\[
\frac{\mu_A^R(x_i) + (1 - \mu_A(x_i))^R}{\sum_{i=1}^n u_i\left(\mu_A^{R+v-1}(x_i) + (1-\mu_A(x_i))^{R+v-1}\right)} \geq D^{-l_i}
\] (3.3.13)

Multiplying both sides \(\sum_{i=1}^n u_i\left(\mu_A^{R+v-1}(x_i) + (1-\mu_A(x_i))^{R+v-1}\right)\) and summing over \(i = 1, 2, \ldots, n\), we get (3.3.3). The last inequality in (3.3.12) gives:

\[
l_i < - \log \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R\right) + \log \frac{\sum_{i=1}^n u_i\left(\mu_A^{R+v-1}(x_i) + (1-\mu_A(x_i))^{R+v-1}\right)}{\sum_{i=1}^n u_i\left(\mu_A^v(x_i) + (1-\mu_A(x_i))^v\right)} + 1
\]

Or

\[
l_i < - \log \frac{\left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R\right)}{\sum_{i=1}^n u_i\left(\mu_A^{R+v-1}(x_i) + (1-\mu_A(x_i))^{R+v-1}\right)} + \log_D D
\]

\[
D^{-l_i} > \frac{\left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R\right)}{\sum_{i=1}^n u_i\left(\mu_A^{R+v-1}(x_i) + (1-\mu_A(x_i))^{R+v-1}\right)} D^{-1}
\]

Taking \(0 < R < 1\) and raising both sides to the power \(\frac{R-1}{R}\), we get;

\[
D^{-l_i}\left(\frac{R-1}{R}\right) < \left(\frac{\left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R\right)}{\sum_{i=1}^n u_i\left(\mu_A^{R+v-1}(x_i) + (1-\mu_A(x_i))^{R+v-1}\right)}\right)^{\frac{R-1}{R}} D^{-\frac{R-1}{R}}
\]

Multiplying both sides by \(\frac{u_i\left(\mu_A^v(x_i) + (1-\mu_A(x_i))^v\right)}{\sum_{i=1}^n u_i\left(\mu_A^v(x_i) + (1-\mu_A(x_i))^v\right)}\) and summing over \(i = 1, 2, \ldots, n\) and after simplifying, gives (3.3.9).

For \(R > 1\), the proof follows along the similar lines.

It is well known that a lower bound on the average length is obtained in terms of Shannon entropy [87] for instantaneous codes in noiseless channel (Abramson[1]). Bernard and Sharma [23] studied variable length codes for noisy channels and presented some combinatorial bounds for this variable length, error correcting codes. Bernard and Sharma [24]
obtained a lower bound on average for variable length error correcting codes satisfying the
criterion of promptness.

In this chapter, we propose a new generalized fuzzy entropy measure using segment
decomposition and effective range and study its particular cases. Also some fuzzy coding
theorems have been established.

4.1. Introduction:-

Incoding theory, it is assumed that \( Q \) is a finite set of alphabets and there are \( D \) code
characters. A codeword is defined as a finite sequence of code characters and a variable length
code \( C \) of size \( K \) is a set of \( K \) code words denoted by \( c_1, c_2, ..., c_k \) with lengths \( n_1, n_2, ..., n_k \)
respectively. Without loss of generality it may be assumed that \( n_1 \leq n_2 \leq \cdots \leq n_k \).

The channel, which is considered here, is not noiseless. In other words, the codes
considered are error correcting codes. The criterion for error correcting is defined in terms of a
mapping \( \alpha \), which depends on the noise characteristics of the channel. This mapping \( \alpha \) is called
the error admissibility mapping. Given codeword ‘\( c \)’ and error admissibility \( \alpha \), the set of
codeword’s received over the channel when \( c \) was sent, denoted by \( \alpha(c) \) is the error range of \( c \).

Various kinds of error pattern can be described in terms of mapping \( \alpha \). In particular, \( \alpha \) may be
defined as (Bernard & Sharma [23])

\[
\alpha_e(c) = \{u|w(c - u) \leq e\},
\]

Where \( e \) is the random substitution error and \( w(c - u) \) is the Humming weight, i.e. the number
of non-zero coordinates of \( (c - u) \). It can be easily verified by Bernard and Sharma [23] that the
number of sequences in \( \alpha_e(c) \) denoted as \( |\alpha_e(c)| \) is given by

\[
|\alpha_e(c)| = \sum_{i=0}^{n} \binom{n}{i}(D - 1)^i,
\]

where \( n \) is the length of codeword \( c \).

We may assume that \( \alpha_0 \) corresponds to the noiseless. In other words, if \( c \) is sent then \( c \) is
received w.r.t. \( \alpha_0 \). Moreover it is clear that \( |\alpha_e(c)| \) depends only on the length \( n \) of \( c \) when \( \alpha \) and
\( D \) are given. In noiseless coding, the class of uniquely decodable instantaneous codes is studied. It
is known that these codes satisfy prefix property (Abramson [1]).
In the same way Hartnett [49] studied variables length code over noisy channel, satisfying the prefix property in the range. These codes are called \( \alpha \)-prompt codes. Such codes have the property that they can decode promptly.

Further, Burnard and Sharma [23] gave a combinational information inequality that must necessarily be satisfied by code word lengths of prompt code codes. Two useful concepts, namely, segment decomposition and the effective range \( r_\alpha(c_i) \) of code words \( c_i \) of length \( n_i \) under error mapping \( \alpha \) as the Cartesian product of ranges of the segment are also given by Bernard and Sharma [23]. The numbers of sequences in effective range of \( c_i \) denoted by \( |r_\alpha|_{n_i} \) depends only on \( \alpha \) and \( n_i \). It is given that:

\[
|r_\alpha|_{n_i} = |\alpha|_{n_1} |\alpha|_{n_2} \cdots |\alpha|_{n_{i-1}} - n_{i-1}.
\]

Also, we adopt the notion \( |\alpha|_0 = 1 \). Moreover, Bernard and Sharma [23] obtained the following inequality.

**Lemma 4.1.1:** For any set of length \( n_1 \leq n_2 \leq \cdots \leq n_k \)

\[
|r_\alpha|_{n_i} = |r_\alpha|_{n_{i-1}} |r_\alpha|_{n_{i-1}}
\]

**Proof:** The proof easily follows from the definition of the effective range.

We have

\[
|r_\alpha|_{n_i} = |\alpha|_{n_1} |\alpha|_{n_2-n_1} \cdots |\alpha|_{n_{i-1}}
\]

and

\[
|r_\alpha|_{n_{i-1}} = |\alpha|_{n_1} |\alpha|_{n_2-n_1} \cdots |\alpha|_{n_{i-2}}
\]

Therefore

\[
|r_\alpha|_{n_i} = |r_\alpha|_{n_{i-1}} |r_\alpha|_{n_{i-1}}
\]

**Theorem 4.1.1:** An \( \alpha \)-prompt code with \( k \) code words of length \( n_1 \leq n_2 \leq \cdots \leq n_k \), satisfies the following inequality.

\[
\sum_{i=1}^{k} |r_\alpha|_{n_i} D^{-n_i} \leq 1 \tag{4.1.1}
\]

**Proof:** Let \( N_i \) denote the number of code words of length \( i \) in the code. Then, since the range of the word of length one has to be disjoint, we have;

\[
N_1 \leq \frac{q}{|r_\alpha|_1} = \frac{q}{|\alpha|_1} = \frac{q}{q} = 1
\]
Next, we know that for a code to be $\alpha$-prompt, no sequence in the range of a code word can be prefix of any sequence in the range of another code word. Since $N_1 \leq 1$, if there are more than one code word and some noise effect is there, then we will not able to get any word of length one and we will have to consider words of length 2 or more only.

The first digit will be one of the code symbols, i.e. for forming words of larger than $N_1 = 0$ and the first position can be filled in just one way for purpose of uniformity of arguments at larger stages. We will see that the first position can be filled in $\left\lfloor \frac{D}{|r_{a_1}|} - N_1 \right\rfloor$ ways.

The number of symbols that may be added at the second position is at most $\frac{D}{|a_1|}$ from Lemma 4.1.1. Thus, we will have;

$$N_2 \leq \left\lfloor \frac{D}{|r_{a_1}|} - N_1 \right\rfloor \left\lfloor \frac{D}{|r_{a_2}|} \right\rfloor$$

$$= \frac{D^2}{|r_{a_2}|} - N_1 \cdot D \frac{|r_{a_1}|}{|r_{a_2}|}$$

Now to form words of length 3, only those sequences of length 2 which are not code words can be accepted as permissible prefix. Their number is;

$$\frac{D^2}{|r_{a_2}|} - N_1 \cdot D \frac{|r_{a_1}|}{|r_{a_2}|} - N_2.$$  

Once again, the number of symbols that may be added in the third position is $\frac{D}{|a_1|}$. From Lemma 4.1.1, we can take $\frac{D}{|a_1|} = D \frac{|r_{a_2}|}{|r_{a_3}|}$.

Thus, $N_3 \leq \left\lfloor \frac{D^2}{|r_{a_2}|} - N_1 D \frac{|r_{a_1}|}{|r_{a_2}|} - N_2 \right\rfloor \left\lfloor \frac{D}{|r_{a_3}|} \right\rfloor$

$$= \frac{D^3}{|r_{a_3}|} - N_1 D^2 \frac{|r_{a_1}|}{|r_{a_2}|} - N_2 D \frac{|r_{a_2}|}{|r_{a_3}|}$$

We may proceed in the same manner to obtain results for various $N_i$'s. For the last length $n_k$, we will have;

$$N_{n_k} \leq \frac{D^{n_k}}{|r_{a_1}|^{n_k} - N_1 D^{n_{k-1}} \frac{|r_{a_1}|}{|r_{a_1}|^{n_k}} - N_2 D^{n_{k-2}} \frac{|r_{a_2}|}{|r_{a_1}|^{n_k}} - \ldots - N_{n_k-1} D \frac{|r_{a_1}|^{n_k-1}}{|r_{a_1}|^{n_k}}}.$$
This can be written as \[ \sum_{i=1}^{k} |r_{a_i}| n_i D^{-i} \leq 1. \]

Changing the summation from the length 1, 2, ..., \(n_k\) to the code word length \(n_1, n_2, ..., n_k\). The above inequality can be equivalently put as \[ \sum_{i=1}^{k} |r_{a_i}| n_i D^{-n_i} \leq 1, \] which proves the theorem.

**Remark 4.1.1:** If the codes of constant length \(n\) are taken, then the average inequality (4.1.1) reduces to Hamming sphere packing bound (Hamming [47]).

**Remark 4.1.2:** If the channel is noiseless, the inequality (4.1.1) reduces to the well known Kraft inequality (Kraft [65]). Bernard and Sharma [24] have obtained a lower bound on average code word length for prompt code using a quantity similar to Shannon entropy.

Campbell [28] considered a code length of order \(t\) defined by;

\[
L(t) = \frac{1}{t} \log_{\mathbb{D}} \sum_{i=1}^{k} (p_i D^{tn_i}); \quad (0 < t < \infty) \quad (4.1.2)
\]

An application of L-Hospitals rule shows that

\[
L(0) = \lim_{t \to 0} L(t) = \sum_{i=1}^{k} n_i p_i \quad (4.1.3)
\]

For large \(t\), \[ \sum_{i=1}^{k} n_i D^{tn_i} \approx p_j D^{tn_j} \], where \(n_j\) is the largest of the numbers \(n_1, n_2, ..., n_k\). Moreover, \(L(t)\) is a monotonic non-decreasing function of \(t\) (Beckenbach and Bellman [21]). Thus \(L(0)\) is the conventional measure of mean length and \(L(\infty)\) is the measure which would be used if the maximum length were of prime importance.

**Definition 4.1.1:** Fuzzy sets are sets whose elements have degrees of membership. Fuzzy sets were introduced by Lotfi A. Zadeh [109] gave an extension of classical notion of set. In classical set theory, the membership of the elements in a set is assessed in binary terms according to a bivalent condition—an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described with the aid of the membership function valued in the real unit interval \([0,1]\) expressed as \(\mu_A(x_i) : U \to [0,1]\), where \(U\) is universe of discourse which represents the grade of membership of \(x \in U\) in \(A\) as follows

\[
\mu_A(x_i) = \begin{cases} 
0, & \text{if } x \notin A \text{ and there is no ambiguity} \\
1, & \text{if } x \in A \text{ and there is no ambiguity} \\
0.5, & \text{if maximum ambiguity, i.e. } x \in A \text{ or } x \notin A
\end{cases}
\]
Let 
\[ A = \{ x_i : 0 < \mu_A(x_i) < 1, \forall i = 1, 2, ..., n \} \]

\[ B = \{ x_i : 0 < \mu_B(x_i) < 1, \forall i = 1, 2, ..., n \} \]

And 
\[ U = \{ u_i : u_i > 0, \forall i = 1, 2, ..., n \}. \]

be two fuzzy sets and \( U \), the set of utilities corresponding to fuzzy membership function \( \mu_A(x_i) \) for any event \( E \). Corresponding to the above membership functions, we have the following fuzzy information scheme.

\[
F.S. = \begin{bmatrix}
E_1 & E_2 & \cdots & E_n \\
\mu_A(x_1) & \mu_A(x_2) & \cdots & \mu_A(x_n) \\
\mu_B(x_1) & \mu_B(x_2) & \cdots & \mu_B(x_n) \\
u_1 & u_2 & \cdots & u_n
\end{bmatrix}
\]

### 4.2 Lower Bound on Code Word Length \( t \):

Suppose that a person believe that the degree of membership of ith event is \( \mu_B(x_i) \) and the code with code length \( n_i \) has been constructed accordingly. But contrary to his belief the true degree of membership is \( \mu_A(x_i) \).

We will now obtain a lower bound of mean length \( L(t) \) under the condition;

\[
\sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{-1}(x_i) + (1 - \mu_B(x_i))^{-1} \right) |r_\alpha| n_i D^{-n_i} \leq 1 \quad (4.2.1)
\]

**Remark 4.2.1:** For a noiseless channel \( |r_\alpha| n_i = 1 \forall i = 1, 2, ..., k \). The inequality (4.2.1) reduces to the fuzzy Inequality corresponding to Autar and Soni [8].

\[
\sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{-1}(x_i) + (1 - \mu_B(x_i))^{-1} \right) D^{-n_i} \leq 1 \quad (4.2.2)
\]

**Remark 4.2.2:** Moreover, if \( \mu_A(x_i) + (1 - \mu_A(x_i)) = \mu_B(x_i) + (1 - \mu_B(x_i)) \) for each \( i \), (4.2.2) reduces to Kraft [65] inequality;

\[
\sum_{i=1}^{k} D^{-n_i} \leq 1 \quad (4.2.3)
\]
Theorem 4.2.1: Let a source S have k messages symbols \(S_1, S_2, \ldots, S_k\) with message degree of membership \(\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_k); \mu_A(x_i) \geq 0\). Let an \(\alpha\)-prompt code encode these messages into a code alphabet of \(D\) symbols and let the length of the code word corresponding to the messages \(S_i\) be \(n_i\). Then the code length of order \(t\), \(L(t)\), shall satisfy the inequality:

\[
L(t) \geq \frac{1}{1-\beta} \log_D \sum_{i=1}^{k} \left( \mu_A^\beta(x_i) + (1 - \mu_A(x_i))^{\frac{\beta}{\tau}} \right) (|r_{\alpha}|n_i)^{1-\beta} \tag{4.2.4}
\]

Proof: In the Holder’s inequality

\[
\left[ \sum_{i=1}^{k} x_i^p \right]^{1/p} \left[ \sum_{i=1}^{k} y_i^q \right]^{1/q} \leq \sum_{i=1}^{k} x_i y_i \tag{4.2.5}
\]

With the equality if and only if \(x_i = cy_i\), where \(c\) is a positive number,

\[
1/p + 1/q = 1 \text{ and } p < 1.
\]

We note the direction of Holder’s inequality is the reverse of the usual one as \(p < 1\) (Backenbach and Bellman [21]).

Substituting

\[
p = -\tau, q = 1 - \beta, \quad x = \left( \frac{1}{\mu_A^\tau(x_i)} + (1 - \mu_A(x_i))^{\frac{1}{\tau}} \right) D^{-n_i}
\]

and

\[
y_i = \left( \frac{1}{\mu_A^\tau(x_i)} + (1 - \mu_A(x_i))^{\frac{1}{\tau}} \right) |r_{\alpha}|n_i
\]

we get;

\[
\left\{ \sum_{i=1}^{k} \left[ \frac{1}{\mu_A^\tau(x_i)} + (1 - \mu_A(x_i))^{\frac{1}{\tau}} \right] D^{-n_i} \right\} \left\{ \sum_{i=1}^{k} \left( \frac{1}{\mu_A^\tau(x_i)} + (1 - \mu_A(x_i))^{\frac{1}{\tau}} \right) |r_{\alpha}|n_i \right\}^{1-\beta} \leq \sum_{i=1}^{k} D^{-n_i} |r_{\alpha}|n_i
\]

or

\[
\left\{ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{n_i} \right\}^{-1/\tau} \left\{ \sum_{i=1}^{k} \left( \mu_A^{\frac{1-\beta}{\tau}}(x_i) + (1 - \mu_A(x_i))^{\frac{1-\beta}{\tau}} \right) |r_{\alpha}|n_i \right\}^{1-\beta} \leq \sum_{i=1}^{k} D^{-n_i} |r_{\alpha}|n_i
\]
\[ \leq \sum_{i=1}^{k} D^{-n_i} |r_{\alpha}|_{n_i} \]

Moreover, \( \frac{1}{p} + \frac{1}{q} = 1 \Rightarrow \beta = (1 + t)^{-1} \), with this substitution the above inequality reduces to

\[ \left\{ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{tn_i} \right\}^{-1/t} \left\{ \sum_{i=1}^{k} \left( \mu_A^\beta(x_i) + (1 - \mu_A(x_i))^\beta \right) \left[ |r_{\alpha}|_{n_i} \right]^{1-\beta} \right\}^{1/(1-\beta)} \leq \sum_{i=1}^{k} D^{-n_i} |r_{\alpha}|_{n_i} \]

Using inequality of Bernard and Sharma [23], viz.

\[ \sum_{i=1}^{k} D^{-n_i} |r_{\alpha}|_{n_i} \leq 1 \]

Which gives;

\[ \left\{ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{tn_i} \right\}^{1/t} \left\{ \sum_{i=1}^{k} \left( \mu_A^\beta(x_i) + (1 - \mu_A(x_i))^\beta \right) \left[ |r_{\alpha}|_{n_i} \right]^{1-\beta} \right\}^{1/(1-\beta)} \]

or

\[ \frac{1}{t} \log_D \left\{ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{tn_i} \right\} \geq \frac{1}{1-\beta} \log_D \left\{ \sum_{i=1}^{k} \left( \mu_A^\beta(x_i) + (1 - \mu_A(x_i))^\beta \right) \left[ |r_{\alpha}|_{n_i} \right]^{1-\beta} \right\} \]

Hence

\[ L(t) \geq \frac{1}{1-\beta} \log_D \left\{ \sum_{i=1}^{k} \left( \mu_A^\beta(x_i) + (1 - \mu_A(x_i))^\beta \right) \left[ |r_{\alpha}|_{n_i} \right]^{1-\beta} \right\} \quad (4.2.6) \]

The quantity

\[ \frac{1}{1-\beta} \log_D \left\{ \sum_{i=1}^{k} \left( \mu_A^\beta(x_i) + (1 - \mu_A(x_i))^\beta \right) \left[ |r_{\alpha}|_{n_i} \right]^{1-\beta} \right\} \]

is similar to fuzzy entropy corresponding to Renyi’s entropy of order \( \beta \) [84].

It can be easily verified that the quantity in (4.2.4) hold if and only if;
\[ n_i = -\beta \log_D \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) + \log_D \left\{ \sum_{i=1}^{k} \left( \mu_A^\beta(x_i) + (1 - \mu_A(x_i))^\beta \right) \left| r_{\alpha} \right|_{n_i}^{1-\beta} \right\} \]

**Particular Cases:**

**a)** For \( t = 0 \) and \( \beta = 1 \), the inequality (4.2.4) reduces to the fuzzy inequality corresponding to the Bernard and Sharma [24].

\[ \bar{n} \geq \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \log_D \left[ \frac{\left| r_{\alpha} \right|_{n_i}}{\sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right]^{1-\beta} \]

**b)** For noiseless channel, \( \left| r_{\alpha} \right|_{n_i} \forall i \), the inequality (4.2.4) reduces to the fuzzy inequality corresponding to the Campbell [28].

\[ L(t) \geq H_\beta(A), \]

where \( H_\beta(A) \) is the fuzzy entropy corresponding to the Renyi’s entropy of order \( \beta \).

**c)** If the channel is noiseless and \( t = 0 \), \( \beta = 1 \), then the inequality reduces the fuzzy entropy corresponding to the well known Shannon’s [87] inequality \( \bar{n} \geq H(A) \), where \( H(A) \) is the fuzzy entropy corresponding to the Shannon’s entropy.

**Theorem 4.2.2:** Let an \( \alpha \)-prompt code encode the \( K \) messages \( S_1, S_2, \ldots, S_k \) into a code alphabet of \( D \) symbols and let the length of the corresponding encoded messages \( S_i \) be \( n_i \). Then the code length of order \( t \), \( L(t) \) shall satisfy the inequality.

\[ L(t) \geq \frac{1}{1-\beta} \log_D \left\{ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{\beta-1}(x_i) + (1 - \mu_B(x_i))^{\beta-1} \right) \left| r_{\alpha} \right|_{n_i}^{1-\beta} \right\} \quad (4.2.7) \]

With equality if and only if;

\[ n_i = -\log \left( \left| r_{\alpha} \right|_{n_i} \right) - \beta \left( \mu_B^\beta(x_i) + (1 - \mu_B(x_i))^\beta \right) \]

\[ + \log_D \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{\beta-1}(x_i) + (1 - \mu_B(x_i))^{\beta-1} \right) \left| r_{\alpha} \right|_{n_i}^{1-\beta} \]

where \( L(t) = \frac{1}{t} \log_D \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{n_i} \).
Proof: In the Holder’s inequality

\[
\left[ \sum_{i=1}^{k} x_i^p \right]^{1/p} \left[ \sum_{i=1}^{k} y_i^q \right]^{1/q} \leq \sum_{i=1}^{k} x_i y_i
\]

With the equality if and only if:

\[
x_i^p = cy_i^q, \text{ where } c \text{ is a positive number, } \frac{1}{p} + \frac{1}{q} = 1 \text{ and } p < 1. \text{ We note that direction of Holder’s inequality is the reverse of the usual one as } p < 1 \text{ (Beckenbach and Bellman [21]).}
\]

Substituting,

\[
p = -t, q = t\beta, \quad x_i = \left( \frac{1}{\mu_A(x_i)} \right)^{\frac{-1}{t}} (1 - \mu_A(x_i))^{\frac{-1}{t}} D^{-ni}
\]

and

\[
y_i = \left( \frac{1}{\mu_B(x_i)} \right)^{\frac{1}{t\beta}} (1 - \mu_B(x_i))^{\frac{1}{t\beta}} (1 - \mu_B(x_i))^{-1} |r_\alpha|_{n_i}
\]

We get;

\[
\left( \sum_{i=1}^{k} (\mu_A(x_i) + (1 - \mu_A(x_i)))^{\frac{-1}{t}} \right) \left( \sum_{i=1}^{k} \left( \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{-\frac{1}{t}}(x_i) + (1 - \mu_B(x_i))^{-\frac{1}{t\beta}} \right) \left( |r_\alpha|_{n_i} \right)^{\frac{1}{t\beta}} \right) \right)^{\frac{1}{1/t\beta}}
\]

\[
\leq \sum_{i=1}^{n} (\mu_A(x_i) + (1 - \mu_A(x_i))) \left( \mu_B^{-1}(x_i) + (1 - \mu_B(x_i))^{-1} \right) |r_\alpha|_{n_i} D^{-ni}
\]

Moreover, \(\frac{1}{p} + \frac{1}{q} = 1\), \(\Rightarrow \beta = (1 + t)^{-1}\), with this substitution the above inequality reduces to;

\[
\left( \sum_{i=1}^{k} (\mu_A(x_i) + (1 - \mu_A(x_i)))^{\frac{-1}{t}} \right) \left( \sum_{i=1}^{k} \left( \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{-\frac{1}{t}}(x_i) + (1 - \mu_B(x_i))^{-\frac{1}{t\beta}} \right) \left( |r_\alpha|_{n_i} \right)^{\frac{1}{t\beta}} \right) \right)^{\frac{1}{1/t\beta}}
\]

\[
\leq \sum_{i=1}^{n} (\mu_A(x_i) + (1 - \mu_A(x_i))) \left( \mu_B^{-1}(x_i) + (1 - \mu_B(x_i))^{-1} \right) |r_\alpha|_{n_i} D^{-ni}
\]
this gives \[
\left( \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{tn_i} \right)^{\frac{1}{t}} \geq \]
\[
\frac{1}{1 - \beta} log_D \left( \sum_{i=1}^{k} \left\{ \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{-1}(x_i) + (1 - \mu_B(x_i))^\beta \right) \right\} \right)
\]
or
\[
\frac{1}{t} log_D \left( \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{tn_i} \right)
\]
\[
\geq \frac{1}{1 - \beta} log_D \left( \sum_{i=1}^{k} \left\{ \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{-1}(x_i) + (1 - \mu_B(x_i))^\beta \right) \right\} \right)
\]
Hence,
\[
L(t) \geq \frac{1}{1 - \beta} log_D \left( \sum_{i=1}^{k} \left\{ \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{-1}(x_i) + (1 - \mu_B(x_i))^\beta \right) \right\} \right)
\]
The quantity;
\[
\frac{1}{1 - \beta} log_D \left( \sum_{i=1}^{k} \left\{ \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \mu_B^{-1}(x_i) + (1 - \mu_B(x_i))^\beta \right) \right\} \right)
\]
is equivalent to fuzzy inaccuracy corresponding to Nath’s inaccuracy [77] of order \(\beta\).

**Particular Cases:**

For \(t = 0\) and \(\beta \rightarrow 1\), the inequality (4.2.7) reduces to;
\[
\bar{\eta} \geq \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \log_D \left( \frac{|r_{\alpha}|_{n_i}}{\mu_B(x_i) + (1 - \mu_B(x_i))} \right) (4.2.8)
\]
For noiseless channel, \(|r_{\alpha}|_{n_i} = 1\); \(\forall i\), the inequality (4.2.8) reduces to;
\[
\bar{\eta} \geq \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \log_D \left( \mu_B(x_i) + (1 - \mu_B(x_i)) \right)
\]
\[ H(\mu_A(x_i), (x_i)) \] (4.2.9).

Where \( H(\mu_A(x_i), (x_i)) \) is a fuzzy measure of inaccuracy corresponding to Kerridge [62] measure of inaccuracy.

a) When \( \mu_A(x_i) = \mu_B(x_i) \), then the R.H.S. of (4.2.9) reduces to the fuzzy inequality corresponding to the Shannon [87] measure of inaccuracy.

For noiseless channel \( (|r|_{n_i}) = 1; \forall i \), the inequality (4.2.7) reduces to fuzzy inequality corresponding to Autar and Soni [8].

\[ L(t) \geq H_\beta(\mu_A(x_i), \mu_B(x_i)) \] (4.2.10)

b) Where \( H_\beta(\mu_A(x_i), \mu_B(x_i)) \) is fuzzy measure of inaccuracy corresponding to Nath [77] of order \( \beta \).

### 4.3 \( \beta \)-measure of Uncertainty Involving Utilities:-

Consider a fuzzy function corresponding to Gill et.al [42] as;

\[
H_\beta^k (A,U) = \frac{\sum_{i=1}^{k} \left(\mu_A(x_i) + \left(1 - \mu_A(x_i)\right)\right) \frac{u_i}{\sum_{i=1}^{k} u_i \left(\mu_A(x_i) + \left(1 - \mu_A(x_i)\right)\right)_i}^{1-\beta} - 1}{1 - 2^{1-\beta}},
\]

\( \beta > 0(\neq 1) \) (4.3.1)

Which is \( \beta \)-measure of uncertainty involving utilities.

**Remark:** When the utility aspect of the scheme is considered (i.e. \( u_i = 1, i = 1, 2, 3, \ldots, k \) as well as \( \beta \rightarrow 1 \), the measure (4.3.1) becomes fuzzy information measure corresponding to Shannon’s [87] measure of information.

Further, define a parametric mean length credited with utilities and membership function \( \mu_A(x_i) \) as;

\[
L(U^\beta) = \left[ \frac{\sum_{i=1}^{k} u_i \left(\mu_A(x_i) + \left(1 - \mu_A(x_i)\right)\right) D^{(\beta^{-1}-1)}_{n_i}}{1 - 2^{1-\beta}} \right]^{\beta} - 1
\]

(4.3.2)

Where \( \beta > 0 (\neq 1) \), \( \mu_A(x_i) \geq 0, i = 1, 2, \ldots, k \) and \( \sum_{i=1}^{k} \mu_A(x_i) = 1 \) which is a generalization fuzzy mean length corresponding to Campbell [28], and for \( \beta \rightarrow 1 \), it reduces to fuzzy mean
The code word length corresponding to Shannon [87] measure and gave a characterization of $HU_K^\beta (A; U)$ under the condition.

$$\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{-n_i} \leq u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)$$  (4.3.3)

**Theorem 4.3.1:** Suppose $n_1, n_2, ..., n_k$ are the lengths of uniquely decodable code words satisfying (4.3.3), then the average code length satisfies:

$$L(U^\beta) \geq H_k^\beta (A, U)$$  (4.3.4)

With the equality in (4.3.4) if and only if;

$$n_i = \beta \log_D \left[ \frac{u_i}{\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right] + \log_D \left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \frac{u_i}{\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right)^{1-\beta} \right]$$  (4.3.5)

**Proof:** In the Holder’s inequality (Beckenback et.al [21]).

$$\left[ \sum_{i=1}^{k} x_i^p \right]^{1/p} \left[ \sum_{i=1}^{k} y_i^q \right]^{1/q} \leq \sum_{i=1}^{k} x_i y_i$$  (4.3.6)

For all $x_i > 0, y_i > 0, i = 1, 2, ..., k$ and $p < 1$, where $\frac{1}{p} + \frac{1}{q} = 1$ with the equality in (4.3.6) if and only if there exists a positive number $c$ such that;

$$x_i^p = cy_i^q$$  (4.3.7)

We substitute

$$x_i = \left( \mu_B^{-1}(x_i) + (1 - \mu_B(x_i)) \right)^{\beta - 1} D^{-n_i};$$

$$y_i = \left( \mu_B^{(1-\beta)^{-1}}(x_i) + (1 - \mu_B(x_i))^{(1-\beta)^{-1}} D^{-n_i} \right) \left( \frac{u_i}{\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right); \forall i$$

$$p = (1 - \beta^{-1}) and q = 1 - \beta, we get;$$
Using the inequality (4.3.3), the above inequality can be written as;

\[
\left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{(\beta^{-1}-1)n_i} \right]^{\beta} - 1
\]

\[
\geq \left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \frac{u_i}{\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right)^{1-\beta} \right] - 1
\]

\[
\leq \frac{\sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D^{(\beta^{-1}-1)n_i}}{1 - 2^{1-\beta}} - 1
\]

Hence,

\[
L(U^\beta) \geq H_k^\beta(A, U).
\]

**Theorem 4.3.2.** Let \( n_1, n_2, ..., n_k \) are the lengths of uniquely decodable code words, then the average code length \( L(U^\beta) \) can be made to satisfy the inequality;

\[
H_k^\beta(A, U) \leq L(U^\beta) \leq D \cdot H_k^\beta(A, U) + \frac{D-1}{1-2^{1-\beta}} \quad (4.3.8)
\]

**Proof:** Suppose

\[
n_i = \beta \log_D \left[ \frac{u_i}{\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right] + \]

77
\[
\log_D \left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \frac{u_i}{\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right)^{1-\beta} \right] \tag{4.3.9}
\]

Clearly, \( \bar{n}_i \) and \( \bar{n}_{i+1} \) satisfy the inequality in Holder’s inequality. Moreover \( \bar{n}_i \) satisfy the inequality (4.3.3).

Let \( n_i \) be the (unique) integer between \( \bar{n}_i \) and \( \bar{n}_{i+1} \). Since \( \beta > 0 \) (\( \neq 1 \)), we have;

\[
\left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D(\beta^{-1-1}n_i) \right]^{\beta} \leq \left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D(\beta^{-1-1}n_i) \right]^{\beta} 
\]

\[
< D \left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D(\beta^{-1-1}n_i) \right]^{\beta} \tag{4.3.10}
\]

We know

\[
\sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \frac{u_i}{\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right)^{1-\beta} \]

\[
= \left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D(\beta^{-1-1}n_i) \right]^{\beta}
\]

Hence, (4.3.10) can be expressed as;

\[
\sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \frac{u_i}{\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right)^{1-\beta} 
\]

\[
\leq \left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) D(\beta^{-1-1}n_i) \right]^{\beta} 
\]

\[
< D \left[ \sum_{i=1}^{k} \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right) \left( \frac{u_i}{\sum_{i=1}^{k} u_i \left( \mu_A(x_i) + (1 - \mu_A(x_i)) \right)} \right)^{1-\beta} \right] 
\]

Thus,

\[
H_k^\beta(A, U) \leq L(U^\beta) \leq D. H_k^\beta(A, U) + \frac{D-1}{1-2^{-\beta}}
\]
4.4 **Fuzzy Directed Divergence Measures and their Bounds:**

Classical information theoretic divergence measures have witnessed the need to study them. Kullback-Leibler [66] first studied the measure of divergence. Jaynes [54] introduced the Principle of Maximum Entropy (PME). He emphasized that “choose a distribution which is consistent to with the information available and is uniform as possible”. For implementation of this consideration another advance was needed in the form of a measure of nearness of two probability distribution and it was already provided by Kullback-Leibler in the form of:

\[
I(P; Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \quad (4.4.1)
\]

If the distribution \( Q \) is uniform. This becomes;

\[
I(P; Q) = \sum_{i=1}^{n} p_i \log p_i + \log n \quad (4.4.2)
\]

Where, \( P, Q \in T_n \) and

\[
T_n = \left\{ P = (p_1, p_2, ..., p_n), p_1 > 0, \sum_{i=1}^{n} p_i = 1 \right\}; \forall i = 1, 2, ..., n, \quad n \geq 2.
\]

Since Shannon’s Entropy

\[
H(P) = \sum_{i=1}^{n} p_i \log p_i \quad (4.4.3)
\]

was already available in the literature, so maximizing \( H \) is equivalent to minimizing \( I(P; Q) \). This is one of the interpretations of PME.

Analyzing (4.4.1) in the following way:

\[
I(P; Q) = \sum_{i=1}^{n} (p_i \log p_i - p_i \log q_i) (4.4.4)
\]

The second term present in (4.4.4) is called the Kerridge Inaccuracy which is;

\[
= -\sum_{i=1}^{n} p_i \log q_i \quad (4.4.5)
\]

Considering Kerridge [62] inaccuracy, we can interpret Kullback-Leibler [66] measure of divergence.
\[ I(P; Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \text{ as i.e.} \]

= difference of Kerridge inaccuracy and Shannon’s entropy

\[ = \sum_{i=1}^{n} \{-p_i \log q_i - (-p_i \log p_i)\} \quad (4.4.6) \]

Since \( I(P; Q) \) provides a measure of nearness of \( P \) from \( Q \). Take the case of Reliability Theory, here we can consider how much the information is reliable. Because the distribution is the revised distribution / strategies to achieve the goal/ objective / target with certain constraints, so optimization theory takes the birth, which is the need of every one.

Hence, whenever we come across divergence measures, we are interested to minimize the divergence to make the information available, reliable. Every walk of life is governed with the reliability of information under certain constraints.

Analogous to information theoretic approach, when we arrive at fuzzy sets or fuzziness, we need to study fuzzy divergence measures. As presently, the vast applications of fuzzy information in life and social sciences, Interpretational communication, Engineering, Fuzzy Aircraft Control, Medicine, Management and Decision making, Computer Sciences, Pattern Recognition and Clustering. Hence the wide applications motivates us to consider Divergence Measures for fuzzy set theory to minimize or maximize or optimize the fuzziness.

Let \( A = \{x_i : \mu_A(x_i), \forall i = 1, 2, ..., n\} \) and \( B = \{x_i : \mu_B(x_i), \forall i = 1, 2, ..., n\} \), where \( 0 < \mu_A(x_i) < 1 \) and \( 0 < \mu_B(x_i) < 1 \), be two fuzzy sets. The fuzzy divergence corresponding to Kullback-Leibler [66] has been defined by Bhandari and Pal [26] as:

\[ D(A \parallel B) = \sum_{i=1}^{n} \left[ \mu_A(x_i) \log \frac{\mu_A(x_i)}{\mu_B(x_i)} + \{1 - \mu_A(x_i)\} \log \frac{1 - \mu_A(x_i)}{1 - \mu_B(x_i)} \right] \quad (4.4.7) \]

The fundamental properties of fuzzy divergence are as follows:

1. Non-negativity, i.e. \( D(A \parallel B) \geq 0 \).
2. \( D(A \parallel B) = 0 \) if \( A = B \).
3. \( D(A \parallel B) \) is a convex function in \( (0, 1) \).
4. \( D(A \parallel B) \) should not change, when \( \mu_A(x_i) \) is changed to \( 1 - \mu_A(x_i) \) and \( \mu_B(x_i) \) to \( 1 - \mu_B(x_i) \).
Bhandari and Pal [26] has established some properties such as:

(a) \( D(A \parallel B) = I(A \parallel B) + I(B \parallel A) \), where \( I(A \parallel B) = [\mu_A(x_i) - \mu_B(x_i)] \log \frac{\mu_A(x_i)}{\mu_B(x_i)} \).

(b) \( D(A \cup B \parallel A \cap B) = D(A \parallel B) \).

(c) \( D(A \cup B \parallel C) \leq D(A \parallel C) + D(B \parallel C) \).

(d) \( D(A \parallel B) \geq D(A \cup B \parallel A) \).

(e) \( D(A \parallel B) \) is maximum if \( B \) is the farthest non-fuzzy set of \( A \).

Havrda-Charvat [50] has given the measure of directed divergence as;

\[
D_\alpha = (P; Q) = \frac{1}{\alpha(\alpha - 1)} \left( \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha} - 1 \right) \tag{4.4.8}
\]

Corresponding to (4.4.8), the average code word length can be taken as

\[
L_\alpha = (P; Q) = \frac{1}{\alpha(\alpha - 1)} \left( \sum_{i=1}^{n} p_i q_i D^{(\alpha-1)n_i} - 1 \right) \tag{4.4.9}
\]

Corresponding to (4.4.8), the fuzzy measure of directed divergence between two fuzzy sets \( \mu_A(x_i) \) and \( \mu_B(x_i) \) can taken as;

\[
D_\alpha = (\mu_A(x_i); \mu_B(x_i))
\]

\[
= \frac{1}{\alpha(\alpha - 1)} \left( \sum_{i=1}^{n} \left\{ (\mu_A(x_i))^\alpha + (1 - \mu_A(x_i))^\alpha (\mu_B(x_i))^{1-\alpha} + (1 - \mu_B(x_i))^{1-\alpha} \right\} - 1 \right)
\]

and its corresponding fuzzy average code word length as;

\[
L_\alpha = (\mu_A(x_i); \mu_B(x_i))
\]

\[
= \frac{1}{\alpha(\alpha - 1)} \left( \sum_{i=1}^{n} \left\{ (\mu_A(x_i))^\alpha + (1 - \mu_A(x_i))^\alpha (\mu_B(x_i))^\alpha + (1 - \mu_B(x_i))^\alpha \right\} D^{(\alpha-1)n_i} - 1 \right)
\]

Remark:

1. As \( \alpha \to 1 \), (4.4.8) tends to (4.4.1).

2. As \( \alpha \to 1 \) and \( q_i = 1 \), (4.4.8) tends to (4.4.3).
3. As \( \alpha \rightarrow 1 \) (4.4.9) tends to average codeword length given as;

\[
L = \sum_{i=1}^{n} p_i q_i n_i
\]  

(4.4.10)

4. As \( \alpha \rightarrow 1 \) and \( q_i = 1 \), (4.4.9) tends to average codeword length corresponding to Shannon’s entropy given as;

\[
L = \sum_{i=1}^{n} p_i n_i
\]  

(4.4.11)

4.5 Noiseless directed divergence Coding Theorems:-

**Theorem 4.5.1:** For all uniquely decipherable codes

\[
D_\alpha \leq L_\alpha
\]  

(4.5.1)

where

\[
L_\alpha = (\mu_A(x_i); \mu_B(x_i))
\]

\[
= \frac{1}{(\alpha - 1)} \left( \sum_{i=1}^{n} \left\{ (\mu_A(x_i)^{\alpha} + (1 - \mu_A(x_i))^{\alpha})^{\mu_B(x_i)} (1 + \mu_B(x_i))^{\mu_B(x_i)} \right\} D^{(\alpha-1)n_i - 1} \right)
\]

**Proof:** By Holder’s inequality, we have;

\[
\sum_{i=1}^{n} x_i y_i \geq \left( \sum_{i=1}^{n} x_i^p \right)^{\frac{1}{p}} \left( \sum_{i=1}^{n} y_i^q \right)^{\frac{1}{q}} ; \quad 0 < p < 1, q < 0 \text{ or } 0 < q < 1, p < 0
\]  

(4.5.2)

Set

\[
x_i = \left[ f(\mu_A(x_i), \mu_B(x_i)) \right]^{\frac{1}{t}} D^{-n_i}
\]

\[
y_i = \left[ f(\mu_A(x_i), \mu_B(x_i)) \right]^{\frac{1}{t}} \text{ and } p = -t \Rightarrow 0 < p < 1, \quad q = \frac{t}{t+1} \Rightarrow q < 0
\]

Thus equation (4.5.2) becomes;

\[
\sum_{i=1}^{n} \left[ f(\mu_A(x_i), \mu_B(x_i)) \right]^{\frac{-t}{t+1}} D^{-n_i} \left[ f(\mu_A(x_i), \mu_B(x_i)) \right]^{\frac{1}{t}} D^{-n_i} \left[ f(\mu_A(x_i), \mu_B(x_i)) \right]^{\frac{1}{t}}
\]

\[
\geq \left[ \sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\}^{\frac{-t}{t+1}} D^{-n_i} \right]^{\frac{-t}{t+1}} \left[ \sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\}^{\frac{1}{t}} \right]^{\frac{t+1}{t}}
\]
Using Kraft’s inequality, we have

\[
\left[ \sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\}^\frac{1}{t} \right]^\frac{t+1}{t} \leq \left[ \sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\}^{-1} D^{-n_i} \right]^{-\frac{1}{t}}
\]

or,

\[
\sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\}^\frac{1}{t} \leq \sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\} D^{n_i t}^{-\frac{1}{t}}
\]

or,

\[
\sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\} \leq \sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\} D^{n_i t} \quad (4.5.3)
\]

dividing both sides by \( t \), we get;

\[
\frac{\sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\}}{t} \leq \frac{\sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\} D^{n_i t}}{t}
\]

Subtracting \( n \) from both sides, we have

\[
\frac{\sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\}}{t} - 1 \leq \frac{\sum_{i=1}^{n} \left\{ f(\mu_A(x_i), \mu_B(x_i)) \right\} D^{n_i t} - 1}{t} \quad (4.5.4)
\]

Taking \( \alpha = t + 1, \; t = \alpha - 1 \)

and

\[
f(\mu_A(x_i), \mu_B(x_i)) = \left\{ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right\},
\]
equation (4.5.4) becomes;

\[
\frac{\sum_{i=1}^{n} \left\{ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right\} - 1}{\alpha - 1} \leq \frac{\sum_{i=1}^{n} \left\{ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right\} D^{n_i (\alpha - 1)} - 1}{\alpha - 1}
\]

(4.5.5)

Dividing both sides by \( \alpha \), we get;
\[
\sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^{\alpha} \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^{\alpha} \right) \right] - 1 
\]
\[\alpha(\alpha - 1)\]

\[
\leq \sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^{\alpha} \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^{\alpha} \right) \right] D_{n_i}^{\alpha - 1} - 1 
\]
\[\alpha(\alpha - 1)\]

that is \( D_\alpha \leq L_\alpha \) which proves the theorem.

**Theorem 4.5.2**: For all uniquely decipherable codes,

\[ D_{\alpha,\beta} \leq L_{\alpha,\beta} \quad (4.5.6) \]

\[
L_{\alpha,\beta} = \frac{1}{\beta - \alpha} \sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^{\alpha} \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^{\alpha} \right) \right] D_{n_i}^{\alpha - 1} - 1 \quad (4.5.7)
\]

Where either \( \alpha \geq 1, \beta \leq 1 \) or \( \alpha \leq 1, \beta \geq 1 \)

**Proof**: Since from (4.5.5), we have;

\[
\sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^{\alpha} \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^{\alpha} \right) \right] - 1 \leq \sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^{\alpha} \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^{\alpha} \right) \right] D_{n_i}^{\alpha - 1} - 1 \quad (4.5.8)
\]

Multiplying both sides by \( (\alpha - 1) \), we get;

\[
\sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^{\alpha} \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^{\alpha} \right) \right] - 1 \leq \sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^{\alpha} \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^{\alpha} \right) \right] D_{n_i}^{\alpha - 1} - 1 \quad (4.5.9)
\]

Changing \( \alpha \) to \( \beta \), (4.5.9) becomes;

\[
\sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\beta + (1 - \mu_A(x_i))^{\beta} \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^{\alpha} \right) \right] - 1 \leq \sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\beta + (1 - \mu_A(x_i))^{\beta} \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^{\alpha} \right) \right] D_{n_i}^{\beta - 1} - 1 \quad (4.5.10)
\]

Subtracting (4.5.10) from (4.5.9), and dividing both sides by \( (\beta - \alpha) \), we have;
\[
\sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right] \\
- \left[ \left( \mu_A(x_i)^\beta + (1 - \mu_A(x_i))^\beta \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\beta \right) \right] \leq \\
\sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right] D \left( \alpha - 1 \right) - \\
\sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\beta + (1 - \mu_A(x_i))^\beta \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\beta \right) \right] D \left( \beta - 1 \right)
\]

That is \( D_{\alpha, \beta} \leq L_{\alpha, \beta} \), which proves the theorem.

**Theorem 4.5.3:**

For all uniquely decipherable codes,

\[
D'_{\alpha, \beta} \leq L'_{\alpha, \beta} \quad (4.5.11)
\]

where

\[
D'_{\alpha, \beta} = \frac{1}{\beta - \alpha} \log_D \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right] (4.5.12)
\]

and

\[
L'_{\alpha, \beta} = \frac{1}{\beta - \alpha} \log_D \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right] D \left( \alpha - 1 \right) (4.5.13)
\]

To prove this theorem, we first prove the following lemma:

**Lemma 4.5.1:** For all uniquely decipherable codes

\[
\log_D \left[ \sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right] \right] \leq \\
\log_D \left[ \sum_{i=1}^{n} \left[ \left( \mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha \right) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right] D \left( \alpha - 1 \right) \right]
\]

**Proof of the Lemma:** From (4.5.3) we have;

\[
\sum_{i=1}^{n} \left[ \left( f(\mu_A(x_i), \mu_B(x_i)) \right) \right] \leq \sum_{i=1}^{n} \left[ \left( f(\mu_A(x_i), \mu_B(x_i)) \right) D_{\alpha, \beta} \right]
\]

Taking logarithm on both sides, we have;

\[
\log_D \left[ \sum_{i=1}^{n} \left[ \left( f(\mu_A(x_i), \mu_B(x_i)) \right) \right] \right] \leq \log_D \left[ \sum_{i=1}^{n} \left[ \left( f(\mu_A(x_i), \mu_B(x_i)) \right) D_{\alpha, \beta} \right] \right]
\]

Taking \( \alpha = t + 1, \ t = \alpha - 1 \)
and
\[ f(\mu_A(x_i), \mu_B(x_i)) = \left( (\mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha) (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right), \]
we have
\[ \log_D \left[ \sum_{i=1}^{n} \left( (\mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha) (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right) \right] \leq \log_D \left[ \sum_{i=1}^{n} \left( (\mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha) (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right) D^{n_i(\alpha-1)} \right] \]

Which proves the Lemma.

**Proof of the Theorem 4.5.3:** Changing \( \alpha \) to \( \beta \) in (4.5.14), we get
\[
\log_D \left[ \sum_{i=1}^{n} \left( (\mu_A(x_i)^\beta + (1 - \mu_A(x_i))^\beta) \left( (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right) \right) \right] \leq \log_D \left[ \sum_{i=1}^{n} \left( (\mu_A(x_i)^\beta + (1 - \mu_A(x_i))^\beta) \left( (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right) \right) D^{n_i(\beta-1)} \right]
\]

subtracting (4.5.15) from (4.5.14), we have;
\[
\log_D \left[ \frac{\sum_{i=1}^{n} \left( (\mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha) (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right)}{\sum_{i=1}^{n} \left( (\mu_A(x_i)^\beta + (1 - \mu_A(x_i))^\beta) (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right)} \right] \leq \log_D \left[ \frac{\sum_{i=1}^{n} \left( (\mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha) (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right) \right] D^{n_i(\alpha-1)} \right] \leq \log_D \left[ \sum_{i=1}^{n} \left( (\mu_A(x_i)^\beta + (1 - \mu_A(x_i))^\beta) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right) \right] D^{n_i(\beta-1)} \right]
\]

Dividing both sides by \( \beta - \alpha \), we have;
\[
\frac{1}{\beta - \alpha} \log_D \left[ \frac{\sum_{i=1}^{n} \left( (\mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha) (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right)}{\sum_{i=1}^{n} \left( (\mu_A(x_i)^\beta + (1 - \mu_A(x_i))^\beta) (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right)} \right] \leq \frac{1}{\beta - \alpha} \log_D \left[ \frac{\sum_{i=1}^{n} \left( (\mu_A(x_i)^\alpha + (1 - \mu_A(x_i))^\alpha) (\mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha) \right) \right] D^{n_i(\alpha-1)} \right] \leq \log_D \left[ \sum_{i=1}^{n} \left( (\mu_A(x_i)^\beta + (1 - \mu_A(x_i))^\beta) \left( \mu_B(x_i)^\alpha + (1 - \mu_B(x_i))^\alpha \right) \right) \right] D^{n_i(\beta-1)} \right]
\]

that is
\[ D'_{\alpha, \beta} \leq L'_{\alpha, \beta}. \] Which proves the theorem.