7):4273-4277.

Received: June 6, 2014; Revised and Accepted: July 14, 2014