National Journal of Multidisciplinary Research and Development
ISSN: 2455-9040; Impact Factor: RJIF 5.22
www.nationaljournals.com
Volume 2; Issue 1; January 2017; Page No. 11-21

Exponential Directed-Divergence Convex Function of 'Type (α, β)'

Dhanesh Garg

Maharishi Markendeshwar University, Mullana-Ambala, Haryana, India

Abstract

In this paper, we introduce a quantity which is called exponential entropy of 'type (α, β)' and discuss its some major properties corresponding to exponential entropy of concave function. Further, we define an exponential directed-divergence convex function of 'type (α, β)'. From this directed-divergence measure a new exponential information measures have also been derived when one of the probability distribution is uniform. Tsallis' -Havrda -Charvat, Takuya Yamano, Sharma and Mittal relative entropies are the particlular cases of the proposed measure.

Keywords: Renyi's entropy, Exponential Shannon's entropy, Convex and concave function, Exponential Tsallis' entropy of 'type (α, β)' Exponential directed-divergence

Introduction

Let $\Delta_n = \{A = (a_1, a_2, ..., a_n) : a_i \geq 0, i = 1, 2, ..., n, \ n \geq 2, \sum_{i=1}^{n} a_i = 1\}$ be a set of n-complete probability distributions. For any probability distribution $A = (a_1, a_2, ..., a_n) \in \Delta_n$, Shannon's entropy [27], is defined as

$$H(A) = -\sum_{i=1}^{n} a_i \log a_i \quad . \tag{1.1}$$

Various generalized entropies have been introduced in the literature, taking the Shannon's entropy as basic and have found applications in various disciplines such as economics, statistics, information processing and computing etc.
Generalizations of Shannon's entropy started with Renyi's entropy [23] of order-α, given by

$$H_\alpha(A) = \frac{1}{1-\alpha} \log\left[\sum_{i=1}^{n} a_i^\alpha\right], \quad \alpha \neq 1, \alpha > 0 \quad . \tag{1.2}$$

Campbell [4] studied exponentials of the Shannon's and Renyi's entropies, given by

$$E(A) = e^{H(A)} \tag{1.3}$$

and

$$E^\alpha(A) = e^{H_\alpha(A)}, \tag{1.4}$$

where $H(A)$ and $H_\alpha(A)$ represent respectively the Shannon's and Renyi's entropies. It may also be mentioned that Koski and Persson [16] studied

$$E_{(\alpha,\beta)}(A) = e^{H_{(\alpha,\beta)}(A)}, \tag{1.5}$$

exponential of Kapur's entropy [14] given by

$$H_{(\alpha,\beta)}(A) = \frac{1}{(\beta-\alpha)} \log \frac{\sum_{i=1}^{n} a_i^\alpha}{\sum_{i=1}^{n} a_i^\beta}, \quad \alpha \neq \beta; \ \alpha, \beta > 0 \tag{1.6}$$

It is interesting to notice that, in the case of discrete uniform distribution $A \in \Delta_n$, (1.3), (1.4) and (1.5) all reduce to n, just the 'size of the sample space of the distribution'.

This paper is organized as follows: Sec.II, define the exponential entropy of 'type (α, β)' and discuss its some major properties corresponding to exponential entropy of concave function. Sec. III, define the measure of exponential relative information of 'type (α, β)'. Sec. IV, discuss the generalized exponential directed-divergence deasure of 'type (α, β)'. Sec. V, discuss the measures of information.

In the next section, we define a new information measure $E_\alpha^\beta(A)$ and study its properties.

2. Exponential Entropy of 'Type (α, β)' and its Properties

Corresponding to Tsallis' entropy [28], the exponential entropy of 'type(α, β)' is defined as follows:

Definition: Exponential 'type (α, β)' entropy of a discrete distribution A is given by:

$$E_\alpha^\beta(A) = \frac{\left[e^{\log_2 \sum_{i=1}^n a_i^\alpha} - e^{\log_2 \sum_{i=1}^n a_i^\beta} \right]}{\beta - \alpha} \quad \alpha, \beta > 0 \; ; \; \alpha \neq \beta \tag{2.1}$$

(i) When $\beta = 1, \alpha \to 1$ or $\alpha = 1, \beta \to 1$ measure (2.1) reduces to Shannon's entropy.

(ii) When $\beta = 1,$ or $\alpha = 1$ (2.1) becomes exponential entropy of type α.

The quantity (2.1) introduced in the present section is entropy. Such a name will be justified, if it shares some major properties with Shannon's and other entropies in the literature. We study some such properties in the next theorem.

Theorem 2.1: The measure of information $E_\alpha^\beta(A), \{A = (a_{1,} a_2, ..., a_n), 0 \le a_i \le 1, \sum_{i=1}^n a_i = 1\}$ has the following properties:

1) Symmetry:

$E_\alpha^\beta(A) = E_\alpha(a_{1,} a_2, ..., a_n)$ is a symmetric function of $(a_{1,} a_2, ..., a_n)$.

2) Non-negative:

$E_\alpha^\beta(A) > 0$, for all $\alpha, \beta > 0; \alpha \neq \beta$.

3) Expansible:

$E_\alpha^\beta(a_{1,} a_2, ..., a_n; 0) = E_\alpha^\beta(a_{1,} a_2, ..., a_n).$

4) Decisive:

$E_\alpha^\beta(0,1) = E_\alpha^\beta(1,0) = 0.$

5) Maximality:

$$E_\alpha^\beta(a_{1,} a_2, ..., a_n) \le E_\alpha^\beta(1/n, 1/n, ..., 1/n) = \frac{e^{(1-\alpha)\log_2 n} - e^{(1-\beta)\log_2 n}}{\beta - \alpha}.$$

6) Concavity:

The measure $E_\alpha^\beta(A)$ is a concave function of the probability distribution $A = (a_{1,} a_2, ..., a_n),\ a_i \ge 0; \sum_{i=1}^n a_i = 1,$ when either $\alpha > 1; \beta \le 1$ or $\alpha \le 1; \beta > 1$.

7) Continuity:

$E_\alpha^\beta(a_{1,} a_2, ..., a_n)$ is continuous in the region $a_i \ge 0$ for all $\alpha, \beta > 0;\ \alpha \neq \beta.$

Proof: (1), (3), (4) and (5); these properties are obvious and can be verified easily for property (7), We know that $\sum_{i=1}^n a_i^\alpha - \sum_{i=1}^n a_i^\beta$ is continuous in the region $a_i \ge 0$ for all $\alpha, \beta > 0$. Hence, $E_\alpha^\beta(A)$, is also continuous in the region $a_i \ge 0$ for all $\alpha, \beta > 0;\ \alpha \neq \beta$.

Property (2): The measure $E_\alpha^\beta(A)$ is non-negative for all $\alpha, \beta > 0;\ \alpha \neq \beta$.

Proof: We consider the following cases:

Case (i): When $\alpha > 1\,; \beta < 1$

$$e^{\log_2\left[\sum_{i=1}^n a_i^\alpha\right]} < 1 \text{ and } e^{\log_2\left[\sum_{i=1}^n a_i^\beta\right]} > 1. \tag{2.2}$$

From (2.2), we get

$$e^{\log_2\left[\sum_{i=1}^{n} a_i^{\alpha}\right]} - e^{\log_2\left[\sum_{i=1}^{n} a_i^{\beta}\right]} < 0 \text{, for } \alpha > 1;\ \beta < 1.$$

Since for $\alpha > 1\,; \beta < 1 \quad \Rightarrow\ \beta - \alpha < 0$.

We get

$$\frac{e^{\log_2\left[\sum_{i=1}^{n} a_i^{\alpha}\right]} - e^{\log_2\left[\sum_{i=1}^{n} a_i^{\beta}\right]}}{\beta - \alpha} > 0.$$

i.e., $E_{\alpha}^{\beta}(A) > 0.$

Case (ii): Similarly, for $\alpha < 1;\ \beta > 1$, we get

$$e^{\log_2\left[\sum_{i=1}^{n} a_i^{\alpha}\right]} - e^{\log_2\left[\sum_{i=1}^{n} a_i^{\beta}\right]} > 0,$$

and $\beta - \alpha > 0,$

we get

$$\frac{e^{\log_2\left[\sum_{i=1}^{n} a_i^{\alpha}\right]} - e^{\log_2\left[\sum_{i=1}^{n} a_i^{\beta}\right]}}{\beta - \alpha} > 0.$$

i.e., $E_{\alpha}^{\beta}(A) > 0.$

Case (iii): When $\alpha > 1$ and $\beta > 1$.

(a) Let $\alpha > \beta > 1$, then

$$e^{\log_2\left[\sum_{i=1}^{n} a_i^{\alpha}\right]} - e^{\log_2\left[\sum_{i=1}^{n} a_i^{\beta}\right]} < 0,$$

and

$\beta - \alpha < 0$, we get $E_{\alpha}^{\beta}(A) > 0.$

(b) Let $\beta > \alpha > 1$, then $e^{\log_2\left[\sum_{i=1}^{n} a_i^{\alpha}\right]} - e^{\log_2\left[\sum_{i=1}^{n} a_i^{\beta}\right]} > 0,$

and

$\beta - \alpha > 0$, we get $E_{\alpha}^{\beta}(A) > 0.$

(c) Let $\alpha > \beta > 1$, then $e^{\log_2\left(\sum_{i=1}^{n} a_i^{\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_i^{\beta}\right)} < 0$

and

$\beta - \alpha < 0$, we get $E_{\alpha}^{\beta}(A) > 0.$

(d) Let $\beta > \alpha > 1$, then $e^{\log_2\left(\sum_{i=1}^{n} a_i^{\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_i^{\beta}\right)} > 0$

and

$\beta - \alpha > 0$, we get $E_{\alpha}^{\beta}(A) > 0.$

Case (iv) : When $\alpha < 1$ and $\beta < 1$.

(a) Let $\alpha < \beta < 1$, then $e^{\log_2\left(\sum_{i=1}^{n} a_i^{\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_i^{\beta}\right)} > 0$

and

$\beta - \alpha > 0$, we get $E_{\alpha}^{\beta}(A) > 0.$

(b) Let $\beta < \alpha < 1$, then $e^{\log_2\left(\sum_{i=1}^{n} a_i^{\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_i^{\beta}\right)} < 0$

and

$\beta - \alpha < 0$, we get $E_\alpha^\beta(A) > 0.$

From case (i), (ii), (iii) and (iv),we conclude that

$E_\alpha^\beta(A) > 0$ for all $\alpha, \beta > 0; \alpha \neq \beta$.

To prove the next property, we shall use the following definition of a concave function.

Definition: (Concave Function): A function $f(.)$ over the points in a convex set $\Re$ is concave if for all $r_1, r_2 \in \Re$ and $\mu \in (0,1)$

$$\mu f(r_1) + (1-\mu) f(r_2) \leq f(\mu r_1 + (1-\mu) r_2) \tag{2.3}$$

The function $f(.)$ is **convex** if the above inequality holds with $\geq$ in place of $\leq$.

Property (6): The measure $E_\alpha^\beta(A)$ is a concave function of the probability distribution $A = (a_1, ..., a_n), a_i \geq 0, \sum_{i=1}^n a_i = 1,$ for all $\alpha, \beta > 0; \alpha \neq \beta$ either $\alpha > 1; \beta \leq 1$ or $\alpha > 1; \beta \leq 1$.

Proof: Associated with the random variable $X = (x_1, x_2, ..., x_n)$, let us consider r distributions

$A_k(X) = \{a_k(x_1), a_k(x_2), ..., a_k(x_n)\},$

Where

$\sum_{i=1}^n a_k(x_i) = 1, k = 1, 2, ..., r.$

Next, let there be r numbers $(\lambda_1, \lambda_2, ..., \lambda_r)$ such that $\lambda_k \geq 0, \sum_{k=1}^r \lambda_k = 1$, and define

$A_0(X) = \{a_0(x_1), a_0(x_2), ..., a_0(x_n)\},$

where

$a_0(x_i) = \sum_{k=1}^r \lambda_k a_k(x_i);\ i = 1, 2, ..., n.$

Obviously $\sum_{i=1}^n a_0(x_i) = 1,$ and thus $A_0(X)$ is a bonafide distribution of X .

If $\alpha > 1; \beta < 1,$ then we have

$$\sum_{k=1}^r \lambda_k E_\alpha^\beta(A_k) - E_\alpha^\beta(A_0)$$

$$= \sum_{k=1}^r \lambda_k E_\alpha^\beta(A_k) - \frac{\left[e^{\log_2\left[\sum_{i=1}^r \lambda_i a_i\right]^\alpha} - e^{\log_2\left[\sum_{i=1}^r \lambda_i a_i\right]^\beta} \right]}{\beta - \alpha}$$

$$\leq \sum_{k=1}^r \lambda_k E_\alpha^\beta(A_k) - \frac{\left[e^{\log_2 \sum_{i=1}^r \lambda_i a_i^\alpha} - e^{\log_2 \sum_{i=1}^r \lambda_i a_i^\beta} \right]}{\beta - \alpha}$$ (by Jensen inequality)

$$\sum_{k=1}^r \lambda_i E_\alpha^\beta(A_k) \leq E_\alpha^\beta(A) \tag{2.4}$$

Similarly, for $\alpha \leq 1;\ \beta > 1$, (2.4) holds. Therefore $E_\alpha^\beta(A)$ is a concave function for all $\alpha, \beta > 0; \alpha \neq \beta$.

Implies

$$\left[\sum_{i=1}^n a_0^\alpha(x_i)\right] \leq \left[\sum_{i=1}^n \left(\sum_{k=1}^r \lambda_k a_k^\alpha(x_i)\right)\right] = \sum_{k=1}^r \lambda_k \left(\sum_{i=1}^n a_k^\alpha(x_i)\right).$$

Implies

$$\left[\sum_{i=1}^n a_0^\alpha(x_i)\right] \leq \left[\sum_{k=1}^r \lambda_k \left(\sum_{i=1}^n a_k^\alpha(x_i)\right)\right].$$

Implies

$$e^{\log_D\left[\sum_{i=1}^{n} a_0^{\alpha}(x_i)\right]} \leq e^{\log_D\left[\sum_{k=1}^{r}\lambda_k\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i)\right)\right]} \quad (2.5)$$

But $$e^{\log_D\left[\sum_{k=1}^{r}\lambda_k\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i)\right)\right]} \leq e^{\sum_{k=1}^{r}\lambda_k \log_D\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i)\right)} \quad (2.6)$$

Also, we have $$e^{\sum_{k=1}^{r}\lambda_k \log_D\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i)\right)} \leq \sum_{k=1}^{r}\lambda_k\, e^{\log_D\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i)\right)} \quad (2.7)$$

Therefore, from (2.5), (2.6), (2.7), we get

$$e^{\log_D\left[\sum_{i=1}^{n} a_0^{\alpha}(x_i)\right]} \leq \sum_{k=1}^{r}\lambda_k\, e^{\log_D\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i)\right)} \quad (2.8)$$

Similarly, for $\beta \geq 1$, we have

$$e^{\log_D\left[\sum_{i=1}^{n} a_0^{\beta}(x_i)\right]} \leq \sum_{k=1}^{r}\lambda_k\, e^{\log_D\left(\sum_{i=1}^{n} a_k^{\beta}(x_i)\right)} \quad (2.9)$$

We shall consider the wo cases :

Case 1: If $\alpha > \beta > 1$, then subtract (2.9) from (2.8), we get

$$\left(e^{\log_D\left(\sum_{i=1}^{n} a_0^{\alpha}(x_i)\right)} - e^{\log_D\left(\sum_{i=1}^{n} a_0^{\beta}(x_i)\right)}\right) \leq \sum_{k=1}^{r}\lambda_k\left(e^{\log_D\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i)\right)} - e^{\log_D\left(\sum_{i=1}^{n} a_k^{\beta}(x_i)\right)}\right) \quad (2.10)$$

Since $\beta - \alpha < 0$ for $\alpha > \beta$, we get

$$(\beta-\alpha)^{-1}\left(e^{\log_D\left(\sum_{i=1}^{n} a_0^{\alpha}(x_i)\right)} - e^{\log_D\left(\sum_{i=1}^{n} a_0^{\beta}(x_i)\right)}\right) \geq \sum_{k=1}^{r}\lambda_k(\beta-\alpha)^{-1}\left(e^{\log_D\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i)\right)} - e^{\log_D\left(\sum_{i=1}^{n} a_k^{\beta}(x_i)\right)}\right) \quad (2.11)$$

i.e., $E_{\alpha}^{\beta}(A_0) \geq \sum_{k=1}^{r}\lambda_k E_{\alpha}^{\beta}(A_k)$.

$E_{\alpha}^{\beta}(A)$ is a concave function for $\alpha > \beta > 1$.

Case 2. $E_{\alpha}^{\beta}(A)$ is also a concave function for $\beta > \alpha > 1$. But the inequality in (iii) is reversed. Applying $\beta - \alpha > 0$ for $\beta > \alpha$ in iii, we get (2.11)

Therefore, $E_{\alpha}^{\beta}(A)$ is a concave function of A for $\alpha, \beta > 1$.This complete the proof of property (6).

3. Measure of Exponential Relative Information of 'Type (α, β)'

Distance measures between two probability distribution play an important role in the probability theory, statistical inference, signal processing, pattern recognition, finance, economics etc. A class of measure which may not satisfied all the conditions of distance measures or metric space is called divergence measures.

Divergence measures based on the idea of information-theoretic entropy was intially introduced in communication theory by Shannon [27]. Kullback and Leibler [12] introduced relative entropy or the divergence measures between two probability distributions as a generalization of the Shannon entropy. During the half past century, various extensions of the Shannon entropy and Kullback-Leibler divergence measures are introduced by various authors like as Renyi [23], Rao [24], Kapur [15], Vajda [29], Lin

[18], Pardo [21], Shioya and Da-te [26], Ali and Silvey [2], Csiszar [5], Sharma and Mittal [25], Burbea and Rao [3], Havrda and Charvat [10], Takuya Yamano [30].

Dragomir [6] introduced the concept of p -Logarithmic and α -power divergence measures and derived a number of basic results. Friedman and Sandow [9] introduced a utility-based generalization of Shannon entropy and Kullback-Leibler information measures. Friedman *et al.* [8] proved various properties for these generalized quanitities similer to the Kullback-Leibler information measure.

Let $A = \{(a_{1,} a_2, ..., a_n), 0 \le a_i \le 1, \sum_{i=1}^{n} a_i = 1\}$ and $B = \{(b_1, b_2, ..., b_n), 0 \le b_i \le 1, \sum_{i=1}^{n} b_i = 1\}$ be probability distributions, then Kullback-Leibler divergence measure is defined as:

$$KL(A;B) = \sum_{i=1}^{n} (a_i) \log(a_i/b_i) \tag{3.1}$$

Generally, metric divergence measures such as Euclidean distance, $D(A;B) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$, satisfy four conditions:

(i) Non-negativity: $D(A;B) \ge 0$.

(ii) Identity: $D(A;B) = 0$ iff $a_i = b_i$ for each i .

(iii) Symmetry: $D(A;B) = D(B;A).$

(iv) Triangular inequality: $D(A;B) + D(B;R) \ge D(A;R).$

Kullback-Leibler directed divergence satisfies the first two conditions of metric measures; but not the third and fourth conditions as they are not essential for a measure of discrepancy. Instead, it possesses two important conditions which are useful for optimization purposes:

(v) $D(A;B)$ is a convex function of $(a_{1,} a_2, ..., a_n)$.

(vi) When this measure is minimized subject to some linear constraints the minimizing probabilities are all non-negative.

Some properties of Kullback-Leibler's measure and their proof can be found in Lexa [17] and Kullback [13].

In addition to the above properties, if B is a uniform distribution. i.e., $B = \left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right)$ then,

$$KL(A;B) = \sum_{i=1}^{n} (a_i) \log\left(\frac{a_i}{1/n}\right) = \sum_{i=1}^{n} (a_i) \log(a_i) + \log(n) = -H(A) + \log(n)$$.

Hence, $\mathrm{KL}(A;B) = -H(A) + \log(n)$.

Where $\log(n)$ is constant. Thus, from this perspective, the Shannon entropy measure can be considered a special case of Kullback-Leibler directed divergence measure, however they are different conceptually as Shannon entropy is an uncertainty concept; but Kullback-Leibler measures the directed -divergence between two probability distributions.

In the next section, we review the generalization of Shannon entropy and Kullback-Leibler information measure. Then, we introduced a generalized the exponential divergence convex function of 'type (α, β) ' and derive a number of basic properties.

4. Generalized Exponential Divergence Measure of 'Type (α, β)

However, we noted that Kullback-Leibler measure satisfies conditions (i), (ii), (v) and, (vi), there are also other measures that satisfy those four conditions and thus qualify as legitimate measures of directed divergence. Even if a measure satisfies only conditions (i), (ii), and (v), but not (vi), it can still be considered as a measure of directed divergence (See Csisz {\a`}r, [5]). These measures are called generalized measures of directed divergence. There are various versions on generalization of Shannon entropy and Kullback-Leibler information measure. For all probability distribution of A and B , we propose the function

$$\mathrm{D}_{\alpha}^{\beta}(A;B) = \frac{1}{\alpha - \beta}\left[e^{\log_2\left[\sum_{i=1}^{n} a_i^{\alpha} b_i^{1-\alpha}\right]} - e^{\log_2\left[\sum_{i=1}^{n} a_i^{\beta} b_i^{1-\beta}\right]}\right], \text{ for all } \alpha, \beta > 1; \alpha \ne \beta \tag{4.1}$$

Remarks

a) When the base of the logrithm is e. Then (4.1) becomes Sharma and Mittal [25] relative information measure.

b) When the base of the logarithm is e and $\beta = 1/\alpha$, then (4.1) becomes

$$\mathrm{D}_{\alpha}^{(1/\alpha)}(A;B) = \frac{1}{\alpha - (1/\alpha)}\left[\sum_{i=1}^{n} a_i^{\alpha} b_i^{1-\alpha} - \sum_{i=1}^{n} a_i^{(1/\alpha)} b_i^{1-(1/\alpha)}\right] \tag{4.2}$$

Which is studied by T.Yamano [30]. The author called it a generalized Kullback-Leibler relative entropy. Further, he present the fundamental properties including positivity, metricity, concavity, bounds and stability. In addition, a connection to shift information and behavior under Liouville dynamics are discussed.

c) If the base of the logarithm is e and $\beta=1$ or $\alpha=1$. Then (4.1) becomes Tsallis' -Havrda – Charvat [10, 28] relative entropy.

d) If the base of the logarithm is e and $\beta=1,\alpha\to 1$ or $\alpha=1,\beta\to 1$. Then (4.1) becomes Kullback-Leibler divergence measure.

It may be noted that $D_\alpha^\beta(A;B)$ is not a metric as it does not satisfy triangle inequality. However, $D_\alpha^\beta(A;B)$ satisfies convexity property which we need to minimize $D_\alpha^\beta(A;B)$ as functions of A and B subject to $\sum_{i=1}^n a_i=\sum_{i=1}^n b_i=1$. This convexity property ensures that in each case, a local minimum will be global. The function $D_\alpha^\beta(A;B)$ satisfies all three conditions (i), (ii) and (v).
For this we need to prove the following results:

Result 1: For all probability distribution A and B it holds, $D_\alpha^\beta(A;B)\ge 0$ for all $\alpha,\beta>1;\alpha\ne\beta$ and the equality holds if $a_i=b_i, \forall\, i=1,2,...,n$.

Proof: By Jensen's inequality for $\alpha>1$, we have

$$\left[\sum_{i=1}^n a_i\left(\frac{b_i}{a_i}\right)\right]^{1-\alpha}\le \sum_{i=1}^n a_i\left(\frac{b_i}{a_i}\right)^{1-\alpha}=\left[\sum_{i=1}^n a_i^\alpha b_i^{1-\alpha}\right]\quad\Rightarrow \sum_{i=1}^n a_i^\alpha b_i^{1-\alpha}\ge 1.$$

$$\Rightarrow e^{\log_2\left[\sum_{i=1}^n a_i^\alpha b_i^{1-\alpha}\right]}\ge 1. \tag{4.3}$$

Let $\beta>1$, then

$$e^{\log_2\left[\sum_{i=1}^n a_i^\beta b_i^{1-\beta}\right]}\ge 1.$$

We consider the following cases

Case (1): Let $\alpha>\beta>1$, then $e^{\log_2\left[\sum_{i=1}^n a_i^\alpha b_i^{1-\alpha}\right]}-e^{\log_2\left[\sum_{i=1}^n a_i^\beta b_i^{1-\beta}\right]}\ge 0$, and $\alpha-\beta>0$. (4.4)

$$\Rightarrow \frac{e^{\log_2\left[\sum_{i=1}^n a_i^\alpha b_i^{1-\alpha}\right]}-e^{\log_2\left[\sum_{i=1}^n a_i^\beta b_i^{1-\beta}\right]}}{\alpha-\beta}\ge 0.$$

Therefore, $D_\alpha^\beta(A;B)\ge 0$.

Case (2): Let $\beta>\alpha>1$, then $e^{\log_2\left[\sum_{i=1}^n a_i^\alpha b_i^{1-\alpha}\right]}-e^{\log_2\left[\sum_{i=1}^n a_i^\beta b_i^{1-\beta}\right]}\le 0$, and $\alpha-\beta<0$, implies $D_\alpha^\beta(A;B)\ge 0$.

As $\beta=1,\alpha\to 1$ or $\alpha=1,\beta\to 1$, $D_\alpha^\beta(A;B)\to\left[\sum_{i=1}^n (a_i)\log(a_i/b_i)\right]$, is a consequence of the fact that the measure (4.1) is a continuous function of α or β which becomes a Kullback-Leibler relative information measure.

Also, for $a_i=b_i$ for each i, $D_\alpha^\beta(A;B)=0$. It is fact that the function (4.1) gives minimum value, if $a_i=b_i$ for $i=1,2,...,n.$

Result 2: $\frac{1}{\alpha-\beta}\left[e^{\log_2\left[\sum_{i=1}^n a_i^\alpha b_i^{1-\alpha}\right]}-e^{\log_2\left[\sum_{i=1}^n a_i^\beta b_i^{1-\beta}\right]}\right]$ is a convex function of A for $\alpha,\beta>1;\alpha\ne\beta$.

Proof: Let $A_0(X)=\{a_0(x_1),a_0(x_2),...,a_0(x_m)\}$ be the probability distribution of X such that $a_0(x_i)=\sum_{k=1}^r \lambda_k a_k(x_i)$, where λ_k are no n-negative numbers which sum to unity and

$A_k(X)=\{a_k(x_i):k=1,2,...,r\}$ are some probability distribution of X. For convex function, we shall use the following Jensen's inequality:

$$\left[\sum_{k=1}^{r} m_k x_k\right]^t \le \left[\sum_{k=1}^{r} m_k x_k^t\right]$$, according as $t>1$, we have

$$\left[\sum_{i=1}^{r} \lambda_k a_k(x_i)\right]^{\alpha} \le \left[\sum_{i=1}^{r} \lambda_k a_k^{\alpha}(x_i)\right]$$, according as $\alpha>1$.

Therefore,

$$\left[\sum_{i=1}^{n}\left[\sum_{k=1}^{r} \lambda_k a_k(x_i)\right]^{\alpha} b_i^{1-\alpha}\right] \le \left[\sum_{i=1}^{n}\left[\sum_{k=1}^{r} \lambda_k a_k^{\alpha}(x_i)\right] b_i^{1-\alpha}\right].$$

Implies

$$\left[\sum_{i=1}^{n} a_0^{\alpha}(x_i) b_i^{1-\alpha}\right] < \left[\sum_{k=1}^{r} \lambda_k \left(\sum_{i=1}^{n} a_k^{\alpha}(x_i) b_i^{1-\alpha}\right)\right].$$

Implies $e^{\log_2\left[\sum_{i=1}^{n} a_0^{\alpha}(x_i) b_i^{1-\alpha}\right]} \le e^{\log_2\left[\sum_{k=1}^{r} \lambda_k \left(\sum_{i=1}^{n} a_k^{\alpha}(x_i) b_i^{1-\alpha}\right)\right]}$.

Now, we know the convexity of log and exponational function as

$$e^{\log_2\left[\sum_{i=1}^{n} a_0^{\alpha}(x_i) b_i^{1-\alpha}\right]} \le e^{\log_2\left[\sum_{k=1}^{r} \lambda_k \left(\sum_{i=1}^{n} a_k^{\alpha}(x_i) b_i^{1-\alpha}\right)\right]} \le e^{\sum_{k=1}^{r} \lambda_k \log_2\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i) b_i^{1-\alpha}\right)} \le \sum_{k=1}^{r} \lambda_k\, e^{\log_2\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i) b_i^{1-\alpha}\right)}$$

$$\Rightarrow \quad e^{\log_2\left[\sum_{i=1}^{n} a_0^{\alpha}(x_i) b_i^{1-\alpha}\right]} \le \sum_{k=1}^{r} \lambda_k\, e^{\log_2\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i) b_i^{1-\alpha}\right)} \quad . (*)$$

Similarly, for $\beta>1$, we have

$$e^{\log_2\left[\sum_{i=1}^{n} a_0^{\beta}(x_i) b_i^{1-\beta}\right]} \le \sum_{k=1}^{r} \lambda_k\, e^{\log_2\left(\sum_{i=1}^{n} a_k^{\beta}(x_i) b_i^{1-\beta}\right)} \quad . (**)$$

We shall consider the wo cases

Case 1: If $\alpha>\beta$, then subtract (**) from (*),we get

$$\left(e^{\log_2\left(\sum_{i=1}^{n} a_0^{\alpha}(x_i) b_i^{1-\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_0^{\beta}(x_i) b_i^{1-\beta}\right)}\right) \le \sum_{k=1}^{r} \lambda_k \left(e^{\log_2\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i) b_i^{1-\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_k^{\beta}(x_i) b_i^{1-\beta}\right)}\right) \quad . iii$$

Since, $\alpha-\beta>0$ for $\alpha>\beta$, we get

$$(\alpha-\beta)^{-1}\left(e^{\log_2\left(\sum_{i=1}^{n} a_0^{\alpha}(x_i) b_i^{1-\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_0^{\beta}(x_i) b_i^{1-\beta}\right)}\right) \le \sum_{k=1}^{r} \lambda_k (\alpha-\beta)^{-1}\left(e^{\log_2\left(\sum_{i=1}^{n} a_k^{\alpha}(x_i) b_i^{1-\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_k^{\beta}(x_i) b_i^{1-\beta}\right)}\right) \quad .1*$$

i.e., $D_\alpha^\beta(A_0;B) \le \sum_{k=1}^{r} \lambda_k D_\alpha^\beta(A_k;B)$.

$D_\alpha^\beta(A;B)$ is a convex function for $\alpha > \beta > 1$.

Case 2. $D_\alpha^\beta(A;B)$ is also a convex function for $\beta > \alpha > 1$.But the inequality in (iii) is reversed. Applying $\alpha - \beta < 0$ for $\beta > \alpha$ in iii,we get 1*

Therefore, $D_\alpha^\beta(A;B)$ is a convex function of A for $\alpha, \beta > 1$.This complete the proof of result 2.

Implies

$$\left[\sum_{i=1}^{n} a_0^\alpha(x_i) b_i^{1-\alpha}\right] \le \left[\sum_{k=1}^{r} \lambda_k \left(\sum_{i=1}^{n} a_k^\alpha(x_i) b_i^{1-\alpha}\right)\right].$$

Implies $$e^{\log_2\left[\sum_{i=1}^{n} a_0^\alpha(x_i) b_i^{1-\alpha}\right]} \le e^{\log_2\left[\sum_{k=1}^{r} \lambda_k \left(\sum_{i=1}^{n} a_0^\alpha(x_i) b_i^{1-\alpha}\right)\right]}.$$

Now, using convexity of log and exponential function, we get

$$e^{\log_2\left[\sum_{i=1}^{n} a_0^\alpha(x_i) b_i^{1-\alpha}\right]} \le e^{\log_2\left[\sum_{k=1}^{r} \lambda_k \left(\sum_{i=1}^{n} a_k^\alpha(x_i) b_i^{1-\alpha}\right)\right]} \le e^{\sum_{k=1}^{r} \lambda_k \log_2\left(\sum_{i=1}^{n} a_k^\alpha(x_i) b_i^{1-\alpha}\right)} \le \sum_{k=1}^{r} \lambda_k e^{\log_2\left(\sum_{i=1}^{n} a_k^\alpha(x_i) b_i^{1-\alpha}\right)}$$

$$\Rightarrow e^{\log_2\left(\sum_{i=1}^{n} a_0^\alpha(x_i) b_i^{1-\alpha}\right)} \le \sum_{k=1}^{r} \lambda_k e^{\log_2\left(\sum_{i=1}^{n} a_k^\alpha(x_i) b_i^{1-\alpha}\right)} \tag{4.5}$$

Similarly, for $\beta > 1$, we have

$$\Rightarrow e^{\log_2\left(\sum_{i=1}^{n} a_0^\beta(x_i) b_i^{1-\beta}\right)} \le \sum_{k=1}^{r} \lambda_k e^{\log_2\left(\sum_{i=1}^{n} a_k^\beta(x_i) b_i^{1-\beta}\right)}. \tag{4.6}$$

We shall consider the two cases

Case (1): If $\alpha > \beta > 1$, then subtract (4.6) from (4.5), we get

$$\left[e^{\log_2\left(\sum_{i=1}^{n} a_0^\alpha(x_i) b_i^{1-\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_0^\beta(x_i) b_i^{1-\beta}\right)}\right] \le \sum_{k=1}^{r} \lambda_k \left[e^{\log_2\left(\sum_{i=1}^{n} a_k^\alpha(x_i) b_i^{1-\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_k^\beta(x_i) b_i^{1-\beta}\right)}\right]. \tag{4.7}$$

Since, $\alpha - \beta > 0$ for $\alpha > \beta$, we get

$$(\alpha-\beta)^{-1}\left(e^{\log_2\left(\sum_{i=1}^{n} a_0^\alpha(x_i) b_i^{1-\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_0^\beta(x_i) b_i^{1-\beta}\right)}\right) \le \sum_{k=1}^{r} \lambda_k (\alpha-\beta)^{-1}\left(e^{\log_2\left(\sum_{i=1}^{n} a_k^\alpha(x_i) b_i^{1-\alpha}\right)} - e^{\log_2\left(\sum_{i=1}^{n} a_k^\beta(x_i) b_i^{1-\beta}\right)}\right).$$

i.e., $$D_\alpha^\beta(A_0;B) \le \sum_{k=1}^{r} \lambda_k D_\alpha^\beta(A_k;B). \tag{4.8}$$

$D_\alpha^\beta(A;B)$ is a convex function for $\alpha > \beta > 1$.

Case (2): $D_\alpha^\beta(A;B)$ is also a convex function for $\beta > \alpha > 1$.But the inequality in (4.7) is reversed and applying $\alpha - \beta < 0$ for $\beta > \alpha$, we get (4.8). Therefore, $D_\alpha^\beta(A;B)$ is a convex function of A for $\alpha, \beta > 1$. This complete the proof of result 2.

5. Measure of Information

Corresponding to the generalized measure of divergence (4.1), there is a unique measure of generalized entropy. let $C = \left(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}\right)$, be the uniform distribution, then generalized measures of directed divergence is defined as

$$D_\alpha^\beta(A;C) = \frac{1}{\alpha-\beta}\left[e^{\log_2\left[\sum_{i=1}^{n} a_i^\alpha n^{\alpha-1}\right]} - e^{\log_2\left[\sum_{i=1}^{n} a_i^\beta n^{\beta-1}\right]} \right]$$

(i)* $D_\alpha^\beta(A;C) \geq 0$.

$(ii)^*$ $D_\alpha^\beta(A;C) = 0, \ if \ A = C$.

$(iii)^*$ $D_\alpha^\beta(A;C)$ is a convex function of A.

We see that

$$D_\alpha^\beta(A;C) = \frac{1}{\alpha-\beta}\left[e^{(\alpha-1)\log_2 n} - e^{(1-\alpha)\log_2 n} + e^{(1-\beta)\log_2 n} - e^{(\beta-1)\log_2 n} + e^{\log_2 \sum_{i=1}^{n} a_i^\alpha n^{\alpha-1}} - e^{\log_2 \sum_{i=1}^{n} a_i^\beta n^{\beta-1}} \right] \quad (5.1)$$

Now we newly define

$$E_\alpha^\beta(A;C) = \frac{1}{\alpha-\beta}\left[e^{\log_2 \sum_{i=1}^{n} a_i^\alpha n^{\alpha-1}} - e^{(\alpha-1)\log_2 n} + e^{(\beta-1)\log_2 n} - e^{\log_2 \sum_{i=1}^{n} a_i^\beta n^{\beta-1}} \right] \quad (5.2)$$

and

$$E_\alpha^\beta(C;C) = \begin{cases} \frac{1}{\alpha-\beta}\left[e^{(\alpha-1)\log_2 n} - e^{(\beta-1)\log_2 n} \right]; & \alpha, \beta > 1; \alpha \neq \beta \\ \log(n)\ ; & if\ \alpha = 1; \beta \to 1\ or\ \beta = 1; \alpha \to 1 \end{cases}, n = 2, 3, \ldots . \quad (5.3)$$

Therefore, $D_\alpha^\beta(A;C) = E_\alpha^\beta(C;C) - E_\alpha^\beta(A;C)$. (5.4)

So to minimize $D_\alpha^\beta(A;C)$ is equivalent to maximize $E_\alpha^\beta(A;C)$.

$(iv)^*$ $E_\alpha^\beta(C;C) \geq E_\alpha^\beta(A;C)$.

$(v)^*$ $E_\alpha^\beta(C;C) = E_\alpha^\beta(A;C)$ if $A = C$

$(vi)^*$ $E_\alpha^\beta(A;C)$ is a concave function of A.

Thus, the function $E_\alpha^\beta(A;C)$ is a new information measure of A and C corresponding to directed- divergence measure $D_\alpha^\beta(A;C)$.

Theorem 5.1: The generalized relative entropy, $D_\alpha^\beta(A;B)$ and the generalized entropy, $E_\alpha^\beta(A;C)$, have the following properties:

$(i)^{**}$ $D_\alpha^\beta(A;B) \geq 0$ with equality if and only if $A = B$.

$(ii)^{**}$ $D_\alpha^\beta(A;B)$ is a convex function of A.

$(iii)^{**}$ $E_\alpha^\beta(A;C) \geq 0$, $E_\alpha^\beta(A;C)$ is a concave function of A.

6. Conclusions

We have investigated properties of a generalized Shannon entropy and Kullback-Leibler divergence in the context of information theory. As a conclusion, we remark that this exponential entropy 'type (α, β)' is also a pertinent information measure which is a generalization of Tsallis' -Havrda -Charvat, Takuya Yamano, Sharma and Mittal.

7. References

1. Arndt C. Information Measure-Information and its description in Science and Engineering, Springer, Berlin, 2001.
2. Ali SM, Silvey SD. A general class of coefficients of divergence of one distribution from another, Journal of the Royal Statistical Society, Series B. 1966; 28:131-142.
3. Burbea J, Rao CR. On the convexity of some divergence measures based on entropy functions, IEEE Transactions on Information Theory. 1982; 28:489-495.
4. Campbell LL. Exponential entropy as a measure of extent of distribution, Z. Wahrscheinlichkeitstheorie verw. Geb. 1966; 5:217-225.
5. Csisz {\a`}r J. Information type measure of differences of probability distribution and indirect observations, Studia Scientifica materia hungary. 1967; 2:299-318.
6. Dragomir SS. On the p -Logarithmic and α -power divergence measures in information theory, PanAmerican mathematical Journal. 2003; 13:1-10.
7. Feinstein A. Foundation of Information Theory, McGraw Hill, New York, 1956.
8. Friedman C, Huang J, Sandow S. A utility-based approach to some information measures, Entropy. 2007; 9:1-26.
9. Friedman C, Sandow S. Model performance measures for expected utility maximizing investors, International Journal of Theoretical and Applied Finance. 2003; 6:355-401.
10. Havrda JH, Charvat F. Quantification method classification process: Concept of structural α -entropy, Kybernetika. 1967; 3:30-35.
11. Kerridge DF. Inaccuracy and inference, J. Royal Statist Soc. Ser. B. 1961; 23:184-194.
12. Kullback S, Leibler RA. On information and sufficiency, Ann. Maths Statistics. 1951; 22:79-86.
13. Kullback S. Information Theory and Statistics, Wiley, New York, 1959.
14. Kapur JN. Measure of information and their applications, 1st ed., New Delhi, Wiley Eastern Limited, 1994.
15. Kapur JN. A comparative assessment of various measures of directed divergence, Advances in management studies. 1984; 3:1-16.
16. Koski T, Persson LE. Some properties of generalized exponential entropies with applications to data compression, Information Sciences. 1992; 62:103-132.
17. Lexa M. Useful facts about the Kullback-Leibler distance, Rice Univ. Houston, Tx. tech. Rep. 2004.
18. Lin J. Divergence measure based on the Shannon entropy, IEEE Transactions on Information Theory. 1991; 37:145-151.
19. Nielsen F, Nock R. On R`enyi and Tsallis' entropies and divergences for exponential families, IEEE Transactions on Information Theory. 2011; 20:10-30.
20. Nath P. An axiomatic characterization of inaccuracy for discrete generalized probability distribution, Opsearch. 1970; 7:115-133.
21. Pardo L. R_h^{φ} -Divergence statistics in applied categorical data analysis with stratified sampling, Utilitas Mathematica. 1993; 44:145-164.
22. Rathie PN, Kannappan PI. A Directed-Divergence function of type β, Inform. and Control. 1972; 20:38-45.
23. Renyi A. On measures of entropy and information, Proc. 4th Barkley Symp. on Math. Stat. and Probability, University of Californa Press. 1961; 1:547-561.
24. Rao CR. Diversity and dissimilarity coefficients: A unified approach, Theoretic Population Biology. 1982; 21:24-43.
25. Sharma BD, Mittal DP. New non-additive measure of relative information, Journal of combinatorics information and systems science. 1977; 2:122-132.
26. Shioya H, Da-te T. A generalization of Lin divergence and the derivative of a new information divergence, Electronics and Communications in Japan. 1995; 78:37-40.
27. Shannon CE. A Mathematical Theory of Communication, The Bell System Tech. J. 1948; 27:379-423, 623-656.
28. Tsallis C. Possible generalization of Boltzmann-Gibbs statistics J. Stat. Phys. 1988; 52:480-487.
29. Vajda I. Theory of Statistical Inference and Information, Kluwer Academic Publishers, Dordrecht-Boston, 1989.
30. Yamano T. A generalization of the Kullback-Leibler divergence and its properties, Journal of Mathematical Physics. 2009; 50:043302. doi: 10.1063/1.3116115.
31. Garg D, Kumar S. Parametric R-norm directed-divergence convex function, Infinite Dimensional Analysis, Quantum Probability and Related Topics, World Science Journals. 2016; 19:2. DOI: 10.1142/S0219025716500144
32. Garg D, Kumar S. An application of generalized Tsalli's-Havrda-Charvat entropy in coding theory through a generalization of Kraft inequality, International Journal of Applied Research. 2016; 1(4):01-05. ISSN: 2394-5869.
33. Garg D, Kumar S. Some Remarks On 'Useful' Renyi's Entropy of Order α in International Journal of trends in Mathematics and Statistics (IJTIMS). 2015; 4(10):178-183. ISSN: 2279-0318.