SOME GENERALIZATION OF FUZZY ENTROPY MEASURE AND ITS APPLICATIONS

Thesis Submitted to University of Kashmir for the Award of the Degree of

Doctor of Philosophy
in
Statistics

By

MOHD JAVID DAR

Under the Supervision of

Dr. M.A.K. Baig
(H.O.D)

Post Graduate Department of Statistics
FACULTY OF PHYSICAL AND MATERIAL SCIENCES
University of Kashmir
Srinagar – 190006, J&K, India
(NAAC Accredited Grade 'A')

(2014)

Post Graduate Department of Statistics

University of Kashmir, Srinagar

Certificate

This is to certify that the work embodied in this thesis is entitled "Some Generalization of Fuzzy Entropy Measure and its Applications" is the original work carried out by the scholar Mr. Mohd Javid Dar under my supervision and the work is suitable for the award of the degree of Doctor of Philosophyin Statistics.

The thesis has reached the standard fulfilling the requirements of regulations relating to the degree. The results contained have not been submitted earlier to this or any other University or Institute for the award of degree or diploma.

(Dr. M.A.K Baig)

Supervisor
Associate Professor,
Department of Statistics
University of Kashmir,
Srinagar

Dedicated to My Parents

(Father: Haji Abdul Gani Dar)
(Mother: MehraHabib)

&

My Affectionate Brother

(Mr.Altaf Ahmad)

Who has supported me all the way since the beginning of my studies.

Acknowledgement

"Praise to ALLAH, Creator of the Universe, the Most Gracious, Most Merciful, Who Bestowed Upon Me the Courage to Accomplish This Work"

Seek knowledge from cradle to the grave.

(Prophet Mohammad (PBUH))

Whoever seek knowledge and find it, will get two rewards; one of them the reward for desiring it, and the other for attaining it; therefore even if he do not attain it, for him is one reward.

(Prophet Mohammad (PBUH))

One learned man is harder on the devil than a thousand ignorant worshipers.

(Prophet Mohammad (PBUH))

This thesis tried to combine and unify different types of fuzzy entropy measures in the field of information theory and statistics. I am deeply indebted to the altruistic influence created by Head Department of Statistics, my supervisor Dr. Mirza Abdul Khalique Baig, a selfless guidance, stimulating suggestions, and encouragement and of course for his tolerance and continuous interest in the field of information theory in the context of fuzzy set measures and its applications. His genuine concern and affectionate attitude always kept me lively to accomplish this work.

I am very grateful to Prof. Aquil Ahmed, former Head Department of Statistics, who supported me in every respect during various stages of accomplishment of this thesis.

I am also very thankful to other faculty members of the Department of Statistics Dr. Tariq Rashid Jan and Dr. Parveiz Ahmed. Sincere thanks to Dr. Mohammad Abdullah Mir, Associate Prof. Department of Mathematics, University of Kashmir and Dr. Mukhtar Ahmad Khandy, Associate Prof. Department of Mathematics, University of Kashmir.

The thesis would not have been possible without the confidence, endurance and support of my family especially my beloved parents Haji Abdul Gani and Mehra Habib, my affectionate brother Mr. Altaf Ahmad, my sympathetic sisters Nayeema and Rubia, my sister-in-law Arifa, my brother –in – law's Mr. Bashir Ahmad and Mr. Nazir Ahmad, my nieces Anjum

Bashir, Uneeza Bashir, Arfeen Bashir, Nuha Jaan, Rabia Altaf and my nephew Numaan Nazir.

I owe lot to my friends and colleagues who helped, inspired and encouraged me especially Mr. Mohammad Akbar (Ph.D scholar Pharmaceutical Science), Miss Nusrat Mushtaq (Ph.D scholar Statistics), Miss Humaira Sultan (Ph.D scholar Statistics), Mr. Kaiser Ahmad (Ph.D scholar Statistics) and my cousin Mr. Waheed – ul – Rasheed.

At last I extend my sincere thanks to all my friends and colleagues with whose prayers have driven me out of depression which I have often faced in the struggle for this academic program.

Mohammad Javid Dar

ΧΟΝΤΕΝΤΣ

PREFACE

This thesis represents only a small section of the different issue and topics that I was involved since 2010. Yet, it shows one of the consequences of my engagements during the past few years. Over these years, I had the historical opportunity to read and witness the rise and the fall of the important theories and results that are cited in this study and my involvements in these topics became a part of my whole academic life.

Fuzzy SetTheory has come a long way since it was formally introduced by L.A. Zadeh in his classic paper entitled 'Fuzzy Sets' published in the journal 'information and Control' in the year 1965. Since that time the subject has been applied to every branch of knowledge. Many research investigations by mathematicians, scientists and social scientists, computer and management scientists and engineers all over the world have been made in the theory and applications of the subject. Applications of fuzzy logic and fuzzy set theory in decision-making, Pattern recognition, Image processing, Control systems, Neural networks, Genetic algorithm and in many other areas has given significant results.

Much work has been done on this branch of Information theory and statistics. It has acquired a great currency in various research journals of statistics and mathematics. In this light, I compiled my thesis on the topic "Some Generalizations of Fuzzy Entropy Measure and its Applications" and chapter wise scheme is as fallows.

Chapter one: Chapter one is devoted to surveying the relevant literature which is required in the development of existing results, the basic concepts and preliminary results. This task is to present a bird's eye view of the following chapters. No effort is made to define the technical vocabulary. Such an undertaking requires a detailed logical presentation. This introductory chapter discusses generalities leaving a more detailed and precise treatment to subsequent chapters.

Chapter Two: Deals with generalizations of fuzzy measures of information and theircode word lengths. We propose some generalized fuzzy average codeword length and establish relationship with generalized fuzzy entropy. Some fuzzy coding theorems have also been developed. The coding theorems obtained in this chapter not only produce new results but also generalizes some well established results in the literature of information theory.

Chapter Three: In this chapter, information scheme and bounds for generalized measure of cast have been considered and their bounds have been obtainedfor suitable generalized mean code word lengths. Also, several coding theorems on fuzzy entropy functions depending upon parameter R and V have been derived.

Chapter Four: This chapter deals with lower bounds on the mean length of code words using the concept of segment decomposition, and the effective range has been established. The bounds obtained provide a measure of optimality for variable length error correcting codes. Also fuzzy directed divergence measures and their corresponding bounds in the form of theorems have been presented, and their particular cases have been studied.

Chapter Five: In this chapter we propose a new characterization result on a life time distribution in terms of fuzzy measure of uncertainty. Based on our proposed measure,a new class of fuzzy life distributions are defined which mimic the increasing failure rate and decreasing failure rate. It is shown that these new classes are different from the known classes of life distributions.

A comprehensive bibliography is given at the end

The intent of this manuscript is to present the work related to generalizations of fuzzy entropy measures of information and their applications. It will be a useful document for the future researchers in this area. The area of generalizations of fuzzy entropy measures of information and its applications is fertile and there is a lot of scope to work on this concept.

The subject matter of the present thesis is fully published in the form of the following research papers written by the author:

[1] "Some Generalizations of Fuzzy Average Codeword Length and Inequalities", International Journal of Statistics and Analysis, ISSN: 2248-9959, Volume 3, No. 4, pp. 393 – 400, (2013).

[2] "Some New Results on Fuzzy Directed Divergence Measures and Their Inequalities", Asian Journal of Mathematics and Statistics, Volume 7, Issue 1, pp. 12 – 20, (2014).

[3] “Some New Generalizations of Fuzzy Average Codeword Length and Their Bounds”, American Journal of Applied Mathematics and Statistics, Volume 2, No. 2, pp. 73 – 76.

[4] “Fuzzy Coding Theorem on Generalized Fuzzy Cost Measure”, Asian Journal of Applied Mathematics (ISSN: 2321–564X), Volume 02 **–** Issue 01, (Feb. 2014)

[5] “Some Coding Theorems on Fuzzy Entropy Function Depending upon Parameter R and V”, IOSR Journal of Mathematics, e-ISSN: 2278-3008, p-ISSN: 2319-7676, Volume 9, Issue 6, pp. 119-123, (jan. 2014)

1.1 Introduction:-

The purpose of this chapter is to clarify the basic concepts of Fuzzy sets theory, Information theory, and “fuzzy entropy”.Uncertainty and fuzziness are the basic nature of human thinking and of many real world objectives. Fuzziness is found in our decision, in our language and in the way we process information. The main use of information is to remove uncertainty and fuzziness. In fact, we measure information supplied by the amount of probabilistic uncertainty removed in an experiment and the measure of uncertainty removed is also called as a measure of information while measure of fuzziness is the measure of vagueness and ambiguity of uncertainties.

By the nineteen sixties it became evident in mathematical systems research that the rigorous treatment based on Aristotelian logic is not appropriate in analyzing real systems.Fuzzy sets were defined by Zadeh [109] to free the mathematical model from the law of the excluded middle. Formally, the characteristic function $\mu_A(x)$ describing the membership of element x in the set A was generalized: in classical mathematics the characteristic function takes either the value 0 or 1; in the case of fuzzy sets the characteristic function may take any value from the real interval [0,1].

The concept of fuzziness has been applied to apparently all phenomena already formalized in systems research: Statistics, Information theory, Clustering and Decision analysis, Medical and Socio-economic predictions, Image processing, etc. This overwhelming success, seen in introspect, is not surprising for the following simple reasons. The mathematical ideas applied in systems research had well known for the workers involved and their applications had prevailed the planning and analysis of information processing systems. Thus both their theoretical clarity and practical relevance had been firmly established. The fundamental but essentially mathematical generalization of these formal ideas posed many challenging questions within the conceptually and methodologically well-known (mathematical) framework.

Within the extremely large field of theories and applications developed from the concept of fuzziness, there has been a relatively small area of dealing with the fuzziness of concepts. The most important questions in this area are: How should we calculate a numerical description of particularlyfuzzy quantifiers like “very”, “more or less”. “rather,” for such categories as “short,” “old,” “many,” and for statements connecting such ideas, like “much older than”?. How should we apply different operations defined on fuzzy sets to formal logic and to conceptual categories?

The concept of fuzziness was made a scientific one in mathematical systems theory by Zadeh’s [109] definition of fuzzy sets. This advantages a “framework which provides a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership”. The introduction of fuzzy sets was motivated by the fuzziness of concepts, i.e., that “More often than not, the classes of objects encountered in the real physical world do not have precisely defined criteria of membership”.

1.1.1 Fuzzy Models:

Fuzzy sets are a generalization of conventional set theory that was introduced by Zadeh [109] as a mathematical way to represent vagueness in everyday life. The basic idea of fuzzy sets is easy to grasp. Suppose, as you approach a red light, you must advise a driving student when to apply the brakes. Would you say, "Begin braking 74 feet from the crosswalk"? Or would your advice be more like, "Apply the brakes pretty soon"? The latter, of course; the former instruction is too precise to be implemented. This illustrates that precision may be quite useless, while vague directions can be interpreted and acted upon. Everyday language is one example of ways vagueness is used and propagated. Children quickly learn how to interpret and implement fuzzy instructions (“go to bed about 10”). We all assimilate and use (act on) fuzzy data, vague rules, and imprecise information, just as we are able to make decisions about situations which seem to be governed by an element of chance. Accordingly, computational models of real systems should also be able to recognize, represent, manipulate, interpret, and use (act on) both fuzzy and statistical uncertainties.

The Process and progress of knowledge unfolds into two stages: an attempt to know the character of the world and a subsequent attempt to know the character of the knowledge itself. The second reflective stage arises from the failures of the first; it generates an enquiry into the possibility of knowledge and into the limits of that possibility. It is in this second stage of enquiry that we find ourselves today. As a result, our concerns with knowledge, perception of problems and attempts at solutions are of a different order than in the past. We want to know not only specific facts or truths but what we cannot know, what we do and do not know, and how we know at all. Our problems have shifted from questions of how to cope with the world (how toprovide ourselves with food, shelter and so on), to questions of how to cope with knowledge (and ignorance) itself. Ours has been called an "information society," and a major portion of our economy is devoted to the handling, processing, selecting, storing, disseminating, protecting, collecting, analyzing and sorting of information, our best tool for this being, of course, the computer.

Our problems are seen in terms of decision, management and prediction; solutions are seen in terms of faster access to more information and of increased aid in analyzing, understanding and utilizing the information that is available and in coping with the information that is not. These two elements, large amounts of information coupled with the large amounts of uncertainty, taken together constitute the ground of many of our problems today: complexity. As we become aware of how much we know and how much we do not know, as information and uncertainty themselves become the focus of our concern, we begin to see our problems as centering on the issue of complexity.

How do we manage to cope with complexity as well as we do, and how could we manage to cope better? The answer seems to lie in the notion of simplifying complexity by making a satisfactory trade-off or compromise between the information available to us and the amount of uncertainty we allow. One option is to increase the amount of allowable uncertainty by sacrificing some of the precise information in favor of a vague but more robust summary. For instance, instead of describing the weather today in terms of the exact percentage of cloud cover (which would be much too complex), we could just say that it is sunny, which is more uncertain and less precise but more useful. In fact, it is important to realize that the imprecision or vagueness that is characteristic of natural language does not necessarily imply a loss of accuracy or meaningfulness. It is, for instance, generally more meaningful to give travel directions in terms of city blocks than in terms of inches, although the former is much less precise than the latter. It is also more accurate to say that it is usually warm in the summer than to say that it is usually 72 degree in the summer. In order for a term such as sunny to accomplish the desired introduction of vagueness, however, we cannot use it to mean precisely 0% cloud cover. Its meaning is not totally arbitrary, however a cloud cover of 100% is not sunny and neither, in fact, is a cloud cover of 80%. We can accept certain intermediate states, such as 10 or 20 %cloud cover, as sunny. But where we draw the line? If, for instance, any cloud cover of 25% or less is considered sunny, does this mean that a cloud cover of 26% is not? This is clearly unacceptable since 1% of cloud cover hardly seems like a distinguishing character between sunny and not sunny. We could therefore add a qualification that any amount of cloud cover1% greater than a cloud cover already considered to be sunny will also be labeled as sunny. We can see, however that this definition eventually leads us to accept all degrees of cloud cover as sunny , no matter how gloomy the weather looks! In order to resolve this paradox, the term sunny may introduce vagueness by allowing some sort of gradual transition from degrees of cloud cover that are considered to be sunny and those that are not. This is, in fact precisely the basic concept of fuzzy set, a concept that is both simple and intuitively pleasing and that forms, in essence, a generalization of the classical or crisp set.

Fuzzy interpretations of data structures are a very natural and intuitively plausible way to formulate and solve various problems. Conventional (crisp) sets contain objects that satisfy precise properties required for membership. The set of numbers H from 6 to 8 is crisp; we write $H = \{r \in \mathcal{R}/6 \leq r \leq 8\}$. Equivalently, H is described by its membership (or characteristic, or indicator) function MF, $\mu_H : \Re \to \{0, 1\}$ defined as

$$\mu_H(r) = \begin{Bmatrix} 1 & 6 \leq r \leq 8 \\ 0 & \text{otherwise} \end{Bmatrix}$$

The crisp set H and the graph of $\mu_H$are shown in the left half of Fig. 1.1.1(a). Every real number (r) either is in H or is not. Since $\mu_H$maps all real numbers $r \in \Re$ onto the two points$(0, 1)$, crisp sets correspond to two-valued logic: is or isn't, on or off, black or white, 1 or 0. In logic, values of $\mu_H$are called truth values with reference to the question, "Is r in H?" The answer is yes if and only if $\mu_H(r) = 1$; otherwise, no.

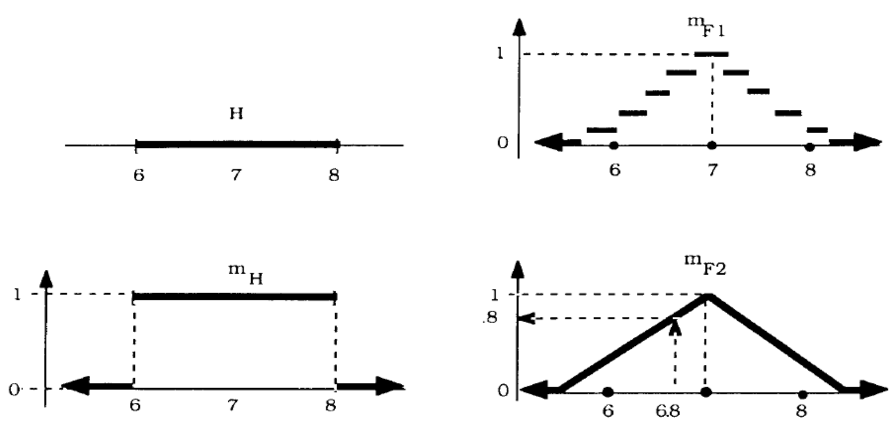

Fig. 1.1.1 (a): Membership functions for hard and fuzzy subsets of $\Re$.

Consider next the set *F*of real numbers that are close to 7. Since the property "close to 7" is fuzzy, there is not aunique membership function for*F*. Rather, the modeler must decide, based on the potential application and propertiesdesired for F, what $\mu_F$should be. Properties that might seem plausible for this F include (i) normality (MF (7) = 1), (ii) monotonicity (the closer r is to 7, the closer $\mu_F(r)$ is to 1, and conversely) and (iii) symmetry (numbers equally far left and right of 7 should have equal memberships). Given these intuitive constraints, either of the functions shown in the right half of Fig. 1.1.1 (a) might be a useful representation of F. $\mu_{F1}$is discrete (the staircase graph), while $\mu_{F2}$is continuous but not smooth (the triangle graph). One can easily construct a MF for F so that every number has some positive membership in F, but we would not expect numbers "far from 7," 20 000 987 for example, to have much! One of the biggest differences between crisp and fuzzy sets is that the former always have unique MFs, whereas every fuzzy set has an infinite number of MFs that may represent it. This is at once both a weakness and strength; uniqueness is sacrificed, but this gives a concomitant gain in terms of flexibility, enabling fuzzy models to be "adjusted" for maximum utility in a given situation.

1.1.2 Fuzzy Sets Theory:

Let X be a space of objects and x be a generic element of X. A classical set A, $A \subseteq X$, is defined as a collection of elements or objects $x \in X$, such that each element (x) can either belong or not to the set A. By defining a characteristic (or membership) function for each element x in X, we can represent a classical set A by a set of ordered pairs $(x, 0)$or $(x, 1)$, which indicates $x \notin A$ or $x \in A$, respectively. Unlike the aforementioned conventional set, a fuzzy set expresses the degree to which an element belongs to a set. Hence the membership function of a fuzzy set is allowed to have values between 0 and 1, which denote the degree of membership of an element in the given set.

1.1.3 Fuzzy sets and membership functions:

If X is a collection of objects denoted generically by x, then a fuzzy setAin X is defined as a set of ordered pairs $A = \{(x, \mu_A(x)/x \in X)\}$, where,$\mu_A(x)$ is called the membershipfunction(or MF for short) for the fuzzy set A. The MF maps each element of X to a membership grade (or membership value) between 0 and 1 (included). Obviously, the definition of a fuzzy set is a simple extension of the definition of a classical (crisp) set in which the characteristic function is permitted to have any values between 0 and 1. If the value of the membership function is restricted to either 0 or 1, then A is reduced to a classical set. For clarity, we shall also refer to classical sets as ordinary sets, crisp sets, non-fuzzy sets, or just sets. Usually X is referred to as the universe of discourse,or simply the universe, and it may consist of discrete (ordered or non-ordered) objects or it can be a continuous space. This can be clarified by the following examples.

1.1.4 Fuzzy sets with a discrete non-ordered universe:

Let X = {San Francisco, Boston, Los Angeles} be the set of cities one may choose to live in. The fuzzy set A = "desirable city to live in" may be described as follows: A = {(San Francisco, 0.9), (Boston, 0.8), (Los Angeles, 0.6)}. Apparently the universe of discourse X is discrete and it contains non-ordered objects - in this case, three big cities in the United States. As one can see, the foregoing membership grades listed above are quite subjective; anyone can come up with three different but legitimate values to reflect his or her preference.

1.1.5 Fuzzy sets with a discrete ordered universe:

Let X = {0, 1, 2, 3, 4, 5, 6} be the set of numbers of children a family may choose to have. Then the fuzzy set B = "desirable number of children in a family" may be described as follows: B = {(0, 0.1), (1, 0.3), (2, 0.7), (3, 1), (4, 0.7), (5, 0.3), (6, 0.1)}. Here we have a discrete ordered universe X; the MF for the fuzzy set B is shown in Fig. 1.1.5(a).

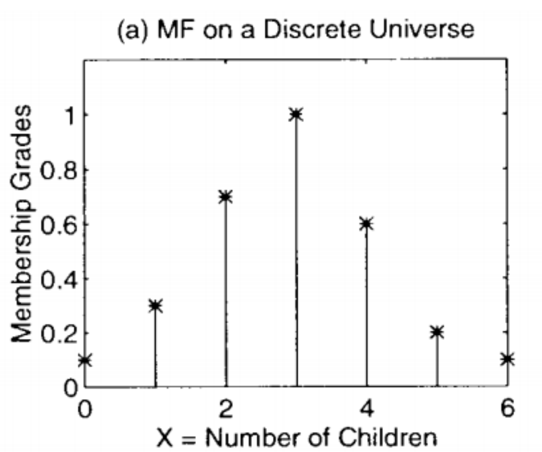

Fig: 1.1.5 (a)

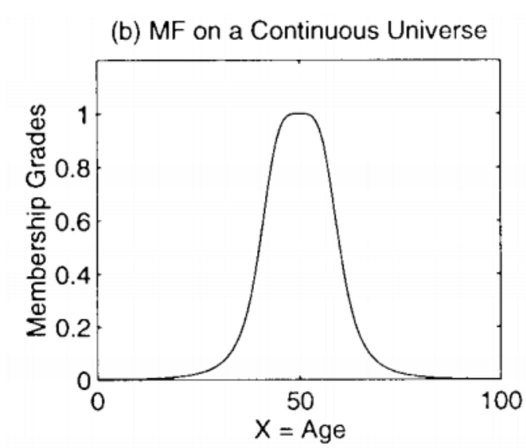

Fig: 1.1.5 (b)

Again, the membership grades of this fuzzy set are obviously subjective measures.

1.1.6 Fuzzy sets with a continuous universe:

Let $X = \mathcal{R}^{+}$ be the set of possible ages for human beings. Then the fuzzy set C = "about 50 years old" may be expressed as $C = \{(x, \mu_c(x)/x \in X)\}$, where

$$\mu_c(x) = 1 \Big/ \left(1 + \left(\frac{x-50}{10}\right)^4\right)$$

This is illustrated in Figure 1.1.5(b). From the preceding examples, it is obvious that the construction of a fuzzy set depends on two things: the identification of a suitable universe of discourse and the specification of an appropriate membership function. The specification of membership functions is subjective, which means that the membership functions specified for the same concept by different persons may vary considerably. This subjectivity comes from individual differences in perceiving or expressing abstract concepts and has little to do with randomness. Therefore, the subjectivityand non randomness of fuzzy sets is the, primary difference between the study of fuzzy sets and probability theory, which deals with objective treatment of random phenomena.

In practice, when the universe of discourse X is a continuous space, we usually partition it into several fuzzy sets whose MFs cover X in a more or less uniform manner. These fuzzy sets, which usually carry names that conform to adjectives appearing in our daily linguistic usage, such as "large," "medium," or "small," are called linguistic values or linguistic labels. Thus, the universe of discourse X is often called the linguistic variable.

1.2 Some nomenclature used in the literature:-

1.2.1 Support of a Fuzzy Set:

The support of a fuzzy set A is the set of all points x in X such that $\mu_A(x) > 0$.

1.2.2 Core of a Fuzzy Set:

The core of a fuzzy set A is the set of all points x in X such that $\mu_A(x) = 1$.

1.2.3 Normality of a Fuzzy Set:

A fuzzy set A is normal if its core is nonempty. In other words, we can always find at least a point $x \in X$ such that $\mu_A(x) = 1$.

1.2.4 Crossover Points:

A crossover point of a fuzzy set A is a point $x \in X$ at which $\mu_A(x) = 0.5$.

1.2.5 Fuzzy Singleton:

A fuzzy set whose support is a single point in x with $\mu_A(x) = 1$ is called a fuzzy singleton.

1.2.6 α– Cut, Strong α– Cut:

The α-cut orα-level set of a fuzzy set A is a crisp set defined by $A_\alpha=\{x/\mu_A(x) \geq \alpha\}$. Strong α-cut or strong α-level set are defined similarly $A'_\alpha=\{x/\mu_A(x) > \alpha\}$. Using this notation, we can express the support and core of a fuzzy set A as support (A) = A'_0 and core $A=A_1$.

1.2.7 Convexity:

A fuzzy set A is convex if and only if for any $x_1, x_2 \in X$and any $\lambda \in[0,1]$,

$$\mu_A(\lambda x_1 +(1-\lambda) x_2) \geq \min \{\mu_A(x_1), \mu_A(x_2)\}.$$

Alternatively, A is convex if all its α-level sets are convex. It is to be noted that the definition of convexity of a fuzzy set is not as strict as the common definition of convexity of a function.

1.2.8 Linguistic variables and linguistic values:

Suppose that X = "age."Then we can define fuzzy sets "young,""middle aged" and "old" that arecharacterized by MFs. Just as a variable canassume various values, a linguistic variable"age" can assume different linguistic values,such as "young," "middle aged "and" old"in this case. If "age" assumes the value of"young," then we have the expression "ageis young," and so forth for the other values.

1.2.9 Fuzzy numbers:

A fuzzy number A is a fuzzy set in the real line that satisfies the conditions for normality and convexity. Most fuzzy sets used in the literature satisfy the conditions for normality and convexity, so fuzzy numbers are the most basic type of fuzzy sets.

Union, intersection, and complement are the most basic operations on classical sets. On the basis of these three operations, a number of identities can be established. Corresponding to the ordinary set operations of union, intersection and complement, fuzzy sets have similar operations, which were initially defined in Zadeh's seminal paper [109]. Before introducing these three fuzzy set operations, first we shall define the notion of containment, which plays a central role in both ordinary and fuzzy sets. This definition of containment is, of course, a natural extension of the case for ordinary sets.

1.2.10 Containment or Subset:

Fuzzy set A is contained in fuzzy set B (or, equivalently, A is a subset of B, or A is smaller than or equal to B, $A \subseteq B$) if and only $\mu_A(x) \le \mu_B(x)$ for all x.

1.2.11 Union (disjunction):

The unionof two fuzzy sets A and B is a fuzzy set C, written as

$$C = A \cup B \text{ or } C = A \text{ OR } B,$$

whose MF is related to those of A and B by

$$\mu_c(x) = \max\big(\mu_A(x),\ \mu_B(x)\big).$$

1.2.12 Intersection (conjunction):

The intersectionof two fuzzy sets A and B is a fuzzy set C, written as $C = A \cap B$ or $C = A \; and \; B$, whose MF is related to those of A and B by

$$\mu_c(x) = \min\big(\mu_A(x),\ \mu_B(x)\big).$$

1.2.13 Complement (negation):

The complement of fuzzy set A, denoted by $\bar{A}$`or NOT A, is defined as

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x).$$

Since, the operations introduced above perform exactly as the corresponding operations for ordinary sets if the values of the membership functions are restricted to either 0 or 1. However, it is understood that these functions are not the only possible generalizations of the crisp set operations. For each of the aforementioned three set operations, several different classes of functions with desirable properties have been proposed subsequently in the literature (e.g. algebraic sum for union and product for intersection). In general, union and intersection of two fuzzy sets can be defined through T-conorm (or S-norm) and T-norm operators respectively. These two operators are functions $S, T: [0,1] \times \big[0, 1 \rightarrow [0,1]\big]$ satisfying some convenient boundary, monotonicity, commutativity and associativity properties. As pointed out by Zadeh

[109], a more intuitive but equivalent definition of union is the, “smallest” fuzzy set containing both A and B. Alternatively, if D is any fuzzy set that contains both A and B, then it also contains $A \cup B$. Analogously, the intersection of A and B is the "largest" fuzzy set which is contained in both A and B. The two fundamental (Aristotelian) laws of crisp set theory are:

(a)Law of Contradiction:

$A \cup \overline{A} = X$(i.e., a set and its complement must comprise the universe of discourse), and

(b)Law of Excluded Middle:

$A \cap \overline{A} = \emptyset$(i.e., an object can either be in its set or its complement; it cannot simultaneously be in both). It can be easily seen that for every fuzzy set that is non-crisp (i.e., whose membership function does not only assume values 0 and 1) both laws are broken (i.e., for fuzzy sets $A \cup \overline{A} \neq X$ and $A \cap \overline{A} \neq \emptyset$. Indeed $\forall\, x \in A$such that

$$\mu_A(x) = \alpha, 0 < \alpha < 1: \mu_{A \cup \bar{A}}(x) = \max\{\alpha, 1 - \alpha\} \neq 1$$

and $$\mu_{A \cap \bar{A}}(x) = \min\{\alpha, 1 - \alpha\} \neq 0.$$

1.3 Interpreting the Membership Function:-

The first point to note is that, like $P(A)$, the probability of a set A, fuzzy set theory does not tell us how to specify $\mu_A(x)$, the membership function of a fuzzy set A. The second point to note is that whereas there is a logical requirement that $P(A) \in [0, 1]$, the fact that $\mu_A(x) \in [0, 1]$is simply a convenience of scaling. The third point to note is that whereas $P(A)$can be interpreted as a two-sided bet (which in principle can be settled when A reveals itself),$\mu_A(x)$reflects an individual’s view of the extent to which $x \in A$; thus $\mu_A(x)$cannot be made operational in the same sense as $P(A)$. Finally, it is not a requirement that $\sum_x \mu_A(x)$be 1, and thus $\mu_A(x)$as a function of x cannot be interpreted as a probability or, for that matter, as a conditional probability, as was done by Loginov [73] and also by Barrett and Woodall [19]. How then can we interpret the membership function$\mu_A(x)$? Because$\mu_A(x)$, as a function of x, reflects the extent to which$x \in A$[i.e., $\mu_A(x)$is an indicator of how likely it is that$x \in A$], we may view$\mu_A(x)$as the likelihood of x for a fixed (i.e., specified)A. Even though the interpretation of a likelihood is almost always derived from a probability model, the likelihood is not a probability (in particular, it does not obey the addition rule) and in statistical inference, the likelihood function reflects the relative degrees of support that a fixed observation provides to several hypotheses. Furthermore, the specification of likelihood is subjective. Thus our interpretation of the membership function is that it is a likelihood function with $\tilde{A}$taking the role of a fixed observation and the values of x taking the role of the hypotheses.

To statisticians specializing in inference, our interpretation of the membership function as a likelihood will appear to be unconventional. This is because in the context of inference, the likelihood entails a fixed observation and a varying parameter. However, our structure for the likelihood is a consequence of the notion of the likelihood from a more philosophical viewpoint, and what we have proposed is in keeping with the foundational notion of likelihood. Basu[20]. The foregoing points are best illustrated via the following example involving two fuzzy sets Aand B, where

$$\mathrm{A} = \{x; x \in X and\ x\ is\ "medium"\}$$

And $$\mathrm{B} = \{x; x \in X\ and\ is\ "small"\}$$

as before, $X = \{1, 2, 3, \dots, 10\}$.

Suppose that an assessor assigns the membership functions $\mu_A(x)$ and $\mu_B(x)$ given in table below.

Table1.3.1 Membership Functions of A and B

X	$\mu_A(x)$	$\mu_B(x)$	$\mu_A(x) + \mu_B(x)$
0	0	1	1
1	0	1	1
2	.2	.8	1
3	.5	.5	1
4	.8	.3	1.1
5	1	.1	1.1
6	.8	0	.8
7	.5	0	.5
8	.2	0	.2
9	0	0	0
10	0	0	0
Col. Sums	**4**	**3.7**	**7.7**

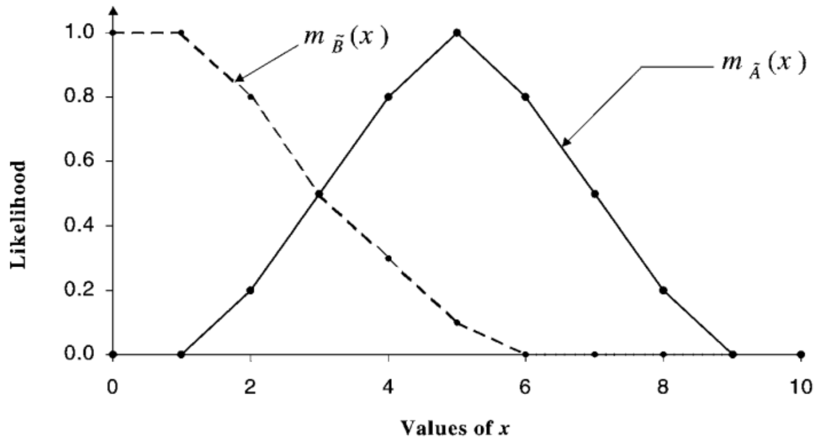

Fig: 1.3.1(a). Membership functions of Ã and B̃

Clearly,$\sum_x \mu_A(x)$and $\sum_x \mu_B(x)$are not 1, nor is it true that $\mu_A(x) + \mu_B(x)$is necessarily 1. A plot of $\mu_A(x)$and$\mu_A(x)$as a function of x, with$\mu_A(x)$and$\mu_B(x)$viewed as likelihoods is shown in Figure 1.3.1(a). The plots reflect the extent to which an x belongs to the sets $A$and B.

1.4 Probability: A Calculus For The uncertainty Of Outcomes:-

The underlying set-theoretic premise for considering probability and its calculus is an experiment, Ɛ, which is yet to be performed. Let x denote a generic outcome of Ɛ, and let $\mathcal{S}$denote the set of all conceived outcomes of Ɛ; thus x $\in \mathcal{S}$. It is important to note that the probability theory does not tell one how to specify $\mathcal{S}$; this choice is subjective and is up to the user. For convenience, we assume that $\mathcal{S}$is a countable set. Let F denote a set whose members are subsets of $\mathcal{S}$; that is, F is a family of sets. However, F is such that it contains $\mathcal{S}$ and φ, where φ is the null set. Furthermore, F is closed under unions and intersections; that is, if $A, B \in F$, then$(A \cup B)$ and $(A \cap B) \in F$. The subsets of $\mathcal{S}$are called events, and in probability theory it is presumed that the events are well defined or "sharp" (also known as "crisp"); that is, there is no ambiguity in declaring whether any outcome x of $\mathcal{S}$belongs to A or to its complement A^c. In contrast, with fuzzy sets there is ambiguity in classifying an x in a subset A or A^c, because A is not sharply defined. If the outcome ofF, say x, is such that x ∈A, then we say that event A has occurred. Because Ɛis yet to be performed, we are uncertain about the occurrence of any particular x. Consequently, we are also uncertain about the occurrence of event A. We describe this uncertainty by a number, P(A), where $0 \leq P(A) \leq 1$; $P(A)$ is the probability of event A, or the probability measure of the set A. There are several interpretations of $P(A)$; the one that is germane to our interest here is that $P(A)$is a two-sided bet (or wager) on the occurrence of event A. Specifically, P(A) is the amount that one is willing to stake out in exchange for a dollar should event A occur or, equivalently, $(1-P(A))$is the amount staked in exchange for a dollar should event A not occur. Furthermore, the individual specifying $P(A)$is required to be indifferent between betting on A orA^c. The two-sided bet will be settled when Ɛis performed and ω is observed, so that the disposition of A is known. An advantage of the foregoing interpretation of $P(A)$is that probability can be made "operational" via the mechanism of betting. This interpretation of probability is a basis for a personalistic (or a subjectivistic) theory of probability. It is important to note that probability theory does not tell us how to arrive at a particular $P(A)$, nor does it in its purely abstract form even attempt to interpret $P(A)$. Many probabilists would declare that the assignment of initial probabilities is a job for a statistician, though some would say that the role of a statistician is to help clients formulate their prior knowledge, because it is the client who knows.

The calculus (or the algebra) of probability tells one how the various uncertainties (i.e., the initial probabilities) combine or cohere. In particular, if $P(B)$denotes the quantification of uncertainty of another event B, then

a). $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, where

b). $P(A \cap B) = \begin{cases} 0 & If A\cap B=\emptyset \\ P(A/B)P(B) & otherwise \end{cases}$

The quantity $P(A/B)$is called the conditional probability of A were B to occur. Like $P(A)$, $P(A/B)$should lie between0 and 1; it represents the amount that one is willing to stake onthe event A should the event B occur but under the proviso thatall bets on A will be called off should B not occur. It is crucialto bear in mind that $P(A/B)$is a bet in the subjunctive mood;this is because the disposition of B is unknown when $P(A/B)$is specified. Finally, ignoring the relevance of a conditioningevent, events A and B are said to be mutually independent if$(A/B) = P(A)$. The calculus given earlier has an axiomaticfoundation based on behavioristic considerations.

Thus, to summarize, a foundation for the theory of probabilityis based on the following ingredients:

a) A well-defined set $\mathcal{S}$ and subsets of $\mathcal{S}$.

b) An adherence to the "law of the excluded middle," the essential import of which is that any outcome ω of Ɛbelongs to a set A or to a set A^c, but not to both.

c) A calculus based on behaviorist axioms involving numbers between 0 and 1 that can be made operational once Ɛ is performed and its outcome observed.

1.4.1 Probability Measures of Fuzzy Events:

In probability theory [101], an event, A, is a member of σ-field ,α, of subsets of a sample space Ω. A probability measure, P, is a normed measure over a measurable space (Ω, α); that is, P is a real valued function which assigns to every $A \ in\ \alpha$; a probability$P(A)$, such that

a) $P(A) \geq 0\ for\ all\ A \in \alpha$;

b) $P(\Omega) = 1$; and

c) P is count ably additive, i.e., if $\{A_i\}$ is any collection of disjoint events then

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \qquad (1.4.1).$$

The notion of an event and its probability constitute the most basic concepts of probability theory. As defined above, an event is a precisely specified collection of points in the sample space. By contrast, in every day experience one frequently encounters situations in which an "event" is a fuzzy rather than sharply defined collection of points. For example, the ill defined events "it is a warm day," "x is approximately equal to 5," "in twenty tosses of a coin there are several more heads than tails," are fuzzy because of imprecision of the meaning of the underlined words.

By using the concepts of a fuzzy set [109], the notion of an event and its probability can be extended in a natural fashion to fuzzy events of the type explained above. It is possible that such an extension may eventually significantly enlarge the domain of applicability of probability theory, especially in those fields in which fuzziness is a pervasive phenomenon.

Let us assume that for simplicity that Ω is an Euclidean n-space R^n. Thus our probability space will be assumed to be a triplet $(R^n, \alpha, P), where\ \alpha$ is σ-field of Borel sets in R^n and P is a probability measure over R^n. A point in R^n will be denoted by x.

Let $A \in \alpha$, then the probability of A can be expressed as

$$P(A) = \int_A dP \quad (1.4.2)$$

Or equivalently

$$P(A) = \int_{R^n} \mu_A(x) dP = E(\mu_A). \quad (1.4.3)$$

Where μ_A denotes the characteristic function of $A\ (\mu_A(x) = 0\ or\ 1)$. And $E(\mu_A)$ is the expectation of μ_A.

The equation (1.4.3) equates the probability of an event A with the expectation of the characteristic function of A. It is this equation that can readily be generalized to fuzzy events through the use of the concept of fuzzy set.

Definition 1.4.1: Let (R^n, α, P) be a probability space in which α is a σ-field of Borel sets in R^n and P is a probability measure over R^n. Then fuzzy event in R^n is a fuzzy set A in R^n whose membership function, $\mu_A: R^n \to [0, 1]$ is Borel measurable. The probability of a fuzzy event A is defined by the Lebesgue-Stieltjes integral

$$P(A) = \int_{R^n} \mu_A(x) dP = E(\mu_A). \quad (1.4.4)$$

Thus as in (1.4.3), the probability of a fuzzy event is the expectation of its membership function. The existence of the Lebesgue-Stieltjes integral is insured by the assumption that μ_A is Borel measurable.

The above definition of a fuzzy event and its probability form a basis for generalizing within the framework of the theory of fuzzy sets a member of the concepts and results of probability theory, information theory and related fields.

1.5 Information theory:-

It is a branch of probability and statistics with extensive potential applications to communication system. Like several other branches of mathematics, information theory has a physical origin. It was initiated by communication scientists C.E. Shannon [87], who were studying the statistical structure of electrical communication equipments. The subject followed by a flood of research papers speculating upon the possible applications to a broad spectrum of research areas, such as pure mathematics, semantics, physics, management, thermodynamics, botany, econometrics, operations research, psychology, epidemiological studies, disease management and related disciplines.

The Mathematical Theory of Communication is the early work of R. V. L. Hartley on the mathematics of information transmission that is recognized. R. A. Fisher introduced notion i.e. Fisher information in 1925 which is closely related to Claude Shannon's notion of entropy. What follows is not intended as a general introduction to information theory through two outstanding contributions to the mathematical theory of communications in 1948 and 1949. Despite several hasty generalization which produces thousands research papers, one thing became evident; this scientific theory has stimulated the interest of thousands of scientists around the world.

1.5.1 Shannon's Information Theory:

Claude E. Shannon's "A Mathematical Theory of Communication" [87] is considered as the "Magna carta" of the Information Age. Shannon's discovery of the fundamental laws of data comprehension and transmission marks the birth of "Information Theory".

Information theory started out as an engineering project. Shannon's simple goal was to find a way to clear up noisy telephone connections. Today, there would be no internet without Shannon's theory. Every new modem upgrade, every compressed file, which includes any in (.gif) or (.jpeg) format, owes something to information theory of Shannon. Even the everyday compact disc would not be possible without error connection based on information theory. To solve the "noise" problem in communications, Shannon developed a new concept, the "channel" and its associated concepts "the channel capacity" and the "redundancy".

Shannon and Weaver [87] suppose a set of possible events whose probabilities of occurrence are$(p_{1,}p_2, \ldots, p_n)$. These probabilities are known but that is all we know concerning which event will occur. Then it is asked: "Can we find a measure of how much 'choice' is involved in the selection of the event or how uncertain we are of the outcome?" If there is such a measure, say $H(p_{1,}p_2, \ldots, p_n)$it is reasonable to require of it the following properties:

(i) It should be continuous in the probabilities (p_i).

(ii) If all the (p_i) are equal, $p_i = 1/n$then H should be monotonic increasing function of *(n).* With equally likely events, there is more choice, or uncertainty, when there are more possible events.

(iii) If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H.

H function was recognized as "Entropy" as in Boltzmann's famous H theorem in statistical mechanics.

1.5.2 Information Function:

Let E_i be the ith event with probability of occurrence p_i,the information function may be defined as

$$h(p_i) = -\log(p_i) \qquad (1.5.1).$$

1.5.3 Shannon's Entropy

Let X be a discrete random variable taking on a finite number of possible values

X = $(x_1, x_2, \ldots, x_n)$ happening with probabilities $P = (p_1, p_2, \ldots, p_n),\ p_i \geq 0, \sum_{i=1}^{n} p_i = 1,\ i = 1, 2, \ldots, n$

we denote

$$\chi = \begin{bmatrix} x_1, x_2, \ldots., x_n \\ p_1, p_2, \ldots, p_n \end{bmatrix} \qquad (1.5.2).$$

and call the scheme (1.5.2) as the information scheme. Shannon [87] proposed the following measure of information for the information scheme (1.5.2) and calls it entropy.

$$H(P) = H(p_1,\ p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i \qquad (1.5.3).$$

Generally, the base of logarithm is taken '2'and it is assume *0log0 = 0*. When the logarithm is taken as a base '2' the unit of information is called a 'bit.' When the natural logarithm is taken, the resulting unit is called a 'nit'. If the logarithm is taken with base 10, the unit of information is known as 'Hartley'.

The information measure (1.5.3) satisfies the following properties.

(1) Non –negativity:

$$H(p_1, p_2, \ldots, p_n) \geq 0$$

The entropy is always non-negative.

(2) Symmetry:

$$H(p_1, p_2, ..., p_n) = H\ (p_{k(1)}, p_{k(2)}, ..., p_{k(n)})\ \ \forall\ (p_1, p_2, ..., p_n) \in P$$

where$(k_{(1)}, k_{(2)} \dots, k_{(n)})$is an arbitrary permutation on *(1, 2,…, n)* $H(p_1, p_2, ..., p_n)$ is a symmetric function on every p_i, $i=1,2,\dots,n$

(3) Normality:

$H\left(\frac{1}{2}, \frac{1}{2}\right) = 1$The entropy becomes unity for two equally probable events.

(4) Expansibility:

$$H_n\ (p_1, p_2, \dots ,p_n) = H_{n+1}\ (0, p_1, p_2, p_3, \dots,p_n)$$

$$= H_{n+1}\ (p_1, p_2, \dots p_i, 0, p_{i+1}, \dots, \dots p_n)$$

$= \dots\dots\dots$

$= H_{n+1}\ (p_1, p_2, \dots, p_n, 0)$

(5) Recursively:

$$H_n(p_1, p_2, ..., p_n) = H_{n-1}(p_1 + p_2, ..., p_n) + (p_1 + p_2) H_2\left(\frac{P_1}{p_1 + p_2}, \frac{P_1}{p_1 + p_2}\right)$$

where $p_i \geq 0$ with $p_1 + p_2 > 0$

(6) Decisively:

$$H_2\ (1, 0) = H_2\ (0, 1) = 0$$

If one of the events is sure to occur then the entropy is zero in the scheme.

(7) Maximality

$$H\ (p_1,p_2, \dots, p_n) \leq H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) = \log n$$

The entropy is maximum when all the events have equal probabilities.

(8) Additivity:

$$H_{nl}\ (\mathrm{PQ}) = H_{nl}\ (p_1q_1, p_1q_2, \dots, p_1q_l, p_2\,q_2, \dots, p_2\,q_l, \dots, p_nq_1, p_nq_2, \dots, p_nq_l)$$

$$= H_n\ (p_1,p_2, \dots p_n) + H_l\ (q_1,q_2, \dots q_l)$$

For all $(p_1, p_2, \ldots, p_n) \in P$ and for all $(q_1, q_2, \ldots, q_l) \in Q$.

If the two experiments are independent then the entropy contained in the experiment is equal to the entropy in the first experiment plus entropy in the second experiment.

(9) Strong Additivity:

$H_{nl}(PQ) = H_{nl}(p_1q_1, p_1q_2, \ldots, p_1q_l, p_2q_2, \ldots, p_2q_l, \ldots, p_nq_1, p_nq_2, \ldots, p_nq_l)$

$= H_n(p_1, p_2, \ldots, p_n) + \sum_{i=1}^{n} p_i H_i(q_{i1}, q_{i2}, \ldots, q_{in})$ for all $(p_1, p_2, \ldots, p_n) \in P$ and for all $(q_1, q_2, \ldots, q_l) \in Q$ and q_{ij} are conditional probabilities i.e., entropy contained in the two experiments is equal to the entropy in the first plus the conditional entropy in the second experiment given that the first experiment given that the first experiment has occurred.

The Shannon's entropy (1.5.3) was characterized by Shannon assuming a set of postulates. There exists several other characterization of the measure (1.5.3) using different set of postulates.

1.6 Coding theorems:-

The elements of a finite set of n input symbols $X = (x_1, x_2, \ldots, x_n)$ be encoded using alphabet of D symbols. The number of symbols in a codeword is called the length of the codeword. It becomes clear that some restriction must be placed on the assignment of codeword's. One of the restrictions may be that the sequence may be decoded accurately. A code is uniquely decipherable if every finite sequence of code character corresponds to at most one message. In other words, we can say uniquely decipherability is to require that no code be prefix of another codeword. We mean by prefix that a sequence 'A" of code character is prefix of a sequence 'B', if 'B' may be written as 'AC' for some sequence 'C'.

A code having the property that no codeword is prefix of another codeword is said to be instantaneous code. Kraft [65] proved that instantaneous/uniquely decipherable code with lengths $(l_1, l_2, \ldots, l_n)$ is possible iff

$$\sum_{i=1}^{n} D^{-l_i} \leq 1 \quad (1.6.1)$$

where D is the size of the code alphabet. Also if

$$L = \sum_{i=1}^{n} l_i p_i \quad (1.6.2)$$

is the average codeword length where p_i is the probability of the i^{th} input symbol to a noiseless channel then for a code which satisfy (1.6.1), the following inequality holds

$$L \geq \frac{H(P)}{\log D}, \quad (1.6.3)$$

by suitable encoding the message, the average code length can be arbitrarily close to H (P).

Shannon's [87] and Renyi's [84] entropies have been studied by several research workers. The study has been carried out from essentially two different points of view. The first is an axiomatic approach and the second is a pragmatic approach. However, these approaches have little connection with the coding theorem of information theory.

Campbell [28] defined a codeword length of order as

$$L(t) = \frac{1}{t}\log\left(\sum_{i=1}^{n} p_i D^{t l_i}\right), \ -1 < t < \infty, \ t \neq 0, \quad (1.6.4)$$

and developed a noiseless coding theorem for Renyi's [84] entropy of order α which is quite similar to the noiseless coding theorem for Shannon's [87] entropy.

By means of prefix code Gurdial and Pessoa [45], Sharma et al [89], Bernard and Sharma [24], Autar and Soni [8], Autar and khan [9] Beig and Zaheerudin [7], Singh, Kumar and Tuteja[92], etc. have established coding theorems for various information measures.

1.6.1 Theorem: A necessary and sufficient condition for the existence of a instantaneous code $S(x_i)$ such that the length of each word $S(x_i)$ should $l_i, i = 1, 2, \ldots, n$ is that the Kraft inequality [65].

$$\sum_{i=1}^{n} D^{-l_i} \leq 1 \quad (1.6.5)$$

should hold, where D is the number of symbols in the code alphabet.

Proof: Necessary part:

First suppose that there exists a code $S(x_i)$ with the word length $l_i, i = 1, 2, \ldots, n$.

Define $m = max\ \{ l_i, i = 1, 2, \ldots, n \}$ and let $u_j,\ j = 1, 2, \ldots, n$ be the number of codeword's with length *j* (some u_j may be zero).Thus the number of codeword's with only one letter cannot be larger than D

$$u_1 \leq D \tag{1.6.6}$$

The number of codeword's of length 2, can use only of the remaining $(D - u_i)$ symbols in their first place, because of prefix property four of our codes, while any of the D symbols can be used in the second place, thus

$$u_2 \leq (D - u_1) = D^2 - u_1 D \qquad (1.6.7).$$

Similarly,

$$u_3 \leq [(D - u_1)D - u_2]D = D^3 - u_1 D^2 - u_2 D \qquad (1.6.8).$$

Finally, if m is the maximum length of the encoded words, one concludes that

$$u_m \leq \quad D^m - u_1 D^{m-1} - u_2 D^{m-2} - \ldots - u_{m-1} D \qquad (1.6.9).$$

Dividing (1.6.9) by D^{-m}, we get

$$0 \leq 1 - u_1 D^{-1} - u_2 D^{-2} - \ldots - u_{m-1} D^{1-m} - u_m D^{-m} \qquad (1.6.10).$$

Or $$\sum_{i=1}^{n} u_i D^{-i} \leq 1. \qquad (1.6.11).$$

It may not be obvious that this condition is identical with (1.6.5) but note that $m \geq l_i$, $i = 1, 2, \ldots n$ and $\sum_{i=1}^{n} u_i D^{-i} \leq 1$ means the sum of 'the members of all sequences of length *i* multiplied by D^{-i}, where the summation extends from 1 to *m*. The left hand side of (1.6.11) can be written as

$$\sum_{i=1}^{n} u_i D^{-i} = \underbrace{\frac{1}{D} + \ldots + \frac{1}{D}}_{u_1} + \underbrace{\frac{1}{D} + \ldots + \frac{1}{D}}_{u_2} + \ldots + \underbrace{\frac{1}{D^m} + \ldots + \frac{1}{D^m}}_{u_m} \qquad (1.6.12).$$

Each bracketed expression corresponds to message x_i , and thereof the total number of term is *n*.

$$\underbrace{1,...,1}_{u_1},\ \underbrace{2,...,2}_{u_2},\ \underbrace{m,...,m}_{u_m}$$

$$u_1 + u_2 + ...+u_3 = n$$

The term in u_k corresponds to the encoded messages of length K. These terms can be considered as $\sum_{i=1}^{n} D^{-l^i}$ when the summation takes place over all those terms with $l_i = k$ .Therefore, by a simple re-assignment of terms, we may equivalently write

$$\sum_{i=1}^{n} u_i D^{-i} = \sum_{i=1}^{n} D^{-l_i} \qquad (1.6.13)$$

Thus

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

The desired set of positive integers $[l_1, l_2,...,l_n]$must satisfy the inequality (1.6.5).This proves the necessity requirement of the theorem.

Sufficient Part:

Suppose now, that inequality (1.6.5) is satisfied for $[l_1, l_2,...,l_n]$then every summand of the left hand side of (1.6.5) being non negative, the partial sums are also at most 1.

$$u_1 D^{-1} \le 1, \quad \text{or} \quad u_1 \le 1$$

$$u_1 D^{-1} + u_2 D^{-2} \le 1, \quad \text{or} \quad u_2 \le D^2 - u_1 D$$

.

.

.

$$u_1 D^{-1} + u_2 D^{-2} + ... + u_n D^{-n} \le 1,$$

or $$u_n \le D^n - u_1 D^{n-1} - u_2 D^{n-2} - ... - u_{n-1} D$$

but these are exactly the conditions that we have to satisfy in order to guarantee that no encoded message can be obtained from any other by the addition of a sequence of letters of the encoding alphabet, thereof, which implies the existence of the instantaneous code.

Remark: For binary case the Kraft inequality tells us that the length l_i must satisfy the equation

$$\sum_{i=1}^{n} 2^{-l_i} =\le 1 \quad (1.6.14)$$

where the summation is over all the words of the block code.

1.6.1 Lemma: For a probability distribution $P = (p_1, p_2, ..., p_n),\ p_i > 0, \sum_{i=1}^{n} p_i = 1$ and incomplete distribution $Q = (q_1, q_2, ..., q_n),\ q_i > 0 \sum_{i=1}^{n} q_i \le 1$. The following inequality holds

$$-\sum_{i=1}^{n} p_i \log p_i \le -\sum_{i=1}^{n} p_i \log q_i \quad (1.6.15)$$

Before proving Lemma (1.6.1) we state the following lemma.

1.6.2 Lemma: If ψ is differentiable concave function in (a, b), then for all $x_i \in (a,b),\ i = 1, 2, ..., n$ and for all $(q_1, q_2, ..., q_n),\ q_i > 0,\ \sum_{i=1}^{n} q_i \le 1,\ i = 1,2,...,n,$ the inequality

$$\psi\left[\sum_{i=1}^{n} q_i\, x_i\right] \ge \sum_{i=1}^{n} q_i\, \psi(x_i).$$

Define the function

$$L(X) = \begin{cases} -x\log x & for\, x \in (0,\infty) \\ 0 & for\, x = 0 \end{cases}$$

It is differentiable concave function of x on $[0, \infty)$ and continuous at 0 (from right), as

$$\frac{\partial^2}{\partial x^2}(x\log x) > 0,\ \lim_{x\to 0} x\log x = 0\log 0 = 0$$

Putting $x_i = \frac{p_i}{q_i}$, $i = 1,2,\ldots,n$ in lemma 1.6.2, we get

$$\sum_{i=1}^{n} q_i\, L\left(\frac{p_i}{q_i}\right) \le L\left(\sum_{i=1}^{n} q_i \frac{p_i}{q_i}\right)$$

$$L\left(\sum_{i=1}^{n} p_i\right) = \text{L }(1) = 0$$

Thus

$$0 \ge -\sum_{i=1}^{n} q_i \frac{p_i}{q_i} \log\frac{p_i}{q_i}$$

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

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

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

1.6.2 Theorem: Let {X} be a discrete message source, without memory, and x_i be any message of this source with probability of transmission p_i. If the {X} ensemble is encoded in a sequences of uniquely decipherable character taken from the alphabet $(a_1, a_2, \ldots, a_n)$ then

$$\text{L} = \sum_{i=1}^{n} p_i l_i \ge \frac{H(P)}{\log D} \quad ,(1.6.16)$$

Proof: The Condition L $\ge \frac{H(P)}{\log D}$ is equivalent to

$$\text{Log D} \sum_{i=1}^{n} p_i l_i \ge -\sum_{i=1}^{n} p_i \log p_i\,,$$

since $p_i l_i \log D = p_i \log D^{l_i} = -p_i \log D^{-l_i}$, the above condition may be written as

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

we define $q_i = -\dfrac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}$, then q_i's add to unity and Lemma 1.6.1 yields

$$-\sum_{i=1}^{n} p_i \log p_i \leq -\sum_{i=1}^{n} p_i \log\left(\frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}\right) \tag{1.6.17}$$

with equality iff $p_i = \dfrac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}$; $\forall\ i = 1, 2, \ldots, n.$

Hence by (1.6.17).

$$H(P) \leq -\sum_{i=1}^{n} p_i \log D^{-l_i} + \sum_{i=1}^{n} p_i \log\left(\sum_{i=1}^{n} D^{-l_i}\right)$$

$$H(P) \leq L \log D + \log\left(\sum_{i=1}^{n} D^{-l_i}\right)$$

With equality iff $\quad p_i = \dfrac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}\ \ \forall\ i = 1, 2, \ldots, n.$

By theorem 1.6.1, $\sum_{i=1}^{n} D^{-l_i} \leq 1$which gives

$$\log\left(\sum_{i=1}^{n} D^{-l_i}\right) \leq 0.$$

Therefore,

$$H(P) \leq \log D$$

$$L \le \frac{H(p)}{\log D}$$

1.6.3 Theorem: Given a random variable $X = (x_1, x_2, ..., x_n)$ having probability distribution $P = (p_1, p_2, ..., p_n)$, $p_i \ge 0$, with entropy (uncertainty) H(P), there exists a base D, instantaneous code for X, whose average code word length $L = \sum_{i=1}^{n} l_i p_i$ satisfies

$$\frac{H(p)}{\log D} \le L \le \frac{H(p)}{\log D} + 1 \qquad (1.6.18)$$

Proof: In general we cannot hope to construct an absolutely optimal code for a given set of probability $P = (p_1, p_2, ..., p_n)$, since if we choose l_i to satisfy $p_i = D^{-l_i}$ then $l_i = \frac{-\log p_i}{\log D}$ may not be an integer. However we can do the next best thing and select the integer l_i such that

$$\frac{-\log p_i}{\log D} \le l_i = \frac{-\log p_i}{\log D} + 1,\ i = 1, 2, \ldots, n \quad (1.6.19)$$

We claim that and instantaneous code can be constructed with word lengths $l_1, l_2, ... l_n$ To prove this we must show that $\sum_{i=1}^{n} D^{-l_i} \le 1$.

For the left hand inequality of (1.6.19) it follows that

$$\log p_i \ge l_i \log D$$

or

$$p_i \ge D^{-l_i}$$

Thus

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

$$\sum_{i=1}^{n} D^{-l_i} \le 1\ ; \quad \sum_{i=1}^{n} p_i = 1$$

To estimate the average codeword length, we multiply (1.6.19) by p_i and sum over $i = 1, 2, \ldots, n,$ to obtain

$$-\sum_{i=1}^{n} p_i \frac{\log p_i}{\log D} \le \sum_{i=1}^{n} p_i\, l_i \ < \ -\sum_{i=1}^{n} p_i \frac{\log p_i}{\log D} \ + \sum_{i=1}^{n} p_i$$

$$\frac{H(p)}{\log D} \le L \le \frac{H(p)}{\log D} + 1.$$

1.7 Fuzzy Entropy:-

The measure of uncertainty is adopted as a measure of information. Hence, the measure of fuzziness is known as fuzzy information measures. The measure of a quantity of fuzzy information gained from a fuzzy set or fuzzy system is known as fuzzy entropy.

A fuzzy subset 'A' in U (universe of discourse) is characterized by a membership function $\mu_A : U \to [0,1]$ which represents the grade of membership of $x \in U$ in A as follows:

$$\mu_x(A) = \begin{cases} 0, if\ x \notin A\ and\ there\ is\ no\ ambiguity \\ 1, if\ x \in A\ and\ there\ is\ no\ ambiguity \\ 0.5, if\ maximum\ ambiguity, i.e., x \in A\ or\ x \notin A \end{cases}$$

In fact $\mu_A(x)$ associated with each $x \in U$, a grade of membership in the set 'A', when $\mu_A(x)$ is values in {0,1}, it is the characteristic function of a crisp (i.e. non-fuzzy) set.

A fuzzy set A^* is called a sharpened version of A if the following conditions are satisfied:

$$\mu_{A^*}(x_1) \le \mu_A(x_1),\ if\ \mu_A(x_1) \le 0.5; \quad \forall i$$

and

$$\mu_{A^*}(x_i) \ge \mu_A(x_i),\ if\ \mu_A(x_i) \le 0.5; \quad \forall i.$$

Since $\mu_A(x)$ and $1 - \mu_A(x)$ gives the same degree of fuzziness, therefore, corresponding to the entropy due to Shannon [87]. De Luca and Termini [33] suggested the following measure of fuzzy entropy:

$$H(A) = -\sum_{i=1}^{n}\Big[\mu_A(x_i)\log\mu_A(x_i) + \big(1-\mu_A(x_i)\big)\log\big(1-\mu_A(x_i)\big)\Big]. \qquad (1.7.1)$$

De Luca and Termini [33] introduced a set of four properties and these properties are widely accepted as a criterion for defining any new fuzzy entropy. In fuzzy set theory, the entropy is a measure of fuzziness which expresses the amount of average ambiguity/difficulty in making a decision whether an element belongs to a set or not. So, a measure of average fuzziness H(A) in a fuzzy set should have at least the following properties to be valid fuzzy entropy:

i)Sharpness:H(A) is minimum if and only if A is a crisp sets, i.e. $\mu_A(x) = 0$ or 1: $\forall$x.

ii) Maximality: H(A) is maximum if and only if A is most fuzzy set, i.e. $\mu_A(x) = 0.5$; $\forall$x.

iii) Resolution:H(A) ≥ H(A*), where A* is sharpened version of A.

iv) Symmetry:H(A)= $H\left(\bar{A}\right)$, where $\bar{A}$ is the complement of A i.e. $\mu_A(x_i) = 1-\mu_A(x_i)$.

1.7.1 Entropy of a Fuzzy Event:

Let x be a random variable which takes the values $\{x_1, x_2, \dots, x_n\}$ with respective probabilities $p_1, p_2, \dots, p_n$. Then, the entropy of the distribution $P = \{p_1, p_2, \dots, p_n\}$ is given by

$$H(x) = -\sum_{i=1}^{n} p_i \log p_i \qquad (1.7.2)$$

This definition suggests that entropy of fuzzy event, A, of the finite set $\{x_1, x_2, \dots, x_n\}$ with respect to a probability distribution $P = \{p_1, p_2, \dots, p_n\}$ be defined as follow

$$H^p(A) = -\sum_{i=1}^{n} \mu_A(x_i) p_i \log p_i \qquad (1.7.3)$$

Where μ_A is the membership function of A. Since (1.7.1) expresses the entropy of a distribution$P$, (1.7.2) represents the entropy of a fuzzy event A with respect to the distribution$P$. Thus, (1.7.1) does not reduce to (1.7.2) when A is non-fuzzy, unless A is taken to be the whole space $\{x_1, x_2, \dots, x_n\}$. Intuitively, $H^p(A)$ may be interpreted as the uncertainty associated with a fuzzy event.

1.8 Fuzzy Reliability:-

System failure engineering is primarily concerned with the failure and related problems. Specifically, by system failure engineering we mean the technological area comprising all failure oriented or failure driven aspects. So, it may compass reliability, safety, security, and so on. If everything went well and met desired requirements, then there would be no dissatisfaction, no failure, and therefore there would be no system failure engineering. Unfortunately, this is not the case. Actually, failure is a nearly unavoidable phenomenon with technological products and systems. One can observe various kind of failure in various circumstances: space shuttle explosion, nuclear reaction accident, airplane crash, chemical plant leak, bridge break and electrical network collapse. One can also observe defective screw, faulty VLSI chip, error us management decision, and so on. Failures can be frequent or rare. The causes of failure are diverse. They can be physical, human, logical and even financial. The effects of failure may be minor or disastrous, and various kinds of criteria and factors can be taken into account to define what a failure means: structure, performance, cost and even subjective intention. However, whatever failure is, if the effect of it tends to be critical, research on it becomes essential.

In conventional reliability theory [18], it is assumed that components and systems have only two abrupt states: good and bad. This implies that the success and failure are precisely defined and there is no intermediate state between them. That is, the failure or success criterion is binary. Even in the research of multi- stat systems [10], the failure or success criterion is also assumed to be binary. In other words, in conventional reliability theory and multi-stat systems, it is assumed that the system states can be binary defined in terms of some structure function (e.g., coherent structure function) of component states. Needless to say, this assumption is valid in extensive cases.

However, the above assumption may not be true in every case. In degradable computing systems the attribute of performance degradation is prominent and should be taken into account in the failure or success criterion [29]. If we treat quality as a body of performance indices (static or dynamic), it is easy to see that quality can be factor of the failure or success criterion. This builds a bridge linking quality control and failure research. Further, it has been argued that other factors like cost, purchasability, etc., should also be taken into account in the definition of failure or success in some cases [86, 96]. After all, besides the structural factors, others like performance, quality, cost, etc., can make contributions to the failure or success criterion. This leads us to a general definition of failure or success.

Let $A = \{a_1, a_2, \dots, a_n\}$ be a set of factors of concern. Let

$$x_{s_i}: a_i \to [0, 1]$$

$$x_{F_i}: a_i \to [0, 1]$$

We call $\{x_{s_i}\}$ success factor variables, and $\{x_{F_i}\}$ failure factor variable. Let

$$\mu_s = \mu_s(x_{s_1}, x_{s_2}, \dots, x_{s_n}):\ [0,1]^n \to [0,1]$$

$$\mu_F = \mu_s(x_{F_1}, x_{F_2}, \dots, x_{F_n}):\ [0,1]^n \to [0,1]$$

we call μ_s (system) success variable or success membership function, and μ_F (system) failure variable or failure variable function. Then system success S and system failure F are defined as fuzzy sets.

$$S = \{(x_1, x_2, \dots, x_n), \mu_s(x_1, x_2, \dots, x_n)\} x_i \in [0,1]$$

$$F = \{(x_1, x_2, \dots, x_n), \mu_z(x_1, x_2, \dots, x_n)\} x_i \in [0,1]$$

Since success and failure factor variables are defined on *A*, we can also define S and F directly on *A*. that is,

$$S = \{(x_1, x_2, \dots, x_n), \mu_s(x_1, x_2, \dots, x_n)\} x_i \in A$$

$$F = \{(x_1, x_2, \dots, x_n), \mu_z(x_1, x_2, \dots, x_n)\} x_i \in A$$

The generality of the above definition can be easily justified. In conventional reliability theory [18], we treat a_i as the ith component in a system, and $x_{s_i}$and x_{F_i} represent its states (0 or 1). Then μ_s and μ_F coincide with the corresponding system structure function. In a degradable computing system [29], $a_1, a_2, \dots, a_n$ can represent the system (non-fuzzy) states and the corresponding success (failure) factor variables represent the relative performance indices. Then μ_s and μ_F can be accordingly determined. For a software system, a_i can represent the ith module, and $x_s(x_{F_i})$ represents its quality index. Then μ_s can be interpreted as a system quality variable. Alternatively, we can treat $\{a_i\}$ as a set of quality factors such as correctness, reliability, efficiency, integrity, usability, maintainability, flexibility, portability, reusability, and so on. The factors in turn determine the quality variable μ_s. Anyway, defining failure and success as fuzzy sets enable them to be widely interpreted.

1.9 Measure and probability:-

1.9.1 Field and sigma (σ) field:

Let Ω be a space of elements *X*. A non empty class, $\Re$ of sets of Ω, closed under complementation and finite union is called a field. i.e. $\Re$ is a field if it satisfies the following axioms:-

i. $\Re$ is non empty.

ii. If $\mathrm{A} \in \Re$ then $\mathrm{A}^c \in \Re$ where A^c is the complement of A relative to Ω.

iii. If $A_1, A_2, \ldots, A_n \in \Re$ then $\bigcup_{i=1}^{n} \mathrm{A}_i \in \Re$ if axiom (iii) is replaced by the axiom.

iv. If $A_1, A_2, \ldots, A_n \in \Re$ then $\bigcup_{i=1}^{n} \mathrm{A}_i \in \Re$ .then $\Re$ is called the σ-field.

Remark 1.9.1: It can be easily verified that the null set ϕ, the space and the countable intersection of sets of field also belongs to $\Re$.

1.9.2 Measurable set and measurable space:

The subset belonging to the σ -field $\Re$ are called measurable set.

The doublet $(\Omega, \Re)$ is called measurable space.

1.9.3 Measure and measure space:

The measurable space $(\Omega, \Re)$ indicates that this is the structure upon which a measure can be defined.

A real valued function μ defined on $(\Omega, \Re)$ is called a measure if it satisfies the fallowing axioms.

(i) $\mu(\phi) = 0,$ where ϕ isnon - empty set.

(ii) $\mu(A) \geq 0,$ for all $A \in \Re$.

(iii) if $A_1, A_2, \ldots$ are disjoint measurable sets, then $\mu\left(\bigcup_{i=1}^{n} \mathrm{A}_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$

Property (iii) is called σ- Additivity.

Remarks:

i) A set of μ - measure zero is said to be a μ - null set and relations valid outside a μ - null set are said to be valid almost everywhere μ .

ii) If $\mu(\Omega)$ is finite then μ is said to be finite measure.

iii) A measure is said to be a σ- finite if the space Ω can be partitioned into a countable numbers of sets in $\Re$, for each of which the value of μ is finite.

iv) The triplet $(\Omega, \Re, \mu)$ is called measure space.

v) μ is called the finite measure on $(\Omega, \Re)$, if an addition the above axioms (i), (ii) and (iii), we have $\mu(\Omega)=1$, where Ω is the space of elementary events or sample space. A probability measure is usually denoted by *p*.

A probability space is the triplet $(\Omega, \Re, P)$, formed by a sample space Ω, a σ-field $\Re$ defined on Ω and a probability measure p defined on $(\Omega, \Re)$. All measure sets $A \in \Re$ are called events.

Thus, with every event *A* consisting of one or more outcomes of an experiment, we associated a numerical quantity, called the probability of *A* denoted by *P (A)* which will measure the chance that event *A* will occur, we take $0 \leq P(A) \leq 1$.

1.9.4 Function:

If *X* and *Y* be two non empty sets, then a function *f* from the set *X* into set *Y* is a correspondence (mapping) such that for each element of *X,* there exists only one element *Y.* This correspondence is generally denoted as $f; X \rightarrow Y$. if $x \in X$ and $y \in Y$ then *y* is said to be a function of *x* and we write $y = f(x)$.

1.9.5 Measurable function and random variable:

A real valued function $f(\cdot)$ defined on Ω, the sample space is said to be an $\Re$ measurable function or simply measurable function if for every real number *r*, $\{X : f(x) \leq r\} \in \Re$.

If $(\Omega, \Re, P)$ is a probability space, then a $\Re$ measurable function $f(\cdot)$ is called a random variable.

1.10 Some Mathematical functions and Inequalities:-

1.10.1 Convex Function:

A real valued function $f(x)$ defined on (*a*, *b*) is said to be convex function if for every α such that $0 \le \alpha \le 1$ and for any two points x_1 and x_2 such that $a < x_1 < x_2 < b$, we have

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2) \tag{1.10.1}$$

If we putα = ½, then (1.10.1) reduces to

$$f\left(\frac{x_1 + x_2}{2}\right) \le \frac{f(x_1) + f(x_2)}{2} \qquad (1.10.2)$$ which is taken as the definition of convexity.

Remark 1.10.1:

If $f''(x) \ge 0$, then $f(x)$ is convex function.

1.10.2 Strictly Convex Function:

A real valued function $f(x)$ defined on *(a, b)* is said to be strictly convex function if for every α such that $0 < \alpha < 1$ and for any two points x_1 and x_2 in *(a, b)* we have

$$f[\alpha x_1 + (1-\alpha)x_2] < \alpha f(x_1) + (1-\alpha) f(x_2) \tag{1.10.3}$$

Remark 1.10.2: If $f''(x) > 0$, then $f(x)$ is strictly convex function.

1.10.3 Joint Convexity:

Let $f\,(0,\infty) \to \Re$ be a convex, then $C_f\,(p,q)$ is jointly convex in *p* and *q*, where $p, q \in \Re_+^n$.

1.10.4 Concave Function:

A function $f(x)$ is said to be concave if $-f(x)$ is convex.

Remark 1.10.3: If $f''(x) \le 0$, then $f(x)$ is concave function.

1.10.5 Strictly Concave Function:

A function $f(x)$ is said to be strictly concave if $-f(x)$ is strictly convex.

Remark 1.10.4: $f''(x) < 0$, then $f(x)$ is strictly concave function.

1.10.6 Log - Concave Function:

A function $f(x)$ is said to be log - concave if every δ,

$$\frac{1}{2} In\ f(x-\delta) + \frac{1}{2} In\ f(x+\delta) \leq In\ f(x)$$

If a density is log – concave, we can always assume that it is log – concave because densities are defined up to a set of measure zero.

1.10.7 Increasing Function:

Let I be an open interval contained in the domain of a real function. The function $f(x)$ is an increasing function on I if $x_1 < x_2$ in I, implies

$$f(x_1) \leq f(x_2).$$

1.10.8 Decreasing Function:

Let I be an open interval contained in the domain of a real function. The function $f(x)$ is a decreasing function on I if $x_1 < x_2$ in I, implies

$$f(x_1) \geq f(x_2).$$

1.10.9 Maximum of a Function:

A function $f(x)$ is said to have a maximum value in an interval I at $x_\circ$, if $f(x_\circ) \geq f(x)$ for all x in I.

1.10.10 Minimum of a Function:

A function $f(x)$ is said to have a minimum value in an interval I at $x_\circ$, if $f(x_\circ) \leq f(x)$ for all x in I.

The following theorems give the working rule for finding the points of local maxima or points of local minima.

1.10.11 Some inequalities:

i) Jensen's inequality:

If X is a random variable such that $E(X)=\mu$ exists and $f(x)$ is a convex function, then

$$E[f(X)] \geq f[E(X)].$$

with equality iff the random variable X has a degenerate distribution at μ.

The following important concept is due to Csiszar and Korner [32].

Let: $f\,(0,\infty) \to \Re$ be a convex function. Then for any $p, q \in \Re_+^n$ with

$$p_n = \sum_{i=1}^{n} p_i > 0, \quad Q_n = \sum_{i=1}^{n} q_i > 0, \quad \text{we have the inequality}$$

$$C_f\,(p,q) \geq Q_n\, f\left(\frac{p_n}{q_n}\right).$$

The equality sign holds iff

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \frac{p_3}{q_3} = \ldots = \frac{p_n}{q_n}$$

In particular, for all we have

$$C_f\,(p,q) \geq f\,(1)$$

With equality iff $P = Q$.

ii) Holder's Inequality:

If $x_i, y_i > 0,\ i = 1,2,\ldots,n$ and $\frac{1}{p}+\frac{1}{q}=1,\ p>1,$ then the following inequality holds

$$\sum_{i=1}^{n} x_i y_i \leq \left[\sum_{i=1}^{n} x_i^p\right]^{\frac{1}{p}} \left[\sum_{i=1}^{n} y_i^q\right]^{\frac{1}{q}}.$$

iii) Chebychev's Inequality:

If X is a random variable with mean μ and variance σ^2, then for any positive number k

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

Or

$$P\{|X-\mu| < k\sigma\} \geq 1-\frac{1}{k^2}.$$

iv) Bienayne - Chebychev's Inequality:

Let $g(x)$ be a non-negative function of a random variable X, then for any $k > 0$,

$$P\{g(x) \geq k\} \leq \frac{E[g(x)]}{k}$$

v) Markov's Inequality:

If we take $g(x) = |x|$ in inequality (iv), then

$$P\{|x| \geq k\} \leq \frac{E|x|}{k}$$

which is Markov's Inequality.

Taking, $g(x) = |x|^r$ and replacing k by k^r in inequality (iv), we get a more generalized form of Markov's inequality.

$$P\{|x|^r \geq k^r\} \leq \frac{E|x|^r}{k^r}$$

vi) Log Sum Inequality:

For non-negative numbers $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ the log sum inequality is given as

$$\sum_{i=1}^{n} a_i \log\frac{a_i}{b_i} \geq \sum_{i=1}^{n} a_i \log\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

with equality, iff $\frac{a_i}{b_i} = k$ where k is a constant.

Shannon [87] has introduced two important ideas in the theory of information in communication engineering. The first idea is that information is a statistical concept. The statistical frequency distribution of the symbol that make-up a message must be considered before the notion can be discussed adequately. The second idea springs from the first and implies that on the basis of the frequency distribution there is an essentially unique function of the distribution which measures the amount of information.

In this chapter, we propose some generalized average codeword length and establish relationship with generalized fuzzy entropy. Some new fuzzy coding theorems have also been proved.

Fuzzy sets play a significant role in many deployed systems because of their capability to model non-statistical imprecision. Consequently, characterization and quantification of fuzziness are important issues that affect the management of uncertainty in many system models and designs. The notion of fuzzy sets was proposed by Zadeh [109] with a view to tackling problems in which indefinites arising from a sort of intrinsic ambiguity plays a fundamental role. Fuzziness, a texture of uncertainty, results from the lack of sharp distinction of the boundary of set. The concept of fuzzy sets in which imprecise knowledge can be used to define an event. A fuzzy set 'A' is represented as

$$A = \{X_i/\mu_A(X_i) : i = 1,2,\dots,n\},$$

Where $\mu_A(x_i)$ gives the degree of belongingness of the element $'x_i'$ to the set 'A'. If every element of the set 'A' is '0' or '1', there is no uncertainty about it and a set is said to be crisp set. On the other hand, a fuzzy set 'A' is defined by a characteristic function

$$\mu_A(x_i) = \{x_1, x_2, x_{3,\dots,}x_n\} \longrightarrow [0,1].$$

The function $\mu_A(x_i)$ associates with each $(x_i) \in R^n$ grade of membership function.

A fuzzy set A^* is called a sharpened version of fuzzy set A if the following conditions are satisfied:

$$\mu_{A^*}(x_i) \le \mu_A(x_i), \quad if \quad \mu_A(x_i) \le 0.5 \ for\ all\ i = 1,2,\dots,n$$

And $$\mu_{A^*}(x_i) \ge \mu_A(x_i), \quad if \quad \mu_A(x_i) \ge 0.5 \ for\ all\ i = 1,2,\dots,n$$

The importance of fuzzy set comes from the fact that it can deal with imprecise and inexact information. Its application areas span from design of fuzzy controller to robotics and artificial intelligence.

2.1. Introduction:-

Let $\Delta_n = \{P = (p_1, \dots, p_n): p_i \geq 0, \sum_i^n p_1 = 1\}$, $n \geq 2$ be a set of n-complete probability distributions. For any probability distribution $P = (p_1, \dots, p_n) \in \Delta_n$, Shannon's entropy is defined as

$$H(P) = -\sum_i^n p_i \log p_i \qquad (2.1.1).$$

Shannon [87] established the first noiseless coding theorem which states that for all uniquely decipherable codes, the lower bound for the average length $\sum_i^n p_i n_i$ lies between $H(P)$ *and* $H(P) + 1$, where $H(P)$ is defined in (2.1.1).

Many fuzzy measures have been discussed and derived by Kapur [58], Lowen [75], Pal and Bezdek [78] etc.

The basic noiseless coding theorems give the lower bound for the mean codeword length of a uniquely decipherable code in terms of Shannon's [87] measure of entropy. Kapur [59] has established relationship between probabilistic entropy and coding. But, there are situations where probabilistic measure of entropy does not work. To tackle such situations, instead of taking the probability, the idea of fuzziness was explored.

De Luca and Termini [33] introduced a measure of fuzzy entropy corresponding to measure Shannon's [87] information theoretic entropy and is given by

$$H(A) = -\sum_i^n \left[\mu_A(x_i) \log \mu_A(x_i) + \left(1 - \mu_A(x_i)\right) \log\left(1 - \mu_A(x_i)\right)\right] \ (2.1.2).$$

A measure of fuzziness in a fuzzy set should have at least the following properties:

$\boldsymbol{P_1(Sharpness)}$: $H(A)$is minimum if and only if A is a crisp set, i.e., $\mu_A(x_i) = 0 \text{ or } 1 \ \forall_{i.}$

$\boldsymbol{P_2(maximality)}$: $H(A)$is maximum, if and only if A is most fuzzy set, i.e,$\mu_A(x_i) = \frac{1}{2} \ \forall_{i.}$

$\boldsymbol{P_3(resolution)}$: $H(A^*) \leq H(A)$, Where A^* is a sharpened version of$A$.

$\boldsymbol{P_4(Symmetry)}$: $H(A) = H(A^c)$, Where A^c is the complement of set$A$.

Bhandari and Pal [25] gave some new information measure for fuzzy sets. Thus corresponding to Renyi's [84] entropy of order α, they suggested that the amount of ambiguity or fuzziness of order α should be:

$$H_\alpha(A) = \frac{1}{1-\alpha}\sum_i^n \log[\mu_A^\alpha(x_i) + (1-\mu_A)^\alpha]; \alpha \neq 1, \alpha > 0 \qquad (2.1.3)$$

Kapur [58] has taken measure of fuzzy entropy corresponding to HavradaCharvat [50] as

$$H^\alpha(A) = \frac{1}{1-\alpha}\sum_i^n \left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right\} - 1\right] \qquad (2.1.4)$$

Corresponding to Campbell's [28] measure of entropy, the fuzzy entropy can be taken as:$H_\alpha^/(A) = \frac{1}{1-\alpha}\log\left[\sum_i^n\left\{\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right\}^\alpha\right];\ \alpha \neq 1, \alpha > 0$ (2.1.5)

Corresponding to Sharma and Taneja [89] measure of entropy of degree(α, β), Kapur [59] has taken the following measure of entropy:

$$H_{\alpha,\beta}^/(A)\frac{1}{\beta-\alpha}\sum_i^n\left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right\} - \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}\right]; \quad \alpha \geq 1,\ \beta \leq 1 \text{ or } \alpha \leq 1, \beta \geq 1 \qquad (2.1.6)$$

Corresponding to Kapur [58] measure of entropy of degree(α, β), Kapur has given measure of entropy for fuzzy sets as:

$$H_{\alpha,\beta}^/(A) = \frac{1}{\alpha+\beta-2}\sum_i^n\left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right\} + \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\} - 2\right]$$

(2.1.7)

Tsalli's,C. [100] has given the following measure of entropy

$$T_\alpha(P) = \frac{\sum_i^n p_i^\alpha - 1}{1-\alpha}, \alpha \neq 1, \alpha > 0 \qquad (2.1.8).$$

Corresponding to (2.1.8), the average codeword length is

$$L_\alpha = \frac{1}{1-\alpha}\left[\sum_i^n p_i D^{(1-\alpha)n_i} - 1\right]; \alpha \neq 1, \alpha > 0 \qquad (2.1.9).$$

Corresponding to (2.1.8), the fuzzy entropy can be taken as

$$T_\alpha(A) = \frac{1}{1-\alpha}\left[\sum_i^n\left\{\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right\} - 1\right];\ \alpha \neq 1, \alpha > 0 \qquad (2.1.10).$$

And its corresponding average codeword length as

$$L_\alpha = \frac{1}{1-\alpha}\left[\sum_i^n \left\{\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right\} D^{(1-\alpha)n_i} - 1\right];\ \alpha \neq 1,\ \alpha > 0 \qquad (2.1.11).$$

Remark:

(i): As $\alpha \to 1$, (2.1.8) tends to (2.1.1)

(ii): As $\alpha \to 1$, (2.1.9) tends to average codeword length given by Shannon a

$$L = \sum_i^n p_i n_i \quad (2.1.12)$$

In section 2.2, we propose some noiseless coding theorems connected with Tsalli's, C. [100] entropy.

2.2 Noiseless Coding Theorems:-

Theorem 2.2.1: For all uniquely decipherable codes

$T_\alpha(A) \le L_\alpha$(2.2.1)

Where

$$L_\alpha = \frac{1}{1-\alpha}\left[\sum_i^n \left\{\mu_A^\alpha(x_i) + \left(1-\mu(x_i)\right)^\alpha\right\} D^{(1-\alpha)(ni)} - 1\right]$$

Proof: By Holders inequality, we have

$$\textstyle\sum_i^n x_i y_i \ \ge \ \left(\sum_i^n x_i^p\right)^{\frac{1}{p}} \left(\sum_i^n y_i^q\right)^{\frac{1}{q}};\quad 0 < p < 1, q < 0 \ \ or \ \ 0 < \ q < 1, p < o \ \ (2.2.2)$$

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

$$y_i = \ \left[f\left(\mu_A(x_i), \mu_B(x_i)\right)\right]^{\frac{1}{t}}$$

and $\qquad p = -t \ \Rightarrow 0 < p < 1, q = \frac{t}{t+1} \Rightarrow q < 0$

Thus equation (2.2.2) becomes

$$\textstyle\sum_i^n \left[\left[f\left(\mu_A(x_i), \mu_B(x_i)\right)\right]^{\frac{-1}{t}} D^{-n_i} \left[f\left(\mu_A(x_i), \mu_B(x_i)\right)\right]^{\frac{1}{t}}\right] \ge$$

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

Using Kraft's inequality, we have

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

Or, $\quad \sum_i^n\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)\right]^{\frac{1}{t}} \leq \sum_i^n\left[\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)\right]D^{n_i t}\right]^{\frac{-1}{t}}$

Or, $\quad \sum_i^n\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)\right] \leq \sum_i^n\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)D^{n_i t}\right]$(2.2.3)

Dividing both sides by t, we get

$$\frac{\sum_i^n\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)\right]}{t} \leq \frac{\sum_i^n\left[\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)\right]D^{n_i t}\right]}{t}$$

Subtracting 'n' from both sides, we get

$$\sum_i^n\frac{\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)-1\right]}{t} \leq \sum_i^n\frac{\left[\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)\right]D^{n_i t}-1\right]}{t} \quad (2.2.4)$$

Taking

$$\alpha = 1-t, \alpha > 0, \qquad t = 1-\alpha$$

and$f\left(\mu_A(x_i),\mu_B(x_i)\right) = \mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha$,

equation (2.1.4) becomes

$$\frac{1}{\alpha-1}\sum_i^n\left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu(x_i)\right)^\alpha\right\}-1\right]$$

$$\leq \frac{1}{\alpha-1}\sum_i^n\left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu(x_i)\right)^\alpha\right\}D^{(1-\alpha)(ni)}-1\right] \quad (2.2.5)$$

That is $\quad T_\alpha(A) \leq L_\alpha$, which proves the theorem.

Theorem 2.2.2: For all uniquely decipherable codes,

$$T_{\alpha,\beta} \leq L_{\alpha,\beta} \;, \qquad (2.2.6)$$

Where

$$L_{\alpha,\beta} = \frac{1}{\beta-\alpha}\sum_i^n \begin{bmatrix} \left\{\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha D^{(1-\alpha)(ni)}\right\} - \\ \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta D^{(1-\beta)(ni)}\right\} \end{bmatrix} \tag{2.2.7}$$

and either $\alpha \geq 1,\ \beta \leq 1\ or\ \beta \geq 1, \alpha \leq 1$

Proof: Since from (2.2.5), we have

$$\frac{1}{\alpha-1}\sum_i^n\left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu(x_i)\right)^\alpha\right\} - 1\right]$$

$$\leq \frac{1}{\alpha-1}\sum_i^n\left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu(x_i)\right)^\alpha\right\}D^{(1-\alpha)(ni)} - 1\right] \tag{2.2.8}$$

Multiplying both sides by$(1-\alpha)$, we have

$$\sum_i^n\left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu(x_i)\right)^\alpha\right\} - 1\right]$$

$$\leq \sum_i^n\left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu(x_i)\right)^\alpha\right\}D^{(1-\alpha)(ni)} - 1\right] \tag{2.2.9}$$

Changing α to β, we get

$$\sum_i^n\left[\left\{\mu_A^\beta(x_i) + \left(1-\mu(x_i)\right)^\beta\right\} - 1\right]$$

$$\leq \sum_i^n\left[\left\{\mu_A^\beta(x_i) + \left(1-\mu(x_i)\right)^\beta\right\}D^{(1-\beta)(ni)} - 1\right] \tag{2.2.10}$$

Subtracting (2.2.10) from (2.2.9), and dividing by $\beta-\alpha$, we get

$$\frac{1}{\beta-\alpha}\sum_i^n\left[\left\{\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right\} - \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}\right]$$

$$\leq \frac{1}{\beta-\alpha}\sum_i^n \begin{bmatrix} \left\{\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha D^{(1-\alpha)(ni)}\right\} - \\ \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta D^{(1-\beta)(ni)}\right\} \end{bmatrix} \tag{2.2.11}$$

That is

$T_{\alpha,\beta} \leq L_{\alpha,\beta}$, which proves the theorem

Theorem 2.2.3: For all uniquely decipherable codes

$$T^{/}{}_{\alpha,\beta} \leq L^{/}{}_{\alpha,\beta} \quad (2.2.12)$$

Where

$$L^{/}{}_{\alpha,\beta} = \frac{1}{\alpha+\beta+2}\sum_i^n \left[\begin{array}{c}\left\{\mu_A^{\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha} D^{(1-\alpha)(ni)}\right\} + \\ \left\{\mu_A^{\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{\beta} D^{(1-\beta)(ni)}\right\} - 2\end{array}\right] (2.2.13)$$

Proof: The result can be easily proved by adding (2.2.9) and (2.2.10), then dividing by

$(\alpha + \beta - 2)$.

Corr.: Taking β = 1 in (2.1.11), we get

$$\frac{1}{\alpha-1}\sum_i^n \left[\left\{\mu_A^{\alpha}(x_i) + \left(1-\mu(x_i)\right)^{\alpha}\right\} - 1\right] \leq \frac{1}{\alpha-1}\sum_i^n \left[\left\{\mu_A^{\alpha}(x_i) + \left(1-\mu(x_i)\right)^{\alpha}\right\} D^{(1-\alpha)(ni)} - 1\right]$$

This is nothing but coding theorem corresponding to Tsalli's measure of fuzzy entropy.

Theorem.2.2.4: For all uniquely decipherable codes

$$T_{\alpha,\beta} \leq L^{\|}{}_{\alpha,\beta}, \qquad (2.2.14)$$

Where $T_{\alpha,\beta} = \frac{1}{\beta-\alpha}\left[\frac{\sum_i^n\left[\left\{\mu_A^{\alpha}(x_i)+\left(1-\mu(x_i)\right)^{\alpha}\right\}-1\right]}{\sum_i^n\left[\left\{\mu_A^{\beta}(x_i)+\left(1-\mu(x_i)\right)^{\beta}\right\}-1\right]}\right]$(2.2.15)

and

$$L^{\|}{}_{\alpha,\beta} = \frac{1}{\beta-\alpha}\left[\frac{\sum_i^n\left[\left\{\mu_A^{\alpha}(x_i)+\left(1-\mu(x_i)\right)^{\alpha}\right\}D^{(1-\alpha)(ni)}-1\right]}{\sum_i^n\left[\left\{\mu_A^{\alpha}(x_i)+\left(1-\mu(x_i)\right)^{\alpha}\right\}D^{(1-\alpha)(ni)}-1\right]}\right] \qquad (2.2.16)$$

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

Lemma 2.2.1: For all uniquely decipherable codes

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

Proof of the lemma:- From equation (2.2.4), we have

$$\sum_i^n D^{-n_i} \geq \left[\sum_i^n f\left(\mu_A(x_i), \mu_B(x_i)\right) D^{n_i t}\right]^{\frac{-1}{t}} \left[\sum_i^n f\left(\mu_A(x_i), \mu_B(x_i)\right)\right]^{\frac{1}{t}}.$$

Using Kraft's inequality, we have

$$\sum_i^n \left[f\left(\mu_A(x_i), \mu_B(x_i)\right)\right] \leq \sum_i^n \left[f\left(\mu_A(x_i), \mu_B(x_i)\right) D^{n_i t}\right].$$

Subtracting 'n' from both sides, we have

$$\sum_i^n \left[\left\{f\left(\mu_A(x_i), \mu_B(x_i)\right)\right\} - 1\right] \leq \sum_i^n \left[\left\{f\left(\mu_A(x_i), \mu_B(x_i)\right)\right\} D^{n_i t} - 1\right].$$

Taking $\alpha = 1 - t,\ \alpha > 0,\ t = 1 - \alpha$

and $f\left(\mu_A(x_i), \mu_B(x_i)\right) = \mu_A^\alpha(x_i) + \left(1 - \mu_A(x_i)\right)^\alpha$, we have

$$\sum_i^n \left[\left\{\mu_A^\alpha(x_i) + \left(1 - \mu(x_i)\right)^\alpha\right\} - 1\right]$$

$$\leq D^{(1-\alpha)(ni)} \sum_i^n \left[\left\{\mu_A^\alpha(x_i) + \left(1 - \mu(x_i)\right)^\alpha\right\} - 1\right] \tag{2.2.17}$$

This proves the lemma.

Proof of the theorem 2.2.4:

Changing α to β, in (2.2.17), we have

$$\sum_i^n \left[\left\{\mu_A^\beta(x_i) + \left(1 - \mu(x_i)\right)^\beta\right\} - 1\right]$$

$$\leq \sum_i^n \left[\left\{\mu_A^\beta(x_i) + \left(1 - \mu(x_i)\right)^\beta\right\} D^{(1-\beta)(ni)} - 1\right] \tag{2.2.18}$$

Dividing (2.2.18) to (2.2.17), we get

$$\frac{\sum_i^n \left[\left\{\mu_A^\alpha(x_i) + \left(1 - \mu(x_i)\right)^\alpha\right\} - 1\right]}{\sum_i^n \left[\left\{\mu_A^\beta(x_i) + \left(1 - \mu(x_i)\right)^\beta\right\} - 1\right]} \leq \frac{\sum_i^n \left[\left\{\mu_A^\alpha(x_i) + \left(1 - \mu(x_i)\right)^\alpha\right\} D^{(1-\alpha)(ni)} - 1\right]}{\sum_i^n \left[\left\{\mu_A^\beta(x_i) + \left(1 - \mu(x_i)\right)^\beta\right\} D^{(1-\beta)(ni)} - 1\right]}.$$

Dividing both side by$(\beta - \alpha)$, we have

$$\frac{1}{\beta-\alpha}\left[\frac{\sum_i^n \left[\left\{\mu_A^\alpha(x_i) + \left(1 - \mu(x_i)\right)^\alpha\right\} - 1\right]}{\sum_i^n \left[\left\{\mu_A^\beta(x_i) + \left(1 - \mu(x_i)\right)^\beta\right\} - 1\right]}\right] \leq \frac{1}{\beta-\alpha}\left[\frac{\sum_i^n \left[\left\{\mu_A^\alpha(x_i) + \left(1 - \mu(x_i)\right)^\alpha\right\} D^{(1-\alpha)(ni)} - 1\right]}{\sum_i^n \left[\left\{\mu_A^\beta(x_i) + \left(1 - \mu(x_i)\right)^\beta\right\} D^{(1-\beta)(ni)} - 1\right]}\right] \tag{2.2.19}$$

That is $T_{\alpha,\beta} \leq L^{\parallel}{}_{\alpha,\beta}$, this proves the theorem.

The R.H.S. of (2.2.19) is a new exponentiated mean code word length of order α and type β, defined by

$$L^{\|}{}_{\alpha,\beta} = \frac{1}{\beta-\alpha}\left[\frac{\sum_i^n\left[\left\{\mu_A^{\alpha}(x_i)+\left(1-\mu(x_i)\right)^{\alpha}\right\}D^{(1-\alpha)(ni)}-1\right]}{\sum_i^n\left[\left\{\mu_A^{\alpha}(x_i)+\left(1-\mu(x_i)\right)^{\alpha}\right\}D^{(1-\alpha)(ni)}-1\right]}\right].$$

It can be easily seen that

$$L^{\|}{}_{\alpha,\beta} = L^{\|}{}_{\beta,\alpha} \text{and} T_{\alpha,\beta} = T_{\beta,\alpha} .$$

So (2.2.9) holds for both when α < 1, β > 1 or α > 1, β < 1.

2.3 Generalized Fuzzy Average Codeword Length and Their Bounds:-

Mathai, A.M.[76] has given the measure of entropy as

$$M_\alpha(P) = \frac{1}{\alpha-1}\left[\sum_i^n p_i^{2-\alpha} - 1\right];\ \alpha \neq 1,\ -\infty < \alpha < 2 \tag{2.3.1}$$

Corresponding to this measure, we propose the following average codeword length as;

$$L_\alpha(P) = \frac{1}{\alpha-1}\left[\sum_i^n p_i D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)} - 1\right];\ \alpha \neq 1,\ \alpha > 0 \tag{2.3.2}$$

Corresponding to equation (2.3.1) we propose the following measure of fuzzy entropy as ;

$$M_\alpha(A) = \frac{1}{\alpha-1}\left[\sum_i^n\left\{\mu_A^{2-\alpha}(x_i) + \left(1-\mu(x_i)\right)^{2-\alpha}\right\} - 1\right] \tag{2.3.3}$$

And the corresponding average codeword length as;

$$L_\alpha = \frac{1}{\alpha-1}\left[\sum_i^n\left\{\mu_A^{2-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\alpha}\right\} D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)} - 1\right] \tag{2.3.4}$$

Remark:

(I) When $\alpha \to 1$, (2.3.1)tends to Shannon's entropy given as;

$$S(P) = -\sum_i^n p_i \log p_i (2.3.5)$$

(II): When $\alpha \to 1$, (2.3.2) tends to average codeword length given by;

$$L = \sum_i^n p_i n_i \tag{2.3.6}$$

In next section, some noiseless coding theorems connected with fuzzy entropy corresponding to Mathai's [76] entropy have been proved.

2.4 Generalized Fuzzy Noiseless Coding Theorems:-

Theorem 2.4.1: For all uniquely decipherable codes

$M_\alpha(A) \leq L_\alpha$(2.4.1)

Where

$$L_\alpha = \frac{1}{\alpha - 1}\left[\sum_i^n \left\{\mu_A^{2-\alpha}(x_i) + \left(1 - \mu(x_i)\right)^{2-\alpha}\right\} D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)} - 1\right]$$

Proof:-By Holders inequality, we have

$$\textstyle\sum_i^n x_i y_i \geq \left(\sum_i^n x_i^p\right)^{\frac{1}{p}}\left(\sum_i^n y_i^q\right)^{\frac{1}{q}} \quad ; 0 < p < 1, q < 0 \ or \ 0 < q < 1, p < 0 \quad (2.4.2)$$

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

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

Thus equation (2.4.2) becomes

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

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

Using Kraft's inequality, we have

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

or $\sum_i^n \left[f\left(\mu_A(x_i), \mu_B(x_i)\right)\right]^{\frac{1}{t}} \leq \sum_i^n \left[\left[f\left(\mu_A(x_i), \mu_B(x_i)\right)\right] D^{n_i t}\right]^{\frac{1}{t}}$

or $\sum_i^n \left[f\left(\mu_A(x_i), \mu_B(x_i)\right)\right] \leq \sum_i^n \left[f\left(\mu_A(x_i), \mu_B(x_i)\right) D^{n_i t}\right]$ (2.4.3)

Dividing both sides by t, we get

$$\frac{\sum_i^n \left[f\left(\mu_A(x_i), \mu_B(x_i)\right)\right]}{t} \leq \frac{\sum_i^n \left[\left[f\left(\mu_A(x_i), \mu_B(x_i)\right)\right] D^{n_i t}\right]}{t}$$

Subtracting n from both sides, we get

$$\sum_i^n \frac{[f(\mu_A(x_i),\mu_B(x_i))-1]}{t} \leq \sum_i^n \frac{[[f(\mu_A(x_i),\mu_B(x_i))]D^{n_it}-1]}{t} \qquad (2.4.4)$$

Taking $\alpha = \frac{1}{1-t},\ t = \frac{\alpha-1}{\alpha}\ ,\ \alpha > 0,\ \alpha \neq 1$ and

$$f(\mu_A(x_i), \mu_B(x_i)) = \mu_A^{2-\alpha}(x_i) + (1 - \mu_A(x_i))^{2-\alpha}$$

Thus equation (2.4.4) becomes

$$\frac{\alpha}{\alpha-1}\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + (1-\mu(x_i))^{2-\alpha}\right\} - 1\right]$$

$$\leq \frac{\alpha}{\alpha-1}\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + (1-\mu(x_i))^{2-\alpha}\right\} D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)} - 1\right]$$

Dividing both sides byα, we get;

$$\frac{1}{\alpha-1}\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + (1-\mu_A(x_i))^{2-\alpha}\right\} - 1\right]$$

$$\leq \frac{1}{\alpha-1}\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + (1-\mu_A(x_i))^{2-\alpha}\right\} D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)} - 1\right] \qquad (2.4.5)$$

That is $M_\alpha(A) \leq L_\alpha$.

Which proves the theorem.

Theorem 2.4.2: For all uniquely decipherable codes,

$$M_{\alpha,\beta} \leq L_{\alpha,\beta} \qquad (2.4.6)$$

Where, $$L_{\alpha,\beta} = \frac{1}{\beta-\alpha}\sum_i^n \left[\begin{array}{l}\left\{\mu_A^{2-\alpha}(x_i) + (1-\mu_A(x_i))^{2-\alpha} D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}\right\} - \\ \left\{\mu_A^{2-\beta}(x_i) + (1-\mu_A(x_i))^{2-\beta} D^{\left(\frac{\beta-1}{\beta}\right)(ni)}\right\}\end{array}\right] \qquad (2.4.7)$$

And $\alpha \geq 1,\ \beta \leq 1\ or\ \beta \geq 1,\ \alpha \leq 1.$

Proof: Since from (2.4.5), we have

$$\frac{1}{\alpha-1}\sum_{i}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}\right\}-1\right]\leq$$

$$\frac{1}{\alpha-1}\sum_{i}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}\right\}D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}-1\right]$$

Multiplying both sides by$(\alpha-1)$, we have

$$\sum_{i}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}\right\}-1\right]\leq\sum_{i=1}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}\right\}D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}-1\right]$$

(2.4.8).

Changing α to β, we have

$$\left[\sum_{i}^{n}\left\{\mu_A^{2-\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\beta}\right\}-1\right]\leq$$

$$\left[\sum_{i}^{n}\left\{\mu_A^{2-\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\beta}\right\}D^{\left(\frac{\beta-1}{\beta}\right)(ni)}-1\right] \quad (2.4.9)$$

Subtract (2.4.9) to (2.4.8), and divide by$(\beta-\alpha)$, we get;

$$\frac{1}{\beta-\alpha}\sum_{i}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}\right\}-\left\{\mu_A^{2-\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\beta}\right\}\right]\leq$$

$$\frac{1}{\beta-\alpha}\sum_{i}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}\right\}\left\{\mu_A^{2-\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\beta}D^{\left(\frac{\beta-1}{\beta}\right)(ni)}\right\}\right]$$

(2.4.10).

That is

$M_{\alpha,\beta}\leq L_{\alpha,\beta}$. This proves the theorem.

Theorem 2.4.3: For all uniquely decipherable codes

$$M^{/}{}_{\alpha,\beta}\leq L^{/}{}_{\alpha,\beta} \quad (2.4.11)$$

Where

$$L^{/}{}_{\alpha,\beta}=\frac{1}{\alpha+\beta+2}\sum_{i}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}\right\}-\left\{\mu_A^{2-\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\beta}D^{\left(\frac{\beta-1}{\beta}\right)(ni)}\right\}-2\right]$$

(2.4.12)

Proof: The result can be easily proved by adding (2.4.8) and (2.4.9) and then dividing by $(\alpha+\beta+2)$**.**

Theorem 2.4.4: For all uniquely decipherable codes

$$M_{\alpha,\beta} \le L'_{\alpha,\beta} \tag{2.4.13}$$

Where

$$M_{\alpha,\beta} = \frac{1}{\beta-\alpha}\left[\frac{\sum_{i=1}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}\right\}-1\right]}{\sum_{i=1}^{n}\left[\left\{\mu_A^{2-\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\beta}\right\}-1\right]}\right] \tag{2.4.14}$$

and

$$L'_{\alpha,\beta} = \frac{1}{\beta-\alpha}\left[\frac{\sum_{i=1}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}\right\}-1\right]}{\sum_{i=1}^{n}\left[\left\{\mu_A^{2-\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\beta}D^{\left(\frac{\beta-1}{\beta}\right)(ni)}\right\}-1\right]}\right] \tag{2.4.15}$$

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

Lemma 1: For all uniquely decipherable codes.

$$\sum_{i=1}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}\right\}-1\right] \le \sum_{i=1}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}\right\}-1\right]$$

Proof of the Lemma. From equation (2.4.3), we have;

$$\sum_{i}^{n}\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)\right] \le \sum_{i}^{n}\left[\left[f\left(\mu_A(x_i),\mu_B(x_i)\right)\right]D^{n_i t}\right]$$

Subtracting ‘n’ from both sides, we get

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

Taking $\alpha = \frac{1}{1-t}$, $t = \frac{\alpha-1}{\alpha}$, and $f\left(\mu_A(x_i),\mu_B(x_i)\right) = \mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}$, we have;

$$\sum_{i}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}\right\}-1\right] \le$$

$$\sum_{i}^{n}\left[\left\{\mu_A^{2-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{2-\alpha}D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}\right\}-1\right] \tag{2.4.16}$$

Which proves the lemma.

Proof of the theorem 2.4.4:

Changing $\alpha\ to\ \beta$ in (2.4.16), we have;

$$\sum_i^n \left[\left\{\mu_A^{2-\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\beta}\right\} - 1\right] \le$$

$$\sum_i^n \left[\left\{\mu_A^{2-\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\beta} D^{\left(\frac{\beta-1}{\beta}\right)(ni)}\right\} - 1\right] \quad (2.4.17)$$

Dividing (2.4.17) to (2.4.16), we get;

$$\frac{\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\alpha}\right\} - 1\right]}{\sum_i^n \left[\left\{\mu_A^{2-\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\beta}\right\} - 1\right]} \le$$

$$\frac{\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\alpha} D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}\right\} - 1\right]}{\sum_i^n \left[\left\{\mu_A^{2-\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\beta} D^{\left(\frac{\beta-1}{\beta}\right)(ni)}\right\} - 1\right]}$$

Dividing both sides by $\beta-\alpha, we\ have$;

$$\frac{1}{\beta-\alpha}\left[\frac{\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\alpha}\right\} - 1\right]}{\sum_i^n \left[\left\{\mu_A^{2-\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\beta}\right\} - 1\right]}\right]$$

$$\le \frac{1}{\beta-\alpha}\left[\frac{\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\alpha} D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}\right\} - 1\right]}{\sum_i^n \left[\left\{\mu_A^{2-\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\beta} D^{\left(\frac{\beta-1}{\beta}\right)(ni)}\right\} - 1\right]}\right]$$

(2.4.18)

$\Rightarrow M_{\alpha,\beta} \le L'_{\alpha,\beta}$. The R.H.S. is a new exponentiated mean codeword length of order α and type β and is defined as;

$$\frac{1}{\beta-\alpha}\left[\frac{\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\alpha}\right\} - 1\right]}{\sum_i^n \left[\left\{\mu_A^{2-\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\beta}\right\} - 1\right]}\right] \le \frac{1}{\beta-\alpha}\left[\frac{\sum_i^n \left[\left\{\mu_A^{2-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\alpha} D^{\left(\frac{\alpha-1}{\alpha}\right)(ni)}\right\} - 1\right]}{\sum_i^n \left[\left\{\mu_A^{2-\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{2-\beta} D^{\left(\frac{\beta-1}{\beta}\right)(ni)}\right\} - 1\right]}\right].$$

The notion of fuzzy sets was proposed by Zadeh [109] with a view to tackling problems in which indefiniteness arising from a sort of intrinsic ambiguity plays a significant role. Fuzziness, a feature of uncertainty, results from the lack of sharp distinction of the boundary of a set, i.e., an individual is neither definitely a member of the set nor definitely not a member of it. The first to qualify the fuzziness was made by Zadeh [107], who based on probabilistic framework introduced the entropy combining probability and membership function of a fuzzy event as weighted Shannon entropy.

In this chapter, several coding theorems have been obtained by considering some parametric fuzzy entropy functions involving utilities. In the literature of information theory several types of coding theorems involving fuzzy entropy functions exists. The coding theorems obtained here are not only new but also generalizes some well known results available in the literature.

3.1. Introduction:-

Let X be discrete random variable taking on a finite number of possible values $X = (x_1, x_2, \ldots, x_n)$with respective membership function $A = \{\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n)\} \rightarrow [0,1], \mu_A(x_i)$gives of the elements the degree of belongingness x_i to the set A.The function $\mu_A(x_i)$ associates with each $x_i \in R^n$ a grade of membership to the set A and is known as membership function.

Denote

$$X = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \\ \mu_A(x_1) & \mu_A(x_2) & \cdots & \mu_A(x_n) \end{bmatrix} \tag{3.1.1}$$

We call the scheme (3.1.1) as a finite fuzzy information scheme. Every finite scheme describes a state of uncertainty. De Luca and termini [33] introduced a quantity which, in a reasonable way to measures the amount of uncertainty (fuzzy entropy) associated with a given finite scheme. This measure is given by

$$H(A) = -\sum_{i}^{n} \left[\mu_A(x_i) \log \mu_A(x_i) + \left(1 - \mu_A(x_i)\right) \log\left(1 - \mu_A(x_i)\right)\right] \tag{3.1.2}$$

The measure (3.1.2) serve as a very suitable measure of fuzzy entropy of the finite information scheme(3.1.1).

Let a finite source of n source symbols $X = (x_1, x_2, \dots, x_n)$ be encoded using alphabet of D symbols, then it has been shown by Feinstein [39] that there is a uniquely decipherable/ instantaneous code with lengths $l_1, l_2 \dots, l_n$ iff the following Kraft [65] inequality is satisfied

$$\sum_i^n D^{-l_i} \le 1 \qquad (3.1.3).$$

Belis and Guiasu [22] observed that a source is not completely specified by the probability distribution P over the source alphabet X in the absence of qualitative character. So it can be assumed (Belis and Guiasu [22]) that the source alphabet letters are assigned weights according to their importance or utilities in view of the experimenter.

Let $U = (u_1, u_{2,} \dots, u_n)$ be the set of positive real numbers, u_i is the utility or importance ofx_i. The utility, in general, is independent of probability of encoding of source symbol x_i, i.e, p_i. The information source is thus given by;

$$X = \begin{bmatrix} X_1 & X_2 \dots & X_n \\ p_1 & p_2 \dots & p_n \\ u_1 & u_2 \dots & u_n \end{bmatrix}, u_i > 0\ p_i \ge 0, \sum_i^n p_i = 1 \qquad (3.1.4)$$

Belis and Guiasu [22] introduced the following quantitative- qualitative measure of information

$$H(P, U) = -\sum_i^n u_i p_i \log p_i \qquad (3.1.5)$$

Which is a measure for the average of quantity of 'variable' or 'useful' information provided by the information source(3.1.4).

Guiasu and Picard [44] considered the problem of encoding the letter output by the source (3.1.4) by means of a single letter prefix code whose codeword's $c_1, c_2, \dots, c_n$ are of lengths $l_1, l_2, \dots, l_n$ respectively and satisfy the Kraft's inequality(3.1.3), they included the following 'useful' mean length of the code.

$$L(U) = \frac{\sum_i^n u_i p_i l_i}{\sum_i^n u_i p_i} \qquad (3.1.6)$$

Further they derived a lower bound for (3.1.6). However, Longo [74] interpreted (3.1.6) as the average transmission cost of the letters x_i and derived the bounds for this cost function.

Now, corresponding to (3.1.5) and(3.1.6), we have the following fuzzy measures;

$$H(A,U) = -\sum_{i=1}^{n} u_i\{\mu_A(x_i) + (1-\mu_A(x_i))\}\log\{\mu_A(x_i) + (1-\mu_A(x_i))\} \quad (3.1.7)$$

and

$$L(U) = \frac{\sum_{i=1}^{n} u_i\{\mu_A(x_i)+(1-\mu_A(x_i))\}l_i}{\sum_{i=1}^{n} u_i\{\mu_A(x_i)+(1-\mu_A(x_i))\}} \quad (3.1.8)$$

respectively.

In the next section, the bounds have been derived in terms of generalized 'useful' fuzzy cost measure and 'useful' fuzzy information measure of order α and type β. The main aim of studying these bounds is to generalize some well known results available in the literature.

3.2. Bounds for Generalized Measure of Cost:-

In the derivation of the cost measure (3.1.8), it is assumed that the cost is linear function of the code length, but this is not always the case. There are occasions when the cost behaves like an exponential function of codeword lengths. Such types of functions occur frequently in market equilibrium and growth models in economics. Thus sometimes it might be more appropriate to choose a code which minimizes the monotonic function of the quantity.

$$C = \sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A{}^{\beta}(x_i) + (1-\mu_A(x_i))^{\beta}\right\} D^{\left(\frac{1-\alpha}{\alpha}\right)l_i}. \quad (3.2.1)$$

Where $\alpha > 0\ (\neq 1)$, $\beta > 0$ are the [parameters related to cost].

In order to make the result more comparable with the usual noiseless coding theorem, instead of minimizing (3.2.1) we minimize

$$L_{\alpha}^{\beta}(U) = \frac{1}{2^{1-\alpha}-1}\left[\left(\frac{\sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A{}^{\beta}(x_i) + (1-\mu_A(x_i))^{\beta}\right\} D^{\left(\frac{1-\alpha}{\alpha}\right)l_i}}{\sum_{i=1}^{n}\left(u_i\{\mu_A(x_i) + (1-\mu_A(x_i))\}\right)^{\beta}}\right)^{\alpha} - 1\right] \quad (3.2.2)$$

where, $\alpha > 0\ (\neq 1),\ \beta > 0.$

Which is monotonic function of C and is the 'useful' fuzzy average code length of order α and type β.

Clearly, if $\alpha \to 1$, $\beta = 1$ (3.2.2) reduces to (3.1.8) which further reduces to ordinary mean length corresponding to Shannon [87], when $u_i = 1,\ \forall\, i = 1,2,\dots,n$. It can also be noted that

(3.2.2) is monotonic non-decreasing function of α and if all the $l_i{}'^s$ are same, say $l_i = l,\ \forall\, i = 1,2,\dots,n$ and $\alpha\to 1$, then $L_\alpha{}^\beta(U) = l$. This is an important property for any measure of length to posses.

Now, consider a function, which is 'useful' fuzzy information measure of order α and type β.

$$H_\alpha^\beta(A,U) = \frac{1}{2^{1-\alpha}-1}\left[\frac{\sum_{i=1}^n u_i^\beta\left\{\mu_A^{\alpha+\beta-1}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^n\left(\mu_i^\beta\{\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\}\right)^\beta} - 1\right] \tag{3.2.3}$$

Where, $\alpha > 0\ (\neq 1), \beta > 0,\ \mu_A(x_i) \geq 0;\ \forall\, i = 1,2,\dots,n;\ \sum_{i=1}^n \mu_A(x_i) \leq 1$.

Remark 3.2.1:

1) When $\beta = 1$, (3.2.3) reduces to the measure of 'useful' fuzzy information corresponding to Hooda and Ram [53].
2) When $\alpha\to 1, \beta = 1$, (3.2.3) reduces to the measure corresponding Belis and Guiasu [22].
3) When $\alpha\to 1,\ \beta = 1$ and $u_i = 1, \forall\, i = 1,2,\dots,n$ (3.2.3) reduces to the Du Luca and Termini [33].

Also the bounds are obtained for the measure (3.2.3) under the condition;

$$\sum_{i=1}^n u_i^\beta\left\{\mu_A^{\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\beta-1}\right\}D^{-l_i} \leq \sum_{i=1}^n u_i^\beta\left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\} \tag{3.2.4}$$

It may be seen that in case $\beta = 1$ and $u_i = 1, \forall\, i = 1,2,\dots,n$ (3.2.4) reduces to the Kraft [65] inequality (3.1.3). Also, D is the size of the code alphabet.

Theorem 3.2.1: For all integers D (D≥2), let l_i satisfies (3.2.4), then the generalized average 'useful' codeword length satisfies;

$$L_\alpha^\beta(U) \geq H_\alpha^\beta(A;U) \tag{3.2.5}$$

Equality holds iff

$$l_i = -\log\left\{\mu_i^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right\} + \log\frac{\sum_{i=1}^n u_i^\beta\left\{\mu_A^{\alpha+\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^n u_i^\beta\left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}}$$

(3.2.6)

Proof: By Holder's inequality.

$$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}} \tag{3.2.7}$$

For all, $x_i, y_i > 0, i = 1,2, \dots, n$; $\frac{1}{p} + \frac{1}{q} = 1, p < 1 (\neq 0), q < 0$ *or* $q < 1(\neq 0), p < 0$.We see the equality holds iff there exists a positive constant C such that

$$x_i^p = c y_i^q \tag{3.2.8}$$

Making the substitution

$$p = \frac{\alpha-1}{\alpha}, q = 1-\alpha.$$

$$x_i = \frac{\left(u_i\left\{\mu_A(x_i) + \left(1-\mu_A(x_i)\right)\right\}\right)^{\frac{\alpha\beta}{\alpha-1}} D^{-l_i}}{\sum_{i=1}^{n}\left(u_i\left\{\mu_A(x_i) + \left(1-\mu_A(x_i)\right)\right\}\right)^{\frac{\alpha\beta}{\alpha-1}}}$$

And $$y_i = \frac{u_i^{\frac{\beta}{1-\alpha}}\left(\mu_A^{\frac{\alpha+\beta-1}{1-\alpha}}(x_i) + \{1-\mu_A(x_i)\}^{\frac{\alpha+\beta-1}{1-\alpha}}\right)}{\sum_{i=1}^{n} u_i\left\{\mu_A(x_i) + \left(1-\mu_A(x_i)\right)\right\}^{\frac{\beta}{1-\alpha}}}$$

In (3.2.7), we get;

$$\frac{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\beta-1}\right\} D^{-l_i}}{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\beta-1}\right\}} \geq$$

$$\left[\frac{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{\beta}\right\} D^{\left(\frac{1-\alpha}{\alpha}\right) l_i}}{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{\beta}\right\}}\right]^{\frac{\alpha}{\alpha-1}} \left[\frac{\sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A^{\alpha+\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i) + \left(1-\mu_A(x_i)\right)^{\beta}\right\}}\right]^{\frac{1}{1-\alpha}}$$

Using the condition (3.2.4), we get;

$$\left[\frac{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}D^{\left(\frac{1-\alpha}{\alpha}\right)l_i}}{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}}\right]^{\frac{\alpha}{\alpha-1}}$$

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

Taking $0<\alpha<1$, and raising power both sides $(1-\alpha)$, we get;

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

Multiplying both sides by $\frac{1}{2^{1-\alpha}-1}>0$ for $0<\alpha<1$ and after simplifying, we get;

$$L_{\alpha}^{\beta}(U)\geq H_{\alpha}^{\beta}(A;U)$$

For all $\alpha > 1$, the proof follows along the similar lines.

Theorem 3.2.2: For every code with lengths $l_1, l_2, \ldots . l_n$ satisfies (3.2.4), $L_{\alpha}^{\beta}(U)$ can be made to satisfy the inequality;

$$L_{\alpha}^{\beta}(U)<H_{\alpha}^{\beta}(A;U)D^{1-\alpha}+\frac{D^{1-\alpha}-1}{2^{1-\alpha}-1} \quad (3.2.9)$$

Proof: Let l_i be the positive integer satisfying the inequality;

$$-\log\{\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\}+\log\frac{\sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A^{\alpha+\beta-1}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^{n}\left(u_i^{\beta}\left\{\mu_A^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}\right)}\leq$$

$$l_i-\log\left\{\mu_A^{\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\right\}+\log\frac{\sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A^{\alpha+\beta-1}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^{n}\left(u_i^{\beta}\left\{\mu_A^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}\right)}+1 \quad (3.2.10)$$

Consider the interval;

$$\delta_i = \left[\begin{array}{c} -\log\left(\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right) + \log\dfrac{\sum_{i=1}^n u_i^\beta \left\{\mu_A^{\alpha+\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^n \left(u_i^\beta \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}\right)}, \\ -\log\left(\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right) + \log\dfrac{\sum_{i=1}^n u_i^\beta \left\{\mu_A^{\alpha+\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^n \left(u_i^\beta \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}\right)} + 1 \end{array}\right] \tag{3.2.11}$$

of length 1. In every δ_i, there lies exactly one positive integer l_i such that

$$\begin{aligned} 0 < -\log\left(\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right) + \log\frac{\sum_{i=1}^n u_i^\beta \left\{\mu_A^{\alpha+\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^n \left(\mu_i^\beta \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}\right)} \le l_i < \\ -\log\left(\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right) + \log\frac{\sum_{i=1}^n u_i^\beta \left\{\mu_A^{\alpha+\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^n \left(u_i^\beta \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}\right)} + 1 \end{aligned} \quad (3.2.12).$$

We will first show that the sequence $l_1, l_2, \ldots, l_n$ thus defined satisfies (3.2.4). From (3.2.12), we have;

$$-\log\left(\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right) + \log\frac{\sum_{i=1}^n u_i^\beta \left\{\mu_A^{\alpha+\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^n \left(u_i^\beta \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}\right)} \le l_i$$

or

$$\frac{\left(\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right)}{\dfrac{\sum_{i=1}^n u_i^\beta \left\{\mu_A^{\alpha+\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^n \left(u_i^\beta \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}\right)}} \ge D^{-l_i}$$

Multiplying both sides by $u_i^\beta\left(\mu_A^{\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\beta-1}\right)$ and summing over $i = 1,2,\ldots,n$, we get (3.2.4).

The last inequality in (3.2.12) gives;

$$l_i < -\log\left(\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)\right)^\alpha\right) + \log\frac{\sum_{i=1}^n u_i^\beta \left\{\mu_A^{\alpha+\beta-1}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^n \left(u_i^\beta \left\{\mu_A^\beta(x_i) + \left(1-\mu_A(x_i)\right)^\beta\right\}\right)} + 1$$

or $$l_i < \log\left(\frac{\left(\mu_A^{\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\right)}{\frac{\sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A^{\alpha+\beta-1}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^{n}\left(u_i^{\beta}\left\{\mu_A^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}\right)}}\right)^{-1} D$$

or $$D^{l_i} < \left(\frac{\left(\mu_A^{\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\right)}{\frac{\sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A^{\alpha+\beta-1}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^{n}\left(u_i^{\beta}\left\{\mu_A^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}\right)}}\right)^{-1} D$$

For $0 < \alpha < 1$, raising power both sides$\frac{1-\alpha}{\alpha}$, we get;

$$D^{l_i\left(\frac{1-\alpha}{\alpha}\right)} < \left(\frac{\left(\mu_A^{\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\right)}{\frac{\sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A^{\alpha+\beta-1}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^{n}\left(u_i^{\beta}\left\{\mu_A^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}\right)}}\right)^{\frac{\alpha-1}{\alpha}} D^{\frac{1-\alpha}{\alpha}}.$$

Multiplying both sides by

$$\frac{u_i^{\beta}\left\{\mu_A^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}}{\sum_{i=1}^{n}\left(u_i^{\beta}\left\{\mu_A^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}\right)}$$

and summing over $i = 1,2,\dots,n$, we get;

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

or

$$\left(\frac{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\} D^{\left(\frac{1-\alpha}{\alpha}\right)l_i}}{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}}\right)^{\alpha} < \left(\frac{\sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A^{\alpha+\beta-1}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}}\right) D^{1-\alpha}$$

Since $2^{1-\alpha} - 1 > 0$ for $0 < \alpha < 1$ and after suitable operations, we get;

$$\frac{1}{2^{1-\alpha}-1}\left[\left(\frac{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}D^{\left(\frac{1-\alpha}{\alpha}\right)l_i}}{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}}\right)^{\alpha}-1\right]$$

$$<\frac{1}{2^{1-\alpha}-1}\left[\left(\frac{\sum_{i=1}^{n} u_i^{\beta}\left\{\mu_A^{\alpha+\beta-1}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha+\beta-1}\right\}}{\sum_{i=1}^{n} u_i{}^{\beta}\left\{\mu_A{}^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}}\right)\right]D^{1-\alpha}+\frac{D^{1-\alpha}-1}{2^{1-\alpha}-1}$$

or we can write

$L_{\alpha}^{\beta}(U) < H_{\alpha}^{\beta}(A;U)D^{1-\alpha}+\frac{D^{1-\alpha}-1}{2^{1-\alpha}-1}$.

As $D \geq 2$, we have $\frac{D^{1-\alpha}-1}{2^{1-\alpha}-1} > 1$ from which it follows that upper bound $L_{\alpha}^{\beta}(U)$ in (3.2.9) is greater than unity.Also, for $\alpha > 1$, the proof follows along the similar lines.

3.3. Some Coding Theorems on Fuzzy Entropy Function Depending Upon Parameter R and V:-

Consider a function

$$H_R\left(\mu_A^{v}(x)_i+\left(1-\mu_A(x_i)\right)^{v},U\right)=\frac{R}{R-1}\left[1-\left(\frac{\sum_{i=1}^{n}u_i\left(\mu_A^{R+v-1}(x)_i+\left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=1}^{n}u_i\left(\mu_A^{v}(x)_i+\left(1-\mu_A(x_i)\right)^{v}\right)}\right)^{\frac{1}{R}}\right] \quad (3.3.1)$$

Where

$$R>0(\neq 1), v>0, \qquad \sum_{i=1}^{n}\left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)=1, \left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)\geq 0.$$

Remark 3.3.1:

- **a)** When$v=1$, (3.3.1) reduces to the ‘useful’ R-norm fuzzy information measure corressponding to Singh, Kumar and Tuteja [92].
- **b)** When $v=1$, $u_i=1\ \forall\ i=1,2,\dots,n$, (3.3.1) reduces to the R-norm fuzzy information measure corresponding to Boekee and Lubbee [27].
- **c)** When$R\to 1$, $v=1\ and u_i=1\ \forall\ i=1,2,\dots,n$, (3.3.1) reduces to the De Luca and Termini [33] measure of fuzzy entropy corresponding to the Shannon [87] measure of entropy.

Further, consider a generalized ‘useful’ codeword mean length;

$$L_R\left(\mu_A^v(x_i)+\left(1-\mu_A(x_i)\right)^v, U\right) = \left[1-\frac{\sum_{i=1}^n u_i\left(\mu_A^v(x)_i+\left(1-\mu_A(x_i)\right)^v\right)D^{-l_i\left(\frac{R-1}{R}\right)}}{\sum_{i=1}^n u_i\left(\mu_A^v(x_i)+\left(1-\mu_A(x_i)\right)^v\right)}\right] \quad (3.3.2).$$

Where

$$R>0(\neq 1), v>0, \quad \sum_{i=1}^n\left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)=1, \left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)\geq 0. D$$

is the size of the code alphabet.

Remark 3.3.2:

(1) When $v = 1$, (3.3.2) reduces to the fuzzy 'useful' codeword mean length corresponding to the Sing, Kumar and Tuteja [92].

(2) When $v = 1,\ u_i = 1\ \forall\ i = 1,2,\dots,n$, (3.3.2) reduces to the fuzzy codeword mean length corresponding to the Boekee et al [27].

(3) when$R \to 1,\ v = 1\ and u_i = 1\ \forall\ i = 1,2,\dots,n$, (3.3.2) reduces to the fuzzy optimal codeword mean length corresponding to the Shannon [87].

We now establish a result that in a sense, gives a characterization of $H_R\left(\mu_A^v(x)_i + \left(1-\mu_A(x_i)\right)^v, U\right)$ Under the condition;

$$\sum_{i=1}^{n} u_i \left(\mu_A^{v-1}(x_i) + \left(1-\mu_A(x_i)\right)^{v-1}\right) D^{-l_i} \leq \sum_{i=1}^{n} u_i\left(\mu_A^v(x_i) + \left(1-\mu_A(x_i)\right)^v\right) \quad (3.3.3)$$

Remark 3.3.3:

When $v = 1,\ u_i = 1\ \forall\ i = 1,2,\dots,n,$ and $\sum_{i=1}^{n}\left(\mu_A(x_i) + \left(1-\mu_A(x_i)\right)\right) = 1$, (3.3.3) is a generalization of (3.1.3) which is Kraft's [65] inequality.

Theorem 3.3.1: For every code whose lengths $l_1, l_2, \dots, l_n$ satisfies (3.3.3), then the average fuzzy codeword length satisfies;

$$L_R\left(\mu_A^v(x_i) + \left(1-\mu_A(x_i)\right)^v, U\right) \geq H_R\left(\mu_A^v(x)_i + \left(1-\mu_A(x_i)\right)^v, U\right) \quad (3.3.4)$$

Equality holds iff

$$l_i = -\,log\left(\mu_A^R(x_i) + \left(1-\mu_A(x_i)\right)^R\right) + log\frac{\sum_{i=1}^{n} u_i\left(\mu_A^{R+v-1}(x_i) + \left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=1}^{n} u_i\left(\mu_A^v(x_i) + \left(1-\mu_A(x_i)\right)^v\right)} \quad (3.3.5)$$

Proof: By holder's inequality

$$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}} \quad (3.3.6)$$

$\forall\ x_i, y_i > 0, i = 1,2,\dots,n and \frac{1}{p} + \frac{1}{q} = 1, p < 1\ (\neq 0), q < 0\ or q < 1\ (\neq 0), p < 0$. We see the equality holds iff there exists a positve constant c such that;

$$x_i^p = c y_i^q \quad (3.3.7)$$

Setting

$$x_i = u_i^{\frac{R}{R-1}}\left(\mu_A^{\frac{vR}{R-1}}(x_i) + \left(1-\mu_A(x_i)\right)^{\frac{vR}{R-1}} D^{-l_i}\right);$$

$$y_i = u_i^{\frac{1}{1-R}}\left(\mu_A^{\frac{R+v-1}{1-R}}(x_i) + \left(1-\mu_A(x_i)\right)^{\frac{R+v-1}{1-R}} D^{-l_i}\right),$$

$$p = \frac{R-1}{R} and q = 1-R$$

In (3.3.6) and using (3.3.3), we get;

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

(3.3.8)

Dividing both sides of (3.3.8) by

$\left(\sum_{i=1}^{n} u_i\left(\mu_A^v(x)_i + \left(1-\mu_A(x_i)\right)^v\right)\right)^{\frac{R}{1-R}}$, we get;

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

Taking $0 < R < 1$, raising both sides to the power $\frac{1-R}{R}, R \neq 1, also \frac{R}{R-1} < 0$ for $0 < R < 1$ and after suitable operations, we obtain the result (3.3.4). For $R > 1$, the inequality (3.3.4) can be obtained in a similar fashion.

Theorem 3.3.2: For every code with lengths $l_1, l_2, \ldots, l_n$ satisfies (3.3.3). Then $L_R\left(\mu_A^v(x_i) + \left(1-\mu_A(x_i)\right)^v, U\right)$ can be made to satisfy the inequality

$$L_R\left(\mu_A^v(x_i) + \left(1-\mu_A(x_i)\right)^v, U\right)$$
$$< H_R\left(\mu_A^v(x)_i + \left(1-\mu_A(x_i)\right)^v, U\right)D^{\left(\frac{R-1}{R}\right)} + \frac{R}{R-1}\left(1-D^{\left(\frac{R-1}{R}\right)}\right) \tag{3.3.9}$$

Proof: Let l_i be the positive integer satisfying the inequality;

$$-log\left(\mu_A^R(x_i) + \left(1-\mu_A(x_i)\right)^R\right) + log\frac{\sum_{i=1}^{n} u_i\left(\mu_A^{R+v-1}(x)_i + \left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=1}^{n} u_i\left(\mu_A^v(x_i) + \left(1-\mu_A(x_i)\right)^v\right)} \leq l_i$$

$$< -\,log\left(\mu_A^R(x_i) + \left(1-\mu_A(x_i)\right)^R\right) + log\frac{\sum_{i=1}^{n} u_i\left(\mu_A^{R+v-1}(x)_i + \left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=1}^{n} u_i\left(\mu_A^{v}(x_i) + \left(1-\mu_A(x_i)\right)^{v}\right)} + 1$$

Consider the interval

$$\delta_i = \left[\begin{array}{l} -\,log\left(\mu_A^R(x_i) + \left(1-\mu_A(x_i)\right)^R\right) + log\dfrac{\sum_{i=1}^{n} u_i\left(\mu_A^{R+v-1}(x_i) + \left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=}^{n} u_i\left(\mu_A^{v}(x_i) + \left(1-\mu_A(x_i)\right)^{v}\right)}, \\ -\,log\left(\mu_A^R(x_i) + \left(1-\mu_A(x_i)\right)^R\right) + log\dfrac{\sum_{i=1}^{n} u_i\left(\mu_A^{R+v-1}(x_i) + \left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=}^{n} u_i\left(\mu_A^{v}(x_i) + \left(1-\mu_A(x_i)\right)^{v}\right)} + 1 \end{array}\right] \tag{3.3.11}$$

of length 1. In every δ_i, there lies exactly one positive integer l_i such that;

$$\begin{aligned} 0 < -\log&\left(\mu_A^R(x)_i + \left(1-\mu_A(x_i)\right)^R\right) \\ &+ log\frac{\sum_{i=1}^{n} u_i\left(\mu_A^{R+v-1}(x)_i + \left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=}^{n} u_i\left(\mu_A^{v}(x)_i + \left(1-\mu_A(x_i)\right)^{v}\right)} \le l_i \\ &< -\,log\left(\mu_A^R(x)_i + \left(1-\mu_A(x_i)\right)^R\right) \\ &+ log\frac{\sum_{i=1}^{n} u_i\left(\mu_A^{R+v-1}(x)_i + \left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=}^{n} u_i\left(\mu_A^{v}(x)_i + \left(1-\mu_A(x_i)\right)^{v}\right)} + 1 \end{aligned} \tag{3.3.12}$$

We will first show that the sequence$\{l_1, l_2, \ldots, l_n\}$, thus defined satisfies (3.3.3).

From (3.3.12), we have;

$$\begin{aligned} 0 < -\,log&\left(\mu_A^R(x_i) + \left(1-\mu_A(x_i)\right)^R\right) \\ &+ log\frac{\sum_{i=1}^{n} u_i\left(\mu_A^{R+v-1}(x)_i + \left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=}^{n} u_i\left(\mu_A^{v}(x_i) + \left(1-\mu_A(x_i)\right)^{v}\right)} \le l_i \end{aligned}$$

or

$$-\,log\frac{\mu_A^R(x)_i + \left(1-\mu_A(x_i)\right)^R}{\left(\frac{\sum_{i=1}^{n} u_i\left(\mu_A^{R+v-1}(x_i)+\left(1-\mu_A(x_i)\right)^{R+v-1}\right)}{\sum_{i=}^{n} u_i\left(\mu_A^{v}(x)_i+\left(1-\mu_A(x_i)\right)^{v}\right)}\right)} \le -\,log_D\, D^{-l_i}$$

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

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

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

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

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

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

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

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

Multiplying both sides by $\frac{u_i\left(\left(\mu_A^v(x_i)+\left(1-\mu_A(x_i)\right)^v\right)\right)}{\sum_{i=}^n u_i\left(\mu_A^v(x_i)+\left(1-\mu_A(x_i)\right)^v\right)}$ 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 $c_1, c_2, \dots, c_k$ with lengths $n_1, n_2, \dots, n_k$ respectively. Without loss of generality it may be assumed that $n_1 \le n_2 \le \cdots \le n_k$.

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 α, which depends on the noise characteristics of the channel. This mapping α is called the error admissibility mapping. Given codeword 'c' and error admissibility α, 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 α. In particular, α may be defined as (Bernard & Sharma [23])

$$\alpha_e(c) = \{\underline{u} | w(c - \underline{u}) \le e\},$$

Where e is the random substitution error and $w(c - \underline{u})$ is the Humming weight, i.e. the number of non-zero coordinates of $(c - \underline{u})$. 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 cord word c.

We may assume that α_0 corresponds to the noiseless. In other words, if c is sent then c is received w.r.t. α_0. Moreover it is clear that $|\alpha_e(c)|$ depends only on the length n of c when α 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 α-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 α 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 α andn_i. It is given that;

$$|r_\alpha|_{n_i} = |\alpha|_{n_1}|\alpha|_{n_2} \dots |\alpha|_{n_i - n_{i-1}}.$$

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 $n_1 \leq n_2 \leq \cdots \leq n_k$

$$|r_\alpha|_{n_i} = |r_\alpha|_{n_{i-1}}.|r_\alpha|_{n_i - n_{i-1}}$$

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

We have

$$|r_\alpha|_{n_i} = |\alpha|_{n_1}.|\alpha|_{n_2 - n_1} \dots |\alpha|_{n_i - n_{i-1}}$$

and $$|r_\alpha|_{n_{i-1}} = |\alpha|_{n_1}.|\alpha|_{n_2 - 1} \dots |\alpha|_{n_{i-1} - n_{i-2}}$$

Therefore $$|r_\alpha|_{n_i} = |r_\alpha|_{n_{i-1}}.|r_\alpha|_{n_i - n_{i-1}}$$

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

$$\sum_{i=1}^{k} |r_\alpha|_{n_i} D^{-n_i} \leq 1 \quad (4.1.1)$$

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 α-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 $\left[\frac{D}{|r_\alpha|_1} - N_1\right]$ ways.

The number of symbols that may be added at the second position is at most $\frac{D}{|\alpha|_1}$ which is equivalent to $D\frac{|r_\alpha|_1}{|r_\alpha|_2}$ from Lemma 4.1.1. Thus, we will have;

$$N_2 \leq \left[\frac{D}{|r_\alpha|_1} - N_1\right]\left[D.\frac{|r_\alpha|_1}{|r_\alpha|_2}\right]$$

$$= \frac{D^2}{|r_\alpha|_2} - N_1.D\frac{|r_\alpha|_1}{|r_\alpha|_2}$$

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;

$$\frac{D^2}{|r_\alpha|_2} - N_1.D\frac{|r_\alpha|_1}{|r_\alpha|_2} - N_2.$$

Once again, the number of symbols that may be added in the third position is $\frac{D}{|\alpha|_1}$. From Lemma 4.1.1, we can take $\frac{D}{|\alpha|_1} = D\frac{|r_\alpha|_2}{|r_\alpha|_3}$.

Thus, $$N_3 \leq \left[\frac{D^2}{|r_\alpha|_2} - N_1 D\frac{|r_\alpha|_1}{|r_\alpha|_2} - N_2\right]\left[D\frac{|r_\alpha|_2}{|r_\alpha|_3}\right]$$

$$= \frac{D^3}{|r_\alpha|_3} - N_1 D^2\frac{|r_\alpha|_1}{|r_\alpha|_3} - N_2 D\frac{|r_\alpha|_2}{|r_\alpha|_3}$$

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

$$N_{n_k} \leq \frac{D^{n_k}}{|r_\alpha|_{n_k}} - N_1 D^{n_{k-1}}\frac{|r_\alpha|_1}{|r_\alpha|_{n_k}} - N_2 D^{n_{k-2}}\frac{|r_\alpha|_2}{|r_\alpha|_{n_k}} \ldots N_{n_{k-1}} D\frac{|r_\alpha|_{n_{k-1}}}{|r_\alpha|_{n_k}}.$$

This can be written as $\sum_{i=1}^{k} |r_\alpha|_i \, N_i D^{-i} \le 1$.

Changing the summation from the length $1, 2, \ldots, n_k$ to the code word length$n_1, n_{2,\ldots,} n_k$. The above inequality can be equivalently put as$\sum_{i=1}^{k} |r_\alpha|_{n_i} \, D^{-n_i} \le 1$, 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;

$$L(t) = {}^{1}/_{t} \, log_D \sum_{i=1}^{k} (p_i D^{tn_i}) \, ; \; (0 < t < \infty) \quad (4.1.2)$$

An application of L-Hospitals rule shows that

$$L(0) = \lim_{t\to 0} L(t) = \sum_{i=1}^{k} n_i p_i \qquad (4.1.3)$$

For large$t$, $\sum_{i=1}^{k} p_i D^{tn_i} \cong p_j D^{tn_j}$, where n_j is the largest of the number$n_1, n_2, \ldots, n_k$. 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 elementeither 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

$$\mu_A(x_i) = \begin{cases} 0, \quad if x \notin A \; and \; there \; is \; noambiguity \\ 1, \quad if x \in A \; and \; there \; is \; no \; ambiguity \\ 0.5, if \; maximum \; ambiguity, i.e. x \in A \; or \; x \notin A \end{cases}$$

Let $A = \{x_i: 0 < \mu_A(x_i) < 1, \forall i = 1, 2, \dots, n\}$

$$B = \{x_i: 0 < \mu_B(x_i) < 1, \forall i = 1, 2, \dots, n\}$$

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

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.

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

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 lengthn_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;

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

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

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

Remark 4.2.2: Moreover, if $\mu_A(x_i) + \left(1 - \mu_A(x_i)\right) = \mu_B(x_i) + \left(1 - \mu_B(x_i)\right)$ 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 $S_1, S_2, \dots, S_k$ with message degree of membership $\mu_A(x_1),\ \mu_A(x_2), \dots,\ \mu_A(x_k);\ \mu_A(x_i) \geq 0$. Let an α-prompt code encode these messages into a code alphabet of D symbols and let the length of the code word corresponding to the messages$S_i$ben_i. Then the code length of order$t$, $L(t)$, shall satisfy the inequality;

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

Proof: In the Holder's inequality

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

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

$$p = -t, q = 1 - \beta,\ \ x = \left(\mu_A^{\frac{-1}{t}}(x_i) + \left(1 - \mu_A(x_i)\right)^{\frac{-1}{t}}\right) D^{-n_i}$$

and $$y_i = \left(\mu_A^{\frac{1}{t}}(x_i) + \left(1 - \mu_A(x_i)\right)^{\frac{1}{t}}\right) |r_\alpha|_{n_i},$$

we get;

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

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

or

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

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

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

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

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

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

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

Which gives;

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

or

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

Hence

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

The quantity

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

is similar to fuzzy entropy corresponding to Renyi's entropy of order β [84].

It can be easily verified that the quantity in (4.2.4) hold if and only if;

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

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].

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

b) For noiseless channel, $|r_\alpha|_{n_{i-1}} \forall i$, 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 β.

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 $\overline{n} \geq H(A)$, where $H(A)$ is the fuzzy entropy corresponding to the Shannon's entropy.

Theorem 4.2.2: Let an α-prompt code encode the K messages $S_1, S_2, \dots, S_k$ 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.

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

With equality if and only if;

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

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

where $L(t) = \frac{1}{t} log_D \sum_{i=1}^{k}\left(\mu_A(x_i) + \left(1-\mu_A(x_i)\right)\right) D^{tn_i}$.

Proof: In the Holder's inequality

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

With the equality if and only if.

$x_i^p = c y_i^q$,where c is a positive number, $1/p + 1/q = 1$ and $p < 1$. We note that direction of Holder's inequality is the reverse of the usual one as $p < 1$ (Beckenbach and Bellman [21]). Substituting.

$$p = -t, q = t\beta, \qquad x_i = \left(\mu_A^{\frac{-1}{t}}(x_i) + \left(1 - \mu_A(x_i)\right)^{\frac{-1}{t}}\right) D^{-n_i}$$

and $$y_i = \left(\mu_A^{\frac{1}{t\beta}}(x_i) + \left(1 - \mu_A(x_i)\right)^{\frac{1}{t\beta}}\right)\left(\mu_B^{-1}(x_i) + \left(1 - \mu_B(x_i)\right)^{-1}\right) |r_\alpha|_{n_i},$$

We get;

$$\left(\sum_{i=1}^{k}\left(\mu_A(x_i) + \left(1 - \mu_A(x_i)\right)\right) D^{tn_i}\right)^{\frac{-1}{t}}$$

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

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

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

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

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

this gives $\left(\sum_{i=1}^{k}\left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)D^{tn_i}\right)^{\frac{1}{t}} \geq$

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

or

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

Hence,

$$L(t) \geq \frac{1}{1-\beta}log_D\left(\sum_{i=1}^{k}\left\{\left(\mu_A(\mathrm{x}_i)+\left(1-\mu_A(x_i)\right)\right)\left(\mu_B^{\beta-1}(x_i)+\left(1-\mu_B(x_i)\right)^{\beta-1}\right)\left(|r_\alpha|_{n_i}\right)^{1-\beta}\right\}\right)$$

The quantity;

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

is equivalent to fuzzy inaccuracy corresponding to Nath's inaccuracy [77] of order β.

Particular Cases:

For $t = 0\ and\beta \to 1$, the inequality (4.2.7) reduces to;

$$\bar{n} \geq \sum_{i=1}^{k}\left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)log_D\left(\frac{|r_\alpha|_{n_i}}{\left(\mu_B(x_i)+\left(1-\mu_B(x_i)\right)\right)}\right) \quad (4.2.8)$$

For noiseless channel, $\left(|r_\alpha|_{n_i}\right) = 1;\ \forall i$, the inequality (4.2.8) reduces to;

$$\bar{n} \geq \sum_{i=1}^{k}\left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)log_D\left(\mu_B(x_i)+\left(1-\mu_B(x_i)\right)\right)$$

$= H\big(\mu_A(x_i),(x_i)\big)$ (4.2.9).

Where$H\big(\mu_A(x_i),(x_i)\big)$ 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 $\left(|r_\alpha|_{n_i}\right) = 1;\ \forall i$, the inequality (4.2.7) reduces to fuzzy inequality corresponding to Autar and Soni [8].

$$L(t) \geq H_\beta\big(\mu_A(x_i), \mu_B(x_i)\big) \tag{4.2.10}$$

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

4.3 $\boldsymbol{\beta}$- measure of Uncertainty Involving Utilities:-

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

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

$$\beta > 0(\neq 1) \tag{4.3.1}$$

Which is β-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 $\beta \to 1$, 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;

$$L(U^\beta) = \frac{\left[\sum_{i=1}^k u_i\left(\mu_A(x_i) + \big(1-\mu_A(x_i)\big)\right) D^{(\beta^{-1}-1)_{n_i}}\right]^\beta - 1}{1-2^{1-\beta}} \tag{4.3.2}$$

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 $\beta \to 1$, it reduces to fuzzy mean code word length corresponding to Shannon [87] measure and gave a characterization of $HU_K^{\beta}(A;U)$ under the condition.

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

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

$$L\left(U^{\beta}\right) \ge H_k^{\beta}(A,U) \qquad (4.3.4)$$

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

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

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

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

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

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

$$x_i^p = c y_i^q \qquad (4.3.7)$$

We substitute

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

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

$p = (1-\beta^{-1}) and q = 1-\beta$, we get;

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

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

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

Using the inequality (4.3.3), the above inequality can be written as;

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

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

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

$$\geq\frac{\left[\sum_{i=1}^{k}\left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)\left(\frac{u_i}{\sum_{i=1}^{k}u_i\left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)}\right)^{1-\beta}\right]-1}{1-2^{1-\beta}}$$

Hence, $\qquad L\left(U^{\beta}\right)\geq H_k^{\beta}(A,U).$

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

$$H_k^{\beta}(A,U)\leq L\left(U^{\beta}\right)\leq D.H_k^{\beta}(A,U)+\frac{D-1}{1-2^{1-\beta}}\quad(4.3.8)$$

Proof: Suppose

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

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

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

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

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

$$< D\left[\sum_{i=1}^{k}\left(\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\right)D^{(\beta^{-1}-1)\tilde{n}_i}\right]^{\beta} \qquad (4.3.10)$$

We know

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

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

Hence, (4.3.10) can be expressed as;

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

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

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

Thus, $\quad H_k^{\beta}(A,U) \leq L\left(U^{\beta}\right) \leq D.\,H_k^{\beta}(A,U)+\frac{D-1}{1-2^{1-\beta}}$

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:

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

If the distribution Q is uniform. This becomes;

$$I(P;Q) = \sum_{i=1}^{n} p_i \log p_i + \log n \qquad (4.4.2)$$

Where, $P, Q \in T_n$ and

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

Since Shannon's Entropy

$$H(P) = \sum_{i=1}^{n} p_i \log p_i \qquad (4.4.3)$$

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;

$$= -\sum_{i=1}^{n} p_i \log q_i \qquad (4.4.5)$$

Considering Kerridge [62] inaccuracy, we can interpret Kullback-Leibler [66] measure of divergence.

$I(P;Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$, as i.e

$=$difference of Kerridge inaccuracy and Shannon's entropy

$$= \sum_{i=1}^{n} \{-p_i \log q_i - (-p_i \log p_i)\} \tag{4.4.6}$$

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 $A = \{x_i : \mu_A(x_i), \forall i = 1,2,\dots,n\}\ and\ B = \{x_i : \mu_B(x_i), \forall i = 1,2,\dots,n\}$,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 :

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

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;

$$D_\alpha = \big(\mu_A(x_i); \mu_B(x_i)\big)$$

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

and its corresponding fuzzy average code word length as;

$$L_\alpha = \big(\mu_A(x_i); \mu_B(x_i)\big)$$

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

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 $\alpha \to 1$ (4.4.9) tends to average codeword length given as;

$$L = \sum_{i=1}^{n} p_i q_i n_i \tag{4.4.10}$$

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

$$L = \sum_{i=1}^{n} p_i n_i \tag{4.4.11}$$

4.5 Noiseless directed divergence Coding Theorems:-

Theorem 4.5.1: For all uniquely decipherable codes

$$D_\alpha \leq L_\alpha \qquad (4.5..1)$$

where

$$L_\alpha = \left(\mu_A(x_i); \mu_B(x_i)\right)$$

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

Proof:-By Holders inequality, we have;

$$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}}; \quad 0 < p < 1, q < 0 \; or \; 0 < q < 1, p < 0 \tag{4.5.2}$$

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

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

Thus equation (4.5.2) becomes;

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

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

Using Kraft's inequality, we have

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

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

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

dividing both sides by t, we get;

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

Subtracting n from both sides, we have

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

Taking $\alpha = t+1,\ t = \alpha - 1$

and

$$f(\mu_A(x_i),\mu_B(x_i)) = \left\{\left(\mu_A(x_i)^{\alpha} + \left(1-\mu_A(x_i)\right)^{\alpha}\right)\left(\mu_B(x_i)^{\alpha} + \left(1-\mu_B(x_i)\right)^{\alpha}\right)\right\},$$

equation (4.5.4) becomes;

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

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

(4.5.5)

Dividing both sides byα, we get;

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

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

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

Theorem 4.5.2:- For all uniquely decipherable codes,

$$D_{\alpha,\beta} \leq L_{\alpha,\beta} \tag{4.5.6}$$

$$L_{\alpha,\beta} = \frac{1}{\beta-\alpha}\sum_{i=1}^{n}\left[\begin{array}{c}\left\{\left(\mu_A(x_i)^\alpha+\left(1-\mu_A(x_i)\right)^\alpha\right)\left(\mu_B(x_i)^\alpha+\left(1-\mu_B(x_i)\right)^\alpha\right)\right\}D^{n_i(\alpha-1)}\\ -\left\{\left(\mu_A(x_i)^\beta+\left(1-\mu_A(x_i)\right)^\beta\right)\left(\mu_B(x_i)^\beta+\left(1-\mu_B(x_i)\right)^\beta\right)\right\}D^{n_i(\beta-1)}\end{array}\right] \tag{4.5.7}$$

Where either $\alpha \geq 1, \beta \leq 1 \; or \; \alpha \leq 1, \beta \geq 1$

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

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

$$\frac{\sum_{i=1}^{n}\left[\left\{\left(\mu_A(x_i)^\alpha+\left(1-\mu_A(x_i)\right)^\alpha\right)\left(\mu_B(x_i)^\alpha+\left(1-\mu_B(x_i)\right)^\alpha\right)\right\}D^{n_i(\alpha-1)}-1\right]}{\alpha-1} \tag{4.5.8}$$

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

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

$$\sum_{i=1}^{n}\left[\left\{\left(\mu_A(x_i)^\alpha+\left(1-\mu_A(x_i)\right)^\alpha\right)\left(\mu_B(x_i)^\alpha+\left(1-\mu_B(x_i)\right)^\alpha\right)\right\}D^{n_i(\alpha-1)}-1\right] \tag{4.5.9}$$

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

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

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

(4.5.10)

Subtracting (4.5.10) from (4.5.9), and dividing both sides by $(\beta-\alpha)$, we have;

$$\sum_{i=1}^{n}\left[\left\{\left(\mu_A(x_i)^{\alpha}+\left(1-\mu_A(x_i)\right)^{\alpha}\right)\left(\mu_B(x_i)^{\alpha}+\left(1-\mu_B(x_i)\right)^{\alpha}\right)\right\}\right.$$
$$\left.-\left\{\left(\mu_A(x_i)^{\beta}+\left(1-\mu_A(x_i)\right)^{\beta}\right)\left(\mu_B(x_i)^{\alpha}+\left(1-\mu_B(x_i)\right)^{\beta}\right)\right\}\right]\leq$$
$$\sum_{i=1}^{n}\left[\begin{array}{l}\left\{\left(\mu_A(x_i)^{\alpha}+\left(1-\mu_A(x_i)\right)^{\alpha}\right)\left(\mu_B(x_i)^{\alpha}+\left(1-\mu_B(x_i)\right)^{\alpha}\right)\right\}D^{n_i(\alpha-1)}-\\ \left\{\left(\mu_A(x_i)^{\beta}+\left(1-\mu_A(x_i)\right)^{\beta}\right)\left(\mu_B(x_i)^{\alpha}+\left(1-\mu_B(x_i)\right)^{\beta}\right)\right\}D^{n_i(\beta-1)}\end{array}\right]$$

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

Theorem 4.5.3:

For all uniquely decipherable codes,

$$D'_{\alpha,\beta}\ \leq L'_{\alpha,\beta} \qquad (4.5.11)$$

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

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

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

Lemma 4.5.1: For all uniquely decipherable codes

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

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

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

Taking logarithm on both sides, we have;

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

Taking $\alpha=t+1,\ t=\alpha-1$

and $\quad f\big(\mu_A(x_i),\mu_B(x_i)\big)=\left\{\left(\mu_A(x_i)^\alpha+\left(1-\mu_A(x_i)\right)^\alpha\right)\left(\mu_B(x_i)^\alpha+\left(1-\mu_B(x_i)\right)^\alpha\right)\right\}$,

we have $\quad \log_D\left[\sum_{i=1}^n\left\{\left(\mu_A(x_i)^\alpha+\left(1-\mu_A(x_i)\right)^\alpha\right)\left(\mu_B(x_i)^\alpha+\left(1-\mu_B(x_i)\right)^\alpha\right)\right\}\right]$

$$\leq \log_D\left[\sum_{i=1}^n\left\{\left(\mu_A(x_i)^\alpha+\left(1-\mu_A(x_i)\right)^\alpha\right)\left(\mu_B(x_i)^\alpha+\left(1-\mu_B(x_i)\right)^\alpha\right)\right\}D^{n_i(\alpha-1)}\right] \tag{4.5.14}$$

Which proves the Lemma.

Proof of the Theorem 4.5.3: Changing $\alpha\ to\ \beta$ in (4.5.14), we get

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

$$\leq \log_D\left[\sum_{i=1}^n\left\{\left(\mu_A(x_i)^\beta+\left(1-\mu_A(x_i)\right)^\beta\right)\left(\mu_B(x_i)^\beta+\left(1-\mu_B(x_i)\right)^\beta\right)\right\}D^{n_i(\beta-1)}\right] \tag{4.5.15}$$

subtracting (4.5.15) from (4.5.14), we have;

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

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

Dividing both sides by $\beta-\alpha$, we have;

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

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

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

The reliable engineering is one of the important engineering tasks in design and development of technical system. The conventional reliability of a system is defined as the probability that the system performs its assigned function properly during a predefined period under the condition that the system behavior can be fully characterized in the context of probability measures. The reliability of a system can be determined on the basis of tests or the acquisition of operational data. However, due to the uncertainty and inaccuracy of this data, the estimation of precise values of probabilities is very difficult in many systems. (e.g., power system, electrical machine, hardware etc., Hammer [46]. For this reason the fuzzy reliability concept has been introduced and formulated in the context of fuzzy measures. The basis for this approach is constituted by the fundamental works on fuzzy set theory of Zadeh [108], Dubois and Prade [34] and other.

5.1 Introduction:-

Let x denote the set of integers between 0 and 10, both inclusive; that is,

$$X = \{0, 1, 2, \ldots., 10\}.$$

Suppose that we are interested in a subset $\tilde{A}\ of\ X$, where $\tilde{A}$ contains all of the “medium” integers of X. Thus

$$\tilde{A} = \{x; x \in X\ and\ x\ is\ "medium"\}.$$

Clearly, to be able to specify $\tilde{A}$, we must be precise as to what we mean by a medium integer; that is, we must be able to operationalize the term “medium integer.”Whereas most would agree that 5 is a medium integer, what is the disposition of an integer like 7? Is 7 a medium integer, or is it a large integer? Our uncertainty (or vagueness) about classifying 7 as a member of the subset $\tilde{A}$makes $\tilde{A}$a fuzzy set. The uncertainty of classification arises because the boundaries of $\tilde{A}$are not sharp. The subset $\tilde{A}$rejects the law of the excluded middle, because an integer like 7 can simultaneously belong to and not belong to $\tilde{A}$.

Membership functions were introduced as a way of dealing with the foregoing form of uncertainty of classification. Specifically, the number $\mu_{\tilde{A}}(x)$which lies between 0 and 1, reflects an assessor's view of the extent to which $x \in \tilde{A}$. As a function of x, $\mu_{\tilde{A}}(x)$is known as the *membership function* of set $\tilde{A}$. Clearly, the membership function is subjective, because it is specific to an individual assessor or a group of assessors. We also assume that for each $x \in X$, the assessor is able to assign an $\mu_{\tilde{A}}(x)$, and that this can be done for all subsets of the type $\tilde{A}$that are of interest.

If $\mu_{\tilde{A}}(x)$= 1 (or 0) for all $x \in X$, then $\tilde{A}$is the usual well defined sharp (or crisp) set. Thus the notion of fuzzy sets incorporates that of crisp sets as a special case, and because it is on crisp sets that probability measures have been defined.

Let $\mu_A(T)$ be a membership function representing the component failure time with failure distribution $\pi_A(t) = P(\mu_A(T) \leq \mathrm{t})$ and survival function $\bar{\pi}_A(t) = 1 - \pi_A(t)$. We shall assume that the component is functioning at $t = 0$ and it will fail at some $t > 0$, so that $\bar{\pi}_A(0) = 1$, and differentiability of $\pi_A(t)$and shall let $f(t) = \pi'_A(t)$ denote its failure density function. The conventional approach to characterize the failure distribution, $\pi_A(t)$, of a component is either by its hazard rate function $\lambda_A(t) = \frac{f(t)}{\pi_A(t)}$for $t < t^*$, where $t^* = sup\{t: \bar{\pi}_A(t) > 0\}$ or by the mean residual lifetime function.

$$\delta_A(t) = E(\mu_A(T) - \mathrm{t}/\mu_A(T) > t) = \begin{Bmatrix} \dfrac{\int_t^\infty \bar{\pi}_A(x)d(x)}{\bar{\pi}_A(t)}, & for\ t < t^* \\ 0, & otherwise. \end{Bmatrix}$$

It is known that each of the function $\bar{\pi}_A, \lambda_{\pi_A}\ and\ \delta_{\pi_A}$ uniquely determines the other two. More specifically,

$$\bar{\pi}_A(t) = exp\left(-\int_0^t \lambda_{\pi_A}(u)du,\ t < t^*\right) \qquad (5.1.1)$$

$$\bar{\pi}_A(t) = \frac{\delta_{\pi_A}(0)}{\delta_{\pi_A}(t)} exp\left(-\int_0^t \frac{1}{\delta_{\pi_A}(x)} dx\right), t < t^* \qquad (5.1.2)$$

$$\lambda_{\pi_A}(t) = \frac{\delta'_{\pi_A}(t) + 1}{\delta_{\pi_A}(t)}, \quad t < t^* \qquad (5.1.3)$$

And $$\delta_{\pi_A}(t) = \int_t^{t^*} exp\left(-\int_t^y \lambda(u)du\right)dy,\ t < t^* \qquad (5.1.4)$$

$\lambda_{\pi_A}(t)$and$\delta_{\pi_A}(t)$ can be used in engineering to describe aging of a component.

In section 5.2, a direct approach to measure fuzziness in the residual life time distribution is proposed. It should be emphasized that our goal in this section is not to come up with a new measure, but try to modify slightly the existing measures in such a way that can be used in the area of "Fuzzy Reliability". The proposed measure gives an alternative characterization of a failure distribution.

5.2 Residual Life Time Distribution:-

The basic uncertainty measure for distribution F is differential entropy

$$H(f) = -\int_0^\infty f(x)\, logf(x)dx = -E(logf(T); \tag{5.2.1}$$

The corresponding fuzzy measure of uncertainty for fuzzy set distribution A is fuzzy differential entropy

$$H(A) = -\int_S \mu_A(x) log\mu_A(x)dP(x) = -E\big(log\mu_A(T)\big)$$

$H(f)$is commonly referred to as the Shannon information measure.[87], and $H(A)$ is referred to as fuzzy information measure. Intuitively speaking $H(A)$ gives expected fuzzy measure of uncertainty contained in $f(t)$ about the predictability of an outcome of $\mu_A(T)$.

Frequently, in survival analysis and in life testing one has information about the current age of component under consideration. In such cases, the age must be taken into account when measuring uncertainty. Obliviously, the measure $H(f)$ in (5.2.1) is unsuitable in such situations and must be modified to take the age into account. A more reliable approach which makes use of the age is given below. Given that a component has served up to time t, we propose to measure fuzzy uncertainty about $\mu_A(T)$, lifetime component, at time t, by

$$H(A;t) = -\int_t^\infty \frac{\mu_A(x)}{\bar{\pi}_A(t)} log \frac{\mu_A(x)}{\bar{\pi}_A(t)} d = -\frac{1}{\bar{\pi}_A(t)}\int_t^\infty \mu_A(x)\log\mu_A(x)\ \mathrm{dx} + \log\bar{\pi}_A(t) \tag{5.2.2}$$

$$= 1 - \frac{1}{\bar{\pi}_A(t)} \int_t^{\infty} \left(log\lambda_{\pi_A}(x)\right) \mu_A(x)dx$$

A natural question to ask is whether $H(A;\, t)$, like δ_A and λ_A characterize $\bar{\pi}(A)$ and consequently $\pi(A)$, where

$$\pi(A) = \int_{\mathcal{S}} \mu_A(T)dP(x) = E[\mu_A(T)].$$

is c.d.f. of fuzzy set A. We explore whether $H(A;\, t)$ characterizes $\pi(A)$, that is for $H(A;\, t)$ does there exists two distinct survival distributions $\pi_{A_1}(x)$ and $\pi_{A_2}(x)$ with corresponding fuzzy density function $\mu_{A_1}(x)$ and $\mu_{A_2}(x)$ respectively, such that for all $t \geq 0$,

$$H\left(\mu_{A_1}(x), t\right) = H\left(\mu_{A_2}(x), t\right).$$

Using (5.1.1) – (5.1.4), we are able to prove the following theorem.

Theorem 5.2.1: Let $\mu_A(T)$ be a membership function corresponding to a fuzzy set A with density function f and with $H(f,t) < \infty, t \geq 0$. Here f is assumed to be continuous. Then $H(f,t)$ uniquely determines $\overline{\pi}(A)$.

Proof: Suppose that $f_1\ and\ f_2$ are density functions with

$$H(f_1, t) = H(f_2, t), t \geq 0 \tag{5.2.3}$$

And both are finite. Using the equation (5.2.2), we get

$$H^{/}(f_i,;t) = \lambda_{\pi_{A_i}}(t) \log \lambda_{A_i}(t) - \frac{\lambda_{\pi_{A_i}}(t)}{\bar{\pi}_{A_i}(t)} \int_t^{\infty} \left(\log\lambda_{\pi_{A_i}}(x)\right) f_i(x)dx$$

$$= \lambda_{\pi_{A_i}}(t) \left[\frac{1}{\bar{\pi}_{A_i}(t)} \int_t^{\infty} \left(\log \lambda_{\pi_{A_i}}(x)\right) f_i(x)dx + \log \lambda_{\pi_{A_i}}(t)\right] \tag{5.2.4}$$

$$= \lambda_{\pi_{A_i}}(t) \left[H(f_i, t) - 1 + \log \lambda_{\pi_{A_i}}(t)\right], \quad i = 1, 2.$$

It follows from (5.2.3) that $H^{/}(f_1,;t) = H^{/}(f_2,;t)$ for all $t \geq 0$, and consequently applying the equation (5.2.4), we have

$$\lambda_{\pi_{A_1}}(t) \left[H(f_1, t) - 1 + \log \lambda_{\pi_{A_1}}(t)\right] = \lambda_{\pi_{A_2}}(t) \left[H(f_2, t) - 1 + \log \lambda_{\pi_{A_2}}(t)\right];$$

(5.2.5)

for all $t \geq 0$.

To prove that (5.2.3) implies $\bar{\pi}_{A_1}(x) = \bar{\pi}_{A_2}(x)$ we need to show that $\lambda_{\pi_{A_1}}(t) = \lambda_{\pi_{A_2}}(t)$ for all $t \geq 0$. (Note that from the equation (5.1.1), the hazard function uniquely determines the survival function). Therefore, it is equivalent to show that (5.2.3), implies $\lambda_{\pi_{A_1}}(t) = \lambda_{\pi_{A_2}}(t)$ for all $t \geq 0$. Upon introducing

$$B = \left\{t: t \geq 0, and\ \lambda_{\pi_{A_1}}(t) \neq \lambda_{\pi_{A_2}}(t)\right\}, \quad (5.2.6)$$

We assume that set B is not empty. Because, if the set $B$is empty, we will have $\lambda_{\pi_{A_1}}(t) = \lambda_{\pi_{A_2}}(t)$ for all $t \geq 0$. And the proof will be complete. If $t_0 \in B$, then $\lambda_{\pi_{A_1}}(t_0) \neq \lambda_{\pi_{A_2}}(t_0)$. Without loss of generality suppose that

$$\lambda_{\pi_{A_1}}(t_0) > \lambda_{\pi_{A_2}}(t_0). \quad (5.2.7)$$

From the equation (5.2.5), therefore we must have either

$$H(f_1, t_0) - 1 + \log \lambda_{\pi_{A_1}}(t_0) < H(f_2, t_0) - 1 + \log \lambda_{\pi_{A_2}}(t_0), \quad (5.2.8)$$

or

$$H(f_1, t_0) - 1 + \log \lambda_{\pi_{A_1}}(t_0) = H(f_2, t_0) - 1 + \log \lambda_{\pi_{A_2}}(t_0) = 0. \quad (5.2.9)$$

Suppose (5.2.8) holds. Using (5.2.3) the inequality (5.2.8) reduces to $\lambda_{\pi_{A_1}}(t_0) < \lambda_{\pi_{A_2}}(t_0)$.If (5.2.9) holds, from the equation (5.2.3), it reduces to $\lambda_{\pi_{A_1}}(t) = \lambda_{\pi_{A_2}}(t)$. Combining these two we get

$$\lambda_{\pi_{A_1}}(t) \leq \lambda_{\pi_{A_2}}(t).$$

This contradicts the equation (5.2.7) and therefore the earlier assumption that B is not empty. Consequently, B is the empty set and this concludes the proof.

The following theorem gives a bound for $H(f, t)\ in\ terms\ of\ \delta_{\pi_A}(t)$.

Theorem 5.2.2: Suppose$\delta_{\pi_A}(t) < \infty$, then

$$H(f, t) \leq 1 + \log \delta_{\pi_A}(t) \quad (5.2.10)$$

Proof: For a given t, let$Y_t^d = Y/Y > t, d$ stands for distribution, and let $g_t(y)$ denote probability density function of fuzzy event with density function Y_t. Then,

$$g_t(y) = \frac{d}{dy}P(Y_t \le y) = \frac{d}{dy}P(Y_t \le y/\, Y > t) = \begin{cases} \frac{d}{dy}\frac{\pi_A(y)}{\bar{\pi}_A(t)} & if \quad y > t \\ 0 & if \quad y \le t \end{cases}$$

$$= \begin{cases} \frac{f(y)}{\bar{\pi}_A(t)} & if\ y > t \\ 0, & otherwise \end{cases}$$

It is clear that $H(f;t) = -\int_t^\infty g_t(x)\log g_t(x)\ \ and\ \int_t^\infty x g_t(x)dx = \delta_{\pi_A}(t) + t.$

Define, $Z_t = Y_t - t$, then the probability density function of fuzzy events with density function of Z_t is $h_t(\eta)$, where $h_t(\eta) = g_t(\eta + t)$ and $E(Z_t) = \delta_{\pi_A}(t)$. Now, the fuzzy entropy of Z_t is

$$-\int_0^\infty h_t(\eta)\log h_t(\eta)d\eta$$

$$= -\int_0^\infty g_t(\eta + t)\log g_t(\eta + t)d\eta$$

$$= -\int_t^\infty g_t(\eta)\log g_t(\eta)d\eta = H(f;t).$$

Given $\delta_{\pi_A}(t)$, if the domain is limited to a half line, the maximum entropy occurs when we have an exponential with mean $\delta_{\pi_A}(t)$. Therefore

$$H(f;t) = -\int_0^\infty h_t(\eta)\log h_t(\eta)d\eta \le 1 + \log\left(\delta_{\pi_A}(t)\right).$$

This completes the proof.

From the theorem 5.2.2, it is clear that the fitness of $H(f;t)$ is guaranteed whenever

$$\delta_{\pi_A}(t) < \infty.$$

5.3 New Class of Fuzzy Life Distributions:-

In this section we propose two new class of fuzzy life distributions based on the notion of fuzziness of residual lifetime described in the previous section. We should mention that, almost all existing classes of life distributions in the literature are based on the notion of aging. Throughout this section decreasing means non-increasing and increasing means non-decreasing.

Definition 5.3.1:$\bar{\pi}_A$ has decreasing (increasing) fuzzy uncertainty of residual life DFURL (IFURL) if $H(f;t)$ is decreasing (increasing) in $t, t \geq 0$.

Intuitively speaking if the component has survival function that belongs to the class of DFURL, then as the component ages the conditional fuzzy probability density function becomes more informative.

I. In definition 5.3.1, $\bar{\pi}_A$ DFURL (IFURL) if,

$$L(f\,;\,t) = -H^{/}(f;t) \geq 0(L(f;t) \leq 0).$$

that is, if we have non-negative (non-positive) local reduction of uncertainty, then $\bar{\pi}_A$ has DFURL (IFURL);

II. Suppose $\bar{\pi}_A$ is both DFURL and IFURL, then;

$$\overline{\pi}_A(t) \log \lambda_{\pi_A}(t) - \int_t^{\infty} f(x) \log \lambda_{\pi_A}(x) dx = 0.$$

and therefore, $\lambda^{/}_{\pi_A}(t) = 0$. That is, $\lambda_{\pi_A}(t) = \lambda$and $\bar{\pi}_A(t) = exp(-\lambda t)$, where λ is some positive constant. This means that exponential distribution is the only distribution which is both DFURL and IFURL. It should be mentioned that many characterizations of exponential distribution have proposed in the literature.

The following theorem gives the relationship between our class and increasing (decreasing) failure rate class of life distributions.

Theorem 5.3.1: If $\bar{\pi}_A$ is an increasing (decreasing) failure rate, IFR (DFR), then it is also a DFURL (IFURL). ($\bar{\pi}_A$is said to be an IFR (DFR) if $\lambda_{\pi_A}(t)$ is increasing (decreasing) in t).

Proof: We will prove it for IFR. Similar arguments can be used for DFR.

Suppose $\lambda_{\bar{\pi}_A}$ is an IFR, then for $t \geq 0$.

$$H^{/}(f;t) = \lambda_{\pi_A}(t)\log\lambda_{\pi_A}(t) - \frac{\lambda_{\pi_A}(t)}{\bar{\pi}_A(t)}\int_t^{\infty}\left(\log\lambda_{\pi_A}(x)\right)f(x)dx.$$

$$\leq \lambda_{\pi_A}(t)\log\lambda_{\pi_A}(t) - \lambda_{\pi_A}(t)\log\lambda_{\pi_A}(t) = 0 \qquad (5.3.1)$$

From (5.3.1) we get that $H(f;t)$ is decreasing in t. That is, $\bar{\pi}_A$ DFURL.

Another class of life distributions is the class of increasing failure rate in average (IFRA). The following example shows that there is no relationship between our class and this class of life distribution. ($\bar{\pi}_A$is said to be IFRA if $-\frac{1}{t}\log\bar{\pi}_A(t)$ is increasing in t.

Example 5.3.1: Define the survival function

$$\bar{\pi}_A(t) = \begin{cases} 1, & if\ 0 \leq t \leq 2 \\ e^{2-t}, & if\ 2 \leq t \leq 3 \\ e^{-1}, & if\ 3 \leq t \leq 4 \\ e^{7-2t}, & if\ \quad t \geq 4 \end{cases}$$

It is easy to verify that $\bar{\pi}_A$ is not an IFRA. However,

$$H(f;t) = \begin{cases} 1 - \dfrac{\log 2}{e}, & if\quad 0 \leq t \leq 2 \\ 1 - \dfrac{\log 2}{e}e^{t-2}, & if\ 2 \leq t \leq 3 \\ 1 - \log 2, & if\quad t \geq 3 \end{cases}$$

and$H(f;t)$ is DFURL. This example shows that DFURL does not imply IFR.

Now we present a lower bound on a DFURL (IFURL) hazard function with known $H(f;t)$.

Theorem 5.3.2: Let $\bar{\pi}_A$ be a DFURL (IFURL), then

$$\lambda_{\pi_A}(t) \leq (\geq) exp\left(1 - H(f;t)\right),\ \ t \geq 0 \qquad (5.3.2)$$

Proof. We will prove it for DFURL. Similar arguments can be used for IFURL. Since $\bar{\pi}_A$ is DFURL, we get that;

$$\log\lambda_{\pi_A}(t) - \frac{1}{\bar{\pi}_A(t)}\int_t^{\infty}\left(\log\lambda_{\pi_A}(x)\right)f(x)dx \leq 0\ for\ all\ t \geq 0$$

That is,

$$\log \lambda_{\pi_A}(t) + H(f;t) \leq 1, \quad (5.3.3)$$

$$for\ all\ t \geq 0$$

From (5.3.3), we get the result.

Remark 5.3.1: Using (5.3.1), (5.1.1) and (5.1.3), if $\bar{\pi}_A$ is a DFURL (IFURL), then

$$\bar{\pi}_A(t) \geq (\leq) - exp \int_0^t \left(1 - H(f;x)\right)dx.$$

and $$\frac{\delta^{/}_{\pi_A}(t)+1}{\delta_{\pi_A}(t)} \leq (\geq) exp\left(1 - H(f;t)\right)$$

for all $t \geq 0$. Furthermore, in Theorem 5.3.2, the equality holds if $\bar{\pi}_A$ is an exponential.

The following corollary gives an upper bound for $H(f;t)$ with known $f(0)$.

Corollary 5.3.1: Let $\bar{\pi}_A$ be a DFURL (IFURL), then,

$$H(f;t) \leq (\geq) 1 - \log \lambda_{\pi_A}(0) = 1 - \log f(0).$$

5.4 A New Class of Generalized Fuzzy entropy Functions:-

Probability Measure Of Fuzzy Events: In probability theory [101], an event is a member of σ-field ,α, of subsets of a sample space Ω. A probability measure, P, is a normed measure over a measurable space (α, Ω); that is, P is a real valued function which assigns to every $A\ in\ \alpha$; a probability measure$P(A)$, such that (a) $P(A) \geq 0 for\ all\ A \in \alpha$; (b) $P(\Omega) = 1$; and (c) P is countably additive, i.e., if $\{A_i\}$ is any collection of disjoint events then;

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

We shall assume that for simplicity that Ω is an Euclidean n-space R^n. Thus our probability space will be assumed to be a triplet $(R^n, \alpha, P), where\ \alpha$ is σ-field of Borel sets in R^n and P is a probability measure over R^n. A point in R^n will be denoted by x.

Let $A \in \alpha$, then the probability of A can be expressed as;

$$P(A) = \int_A dP.$$

or equivalently,

$$P(A) = \int_{R^n} \mu_A(x) dP = E(\mu_A).$$

Where μ_A denotes the characteristic function of $A(\mu_A(x) = 0 \; or \; 1)$. And $E(\mu_A)$ is the expectation of μ_A.

The equation $P(A) = \int_{R^n} \mu_A(x) dP = E(\mu_A)$ equates the probability of an event A with the expectation of the characteristic function of A. It is this equation that can readily be generalized to fuzzy events through the use of the concept of fuzzy set.

Fuzzy Set and Membership Function: A fuzzy set $A \; in \; R^n$ is defined by a characteristic function $\mu_A: R^n \to [0, 1]$ which associates with each $x \; in \; R^n$ its "grade of membership," $\mu_A(x), in \; A$. To distinguish between the characteristic function of a non-fuzzy set and the characteristic function of a fuzzy set, the latter will be referred to as a membership function.

Definition: Let (R^n, α, P) be a probability space in which α is a σ-field of Borel sets in $R^n \; and \; P$ is a probability measure over R^n. Then fuzzy event in R^n is a fuzzy set A in R^n whose membership function, $\mu_A: R^n \to [0, 1]$ is Borel measurable.

The probability of a fuzzy event A is defined by the Lebesgue-Stieltjes integral

$$P(A) = \int_{R^n} \mu_A(x) dP = E(\mu_A).$$

Thus the probability of a fuzzy event is the expectation of its membership function. The existence of the Lebesgue-Stieltjes integral is insured by the assumption that μ_A is Borel measurable.

Let $\mu_A(T)$ be a membership function with density function $P(A)$. Then corresponding to verma's entropy of order$\alpha \; and \; type \; \beta$, the fuzzy entropy is defined as;

$$H(\alpha, \beta) = \frac{1}{\beta - \alpha} \log \int \left(\mu_A(t)^{\alpha+\beta-1} + (1 - \mu_A(t))^{\alpha+\beta-1}\right) dP$$

for $\beta - 1 < \alpha < \beta, \beta \geq 1$ (5.4.1)

and in discrete case;

$$H(\alpha,\beta)=\frac{1}{\beta-\alpha}\log\sum_{k=1}^{n}\left(\mu_A(t)^{\alpha+\beta-1}+(1-\mu_A(t))^{\alpha+\beta-1}\right)$$

for $\beta-1<\alpha<\beta,\beta\geq 1$ (5.4.2)

also $\lim_{\alpha\to 1,\beta=1} H(\alpha,\beta)=-\int\left[\mu_A(t)\log\mu_A(t)+\left(1-\mu_A(t)\right)\log\left(1-\mu_A(t)\right)\right]dt$

(5.4.3)

and in discrete case

$$\lim_{\alpha\to 1,\beta=1} H(\alpha,\beta)=-\sum_{i=1}^{n}\left[\mu_A(t_i)\log\mu_A(t_i)+\left(1-\mu_A(t_i)\right)\log\left(1-\mu_A(t_i)\right)\right]$$

(5.4.4)

This is measure of fuzzy entropy due to Luca and termini [33] in both the cases.

As argued by Ebrahimi [37], if a unit is known to have survived up to an age t, then$H(t)$ is no longer useful in measuring the uncertainty about the remaining life time of the unit. The idea is that a unit with great uncertainty is less reliable than a unit with low uncertainty. Accordingly, he introduced a measure of uncertainty known as residual entropy for the residual life time distribution. The residual entropy of a fuzzy set A is defined as,

$$H(\mu_A(T),t)=-\int_t^\infty\frac{\mu_A(x)}{\bar{\pi}_A(t_i)}\log\frac{\mu_A(x)}{\bar{\pi}_A(t_i)} \quad (5.4.5)$$

$$H\left(\mu_A(t_j)\right)=-\sum_{k=j}^{n}\frac{\mu_A(t_k)}{\bar{\pi}_A(t_j)}\log\frac{\mu_A(t_k)}{\bar{\pi}_A(t_j)} \quad (5.4.6)$$

5.5 Generalized Fuzzy Residual Entropy Function:-

Let $\mu_A(T)$ be a membership function representing the component failure time with failure distribution $\pi_A(t)=P(\mu_A(T)\leq \mathrm{t})$ and survival function $\bar{\pi}_A(t)=1-\pi_A(t)$. We shall assume that the component is functioning at $t=0$ and it will fail at some $t>0$, so that $\bar{\pi}_A(0)=1$, and differentiability of $\pi_A(t)$and shall let $f\left(\mu_A(t)\right)=\pi'_A(t)$ denote its failure density function. We define fuzzy entropy corresponding to Verma's entropy for residual life as,

$$H(\alpha,\beta,t)=\frac{1}{\beta-\alpha}\log\left(\frac{\int_t^\infty \mu_A^{\alpha+\beta-1}(x)}{\bar{\pi}_A^{\alpha+\beta-1}(t)}dx\right), \beta-1<\alpha<\beta, \beta\geq 1 \quad (5.5.1)$$

or $\quad (\beta-\alpha)H(\alpha,\beta,t)=\log\left(\int_t^\infty \mu_A^{\alpha+\beta-1}(x)dx\right)-(\alpha+\beta-1)\log\bar{\pi}_A(t)$;

$$\beta-1<\alpha<\beta, \beta\geq 1. \quad (5.5.2)$$

for $\quad \beta=1, \alpha\to 1$ (5.5.1) tends to (5.4.5)

We now show that $H(\alpha,\beta,t)$ uniquely determines $\bar{\pi}_A(t)$.

Theorem 5.5.1: Let $\mu_A(T)$ be a random membership function having density function f and distribution function $\pi_A(t)=P(\mu_A(T)\leq t)$ and survival function

$$\bar{\pi}_A(t)=1-\pi_A(t).$$

Assume

$$H(\alpha,\beta,\mu_A(t))<\infty, t\geq 0, \beta-1<\alpha<\beta, \beta\geq 1$$

and increasing in t, then $H(\alpha,\beta,t)$ uniquely determines $\bar{\pi}_A(t)$.

Proof: differentiating (5.5.2) with respect to t, we get,

$$(\beta-\alpha)H^{/}(\alpha,\beta,t)=(\alpha+\beta-1)h(t)-\frac{\pi_A^{\alpha+\beta-1}(t)}{\int_t^\infty \pi_A^{\alpha+\beta-1}(x)dx} \quad (5.5.3)$$

Where $h(t)=\frac{\mu_A(t)}{\bar{\pi}_A(t)}$ is the failure rate function.

From (5.5.2) and (5.5.3), we have;

$$h^{\alpha+\beta-1}(t)=(\alpha+\beta-1)h(t)exp\big((\beta-\alpha)H(\alpha,\beta,t)\big) \quad (5.5.4)$$

$$-(\beta-\alpha)H^{/}(\alpha,\beta,t)exp\big((\beta-\alpha)H(\alpha,\beta,t)\big)$$

Hence for fixed $t>0, h(t)$ is a solution of

$$g(x)=(x)^{\alpha+\beta-1}-(\alpha+\beta-1)x\ exp\big((\beta-\alpha)H(\alpha,\beta,t)\big) \quad (5.5.5)$$

$$+(\beta-\alpha)H^{/}(\alpha,\beta,t)\ exp\big((\beta-\alpha)H(\alpha,\beta,t)\big)=0$$

Differentiating both sides with respect to x, we get;

$$g^{/}(x)=(\alpha+\beta-1)(x)^{\alpha+\beta-2}-(\alpha+\beta-1)\ exp\big((\beta-\alpha)H(\alpha,\beta,t)\big) \quad (5.5.6)$$

For extreme value of $g(x)$, we have;

$g^{/}(x) = 0$, which gives

$$x = exp\left(\frac{\beta - \alpha}{\alpha + \beta - 2} H(\alpha, \beta, t)\right) = x_t$$

also

$$g^{//}(x) = (\alpha + \beta - 1)(\alpha + \beta - 2)x^{\alpha+\beta-3}$$

Case I: Let $\alpha + \beta > 2$, then $g^{//}(x_t) > 0$. thus $g(x)$ attains maximum at x_t. Also, $g(0) > 0$and $g(\infty) = \infty$. Further, $g(x)$ decreases for $0 < x < x_t$ and hence increases for $x > x_t$. So, $x = h(t)$ is the unique solution to $g(x) = 0$.

Case II: If $\alpha + \beta < 2$, then $g^{//}(x_t) < 0$. Thus $g(x)$ attains maximum at x_t. Also,

$g(0) > 0$and $g(\infty) = -\infty$.

Further it can be seen that $g(x)$ decreases for $x > x_t$ and increases for

$0 < x < x_t$. So $x = h(t)$.

is the unique solution to $g(x) = 0$.

Theorem 5.5.2: Let $\mu_A(T)$ be a random membership function having fuzzy residual entropy

$$H(\alpha, \beta, t) = \frac{1}{\beta-\alpha}\log(k) - \frac{2-\alpha-\beta}{\beta-\alpha}\log h(t) \tag{5.5.7}$$

Where $h(t)$ is the failure rate function of$\mu_A(T)$, then $\mu_A(T)$ has

I. Exponential distribution iff$k = \frac{1}{\alpha+\beta-1}$.

II. Pareto distribution iff$k < \frac{1}{\alpha+\beta-1}$.

III. Finite range distribution iff$k > \frac{1}{\alpha+\beta-1}$.

Proof (I): Let $\mu_A(T)$ has exponential distribution with distribution function

$$\mu_A(t) = \frac{1}{\theta} exp\left(-\frac{t}{\theta}\right), t > 0, \theta > 0$$

The reliability function is given by

$$\bar{\pi}_A(t) = exp\left(-\frac{1}{\theta}\right).$$

The failure rate function

$$h(t) = \frac{1}{\theta}.$$

therefore $\qquad H(\alpha,\beta,t) = \frac{1}{\beta-\alpha}\log\left(\frac{\int_t^\infty \mu^{\alpha+\beta-1}(x)}{\bar{\pi}^{\alpha+\beta-1}(t)}dx\right), \beta-1<\alpha<\beta, \beta\geq 1$

or $\qquad H(\alpha,\beta,t) = \frac{1}{\beta-\alpha}\log(k) - \frac{2-\alpha-\beta}{\beta-\alpha}\log h(t)$

Where $\qquad k = \frac{1}{\alpha+\beta-1}, h(t) = \frac{1}{\theta}.$

Thus (5.5.7) holds.

Conversely, suppose $k = \frac{1}{\alpha+\beta-1}$, then

$$\frac{1}{\beta-\alpha}\log(k) - \frac{2-\alpha-\beta}{\beta-\alpha}\log h(t) = \frac{1}{\beta-\alpha}\log\left(\frac{\int_t^\infty \mu^{\alpha+\beta-1}(x)}{\bar{\pi}^{\alpha+\beta-1}(t)}dx\right)$$

or

$$\int_t^\infty \mu^{\alpha+\beta-1}(x) = \bar{\pi}^{\alpha+\beta-1}(t)\; exp(\log(k) - (2-\alpha-\beta)\log h(t))$$

Differentiating both sides with respect to t, we get;

$$\frac{h^2(t)}{h^/(t)} = \frac{k(2-\alpha-\beta)}{1-k(\alpha+\beta-1)}$$

or

$$h^{-2}(t)h^/(t) = \frac{1-k(\alpha+\beta-1)}{k(2-\alpha-\beta)}$$

or

$$h(t) = \left(\frac{1-k(\alpha+\beta-1)}{k(2-\alpha-\beta)}t + \frac{1}{h(0)}\right)^{-1} = (at+b)^{-1} \qquad (5.5.8)$$

where $a = \frac{1-k(\alpha+\beta-1)}{k(2-\alpha-\beta)}$ *and* $b = \frac{1}{h(0)}$.

now $k = \frac{1}{\alpha+\beta-1}$, therefore $a = 0$

Cleary (5.5.8) is the failure rate function of the exponential distribution.

(II) The density function of the Pareto distribution is given by;

$$\mu_A(t) = \frac{(b)^{\frac{1}{a}}}{(at+b)^{1+\frac{1}{a}}}, t \geq 0, a > 0, b > 0$$

The reliability function is given by;

$$\bar{\pi}_A(t) = \frac{(b)^{\frac{1}{a}}}{(at+b)^{\frac{1}{a}}}, t \geq 0, a > 0, b > 0$$

The failure rate is given by;

$$h(t) = (at+b)^{-1} \tag{5.5.9}$$

and

$$H(\alpha,\beta,t) = \frac{1}{\beta-\alpha}\log(k) - \frac{2-\alpha-\beta}{\beta-\alpha}\log h(t)$$

Where $k = \frac{1}{(\alpha+\beta-1)+\alpha(\alpha+\beta-2)}$ and $h(t) = (at+b)^{-1}$. Since $\alpha+\beta > 2$,therefore $k < \frac{1}{\alpha+\beta-1}$, thus (5.5.7) holds.

Conversely, suppose $k < \frac{1}{\alpha+\beta-1}$, proceeding as in (I), (5.5.8) gives

$$h(t) = \left(\frac{1-k(\alpha+\beta-1)}{k(2-\alpha-\beta)}t + \frac{1}{h(0)}\right)^{-1} = (at+b)^{-1} \tag{5.5.10}$$

where $a = \frac{1-k(\alpha+\beta-1)}{k(2-\alpha-\beta)}$ *and* $b = \frac{1}{h(0)}$.

since $k < \frac{1}{\alpha+\beta-1}$ *and* $\alpha+\beta > 2, therefore\ a > 0.$

Clearly (5.5.10) is the failure rate function of the Pareto distribution given in (5.5.9)

(III) The density function of the finite range distribution is given by;

$$\mu_A(t) = \frac{\beta_1}{v}\left(1 - \frac{t}{v}\right)^{\beta_1 - 1}, \beta_1 > 1, 0 \le t \le v < \infty$$

The reliability function is given by;

$$\bar{\pi}_A(t) = \left(1 - \frac{t}{v}\right)^{\beta_1}, \beta_1 > 1, 0 \le t \le v < \infty$$

The failure rate function is;

$$h(t) = \frac{\beta_1}{v}\left(1 - \frac{t}{v}\right)^{-1} \qquad (5.5.11)$$

and

$$H(\alpha,\beta,t) = \frac{1}{\beta-\alpha}\log(k) - \frac{2-\alpha-\beta}{\beta-\alpha}\log h(t)$$

Where $\quad k = \frac{\beta_1}{(\alpha+\beta-1)(\beta_1-1)+1}$ and $h(t) = \left(\frac{\beta_1}{v}\right)\left(1 - \frac{t}{v}\right)^{-1}$

Since $\quad \alpha + \beta > 2$, therefore $k > \frac{1}{\alpha+\beta-1}$. Proceeding as in (I), (5.5.8) gives

$$h(t) = h(0)\left(1 - \frac{k(\alpha+\beta-1)}{k(\alpha+\beta-2)}h(0)t\right)^{-1} \qquad (5.5.12)$$

which is the failure rate function of the distribution given in (5.5.11).

5.6 New Class of Life Time Distribution:-

The survival function has increasing (decreasing) fuzzy entropy corresponding to Verma's entropy for residual life of order α and type β, if $H(\alpha,\beta,t)$ is increasing (decreasing) in $t, t > 0$. This implies that $\bar{\pi}_A(t)$**IFERL** (α,β) , **DFERL**(α,β) if

$$H^{/}(\alpha,\beta,t) \ge 0$$

$$\le 0$$

Theorem 5.6.1: If a distribution is **IFERL**(α,β) as well as **DFERL** (α,β) for some constant, then it must be exponential.

Proof: Since the membership function $\mu_A(T)$ is both **IFERL** (α,β) and **DFERL** (α,β), therefore;

$$H(\alpha,\beta,t) = \text{constant}.$$

$$\frac{1}{\beta-\alpha}\log\left(\frac{\int_t^\infty \mu^{\alpha+\beta-1}(x)}{\bar{\pi}^{\alpha+\beta-1}(t)}dx\right) = k.$$

or

$$\int_t^\infty \mu^{\alpha+\beta-1}(x)dx = \bar{\pi}^{\alpha+\beta-1}(t)exp\big(k(\beta-\alpha)\big).$$

Differentiating both sides with respect to t,we get;

$$\frac{\mu(t)}{h(t)} = constant.$$

or

$$h(t) = constant.$$

This means that the distribution is exponential.

The next theorem gives upper (lower) bounds to the failure rate function.

Theorem 5.6.2: If $\mu_A(T)$ is **IFERL** (α,β)**DFERL** (α,β), then

(1) $$\left(h(t) \le (\ge)(\alpha+\beta-1)^{\frac{1}{\alpha+\beta-2}}\right)exp\left(-\frac{\alpha-\beta}{\alpha+\beta-2}H(\alpha,\beta,t)\right)$$

If $\alpha+\beta > 2$.

(2) $$h(t) \ge (\le)(\alpha+\beta-1)^{\frac{1}{\alpha+\beta-2}}exp\left(-\frac{\alpha-\beta}{\alpha+\beta-2}H(\alpha,\beta,t)\right)$$

If $\alpha+\beta < 2$.

Proof: If $\mu_A(T)$ is **IFERL** (α,β), then

$$H^/(\alpha,\beta,t) \ge 0$$

Which gives $h^{\alpha+\beta-2}(t) \le (\alpha+\beta-1)exp\big((\beta-\alpha)H(\alpha,\beta,t)\big).$

Similarly, if $\mu_A(T)$ is **DFERL** (α,β), then

$$h^{\alpha+\beta-2}(t) \geq (\alpha+\beta-1)exp\big((\beta-\alpha)H(\alpha,\beta,t)\big).$$

Case I: If $\alpha+\beta > 2$ and $\mu_A(T)$ is **IFERL**$(\alpha,\beta)\big($**DFERL** $(\alpha,\beta)\big)(\alpha,\beta)$, then

$$h(t) \leq (\geq)(\alpha+\beta-1)^{\frac{1}{\alpha+\beta-2}} exp\left(-\frac{\alpha-\beta}{\alpha+\beta-2}H(\alpha,\beta,t)\right) \tag{5.6.1}$$

CaseII: If $\alpha+\beta < 2$ and $\mu_A(T)$ is **IFERL**$(\alpha,\beta)\big($**DFERL** $(\alpha,\beta)\big)$, then

$$h(t) \geq (\leq)(\alpha+\beta-1)^{\frac{1}{\alpha+\beta-2}} exp\left(-\frac{\alpha-\beta}{\alpha+\beta-2}H(\alpha,\beta,t)\right) \tag{5.6.2}$$

BIBLIOGRAPHY

[1] ***Abramson, N. [1963]: "****Information theory and coding"; Mc.Graw Hill, New York. and statistical inference, Metrika, vol. 36, pp.129-147.*

[2] ***Aczel, J. [1975]:****"On Shannon's inequality, optimal coding and characterization of Shannon's and Renyi's entropies"; Institute Novonal De Alta Mathematics, Symposia Maths Vol.15, pp.153-179.*

[3] ***Asadi, M. Ebrahimi N. [2000]:*** *"Residualentropyand its characterization interms of hazard function and mean residual life time function"; Statist. Prob. Lett. Vol. 49, pp 263-269.*

[4] ***Ash, B.R. [1990]:****"Information theory"; Dover, New York.*

[5] ***Atanassov,K. [1983]:****"Intuitionistic fuzzy sets"; VII ITKR Session, Sofia (Deposed in Central Library of the Bulgarian Academy Of Sciences, (1697/84), (in Bulgarian).*

[6] ***Atanassov, K. [1999]:****"Intuitionistic Fuzzy Sets: Theory and Applications";Physica, Heidelberg, Germany.*

[7] ***Atanassov, K. [1989]:****"More on intuitionistic fuzzy sets, Fuzzy Sets and Systems"; Vol. 33, No.1, pp.37-45.*

[8] ***Autar, R and Soni, R.S. [1975]: "****Inaccuracy and coding theorem"; Journ. Appl. Prob. Vol.12, pp. 845-851.*

[9] ***Autar, R and Khan, A.B. [1989]:“****On Generalized Useful Information for Incomplete Distribution”; Journal of Combinatorics, Information and System Sciences, Vol. 14(4), pp. 187 – 191.*

[10] ***Aven, T. [1985]: “****Reliability Evaluation of Multistate Components”; IEEE Transactions on Reliability, Vol. R – 34, No. 5, pp. 473 – 479.*

[11] ***Badaloni, S. and Falda, M. [2010]:****“Temporal-based medical diagnoses using a fuzzy temporal reasoning system”;Journal of Intelligent Manufacturing, Vol. 21, No. 1, pp.145-153.*

[12] ***Baig, M.A.K. and Javid, M. [2014]:“****Fuzzy Coding Theorems on Generalized Fuzzy Cost Measure”; Asian Journal of Fuzzy and Applied Mathematics (ISSN: 2321 – 564X), Vol. 02, Issue, 01*

[13] ***Baig, M.A.K. and Javid, M. [2014]: “****Some New Results on Fuzzy Directed Divergence Measures and Their Inequalities”; Asian Journal of Mathematics and Statistics. Vol. 7.pp 12 – 20.*

[14] ***Baig, M.A.K. and Javid, M. [2014]:“****Some New Generalization of Fuzzy Average Codeword Length and Their Bounds”; American Journal of Applied Mathematics and Statistics. Vol. 2, pp.73 – 76.*

[15] ***Baig, M.A.K. and Javid, M. [2014]:“****Some Coding Theorems on Fuzzy Entropy Function depending Upon Parameter R and V”; IOSR Journal of Mathematics. Vol. 9, pp. 119 – 123.*

[16] **Baig, M.A.K. and Javid, M. [2013]:** *"Some Generalization of Fuzzy Average Length and Inequalities"; International Journal of Statistics and Analysis. Vol. 3, pp. 393 – 400.*

[17] **Baig, M.A.K. and Zaheeruddin [1993]:** *"Coding Theorem and a Measure of Inaccuracy"; Soochow Journal of Mathematics, Vol. 19(4), pp. 1- 7.*

[18] **Barlow, R.E. [1975]:** *"Statistical Theory of Reliability and Life Testing"; Probability Models, Holt, Rinbart and Winston, USA.*

[19] **Barrett, J.D., and Woodall, W.H. [1997]:** *"A Probabilistic Alternative to Fuzzy Logic Controllers"; IEEE Transactions, Vol. 29, pp. 459 – 467.*

[20] **Basu, D. [1975]:** *"Statistical Information and Likelihood";Sankhya, Ser. A, Vol. 37, pp. 1 – 71.*

[21] **Beckenback, E.F. and Bellman, R. [1961]:** *"Inequalities"; Springer, Berlin.*

[22] **Belis, M.and Guiasu, S. [1968]:** *"A Quantitative and Qualitative Measure of Information in cybernetic System"; IEE Transaction on information theory, Vol. IT-14, pp. 593 – 594.*

[23] **Bernard, M.A. and Sharma, B.D. [1988]:** *"Some Combinatorial results on variable length error correcting Codes"; ARS Combinatorial, Vol. 25B, pp. 181 - 194*

[24] ***Bernard, M.A. and Sharma, B.D. [1990]:*** *"A lower bound on average code word length of variable length error correcting code"; IEEE trans. Information theory, Vol. 36, pp. 1474 – 1475.*

[25] ***Bhandari, D. and N. R. Pal [1993]:*** *"Some new information measures for fuzzy sets"; Information Science, Vol. 67, pp. 204-228.*

[26] ***Bhandari, D., Pal N.R. andMajumdar, D.D. [1992,1993]:*** *"Fuzzy Divergence, Probability Measure of Fuzzy Events and Image Thresholding"; Pattern Recognition Letters , pp. 857-868.*

[27] ***Boekee, E. and Van Der Lubbe, J.C.A. [1980]:*** *"The R- Norm Information measure"; Information and Control, Vol. 45, pp. 136-155.California Press,Vol. 1, pp. 547-561.*

[28] ***Campbell, L.L. [1965]: "****A coding theorem and Renyi's entropy"; Information and control, vol. 8, pp.423-429. Combinatorial, 25B, pp. 181-194.*

[29] ***Cai, K.Y., Wen, C.Y. [1991]: "****Fuzzy Reliability Modeling of Gracefully Degradable Computing Systems"; Reliability Engineering and System Safety, Vol. 33. pp. 141 – 157.*

[30] ***Chaudary, B.B. and Rosenfeld, A. [1996]: "****On a metric distance between fuzzy sets"; Pattern Recognition Letters, Vol.17, No.1, pp.1157-1160.*

[31] ***Cisaszer, I. [1967b]:*** *"On Topological Properties of Divergence";Studia Math. Hungarica, Vol.2, pp.329-339.*

[32] ***Csiszar, I. [1967a]:****"Information type measures of difference of probabililty distribution and indirect observation";Studia Math. Hungarica, Vol.2, pp.229-318.*

[33] ***De Luca and Termini S. [1972]:****"A Definition of a Non-probabilistic Entropy in the Setting of fuzzy sets theory"; Information and Control, Vol.20, pp.301-312.*

[34] ***Dubois, D. and Prade, H. [1980]: "****Fuzzy set and Systems"; Theory and Applications. Academic Press.*

[35] ***Dubois, D. and Prade, H. [1987]:****"Properties of information in Evidence and Possibility Theories"; Fuzzy sets and systems, Vol.24, pp.161-182.*

[36] ***Ebanks, B.R. [1983]:****"On measures of fuzziness and their representations"; Journal of Math. Anal and Appl., Vol. 94, pp. 24-37.*

[37] ***Ebrahimi N. [1996]:*** *"How to Measure Uncertainty in the Life Time Distributions";Sankya, vol. 4, pp, 48 – 57.*

[38] ***El-Hawary, M.E. [2000]: "****Electric Power Applications of fuzzy systems"; IEEE Press Series on Power Engineering.*

[39] ***Feinstin, A. [1958]:****"Foundations of information theory"; McGraw-Hill, New York.*

[40] ***Fisher, R.A. [1925]: "****Theory of Statistical Estimation"; Mathematical Proceeding of the Cambridge Philosophical Society, Vol. 25, N0. 5, pp. 700, 725.*

[41] ***Gadaras and Mikhailov, L. [2009]:***"*An interpretable fuzzy rule-based classification methodology for medical diagnosis";Artificial Intelligence in Medicine, Vol. 47, No. 1, pp. 25–41.*

[42] ***Gill, M.A., Perez, R. and Gill, P. [1989]:*** *"A family of measure of uncertainty involving utilities; Definition, properties, application and Statistical inference"; Metrika, Vol. 36, pp 129 - 147*

[43] ***Guiasu, S. [1971]:***"*Weighted Entropy"; Reports on Math., Physics, Vol. 2, pp.165-179.*

[44] ***Guiasu, S. and Picard, C.F. [1971]:***"*Borne inferieure de la longer de certain codes"; C.R. Academic Sciences, Paris,Vol. 273, pp. 248-251*

[45] ***Gurdial and Pessoa, F. [1977]: "****On Useful Information of Order α, Journal of Combinatorics"; Information and System Sciences, Vol. 2, pp. 158 - 162*

[46] ***Hammer, M. [2001]: "****Application of fuzzy theory to electrical machine reliability";ZeszytyNaukowPolitechnikiSlaskiej, Seria: Elektryka, Vol. 77, pp. 161 – 166.*

[47] ***Hamming, R.W. [1950]: "****Error detecting and error correcting codes"; Bell System Tech. Jour. Vol. 29, pp.147-160.*

[48] ***Hartley, R.V.L. [1928]: "****Transmission of Information"; Bell Systems and Technical Journal. Vol. 7, pp. 535 – 563.*

[49] ***Hartnett, W.E. [1974]: "****Foundation of coding theory"; D. Riddling Publishing Co., Dordietcht, Holland*

[50] ***Havrada, J. H. And Charvat, F. [1967]:****"Quantification methods of classificatory processes, the concepts of structural α entropy";Kybernetika, Vol.3, pp. 30-35.*

[51] ***Hooda D. S. [2004]:****"On Generalized Measures of Fuzzy Entropy";MathematicaSlovaca, Vol.54, pp. 315 - 325.*

[52] ***Hooda D. S. and. Bajaj R. K [2008]:****"On Generalized R-norm Measures of Fuzzy Information"; Journal of Applied Mathematics, Statistics and Informatics, Vol.4, No.2, pp.199-212.*

[53] ***Hooda, D. S. and Ram, A. [1998]: "****Characterization of Non – additive useful information measure, Recent advances in information theory; statistics and Computer applications"; CCC Haryana, Agriculture University , Hisar, pp. 248 – 251.*

[54] ***Jaynes, E. T. [1957]:****"Information Theory and statistical Mechanics"; Physical Review, Vol.106, pp.620-630, 108, pp.171-193.*

[55] ***Kapur, J. N. [1986]:****"A generalization of Campbell's noiseless coding theorem"; Jour. Bihar Math, Society, Vol.10, pp.1-10.*

[56] ***Kapur, J. N. [1990]:****"Maximum entropy models in Science and Engineering"; Wiley East, New Delhi.*

[57] ***Kapur, J. N. [1997]:****"Inequalities Theory, Applications and Measurements"; Mathematical Science Trust Society, New Delhi.*

[58] ***Kapur, J. N. [1997]:*** *"Measures of Fuzzy Information"; Mathematical Science Trust Society, New Delhi.*

[59] ***Kapur, J. N. [1998]:*** *"Entropy and Coding"; Mathematical Science Trust Society, New Delhi.*

[60] ***Kaufmann, A. [1957]:*** *"Introduction to Theory of Fuzzy Subsets"; New York, Academic.*

[61] ***Kerridge D. E. F. [1961]:*** *"Inaccuracy and Inference"; J. Royal Statist. Society, Vol.23(A), pp.184-194.*

[62] ***Kerridge, D. E. F. [1961]:*** *"Inaccuracy and Inference"; J. Royal Statist. Society, Vol.23(A), pp.184-194.*

[63] ***Kerridge, D.F. [1961]: "*** *Inaccuracy and Inference"; J.R. Statist. Soc. B, Vol. 23, pp. 184-194.*

[64] ***Khan, A.B., Autar, R. and Ahmad, H. [1981]:*** *"Noiseless Coding Theorems for Generalized Non-additive Entropy"; Tam Kong J. Math, Vol.1211, pp.15-20.*

[65] ***Kraft, L.J. [1949]:*** *"A device for quantizing grouping and coding amplitude modulates pulses"; M. S. Thesis, Department of Electrical Engineering, MIT, Cambridge.*

[66] ***Kulback-Leibler, R.A. [1951]:*** *"On Fuzzy symmetric Divergence"; AFSS, pp. 904-908*

[67] ***Kullback, S. [1959]:*** *"Information Theory and Statistics"; Willey and Sons, New Delhi, India.*

[68] ***Kullback, S. [1959]:*** *"Information Theory and Statistics"; Willey and Sons, New Delhi.*

[69] ***Kullback, S. [1968]:*** *"Information Theory and Statistics"; Dover, New York, NY, USA.*

[70] ***Kullback, S. and Leibler, R.A. [1951]:*** *"On Information and Sufficiency"; Annals of Mathematical Statistics, Vol.22, pp.79-86.*

[71] ***Kumar, A. [2011]:*** *"Some Generalized Measures of Fuzzy Entropy"; Vol. 1, No.2.*

[72] ***Lin, J. [1991]:*** *"Divergence measures based on the Shannon entropy";IEEE Transactions on Information Theory, Vol. 37, No. 1, pp.145-151.*

[73] ***Loginov, V.I. [1966]: "****Probability Treatment of Zadeh Membership Functions and Their Use in Pattern Recognition"; Engineering Cybernetics, pp. 68 – 69.*

[74] ***Longo, G. [1976]:*** *"A noiseless coding theorem for source having utilities"; SIAM Journal of Applied Mathematics, Vol.30, pp.739-748. Massach.Inst.Of Tech.*

[75] ***Lowen, R. [1996]:*** *"Fuzzy Set Theory–Basic Concepts"; Techniques and Bibliography, Kluwer Academic Publication.Applied Intelligence, Vol. 31, No. 3, pp.283-291.*

[76] ***Mathai, A.M. and Rathie, P.N. [1975]:***_“Basic Concept in Information Theory and Statistics”; Wiley Eastern Limited, New Delhi._

[77] ***Nath, P. [1970]:***_“An axiomatic characterization of inaccuracy for discrete generalized probability distributions”; operation Search, Vol. 7, pp. 115 - 133_

[78] ***Pal and Bezdek [1994]:***_“Measuring Fuzzy Uncertainty”; IEEE Trans. of fuzzy systems, Vol. 2, No. 2, pp.107-118._

[79] ***Parkash, O. [1998]: “***_A new parametric measure of fuzzy entropy”;Information Processing and Management of Uncertainty Vol.2, pp.1732-1737._

[80] ***Parkash, O. and Sharma, P.K. [2004]:***_“Measures of fuzzy entropy and their relations”;Inernationa. Journal of Management & Systems, Vol.20, pp.65-72._

[81] ***Parkash, O., Sharma, P. K. and Kumar, J. [2008]:***_“Characterization of fuzzy measures via concavity and recursivity”;Oriental Journal of Mathematical Sciences. Vol.1, pp.107-117._

[82] ***Parkash. O, And Sharma, P. K. [2004]:***_“Noiseless coding theorems corresponding to fuzzy entropies”; Asian Bulletine of Mathematical Society (Accepted)._

[83] ***Rathie, P.N. [1968]: “***_On Generalized Entropy and Coding Theorem”; Queen’s University Kingston Pre print, No. 17._

[84] ***Renyi, A. [1961]:*** *"On measures of entropy and information";Proceedings 4^{th} Berkeley Symposium on Mathematical Statistics and Probability, Vol.1, pp.547-561.*

[85] ***Rudas, I. J. [2001]:*** *"Measures of fuzziness"; theory and applications. Advances in fuzzy systems and evolutionary computation.pp.187-192. World Scientific and Engineering Society Press, Athens.*

[86] ***Sevart, J.B. [1983]: "****On a Broader Definition of Machine Failure"; American Society of Mechanical Engineering, pp. 125 – 129.*

[87] ***Shannon, C. E. [1948]:*** *"A mathematical theory of communication";Bell System Technical Journal, Vol.27, pp.379-423, 623-659.*

[88] ***Sharma D.and Mittal D.P. [1975]:*** *"New non-additive measures of entropy for discrete probability distribution"; J. Math. Sci. (Calcutta), Vol.10, pp.28-40.*

[89] ***Sharma, B.D. and Taneja, I. J. [1975]:*** *"Entropies of type α, β and other generalized measures of information theory";Mathematika, Vol.22, 3 pp. 205-215.*

[90] ***Shisha, O. [1967]:*** *"Inequalities", Academic press, New York.*

[91] ***Sing and Tomar [2009]:*** *"On Fuzzy Measures of Divergences and their Inequalities"; Proceeding 10^{th} Annual Confrence of Indian Society of Information Theory and Applications (ISITA), pp.31-43.*

[92] Sing, R.P., Kumar, R. and Tuteja, R.K. [2003]: "*Application of Holder's inequality in information theory"; Information Sciences, Vol. 152, pp. 145-154 Word Length and their Bonds, (communicated).*

[93] Smets, P. [1981]:*"Medical diagnosis"; Fuzzy sets and degrees of belief, Fuzzy sets and systems, Vol.5, pp. 259-266.*

[94] Steimann F. andAdlassnig K. P. [2001]:*"Fuzzy Medical Diagnosis";*

[95] Straszecka E. [2007]:*"Combining uncertainty and imprecision in models of medical diagnosis";Information Sciences, Vol. 176, No. 20, pp. 3026–3059.*

[96] Sturgeon, R.C., Rudy, R.J. [1983]: "*Everyone's Failure are Different – So What"; in: G, M. Kurajian (ed), Failure Prevention and Reliability, American Society of Mechanical Engineering, pp. 131 – 135.*

[97] Szmidt E. and Kacprzyk J. [2000]:*"Distances between intuitionistic fuzzy sets";Fuzzy Sets and Systems, Vol. 114, No. 3, pp. 505–518.*

[98] Szmidt E. and Kacprzyk J. [2001]:*"Intuitionistic fuzzy sets in some medical applications"; in Proceedings of the 7th Fuzzy Days Dortmund, B. Reusch, Ed., Vol. 2206 of Computational intelligence: theory and applications, pp. 148–151. Springer, Berlin, Germany.*

[99] Teneja, I. J. [2004]:*"Generalized Symmetric Divergence Measures and Inequalities";RGMIA Research Report Collection, Art 9.*

[100] Tsalli's, C.[1988]:*"Possible generalization of Boltzmann-Gibbs statistics"; Vol. 52, pp. 479-487. Van Der Lubbe, J.C.A. [1978]: On certain coding* *theorems for the information of order α and type β, In. Trans, 8^{th} Prague Conf. Inf. Theory, D. Reidell, Dordrecht, pp.253-256.*

[101] Tucker, H. G.[1967] *"A Graduate Course in Probability"; Academic Press, New Yark,*

[102] Xu, Z. S. [2006]:*"On correlation measures of intuitionistic fuzzy sets";Lecture Notes in Computer Science, Vol. 4224, pp.16-24.*

[103] Yager, R. R. [1979]:*"On measures of fuzziness and negation"; Part-I: membership in the unit interval, International Journal of General Systems, Vol.5, pp. 221-229.*

[104] Yager, R.R. [1980]:*"On Measures of Fuzziness and Negation"; Part II.Latices, Inform & Control, Vol.44, pp. 236-260.*

[105] Yao, F.and Yao, J. S. [2001]:*"Fuzzy decision making for medical diagnosis based on fuzzy number and compositional rule of inference";Fuzzy Sets and Systems, Vol. 120, No. 3, pp.351-366.*

[106] Zadeh, L. A. [1966]:*"Fuzzy Sets"; Information, and Control, Vol.8, pp.94-102.*

[107] Zadeh, L. A. [1968]:*"Probability measures of fuzzy events";Journal of Mathematical Analysis and Applications, Vol.23, pp.421-427.*

[108] Zadeh, L. A. [1978]: "*Fuzzy Set as a basis for theory of possibility"; Fuzzy set and systems, Vol. 1, pp. 3 – 28.*

[109] Zadeh. L.A. [1965]: *“Fuzzy sets”;Information and Control, Vol. 8, No.3, pp. 338-353.*

[110] Zhang Q. S.and S. Y. Jiang [2008]: *“A note on information entropy measures for vague sets and its applications”;Information Sciences, Vol. 178, No. 21, pp.4184-4191.*