Editors

Valeri Mladenov

Nikos Mastorakis

WSEAS

Information Theoretic Models and Their Applications

by Prof. Om Parkash

ISBN: 978-1-61804-371-9

Information Theoretic Models and Their Applications

Editors
Prof. Valeri Mladenov
Prof. Nikos Mastorakis

Author

Om Parkash

Published by WSEAS Press
www.wseas.org

ISBN: 978-1-61804-371-9

Information Theoretic Models and their Applications

Published by WSEAS Press
www.wseas.org

All papers of the present volume were peer reviewed by two independent reviewers. Acceptance was granted when both reviewers' recommendations were positive.
See also: http://www.worldses.org/review/index.html

ISBN: 978-1-61804-371-9

World Scientific and Engineering Academy and Society

Preface

The objective of the present book entitled "Information Theoretic Models and Their Applications" is to acquaint the readers with the quantitative measure of information theoretic entropy discovered by well known American Mathematician C.E. Shannon. This discovery has played an increasingly significant role towards its applications in various disciplines of Science and Engineering. On the other hand, peculiar to information theory, fuzziness is a feature of imperfect information which gave birth to fuzzy entropy, loosely representing the information of uncertainty, and was introduced by an eminent American Electrical Engineer, Lofti Zadeh. The measures of entropy for probability and fuzzy distributions have a great deal in common and the knowledge of one may be used to enrich the literature on the other and vice-versa. The present manuscript provides the contribution of both types of entropy measures.

The two basic concepts, viz, entropy and coding are closely related to each other. In coding theory, we develop optimal and uniquely decipherable codes by using various measures of entropy, and these codes find tremendous applications in defense and banking industry. Another idea providing a holistic view of problems comes under the domain of Jaynes "Maximum Entropy Principle" which deals with the problems of obtaining the most unbiased probability distributions under a set of specified constraints. The contents of the book provide a study of uniquely decipherable codes and the maximum entropy principle.

It is worth mentioning here that engineers, scientists, and mathematicians want to experience the sheer joy of formulating and solving mathematical problems and thus have very practical reasons for doing mathematical modeling. The mathematical models find tremendous applications through their use in a number of decision-making contexts. This is to be emphasized that the use of mathematical models avoids intuition and, in certain cases, the risk involved, time consumed and the cost associated with the study of primary research. The book provides a variety of mathematical models dealing with discrete probability and fuzzy distributions.

I am thankful to Guru Nanak Dev University, Amritsar, India, for providing me sabbatical leave to write this book. I am also thankful to my wife Mrs. Asha, my daughter Miss Tanvi and my son Mr. Mayank for their continuous encouragements towards my

academic activities and also for providing the congenial atmosphere in the family for writing this book. I would like to express my gratitude to my research scholars, Mr. Mukesh and Ms. Priyanka Kakkar, Department of Mathematics, Guru Nanak Dev University, Amritsar, India, for their fruitful academic discussions and efforts made in meticulous proof reading for the completion of the book project. I shall be failing in my duty if I do not thank the WSEAS publishing team for their help and cooperation extended in publishing the present book.

I have every right to assume that the contents of this reference book will be useful to the scientists interested in information theoretic measures, and using entropy optimization problems in a variety of disciplines. I would like to express my gratitude for the services rendered by eminent reviewers for carrying out the reviewing process and their fruitful suggestions for revising the present volume. I sincerely hope that the book will be a source of inspiration to the budding researchers, teachers and scientists for the discovery of new principles, ideas and concepts underlying a variety of disciplines of Information Theory. Also, it will go a long way, I expect, in removing the cobwebs in the existing ones. I shall be highly obliged and gratefully accept from the readers any criticism and suggestions for the improvement of the present volume.

Om Parkash
Professor, Department of Mathematics
Guru Nanak Dev University, Amritsar, India

Forward

The book "Information Theoretic Models and their Applications" written by Dr. Om Parkash, Professor of Mathematics, Guru Nanak Dev University, Amritsar, India, is an advanced treatise in information theory. This volume will serve as a reference material to research scholars and students of mathematics, statistics and operations research. The scholarly aptitude of Dr. Om Parkash is evident from his high rated contributions in the field of information theory. He is a meticulous, methodical and mellowed worker, with an in depth knowledge on the subject.

Dr. R.K.Tuteja
Ex-Professor of Mathematics
Maharshi Dayanand University, Rohtak, India
President, Indian Society of Information
Theory and Applications

Table of Contents

Preface

Forward

1. Information Theoretic Measures Based Upon Discrete Probability Distributions 1

1.1 Introduction 1

1.2 A new generalized probabilistic measure of entropy 7

1.3 New generalized measures of weighted entropy 10

1.4 Probabilistic measures of directed divergence 17

2. Generalized Measures of Fuzzy Entropy-Their Properties and Contribution 28

2.1 Introduction 28

2.2 New measures of fuzzy entropy for discrete fuzzy distributions 31

2.3 Monotonic character of new measures of fuzzy entropy 35

2.4 Partial information about a fuzzy set-a measurement 38

2.5 Generating functions for measures of fuzzy entropy 42

2.6 Normalizing measures of fuzzy entropy 45

3. New Generalized Measures of Fuzzy Divergence and Their Detailed Properties 50

3.1 Introduction 50

3.2 New generalised measures of weighted fuzzy divergence 52

3.3 Generating measures of fuzzy entropy through fuzzy divergence measures 60

3.4 Some quantitative-qualitative measures of crispness 65

3.5 Generating functions for various weighted measures of fuzzy divergence 68

4. Optimum Values of Various Measures of Weighted Fuzzy Information 73

4.1 Introduction 73

4.2 Optimum values of generalized measures of weighted fuzzy entropy 76

4.3 Optimum values of generalized measures of weighted fuzzy divergence 82

5. Applications of Information Measures to Portfolio Analysis and Queueing Theory 94

5.1 Introduction 94

5.2 Development of new optimization principle 99

5.3 Measuring risk in portfolio analysis using parametric measures of divergence 102

5.4 Applications of information measures to the field of queueing theory 104

6. New Mean Codeword Lengths and Their Correspondence with Information Measures 114

6.1 Introduction 114

6.2 Development of exponentiated codeword lengths through measures of divergence 120

6.3 Derivations of well known existing codeword lengths 125

6.4 Development of information theoretic inequalities via coding theory and measures of divergence 130

6.5 Generating possible generalized measures of weighted entropy via coding theory 133

6.6 Measures of entropy as possible lower bounds 145

7. A Study of Maximum Entropy Principle for Discrete Probability Distributions 151

7.1 Introduction 151

7.2 Maximum entropy principle for discrete probability distributions 156

7.3 Maximum entropy principle for approximating a given probability distribution 171

Index 178

CHAPTER-I

INFORMATION THEORETIC MEASURES BASED UPON DISCRETE PROBABILITY DISTRIBUTIONS

ABSTRACT

In this chapter, we have investigated and introduced a new generalized parametric measure of entropy depending upon a probability distribution and studied its essential and desirable properties. By taking into consideration the concept of weighted entropy, we have proposed some new generalized weighted parametric measures of entropy depending upon a probability and weighted distribution and studied their important properties. We have explained the necessity for a new concept of distance for the disciplines other than science and engineering and taking into consideration the application areas of distance measures, we have investigated and developed some new parametric and non-parametric measures of divergence for complete finite discrete probability distributions.

Keywords: ***Entropy, Cross entropy, Concave function, Additivity, Convex function, Symmetric directed divergence, Hessian matrix.***

1.1 INTRODUCTION

Introduced by well known American mathematician cum communication engineer Claude Shannon, also known as the father of the digital age, information theory is one of the few scientific fields fortunate enough to have an identifiable beginning. The path-breaking work of Shannon who published his first paper "a mathematical theory of communication" in 1948 is the Magna Carta of the information age. In the beginning of his paper, Shannon acknowledged the work done before him, by such pioneers as Harry Nyquist and RVL Hartley, working at the American Bell Laboratories in 1920s. Though their influence was profound, the work of those early pioneers had limited applications in their fields of interest. It was Shannon's unifying vision that revolutionized communication, and spawned a multitude of communication research that we now define as the field of Information Theory.

This theory is not just a product of the Shannon's [24] work only but the result of crucial contributions made by many well known distinct individuals, from a variety of disciplines. The direction of these pioneers, their perspectives and interests had provided a well behaved shape to "Information Theory" dealing with uncertainty, and was sponsored in anticipation of what it could provide. This perseverance and continued interest eventually resulted in the multitude of technologies we have today. Before Shannon's theory, there was only the fuzziest idea about a message and rudimentary understanding of how to transmit a waveform but there was essentially no understanding of how to turn a message into a transmitted waveform. After the publication of his first paper, Shannon [24] showed how information could be quantified with absolute precision, demonstrated the essential unity of all information media, and proved that every mode of communication could be encoded in bits. This paper provided a "blueprint for the digital age".

Information theory has mainly two primary objectives: The first one is the development of the fundamental theoretical limits on the achievable performance when communicating given information source over a given communication channel using coding schemes. The second object is the development of coding schemes providing reasonably good performance. Shannon's [24] original paper contained the basic results for simple memoryless sources and channels. Zadeh [27] remarked that uncertainty is an attribute of information and the theory provided by Shannon which has profound intersections with Probability, Statistics, Computer Science, and other allied fields of Science and Engineering, has led to a universal acceptance that information is statistical in nature. A logical consequence of Shannon's theory of uncertainty, in whatever form it is, is that it should be dealt with through the use of probability theory. Thus, information theory can be viewed as a branch of applied probability theory and it studies all theoretical problems connected with the transmission of information over communication channels. It was Shannon [24] who firstly developed a mathematical function to measure the uncertainty contained in a probabilistic experiment. By associating uncertainty with every probability distribution $P=\left(p_1, p_2, ..., p_n\right)$, Shannon introduced the concept of information theoretic entropy and developed a unique function that can measure the uncertainty, given by

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i \tag{1.1.1}$$

This is to be noted that unless specified, all logarithms are taken to the base 2. Shannon [24] called the expression (1.1.1) as entropy and the inspiration behind adopting the nomenclature "entropy" came from the close resemblance between Shannon's mathematical formula and very similar known formulae from thermodynamics. In statistical thermodynamics, the most general formula for the thermodynamic entropy S of a system is the Gibbs entropy, given by

$$S = -k_B \sum_{i=1}^{n} p_i \log p_i \tag{1.1.2}$$

where k_B is the Boltzmann constant, and p_i is the probability of a microstate.

Practically speaking, the links between information theoretic entropy and thermodynamic entropy are not very much evident. The researchers working in various disciplines of Physics and Chemistry are usually more interested in entropy changes as a system spontaneously evolves away from its initial conditions, in accordance with the second law of thermodynamics, rather than an unchanging probability distribution. The presence of Boltzmann's constant k_B indicates, the changes in S/k_B for even small amounts of substances in chemical and physical processes represent amounts of entropy which are extremely large compared to anything seen in data compression or signal processing. Moreover, thermodynamic entropy is defined in terms of macroscopic measurements and makes no reference to any probability distribution, which is central theme to the definition of information theoretic entropy. However, the possibility of connections between the two different types of entropies, that is, thermodynamic entropy and information entropy cannot be ruled out. In fact, it was Jaynes [14], who pointed out that thermodynamic entropy should be seen as an application of Shannon's information theory and this thermodynamic entropy should be interpreted as being proportional to the amount of Shannon information needed to define the detailed microscopic state of the system, that remains uncommunicated by a description solely in terms of the macroscopic variables of classical thermodynamics.

When Shannon [24] introduced the concept of entropy, it was then realized that entropy is a property of any stochastic system and the concept is now used widely in many fields. The tendency of the systems to become more disordered over time is best described by the second law of thermodynamics, which states that the entropy of the system cannot spontaneously decrease. Any discipline that can assist us in understanding it, measuring it, regulating it, maximizing or minimizing it and ultimately controlling it to the extent possible, should certainly be considered an important contribution to our scientific understanding of complex phenomena. Today, information theory which is one of such disciplines, is principally concerned with communication systems, but there are widespread applications in statistics, information processing and computing. A great deal of insight is obtained by considering entropy equivalent to uncertainty, where we attach the ordinary dictionary meaning to the later term. A generalized theory of uncertainty, playing a significant role in our perceptions about the external world has well been explained by Zadeh [27].

The measure of entropy (1.1.1) possesses a number of interesting properties as discussed below:

1. Non-negativity

$H(P)$ is always non-negative, that is,

$$H(P)=-\sum_{i=1}^{n} p_i \log p_i \geq 0$$

Since $-p_i \log p_i \geq 0$ for all i, the result is obvious. It is zero, if one $p_i = 1$ and rest are zeros.

2. Maxima

$H(p_1, p_2, ..., p_n) \leq \log n$, with equality if, and only if $p_i = \frac{1}{n}$, for all i.

3. Minima

Its minimum value is zero and it occurs when one of the probabilities in unity and all others are zero.

4. Continuity

$H(p_1, p_2, ..., p_n)$ is a continuous function of p_i's, that is, a slight change in the probabilities p_i results in the uncertainty measure also.

5. Symmetry

$H(p_1, p_2, ..., p_n)$ is a symmetric function of p_i's, that is, it is invariant with respect to the order of the outcomes.

6. Grouping or Branching Property

$$H(p_1, p_2, ..., p_n) = H(p_1 + ... + p_r, \quad p_{r+1} + ... + P_n) +$$

$$(p_1 + p_2 + ... + p_r) H\left(\frac{p_1}{\sum_{i=1}^{r} p_i}, ..., \frac{p_r}{\sum_{i=1}^{r} p_i}\right) + (p_{r+1} + ... + p_n) H\left(\frac{p_{r+1}}{\sum_{i=r+1}^{n} p_i}, ..., \frac{p_n}{\sum_{i=r+1}^{n} p_i}\right) \text{ for } r = 1, 2, ..., n-1.$$

7. Additivity

If $P = \{p_1, p_2, ..., p_n\}$ and $Q = \{q_1, q_2, ..., q_m\}$ are two independent probability distributions, then

$H(P*Q) = H(P) + H(Q),$

where $P*Q$ is the joint probability distribution.

The entropy measure has many other additional properties, for example:

(i) The maximum value increases with n, that is, we do expect uncertainty to increase with the number of outcomes.

(ii) $H_{n+1}(p_1, p_2, ..., p_n, 0) = H_n(p_1, p_2, ..., p_n)$, that is, uncertainty measure is not changed by adding an impossible outcome.

(iii) $H(P*Q) = H(P) + \sum_{i=1}^{n} p_i H(Q/A_i)$,

where $H(Q/A_i)$ is the conditional entropy of Q when the i[th] outcome A_i corresponding to P has happened.

Immediately, after Shannon introduced his measure, researchers realized the potential of its applications in a variety of disciplines and a large number of other information theoretic measures were investigated and characterized.

Firstly, Renyi [23] defined entropy of order α, given by

$$H_\alpha(P) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^\alpha \Big/ \sum_{i=1}^{n} p_i\right), \ \alpha \neq 1,\ \alpha > 0 \tag{1.1.3}$$

The entropy measure (1.1.3) includes Shannon's [24] entropy as a limiting case as α→ 1. Zyczkowski [28] explored the relationships between the Shannon's entropy and Renyi's entropies of integer order. The author established lower and upper bound for Shannon entropy in terms of Renyi entropies of order 2 and 3.

Havrada and Charvat [12] introduced first non-additive entropy, given by

$$H^\alpha(P) = \frac{\left[\sum_{i=1}^{n} p_i^\alpha\right] - 1}{2^{1-\alpha} - 1}, \ \alpha \neq 1, \alpha > 0 \tag{1.1.4}$$

Aczel and Daroczy [1] developed the following measures of entropy:

$$\phi_1(P) = -\frac{\sum_{i=1}^{n} p_i^r \log p_i}{\sum_{i=1}^{n} p_i^r}, \quad r > 0, \tag{1.1.5}$$

$$\phi_2(P) = \{s-r\}^{-1} \log\left\{\frac{\sum_{i=1}^{n} p_i^r}{\sum_{i=1}^{n} p_i^s}\right\}, \quad r \neq s, r > 0, s > 0 \tag{1.1.6}$$

The following measures of entropy are due to Sharma and Taneja [25]:

$$\phi_3(P) = -2^{r-1} \sum_{i=1}^{n} p_i^r \log p_i, \quad r > 0, \tag{1.1.7}$$

$$\phi_4(P) = \left\{2^{1-r} - 2^{1-s}\right\}^{-1} \sum_{i=1}^{n} \left\{p_i^r - p_i^s\right\}, \quad r \neq s,\ r > 0, s > 0 \tag{1.1.8}$$

Kapur [15] introduced a generalized measure of entropy of order 'α' and type 'β', viz.,

$$H_{\alpha}^{\beta}(P) = \frac{1}{\alpha+\beta-2}\left[\sum_{i=1}^{n} p_i^{\alpha} + \sum_{i=1}^{n} p_i^{\beta} - 2\right], \alpha \geq 1, \beta \leq 1 \, or \, \alpha \leq 1, \beta \geq 1 \tag{1.1.9}$$

Kapur [16] also introduced the following non-additive measures of entropy:

$$H_a(P) = -\sum_{i=1}^{n} p_i \log p_i + \sum_{i=1}^{n} [(1+ap_i)\log(1+ap_i) - ap_i], a > 0 \tag{1.1.10}$$

$$H_b(P) = -\sum_{i=1}^{n} p_i \log p_i + \frac{1}{b}\sum_{i=1}^{n} [(1+bp_i)\log(1+bp_i) + (1+b)\log(1+b)], b > 0 \tag{1.1.11}$$

In his theory, Burgin [4] claimed that it is not the measure of information that counts, but the operational significance given to such measures by the development of certain theorems such as the source coding theorems and channel coding theorems due to Shannon and his other successors. Similarly, some of the other theories have solid applications, including the Fisher information in statistics and the Renyi information in some noiseless source coding problem. Brissaud [3] defined that "Entropy is a basic physical quantity that has led to various, and sometimes apparently conflicting, interpretations". It has been successively assimilated to different concepts such as disorder and information. In his communication, the author considered these conceptions and established the following results:

(1) Entropy measures lack of information; it also measures information. These two concepts are complementary.

(2) Entropy measures freedom and this allows a coherent interpretation of entropy formulas and experimental facts.

(3) To associate entropy and disorder implies defining order.

Dehmer and Mowshowitz [7] described methods for measuring the entropy of graphs and to demonstrate the wide applicability of entropy measures. The authors discussed the graph entropy measures which play an important role in a variety of problem areas, including biology, chemistry and sociology, and moreover, developed relationships between certain selected entropy measures, illustrating differences quantitatively with concrete numerical examples. Some applications of the entropy measures to the field of linguistics have been extended by Parkash, Singh and Sunita [20] whereas certain important and desirable developments regarding entropy measures and their classification have been provided by Hillion [13] and Garrido [10].

1.2 A NEW GENERALIZED PROBABILISTIC MEASURE OF ENTROPY

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

$$H_\beta\left(P\right) = -\frac{\log_D \sum_{i=1}^{n} p_i \beta^{\log_D p_i}}{\log_D \beta}, \beta > 1, \beta \neq 1 \tag{1.2.1}$$

Obviously, we have $\lim_{\beta \to 1} H_\beta(P) = -\sum_{i=1}^{n} p_i \log_D p_i$

Thus, $H_\beta\left(P\right)$ can be taken as a generalization of Shannon's [24] well known measure of entropy.

Next, to prove that the measure (1.2.1) is a valid measure of entropy, we have studied its essential properties as follows:

(i) Clearly $H_\beta\left(P\right) \geq 0$

(ii) $H_\beta\left(P\right)$ is permutationally symmetric as it does not change if $p_1, p_2, p_3, ..., p_n$ are re-ordered among themselves.

(iii) $H_\beta\left(P\right)$ is a continuous function of p_i for all p_i's.

(iv) **Concavity:** To prove concavity property, we proceed as follows:

Let $\phi\left(p\right) = p_i \beta^{\log_D p_i}$. Then $\phi'\left(p\right) = \beta^{\log_D p_i}\left(1 + \log_D \beta\right)$

$$\phi''\left(p\right) = \frac{1}{p_i} \beta^{\log_D p_i} \log_D \beta\left(1 + \log_D \beta\right) > 0 \text{ for all } \beta > 1$$

Thus, $\phi\left(p\right)$ is a convex function of p. Now, since the sum of convex functions is also a convex function, $\sum_{i=1}^{n} p_i \beta^{\log_D p_i}$ is a convex function of $p_1, p_2, p_3, ..., p_n$. Also, since log of a convex function is convex, $-\log_D \sum_{i=1}^{n} p_i \beta^{\log_D p_i}$ is a concave function of $p_1, p_2, p_3, ..., p_n$.

Thus $H_\beta\left(P\right) = -\frac{\log_D \sum_{i=1}^{n} p_i \beta^{\log_D p_i}}{\log_D \beta}, \beta > 1, \beta \neq 1$ is a concave function.

Moreover, with the help of numerical data, we have presented $H_\beta(P)$ as shown in the following Fig.-1.2.1 which shows that the generalized measure (1.2.1) is concave.

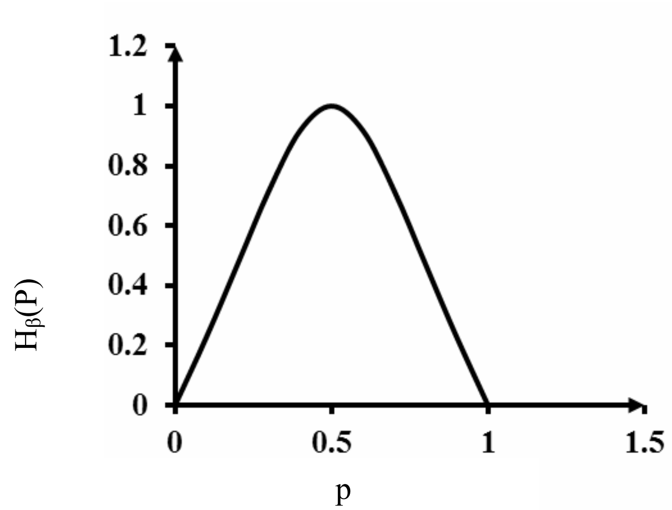


Fig.-1.2.1 Concavity of $H_\beta(P)$ with respect to P.

Hence, under the above conditions, the function $H_\beta(P)$ is a correct measure of entropy.

Next, we study the some most desirable properties of $H_\beta(P)$.

(i) $H_\beta(p_1, p_2, p_3, \ldots, p_n, 0) = H_\beta(p_1, p_2, p_3, \ldots, p_n)$

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

(ii) For degenerate distributions, $H_\beta(P) = 0$.

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

(iii) **Maximization of entropy:** We use Lagrange's method to maximize the entropy measure (1.2.1) subject to the natural constraint $\sum_{i=1}^{n} p_i = 1$.

In this case, the corresponding Lagrangian is given by

$$L \equiv -\frac{\log_D \sum_{i=1}^{n} p_i \beta^{\log_D p_i}}{\log_D \beta} - \lambda\left(\sum_{i=1}^{n} p_i - 1\right) \tag{1.2.2}$$

Differentiating equation (1.2.2) with respect to $p_1, p_2, p_3, \ldots, p_n$ and equating the derivatives to zero, we get $p_1 = p_2 = \ldots = p_n$. This further gives $p_i = \frac{1}{n} \forall i$

Thus, we observe that the maximum value of $H_\beta(P)$ arises for the uniform distribution and this result is most desirable.

(iv) The maximum value of the entropy is given by

$$H\left(\frac{1}{n},\frac{1}{n},\ldots,\frac{1}{n}\right)=\log_D n$$

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

(v) **Additivity property:** Let $P=\left(p_1,p_2,\ldots,p_n\right)$ and $Q=\left(q_1,q_2,\ldots,q_m\right)$ be any two independent discrete probability distributions of two random variables X and Y, so that

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

and

$$P\left(X=x_i,Y=y_j\right)=P\left(X=x_i\right)P\left(Y=y_j\right)=p_iq_j$$

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

$$H_\beta^{mn}\left(P*Q\right)=-\frac{\log_D \sum_{i=1}^{n}\sum_{j=1}^{m} p_iq_j\beta^{\log_D p_iq_j}}{\log_D \beta}$$

$$=-\frac{\log_D \sum_{i=1}^{n}\sum_{j=1}^{m} p_iq_j\beta^{\log_D p_i+\log_D q_j}}{\log_D \beta}$$

$$=-\frac{\log_D \sum_{i=1}^{n}\sum_{j=1}^{m} p_iq_j\beta^{\log_D p_i}\beta^{\log_D q_j}}{\log_D \beta}$$

$$=-\frac{\log_D\left(\sum_{i=1}^{n} p_i\beta^{\log_D p_i}\sum_{j=1}^{m} q_j\beta^{\log_D q_j}\right)}{\log_D \beta}$$

$$=-\frac{\log_D \sum_{i=1}^{n} p_i\beta^{\log_D p_i}}{\log_D \beta}-\frac{\log_D \sum_{j=1}^{m} q_j\beta^{\log_D q_j}}{\log_D \beta}$$

Thus, we have the following equality:

$$H_{\beta}^{mn}(P*Q)=H_{\beta}^{n}(P)+H_{\beta}^{m}(Q) \qquad (1.2.3)$$

Equation (1.2.3) shows that for two independent distributions, the entropy of the joint distribution is the sum of the entropies of the two distributions.

Thus, we claim that the new generalized measure of entropy of order β introduced in (1.2.1) satisfies all the essential as well as desirable properties of being an entropy measure; it is a new generalized measure of entropy.

Proceeding as above, many new probabilistic measures of entropy can be developed.

1.3 NEW GENERALIZED MEASURES OF WEIGHTED ENTROPY

It is to be noted that the measure of entropy introduced by Shannon [24] takes into account only the probabilities associated with the events and not their significance or importance. But, there exist many fields dealing with random events where it is necessary to take into account both these probabilities and some qualitative characteristics of the events. For instance, in two-handed game, one should keep in mind both the probabilities of different variants of the game, that is, the random strategies of the players and the wins corresponding to these variants. Thus, it is necessary to associate with every elementary event both the probability with which it occurs and its weight.

Innovated by the idea, Belis and Guiasu [2] proposed to measure the utility or weighted aspect of the outcomes by means of weighted distribution $W=\{w_1, w_2, ..., w_n\}$ where each of w_i is a non-negative real number accounting for the utility or importance or weight of its outcome. Weighted entropy is the measure of information supplied by a probabilistic experiment whose elementary events are characterized both by their objective probabilities and by some qualitative characteristics, called weights. To explain the concept of weighted entropy, let $E_1, E_2, ..., E_n$ denote n possible outcomes with $p_1, p_2, ..., p_n$ as their probabilities and let $w_1, w_2, ..., w_n$ be non-negative real numbers representing their weights. Then, it was Guiasu [11], who developed the following qualitative-quantitative measure of entropy:

$$H(P:W)=-\sum_{i=1}^{n} w_i\, p_i \log p_i \qquad (1.3.1)$$

and called it weighted entropy.

Some interesting results regarding weighted information have been investigated by Parkash and Singh [19], Parkash and Taneja [21], Kapur [16] etc.

Kapur [16] remarked that the expression given in equation (1.3.1) should be an appropriate measure of weighted entropy, if it satisfies certain desirable and important properties, given below:

(i) It must be continuous, and permutationally symmetric function of $(p_1, w_1);(p_2, w_2);....;(p_n, w_n)$.

(ii) $H(P:W) \geq 0$.

(iii) It must be maximum subject to $\sum_{i=1}^{n} p_i = 1$ when each p_i is some function of w_i that is, when $p_1 = g(w_1); p_2 = g(w_2);....; p_n = g(w_n)$.

(iv) Its minimum value should be zero and this should arise for a degenerate distribution.

(v) It should be a concave function of $p_1, p_2, ..., p_n$ because entropy function should always be concave.

(vi) When w_i 's are equal, it must reduce to some standard measure of entropy.

(vii) When it is maximized subject to linear constraints, the maximizing probabilities should be non-negative.

In this section, we have proposed some new generalized weighted parametric measures of entropy depending upon a probability distribution $P = \left\{(p_1, p_2, ..., p_n), p_i \geq 0, \sum_{i=1}^{n} p_i = 1\right\}$ and the weighted distribution $W = \{(w_1, w_2, ..., w_n), w_i \geq 0\}$, and studied their important properties.

I. A new generalized measure of weighted entropy of order α

We first introduce a new weighted generalized measure of entropy of order α given by the following mathematical expression:

$$H^{\alpha}(P;W) = \frac{1}{1-\alpha}\left[\sum_{i=1}^{n} w_i p_i^{\alpha} - \sum_{i=1}^{n} w_i p_i\right], \alpha \neq 1, \alpha > 0 \qquad (1.3.2)$$

To prove that the measure (1.3.2) is a valid measure of weighted entropy, we study its essential properties as follows:

(i) $H^{\alpha}(P;W)$ is a continuous function of p_i for all p_i 's.

(ii) $H^{\alpha}(P;W)$ is permutationally symmetric function of $p_1, p_2, \ldots, p_n$; $w_1, w_2, \ldots, w_n$ in the sense that it must not change when the pairs $(p_1, w_1),(p_2, w_2),\ldots,(p_n, w_n)$ are permuted among themselves.

(iii) For degenerate distributions, $H^{\alpha}(P;W)=0$.

Thus, $H^{\alpha}(P;W)\geq 0$.

(iv) **Concavity:** To prove concavity property, we proceed as follows:

We have

$$\frac{\partial H^{\alpha}(P;W)}{\partial p_i}=\frac{w_i}{1-\alpha}\left\{\alpha\, p_i^{\alpha-1}-1\right\}$$

and

$$\frac{\partial^2 H^{\alpha}(P;W)}{\partial p_i^2}=-\alpha\, w_i\, p_i^{\alpha-2}<0\ \ \forall\alpha$$

Thus, $H^{\alpha}(P;W)$ is a concave function of $p_1, p_2, \ldots, p_n$.

(v) **Maximization:** We use Lagrange's method to maximize the weighted entropy (1.3.2) subject to the natural constraint $\sum_{i=1}^{n} p_i = 1$. In this case, the corresponding Lagrangian is

$$L\equiv\frac{1}{1-\alpha}\left[\sum_{i=1}^{n} w_i\, p_i^{\alpha}-\sum_{i=1}^{n} w_i\, p_i\right]-\lambda\left[\sum_{i=1}^{n} p_i-1\right] \tag{1.3.3}$$

Differentiating equation (1.3.3) with respect to $p_1, p_2, \ldots, p_n$ and equating the derivatives to zero, we get

$$\frac{\partial L}{\partial p_1}=\frac{w_1}{1-\alpha}\left\{\alpha\, p_1^{\alpha-1}-1\right\}-\lambda=0$$

$$\frac{\partial L}{\partial p_2}=\frac{w_2}{1-\alpha}\left\{\alpha\, p_2^{\alpha-1}-1\right\}-\lambda=0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\frac{\partial L}{\partial p_n}=\frac{w_n}{1-\alpha}\left\{\alpha\, p_n^{\alpha-1}-1\right\}-\lambda=0$$

From the above set of equations, we have

$$p_1=\frac{1}{\alpha}\left(\frac{\lambda(1-\alpha)}{w_1}+1\right)^{1/(\alpha-1)}$$

$$p_2=\frac{1}{\alpha}\left(\frac{\lambda(1-\alpha)}{w_2}+1\right)^{1/(\alpha-1)}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$p_n = \frac{1}{\alpha}\left(\frac{\lambda(1-\alpha)}{w_n}+1\right)^{1/(\alpha-1)}$$

that is, each p_i is a function of w_i.

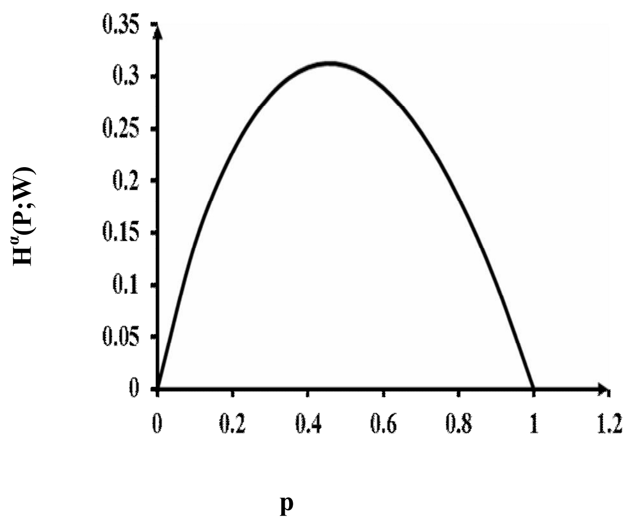
In particular, when the weights are ignored, then

$p_1 = p_2 = \cdots = p_n$ and applying the condition that $\sum_{i=1}^{n} p_i = 1$, we get $p_i = \frac{1}{n}\ \forall i$

Now $f(n) = \frac{n}{1-\alpha}\left[n^{1-\alpha}-1\right]$ where $f(n)$ denotes the maximum value.

Then $f'(n) = \frac{1}{n^{\alpha}} > 0\ \forall\alpha$

which shows that the maximum value is an increasing function of n and this result is most desirable. Moreover, with the help of numerical data, we have presented $H^{\alpha}(P;W)$ as shown in Fig.-1.3.1 which clearly shows that the measure (1.3.2) is a concave function in nature.



<u>**Fig.-1.3.1**</u> Concavity of $H^{\alpha}(P;W)$ with respect to P.

Note: If $w_i = 1\ \forall i$, then

$$H^{\alpha}(P) = \frac{1}{1-\alpha}\left[\sum_{i=1}^{n} p_i^{\alpha} - 1\right]; \alpha \neq 1, \alpha > 0$$ which is Havrada- Charvat's [12] measure of entropy.

Hence, under the above conditions, the function $H^{\alpha}(P;W)$ proposed in equation (1.3.2) is a correct generalized measure of weighted entropy.

II. A new generalized measure of weighted entropy of order α and type β

We now introduce a new generalized weighted measure of entropy of order α and type β, given by

$$H_\alpha^\beta(P;W)=\frac{1}{\beta-\alpha}\left[\sum_{i=1}^{n} w_i p_i^\alpha-\sum_{i=1}^{n} w_i p_i^\beta\right], \alpha>1, \beta<1 \text{ or } \alpha<1, \beta>1 \tag{1.3.4}$$

To prove that the measure (1.3.4) is a valid measure of weighted entropy, we study its essential properties as follows:

(i) $H_\alpha^\beta(P;W)$ is a continuous function of p_i for all p_i's.

(ii) $H_\alpha^\beta(P;W)$ is permutationally symmetric function of $p_1, p_2, \ldots, p_n$; $w_1, w_2, \ldots, w_n$ in the sense that it must not change when the pairs $(p_1, w_1), (p_2, w_2), \ldots, (p_n, w_n)$ are permuted among themselves.

(iii) For degenerate distributions $H_\alpha^\beta(P;W)=0$.

Thus, $H_\alpha^\beta(P;W)\geq 0$.

(iv) **Concavity:** To prove concavity property, we proceed as follows:

We have $\dfrac{\partial H_\alpha^\beta(P;W)}{\partial p_i}=\dfrac{w_i}{\beta-\alpha}\left[\alpha p_i^{\alpha-1}-\beta p_i^{\beta-1}\right]$

and

$$\frac{\partial^2 H_\alpha^\beta(P;W)}{\partial p_i^2}=\frac{w_i}{\beta-\alpha}\left[\alpha(\alpha-1)p_i^{\alpha-2}-\beta(\beta-1)p_i^{\beta-2}\right]<0 \text{ for all } \alpha>1, \beta<1 \text{ or } \alpha<1, \beta>1.$$

Thus, $H_\alpha^\beta(P;W)$ is a concave function of $p_1, p_2, \ldots, p_n$.

(v) **Maximization:** We use Lagrange's method to maximize the weighted entropy (1.3.4) subject to the natural constraint $\sum_{i=1}^{n} p_i=1$. In this case, the corresponding Lagrangian is

$$L\equiv\frac{1}{\beta-\alpha}\left[\sum_{i=1}^{n} w_i p_i^\alpha-\sum_{i=1}^{n} w_i p_i^\beta\right]-\lambda\left[\sum_{i=1}^{n} p_i-1\right] \tag{1.3.5}$$

Differentiating equation (1.3.5) with respect to $p_1, p_2, \ldots, p_n$ and equating the derivatives to zero, we get

$$\frac{\partial L}{\partial p_1} = \frac{w_1}{\beta - \alpha}\left[\alpha p_1^{\alpha-1} - \beta p_1^{\beta-1}\right] - \lambda = 0$$

$$\frac{\partial L}{\partial p_2} = \frac{w_2}{\beta - \alpha}\left[\alpha p_2^{\alpha-1} - \beta p_2^{\beta-1}\right] - \lambda = 0$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\frac{\partial L}{\partial p_n} = \frac{w_n}{\beta - \alpha}\left[\alpha p_n^{\alpha-1} - \beta p_n^{\beta-1}\right] - \lambda = 0$$

From the above set of equations, we have

$$\alpha p_1^{\alpha-1} - \beta p_1^{\beta-1} = \frac{\lambda(\beta-\alpha)}{w_1}$$

$$\alpha p_2^{\alpha-1} - \beta p_2^{\beta-1} = \frac{\lambda(\beta-\alpha)}{w_2}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\alpha p_n^{\alpha-1} - \beta p_n^{\beta-1} = \frac{\lambda(\beta-\alpha)}{w_n}$$

that is, each p_i is a function of w_i.

In particular, when the weights are ignored, then

$$\alpha p_1^{\alpha-1} - \beta p_1^{\beta-1} = \alpha p_2^{\alpha-1} - \beta p_2^{\beta-1} = \cdots = \alpha p_n^{\alpha-1} - \beta p_n^{\beta-1}$$

which is possible only if

$$p_1 = p_2 = \cdots = p_n = \frac{1}{n}.$$

Now $f(n) = \dfrac{1}{\beta-\alpha}\left[n^{1-\alpha} - n^{1-\beta}\right]$

Then

$$f'(n) = \frac{1}{\beta-\alpha}\left[\frac{1-\alpha}{n^{\alpha}} - \frac{1-\beta}{n^{\beta}}\right] > 0 \ \forall \alpha > 1, \beta < 1 \text{ or } \alpha < 1, \beta > 1$$

which shows that the maximum value is an increasing function of n and this result is most desirable since the maximum value of entropy should always increase.

Moreover, the graphical presentation of $H_{\alpha}^{\beta}(P;W)$ as shown in Fig.-1.3.2 clearly shows that the measure (1.3.4) is a concave function of p_i.

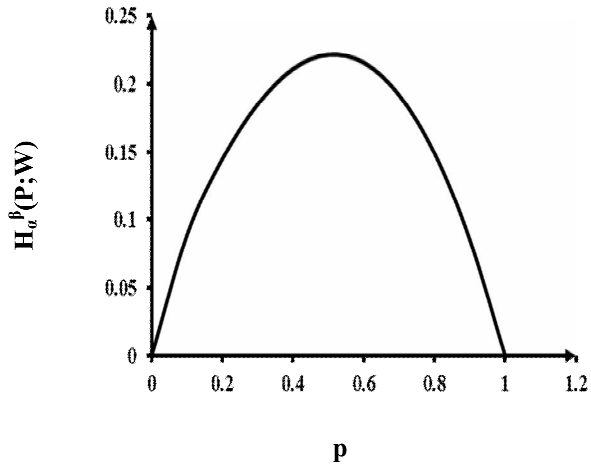


Fig.-1.3.2 Concavity of $H_\alpha^\beta(P;W)$ with respect to P.

Hence, under the above conditions, the function $H_\alpha^\beta(P;W)$ is a correct measure of weighted entropy.

Note: If $w_i = 1 \ \forall i$, then

$$H_\alpha^\beta(P) = \frac{1}{\beta-\alpha}\left[\sum_{i=1}^{n} p_i^\alpha - \sum_{i=1}^{n} p_i^\beta\right]$$

which is Sharma and Taneja's [25] entropy.

Thus, we claim that under the above conditions, the function $H_\alpha^\beta(P;W)$ proposed in equation (1.3.4) is a valid generalized measure of weighted entropy.

Note: In Biological Sciences, we have observed that researchers frequently use a single measure, that is, Shannon's [24] measure of entropy for measuring diversity in different species. But, if we have a variety of information measures, then we shall be more flexible in applying a standard measure depending upon the prevailing situation. Keeping this idea in mind, Parkash and Thukral [22] have developed some information theoretic measures depending upon well known statistical constants existing in the literature of statistics and concluded that for the known values of arithmetic mean, geometric mean, harmonic mean, power mean, and other measures of dispersion, the information content of a discrete frequency distribution can be calculated and consequently, new probabilistic information theoretic measures can be investigated and developed. Some of the measures developed by the authors are:

$$H^1(P) = n\log\left(\frac{G}{nM}\right) \tag{1.3.6}$$

$$H^2(P) = \frac{n^2}{\sum_{i=1}^{n}\frac{1}{p_i}} \tag{1.3.7}$$

$$H^3(P) = \frac{\sigma^2 + M^2}{nM^2} \tag{1.3.8}$$

$$H^4(P) = \left(\frac{M_p}{M}\right)^r \tag{1.3.9}$$

where the notations M, G, M_p and σ^2 stand for arithmetic mean, geometric mean, power mean and standard deviation respectively.

It is to be further observed that in statistics, the coefficient of determination r^2 is used in the context of statistical models whose main purpose is the prediction of future outcomes based upon certain related information. The coefficient of determination is the proportion of variability in a data set that is accounted for by the statistical model which provides a measure of how well future outcomes are likely to be predicted by the model. On the other hand, the coefficient of non-determination, $1-r^2$, the complement of the coefficient of determination, gives the unexplained variance, as against the coefficient of determination which gives explained variance as a ratio of total variance between the regressed variables. Since coefficient of determination measures association between two variables, there is a need to develop a measure which gives the randomness in linear correlation. Keeping this idea in mind, Parkash, Mukesh and Thukral [18] have proved that the coefficient of non-determination can act as an information measure and developed a mathematical model for its measurement.

1.4 PROBABILISTIC MEASURES OF DIRECTED DIVERGENCE

The idea of probabilistic distances, also called divergence measures, which in some sense assess how 'close' two probability distributions are from one another, has been widely employed in probability, statistics, information theory, and other related fields. In information theory, the Kullback–Leibler's [17] divergence measure, also known as information divergence or information gain or relative entropy, is a non-symmetric measure of the difference between two probability distributions P and Q. Specifically, the Kullback-Leibler (KL) divergence of Q from P is a measure of the information lost when Q is used to approximate P. Although it is often intuited as a metric or distance, the KL divergence is not a true metric- for example, it is not symmetric: the KL divergence from P to Q is generally not the same as the KL divergence from Q to P.

To make the proper understanding of the concept of distance in probability spaces, let $P=(p_1,p_2,\ldots,p_n)$ and $Q=(q_1,q_2,\ldots,q_n)$ be any two probability distributions, then we can use the following distance measure as usually defined in metric spaces:

$$^1D_1(P;Q)=\left[\sum_{i=1}^{n}(p_i-q_i)^2\right]^{\frac{1}{2}} \tag{1.4.1}$$

But, the Kullback–Leibler's [17] divergence measure defined in probability spaces is given by

$$^1D_2(P;Q)=\sum_{i=1}^{n}p_i\log\frac{p_i}{q_i} \tag{1.4.2}$$

We, now observe some special features related with both types of measures (1.4.1) and (1.4.2) as discussed below:

(i) $p_1,p_2,\ldots,p_n$; $q_1,q_2,\ldots,q_n$ are not any n real numbers, these have to lie between 0 and 1 and also satisfy $\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}q_i$ whereas in metric spaces, the coordinates are always any n real numbers.

(ii) The condition of symmetry essential for metric spaces is not so necessary in probability spaces. This is because of the reason that we have one distribution fixed and we have to find the distance of other distribution from it. Thus, we are essentially interested in the distances or divergences in one direction only. Hence, the condition of symmetry is restricted and we should not necessarily impose it on the distance measure in probability spaces.

(iii) We do not have much use for the triangle inequality because this inequality arises from the geometrical fact that the sum of the lengths of two sides of a triangle is always greater than the length of the third side. Here, since we are not dealing with geometrical distances, but with social, political, economic, genetic distances etc., we need not to satisfy the triangle inequality.

(iv) We want to minimize the distances in many applications and as such we should like the distance function to be convex function so that when it is minimized subject to linear constraints, its local minimum is global minimum. Thus, we want the distance measure to be minimized subject to some linear constraints on p_i's by the use of Lagrange's method and the minimizing probabilities should always be non-negative. In case Lagrange's method gives negative probabilities, we have to minimize this measure subject to linear constraints and non-negative conditions $p_i \geq 0$ and for this purpose; we have to use more complicated mathematical programming techniques.

We now compare the two measures, that is, (1.4.1) and (1.4.2).

(i) Both are ≥ 0 and vanish iff $p_i = q_i \ \forall i$, that is, iff $P = Q$.

(ii) ${}^1D_1(P;Q)$ is symmetric but ${}^1D_2(P;Q)$ will not in general be symmetric.

(iii) ${}^1D_1(P;Q)$ satisfies triangle inequality but ${}^1D_2(P;Q)$ may not.

(iv) Both are convex functions of $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$.

(v) When minimized subject to linear constraints on p_i's by using Lagrange's method, ${}^1D_1(P;Q)$ can lead to negative probabilities but due to the presence of logarithmic term, ${}^1D_2(P;Q)$ will always lead to positive probabilities.

(vi) Because of the presence of square root in the expression ${}^1D_1(P;Q)$, it is more complicated in applications than the expression ${}^1D_2(P;Q)$.

(vii) Both measures do not change if the n pairs $(p_1, q_1), (p_2, q_2), \ldots, (p_n, q_n)$ are permuted among themselves.

From the above comparison, we conclude that in geometrical applications, where symmetry of distance and triangle inequality are essential, the first measure is to be used but in probability spaces, when these conditions are not essential, the second measure is to be preferred. In the present book, since we have to deal with probability spaces only, we shall always use Kullback-Leibler's [17] divergence measure only.

Recently, Cai, Kulkarni and Verdu [5] remarked that Kullback-Leibler's [17] divergence is a fundamental information measure, special cases of which are mutual information and entropy, but the problem of divergence estimation of sources whose distributions are unknown has received relatively little attention.

Some parametric measures of directed divergence are:

$$ {}_{\alpha}D(P;Q) = \frac{1}{\alpha - 1} \log \sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha}, \alpha \neq 1, \alpha > 0 \tag{1.4.3} $$

which is Renyi's [23] probabilistic measure of directed divergence.

$$ D^{\alpha}(P;Q) = \frac{1}{\alpha - 1} \left[\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} - 1 \right], \alpha \neq 1, \alpha > 0 \tag{1.4.4} $$

which is Havrada and Charvat's [12] probabilistic measure of divergence.

$$D_a(P\,;Q)=\sum_{i=1}^{n} p_i \log\frac{p_i}{q_i}-\frac{1}{a}\sum_{i=1}^{n}(1+ap_i)\log\frac{(1+ap_i)}{1+aq_i},\ \ a\geq -1 \tag{1.4.5}$$

which is Kapur's [16] probabilistic measure of divergence.

$$D_\lambda(P;Q)=\frac{1}{\lambda}\sum_{i=1}^{n}(1+\lambda\ p_i)\log\frac{1+\lambda\ p_i}{1+\lambda\ q_i},\ \lambda>0 \tag{1.4.6}$$

which is Ferreri's [9] probabilistic measure of divergence.

Eguchi and Kato [8] have remarked that in statistical physics, Boltzmann-Shannon entropy provides good understanding for the equilibrium states of a number of phenomena. In statistics, the entropy corresponds to the maximum likelihood method, in which Kullback-Leibler's [17] divergence connects Boltzmann-Shannon entropy and the expected log-likelihood function. The maximum likelihood estimation has been supported for the optimal performance, which is known to be easily broken down in the presence of a small degree of model uncertainty. To deal with this problem, the authors have proposed a new statistical method.

Recently, Taneja [26] remarked that the arithmetic, geometric, harmonic and square- root means are all well covered in the literature of information theory. In this paper, the author has constructed divergence measures based on non-negative differences between these means, and established an associated inequality by use of properties of Csiszar's *f*-divergence. An improvement over this inequality has also been presented and comparisons of new mean divergence measures with some classical divergence measures are also made. Chen, Kar and Ralescu [6] have observed that divergence or cross entropy is a measure of the difference between two distribution functions and in order to deal with the divergence of uncertain variables via uncertainty distributions, the authors introduced the concept of cross entropy for uncertain variables based upon uncertain theory and investigated some mathematical properties of this concept. Today, it is well known that in different disciplines of science and engineering, the concept of distance has been proved to be very useful but its application areas can be extended to other emerging disciplines of social, economic, physical and biological sciences by the modification of the concept of distance. To explain the necessity for a new concept of distance for the disciplines other than that of science and engineering, we consider the following typical problems which are usually encountered in these emerging fields:

1. Find the distance between income distributions in two countries whose proportions of persons in n income groups are $(p_1, p_2, ..., p_n)$ and $(q_1, q_2, ..., q_n)$.

2. Find the distance between balance sheets of two companies or between balance sheets of same company in two different years.
3. Find a measure for the improvement in income distribution when income distribution $(q_1, q_2, ..., q_n)$ is changed to $(r_1, r_2, ..., r_n)$ by some government measures when the ideal distribution is assumed to be $(p_1, p_2, ..., p_n)$.
4. Find a measure for comparing the distribution of industry or poverty in different regions of a country.
5. Find a measure for comparing the distribution of voters among different political parties in two successive general elections.

Taking into consideration the above mentioned problems, we now introduce some new parametric and non-parametric measures of directed divergence by considering the following set of all complete finite discrete probability distributions:

$$\Omega_n = \left\{ P = (p_1, p_2, \dots, p_n) : p_i > 0, \sum_{i=1}^{n} p_i = 1 \right\}, n \geq 2$$

New parametric measures of directed divergence

I. For $P, Q \in \Omega_n$, we propose a new measure of symmetric directed divergence given by

$$D_1(P;Q) = \sum_{i=1}^{n} \left(\frac{p_i^2}{q_i} + \frac{q_i^2}{p_i} - 2q_i \right). \tag{1.4.7}$$

Some of the important properties of this directed divergence are:

(1) $D_1(P;Q)$ is a continuous function of $p_1, p_2, \dots, p_n$ and of $q_1, q_2, \dots, q_n$.

(2) $D_1(P;Q)$ is symmetric with respect to P and Q.

(3) $D_1(P;Q) \geq 0$ and vanishes if and only if $P = Q$.

(4) We can deduce from condition (3) that the minimum value of $D_1(P;Q)$ is zero.

(5) We shall now prove that $D_1(P;Q)$ is a convex function of both P and Q. This result is important in establishing the property of global minimum.

Let $D_1(P;Q) = f(p_1, p_2, \dots, p_n; q_1, q_2, \dots, q_n) = \sum_{i=1}^{n} \left(\frac{p_i^2}{q_i} + \frac{q_i^2}{p_i} - 2q_i \right)$

Thus $\frac{\partial f}{\partial p_i} = \frac{2p_i}{q_i} - \frac{q_i^2}{p_i^2}$, $\frac{\partial^2 f}{\partial p_i^2} = \frac{2}{q_i} + \frac{2q_i^2}{p_i^3}$ $\forall i = 1, 2, \dots, n$, and $\dfrac{\partial^2 f}{\partial p_i \partial p_j} = 0$ for $i, j = 1,2,\ldots,n; i \neq j$

Similarly,

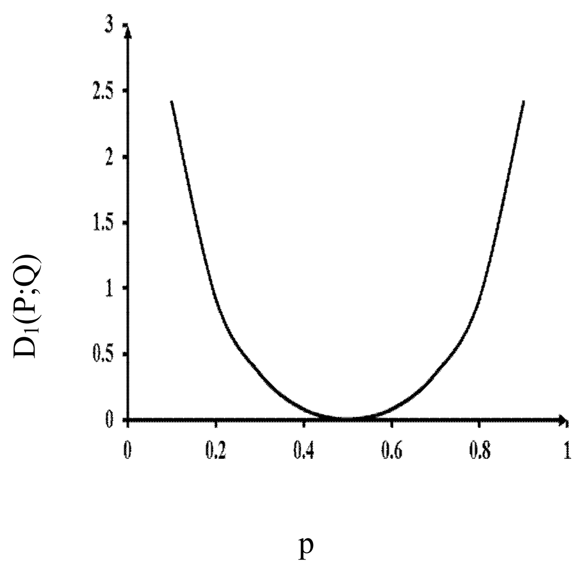
$$\frac{\partial f}{\partial q_i} = -\frac{p_i^2}{q_i^2} + \frac{2q_i}{p_i} - 2\,,\ \frac{\partial^2 f}{\partial q_i^2} = 2\frac{p_i^2}{q_i^3} + \frac{2}{p_i} \quad \forall\, i = 1,2,\ldots,n$$

and $\dfrac{\partial^2 f}{\partial q_i \partial q_j} = 0$ for $i, j = 1,2,\ldots,n; i \neq j$

Hence, the Hessian matrix of second order partial derivatives of f with respect to $p_1, p_2, \ldots, p_n$ is given by

$$\begin{bmatrix} \frac{2}{q_1} + \frac{2q_1^2}{p_1^3} & 0 & \cdots & 0 \\ 0 & \frac{2}{q_2} + \frac{2q_2^2}{p_2^3} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \frac{2}{q_n} + \frac{2q_n^2}{p_n^3} \end{bmatrix}$$

which is positive definite. A similar result is also true for the second order partial derivatives of f with respect to $q_1, q_2, \ldots, q_n$. Thus, we conclude that $D_1(P;Q)$ is a convex function of both $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$.



<u>Fig.-1.4.1</u>: Convexity of $D_1(P;Q)$ with respect to P.

Moreover, with the help of numerical data we have presented $D_1(P;Q)$ as shown in the Fig.-1.4.1 which clearly shows that $D_1(P;Q)$ is convex.

Under the above conditions, the function $D_1(P;Q)$ proposed above is a valid measure of symmetric directed divergence.

II. For any $P,Q \in \Omega_n$, we propose a new parametric measure of cross-entropy given by the following expression:

$$D_\alpha(P;Q) = \frac{\sum_{i=1}^{n} p_i \alpha^{\log \frac{p_i}{q_i}} - 1}{\alpha - 1}, \alpha > \frac{1}{2}, \alpha \neq 1. \qquad (1.4.8)$$

where α is a real parameter.

To prove that (1.4.8) is an appropriate measure of cross-entropy, we study its following properties:

(1) Clearly, $D_\alpha(P;Q)$ is a continuous function of $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$.

(2) We shall prove that $D_\alpha(P;Q)$ is a convex function of both P and Q.

Let $D_\alpha(P;Q) = f(p_1, p_2, \ldots, p_n; q_1, q_2, \ldots, q_n) = \dfrac{\sum_{i=1}^{n} p_i \alpha^{\log \frac{p_i}{q_i}} - 1}{\alpha - 1}$.

Thus $\dfrac{\partial f}{\partial p_i} = \dfrac{(1+\log\alpha)\alpha^{\log \frac{p_i}{q_i}}}{\alpha - 1}$,

$$\frac{\partial^2 f}{\partial p_i^2} = \frac{\log\alpha(1+\log\alpha)\alpha^{\log \frac{p_i}{q_i}}}{(\alpha-1)p_i}, \ \forall\ i = 1,2,\ldots,n,$$

and

$$\frac{\partial^2 f}{\partial p_i \partial p_j} = 0 \ \forall\ i, j = 1,2,\ldots,n;\ i \neq j.$$

Similarly,

$$\frac{\partial f}{\partial q_i} = -\frac{\log\alpha\ \alpha^{\log \frac{p_i}{q_i}} p_i}{(\alpha-1)q_i}, \ \frac{\partial^2 f}{\partial q_i^2} = \frac{\log\alpha(1+\log\alpha)\alpha^{\log \frac{p_i}{q_i}} p_i}{(\alpha-1)q_i^2} \quad \forall i = 1,2,\ldots,n,$$

and

$$\frac{\partial^2 f}{\partial q_i \partial q_j} = 0 \ \forall i, j = 1,2,\ldots,n;\ i \neq j.$$

Hence, the Hessian matrix of the second order partial derivatives of f with respect to $p_1, p_2, \ldots, p_n$ is given by

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

which is positive definite. Similarly, we can prove that the Hessian matrix of second order partial derivatives of f with respect to $q_1, q_2, \ldots, q_n$ is positive definite. Thus, we conclude that $D_\alpha(P;Q)$ is a convex function of both $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$. Moreover, with the help of numerical data we have presented $D_\alpha(P;Q)$ as shown in the Fig-1.4.2.

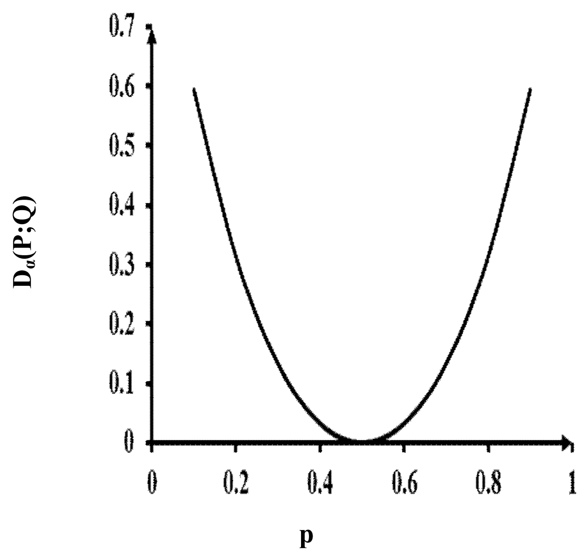


Fig.-1.4.2: Convexity of $D_\alpha(P;Q)$ with respect to P.

(3) Since P and Q are two different probability distributions, their difference or discrepancy or distance will be minimum only if $q_i = p_i$. Thus, for $q_i = p_i$, equation (1.4.8) gives $D_\alpha(P;Q) = 0$.

Moreover, since $D_\alpha(P;Q)$ is a convex function and its minimum value is zero, we must have $D_\alpha(P;Q) \geq 0$.

Under the above conditions, the function $D_\alpha(P;Q)$ is a valid parametric measure of cross-entropy.

Note: We have

$$\lim_{\alpha \to 1} D_\alpha(P;Q) = \lim_{\alpha \to 1} \frac{\sum_{i=1}^{n} p_i \alpha^{\log \frac{p_i}{q_i}} - 1}{\alpha - 1} = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$, which is Kullback-Leibler's [17] measure of cross-entropy.

Thus, $D_\alpha(P;Q)$ is a generalized measure of cross-entropy.

Concluding Remarks: In the existing literature of information theory, we find many probabilistic, weighted, parametric and non-parametric measures of entropy and directed divergence, each with its own merits, demerits and limitations. It has been observed that generalized measures of entropy and divergence should be introduced because upon optimization, these measures lead to useful probability distributions and mathematical models in various disciplines and also introduce flexibility in the system. Moreover, there exist a variety of innovation models, diffusion models and a large number of models applicable in economics, social sciences, biology and even in physical sciences, for each of which, a single measure of entropy or divergence cannot be adequate from application point of view. Thus, we need a variety of generalized parametric measures of information to extend the scope of their applications. But, one should be interested to develop only those measures which can be successfully applied to various disciplines of mathematical, social, biological and engineering sciences. Keeping this idea in mind, we have developed only those measures which find their applications in next chapters of the present book.

REFERENCES

[1] Aczel, J., & Daroczy, Z. (1963). Characterisierung der entropien positiver ordnung under Shannonschen entropie. *Acta Mathematica Hungarica, 14*, 95-121.

[2] Belis, M., & Guiasu, S. (1968). A quantitative-qualitative measure of information in cybernetic systems. *IEEE Transactions on Information Theory, 14*, 593-594.

[3] Brissaud, J. B. (2005). The meaning of entropy. *Entropy, 7*, 68-96.

[4] Burgin, M. (2003). Information theory: A multifaceted model of Information. *Entropy, 5*, 147-160.

[5] Cai, H., Kulkarni, S., & Verdu, S. (2006). Universal divergence estimation for finite-alphabet sources. *IEEE Transactions on Information Theory, 52*, 3456-3475.

[6] Chen, X., Kar, S., & Ralescu, D. A. (2012). Cross-entropy measure of uncertain variables. *Information Sciences, 201*, 53-60.

[7] Dehmer, M., & Mowshowitz, A. (2011). A history of graph entropy measures. *Information Sciences, 181*(1), 57-78.

[8] Eguchi, S., & Kato, S. (2010). Entropy and Divergence Associated with Power Function and the Statistical Application. *Entropy, 12*(2), 262-274.

[9] Ferreri, C. (1980). Hypoentropy and related heterogeneity divergence measures. *Statistica, 40*, 155-168.

[10] Garrido, A. (2011). Classifying entropy measures. *Symmetry, 3*(3), 487-502.

[11] Guiasu, S. (1971). Weighted entropy. *Reports on Mathematical Physics, 2*, 165-171.

[12] Havrada, J. H., & Charvat, F. (1967). Quantification methods of classification process: Concept of structural α-entropy. *Kybernetika, 3*, 30-35.

[13] Hillion, E. (2012). Concavity of entropy along binomial convolutions. *Electronic Communications in Probability, 17*(4), 1-9.

[14] Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Review, 106,* 620-630.

[15] Kapur, J. N. (1986). Four families of measures of entropy. *Indian Journal of Pure and Applied Mathematics, 17*, 429-449.

[16] Kapur, J. N. (1995). *Measures of information and their applications*. New York: Wiley Eastern.

[17] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*, 79-86.

[18] Parkash, O., Mukesh, &.Thukral, A. K (2010). Measuring randomness of association through coefficient of non-determination. In O. Parkash (Ed.), *Advances in Information Theory and Operations Research* (pp. 46-55). Germany: VDM Verlag.

[19] Parkash, O., & Singh, R. S. (1987). On characterization of useful information theoretic measures. *Kybernetika, 23*(3), 245-251.

[20] Parkash, O., Singh, Y. B., & Sunita (1996). Entropy and internal information of printed Punjabi. *International Journal of Management and Systems*, *12*(1), 25-32.

[21] Parkash, O., & Taneja, H. C. (1986).Characterization of quantitative-qualitative measure of inaccuracy for discrete generalized probability distributions. *Communications in Statistics-Theory and Methods, 15*(12), 3763-3772.

[22] Parkash, O., & Thukral, A. K. (2010). Statistical measures as measures of diversity. *International Journal of Biomathematics*, *3*(2), 173-185.

[23] Renyi, A. (1961). On measures of entropy and information. *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability* (pp.547-561).

[24] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-659.

[25] Sharma, B. D., & Taneja, I. J. (1975). Entropies of type (α,β) and other generalized measures of information theory. *Metrika, 22*, 202-215.

[26] Taneja, I. J. (2008). On the divergence measures. *Advances in Inequalities Theory and Statistics,* 169-186.

[27] Zadeh, L. A. (2005). Towards a generalized theory of uncertainty (GTU)- an outline. *Information Sciences, 172*, 1-40.

[28] Zyczkowski, K. (2003). Renyi extrapolation of Shannon entropy. *Open Systems and Information Dynamics, 10*, 297-310.

CHAPTER-II

GENERALIZED MEASURES OF FUZZY ENTROPY-THEIR PROPERTIES AND CONTRIBUTION

ABSTRACT

In the existing literature of theory of fuzzy information, there is a huge availability of parametric and non-parametric measures of fuzzy entropy each with its own merits and limitations. Keeping in view the flexibility in the system and the application areas, some new generalized parametric measures of fuzzy entropy have been proposed and their essential and desirable properties have been studied. Moreover, we have measured the partial information about a fuzzy set when only partial knowledge of fuzzy values is given. The generating functions for various measures of fuzzy entropy have been obtained for fuzzy distributions. The necessity for normalizing measures of fuzzy entropy has been discussed and some normalized measures of fuzzy entropy have been obtained.

Keywords: ***Uncertainty, Fuzzy set, Fuzzy entropy, Concave function, Monotonicity, Partial information, Generating function.***

2.1 INTRODUCTION

It is often seen that in real life situations, uncertainty arises in decision-making problem and this uncertainty is either due to lack of knowledge or due to inherent vagueness. Such types of problems can be solved using probability theory and fuzzy set theory respectively. Fuzziness, which is a feature of imperfect information, results from the lack of crisp distinction between the elements belonging and not belonging to a set. However, the two functions measure fundamentally different types of uncertainty. Basically, the Shannon's [19] entropy measures the average uncertainty in bits associated with the prediction of outcomes in a random experiment.

The concept of fuzzy entropy which is peculiar to mathematics, information theory, computer science, and other branches of mathematical sciences tracing to the fuzzy set theory of Iranian-born American electrical engineer and computer scientist Lotfi Zadeh, who extended the Shannon's [19] entropy theory to be applied as a fuzzy entropy of a fuzzy subset for a finite set, is the entropy of a fuzzy set, loosely representing the information of uncertainty. After the introduction of the theory of fuzzy sets, it received recognition from different quarters and a considerable body of literature blossomed around the concept of fuzzy sets. A fuzzy set "A" is a subset of universe of discourse U, characterized by a membership function $\mu_A(x)$ which associates to each $x \in$U, a membership value from [0, 1], that is, $\mu_A(x)$ represents the grade of membership of x in "A". When $\mu_A(x)$ takes a value only 0 or 1, "A" reduces to a crisp or non fuzzy set and $\mu_A(x)$ represents the characteristic function of the set "A". The role of membership functions in probability measures of fuzzy sets has well been explained by Singpurwalla and Booker [21] whereas Zadeh [24] defined the entropy of a fuzzy event as weighted Shannon [19] entropy. Kapur [8] explained the concept of fuzzy uncertainty as follows:

The fuzzy uncertainty of grade x is a function of x with the following properties:

I. $f(x) = 0$ when x = 0 or 1.

II. $f(x)$ increases as x goes from 0 to 0.5

III. $f(x)$ decreases as x goes from 0.5 to 1.0

IV. $f(x) = f(1-x)$

It is desirable that $f(x)$ should be a continuous and differentiable function but not necessarily.

Some Definitions

Consider the vector $\{\mu_A(x_1), \mu_A(x_2), ..., \mu_A(x_n)\}$

If $\mu_A(x_i) = 0$ then the ith element certainly does not belong to set A and if $\mu_A(x_i) = 1$, it definitely belongs to set A. If $\mu_A(x_i) = 0.5$, there is maximum uncertainty whether x_i belongs to set A or not. A vector of the type in which every element lies between 0 and 1 and has the interpretation given above, is called a fuzzy vector and the set A is called a fuzzy set.

If every element of the set is 0 or 1, there is no uncertainty about it and the set is said to be a crisp set. Thus, there are 2^n crisp sets with n elements and infinity of fuzzy sets with n elements. If some elements are 0 and 1 and others lie between 0 and 1, the set will still said to be a fuzzy set.

With the ith element, we associate a fuzzy uncertainty $f(\mu_A(x_i))$, when $f(x)$ is any function which has the four properties mentioned above. If the n elements are independent, the total fuzzy uncertainty is given by the following expression:

$$\sum_{i=1}^{n} f\{\mu_A(x_i)\}.$$

This fuzzy uncertainty is very often called fuzzy entropy.

Keeping in view the fundamental properties which a valid measure of fuzzy entropy should satisfy, a large number of measures of fuzzy entropy were discussed, investigated, characterized and generalized by various authors, each with its own objectives, merits and limitations.

Thus, De Luca and Termini [2] suggested that corresponding to Shannon's [19] probabilistic entropy, the measure of fuzzy entropy should be:

$$H(A)=-\sum_{i=1}^{n}\left[\mu_A(x_i)\log\mu_A(x_i)+(1-\mu_A(x_i))\log(1-\mu_A(x_i))\right] \tag{2.1.1}$$

Bhandari and Pal [1] developed the following measure of fuzzy entropy corresponding to Renyi's [17] probabilistic entropy:

$$H_\alpha(A)=\frac{1}{1-\alpha}\sum_{i=1}^{n}\log[\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}],\ \alpha\neq 1,\alpha>0 \tag{2.1.2}$$

Kapur [8] took the following measure of fuzzy entropy corresponding to Havrada and Charvat's [6] probabilistic entropy:

$$H^{\alpha}(A)=(1-\alpha)^{-1}\sum_{i=1}^{n}\left[\left\{\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right\}-1\right],\alpha\neq 1,\alpha>0 \tag{2.1.3}$$

Parkash [13] introduced a new measure of fuzzy entropy involving two real parameters, given by

$$H_\alpha^{\beta}(A)=[(1-\alpha)\,\beta]^{-1}\sum_{i=1}^{n}\left[\left\{\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right\}^{\beta}-1\right],\alpha\neq 1,\alpha>0,\beta\neq 0 \tag{2.1.4}$$

and called it (α,β) fuzzy entropy which includes some well known measures of fuzzy entropy.

Rudas [18] discussed elementary entropy function based upon fuzzy entropy construction and fuzzy operations and also defined new generalized operations as minimum and maximum entropy operations. Hu and Yu [7] presented an interpretation of Yager's [22] entropy from the point of view of the discernibility power of a relation and consequently, based upon Yager's entropy, generated some basic definitions in Shannon's information theory. The authors introduced the definitions of joint entropy, conditional entropy, mutual entropy and relative entropy to compute the information changes for fuzzy indiscernibility relation operations. Moreover, for measuring the information increment, the authors proposed the concepts of conditional entropy and relative conditional entropy. Yang and Xin [23] introduced the axiom definition of information entropy of vague sets which is different from the ones in the existing literature and revealed the connection between the notions of entropy for vague sets and fuzzy sets and also provided the applications of the new information entropy measures for vague sets to pattern recognition and medical diagnosis.

Guo and Xin [5] studied some new generalized entropy formulas for fuzzy sets and proposed new idea for the further development of the subject. After attaching the weights to the probability distribution, Parkash, Sharma and Mahajan [15, 16] delivered the applications of weighted measures of fuzzy entropy for the study of maximum entropy principles. Osman, Abdel-Fadel, El-Sersy and Ibraheem [11] have provided the applications of fuzzy parametric measures to programming problems. Li and Liu [10] have extended the applications of fuzzy variables towards maximum entropy principle and proved that the exponentially distributed fuzzy variable has the maximum entropy when the non-negativity is assumed and the expected value is given in advance whereas the normally distributed fuzzy variable has the maximum entropy when the variance is prescribed.

Some other developments, proposals and other interesting findings related with the theoretical measures of fuzzy entropy and their applications have been investigated by Ebanks [3], Klir and Folger [9], Kapur [8], Pal and Bezdek [12], Hu and Yu [7], Parkash and Sharma [14], Singh and Tomar [20] etc. In section 2, we have introduced some new generalized measures of fuzzy entropy and studied their detailed properties.

2.2 NEW MEASURES OF FUZZY ENTROPY FOR DISCRETE FUZZY DISTRIBUTIONS

In this section, we propose some new generalized measures of fuzzy entropy. These measures are given by the following mathematical expressions:

I) $$H_1^\alpha(A) = -\frac{1}{\alpha}\left\{\sum_{i=1}^{n}\left[\mu_A^{\alpha\mu_A(x_i)}(x_i) + (1-\mu_A(x_i))^{\alpha(1-\mu_A(x_i))} - 2\right]\right\}; \alpha > 0 \tag{2.2.1}$$

Under the assumption $0^{0.\alpha} = 1$, we study the following properties:

(i) When $\mu_A(x_i) = 0\, or\, 1$, $H_1^\alpha(A) = 0$

(ii) When $\mu_A(x_i) = \frac{1}{2}$, $H_1^\alpha(A) = \dfrac{2n\left(2^{\frac{\alpha}{2}} - 1\right)}{\alpha 2^{\frac{\alpha}{2}}}$

Hence, $H_1^\alpha(A)$ is an increasing function of $\mu_A(x_i)$ for $0 \le \mu_A(x_i) \le \frac{1}{2}$

(iii) $H_1^\alpha(A)$ is a decreasing function of $\mu_A(x_i)$ for $\frac{1}{2} \le \mu_A(x_i) \le 1$

(iv) $H_1^\alpha(A) \ge 0$

(v) $H_1^\alpha(A)$ does not change when $\mu_A(x_i)$ is changed to $(1-\mu_A(x_i))$.

(vi) **Concavity:** To verify that the proposed measure is concave, we proceed as follows:

We have

$$\frac{\partial H_1^{\alpha}(A)}{\partial \mu_A(x_i)} = -\left[\mu_A^{\alpha\mu_A(x_i)} \left\{1+\log \mu_A(x_i)\right\} - \left\{1-\mu_A(x_i)\right\}^{\alpha(1-\mu_A(x_i))} \left\{1+\log(1-\mu_A(x_i))\right\} \right]$$

Also

$$\frac{\partial^2 H_1^{\alpha}(A)}{\partial \mu_A^2(x_i)} = -\left[\begin{array}{l} \mu_A^{\alpha\mu_A(x_i)-1} + \mu_A^{\alpha\mu_A(x_i)}\alpha \left\{1+\log \mu_A(x_i)\right\}^2 + \left\{1-\mu_A(x_i)\right\}^{\alpha(1-\mu_A(x_i))-1} \\ \\ +\alpha \left\{1-\mu_A(x_i)\right\}^{\alpha(1-\mu_A(x_i))} \left\{1+\log(1-\mu_A(x_i))\right\}^2 \end{array} \right] \le 0$$

Hence, $H_1^{\alpha}(A)$ is a concave function of $\mu_A(x_i)$.

Under the above conditions, the generalized measure proposed in equation (2.2.1) is a valid measure of fuzzy entropy. Next, we have presented the generalized measure $H_1^{\alpha}(A)$ graphically in Fig.-2.2.1 which clearly shows that the fuzzy entropy is a concave function.

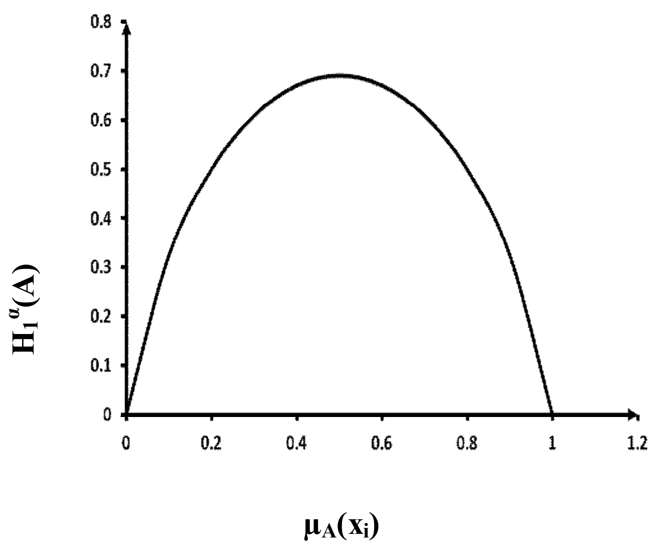


Fig.-2.2.1 Concavity of $H_1^{\alpha}(A)$ with respect to $\mu_A(x_i)$

Next, we study a more desirable property of the generalized measure of fuzzy entropy $H_1^{\alpha}(A)$.

Maximum value

Taking $\frac{\partial H_1^{\alpha}(A)}{\partial \mu_A(x_i)} = 0$, we get $\mu_A(x_i) = \frac{1}{2}$

Also $\left[\dfrac{\partial^2 H_1^\alpha(A)}{\partial \mu_A^2(x_i)}\right]_{\mu_A(x_i)=\frac{1}{2}} = -\left[2^{2-\frac{\alpha}{2}} + \alpha.2^{1-\frac{\alpha}{2}}\left(1-\log 2\right)^2\right] < 0$

Hence, the maximum value exists at $\mu_A(x_i) = \dfrac{1}{2}$. If we denote the maximum value by $f\left(n\right)$, we get

$$f(n) = \frac{2n}{\alpha}\left(1-2^{-\frac{\alpha}{2}}\right)$$

Also $f^{'}(n) = \dfrac{2}{\alpha}\left(1-2^{-\frac{\alpha}{2}}\right) > 0$

Thus, the maximum value is an increasing function of n and this result is most desirable.

II) $H_1^\alpha(A;W) = -\dfrac{1}{\alpha}\left\{\sum_{i=1}^{n} w_i \left[\begin{array}{l}\mu_A^{\alpha\mu_A(x_i)} + \left\{1-\mu_A(x_i)\right\}^{\alpha(1-\mu_A(x_i))} - 2 + \mu_A(x_i)\log\mu_A(x_i) \\ +\left\{1-\mu_A(x_i)\right\}\log\left\{1-\mu_A(x_i)\right\}\end{array}\right]\right\}; w_i \geq 0, \alpha > 0$ (2.2.2)

Under the assumption that $0^{0.\alpha} = 1$, we study the following properties

(i) When $\mu_A(x_i) = 0\, or\, 1$, $H_1^\alpha(A;W) = 0$

(ii) When $\mu_A(x_i) = \dfrac{1}{2}$,

$$H_1^\alpha(A;W) = \left\{\frac{2\left(2^{\frac{\alpha}{2}}-1\right)}{\alpha.2^{\frac{\alpha}{2}}} + \frac{1}{\alpha}\log 2\right\}\sum_{i=1}^{n} w_i > 0$$

Hence, $H_1^\alpha(A;W)$ is an increasing function of $\mu_A(x_i)$ for $0 \leq \mu_A(x_i) \leq \dfrac{1}{2}$

(iii) $H_1^\alpha(A;W)$ is a decreasing function of $\mu_A(x_i)$ for $\dfrac{1}{2} \leq \mu_A(x_i) \leq 1$

(iv) $H_1^\alpha(A;W) \geq 0$

(v) $H_1^\alpha(A;W)$ does not change when $\mu_A(x_i)$ is changed to $(1-\mu_A(x_i))$.

(vi) **Concavity**: To verify that the proposed weighted measure is concave, we proceed as follows:

We have

$$\frac{\partial H_1^\alpha(A;W)}{\partial \mu_A(x_i)} = -\frac{1}{\alpha} w_i \left[\begin{array}{l} \mu_A^{\alpha\mu_A(x_i)} \alpha \{1+\log \mu_A(x_i)\} - \{1-\mu_A(x_i)\}^{\alpha(1-\mu_A(x_i))} \alpha \{1+\log(1-\mu_A(x_i))\} \\ + \log \mu_A(x_i) - \log(1-\mu_A(x_i)) \end{array} \right]$$

Also

$$\frac{\partial^2 H_1^\alpha(A;W)}{\partial \mu_A^2(x_i)} = -\frac{1}{\alpha} w_i \left[\begin{array}{l} \alpha \mu_A^{\alpha\mu_A(x_i)-1} + \alpha^2 \mu_A^{\alpha\mu_A(x_i)} \{1+\log \mu_A(x_i)\}^2 + \alpha \{1-\mu_A(x_i)\}^{\alpha(1-\mu_A(x_i))-1} \\ +\alpha^2 \{1-\mu_A(x_i)\}^{\alpha(1-\mu_A(x_i))} \{1+\log(1-\mu_A(x_i))\}^2 + \dfrac{1}{\mu_A(x_i)} + \dfrac{1}{1-\mu_A(x_i)} \end{array} \right] \le 0$$

Hence, $H_1^\alpha(A;W)$ is a concave function of $\mu_A(x_i)$.

Under the above conditions, the generalized measure proposed in equation (2.2.2) is a valid measure of weighted fuzzy entropy. Next, we have presented the fuzzy entropy $H_1^\alpha(A;W)$ in Fig.-2.2.2 which clearly shows that the fuzzy entropy is a concave function.

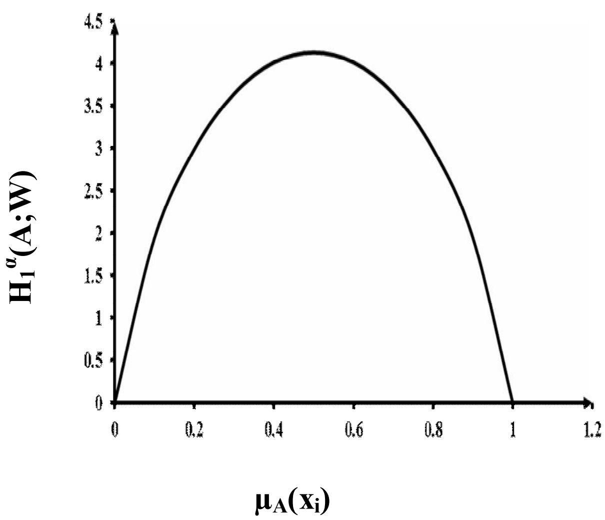


Fig.-2.2.2 Concavity of $H_1^\alpha(A;W)$ with respect to $\mu_A(x_i)$

Moreover, we have studied the following desirable property of the generalized measure of fuzzy entropy:

Maximum value

Taking $\dfrac{\partial H_1^\alpha(A;W)}{\partial \mu_A(x_i)} = 0$, we get $\mu_A(x_i) = \dfrac{1}{2}$

Also $\left[\dfrac{\partial^2 H_1^\alpha(A;W)}{\partial \mu_A^2(x_i)}\right]_{\mu_A(x_i)=\frac{1}{2}} < 0$

Hence, the maximum value exists at $\mu_A(x_i) = \dfrac{1}{2}$ and is given by

$$Max.H_1^\alpha(A;W) = \left\{\frac{2}{\alpha}\left(1-2^{-\frac{\alpha}{2}}\right)+\frac{\log 2}{\alpha}\right\}\sum_{i=1}^{n} w_i \tag{2.2.3}$$

Note: If we ignore the weights and denote the maximum value of the fuzzy entropy by $f(n)$, then equation (2.2.3) becomes

$$f(n) = n\left[\frac{2}{\alpha}\left(1-2^{-\frac{\alpha}{2}}\right)+\frac{\log 2}{\alpha}\right]$$

Then

$$f^{'}(n) = \left[\frac{2}{\alpha}\left(1-2^{-\frac{\alpha}{2}}\right)+\frac{\log 2}{\alpha}\right] > 0$$

Thus, the maximum value is an increasing function of *n* and this result is most desirable.

2.3 MONOTONIC CHARACTER OF NEW MEASURES OF FUZZY ENTROPY

In this section, we discuss the monotonic character of the measures of fuzzy entropy proposed in the above section.

I. From equation (2.2.1), we have

$$\frac{dH_1^\alpha(A)}{d\alpha} = -\frac{1}{\alpha}\left\{\sum_{i=1}^{n}\left[\mu_A^{\alpha\mu_A(x_i)+1}(x_i)\log\mu_A(x_i)+(1-\mu_A(x_i))^{\alpha(1-\mu_A(x_i))+1}\log(1-\mu_A(x_i))\right]\right\}$$
$$+\frac{1}{\alpha^2}\sum_{i=1}^{n}\left\{\mu_A^{\alpha\mu_A(x_i)}(x_i)+(1-\mu_A(x_i))^{\alpha(1-\mu_A(x_i))}-2\right\}$$

or

$$\alpha^2\frac{dH_1^\alpha(A)}{d\alpha} = -\alpha\left\{\sum_{i=1}^{n}\left[\mu_A^{\alpha\mu_A(x_i)+1}(x_i)\log\mu_A(x_i)+(1-\mu_A(x_i))^{\alpha(1-\mu_A(x_i))+1}\log(1-\mu_A(x_i))\right]\right\}$$
$$+\sum_{i=1}^{n}\left\{\mu_A^{\alpha\mu_A(x_i)}(x_i)+(1-\mu_A(x_i))^{\alpha(1-\mu_A(x_i))}-2\right\}$$

$$=\left\{\sum_{i=1}^{n}\left[\begin{array}{l}\mu_A^{\alpha\mu_A(x_i)}(x_i)\left\{1-\alpha\mu_A(x_i)\log\mu_A(x_i)\right\}\\ \\ +(1-\mu_A(x_i))^{\alpha(1-\mu_A(x_i))}\left\{1-\alpha(1-\mu_A(x_i))\log(1-\mu_A(x_i))\right\}-2\end{array}\right]\right\} \tag{2.3.1}$$

Let $f(x)=\left[x^{\alpha x}\left\{1-\alpha x\log x\right\}+(1-x)^{\alpha(1-x)}\left\{1-\alpha(1-x)\log(1-x)\right\}-2\right];\, x\le 1$

Also, we have the following inequality:

$$x-x\log x-1\le 0, \tag{2.3.2}$$

Rewriting equation (2.3.1) and applying the inequality (2.3.2), we get

$$f(x)=\left[\left\{x^{\alpha x}-x^{\alpha x}\log x^{\alpha x}-1\right\}+\left\{(1-x)^{\alpha(1-x)}-(1-x)^{\alpha(1-x)}\log(1-x)^{\alpha(1-x)}-1\right\}\right]\le 0$$

Hence equation (2.3.1) gives that $\dfrac{dH_1^{\alpha}(A)}{d\alpha}\le 0$

Thus, $H_1^{\alpha}(A)$ is a monotonically decreasing function $\forall\alpha$.

II. From equation (2.2.2), we have

$$\frac{dH_1^{\alpha}(A;W)}{d\alpha}=-\frac{1}{\alpha}\left\{\sum_{i=1}^{n}w_i\left[\mu_A^{\alpha\mu_A(x_i)+1}(x_i)\log\mu_A(x_i)+\left\{1-\mu_A(x_i)\right\}^{\alpha(1-\mu_A(x_i))+1}\log\left\{1-\mu_A(x_i)\right\}\right]\right\}$$

$$+\frac{1}{\alpha^2}\left\{\sum_{i=1}^{n}w_i\left[\begin{array}{l}\mu_A^{\alpha\mu(x_i)}(x_i)+\left\{1-\mu_A(x_i)\right\}^{\alpha(1-\mu_A(x_i))}-2\\ \\ \quad+\mu_A(x_i)\log\mu_A(x_i)+\left\{1-\mu_A(x_i)\right\}\log\left\{1-\mu_A(x_i)\right\}\end{array}\right]\right\}$$

or

$$\alpha^2\frac{dH_1^{\alpha}(A;W)}{d\alpha}=\left\{\sum_{i=1}^{n}w_i\left[\begin{array}{l}\mu_A^{\alpha\mu_A(x_i)}(x_i)\left\{1-\alpha\mu_A(x_i)\log\mu_A(x_i)\right\}\\ \\ \quad+\left\{1-\mu_A(x_i)\right\}^{\alpha(1-\mu_A(x_i))}\left\{1-\alpha(1-\mu_A(x_i)\right\}\log\left\{1-\mu_A(x_i)\right\}\\ \\ \quad\quad+\mu_A(x_i)\log\mu_A(x_i)+\left\{1-\mu_A(x_i)\right\}\log\left\{1-\mu_A(x_i)\right\}-2\end{array}\right]\right\} \tag{2.3.3}$$

Taking $g(x)=x^{\alpha x}(1-\alpha x\log x)+(1-x)^{\alpha(1-x)}\{1-\alpha(1-x)\log(1-x)\}+x\log x+(1-x)\log(1-x)-2\,;x\le 1$

Rewriting the above equation, we get

$$g(x)=\begin{bmatrix}\{x^{\alpha x}-x^{\alpha x}\log x^{\alpha x}-1\}+\{(1-x)^{\alpha(1-x)}-(1-x)^{\alpha(1-x)}\log(1-x)^{\alpha(1-x)}-1\}\\ \\ +\{x\log x+(1-x)\log(1-x)\}\end{bmatrix} \tag{2.3.4}$$

Further, let $f(x)=x\log x+(1-x)\log(1-x)$, then

$f(x)=f(1-x)$ and $f^{'}(x)=-f^{'}(1-x)$

Also $f(0)=f(1)=0$. Since $f(x)$ is convex, we have

$f(x)\le 0$ for all $0\le x\le 1$ (2.3.5)

Applying (2.3.2) and (2.3.5), we get $g(x)\le 0$.

Hence equation (2.3.3) gives that $\dfrac{dH_1^{\alpha}(A;W)}{d\alpha}\le 0$

Thus, $H_1^{\alpha}(A;W)$ is a monotonic decreasing function of α.

Next, we have presented $H_1^{\alpha}(A;W)$ graphically and obtained the following figures 2.3.2 and 2.3.3.

Case-I: For $\alpha>1$, we have

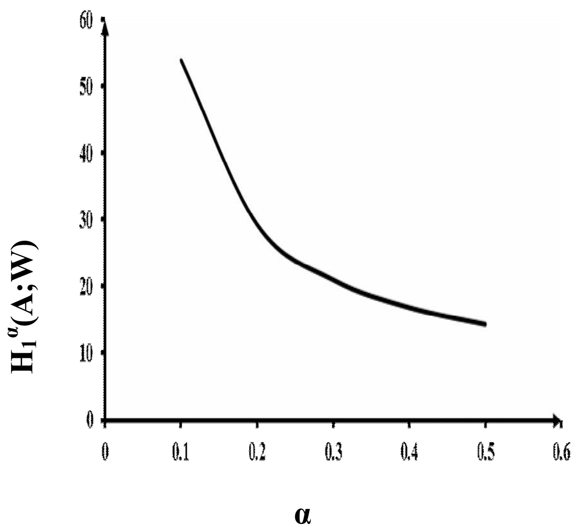


Fig.-2.3.2 Monotonicity $H_1^{\alpha}(A;W)$ for $\alpha>1$

Case-II: For $0<\alpha<1$, we have

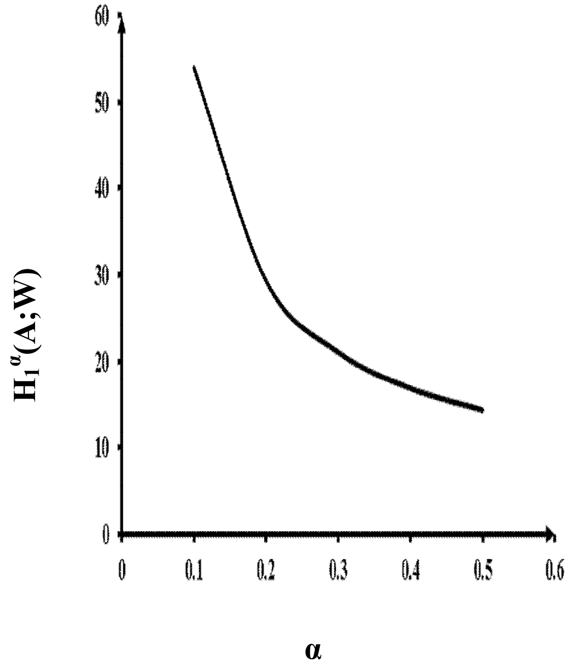


Fig.-2.3.3 Monotonicity $H_1^\alpha(A;W)$ for $0 < \alpha < 1$

The above figures clearly show that the generalized weighted fuzzy entropy $H_1^\alpha(A;W)$ is a monotonically decreasing function of α for each α.

2.4 PARTIAL INFORMATION ABOUT A FUZZY SET-A MEASUREMENT

We know that if we know all the fuzzy values $\mu_A(x_1), \mu_A(x_2), \mu_A(x_3), ..., \mu_A(x_n)$ of a fuzzy set A, then we have the complete knowledge or full information about the fuzzy set A. If we have some knowledge about these fuzzy values $\mu_A(x_1), \mu_A(x_2), \mu_A(x_3), ..., \mu_A(x_n)$, then this knowledge will not enable us to determine these fuzzy values uniquely. Thus, we have only partial information about the fuzzy set. We know that the information supplied by this incomplete knowledge is always less than the full information and this gives birth to a new problem that how to find a quantitative measure for this information which will show how less is this partial information. To solve this problem, we make use of the parametric measure of fuzzy entropy introduced in equation (2.2.1).

We know that the fuzzy entropy (2.2.1) is maximum when each $\mu_A(x_i) = \frac{1}{2}$, that is, when the fuzzy set A is most fuzzy set. Also, it is minimum when $\mu_A(x_i) = 0$ or 1, that is, when the set is crisp set. Thus, we see that the model (2.2.1) measures the fuzziness of the set A.

Now, if we do not know $\mu_A(x_1), \mu_A(x_2), \mu_A(x_3), ..., \mu_A(x_n)$, each of these can take any value between 0 and 1. In this case, the maximum value of the fuzzy entropy is given by

$$\left[H_1^{\alpha}(A)\right]_{\max} = \frac{2n}{\alpha}\left[1 - \frac{1}{2^{\alpha/2}}\right] \tag{2.4.1}$$

and the minimum entropy is given by

$$\left[H_1^{\alpha}(A)\right]_{\min} = 0 \tag{2.4.2}$$

Thus, the expression for fuzziness gap is given by

$$\text{Fuzziness gap} = \left[H_1^{\alpha}(A)\right]_{\max} - \left[H_1^{\alpha}(A)\right]_{\min} = \frac{2n}{\alpha}\left[1 - \frac{1}{2^{\alpha/2}}\right] \tag{2.4.3}$$

Now, if we are given some knowledge about the fuzzy values, it can only decrease $\left[H_1^{\alpha}(A)\right]_{\max}$ and increase $\left[H_1^{\alpha}(A)\right]_{\min}$. From this discussion, it is obvious that this knowledge reduces the fuzziness gap and this reduction is due to the information given by the partial knowledge. Thus, we conclude that the information given by the partial knowledge is the difference between fuzziness gap before and after we use this information. Thus, the information provided by the partial knowledge can be calculated and to calculate this information, we discuss the following cases:

Case-I: When one or more fuzzy values are known

Suppose only one fuzzy value $\mu_A(x_k)$ is known and all others are unknown. Then, we have

$$\left[H_1^{\alpha}(A)\right]_{\max} = -\frac{1}{\alpha}\left\{\mu_A^{\alpha\mu_A(x_k)}(x_k) + (1-\mu_A(x_k))^{\alpha(1-\mu_A(x_k))} - 2\right\} + \frac{2(n-1)}{\alpha}\left[1 - \frac{1}{2^{\alpha/2}}\right] \tag{2.4.4}$$

and

$$\left[H_1^{\alpha}(A)\right]_{\min} = -\frac{1}{\alpha}\left\{\mu_A^{\alpha\mu_A(x_k)}(x_k) + (1-\mu_A(x_k))^{\alpha(1-\mu_A(x_k))} - 2\right\} \tag{2.4.5}$$

From equations (2.4.4) and (2.4.5), we have

$$\left[H_1^{\alpha}(A)\right]_{\max} - \left[H_1^{\alpha}(A)\right]_{\min} = \frac{2(n-1)}{\alpha}\left[1 - \frac{1}{2^{\alpha/2}}\right] \tag{2.4.6}$$

Thus, from equations (2.4.3) and (2.4.6), we have

$$\text{Reduction in fuzziness gap} = \frac{2}{\alpha}\left[1 - \frac{1}{2^{\alpha/2}}\right]$$

Thus, the fuzzy information received by the partial knowledge, when only one fuzzy value is known, is given by the following equation:

$$I_1(A) = \frac{2}{\alpha}\left[1 - \frac{1}{2^{\alpha/2}}\right] \tag{2.4.7}$$

Case-II: When only a range of fuzzy values is known

Suppose we are given only that $\mu_A(x_k)$ lies between two fuzzy values a and b, and no other knowledge about the fuzzy values is given. Then, the maximum fuzzy information is given by

$$\left[H_1^\alpha(A)\right]_{\max} = \max_{a\le\mu_A(x_k)\le b}\left[-\frac{1}{\alpha}\left\{\mu_A^{\alpha\mu_A(x_k)}(x_k)+(1-\mu_A(x_k))^{\alpha(1-\mu_A(x_k))}-2\right\}\right]+\frac{2(n-1)}{\alpha}\left[1-\frac{1}{2^{\alpha/2}}\right] \tag{2.4.8}$$

and the minimum fuzzy information is given by

$$\left[H_1^\alpha(A)\right]_{\min} = \min_{a\le\mu_A(x_k)\le b}\left[-\frac{1}{\alpha}\left\{\mu_A^{\alpha\mu_A(x_k)}(x_k)+(1-\mu_A(x_k))^{\alpha(1-\mu_A(x_k))}-2\right\}\right] \tag{2.4.9}$$

Thus fuzziness gap $=\dfrac{2(n-1)}{\alpha}\left[1-\dfrac{1}{2^{\alpha/2}}\right]+\left\{H_1(A)-H_2(A)\right\}$ (2.4.10)

where

$$H_1(A)=\max_{a\le\mu_A(x_k)\le b}\left[-\frac{1}{\alpha}\left\{\mu_A^{\alpha\mu_A(x_k)}(x_k)+(1-\mu_A(x_k))^{\alpha(1-\mu_A(x_k))}-2\right\}\right]$$

and

$$H_2(A)=\min_{a\le\mu_A(x_k)\le b}\left[-\frac{1}{\alpha}\left\{\mu_A^{\alpha\mu_A(x_k)}(x_k)+(1-\mu_A(x_k))^{\alpha(1-\mu_A(x_k))}-2\right\}\right]$$

Now, as shown in Fig.-2.4.1, the following three cases arise:

(a) If $a\le 0.5$ and $b\le 0.5$, then the maximum value arises when $\mu_A(x_k)=b$ and the minimum value arises when $\mu_A(x_k)=a$.

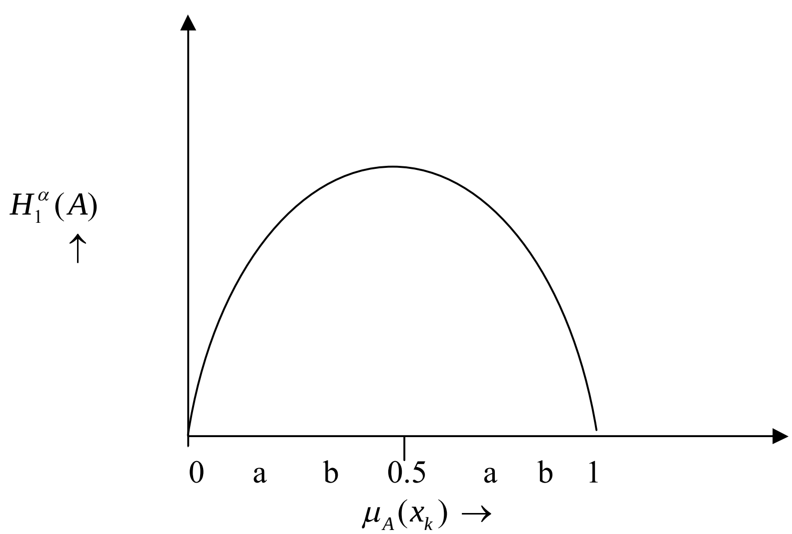


Fig.-2.4.1 Presentation of $H_1^\alpha(A)$ for optimality

(b) If $a \geq 0.5$ and $b \geq 0.5$, then the maximum value arises when $\mu_A(x_k) = a$ and the minimum value arises when $\mu_A(x_k) = b$.

(c) If $a < 0.5$ and $b > 0.5$, and also $|a - b| < 0.5$, then

(i) The maximum occurs at a if $a > 1-b$ and at b if $a < 1-b$.

(ii) The minimum occurs at a if $b < 1-a$ and at b if $b > 1-a$.

Thus, we see that in every case, the optimum values can be calculated and consequently, the fuzziness gap can be obtained.

Example: For $n = 10$ and $\alpha = 2$, let $\mu_A(x_k)$ lies between 0.1 and 0.7, then maximum will take place at $\mu_A(x_k) = 0.7$ and the minimum will take place at $\mu_A(x_k) = 0.1$.

Thus $\left[H_1^\alpha(A)\right]_{\max} = 4.95375$ and

$\left[H_1^\alpha(A)\right]_{\min} = 0.27095$

Thus, the fuzziness gap $= 4.6828$

Hence, we see that the fuzzy information received by the partial knowledge, when only a range of fuzzy values is known, is given by the following equation:

$$I_2(A) = \frac{2n}{\alpha}\left[1 - \frac{1}{2^{\alpha/2}}\right] - 4.68345 = 0.3172$$

whereas

$$I_1(A) = \frac{2}{\alpha}\left[1 - \frac{1}{2^{\alpha/2}}\right] = 0.5$$

Thus $I_2(A) < I_1(A)$

Hence, we conclude that the information supplied by the partial knowledge of the range of fuzzy value $\mu_A(x_k)$ is always less than the information supplied by the exact value of $\mu_A(x_k)$ and this result finds total compatibility with real life problems.

Case-III: When a relation is specified between the fuzzy values

Suppose that the fuzzy values are connected by the following relation:

$$\mu_A(x_1) + \mu_A(x_2) + \mu_A(x_3) + \ldots + \mu_A(x_n) = k \tag{2.4.11}$$

In this case, we see that, maximum value arises when

$$\mu_A(x_1) = \mu_A(x_2) = \mu_A(x_3) = \ldots = \mu_A(x_n) = \frac{k}{n} \tag{2.4.12}$$

and this maximum value is given by

$$\left[H_1^{\alpha}(A)\right]_{\max} = -\frac{n}{\alpha}\left[\left\{\frac{k}{n}\right\}^{\alpha\frac{k}{n}} + \left\{1-\frac{k}{n}\right\}^{\alpha\left\{1-\frac{k}{n}\right\}} - 2\right] \tag{2.4.13}$$

For obtaining the minimum value, we consider the following two cases:

(a) The minimum value of this entropy arises when some of the values are unity and the others are zero, that is, when the set is a crisp set and this is possible only if k is positive integer.

Thus, we have $\left[H_1^{\alpha}(A)\right]_{\min} = 0$.

Hence, in this case, we have

$$\left[H_1^{\alpha}(A)\right]_{\max} - \left[H_1^{\alpha}(A)\right]_{\min} = -\frac{n}{\alpha}\left[\left\{\frac{k}{n}\right\}^{\alpha\frac{k}{n}} + \left\{1-\frac{k}{n}\right\}^{\alpha\left\{1-\frac{k}{n}\right\}} - 2\right]$$

Consequently, the weighted information supplied by the constraint (2.4.11) is given by the following relation:

$$I_3(A) = -\frac{2n}{\alpha}\left[\left\{1-\frac{1}{2^{\alpha/2}}\right\} + \frac{1}{2}\left\{\left\{\frac{k}{n}\right\}^{\alpha\frac{k}{n}} + \left\{1-\frac{k}{n}\right\}^{\alpha\left\{1-\frac{k}{n}\right\}} - 2\right\}\right] \tag{2.4.14}$$

(b) In case, k is not an integer, then minimum value is obtained by

$$\left[H_1^{\alpha}(A)\right]_{\min} = \frac{n}{\alpha}\left[\left\{\left\{k-\lceil k\rceil\right\}^{\alpha\{k-\lceil k\rceil\}} + \left\{1-k+\lceil k\rceil\right\}^{\alpha\{1-k+\lceil k\rceil\}} - 2\right\}\right]$$

Thus, we have

$$\left[H_1^{\alpha}(A)\right]_{\max} - \left[H_1^{\alpha}(A)\right]_{\min} = \frac{n}{\alpha}\left[\left\{\left\{k-\lceil k\rceil\right\}^{\alpha\{k-\lceil k\rceil\}} + \left\{1-k+\lceil k\rceil\right\}^{\alpha\{1-k+\lceil k\rceil\}} - \left\{\frac{k}{n}\right\}^{\alpha\frac{k}{n}} - \left\{1-\frac{k}{n}\right\}^{\alpha\left\{1-\frac{k}{n}\right\}}\right\}\right]$$

Consequently, the information supplied by the constraint (2.4.11) is given by the following relation:

$$I_4(A) = \frac{n}{\alpha}\left[2\left\{1-\frac{1}{2^{\alpha/2}}\right\} - \left\{\left\{k-\lceil k\rceil\right\}^{\alpha\{k-\lceil k\rceil\}} + \left\{1-k+\lceil k\rceil\right\}^{\alpha\{1-k+\lceil k\rceil\}} - \left\{\frac{k}{n}\right\}^{\alpha\frac{k}{n}} - \left\{1-\frac{k}{n}\right\}^{\alpha\left\{1-\frac{k}{n}\right\}}\right\}\right] \tag{2.4.15}$$

Proceeding on similar lines, the partial information about a fuzzy set can be measured by using various measures of fuzzy entropy.

2.5 GENERATING FUNCTIONS FOR MEASURES OF FUZZY ENTROPY

Golomb [4] defined an information generating function for a probability distribution, given by

$$f(t) = -\sum_{i=1}^{n} p_i^{\,t} \tag{2.5.1}$$

with the property that,

$$f'(1) = -\sum_{i=1}^{n} p_i \log p_i \tag{2.5.2}$$

which is Shannon's [19] measure of entropy for the probability distribution P and thus showed that $f(t)$ gives the expression for the generating function for Shannon's [19] measure of entropy. Similar expressions of generating functions for measures of fuzzy entropy can be obtained upon extending the idea for fuzzy distributions. Let the fuzzy information scheme defined on the fuzzy set A is given by

$$S_1 = \begin{bmatrix} E_1 & E_2 & \ldots E_n \\ \\ \mu_A(x_1) & \mu_A(x_2) & \ldots \mu_A(x_n) \end{bmatrix}$$

where $\mu_A(x_i); i = 1, 2, ..., n$ is a membership function.

We define $f(t) = -\sum_{i=1}^{n}\left[\mu_A^t(x_i) + (1-\mu_A(x_i))^t\right]$ so that

$$f'(1) = -\sum_{i=1}^{n}\left[\mu_A(x_i)\log\mu_A(x_i) + (1-\mu_A(x_i))\log(1-\mu_A(x_i))\right]$$

Thus, $f(t)$ gives the expression for the generating function for De Luca and Termini's [2] measure of fuzzy entropy.

Below, we discuss some generating functions for different measures of fuzzy entropy:

I. Generating function for Kapur's [8] measure of fuzzy entropy of order α and type β

We know that Kapur's [8] measure of fuzzy entropy of type α, β is given by

$$H_{\alpha,\beta}(A) = \frac{1}{\beta-\alpha}\sum_{i=1}^{n}\log\frac{\left\{\mu_A{}^{\alpha}(x_i) + (1-\mu_A(x_i))^{\alpha}\right\}}{\left\{\mu_A{}^{\beta}(x_i) + (1-\mu_A(x_i))^{\beta}\right\}}; \alpha \geq 1, \beta \leq 1 \, or \, \alpha \leq 1, \beta \geq 1 \tag{2.5.3}$$

Let us now consider the following function:

$$f_{\alpha,\beta}(t) = \frac{1}{\beta-\alpha}\sum_{i=1}^{n}\left[\left\{\mu_A{}^{\alpha}(x_i) + (1-\mu_A(x_i))^{\alpha}\right\}^t - \left\{\mu_A{}^{\beta}(x_i) + (1-\mu_A(x_i))^{\beta}\right\}^t\right]; \alpha \neq \beta \tag{2.5.4}$$

Taking $t=1$, equation (2.5.4) gives

$$f_{\alpha,\beta}(1)=\frac{1}{\beta-\alpha}\sum_{i=1}^{n}\left[\mu_A^{\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}-\mu_A^{\beta}(x_i)-\left(1-\mu_A(x_i)\right)^{\beta}\right];\alpha\neq\beta \tag{2.5.5}$$

which is fuzzy entropy introduced by Kapur [8].

Now differentiating equation (2.5.4) with respect to t, we have

$$\begin{aligned}f'_{\alpha,\beta}(t)=\frac{1}{\beta-\alpha}\sum_{i=1}^{n}\Big[&\left\{\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right\}^{t}\log\left\{\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right\}\\&-\left\{\mu_A^{\beta}(x_i)+(1-\mu_A(x_i))^{\beta}\right\}^{t}\log\left\{\mu_A^{\beta}(x_i)+(1-\mu_A(x_i))^{\beta}\right\}\Big]\end{aligned} \tag{2.5.6}$$

Taking $t=0$, equation (2.5.6) gives the following expression:

$$f'_{\alpha,\beta}(0)=\frac{1}{\beta-\alpha}\sum_{i=1}^{n}\left[\log\left\{\mu_A^{\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\right\}-\log\left\{\mu_A^{\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\beta}\right\}\right]$$

which is fuzzy entropy (2.5.3) introduced by Kapur [8]. Thus, $f_{\alpha,\beta}(t)$ can be taken as a generating function for Kapur's [8] measure of fuzzy entropy.

Also we have the following results:

(a) $$f_{\alpha,1}(1)=\frac{1}{1-\alpha}\sum_{i=1}^{n}\left[\mu_A^{\alpha}(x_i)-(1-\mu_A(x_i))^{\alpha}-1\right]$$

which is Kapur's [8] fuzzy entropy.

(b) $$f'_{\alpha,1}(0)=\frac{1}{1-\alpha}\sum_{i=1}^{n}\log\left[\mu_A^{\alpha}(x_i)-(1-\mu_A(x_i))^{\alpha}\right]$$

which is Bhandari and Pal's [1] measure of fuzzy entropy.

II. Generating function for Bhandari and Pal's [1] measure of fuzzy entropy

We have already expressed Bhandari and Pal's [1] measure of fuzzy entropy in equation (2.1.2).

Let us now consider a function

$$f_{\alpha,\beta}(t)=\frac{1}{\beta-\alpha}\sum_{i=1}^{n}\left[\left\{\mu_A^{\alpha/\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha/\beta}\right\}^{t}-\beta\right];\alpha\neq\beta,\alpha\neq1,\beta>0 \tag{2.5.7}$$

Then

$$f_{\alpha,1}(1)=\frac{1}{1-\alpha}\sum_{i=1}^{n}\left[\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}-1\right];\alpha\neq1 \tag{2.5.8}$$

which is a measure of fuzzy entropy introduced by Kapur [8] and it corresponds to Havrada Charvat's [6] probabilistic entropy.

Now differentiating equation (2.5.7) with respect to t, we have

$$f'_{\alpha,\beta}(t) = \frac{1}{\beta-\alpha}\sum_{i=1}^{n}\left[\left\{\mu_A^{\alpha/\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha/\beta}\right\}^t \log\left\{\mu_A^{\alpha/\beta}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha/\beta}\right\}\right] \tag{2.5.9}$$

Taking $t = 0, \beta = 1$, equation (2.5.9) gives the following expression:

$$f'_{\alpha,1}(0) = \frac{1}{1-\alpha}\sum_{i=1}^{n}\left[\log\left\{\mu_A^{\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\right\}\right]$$

which is Bhandari and Pal's [1] measure of fuzzy entropy and it corresponds to Renyi's [17] measure of probabilistic entropy.

Thus, the function $f_{\alpha,\beta}(t)$ can be taken as a generating function for Bhandari and Pal's [1] measure of fuzzy entropy.

Proceeding on similar lines, the generating functions for various measures of fuzzy entropy can be developed.

2.6 NORMALIZING MEASURES OF FUZZY ENTROPY

We first of all discuss the need for normalizing measures of fuzzy entropy. We know that De Luca and Termini's [2] measure of fuzzy entropy is given by

$$H(A) = -\sum_{i=1}^{n}\left[\mu_A(x_i)\log\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\log\left(1-\mu_A(x_i)\right)\right] \tag{2.6.1}$$

This entropy measures the degree of equality among the fuzzy values, that is, the greater the equality among, $\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n)$, the greater is the value of $H(A)$ and this entropy has its maximum value $n\log 2$ when all the fuzzy values are equal, that is, when each $\mu_A(x_i) = \frac{1}{2}$.

Thus, we have

$$H(F) = n\log 2 \tag{2.6.2}$$

where $F = \left\{\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}\right\}$ is the most fuzzy distribution.

The entropy (2.6.1) also measures the uniformity of A or 'closeness' of A to the most fuzzy distribution F , since according to Bhandari and Pal's [1] measure, the fuzzy directed divergence of A from F is given by

$$\begin{aligned} D(A,F) &= \sum_{i=1}^{n}\left[\mu_A(x_i)\log\frac{\mu_A(x_i)}{1/2} + (1-\mu_A(x_i))\log\frac{1-\mu_A(x_i)}{1/2}\right] \\ &= n\log 2 - H(A) \end{aligned} \tag{2.6.3}$$

(This is to be remarked that we shall discuss Bhandari and Pal's [1] measure of fuzzy directed divergence in Chapter-III)

Thus, greater the value of entropy $H(A)$, the nearer is A to F. This entropy provides a measure of equality or uniformity of the fuzzy values $\mu_A(x_1), \mu_A(x_2), ..., \mu_A(x_n)$ among themselves. Now, let us consider the following fuzzy distributions:

$\mu_A(x_i) = (0.4, 0.4, 0.4, 0.4)$ and $\mu_B(x_i) = (0.3, 0.3, 0.4, 0.4, 0.4)$

Then, we have

$H(A) = 1.2734$ and $H(B) = 1.5825 > H(A)$

We want to check which fuzzy distribution is more uniform or in which distribution the fuzzy values are more equal? The obvious answer is that fuzzy values of A are more equal than the fuzzy values of B. Thus, A is more uniformly distributed than B but still $H(A) < H(B)$.

From the values of the two fuzzy entropies, it appears that B is more uniform than A which is obviously wrong and this fallacy arises due to the fact that the fuzzy entropy depends not only on the degree of equality among the fuzzy values; it also depends on the value of n. So long as n is the same, entropy can be used to compare the uniformity of the fuzzy distributions. But, if the number of outcomes are different, then fuzzy entropy is not a satisfactory measure of uniformity. In this case, we try to eliminate the effect of n by normalizing the fuzzy entropy, that is, by defining a normalized measure of fuzzy entropy as

$$\overline{H}(A) = \frac{H(A)}{\max H(A)} \tag{2.6.4}$$

It is obvious that

$$0 \le \overline{H}(\text{A}) \le 1 \tag{2.6.5}$$

For De Luca and Termini's [2] measure of fuzzy entropy, we have

$$\overline{H}(\text{A}) = -\frac{\sum_{i=1}^{4} [\mu_{\text{A}}(x_{\text{i}}) \ \log \mu_{\text{A}}(x_{\text{i}}) + (1-\mu_{\text{A}}(x_{\text{i}}))\log(1-\mu_{\text{A}}(x_{\text{i}}))]}{4\log 2} = 1.05754$$

and

$\overline{H}(\text{B}) = 1.05149$

Obviously, $\overline{H}(\text{A}) > \overline{H}(\text{B})$

Thus, A is more uniform than B. This gives the correct result that B is less uniform than A. Thus, to compare the uniformity, or equality or uncertainty of two fuzzy distributions, we should compare their normalized values and not the exact ones. Some of the normalized measures of fuzzy entropy are discussed below:

I. Normalization of Parkash's [13] measure of fuzzy entropy

Parkash [13] introduced the following measure of fuzzy entropy:

$$H_\alpha{}^\beta(A) = [(1-\alpha)\,\beta]^{-1} \sum_{i=1}^{n} \left[\left\{ \mu_A{}^\alpha(x_i) + (1-\mu_A(x_i))^\alpha \right\}^\beta - 1 \right], \alpha \neq 1, \alpha > 0, \beta \neq 0 \tag{2.6.6}$$

The maximum value of (2.6.6) is given by

$$\left[H_\alpha^\beta(A)\right]_{\max} = \frac{n}{(1-\alpha)\beta}\left\{2^{(1-\alpha)\beta} - 1\right\} \tag{2.6.7}$$

Thus, the expression for Parkash's [13] normalized measure is given by

$$\left[H_\alpha^\beta(A)\right]_N = \frac{\sum_{i=1}^{n} \left[\left\{ \mu_A{}^\alpha(x_i) + (1-\mu_A(x_i))^\alpha \right\}^\beta - 1 \right]}{n\left\{2^{(1-\alpha)\beta} - 1\right\}} \tag{2.6.8}$$

II. Normalization of Parkash and Sharma's [14] measure of fuzzy entropy of order a

Parkash and Sharma [14] introduced the following measure of fuzzy entropy:

$$H_a(A) = \sum_{i=1}^{n} \left[\log\left(1 + a\mu_A(x_i)\right) + \log\left(1 + a\left(1 - \mu_A(x_i)\right)\right) - \log(a+1) \right]; \; a > 0 \tag{2.6.9}$$

The maximum value of (2.6.9) is given by

$$\left[H_a(A)\right]_{\max} = n\log\frac{\left\{\frac{a+2}{2}\right\}^2}{1+a} \tag{2.6.10}$$

Thus, the expression for Parkash and Sharma's [14] normalized measure is given by

$$\left[H_a(A)\right]_N = \frac{\sum_{i=1}^{n} \left[\log\left(1 + a\mu_A(x_i)\right) + \log\left(1 + a\left(1 - \mu_A(x_i)\right)\right) - \log(a+1) \right]}{n\log\frac{\left\{\frac{a+2}{2}\right\}^2}{1+a}} \tag{2.6.11}$$

Proceeding on similar lines, we can obtain maximum values for different fuzzy entropies and consequently, develop many other expressions of the normalized measures of fuzzy entropy.

Concluding Remarks: Information theory deals with the development of probabilistic and fuzzy measures of information. The necessity of developing fuzzy measures arises when the experiment under consideration is non-probabilistic in nature. In such cases, we explore the idea of fuzzy distributions instead of probability distributions and consequently, develop fuzzy measures of entropy which can be successfully applied to different fields. In the present chapter, we have introduced some new parametric measures of fuzzy entropy by considering fuzzy distributions and proved that these measures satisfy all the necessary and desirable properties of being measures of fuzzy entropy. Moreover, with the help of these measures, we have measured the partial information about a fuzzy set when only partial knowledge of the fuzzy values is provided. Regarding these fuzzy distributions, we have shown that while comparing two or more fuzzy distributions, we should compare their normalized values and not their exact values.

REFERENCES

[1] Bhandari, D., & Pal, N. R. (1993). Some new information measures for fuzzy sets. *Information Sciences, 67*, 209-228.

[2] De Luca, A., & Termini, S. (1972). A definition of non-probabilistic entropy in setting of fuzzy set theory. *Information and Control, 20*, 301-312.

[3] Ebanks, B. R. (1983). On measures of fuzziness and their representations. *Journal of Mathematical Analysis and Applications, 94*, 24-37.

[4] Golomb, S. (1966). The information generating function of a probability distribution. *IEEE Transactions on Information Theory,* 12, 75-79.

[5] Guo, X. Z., & Xin, X. L. (2006). Some new generalized entropy formulas of fuzzy sets. *Journal of the Northwest University, 36*, 529-532.

[6] Havrada, J. H., & Charvat, F. (1967). Quantification methods of classification process: Concept of structural α-entropy. *Kybernetika, 3*, 30-35.

[7] Hu, Q., & Yu, D. (2004). Entropies of fuzzy indiscernibility relation and its operations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 12*, 575-589.

[8] Kapur, J. N. (1997). *Measures of fuzzy information*. New Delhi: Mathematical Sciences Trust Society.

[9] Klir, G. J., & Folger, T. A. (1988). *Fuzzy sets, uncertainty and indetermination*. New York: Prentice Hall.

[10] Li, X., & Liu, B. (2007). Maximum fuzzy entropy principle for fuzzy variables. *International Journal of Uncertainty, Fuzziness and Knowledge-based systems*, *15*, 43-52.

[11] Osman, M. S. A., Abdel-Fadel, A. M. A., El-Sersy, K. F., & Ibraheem, N. M. A. (2008). On the solution of fuzzy parametric bicriterion programming problems. *Opsearch*, *45*(2), 119-138.

[12] Pal, N. R., & Bezdek, J. C. (1994). Measuring fuzzy uncertainty. *IEEE Transaction on Fuzzy Systems, 2*, 107-118.

[13] Parkash, O. (1998). A new parametric measure of fuzzy entropy. *Information Processing and Management of Uncertainty, 2*, 1732-1737.

[14] Parkash, O., & Sharma, P. K. (2004). Measures of fuzzy entropy and their relations. *International. Journal of Management & Systems, 20*, 65-72.

[15] Parkash, O., Sharma, P. K., & Mahajan, R. (2008). New measures of weighted fuzzy entropy and their applications for the study of maximum weighted fuzzy entropy principle. *Information Sciences, 178*, 2389-2395.

[16] Parkash, O., Sharma, P. K., & Mahajan, R. (2010). Optimization principle for weighted fuzzy entropy using unequal constraints. *Southeast Asian Bulletin of Mathematics, 34*, 155-161.

[17] Renyi, A. (1961). On measures of entropy and information. *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability* (pp.547-561).

[18] Rudas, I. J. (2001). Measures of fuzziness: theory and applications. *Advances in fuzzy systems and evolutionary computation*. Athens: World Scientific and Engineering Society Press.

[19] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-659.

[20] Singh, R. P., & Tomar, V. P. (2006). Fuzzy information measures through generating functions and some comparison results. *Information Processing and Management of Uncertainty*, 21-28.

[21] Singpurwalla, N. D., & Booker, J. M. (2004). Membership functions and probability measures of fuzzy sets. *Journal of the American Statistical Association, 99*, 867-889.

[22] Yager, R. R. (1979). On measures of fuzziness and negation, Part-I: membership in the unit interval. *International Journal of General Systems, 5*, 221-229.

[23] Yang, Y. W., & Xin, X. L. (2012). Information entropy measure for vague sets and its application. *Mohu Xitong yu Shuxue, 26*(1), 161-165.

[24] Zadeh, L. A. (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications, 23*, 421-427.

CHAPTER-III

NEW GENERALIZED MEASURES OF FUZZY DIVERGENCE AND THEIR DETAILED PROPERTIES

ABSTRACT

It is known fact that the existing literature of mathematics contains a variety of distance measures applicable to various disciplines of science and engineering. Similarly, in the field of information theory, especially while dealing with fuzzy distributions, we have a large number of mathematical models representing measures of fuzzy divergence, each with its own merits and limitations. Keeping in view the application areas of these divergence measures, we have introduced some new generalized parametric measures of fuzzy divergence and studied their important properties. These fuzzy divergence measures can be used to generate new measures of fuzzy entropy, the findings of which have been presented in the present chapter. The concepts of crispness and generating functions have been explained, and some crispy measures and the generating functions for some divergence measures have been obtained.

Keywords: ***Directed divergence, Fuzzy divergence, Cross entropy, Fuzzy entropy, Convex function, Probability distribution, Crispness, Generating functions.***

3.1 INTRODUCTION

In various disciplines of science and engineering, the concept of distance has been proved to be very important and useful because of its applications towards the development of various mathematical models. In the literature of information theory, one such a measure of distance in probability spaces, usually known as directed divergence has been provided by Kullback and Leibler [13]. In case of fuzzy distributions, distance measure is a term that describes the difference between fuzzy sets and can be considered as a dual concept of similarity measure.

Many researchers, such as Yager [23], Kosko [12] and Kaufmann [11] have used distance measure to define fuzzy entropy. Using the axiom definition of distance measure, Fan, Ma and Xie [6] developed some new formulas of fuzzy entropy induced by distance measure and studied some new properties of distance measure. Dubois and Prade [5] defined the distance between two fuzzy subsets on a fuzzy subset of R^+. Rosenfeld [20] defined the shortest distance between two fuzzy sets as a density function on the non-negative reals, and this definition generalizes the definition of shortest distance for crisp sets in a natural way.

Thus, corresponding to Kullback and Leibler [13] divergence measure, Bhandari and Pal [2] introduced the following measure of fuzzy directed divergence:

$$D(A:B)=\sum_{i=1}^{n}\left[\mu_A(x_i)\log\frac{\mu_A(x_i)}{\mu_B(x_i)}+(1-\mu_A(x_i))\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right] \tag{3.1.1}$$

Corresponding to Renyi's [19] and Havrada and Charvat's [9] probabilistic divergence measures, Kapur [10] took the following expressions of fuzzy divergence measures:

$$D_\alpha(A:B)=\frac{1}{\alpha-1}\sum_{i=1}^{n}\log\left[\mu_A^{\alpha}(x_i)\,\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right];\alpha\neq 1,\alpha>0 \tag{3.1.2}$$

and

$$D^\alpha(A:B)=\frac{1}{\alpha-1}\sum_{i=1}^{n}\left[\mu_A{}^{\alpha}(x_i)\,\mu_B{}^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\,(1-\mu_B(x_i))^{1-\alpha}-1\right];\alpha\neq 1,\alpha>0 \tag{3.1.3}$$

Tran and Duckstein [22] developed a new approach for ranking fuzzy numbers based on a distance measure and introduced a new class of distance measures for interval numbers that takes into account all the points in both intervals, and then used it to formulate the distance measure for fuzzy numbers. The approach has been illustrated by numerical examples. Motivated by the standard measures of probabilistic divergence, Parkash and Sharma [17] introduced the following model for fuzzy directed divergence corresponding to Ferrari's [7] probabilistic divergence:

$${}_1D_a(A:B)=\frac{1}{a}\sum_{i=1}^{n}\left[\left(1+a\mu_A(x_i)\right)\log\frac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}+\left\{1+a\left(1-\mu_A(x_i)\right)\right\}\log\frac{1+a\left(1-\mu_A(x_i)\right)}{1+a\left(1-\mu_B(x_i)\right)}\right],a>0 \tag{3.1.4}$$

Kapur [10] introduced the following measures of fuzzy divergence:

$${}_2D_{\alpha,\beta}(A:B)=\frac{1}{\beta-\alpha}\log\frac{\sum_{i=1}^{n}\left\{\mu_A{}^{\alpha}(x_i)\mu_B{}^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right\}}{\sum_{i=1}^{n}\left\{\mu_A{}^{\beta}(x_i)\mu_B{}^{1-\beta}(x_i)+(1-\mu_A(x_i))^{\beta}(1-\mu_B(x_i))^{1-\beta}\right\}};\alpha\neq\beta,\alpha\leq 1,\beta\geq 1\,or\,\alpha\geq 1,\beta\leq 1 \tag{3.1.5}$$

$${}_3D_{\alpha,\beta}(A:B)=\frac{1}{2-\beta-\alpha}\sum_{i=1}^{n}\left\{\begin{array}{r}\mu_A{}^{\alpha}(x_i)\mu_B{}^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}-\mu_A{}^{\beta}(x_i)\mu_B{}^{1-\beta}(x_i)\\-(1-\mu_A(x_i))^{\beta}(1-\mu_B(x_i))^{1-\beta}-2\end{array}\right\}; \tag{3.1.6}$$

$$\alpha\neq\beta,\alpha\leq 1,\beta\geq 1\,or\,\alpha\geq 1,\beta\leq 1$$

and

$$ {}_{4}D_{a}(A:B)=\sum_{i=1}^{n}\left[\mu_A(x_i)\log\frac{\mu_A(x_i)}{\mu_B(x_i)}+(1-\mu_A(x_i))\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right] $$

$$ -\frac{1}{a}\sum_{i=1}^{n}\left[\begin{array}{l}\left(1+a\mu_A(x_i)\right)\log\dfrac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}\\ \qquad+\left\{1+a\left(1-\mu_A(x_i)\right)\right\}\log\dfrac{1+a\left(1-\mu_A(x_i)\right)}{1+a\left(1-\mu_B(x_i)\right)}\end{array}\right];a>0 \tag{3.1.7} $$

Parkash [16] introduced a new generalized measure of fuzzy directed divergence involving two real parameters, given by

$$ {}_{1}D_{\alpha}^{\beta}(A:B)=[(\alpha-1)\,\beta]^{-1}\sum_{i=1}^{n}\left[\left\{\ \mu_A{}^{\alpha}(x_i)\mu_B{}^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right\}^{\beta}-1\right];\alpha\neq 1,\alpha>0,\beta\neq 0 \tag{3.1.8} $$

Many measures of fuzzy divergence have been developed by Kapur [10], Parkash and Sharma [17], Pal and Bezdek [15] etc. In fact, Kapur [10] has developed many expressions for the measures of fuzzy directed divergence corresponding to probabilistic measures of divergence due to Harvada and Charvat [9], Renyi [19], Sharma and Taneja [21] etc. Many other measures of fuzzy divergence have been discussed and developed by Lowen [14], Bhandari, Pal and Majumder [3], Parkash [16], Dubois and Prade [5], Parkash and Tuli [18] etc.

In the next section, we have used the concept of weighted information provided by Belis and Guiasu [1] and consequently, introduced some new measures of weighted fuzzy divergence depending upon some real parameters. The importance of the weighted measures arise because when we have a number of fuzzy elements, some of these elements are supposed to be more important than the others for the purpose of decision making.

3.2 NEW GENERALISED MEASURES OF WEIGHTED FUZZY DIVERGENCE

We propose the following measures of weighted fuzzy directed divergence:

I. $$ D^{\alpha}(A:B;W)=-\frac{1}{\alpha}\sum_{i=1}^{n}w_i\left[2-\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)}-\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))}\right];\alpha\neq 0 \tag{3.2.1} $$

To prove that (3.2.1) is a correct measure of weighted fuzzy divergence, we study the following properties:

(i) Convexity of $D^{\alpha}(A:B;W)$

We have $\dfrac{\partial D^{\alpha}(A:B;W)}{\partial \mu_A(x_i)} = w_i\left[\left(\dfrac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)}\left(1+\log\dfrac{\mu_A(x_i)}{\mu_B(x_i)}\right)-\left(\dfrac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))}\left(1+\log\dfrac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)\right]$

Also

$$\frac{\partial^2 D^{\alpha}(A:B;W)}{\partial \mu_A^2(x_i)} = w_i\left[\begin{array}{c}\left(\dfrac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)}\alpha\left(1+\log\dfrac{\mu_A(x_i)}{\mu_B(x_i)}\right)^2+\left(\dfrac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)}\dfrac{1}{\mu_A(x_i)}\\ +\left(\dfrac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))}\alpha\left(1+\log\dfrac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^2\\ +\left(\dfrac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))}\dfrac{1}{1-\mu_A(x_i)}\end{array}\right]>0$$

which shows that $D^{\alpha}(A:B;W)$ is a convex function of $\mu_A(x_i)$.

Similarly, we have

$$\frac{\partial^2 D^{\alpha}(A:B;W)}{\partial \mu_B^2(x_i)} = w_i\left[\frac{(\mu_A(x_i))^{(1+\alpha\mu_A(x_i))}}{(\mu_B(x_i))^{(2+\alpha\mu_A(x_i))}}\{\alpha\mu_A(x_i)+1\}+\frac{(1-\mu_A(x_i))^{\alpha(1-\mu_A(x_i))+1}}{(1-\mu_B(x_i))^{\alpha(1-\mu_A(x_i))+2}}\{\alpha(1-\mu_A(x_i))+1\}\right]>0$$

which shows that $D^{\alpha}(A:B;W)$ is a convex function of $\mu_B(x_i)$.

(ii) Non-negativity of $D^{\alpha}(A:B;W)$

Firstly, we find the minimum value of $D^{\alpha}(A:B;W)$. For minimum value, we put

$\dfrac{\partial D^{\alpha}(A:B;W)}{\partial \mu_A(x_i)}=0$ which implies that

$$\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)}\left(1+\log\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)=\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))}\left(1+\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)$$

which is possible only if

$$\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)=\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)$$

and $\left(1+\log\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)=\left(1+\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)$

that is, only if

$\mu_A(x_i)=\mu_B(x_i)\ \forall i$

Now, when $\mu_A(x_i)=\mu_B(x_i)\ \forall i$, we have $D^{\alpha}(A:B;W)=0$. Since, the minimum value of the function $D^{\alpha}(A:B;W)$ is 0 and the function itself is convex, we must have $D^{\alpha}(A:B;W)\geq 0$.

Thus, $D^{\alpha}(A:B;W)$ satisfies the following properties:

(i) $D^{\alpha}(A:B;W)\geq 0$

(ii) $D^{\alpha}(A:B;W)=0$ iff $\mu_A(x_i)=\mu_B(x_i)\ \forall i$

(iii) $D^{\alpha}(A:B;W)$ is a convex function of both $\mu_A(x_i)$ and $\mu_B(x_i)$ for each α.

Hence, the proposed measure $D^{\alpha}(A:B;W)$ is a valid measure of weighted fuzzy divergence.

Next, we have presented the measure (3.2.1) graphically with the help of data and consequently, obtained Fig.-3.2.1 which shows that the generalized measure of weighted fuzzy divergence (3.2.1) is a convex function.

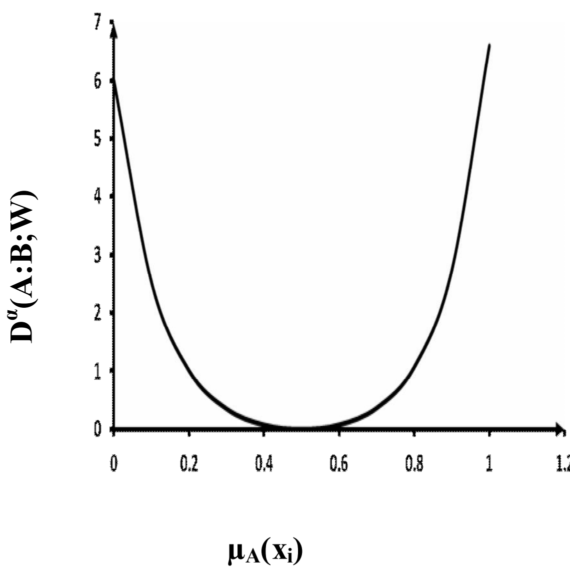


Fig.-3.2.1 Convexity of $D^{\alpha}(A:B;W)$ with respect to $\mu_A(x_i)$

II. Next, we propose another generalized measure of weighted directed divergence, given by

$$D_\alpha(A:B;W) = -\frac{1}{\alpha}\sum_{i=1}^{n} w_i \left[\begin{array}{l} 4 - \left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)} - \left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))} - \frac{\mu_A(x_i)}{\mu_B(x_i)} \\ \\ + \log\frac{\mu_A(x_i)}{\mu_B(x_i)} - \frac{1-\mu_A(x_i)}{1-\mu_B(x_i)} + \log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)} \end{array} \right] \quad (3.2.2)$$

To prove that the expression introduced in (3.2.2) is a correct measure of divergence, we study the following properties:

(i) Convexity of $D_\alpha(A:B;W)$

We have

$$\frac{\partial D_\alpha(A:B;W)}{\partial \mu_A(x_i)} = -\frac{1}{\alpha} w_i \left[\begin{array}{l} -\alpha\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)}\left(1+\log\frac{\mu_A(x_i)}{\mu_B(x_i)}\right) - \frac{1}{\mu_B(x_i)} + \frac{1}{\mu_A(x_i)} \\ \\ +\alpha\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))}\left(1+\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right) \\ \\ +\frac{1}{1-\mu_B(x_i)} - \frac{1}{1-\mu_A(x_i)} \end{array} \right]$$

Also

$$\frac{\partial^2 D_\alpha(A:B;W)}{\partial \mu_A^2(x_i)} = -\frac{1}{\alpha} w_i \left[\begin{array}{l} -\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)}\frac{\alpha}{\mu_A(x_i)} - \left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)}\alpha^2\left(1+\log\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^2 \\ \\ -\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))}\frac{\alpha}{1-\mu_A(x_i)} - \left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))} \\ \\ \alpha^2\left(1+\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^2 - \frac{1}{(\mu_A(x_i))^2} - \frac{1}{(1-\mu_A(x_i))^2} \end{array} \right]$$

$$> 0$$

which shows that $D_\alpha(A:B;W)$ is a convex function of $\mu_A(x_i)$.

Similarly, we can show that $D_\alpha(A:B;W)$ is a convex function of $\mu_B(x_i)$ $\forall i$

(ii) Non-negativity of $D_\alpha(A:B;W)$

Firstly, we find the minimum value of $D_\alpha(A:B;W)$ and for this purpose, we put

$$\frac{\partial D_\alpha(A:B;W)}{\partial \mu_A(x_i)}=0$$

This implies that

$$\alpha\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha\mu_A(x_i)}\left(1+\log\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)+\frac{1}{\mu_B(x_i)}+\frac{1}{1-\mu_A(x_i)}=\alpha\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha(1-\mu_A(x_i))}\left(1+\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)$$

$$+\frac{1}{\mu_A(x_i)}+\frac{1}{1-\mu_B(x_i)}$$

which is possible only iff

$$\mu_A(x_i)=\mu_B(x_i)\ \forall i$$

Now, when $\mu_A(x_i)=\mu_B(x_i)$ $\forall i$, we have $D_\alpha(A:B;W)=0$. Since, the minimum value of the function $D_\alpha(A:B;W)$ is 0 and the function itself is convex, we must have $D_\alpha(A:B;W)\geq 0$.

Thus, $D_\alpha(A:B;W)$ satisfies the following properties:

(i) $D_\alpha(A:B;W)\geq 0$

(ii) $D_\alpha(A:B;W)=0$ iff $\mu_A(x_i)=\mu_B(x_i)$ $\forall i$

(iii) $D_\alpha(A:B;W)$ is a convex function of both $\mu_A(x_i)$ and $\mu_B(x_i)$ for each α.

Hence, the proposed measure $D_\alpha(A:B;W)$ is a valid measure of weighted fuzzy divergence.

Next, we have presented the generalized weighted measure of divergence (3.2.2) graphically and obtained the following Fig.-3.2.2 which clearly shows that the measure of weighted divergence (3.2.2) is a convex function.

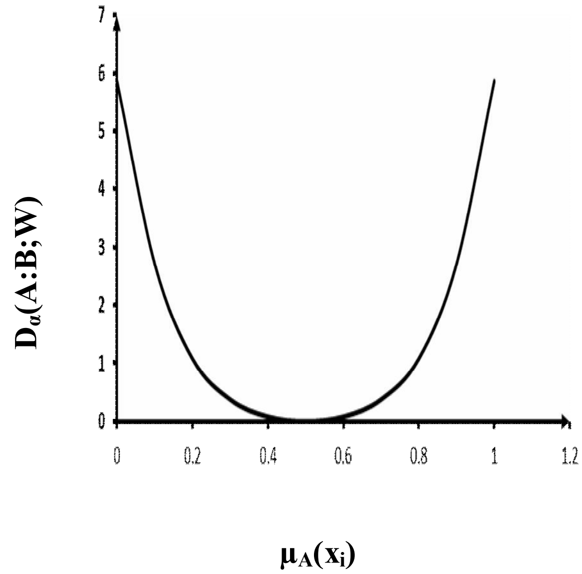


Fig.-3.2.2 Convexity of $D_\alpha(A:B;W)$ with respect to $\mu_A(x_i)$

III. Next, we first introduce a more generalized form of a non-parametric measure of weighted fuzzy divergence, given by

$$D_1^a\ (A:B;W)=\sum_{i=1}^{n} w_i\left[\mu_B(x_i)\,\phi\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)+(1-\mu_B(x_i))\,\phi\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)\right]+$$

$$\sum_{i=1}^{n} w_i\left[(1+a\ \mu_B(x_i))\,\phi\left(\frac{1+a\ \mu_A(x_i)}{1+a\ \mu_B(x_i)}\right)+\{1+a\ (1-\mu_B(x_i))\}\,\phi\left\{\frac{1+a\ (1-\mu_A(x_i))}{1+a\ (1-\mu_B(x_i))}\right\}\right];\ a\geq -1 \qquad (3.2.3)$$

where $\phi(.)$ is a twice differential convex function with $\phi(1)=0$.

Now differentiating (3.2.3) w.r.t. $\mu_A(x_i)$, we get

$$\frac{\partial\ D_1^a\ (A:B;W)}{\partial\mu_A(x_i)}=\ w_i\left[\phi^{/}\left\{\frac{\mu_A(x_i)}{\mu_B(x_i)}\right\}-\phi^{/}\left\{\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right\}+a\phi^{/}\left\{\frac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}\right\}-a\phi^{/}\left\{\frac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))}\right\}\right]$$

Also

$$\frac{\partial^2\ D_1^a\ (A:B;W)}{\partial\mu_A^2(x_i)}=w_i\left[\begin{array}{l}\dfrac{1}{\mu_B(x_i)}\phi^{//}\left\{\dfrac{\mu_A(x_i)}{\mu_B(x_i)}\right\}+\dfrac{1}{1-\mu_B(x_i)}\phi^{/}\left\{\dfrac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right\}\\ \\ +\dfrac{a^2}{1+a\mu_B(x_i)}\phi^{/}\left\{\dfrac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}\right\}+\dfrac{a^2}{1+a(1-\mu_B(x_i))}\phi^{//}\left\{\dfrac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))}\right\}\end{array}\right]>0$$

Thus, we observe that, $D_1^a(A:B;W)$ is a convex function of $\mu_A(x_i)$. Similarly, we can prove that $D_1^a(A:B;W)$ is a convex function of $\mu_B(x_i)$. Also, for minimum value, we have

$$\frac{\partial D_1^a(A:B;W)}{\partial \mu_A(x_i)} = 0$$ which implies that

$$\phi^{/}\left\{\frac{\mu_A(x_i)}{\mu_B(x_i)}\right\} - \phi^{/}\left\{\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right\} + a\phi^{/}\left\{\frac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}\right\} - a\phi^{/}\left\{\frac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))}\right\} = 0$$

which is possible only if

$$\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right) = \left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right) \text{and} \left(\frac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}\right) = \left(\frac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))}\right)$$

that is, only if $\mu_A(x_i) = \mu_B(x_i)\ \forall i$ and this gives $Min.\, D_1^a\ (A:B;W) = 0$.

Thus, for all values of $\mu_A(x_i)$ and $\mu_B(x_i)$, we have

(*i*) $D_1^a\ (A:B;W) \geq 0$

(*ii*) $D_1^a\ (A:B;W) = 0$ when $A = B$

(*iii*) $D_1^a\ (A:B;W)$ is convex.

(iv) $D_1^a\ (A:B;W)$ does not change when $\mu_A(x_i)$ is changed to $1-\mu_A(x_i)$

Thus, we see that the measure (3.2.3) is a generalized weighted parametric measure of divergence. Next, we discuss the following particular cases:

Case-I: If we take $\phi(x) = \log x - x + 1$ in (3.2.3), we get

$$D_2^a\ (A:B;W) = \sum_{i=1}^n w_i \left[\mu_B(x_i)\left(\log\frac{\mu_A(x_i)}{\mu_B(x_i)} - \frac{\mu_A(x_i)}{\mu_B(x_i)} + 1\right) + (1-\mu_B(x_i))\left(\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)} - \frac{1-\mu_A(x_i)}{1-\mu_B(x_i)} + 1\right)\right]$$

$$+ \sum_{i=1}^n w_i \left[\begin{array}{l}(1+a\mu_B(x_i))\left(\log\frac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)} - \frac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)} + 1\right) + (1+a(1-\mu_B(x_i)) \\ \left(\log\frac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))} - \frac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))} + 1\right)\end{array}\right]$$

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

$$+ \sum_{i=1}^n w_i \left[\{1+a\mu_B(x_i)\}\log\frac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)} + \{1+a(1-\mu_B(x_i)\}\log\frac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))}\right] \tag{3.2.4}$$

which is a new weighted measure of fuzzy divergence.

When $a=-1$, equation (3.2.3) gives

$$D_3\,(A:B;W)=2\sum_{i=1}^{n}w_i\left[\mu_B(x_i)\,\phi\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)+(1-\mu_B(x_i))\,\phi\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)\right] \tag{3.2.5}$$

which is another form of a generalized weighted divergence measure.

Case-II: If we take $\phi(x)=\dfrac{x^\alpha-\alpha x+\alpha-1}{\alpha(\alpha-1)};\alpha\neq 0,1$ in (3.2.3), we get

$$D_4^a(A:B;W)=\frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{n}w_i\left[\begin{array}{l}\mu_B(x_i)\left\{\left\{\dfrac{\mu_A(x_i)}{\mu_B(x_i)}\right\}^\alpha-\alpha\left\{\dfrac{\mu_A(x_i)}{\mu_B(x_i)}\right\}+\alpha-1\right\}\\ \\ +\left\{1-\mu_B(x_i)\right\}\left\{\left\{\dfrac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right\}^\alpha-\alpha\left\{\dfrac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right\}+\alpha-1\right\}\end{array}\right]$$

$$+\frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{n}w_i\left[\begin{array}{l}\left\{1+a\mu_B(x_i)\right\}\left\{\left\{\dfrac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}\right\}^\alpha-\alpha\left\{\dfrac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}\right\}+\alpha-1\right\}\\ \\ +\left\{1+a(1-\mu_B(x_i))\right\}\left\{\left\{\dfrac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))}\right\}^\alpha-\alpha\left\{\dfrac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))}\right\}+\alpha-1\right\}\end{array}\right]$$

Thus, we have

$$\begin{aligned}D_4^a(A:B;W)&=\frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{n}w_i\left[\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}-1\right]+\frac{1}{\alpha(\alpha-1)}\\&\sum_{i=1}^{n}w_i\left[(1+a\mu_A(x_i))^\alpha(1+a\mu_B(x_i))^{1-\alpha}+(1+a(1-\mu_A(x_i)))^\alpha(1+a(1-\mu_B(x_i)))^{1-\alpha}-2-a\right]\end{aligned} \tag{3.2.6}$$

which is a new weighted measure of fuzzy divergence.

When $a=-1$, (3.2.6) gives

$$D_5^\alpha(A:B;W)=\frac{2}{\alpha(\alpha-1)}\sum_{i=1}^{n}w_i\left[\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i)+\left\{1-\mu_A(x_i)\right\}^\alpha\left\{1-\mu_B(x_i)\right\}^{1-\alpha}-1\right] \tag{3.2.7}$$

which is another generalized weighted measure of fuzzy divergence.

Note-1: Upon ignoring weights, (3.2.6) gives

$$D_4^{a}(A:B)=\frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{n}\left[\begin{array}{l}\left\{\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}-1\right\}\\ +\left\{\begin{array}{l}(1+a\mu_A(x_i))^{\alpha}(1+a\mu_B(x_i))^{1-\alpha}\\ \qquad+(1+a(1-\mu_A(x_i)))^{\alpha}(1+a(1-\mu_B(x_i)))^{1-\alpha}-2-a\end{array}\right\}\end{array}\right]$$

Taking limit as $\alpha\to 1$, we get

$$D_4(A:B)=\sum_{i=1}^{n}\left[\mu_B(x_i)\log\left\{\frac{\mu_A(x_i)}{\mu_B(x_i)}\right\}+\{1-\mu_A(x_i)\}\log\left\{\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right\}+\{1+a\mu_A(x_i)\}\log\left\{\frac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}\right\}\right]$$

In particular if $a=0$, we get

$$D_4(A:B)=\sum_{i=1}^{n}\left[\mu_B(x_i)\log\left\{\frac{\mu_A(x_i)}{\mu_B(x_i)}\right\}+\{1-\mu_A(x_i)\}\log\left\{\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right\}\right]$$

which is Bhandari and Pal's [2] measure of fuzzy divergence.

Note-2: It is observed that

$$Lt._{\alpha\to 1}D_5^{\alpha}(A:B;W)=2\sum_{i=1}^{n}w_i\left[\mu_A(x_i)\log\frac{\mu_A(x_i)}{\mu_B(x_i)}+\{1-\mu_A(x_i)\}\log\frac{\{1-\mu_A(x_i)\}}{\{1-\mu_B(x_i)\}}\right] \tag{3.2.8}$$

The result (3.2.8) upon ignoring weights gives a measure of fuzzy directed divergence due to Bhandari and Pal [2] except for a multiplicative constant.

Proceeding as above, many new parametric and non-parametric generalized weighted and non-weighted measures of fuzzy divergence can be investigated.

3.3 GENERATING MEASURES OF FUZZY ENTROPY THROUGH FUZZY DIVERGENCE MEASURES

In this section, we have used the concept of fuzzy divergence measure to obtain many new measures of fuzzy entropy. We know that the greater the distance of a given fuzzy set A from the fuzziest set F, the greater should be the deviation of fuzzy uncertainty of the set A from the maximal uncertainty, that is, the smaller is the uncertainty of A. Thus, the fuzzy uncertainty of the set A can be considered as a monotonic decreasing function of the distance of A from F. In other words

$$H(A) = f\{D(A:F)\} \tag{3.3.1}$$

where $H(A)$ is fuzzy entropy of the set A, $D(A:F)$ is the divergence of A from F and $f(.)$ is any monotonic decreasing function. We have chosen the above function $f(.)$ in such a way that it satisfies the following desirable properties:

(i) $H(A) \geq 0$

(ii) $H(C) = 0$ where C is any crisp set.

(iii) $H(A)$ should be a concave function of fuzzy values $\mu_A(x_i)$.

The simplest function satisfying all the above three conditions is given by

$$H_1(A) = D(C:F) - D(A:F) \tag{3.3.2}$$

The result (3.3.2) can be used to obtain new measures of fuzzy entropy for every well-known measure of fuzzy directed divergence.

(a) Measures of fuzzy entropy based upon Parkash's [16] fuzzy divergence of order α and type β

Parkash [16] introduced the following measure of fuzzy divergence:

$$D_\alpha^\beta(A:B) = \frac{1}{(\alpha-1)\beta}\sum_{i=1}^{n}\left[\left\{\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i) + (1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}\right\}^\beta - 1\right]; \quad \alpha \neq 1, \alpha > 0, \beta > 0 \tag{3.3.3}$$

Next, we use the divergence measure (3.3.3) to generate new measure of fuzzy entropy:

(1) Using (3.3.3), equation (3.3.2) gives

$$H_1(A) = \frac{1}{(\alpha-1)\beta}\sum_{i=1}^{n}\left[\left\{\mu_C^\alpha(x_i)2^{\alpha-1} + (1-\mu_C(x_i))^\alpha\, 2^{\alpha-1}\right\}^\beta - 1\right]$$

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

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

$$+\frac{2^{(\alpha-1)\beta}.n}{(1-\alpha)\beta} - \frac{n}{(1-\alpha)\beta}$$

$$= \frac{n}{(\alpha-1)\beta}\left(2^{\alpha\beta}-1\right) - \frac{2^{(\alpha-1)\beta}.n}{(\alpha-1)\beta} - \frac{n}{(1-\alpha)\beta} + 2^{(\alpha-1)\beta}H_\alpha^\beta(A)$$

Thus $H_1(A) = C + 2^{(\alpha-1)\beta} H_\alpha{}^\beta(A)$ (3.3.4)

where

$C = \dfrac{n2^{(\alpha-1)\beta}(2^\beta - 1)}{(\alpha-1)\beta}$ is a positive constant and $H_\alpha{}^\beta(A) = \dfrac{1}{(1-\alpha)\beta}\sum_{i=1}^{n}\left[\left\{\mu_A{}^\alpha + (1-\mu_A)^\alpha\right\}^\beta - 1\right]$

is a measure of fuzzy entropy introduced by Parkash [16].

Thus, $H_1(A)$ introduced in (3.3.4) is a new measure of fuzzy entropy involving Parkash's [16] measure of fuzzy entropy $H_\alpha{}^\beta(A)$.

(b) Measures of fuzzy entropy based upon Parkash and Sharma's [17] fuzzy directed divergence

Here, we consider a more generalized form of fuzzy directed divergence introduced by Parkash and Sharma [17]. This measure is given by

$$D_\lambda(\text{A:B}) = \sum_{i=1}^{n}\left[\begin{array}{l}\left(\lambda\,\mu_A(x_i) + (1-\lambda)\,\mu_B(x_i)\right)\phi\left(\dfrac{\mu_A(x_i)}{\lambda\,\mu_A(x_i) + (1-\lambda)\,\mu_B(x_i)}\right)\\ \\ +\left\{(\lambda(1-\mu_A(x_i)) + (1-\lambda)(1-\mu_B(x_i)\right\}\phi\left(\dfrac{1-\mu_A(x_i)}{\lambda\left(1-\mu_A(x_i)\right)+(1-\lambda)\left(1-\mu_B(x_i)\right)}\right)\end{array}\right];0 \le \lambda \le 1$$

(3.3.5)

Next, we use the divergence measure (3.3.5) to generate the different measures of fuzzy entropy. For this, we consider the following cases:

Case-I: When $\phi(x) = x\log x$, then equation (3.3.5) gives

$$D_{1,\lambda}(\text{A:B}) = \sum_{i=1}^{n}\left[\begin{array}{l}\mu_A(x_i)\log\left(\dfrac{\mu_A(x_i)}{\lambda\,\mu_A(x_i) + (1-\lambda)\,\mu_B(x_i)}\right)\\ \\ \qquad\qquad +(1-\mu_A(x_i))\log\left(\dfrac{1-\mu_A(x_i)}{\lambda\left(1-\mu_A(x_i)\right)+(1-\lambda)\left(1-\mu_B(x_i)\right)}\right)\end{array}\right] \tag{3.3.6}$$

Now, using (3.3.6), equation (3.3.2) gives

$$H_{1,\lambda}(\text{A}) = \text{n}\log\left\{\frac{2}{1+\lambda}\right\}^2 - \sum_{i=1}^{n}\left[\mu_A(x_i)\log\left(\frac{\mu_A(x_i)}{\lambda\,\mu_A(x_i) + \dfrac{1-\lambda}{2}}\right) + (1-\mu_A(x_i))\log\left(\frac{1-\mu_A(x_i)}{\lambda\left(1-\mu_A(x_i)\right) + \dfrac{1-\lambda}{2}}\right)\right]$$

(3.3.7)

When $\lambda = 0$, equation (3.3.7) gives

$$H_{1,0}(\mathrm{A})=\mathrm{n}\log 2-\mathrm{H}(\mathrm{A}) \tag{3.3.8}$$

where $H(A)=-\sum_{i=1}^{n}\left[\mu_A(x_i)\log\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\log\left(1-\mu_A(x_i)\right)\right]$ is De Luca and Termini's [4] measure of fuzzy entropy.

Thus, $H_{1,0}(\mathrm{A})$ introduced in (3.3.8) is a new measure of fuzzy entropy involving standard measure of fuzzy entropy due to De Luca and Termini [4].

Case-II: When $\phi(x)=\dfrac{x^\alpha-x}{\alpha(\alpha-1)};\alpha\neq 0,1$, then equation (3.3.5) gives

$$D_{2,\lambda}(\mathrm{A{:}B})=\frac{1}{\alpha(\alpha\text{-}1)}\sum_{i=1}^{n}\left[\begin{array}{l}\mu_A^\alpha(x_i)\left\{\lambda\,\mu_A(x_i)+(1-\lambda)\,\mu_B(x_i)\right\}^{1-\alpha}\\ \qquad +(1-\mu_A(x_i))^\alpha\left\{\lambda\left(1-\mu_A(x_i)\right)+(1-\lambda)\left(1-\mu_B(x_i)\right)\right\}^{1-\alpha}-1\end{array}\right] \tag{3.3.9}$$

Now, using (3.3.9), equation (3.3.2) gives

$$\begin{aligned}H_{2,\lambda}(\mathrm{A})=&\frac{1}{\alpha(\alpha\text{-}1)}\sum_{i=1}^{n}\left[\begin{array}{l}\mu_C^\alpha(x_i)\left\{\lambda\,\mu_C(x_i)+(1-\lambda)/2\right\}^{1-\alpha}\\ \qquad +\left\{1-\mu_C(x_i)\right\}^\alpha\left\{\lambda\left(1-\mu_C(x_i)\right)+(1-\lambda)/2\right\}^{1-\alpha}-1\end{array}\right]\\ &-\frac{1}{\alpha(\alpha\text{-}1)}\sum_{i=1}^{n}\left[\begin{array}{l}\mu_A^\alpha(x_i)\left\{\lambda\,\mu_A(x_i)+(1-\lambda)/2\right\}^{1-\alpha}\\ \qquad +(1-\mu_A(x_i))^\alpha\left\{\lambda\left\{1-\mu_A(x_i)\right\}+(1-\lambda)/2\right\}^{1-\alpha}-1\end{array}\right]\end{aligned} \tag{3.3.10}$$

When $\lambda=0$, equation (3.3.10) gives

$$H_{2,0}(\mathrm{A})=\frac{\mathrm{n}(2^{\alpha\text{-}1}-1)}{\alpha(\alpha\text{-}1)}+\frac{2^{\alpha\text{-}1}}{\alpha(\alpha\text{-}1)}\sum_{i=1}^{n}\left[\mu_A^\alpha(x_i)+\left\{1-\mu_A(x_i)\right\}^\alpha-1\right]+\frac{\mathrm{n}(1\text{-}2^{\alpha\text{-}1})}{\alpha(\alpha\text{-}1)}$$

Thus

$$H_{2,0}(\mathrm{A})=2^{\alpha\text{-}1}H^\alpha(A) \tag{3.3.11}$$

where

$$H^\alpha(A)=\frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{n}\left[\left\{\mu_A{}^\alpha(x_i)+(1-\mu_A(x_i))^\alpha\right\}-1\right],\alpha\neq 0,1,\alpha>0$$

is an extended version of fuzzy entropy undertaken by Kapur [10] and it corresponds to Havrada-Charvat's [9] probabilistic entropy.

Thus, $H_{2,0}$(A) introduced in (3.3.11) is a standard generalized measure of fuzzy entropy $H^{\alpha}(A)$ except a multiplicative constant.

Case-III: When $\phi(x)=\dfrac{x^{\alpha}-x^{\beta}}{\alpha(\alpha-1)};\alpha\neq\beta,\alpha>1,\beta<1\,or\,\alpha<1,\beta>1$, then equation (3.3.5) gives

$$D_{3,\lambda}(\text{A:B})=\frac{1}{(\alpha\text{-}\beta)}\sum_{i=1}^{n}\left[\begin{array}{l}\mu_A^{\alpha}(x_i)\{\lambda\,\mu_A(x_i)+(1-\lambda)\,\mu_B(x_i)\}^{1-\alpha}-\mu_A^{\beta}(x_i)\{\lambda\,\mu_A(x_i)+(1-\lambda)\,\mu_B(x_i)\}^{1-\beta}\\ \quad+\{1-\mu_A(x_i)\}^{\alpha}\{\lambda\{1-\mu_A(x_i)\}+(1-\lambda)\{1-\mu_B(x_i)\}\}^{1-\alpha}\\ \quad-\{1-\mu_A(x_i)\}^{\beta}\{\lambda\{1-\mu_A(x_i)\}+(1-\lambda)\{1-\mu_B(x_i)\}\}^{1-\beta}\end{array}\right] \tag{3.3.12}$$

Now, using (3.3.12), equation (3.3.2) gives

$$H_{3,\lambda}(\text{A})=\frac{1}{(\alpha\text{-}\beta)}\sum_{i=1}^{n}\left[\begin{array}{l}\mu_C^{\alpha}(x_i)\{\lambda\,\mu_C(x_i)+(1-\lambda)/2\}^{1-\alpha}-\mu_C^{\beta}(x_i)\{\lambda\,\mu_C(x_i)+(1-\lambda)/2\}^{1-\beta}\\ \quad+\{1-\mu_C(x_i)\}^{\alpha}\{\lambda\{1-\mu_C(x_i)\}+(1-\lambda)/2\}^{1-\alpha}\\ \quad-\{1-\mu_C(x_i)\}^{\beta}\{\lambda\{1-\mu_C(x_i)\}+(1-\lambda)/2\}^{1-\beta}\end{array}\right]$$

$$-\frac{1}{(\alpha\text{-}\beta)}\sum_{i=1}^{n}\left[\begin{array}{l}\mu_A^{\alpha}(x_i)\{\lambda\,\mu_A(x_i)+(1-\lambda)/2\}^{1-\alpha}-\mu_A^{\beta}(x_i)\{\lambda\,\mu_A(x_i)+(1-\lambda)/2\}^{1-\beta}\\ \quad+\{1-\mu_A(x_i)\}^{\alpha}\{\lambda\{1-\mu_A(x_i)\}+(1-\lambda)/2\}^{1-\alpha}\\ \quad-\{1-\mu_A(x_i)\}^{\beta}\{\lambda\{1-\mu_A(x_i)\}+(1-\lambda)/2\}^{1-\beta}\end{array}\right] \tag{3.3.13}$$

When $\lambda=0$, equation (3.3.13) gives

$$H_{3,0}(\mathrm{A})=\frac{\mathrm{n}(2^{\alpha}-2^{\beta})}{(\alpha\text{-}\beta)}-\frac{2^{\alpha-1}}{(\alpha\text{-}\beta)}\sum_{i=1}^{n}\left[\mu_A{}^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right]+\frac{2^{\beta-1}}{(\alpha\text{-}\beta)}\sum_{i=1}^{n}\left[\mu_A{}^{\beta}(x_i)+(1-\mu_A(x_i))^{\beta}\right]$$

Thus, we have

$$H_{3,0}(\mathrm{A})=\frac{2^{\alpha-1}(1-\alpha)}{(\alpha\text{-}\beta)(1-\alpha)}\sum_{i=1}^{n}\left[\mu_A{}^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}-1\right]+\frac{2^{\beta-1}(1-\beta)}{(\alpha\text{-}\beta)(1-\beta)}\sum_{i=1}^{n}\left[\mu_A{}^{\beta}(x_i)+(1-\mu_A(x_i))^{\beta}-1\right]$$

$$-\frac{\mathrm{n}2^{\alpha-1}}{(\alpha\text{-}\beta)}+\frac{\mathrm{n}2^{\beta-1}}{(\alpha\text{-}\beta)}$$

$$=C_1+C_2H^{\alpha}(A)+C_3H^{\beta}(A) \qquad (3.3.14)$$

where

$C_1=\dfrac{\mathrm{n}\left\{2^{\alpha-1}-2^{\beta-1}\right\}}{(\alpha\text{-}\beta)}$ is any constant,

$C_2=\dfrac{2^{\alpha-1}(\alpha-1)}{(\alpha\text{-}\beta)}$ is any positive constant for $\alpha>1,\beta<1\,or\,\alpha<1,\beta>1$,

$C_3=-\dfrac{2^{\beta-1}(\beta-1)}{(\alpha\text{-}\beta)}$ is any positive constant for $\alpha>1,\beta<1\,or\,\alpha<1,\beta>1$

and $H^{\alpha}(A)=(1-\alpha)^{-1}\sum_{i=1}^{n}\left[\left\{\mu_A{}^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right\}-1\right]$ is Kapur's [10] measure of fuzzy entropy.

Thus, $H_{3,0}(\mathrm{A})$ introduced in (3.3.14) is a new measure of fuzzy entropy involving standard measure of fuzzy entropy $H^{\alpha}(A)$. Proceeding as above, many new generalized measures of fuzzy entropy can be generated via fuzzy divergence measures.

3.4 SOME QUANTITATIVE-QUALITATIVE MEASURES OF CRISPNESS

We know that a measure of fuzzy entropy of a fuzzy set is a measure of fuzziness of that set and a set is most fuzzy if every element of that set has membership value $\frac{1}{2}$, that is, it is the set having fuzzy vector $F=\left(\frac{1}{2},\frac{1}{2},\frac{1}{2},...,\frac{1}{2}\right)$. If $D(A:B)$ represents the distance between any two fuzzy sets A and B, then $D(A:F)$ gives the distance or divergence between the set A and the most fuzzy set F. Thus, we see that lesser the distance between A and F, fuzzier would be the set A. In the above section, we have explained that the fuzzy uncertainty of the set A can be considered as a monotonic decreasing function of the distance of A from F. In the same way, crispness of a fuzzy set can be measured by taking any monotonic increasing function of the distance of A from F so that larger is the distance, less fuzzy would be the set A and consequently, more crispy. This gives a method of obtaining crispy measures. Below, we have obtained some measures of crispness.

Let A and B be any two fuzzy sets with same support points $x_1, x_2, x_3, ..., x_n$ and with different fuzzy vectors $\mu_A(x_1), \mu_A(x_2), ..., \mu_A(x_n)$ and $\mu_B(x_1), \mu_B(x_2), ..., \mu_B(x_n)$ respectively. Then, we use different measures of fuzzy directed divergence to obtain the measures of crispness as follows:

I. Crispy measures based upon Parkash and Sharma's [17] fuzzy directed divergence

To obtain this measure, we consider a more generalized form (3.3.5) of the fuzzy directed divergence introduced by Parkash and Sharma [17].

Taking $\mu_B(x_i) = \frac{1}{2} \forall i$, equation (3.3.5) gives

$$D_\lambda\left\{A:\frac{1}{2}\right\} = \sum_{i=1}^{n}\left[\begin{array}{l}\left(\lambda\,\mu_A(x_i)+\frac{1-\lambda}{2}\right)\phi\left(\dfrac{\mu_A(x_i)}{\lambda\,\mu_A(x_i)+\frac{1-\lambda}{2}}\right)\\ \qquad +\left(\lambda(1-\mu_A(x_i))+\frac{1-\lambda}{2}\right)\phi\left(\dfrac{1-\mu_A(x_i)}{\lambda\left(1-\mu_A(x_i)\right)+\frac{1-\lambda}{2}}\right)\end{array}\right] \tag{3.4.1}$$

Now, we discuss the following cases:

Case-I: When $\phi(x) = x\log x$, then equation (3.4.1) gives

$$D_{1,\lambda}\left\{A:\frac{1}{2}\right\} = \sum_{i=1}^{n}\left[\begin{array}{l}\left(\lambda\,\mu_A(x_i)+\frac{1-\lambda}{2}\right)\left(\dfrac{\mu_A(x_i)}{\lambda\,\mu_A(x_i)+\frac{1-\lambda}{2}}\log\dfrac{\mu_A(x_i)}{\lambda\,\mu_A(x_i)+\frac{1-\lambda}{2}}\right)\\ +\left(\lambda(1-\mu_A(x_i))+\frac{1-\lambda}{2}\right)\left(\dfrac{1-\mu_A(x_i)}{\lambda\left(1-\mu_A(x_i)\right)+\frac{1-\lambda}{2}}\log\dfrac{1-\mu_A(x_i)}{\lambda\left(1-\mu_A(x_i)\right)+\frac{1-\lambda}{2}}\right)\end{array}\right] \tag{3.4.2}$$

(a) When $\lambda = 0$, equation (3.4.2) gives

$$D_{1,0}\left\{\mathrm{A}:\frac{1}{2}\right\}=\sum_{i=1}^{n}\left[\mu_A(x_i)\log 2\,\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\log 2\left(1-\mu_A(x_i)\right)\right]$$

Thus, the first measure of crispness is given by

$$C_1(A)=n\log 2+\sum_{i=1}^{n}\left[\mu_A(x_i)\log\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\log\left(1-\mu_A(x_i)\right)\right] \tag{3.4.3}$$

(b) When $\lambda=\frac{1}{2}$, equation (3.4.2) gives the second measure of crispness as given by

$$C_2(A)=\sum_{i=1}^{n}\left[\frac{2\mu_A(x_i)+1}{4}\left\{\frac{4\mu_A(x_i)}{2\mu_A(x_i)+1}\log\frac{4\mu_A(x_i)}{2\mu_A(x_i)+1}+\frac{4\{1-\mu_A(x_i)\}}{2\{1-\mu_A(x_i)\}+1}\log\frac{4\{1-\mu_A(x_i)\}}{2\{1-\mu_A(x_i)\}+1}\right\}\right] \tag{3.4.4}$$

Case-II: When $\phi(x)=\dfrac{x^\alpha-x}{\alpha(\alpha-1)};\alpha\neq 0,1$, then equation (3.4.1) gives

$$D_{2,\lambda}\left\{\mathrm{A}:\frac{1}{2}\right\}=\frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{n}\left[\begin{array}{l}\left(\lambda\,\mu_A(x_i)+\frac{1-\lambda}{2}\right)\left(\left\{\dfrac{\mu_A(x_i)}{\lambda\,\mu_A(x_i)+\frac{1-\lambda}{2}}\right\}^{\alpha}-\dfrac{\mu_A(x_i)}{\lambda\,\mu_A(x_i)+\frac{1-\lambda}{2}}\right)\\[2ex] +\left(\lambda(1-\mu_A(x_i))+\frac{1-\lambda}{2}\right)\left(\left\{\dfrac{1-\mu_A(x_i)}{\lambda\left(1-\mu_A(x_i)\right)+\frac{1-\lambda}{2}}\right\}^{\alpha}-\dfrac{1-\mu_A(x_i)}{\lambda\left(1-\mu_A(x_i)\right)+\frac{1-\lambda}{2}}\right)\end{array}\right] \tag{3.4.5}$$

(a) When $\lambda=0$, equation (3.4.5) gives the following measure of crispness:

$$C_3(A)=\frac{n}{\alpha(1-\alpha)}+\frac{2^{\alpha-1}}{\alpha(1-\alpha)}\sum_{i=1}^{n}\left[\mu_A{}^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right] \tag{3.4.6}$$

(b) When $\lambda=\frac{1}{2}$, equation (3.4.5) gives the following measure of crispness:

$$C_4(A)=\frac{1}{\alpha(\alpha-1)}\sum_{i=1}^{n}\left[\begin{array}{l}\dfrac{2\mu_A(x_i)+1}{4}\left\{\left\{\dfrac{4\mu_A(x_i)}{2\mu_A(x_i)+1}\right\}^{\alpha}-\dfrac{4\mu_A(x_i)}{2\mu_A(x_i)+1}\right\}\\[2ex] +\dfrac{2\{1-\mu_A(x_i)\}+1}{4}\left\{\left\{\dfrac{4\{1-\mu_A(x_i)\}}{2\{1-\mu_A(x_i)\}+1}\right\}^{\alpha}-\dfrac{4\{1-\mu_A(x_i)\}}{2\{1-\mu_A(x_i)\}+1}\right\}\end{array}\right] \tag{3.4.7}$$

Moreover, in real life situations, the importance of the events cannot be ignored and thus, the expressions for crispy measures can be obtained by attaching weighted distribution introduced by Belis and Guiasu [1]. In such cases, the weighted fuzzy information scheme is given by

$$S = \begin{bmatrix} E_1 & E_2 & \dots E_n \\ \mu_A(x_1) & \mu_A(x_2) & \dots \mu_A(x_n) \\ \mu_B(x_1) & \mu_B(x_2) & \dots \mu_B(x_n) \\ w_1 & w_2 & w_n \end{bmatrix} \tag{3.4.8}$$

Here, we make use of our own weighted measures introduced in (3.2.1) and (3.2.2). Thus, using (3.2.1), the first weighted crispy measure is given by

$$C_5(A;W) = -\frac{1}{\alpha}\sum_{i=1}^{n} w_i\left[2-\left(2\mu_A(x_i)\right)^{\alpha\mu_A(x_i)}-\left(2(1-\mu_A(x_i))\right)^{\alpha(1-\mu_A(x_i))}\right] \tag{3.4.9}$$

Using (3.2.2), the second weighted crispy measure is given by

$$C_6(A;W) = \frac{2(1-\log 2)}{\alpha}\sum_{i=1}^{n} w_i - \frac{1}{\alpha}\sum_{i=1}^{n} w_i\left[4-\left(2\mu_A(x_i)\right)^{\alpha\mu_A(x_i)}-\left(2(1-\mu_A(x_i))\right)^{\alpha(1-\mu_A(x_i))}+\log\frac{\mu_A(x_i)}{1-\mu_A(x_i)}\right] \tag{3.4.10}$$

Proceeding on similar lines, many new crispy measures can be constructed.

3.5 GENERATING FUNCTIONS FOR VARIOUS WEIGHTED MEASURES OF FUZZY DIVERGENCE

The concept of generating functions for the probability distributions was first discussed by Guiasu and Reisher [8]. To explain the concept, the authors considered two probability distributions P and Q, and then defined the generating function $g(t)$ for relative information or cross-entropy or directed-divergence of P from Q given by

$$g(t) = \sum_{i=1}^{n} q_i\left\{\frac{p_i}{q_i}\right\}^t \tag{3.5.1}$$

and proved that this function $g(t)$ gives an expression for the generating function for an important measure of directed divergence known as Kullback-Leibler's [13] probabilistic measure of divergence.

But, there are situations where probabilistic measures do not work and we explore the possibility of fuzzy measures.

Below, we have obtained some generating functions for different measures of fuzzy divergence.

I. Generating function for Kapur's [10] measure of fuzzy directed divergence

To develop generating function for Kapur's [10] measure of fuzzy directed divergence, we consider the following function:

$$f_{\alpha,\beta}(t)=\frac{1}{(\alpha-1)\beta}\sum_{i=1}^{n}\left[\left\{\mu_A{}^{\alpha}(x_i)\mu_B{}^{1-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\left(1-\mu_B(x_i)\right)^{1-\alpha}\right\}^{\beta t}-1\right] \tag{3.5.2}$$

$$;\alpha\neq 1,\alpha>0,\beta\neq 1$$

Clearly for $t=1$, we have

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

which is a measure of fuzzy divergence introduced by Parkash [16].

Now

$$f'_{\alpha,\beta}(t)=\frac{1}{\alpha-1}\sum_{i=1}^{n}\left[\left\{\mu_A{}^{\alpha}(x_i)\mu_B{}^{1-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\left(1-\mu_B(x_i)\right)^{1-\alpha}\right\}^{\beta t}\right.$$

$$\left.\log\left\{\mu_A{}^{\alpha}(x_i)\mu_B{}^{1-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\left(1-\mu_B(x_i)\right)^{1-\alpha}\right\}\right]$$

Thus

$$f'_{\alpha,\beta}(0)=\frac{1}{\alpha-1}\sum_{i=1}^{n}\log\left\{\mu_A{}^{\alpha}(x_i)\mu_B{}^{1-\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}\left(1-\mu_B(x_i)\right)^{1-\alpha}\right\}$$

which is Kapur's [10] measure of fuzzy directed divergence.

Hence, $f_{\alpha,\beta}(t)$ can be taken as an expression of generating function for Kapur's [10] measure of fuzzy divergence and it corresponds to Renyi's [19] measure of probabilistic directed divergence. Moreover, we have the following results:

(a) From equation (3.5.2), we have

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

which is Kapur's [10] measure of fuzzy directed divergence.

(b) $f_{1,\beta}(1)=\underset{\alpha\to 1}{Lt}\,f_{\alpha,\beta}(1)$

$$= \underset{\alpha \to 1}{Lt} \sum_{i=1}^{n} \Bigg[\left\{ \mu_A{}^{\alpha}(x_i)\,\mu_B{}^{1-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha} \left(1-\mu_B(x_i)\right)^{1-\alpha} \right\}^{\beta-1}$$

$$\left\{ \mu_B{}^{1-\alpha}(x_i)\,\mu_A{}^{\alpha}(x_i) \log \mu_A(x_i) - \mu_A{}^{\alpha}(x_i)\,\mu_B{}^{1-\alpha}(x_i) \log \mu_B(x_i) \right.$$

$$+ \left(1-\mu_B(x_i)\right)^{1-\alpha} \left(1-\mu_A(x_i)\right)^{\alpha} \log\left(1-\mu_A(x_i)\right)$$

$$\left. - \left(1-\mu_A(x_i)\right)^{\alpha} \left(1-\mu_B(x_i)\right)^{1-\alpha} \log\left(1-\mu_B(x_i)\right) \right\} \Bigg]$$

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

which is Bhandari and Pal's [2] measure of fuzzy directed divergence corresponding to Kullback-Leibler's [13] probabilistic measure of divergence.

(c) $f_{\alpha,0}(1) = \underset{\beta \to 0}{Lt}\, f_{\alpha,\beta}(1)$

$$= \underset{\beta \to 0}{Lt} \frac{1}{\alpha-1} \sum_{i=1}^{n} \Bigg[\left\{ \mu_A{}^{\alpha}(x_i)\,\mu_B{}^{1-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha} \left(1-\mu_B(x_i)\right)^{1-\alpha} \right\}^{\beta}$$

$$\log\left\{ \mu_A{}^{\alpha}(x_i)\,\mu_B{}^{1-\alpha}(x_i) + \left(1-\mu_A(x_i)\right)^{\alpha} \left(1-\mu_B(x_i)\right)^{1-\alpha} \right\} \Bigg]$$

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

which is a measure of fuzzy divergence introduced by Kapur [10] and it corresponds to Renyi's [19] measure of directed divergence.

II. Generating function for Parkash and Sharma's [17] measure of fuzzy directed divergence

To develop generating function for Parkash and Sharma's [17] measure of fuzzy directed divergence, we consider the following function:

$$f_{\alpha}(t) = \sum_{i=1}^{n} \left[\mu_{\mathrm{A}}(x_i) \left(\frac{2\,\mu_{\mathrm{A}}(x_i)}{(\mu_{\mathrm{A}}(x_i) + \mu_{\mathrm{B}}(x_i))} \right)^{t} + \{1-\mu_{\mathrm{A}}(x_i)\} \left(\frac{2\,(1-\mu_{\mathrm{A}}(x_i))}{(1-\mu_{\mathrm{A}}(x_i) + 1-\mu_{\mathrm{B}}(x_i))} \right)^{t} \right] \qquad (3.5.3)$$

Differentiating (3.5.3), we have

$$f_{\alpha}^{'}(t) = \sum_{i=1}^{n} \Bigg[\mu_{\mathrm{A}}(x_i) \left(\frac{2\,\mu_{\mathrm{A}}(x_i)}{(\mu_{\mathrm{A}}(x_i) + \mu_{\mathrm{B}}(x_i))} \right)^{t} \log\left(\frac{2\,\mu_{\mathrm{A}}(x_i)}{(\mu_{\mathrm{A}}(x_i) + \mu_{\mathrm{B}}(x_i))} \right)$$

$$+ \{1-\mu_{\mathrm{A}}(x_i)\} \left(\frac{2\,(1-\mu_{\mathrm{A}}(x_i))}{(1-\mu_{\mathrm{A}}(x_i) + 1-\mu_{\mathrm{B}}(x_i))} \right)^{t} \log\left(\frac{2\,(1-\mu_{\mathrm{A}}(x_i))}{(1-\mu_{\mathrm{A}}(x_i) + 1-\mu_{\mathrm{B}}(x_i))} \right) \Bigg]$$

Thus, we have

$$f_{\alpha}^{'}(0)=\sum_{i=1}^{n}\left[\mu_{A}(x_i)\log\left(\frac{2\,\mu_{A}(x_i)}{(\mu_{A}(x_i)+\mu_{B}(x_i))}\right)+\{1-\mu_{A}(x_i)\}\log\left(\frac{2\,(1-\mu_{A}(x_i))}{(1-\mu_{A}(x_i)+1-\mu_{B}(x_i))}\right)\right]$$

which is Parkash and Sharma's [17] measure of fuzzy directed divergence.

Thus, we conclude that the function $f_{\alpha}(t)$ can be taken as an expression of generating function for Parkash and Sharma's [17] measure of fuzzy directed divergence.

Concluding Remarks: Information theory deals with two basic concepts, viz, entropy and directed divergence. Both the concepts are related to each others and given the one, the other can be calculated. In the present chapter, we have used the fuzzy divergence measures to generate new measures of fuzzy entropy. The necessity for developing new measures of fuzzy entropy and fuzzy divergence measures arises from the point of view that we need a variety of generalized parametric and non-parametric measures of information to extend the scope of their applications in a variety of disciplines. The measures developed in this chapter deal with discrete fuzzy distributions only but the present work can also be extended to continuous type of fuzzy distributions.

REFERENCES

[1] Belis, M., & Guiasu, S. (1968). A quantitative-qualitative measure of information in cybernetic systems. *IEEE Transactions on Information Theory, 14*, 593-594.

[2] Bhandari, D., & Pal, N. R. (1993). Some new information measures for fuzzy sets. *Information Sciences, 67*, 209-228.

[3] Bhandari, D., Pal, N. R., & Majumder, D. D. (1992). Fuzzy divergence, probability measure of fuzzy events and image thresholding. *Pattern Recognition Letters, 1*, 857-867.

[4] De Luca, A., & Termini, S. (1972). A definition of non-probabilistic entropy in setting of fuzzy set theory. *Information and Control, 20*, 301-312.

[5] Dubois, D., & Prade, H. (1983). On distances between fuzzy points and their use for plausible reasoning. *Proceedings of International Conference on Systems, Man and Cybernetics* (pp.300-303).

[6] Fan, J. L., Ma, Y. L., & Xie, W. X. (2001). On some properties of distance measure. *Fuzzy Sets and Systems, 117*, 355-361.

[7] Ferreri, C. (1980). Hypoentropy and related heterogeneity divergence measures. *Statistica, 40*, 155-168.

[8] Guiasu, S., & Reisher, C. (1985). The relative information generating function. *Information Sciences,* 35, 235-241.

[9] Havrada, J. H., & Charvat, F. (1967). Quantification methods of classification process: Concept of structural α-entropy. *Kybernetika, 3*, 30-35.

[10] Kapur, J. N. (1997). *Measures of fuzzy information*. New Delhi: Mathematical Sciences Trust Society.

[11] Kaufmann, A. (1975). *Introduction to theory of fuzzy subsets.* New York: Academic Press.

[12] Kosko, B. (1991). Fuzzy entropy and conditioning. *Information Sciences, 42*, 165-174.

[13] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*, 79-86.

[14] Lowen, R. (1996). *Fuzzy set theory-basic concepts, techniques and bibliography*. Boston: Kluwer Academic Publishers.

[15] Pal, N. R., & Bezdek, J. C. (1994). Measuring fuzzy uncertainty. *IEEE Transaction on Fuzzy Systems, 2*, 107-118.

[16] Parkash, O. (2000). On fuzzy symmetric divergence. *The Fourth Asian Fuzzy System Symposium, 2*, 904-908.

[17] Parkash, O., & Sharma, P. K. (2005). Some new measures of fuzzy directed divergence and their generalization. *Journal of the Korean Society of Mathematical Education Series B*, *12*, 307-315.

[18] Parkash, O., & Tuli, R. K. (2005). On generalized measures of fuzzy directed divergence. *Ultra Scientist of Physical Sciences, 17*, 27-34.

[19] Renyi, A. (1961). On measures of entropy and information. *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability* (pp.547-561).

[20] Rosenfeld, A. (1985). Distance between fuzzy sets. *Pattern Recognition Letters, 3*, 229-231.

[21] Sharma, B. D., & Taneja, I. J. (1975). Entropies of type (α,β) and other generalized measures of information theory. *Metrika, 22*, 202-215.

[22] Tran, L., & Duckstein, L. (2002). Comparison of fuzzy numbers using a fuzzy distance measure. *Fuzzy Sets and Systems, 130*, 331-341.

[23] Yager, R. R. (1979). On measures of fuzziness and negation, Part-I: membership in the unit interval. *International Journal of General Systems, 5*, 221-229.

CHAPTER-IV

OPTIMUM VALUES OF VARIOUS MEASURES OF WEIGHTED FUZZY INFORMATION

ABSTRACT

In real life situations as well as in several disciplinary pursuits like coding theory, portfolio analysis, linear programming and many other disciplines of operations research, we are frequently concerned with the problems of finding optimal solutions of constrained optimization problems. In the literature of mathematics, there exist several techniques of optimization including the method of inequalities, the method of differentiation, the method of calculus of variation and the Lagrange's method of maximum multipliers. In the present chapter, we have proposed a method different than the existing ones to optimize generalized measures of weighted fuzzy information.

Keywords: ***Fuzzy set, Crisp set, Fuzzy entropy, Fuzzy cross entropy, Concavity, Piecewise convex.***

4.1 INTRODUCTION

In actual practice it is usually observed that exact values of model parameters are rare in most of the problems related with engineering sciences and biological systems. Normally, uncertainties arise due to incomplete information reflected in uncertain model parameters. This is often the case in price and bidding in market oriented power system operation and planning, in internet search engines, and with the amount of carbohydrates, proteins and fats in ingested meals and gastroparese factor in human glucose metabolic models. A fruitful approach to handle parametric uncertainties is the use of fuzzy numbers which capture our intuitive conceptions of approximate numbers and imprecise quantities, and play a significant role in applications, such as estimation, prediction, classification, decision-making optimization and control.

A measure of fuzziness often used and citied in the literature of information theory, known as fuzzy entropy, was first introduced by Zadeh [25] and the name entropy was chosen due to an intrinsic similarity of equations to the ones in the Shannon entropy [22]. However, the two functions measure fundamentally different types of uncertainties. Basically, the Shannon entropy measures the average uncertainty associated with the prediction of outcomes in a random experiment whereas fuzzy entropy is the quantitative description of fuzziness in fuzzy sets. After the introduction of the theory of fuzzy sets by Zadeh [24], many researchers started working around this concept. Thus, keeping in view the idea of fuzzy sets, De Luca and Termini [2] introduced a measure of fuzzy entropy corresponding to Shannon's [22] well known measure of probabilistic entropy, given by

$$H(A)=-\sum_{i=1}^{n}\left[\mu_A(x_i)\log\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\log\left(1-\mu_A(x_i)\right)\right] \tag{4.1.1}$$

After this development, a large number of measures of fuzzy entropy were discussed, characterized and generalized by various authors. Bhandari and Pal [1] developed the following measure of fuzzy entropy corresponding to Renyi's [20] probabilistic entropy:

$$H_\alpha(A)=\frac{1}{1-\alpha}\sum_{i=1}^{n}\log\left[\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right];\alpha\neq 1,\alpha>0 \tag{4.1.2}$$

Kapur [11] took the following measure of fuzzy entropy corresponding to Havrada and Charvat's [9] probabilistic entropy:

$$H^{\alpha}(A)=(1-\alpha)^{-1}\sum_{i=1}^{n}\left[\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}-1\right];\alpha\neq 1,\alpha>0 \tag{4.1.3}$$

Kapur [11] also introduced the following measure of fuzzy entropy:

$$H_{\alpha,\beta}(A)=\frac{1}{\beta-\alpha}\log\frac{\sum_{i=1}^{n}\left\{\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right\}}{\sum_{i=1}^{n}\left\{\mu_A^{\beta}(x_i)+(1-\mu_A(x_i))^{\beta}\right\}};\alpha\geq 1,\beta\leq 1\ or\ \alpha\leq 1,\beta\geq 1 \tag{4.1.4}$$

Parkash [14] introduced a new measure of fuzzy entropy depending upon two parameters, given by

$$H_\alpha^{\beta}(A)=[(1-\alpha)\,\beta]^{-1}\sum_{i=1}^{n}\left[\left\{\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right\}^{\beta}-1\right];\alpha\neq 1,\alpha>0,\beta\neq 0 \tag{4.1.5}$$

and called it (α,β) fuzzy entropy which includes some well known measures of fuzzy entropy.

Parkash and Sharma [16] introduced two new measures of fuzzy entropy, given by

$$\mathrm{K}_a(A)=\sum_{i=1}^{n}\left[\log(1+a\mu_A(x_i))+\log(1+a(1-\mu_A(x_i)))-\log(1+a)\right];a\geq 0 \tag{4.1.6}$$

and

$$\begin{aligned}\mathrm{H}_a(A)=&-\sum_{i=1}^{n}\left[\mu_A(x_i)\log\mu_A(x_i)+\left(1-\mu_A(x_i)\right)\log\left(1-\mu_A(x_i)\right)\right]\\&-\frac{1}{a}\sum_{i=1}^{n}\begin{bmatrix}\left(1+a\mu_A(x_i)\right)\log\left(1+a\mu_A(x_i)\right)\\+\left\{1+a\left(1-\mu_A(x_i)\right)\right\}\log\left\{1+a\left(1-\mu_A(x_i)\right)\right\}-(1+a)\log(1+a)\end{bmatrix};a>0\end{aligned} \tag{4.1.7}$$

The detailed properties of measures of fuzziness along with their representations have been provided by Ebanks [4] whereas entropy formulas for fuzzy sets have been given by Guo and Xin [8]. If the relative importance of the ith element is given by the weight w_i then the weighted measures of fuzzy entropy can be obtained. Recently, Parkash, Sharma and Mahajan [18] introduced the following two new trigonometric measures of weighted fuzzy entropy, given by

$$H_1(A;W)=\sum_{i=1}^{n} w_i\left[Sin\frac{\pi\mu_A(x_i)}{2}+Sin\frac{\pi(1-\mu_A(x_i))}{2}-1\right] \tag{4.1.8}$$

and

$$H_2(A;W)=\sum_{i=1}^{n} w_i\left[Cos\frac{\pi\mu_A(x_i)}{2}+Cos\frac{\pi(1-\mu_A(x_i))}{2}-1\right] \tag{4.1.9}$$

and provided the applications of their findings for the study of maximum weighted fuzzy entropy principle. Further, Parkash, Sharma and Mahajan [19] using the existing weighted measures of fuzzy entropy, studied the principle of maximum weighted fuzzy entropy with unequal constraints. For this purpose, the authors explained their algorithm with the two techniques and proved that both techniques lead to the same solution.

Another important measure of information is the measure of distance or directed divergence and it describes the difference between fuzzy sets, and many researchers have used this distance measure to define fuzzy entropy. Using the axiom definition of distance measure, Fan, Ma and Xie [5] developed some new formulas of fuzzy entropy induced by distance measure and studied some new properties of this measure whereas Dubois and Prade [3] and Rosenfeld [21] defined differently the measure of distance between two fuzzy subsets. While dealing with fuzzy distributions and keeping in view the probabilistic measure of divergence due to Kullback and Leibler [13], Bhandari and Pal [1] introduced a new measure of fuzzy directed divergence and studied its important properties. This measure of fuzzy divergence is given by the following mathematical model:

$$D(A:B)=\sum_{i=1}^{n}\left[\mu_A(x_i)\log\frac{\mu_A(x_i)}{\mu_B(x_i)}+(1-\mu_A(x_i))\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right] \tag{4.1.10}$$

Corresponding to Renyi's [20] probabilistic divergence, Bhandari and Pal [1] took the following mathematical expression for the measure of fuzzy divergence:

$$D_\alpha(A:B)=\frac{1}{\alpha-1}\sum_{i=1}^{n}\log\left[\mu_A^{\alpha}(x_i)\,\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right];\alpha\neq 1,\alpha>0 \tag{4.1.11}$$

Corresponding to Havrada and Charvat's [9] probabilistic divergence measures, Kapur [11] took the following expressions of fuzzy divergence measures:

$$D^{\alpha}(A:B)=\frac{1}{\alpha-1}\sum_{i=1}^{n}\left[\mu_A{}^{\alpha}(x_i)\,\mu_B{}^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\,(1-\mu_B(x_i))^{1-\alpha}-1\right];\alpha\neq 1,\alpha>0 \qquad (4.1.12)$$

Parkash [15] introduced a generalized measure of fuzzy divergence, given by

$$D_{\alpha}^{\beta}(A:B)=[(\alpha-1)\,\beta]^{-1}\sum_{i=1}^{n}\left[\left\{\mu_A{}^{\alpha}(x_i)\mu_B{}^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right\}^{\beta}-1\right];\alpha\neq 1,\alpha>0,\beta\neq 0 \qquad (4.1.13)$$

Motivated by the standard measures of divergence, Parkash and Sharma [17] introduced the following measure of fuzzy directed divergence corresponding to Ferrari's [6] probabilistic divergence:

$$D^{\alpha}(A:B)=\frac{1}{a}\sum_{i=1}^{n}\left[\begin{aligned}&(1+a\mu_A(x_i))\log\frac{1+a\mu_A(x_i)}{1+a\mu_B(x_i)}\\&\qquad+[1+a(1-\mu_A(x_i))]\log\frac{1+a(1-\mu_A(x_i))}{1+a(1-\mu_B(x_i))}\end{aligned}\right];a>0 \qquad (4.1.14)$$

Many other desirable measures of fuzzy directed divergence satisfying the necessary properties have been discussed and developed by various authors. Keeping in view the existing measures of probabilistic divergence, such as Sharma and Taneja's [23] measure, Ferreri's [6] measure etc., Kapur [11] has developed many expressions for the measures of fuzzy directed divergence to be used while dealing with fuzzy distributions. Related with optimization principles, some work has been done by Guiasu and Shenitzer [7] whereas a large number of applications of maximum entropy principle and minimum cross entropy principle in science and engineering have been investigated by Kapur [10] and Kapur and Kesavan [12]. In the next section, we have introduced a measure of weighted fuzzy entropy, obtained its optimum values, and the results so obtained have been presented graphically.

4.2 OPTIMUM VALUES OF GENERALIZED MEASURES OF WEIGHTED FUZZY ENTROPY

Consider a given fuzzy set *A* with *n* supporting points $(x_1,x_2,...,x_n)$ corresponding to the fuzzy vector $\{\mu_A(x_1),\mu_A(x_2),...,\mu_A(x_n)\}$ where $\mu_A(x_i)$ is the degree of membership of the elements x_i of the set *A*. Our purpose is to find the maximum and minimum values of generalized measures of weighted fuzzy entropy subject to the constraint of total fuzziness, that is, $\sum_{i=1}^{n}\mu_A(x_i)=k,\ 0\leq k\leq n$. For this purpose, we first propose a generalized measure of weighted fuzzy entropy corresponding to Havrada and Charvat's [9] probabilistic entropy given by

$$H^{\alpha}(A;W)=\frac{1}{1-\alpha}\sum_{i=1}^{n}w_i\left[\mu_A^{\alpha}(x_i)+\left(1-\mu_A(x_i)\right)^{\alpha}-1\right];\alpha\neq 1,\alpha>0 \tag{4.2.1}$$

where w_i are non-negative real numbers.

To prove its authenticity, we study the following properties:

(i) We have $\dfrac{\partial H^{\alpha}(A;W)}{\partial\mu_A(x_i)}=\dfrac{\alpha}{1-\alpha}w_i\left[\mu_A^{\alpha-1}(x_i)-\left(1-\mu_A(x_i)\right)^{\alpha-1}\right]$

Also $\dfrac{\partial^2 H^{\alpha}(A;W)}{\partial\mu_A^{\;2}(x_i)}=-\alpha w_i\left\{\mu_A^{\alpha-2}(x_i)+(1-\mu_A(x_i))^{\alpha-2}\right\}<0$

Thus $H^{\alpha}(A;W)$ is a concave function of $\mu_A(x_i)$ $\forall$ i

(ii) $H^{\alpha}(A;W)$ is an increasing function of $\mu_A(x_i)$ for $0\leq\mu_A(x_i)\leq ½$ because

$\left\{H^{\alpha}(A;W)/\mu_A(x_i)=0\right\}=0$ and $\left\{H^{\alpha}(A;W)/\mu_A(x_i)=1/2\right\}=\left\{\dfrac{2^{1-\alpha}-1}{1-\alpha}\right\}\sum_{i=1}^{n}w_i>0$

(iii) $H^{\alpha}(A;W)$ is decreasing function of $\mu_A(x_i)$ for $½\leq\mu_A(x_i)\leq 1$ because

$\left\{H^{\alpha}(A;W)/\mu_A(x_i)=1/2\right\}=\left\{\dfrac{2^{1-\alpha}-1}{1-\alpha}\right\}\sum_{i=1}^{n}w_i>0$ and $\left\{H^{\alpha}(A;W)/\mu_A(x_i)=1\right\}=0$

(iv) $H^{\alpha}(A;W)=0$ for $\mu_A(x_i)=0$ or 1

(v) $H^{\alpha}(A;W)\geq 0$

(vi) $H^{\alpha}(A;W)$ is permutationally symmetric.

Under these conditions, $H^{\alpha}(A;W)$ is a valid measure of weighted fuzzy entropy.

Next, we find the optimum values of the weighted measure (4.2.1).

(a) Maximum value of $H^{\alpha}(A;W)$

Since $H^{\alpha}(A;W)$ is a concave function of $\mu_A(x_i)$ for each α, its maximum value exists.

For maximum value, we put

$\dfrac{\partial H^{\alpha}(A;W)}{\partial\mu_A(x_i)}=0$ which gives $\mu_A(x_i)=1-\mu_A(x_i) \quad\Rightarrow\quad \mu_A(x_i)=\dfrac{1}{2}$

Also $\sum_{i=1}^{n}\mu_A(x_i)=k$ gives $\dfrac{k}{n}=\dfrac{1}{2}=\mu_A(x_i)$

Therefore, the maximum value of $H^{\alpha}(A;W)$ occurs at $k=\frac{n}{2}$ and is given by

$$Max.H^{\alpha}(A;W)=\frac{1}{1-\alpha}\sum_{i=1}^{n}w_i\left[\left(\frac{k}{n}\right)^{\alpha}+\left(1-\frac{k}{n}\right)^{\alpha}-1\right]$$

$$=\frac{1}{n^{\alpha}(1-\alpha)}\left[k^{\alpha}+(n-k)^{\alpha}-n^{\alpha}\right]\sum_{i=1}^{n}w_i \tag{4.2.2}$$

$$=\frac{\sum_{i=1}^{n}w_i}{n^{\alpha}(1-\alpha)}h(k)\text{, where}$$

$$h(k)=k^{\alpha}+(n-k)^{\alpha}-n^{\alpha};$$

$$h'(k)=\alpha\left(k^{\alpha-1}-(n-k)^{\alpha-1}\right);$$

$$h''(k)=\left(\alpha^2-\alpha\right)\left(k^{\alpha-2}+(n-k)^{\alpha-2}\right)$$

Now, we consider the following cases:

(i) When $\alpha>1$, $h''(k)>0\Rightarrow h(k)$ is a convex function of *k* for each $\alpha>1$

$$\Rightarrow\frac{\sum_{i=1}^{n}w_i}{n^{\alpha}(1-\alpha)}h(k)\text{ is a concave function of } k \text{ for each } \alpha>1$$

(ii) When $0<\alpha<1$, $h''(k)<0\Rightarrow h(k)$ is a concave function of *k* for each $0<\alpha<1$

$$\Rightarrow\frac{\sum_{i=1}^{n}w_i}{n^{\alpha}(1-\alpha)}h(k)\text{ is a concave function of } k \text{ for each } 0<\alpha<1$$

Hence, $Max.H^{\alpha}(A;W)$ is a concave function of *k*. Therefore, its maximum value exists and for maximum value, we put

$\frac{d}{dk}\left[Max.H^{\alpha}(A;W)\right]=0$ which gives

$$\Rightarrow\frac{\alpha\sum_{i=1}^{n}w_i}{n^{\alpha}(1-\alpha)}\left[k^{\alpha-1}-(n-k)^{\alpha-1}\right]=0\Rightarrow k^{\alpha-1}-(n-k)^{\alpha-1}=0$$

$$\Rightarrow k=n-k\Rightarrow\frac{k}{n}=\frac{1}{2}$$

Thus, the maximum value of $Max.H^{\alpha}(A;W)$ exists at $k=\frac{n}{2}$ and is given by

$$Max.\left[Max.H^{\alpha}(A;W)\right]=\frac{\sum_{i=1}^{n}w_i}{n^{\alpha}(1-\alpha)}\left[\left(\frac{n}{2}\right)^{\alpha}+\left(\frac{n}{2}\right)^{\alpha}-n^{\alpha}\right]$$

$$=\frac{1}{\alpha-1}\left(\frac{2^{\alpha}-2}{2^{\alpha}}\right)\sum_{i=1}^{n}w_i$$

Further, when

(i) $k=0,\ Max.H^{\alpha}(A;W)=0$ and $\frac{d}{dk}\left(Max.H^{\alpha}(A;W)\right)=\frac{\alpha}{n(\alpha-1)}\sum_{i=1}^{n}w_i\begin{cases}>0, & \alpha>1\\<0, & 0<\alpha<1\end{cases}$

(ii) $k=\frac{n}{2},\ Max.H^{\alpha}(A;W)=\frac{1}{\alpha-1}\left(\frac{2^{\alpha}-2}{2^{\alpha}}\right)\sum_{i=1}^{n}w_i$ and $\frac{d}{dk}\left(Max.H^{\alpha}(A;W)\right)=0$

(iii) $k=n,\ Max.H^{\alpha}(A;W)=0$ and $\frac{d}{dk}\left[Max.H^{\alpha}(A;W)\right]=\frac{\alpha}{n(1-\alpha)}\sum_{i=1}^{n}w_i\begin{cases}<0, & \alpha>1\\>0, & 0<\alpha<1\end{cases}$

Thus, we see that $Max.H^{\alpha}(A;W)$ increases from 0 to $\frac{1}{\alpha-1}\left(\frac{2^{\alpha}-2}{2^{\alpha}}\right)\sum_{i=1}^{n}w_i$ as k increases from 0 to $\frac{n}{2}$ and it decreases from $\frac{1}{\alpha-1}\left(\frac{2^{\alpha}-2}{2^{\alpha}}\right)\sum_{i=1}^{n}w_i$ to 0 as k further increases from $\frac{n}{2}$ to n. Thus, $Max.H^{\alpha}(A;W)$ against the values $0\le k\le n$ is depicted in the following Fig.-4.2.1.

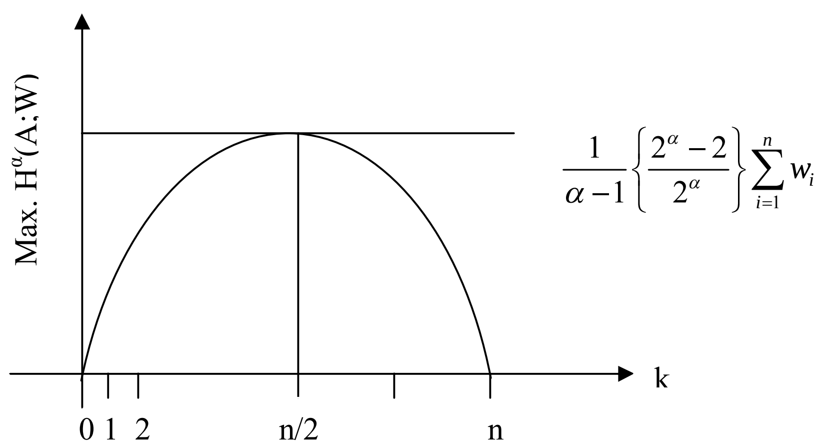


Fig.-4.2.1 $Max.H^{\alpha}(A;W)$

From Fig.-4.2.1, we conclude that $Max.H^{\alpha}(A;W)$ is a concave function of $k(0\le k\le n)$.

(b) Minimum value of $H^{\alpha}(A;W)$

To obtain the minimum value of $H^{\alpha}(A;W)$, we consider the following cases:

Case-I: When k is any non-negative integer, say, $k = m$, then we can have m values of $\mu_A(x_i)$ as unity, and other $(n-m)$ values of $\mu_A(x_i)$ as zero, that is, $\mu_A(x_i) = (1,1,...,1,0,0,...,0)$.

In this case, the minimum value of $H^\alpha(A;W)$ can be obtained as follows:

From equation (4.2.1), we have

$$H^\alpha(A;W) = \frac{1}{1-\alpha}\left[\sum_{i=1}^{m} w_i\left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha - 1\right] + \sum_{i=m+1}^{n} w_i\left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha - 1\right]\right]$$

Thus, $Min.H^\alpha(A;W) = 0$

Case-II: When k is any fraction, then we can write $k = m + \xi$, where m is any non-negative integer and ξ is a positive fraction. Then, we can choose m values $\mu_A(x_i)$ as unity, $(m+1)^{th}$ value as ξ and the remaining $(n-m-1)$ values of $\mu_A(x_i)$ as zero, that is, $\mu_A(x_i) = (1,1,...,1,\xi,0,...,0)$.

In this case, the minimum value of $H^\alpha(A;W)$ can be obtained as follows:

From equation (4.2.1), we have

$$H^\alpha(A;W) = \frac{1}{1-\alpha}\left[\begin{array}{l}\sum_{i=1}^{m} w_i\left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha - 1\right] + w_{m+1}\left(\mu_A^\alpha(x_{m+1}) + (1-\mu_A(x_{m+1}))^\alpha - 1\right) \\ \qquad\qquad + \sum_{i=m+2}^{n} w_i\left[\mu_A^\alpha(x_i) + \left(1-\mu_A(x_i)^\alpha - 1\right)\right]\end{array}\right]$$

Thus, $Min.H^\alpha(A;W) = \frac{1}{1-\alpha} w_{m+1}\left[\xi^\alpha + (1-\xi)^\alpha - 1\right]$

$$= \frac{w_{m+1}}{1-\alpha} k(\xi), \text{ where } k(\xi) = \xi^\alpha + (1-\xi)^\alpha - 1$$

$$k'(\xi) = \alpha\left[\xi^{\alpha-1} - (1-\xi)^{\alpha-1}\right] \ ; \ k''(\xi) = (\alpha^2 - \alpha)\left[\xi^{\alpha-2} + (1-\xi)^{\alpha-2}\right]$$

(i) When $\alpha > 1$, then $k''(\xi) > 0 \Rightarrow k(\xi)$ is a convex function of ξ

$\Rightarrow \frac{w_{m+1}}{1-\alpha} k(\xi)$ is a concave function of ξ for each $\alpha > 1$.

Therefore, in this case, $Min.H^\alpha(A;W)$ is a concave function of ξ for each $\alpha > 1$

(ii) When $0 < \alpha < 1$, $k''(\xi) < 0 \Rightarrow k(\xi)$ is a concave function of ξ

$\Rightarrow \frac{w_{m+1}}{1-\alpha} k(\xi)$ is a concave function of ξ for each $0 < \alpha < 1$. Hence, we see that $Min.H^\alpha(A;W)$ is a concave function of ξ for each α and its maximum value exists. For maximum value, we put

$\frac{d}{d\xi}\left(Min.H^{\alpha}\left(A;W\right)\right)=0$, which gives

$$\frac{\alpha}{1-\alpha}w_{m+1}\left(\xi^{\alpha-1}-\left(1-\xi\right)^{\alpha-1}\right)=0\Rightarrow\xi=\frac{1}{2}$$

Thus, the maximum value of $Min.H^{\alpha}\left(A;W\right)$ exists at $\xi=\frac{1}{2}$ and is given by

$$Max.Min.H^{\alpha}\left(A;W\right)=\frac{w_{m+1}}{1-\alpha}\left[\left(\frac{1}{2}\right)^{\alpha}+\left(\frac{1}{2}\right)^{\alpha}-1\right]=\frac{w_{m+1}(2^{\alpha}-2)}{2^{\alpha}\left(\alpha-1\right)}$$

Further, when (i) $\xi=0$, $Min.H^{\alpha}\left(A;W\right)=0$

(ii) $\xi=\frac{1}{2}$, $Min.H^{\alpha}\left(A;W\right)=\frac{w_{m+1}(2^{\alpha}-2)}{2^{\alpha}\left(\alpha-1\right)}$

(iii) $\xi=1$, $Min.H^{\alpha}\left(A;W\right)=0$

From the above numerical values, we observe that as ξ increases from 0 to $\frac{1}{2}$, the value of $Min.H^{\alpha}\left(A;W\right)$ increases from 0 to $\frac{w_{m+1}(2^{\alpha}-2)}{2^{\alpha}\left(\alpha-1\right)}$ and when ξ further increases from $\frac{1}{2}$ to 1, the value of $Min.H^{\alpha}\left(A;W\right)$ decreases from $\frac{w_{m+1}(2^{\alpha}-2)}{2^{\alpha}\left(\alpha-1\right)}$ to 0.

The presentation of $Min.H^{\alpha}\left(A;W\right)$ against the values $0\le\xi\le1$ is shown in Fig.- 4.2.1 which proves that the $Min.H^{\alpha}\left(A;W\right)$ is a concave function of *k* when *k* is any non-negative fractional value.

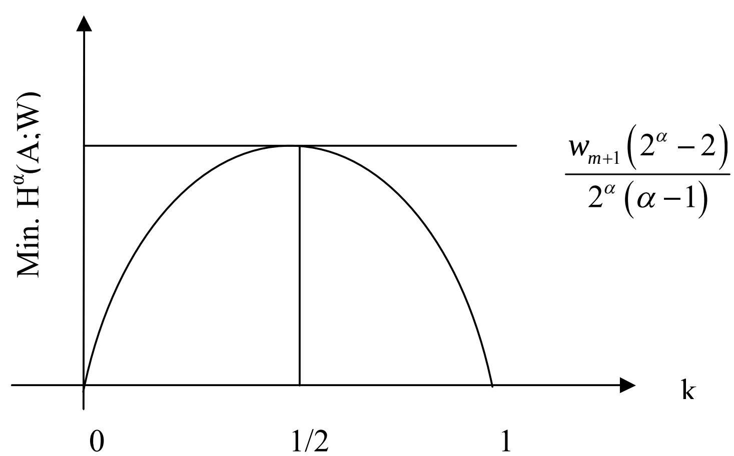


Fig.-4.2.1 $Min.H^{\alpha}\left(A;W\right)$

From the above discussion, we see that $Min.H^{\alpha}\left(A;W\right)$ is a concave function of k which vanishes for every non-negative integer k and has a maximum value of $\frac{w_{m+1}(2^{\alpha}-2)}{2^{\alpha}\left(\alpha-1\right)}$ which occurs for every positive fractional value of k.

Proceeding on similar lines, the optimum values of various existing measures of fuzzy entropy can be obtained.

In the next section, we have considered a generalized weighted measure of fuzzy cross entropy, obtained its optimum values and presented the results graphically.

4.3 OPTIMUM VALUES OF GENERALIZED MEASURES OF WEIGHTED FUZZY DIVERGENCE

Let A be any fuzzy set with n supporting points $\left(x_1,x_2,...,x_n\right)$ corresponding to the fuzzy vector $\left\{\mu_A\left(x_1\right),\mu_A\left(x_2\right),...,\mu_A\left(x_n\right)\right\}$, and B is any fuzzy set with the same n supporting points corresponding to the fuzzy vector $\left\{\mu_B\left(x_1\right),\mu_B\left(x_2\right),...,\mu_B\left(x_n\right)\right\}$. Our purpose is to find the optimum values of some well known measures of generalized weighted fuzzy cross entropy subject to the constraint of total fuzziness, that is, $\sum_{i=1}^{n}\mu_A\left(x_i\right)=k, 0\le k\le n$.

For this purpose, we first introduce a weighted measure of fuzzy directed divergence which corresponds to Havrada and Charvat's [9] measure of probabilistic divergence. This measure of fuzzy cross entropy is given by the following mathematical expression:

$$D^{\alpha}\left(A:B;W\right)=\frac{1}{\alpha-1}\sum_{i=1}^{n}w_i\left[\mu_A^{\alpha}\left(x_i\right)\mu_B^{1-\alpha}\left(x_i\right)+\left(1-\mu_A\left(x_i\right)\right)^{\alpha}\left(1-\mu_B\left(x_i\right)\right)^{1-\alpha}-1\right];\alpha\ne1,\alpha>0 \qquad (4.3.1)$$

where w_i are non-negative real numbers.

We first claim that the proposed measure given by equation (4.3.1) is a valid measure of weighted fuzzy cross entropy. To prove its authenticity, we study the following properties:

I. Convexity of $D^{\alpha}\left(A:B;W\right)$

To prove the convexity of the measure (4.3.1), we proceed as follows:

Let $f\left(x,y\right)=\left\{x^{\alpha}y^{1-\alpha}+\left(1-x\right)^{\alpha}\left(1-y\right)^{1-\alpha}-1\right\};0\le x\le1,\ 1\le y\le1$

Differentiating the above equation partially w.r.t. x, we get

$$\frac{\partial f}{\partial x}=\alpha\left(x^{\alpha-1}y^{1-\alpha}-(1-x)^{\alpha-1}(1-y)^{1-\alpha}\right);\frac{\partial^2 f}{\partial x^2}=\left(\alpha^2-\alpha\right)\left(x^{\alpha-2}y^{1-\alpha}+(1-x)^{\alpha-2}(1-y)^{1-\alpha}\right)$$

(i) When $\alpha>1$, $\frac{\partial^2 f}{\partial x^2}>0\ \forall\ 0\le x\le 1, 0\le y\le 1 \Rightarrow f(x,y)$ is a convex function of x. Since, $w_i\ge 0$, we claim that $\frac{1}{\alpha-1}\sum_{i=1}^{n} w_i f\left(\mu_A(x_i),\mu_B(x_i)\right)$ is also a convex function of $\mu_A(x_i)$ for all $0\le\mu_A(x_i)\le 1$.

(ii) When $0<\alpha<1$, then $\frac{\partial^2 f}{\partial x^2}<0\ \forall\ 0\le x\le 1,\ 0\le y\le 1 \Rightarrow f(x,y)$ is a concave function of x

$\Rightarrow \frac{1}{\alpha-1}\sum_{i=1}^{n} w_i f\left(\mu_A(x_i),\mu_B(x_i)\right)$ is a convex function of $\mu_A(x_i)$ for each $0\le\mu_A(x_i)\le 1$.

Thus, $D^{\alpha}(A:B;W)$ is a convex function of $\mu_A(x_i)$ for each α. Similarly, it can be proved that $D^{\alpha}(A:B;W)$ is a convex function of $\mu_B(x_i)$ for each α.

II. Non-negativity: To prove the non-negativity property, we first find the minimum value of $D^{\alpha}(A:B;W)$. For minimum value, we put

$\frac{\partial D^{\alpha}(A:B;W)}{\partial\mu_A(x_i)}=0$ which gives $\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)=\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)$

that is, only if $\mu_A(x_i)=\mu_B(x_i)\ \forall i$

Now, when $\mu_A(x_i)=\mu_B(x_i)\ \forall i$, we have $D^{\alpha}(A:B;W)=0$**.** Since, the minimum value of the function $D^{\alpha}(A:B;W)$ is 0 and the function itself is convex, we must have $D^{\alpha}(A:B;W)\ge 0$.

Thus, $D^{\alpha}(A:B;W)$ satisfies the following properties:

(i) $D^{\alpha}(A:B;W)\ge 0$

(ii) $D^{\alpha}(A:B;W)=0$ iff $\mu_A(x_i)=\mu_B(x_i)\ \forall i$

(iii) $D^{\alpha}(A:B;W)$ is a convex function of both $\mu_A(x_i)$ and $\mu_B(x_i)$ for each α.

Hence, the proposed measure $D^{\alpha}(A:B;W)$ is a valid measure of weighted fuzzy cross entropy.

Next, we obtain the optimum values of the measure (4.3.1).

(a) Minimum value of $D^{\alpha}(A:B;W)$

Since $D^{\alpha}(A:B;W)$ is a convex function of both $\mu_A(x_i)$ and $\mu_B(x_i)$, its minimum value exists. For minimum value, we put

$\frac{\partial D^{\alpha}(A:B;W)}{\partial \mu_A(x_i)}=0$ which gives $\frac{\alpha}{\alpha-1}w_i\left[\mu_A^{\alpha-1}(x_i)\mu_B^{1-\alpha}(x_i)-(1-\mu_A(x_i))^{\alpha-1}(1-\mu_B(x_i))^{1-\alpha}\right]=0$

This gives that $\mu_A^{\alpha-1}(x_i)\mu_B^{1-\alpha}(x_i)-(1-\mu_A(x_i))^{\alpha-1}(1-\mu_B(x_i))^{1-\alpha}=0$

Now, since $\mu_B(x_i)$ is a fuzzy value, it is fixed and known. Let it be C.

Thus, we have $C\left[\mu_A^{\alpha-1}(x_i)-(1-\mu_A(x_i))^{\alpha-1}\right]=0$ since $\mu_B(x_i)=1-\mu_B(x_i)$

The above equation holds if $\mu_A(x_i)=1-\mu_A(x_i) \Rightarrow \mu_A(x_i)=\frac{1}{2}$

Also $\sum_{i=1}^{n}\mu_A(x_i)=k$ gives $\frac{n}{2}=k \Rightarrow \frac{n}{k}=\frac{1}{2}=\mu_A(x_i)$

Thus, the minimum value of $D^{\alpha}(A:B;W)$ exists at $k=\frac{n}{2}$ and is given by

$$Min.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\sum_{i=1}^{n}w_i\left[\left(\frac{k}{n}\right)^{\alpha}\mu_B^{1-\alpha}(x_i)+\left(1-\frac{k}{n}\right)^{\alpha}(1-\mu_B(x_i))^{1-\alpha}-1\right]$$

Further, when

(i) $k=0,\ Min.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\sum_{i=1}^{n}w_i\left[(1-\mu_B(x_i))^{1-\alpha}-1\right]$

(ii) $k=\frac{n}{2},\ Min.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\sum_{i=1}^{n}w_i\left[\left(\frac{1}{2}\right)^{\alpha}\mu_B^{1-\alpha}(x_i)+\left(\frac{1}{2}\right)^{\alpha}(1-\mu_B(x_i))^{1-\alpha}-1\right]$

(iii) $k=n,\quad Min.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\sum_{i=1}^{n}w_i\left[\mu_B^{1-\alpha}(x_i)-1\right]$

Illustration: Suppose $\mu_B(x_i)$ is monotonically decreasing function of its fuzzy values, that is, $\mu_B(x_i)=\left(\frac{1}{2},\frac{1}{3},\frac{1}{4},\ldots,\frac{1}{n+1}\right)$. Also without any loss of generality, we can assume the weighted distribution to be $W=\{1,2,3,\ldots,n\}$.

Thus, we have the following cases:

(i) When $k=0, Min.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\sum_{i=1}^{n}w_i\left[(1-\mu_B(x_i))^{1-\alpha}-1\right]$

$$=\frac{1}{\alpha-1}\left[2^{\alpha-1}+2\left(\frac{3}{2}\right)^{\alpha-1}+3\left(\frac{4}{3}\right)^{\alpha-1}+\ldots.+n\left(\frac{n+1}{n}\right)^{\alpha-1}-\frac{n(n+1)}{\lfloor 2}\right]$$

$=\phi_1(n)$ (say)

(ii) When $k=\frac{n}{2}$, $Min.D^{\alpha}\left(A:B;W\right)=\frac{1}{\alpha-1}\sum_{i=1}^{n}w_i\left[\left(\frac{1}{2}\right)^{\alpha}\mu_B^{1-\alpha}\left(x_i\right)+\left(\frac{1}{2}\right)^{\alpha}\left(1-\mu_B\left(x_i\right)\right)^{1-\alpha}-1\right]$

$$=\frac{1}{(\alpha-1)}\left[\begin{array}{l}1+2^{-\alpha}.2\left(3^{\alpha-1}+\left\{\frac{3}{2}\right\}^{\alpha-1}\right)+2^{-\alpha}.3\left(4^{\alpha-1}+\left\{\frac{4}{3}\right\}^{\alpha-1}\right)+\ldots+\\ 2^{-\alpha}.n\left((n+1)^{\alpha-1}+\left\{\frac{n+1}{n}\right\}^{\alpha-1}\right)-\frac{n(n+1)}{\lfloor 2}\end{array}\right]=\phi_2(n)\quad \text{(say)}$$

(iii) When $k=n, Min.D^{\alpha}\left(A:B;W\right)=\frac{1}{\alpha-1}\sum_{i=1}^{n}w_i\left[\mu_B^{1-\alpha}\left(x_i\right)-1\right]$

$$=\frac{1}{\alpha-1}\left(2^{\alpha-1}+2.3^{\alpha-1}+3.4^{\alpha-1}+....+n.\left(n+1\right)^{\alpha-1}-\frac{n(n+1)}{\lfloor 2}\right)=\phi_3(n)\quad \text{(say)}$$

From the above discussion, we observe the behavior of the $Min.D^{\alpha}\left(A:B;W\right)$ and conclude that the $Min.D^{\alpha}\left(A:B;W\right)$ first increases at $k=0$ and then decreases at $k=n/2$ and again increases at $k=n$.

Numerical Verification

Case-I: For $n=2$ and for different values of the parameter $\alpha>1$, we have computed different values of $Min.D^{\alpha}\left(A:B;W\right)$ as shown in Table-4.3.1.

Table-4.3.1

α	$\phi_1(n)$	$\phi_2(n)$	$\phi_3(n)$
2	2.00	0.25	5.00
3	2.75	0.41	9.50
4	3.92	0.60	19.67
5	5.78	0.84	43.75
6	8.84	1.17	103.00

Case-II: For $n = 3$ and for different values of $\alpha < 1$, we have computed different values of $Min.D^{\alpha}\left(A:B;W\right)$ as shown in Table-4.3.2.

Table-4.3.2

α	$\phi_1(n)$	$\phi_2(n)$	$\phi_3(n)$
0.2	2.00	0.10	4.51
0.3	2.04	0.15	4.74
0.4	2.08	0.21	5.00
0.5	2.12	0.26	5.28
0.6	2.17	0.32	5.58

From the above tables, we present the $Min.D^{\alpha}\left(A:B;W\right)$ graphically against the values $0 \le k \le n$ as shown in the following Fig.-4.3.1.

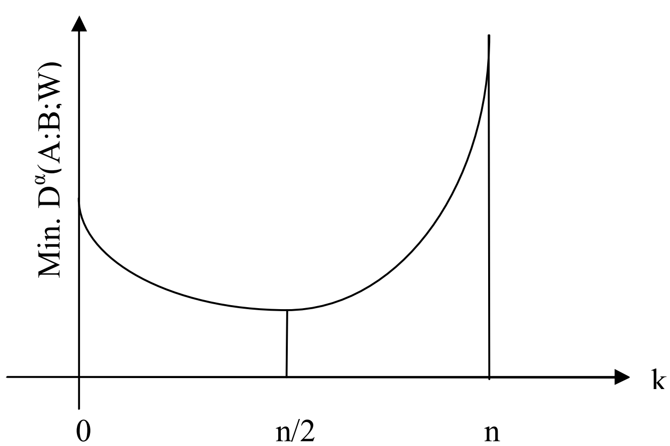


Fig.-4.3.1 $Min.D^{\alpha}\left(A:B;W\right)$

The above Fig.-4.3.1 clearly indicates that $Min.D^{\alpha}\left(A:B;W\right)$ is a continuous and piecewise convex function of k.

(b) Maximum Value of $D^{\alpha}(A:B;W)$

To obtain the maximum value of $D^{\alpha}(A:B;W)$, we consider the following cases:

Case I : When k is any positive integer, say $k=m$, then, we can choose m values of $\mu_A(x_i)$ as unity and remaining $n-m$ values as zero, that is, $\mu_A(x_i)=(1,1,1,...,1,0,0,...,0)$

In this case, from equation (4.3.1), we have

$$D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\left[\begin{array}{l}\sum_{i=1}^{m} w_i\left[\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}-1\right]\\ +\sum_{i=m+1}^{n} w_i\left[\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}-1\right]\end{array}\right]$$

Thus,

$$Max.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\left[\sum_{i=1}^{m} w_i\left\{\mu_B^{1-\alpha}(x_i)-1\right\}+\sum_{i=m+1}^{n} w_i\left\{(1-\mu_B(x_i))^{1-\alpha}-1\right\}\right] \qquad (4.3.2)$$

Illustration: Suppose $\mu_B(x_i)$ is monotonically decreasing function of its fuzzy values, that is, $\mu_B(x_i)=\left(\frac{1}{2},\frac{1}{3},\frac{1}{4},...,\frac{1}{n+1}\right)$ and without any loss of generality, we can assume that $W=(1,2,3,...,n)$.

Now, we discuss the following cases:

(i) When $m=1$, equation (4.3.2) gives

$$\begin{aligned}Max.D^{\alpha}(A:B;W)&=\frac{1}{\alpha-1}\left[w_1\left\{\mu_B^{1-\alpha}(x_i)-1\right\}+\sum_{i=2}^{n} w_i\left[(1-\mu_B(x_i))^{1-\alpha}-1\right]\right]\\ &=\frac{1}{\alpha-1}\left[2^{\alpha-1}+2\left(\frac{3}{2}\right)^{\alpha-1}+3\left(\frac{4}{3}\right)^{\alpha-1}+....+n\left(\frac{n+1}{n}\right)^{\alpha-1}-\frac{n(n+1)}{\lfloor 2}\right]\\ &=\phi_1(n)\text{, say}\end{aligned}$$

(ii) When $m=2$, equation (4.3.2) gives

$$\begin{aligned}Max.D^{\alpha}(A:B;W)&=\frac{1}{\alpha-1}\left\{w_1\left\{\mu_B^{1-\alpha}(x_1)-1\right\}+w_2\left\{\mu_B^{1-\alpha}(x_2)-1\right\}+\sum_{i=3}^{n} w_i\left[(1-\mu_B(x_i))^{1-\alpha}-1\right]\right\}\\ &=\frac{1}{\alpha-1}\left[2^{\alpha-1}+2.3^{\alpha-1}+3.\left(\frac{4}{3}\right)^{\alpha-1}+4.\left(\frac{5}{4}\right)^{\alpha-1}+....+n.\left(\frac{n+1}{n}\right)^{\alpha-1}-\frac{n(n+1)}{\lfloor 2}\right]\\ &=\phi_2(n)\text{, say}\end{aligned}$$

(iii) When $m=3$, equation (4.3.2) gives

$$Max.D^{\alpha}\left(A:B;W\right)=\frac{1}{\alpha-1}\left[2^{\alpha-1}+2.3^{\alpha-1}+3.4^{\alpha-1}+4.\left(\frac{5}{4}\right)^{\alpha-1}+5.\left(\frac{6}{5}\right)^{\alpha-1}+....+n\left(\frac{n+1}{n}\right)^{\alpha-1}-\frac{n(n+1)}{\lfloor 2}\right]$$

$$=\phi_3(n)\text{, say}$$

(iv) When $m=n$, equation (4.3.2) gives

$$Max.D^{\alpha}\left(A:B;W\right)=\frac{1}{\alpha-1}\left[2^{\alpha-1}+2.3^{\alpha-1}+3.4^{\alpha-1}+...+n\left(n+1\right)^{\alpha-1}-\frac{n(n+1)}{\lfloor 2}\right]$$

$$=\phi_n(n)\text{, say}$$

From the above discussion, we observe that for every positive value of α and for different values of m, the $Max.D^{\alpha}\left(A:B;W\right)$ goes on increasing.

The graph of $Max.D^{\alpha}\left(A:B;W\right)$ against the values $0\le k\le n$ is shown in the following Fig.-4.3.2.

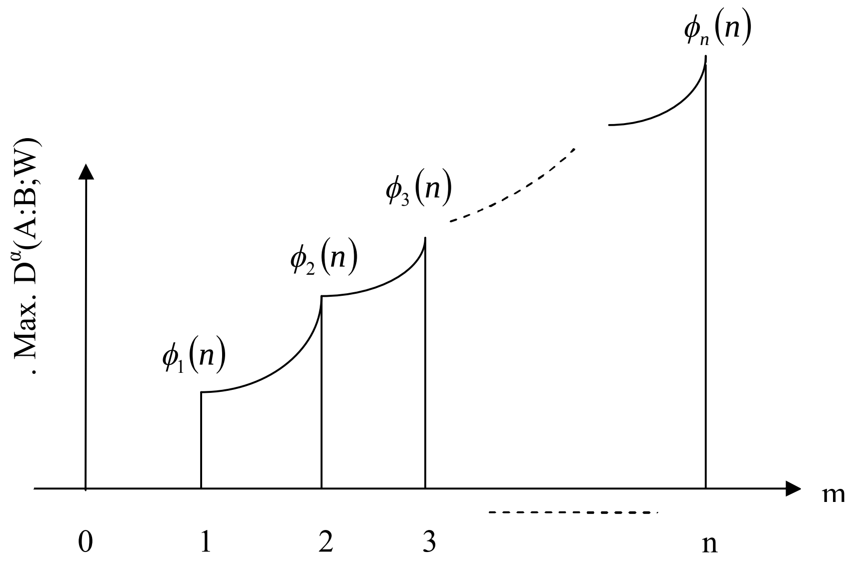


Fig.-4.3.2 $Max.D^{\alpha}\left(A:B;W\right)$

From Fig.-4.3.2, we conclude that $Max.D^{\alpha}\left(A:B;W\right)$ is monotonically increasing function and is continuous and piecewise convex function of k when $\mu_B\left(x_i\right)$ is monotonically decreasing function of its fuzzy values.

Case II. When $k=m+\xi$, where m is any non-negative integer and ξ is a positive fraction. We put m values of $\mu_A\left(x_i\right)$ as unity, $\left(m+1\right)^{th}$ value as ξ and the remaining $\left(n-m-1\right)$ values as zero, that is,

$\mu_A\left(x_i\right)=\left(1,1,...,1,\xi,0,0,...,0\right)$.

Now, using equation (4.3.1), we have

$$D^{\alpha}\left(A:B;W\right)=\frac{1}{\alpha-1}\left[\begin{array}{l}\sum_{i=1}^{m}w_i\left[\mu_A^{\alpha}\left(x_i\right)\mu_B^{1-\alpha}\left(x_i\right)+\left(1-\mu_A\left(x_i\right)\right)^{\alpha}\left(1-\mu_B\left(x_i\right)\right)^{1-\alpha}-1\right] \\ +w_{m+1}\left\{\mu_A^{\alpha}\left(x_{m+1}\right)\mu_B^{1-\alpha}\left(x_{m+1}\right)+\left(1-\mu_A\left(x_{m+1}\right)\right)^{\alpha}\left(1-\mu_B\left(x_{m+1}\right)\right)^{1-\alpha}-1\right\} \\ +\sum_{i=m+2}^{n}w_i\left[\mu_A^{\alpha}\left(x_i\right)\mu_B^{1-\alpha}\left(x_i\right)+\left(1-\mu_A\left(x_i\right)\right)^{\alpha}\left(1-\mu_B\left(x_i\right)\right)^{1-\alpha}-1\right]\end{array}\right]$$

Thus, we have

$$Max.D^{\alpha}\left(A:B;W\right)=\frac{1}{\alpha-1}\left[\begin{array}{l}\sum_{i=1}^{m}w_i\left\{\mu_B^{1-\alpha}\left(x_i\right)-1\right\}+w_{m+1}\left\{\xi^{\alpha}\mu_B^{1-\alpha}\left(x_{m+1}\right)+\left(1-\xi\right)^{\alpha}\left(1-\mu_B\left(x_{m+1}\right)\right)^{1-\alpha}-1\right\} \\ +\sum_{i=m+2}^{n}w_i\left[\left(1-\mu_B\left(x_i\right)\right)^{1-\alpha}-1\right]\end{array}\right] \tag{4.3.3}$$

Illustration: Suppose $\mu_B\left(x_i\right)$ is monotonically decreasing function of its fuzzy values, that is,

$\mu_B\left(x_i\right)=\left(\frac{1}{2},\frac{1}{3},\frac{1}{4},...,\frac{1}{m+1},\frac{1}{m+2},\frac{1}{m+3},...,\frac{1}{n+1}\right)$, and without any loss of generality, we assume that $W=\left(1,2,3,...,m,m+1,m+2,...,n\right)$, then from equation (4.3.3), we have

$$Max.D^{\alpha}\left(A:B;W\right)=\frac{1}{\alpha-1}\left(2^{\alpha-1}+2.3^{\alpha-1}+....+m\left(m+1\right)^{\alpha-1}\right)$$

$$+\frac{1}{\alpha-1}\left[(m+2)\left(\frac{m+3}{m+2}\right)^{\alpha-1}+(m+3)\left(\frac{m+4}{m+3}\right)^{\alpha-1}+...+n\left(\frac{n+1}{n}\right)^{\alpha-1}-\frac{n(n+1)}{\underline{|2}}\right]+\phi\left(\xi\right)$$

$$where\ \phi\left(\xi\right)=\frac{m+1}{\alpha-1}\left(\xi^{\alpha}\left(m+2\right)^{\alpha-1}+\left(1-\xi\right)^{\alpha}\left(\frac{m+2}{m+1}\right)^{\alpha-1}\right)$$

Convexity of $\phi\left(\xi\right)$: We first prove that the function $\phi\left(\xi\right)$ is a convex function of ξ for each value of α. For this, we have

$$\phi^{'}\left(\xi\right)=\frac{\alpha(m+1)}{\alpha-1}\left(\xi^{\alpha-1}\left(m+2\right)^{\alpha-1}-\left(1-\xi\right)^{\alpha-1}\left(\frac{m+2}{m+1}\right)^{\alpha-1}\right)$$

Also

$$\phi^{''}\left(\xi\right)=\alpha(m+1)\left(\xi^{\alpha-2}\left(m+2\right)^{\alpha-1}+\left(1-\xi\right)^{\alpha-2}\left(\frac{m+2}{m+1}\right)^{\alpha-1}\right)>0$$

This shows that $\phi(\xi)$ is a convex function of ξ. Hence, its minimum value exits. For minimum value, we have $\phi^{'}(\xi)=0$ which gives $\xi=\frac{1}{m+2}$.

Further, we have

$$Min.Max.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\left[\begin{array}{l}2^{\alpha-1}+2.3^{\alpha-1}+3.4^{\alpha-1}+....+m(m+1)^{\alpha-1}+(m+2)\left(\frac{m+3}{m+2}\right)^{\alpha-1}\\ +(m+3)\left(\frac{m+4}{m+3}\right)^{\alpha-1}+....+n\left(\frac{n+1}{n}\right)^{\alpha-1}-\frac{n(n+1)}{\lfloor 2}\end{array}\right]+\frac{m+1}{\alpha-1} \tag{4.3.4}$$

(i) When $m=0$, equation (4.3.4) gives

$$Min.Max.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\left[2\left(\frac{3}{2}\right)^{\alpha-1}+3\left(\frac{4}{3}\right)^{\alpha-1}+...+n\left(\frac{n+1}{n}\right)^{\alpha-1}-\left\{1-\frac{n(n+1)}{\lfloor 2}\right\}\right]$$

$$=g_0(\alpha),\text{ say}$$

(ii) When $m=1$, equation (4.3.4) gives

$$Min.Max.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\left[2^{\alpha-1}+3\left(\frac{4}{3}\right)^{\alpha-1}+4\left(\frac{5}{4}\right)^{\alpha-1}+...+n\left(\frac{n+1}{n}\right)^{\alpha-1}-\left\{2-\frac{n(n+1)}{\lfloor 2}\right\}\right]$$

$$=g_1(\alpha),\text{ say}$$

(iii) When $m=2$, equation (4.3.4) gives

$$Min.Max.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\left[2^{\alpha-1}+2.3^{\alpha-1}+4\left(\frac{5}{4}\right)^{\alpha-1}+5\left(\frac{6}{5}\right)^{\alpha-1}+...+n\left(\frac{n+1}{n}\right)^{\alpha-1}-\left\{3-\frac{n(n+1)}{\lfloor 2}\right\}\right]$$

$$=g_2(\alpha),\text{ say}$$

(iv) When $m=3$, equation (4.3.4) gives

$$Min.Max.D^{\alpha}(A:B;W)=\frac{1}{\alpha-1}\left[\begin{array}{l}2^{\alpha-1}+2.3^{\alpha-1}+3.4^{\alpha-1}+5\left(\frac{6}{5}\right)^{\alpha-1}+6\left(\frac{7}{6}\right)^{\alpha-1}+...+n\left(\frac{n+1}{n}\right)^{\alpha-1}\\ -\left\{4-\frac{n(n+1)}{\lfloor 2}\right\}\end{array}\right]$$

$$=g_3(\alpha),\text{ say}$$

(v) When $m=n$, equation (4.3.4) gives

$$Min.Max.D^{\alpha}\left(A:B;W\right)=\frac{1}{\alpha-1}\left[2^{\alpha-1}+2.3^{\alpha-1}+3.4^{\alpha-1}+...+n\left(n+1\right)^{\alpha-1}-\left\{(n+1)-\frac{n(n+1)}{\underline{|2}}\right\}\right]$$

$$=g_n\left(\alpha\right),\text{ say}$$

From the above discussion, we observe that for every term of the $Min.Max.D^{\alpha}\left(A:B;W\right)$ goes on increasing. The graph of $Min.Max.D^{\alpha}\left(A:B;W\right)$ against the values of $0\le k\le n$ is shown in the following Fig.-4.3.3.

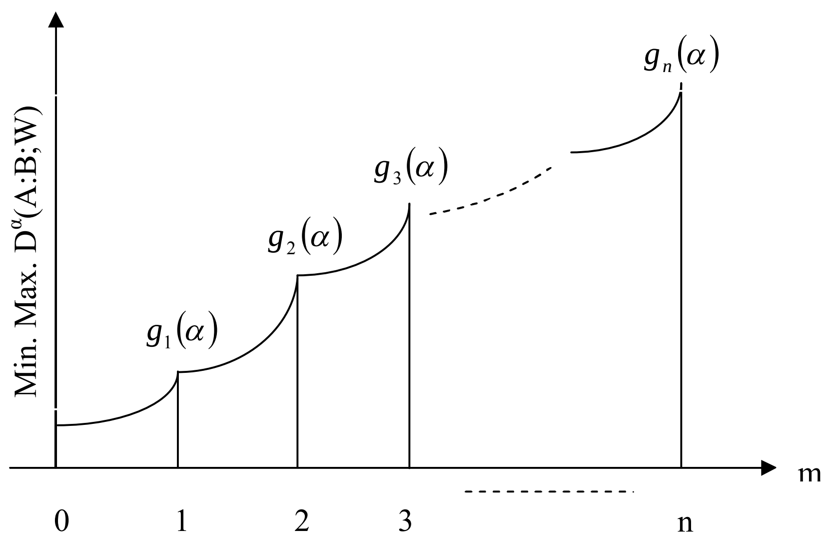


Fig.-4.3.3 $Min.Max.D^{\alpha}\left(A:B;W\right)$

From Fig.-4.3.3, we conclude that $Min.Max.D^{\alpha}\left(A:B;W\right)$ is monotonically increasing function when $\mu_B\left(x_i\right)$ is monotonically decreasing function and also this function is a continuous and piecewise convex function of k.

Concluding Remarks: In the present chapter, it has been proved that the maximum value of the generalized weighted fuzzy entropy subject to the constraint of total fuzziness, that is, $\sum_{i=1}^{n}\mu_A\left(x_i\right)=k$, is a continuous and concave function of k while the minimum value of the generalized weighted fuzzy entropy is continuous and concave function which vanishes for every non-negative value of k and has a maximum value for any non-negative fractional value of k. Further for the case of cross entropy, it has been proved that the maximum value of the generalized weighted fuzzy cross entropy is a continuous and piecewise convex function of k and is monotonically increasing function when it is given that

$\mu_B(x_i)$ is monotonically decreasing function of its fuzzy values. With similar discussions, the optimum values of the other measures of entropy and cross entropy can be obtained.

REFERENCES

[1] Bhandari, D., Pal, N. R., & Majumder, D. D. (1992). Fuzzy divergence, probability measure of fuzzy events and image thresholding. *Pattern Recognition Letters, 1*, 857-867.

[2] De Luca, A., & Termini, S. (1972). A definition of non-probabilistic entropy in setting of fuzzy set theory. *Information and Control, 20*, 301-312.

[3] Dubois, D., & Prade, H. (1983). On distances between fuzzy points and their use for plausible reasoning. *Proceedings of International Conference on Systems, Man and Cybernetics* (pp.300-303).

[4] Ebanks, B. R. (1983). On measures of fuzziness and their representations. *Journal of Mathematical Analysis and Applications, 94*, 24-37.

[5] Fan, J. L., Ma, Y. L., & Xie, W. X. (2001). On some properties of distance measure. *Fuzzy Sets and Systems, 117*, 355-361.

[6] Ferreri, C. (1980). Hypoentropy and related heterogeneity divergence measures. *Statistica, 40*, 155-168.

[7] Guiasu, S., & Shenitzer, A. (1975). The principle of maximum entropy. *Mathematical Intelligencer, 7*, 42-48.

[8] Guo, X. Z., & Xin, X. L. (2006). Some new generalized entropy formulas of fuzzy sets. *Journal of the Northwest University, 36*, 529-532.

[9] Havrada, J. H., & Charvat, F. (1967). Quantification methods of classification process: Concept of structural α-entropy. *Kybernetika, 3*, 30-35.

[10] Kapur, J. N. (1989). *Maximum entropy models in science and engineering*. New Delhi: Wiley Eastern.

[11] Kapur, J. N. (1997). *Measures of fuzzy information*. New Delhi: Mathematical Sciences Trust Society.

[12] Kapur, J. N., & Kesavan, H. K. (1987). *The generalized maximum entropy Principle*. Canada: Standard Educational Press, University of Waterloo.

[13] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*, 79-86.

[14] Parkash, O. (1998). A new parametric measure of fuzzy entropy. *Information Processing and Management of Uncertainty, 2*, 1732-1737.

[15] Parkash, O. (2000). On fuzzy symmetric divergence. *The Fourth Asian Fuzzy System Symposium, 2*, 904-908.

[16] Parkash, O., & Sharma, P. K. (2004). Measures of fuzzy entropy and their relations. *Inernationa. Journal of Management & Systems, 20*, 65-72.

[17] Parkash, O., & Sharma, P. K. (2005). Some new measures of fuzzy directed divergence and their generalization. *Journal of the Korean Society of Mathematical Education Series B*, *12*, 307-315.

[18] Parkash, O., Sharma, P. K., & Mahajan, R. (2008). New measures of weighted fuzzy entropy and their applications for the study of maximum weighted fuzzy entropy principle. *Information Sciences, 178*, 2389-2395.

[19] Parkash, O., Sharma, P. K., & Mahajan, R. (2010). Optimization principle for weighted fuzzy entropy using unequal constraints. *Southeast Asian Bulletin of Mathematics, 34*, 155-161.

[20] Renyi, A. (1961). On measures of entropy and information. *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability* (pp.547-561).

[21] Rosenfeld, A. (1985). Distance between fuzzy sets. *Pattern Recognition Letters, 3*, 229-231.

[22] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-659.

[23] Sharma, B. D., & Taneja, I. J. (1975). Entropies of type (α,β) and other generalized measures of information theory. *Metrika, 22*, 202-215.

[24] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338-353.

[25] Zadeh, L. A. (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications, 23*, 421-427.

CHAPTER-V

APPLICATIONS OF INFORMATION MEASURES TO PORTFOLIO ANALYSIS AND QUEUEING THEORY

ABSTRACT

In the present chapter, we have developed optimizational principles using new divergence measures and consequently, provided the applications of these measures for the development of measures of risk in portfolio analysis for minimizing risk. We have observed that minimizing these measures implies the minimization of the expected utility of a risk-prone person and maximization of the expected utility of a risk-averse person. Moreover, we have provided the applications of information measures to the field of queueing theory and proved that in case of steady state queueing process, as the arrival rate increases relatively to service rate, the uncertainty increases whereas in the case of non-steady birth-death process, the uncertainty measure first increases and attains its maximum value and then with the passage of time, it decreases and attains its minimum value.

***Keywords:* Portfolio analysis, Divergence measure, Covariance matrix, Mean–Variance efficient frontier, Birth-death process, Steady state, Non-steady state.**

5.1 INTRODUCTION

The portfolio theory deals with the investments which attempt to maximize portfolio expected return for a given amount of portfolio risk, or equivalently minimize risk for a given level of expected return, by carefully choosing the proportions of various assets. In portfolio analysis, which is usually divided into two stages, an investor has a great deal of choice in selecting his strategy for investment and wishes to maximize profits. The first stage starts with observation and experience and ends with the beliefs about the future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with the choice of portfolio.

The modern portfolio selection theory founded by Markowitz [13], deals with the second stage in which we first feel that the investor should maximize discounted expected returns. We next consider the rule that the investor should consider expected return a desirable thing and variance of return an undesirable one. Markowitz [13] illustrated geometrically relations between beliefs and choice of portfolio according to the "expected returns-variance of returns" rule. Classical formulations of the portfolio optimization problem, such as mean-variance can result in a portfolio extremely sensitive to errors in the data, such as mean and covariance matrix of the returns.

It is worth mentioning here that some of the investments made by the investor may yield low returns, but these returns may be compensated by considerations of relative safety because of a proven record of non-volatility in price fluctuations. On the other hand, there might be some better investments which would be promising ones and achieve high expected returns, but these may be prone to a great deal of risk. The process may not be much difficult in maximizing the expected return since the various outcomes and the probabilities of their outcomes and the return on a unit amount invested in each security are known. However, investor’s major problem is to find a measure of risk which can be most satisfactory according to his investments. The earliest measure proposed for the return on all investments was variance and its proposal was based upon the fundamental argument that the risk increases with variance. Accordingly, in his seminal work, Markowitz [13] presented the concept of mean-variance efficient frontier, which enabled him to find all the possible efficient portfolios that simultaneously maximize the expected returns and minimize the variance.

Jianshe [9] developed a new theory of portfolio and risk based on incremental entropy and Markowitz's [13] theory by replacing arithmetic mean return adopted by Markowitz [13], with geometric mean return as a criterion for assessing a portfolio. The new theory emphasizes that there is an objectively optimal portfolio for given probability of returns. Some portfolio optimization methodology has been discussed by Bugár and Uzsoki [4] whereas other work related with diversification of investments has been provided by Markowitz [14]. More regarding with the study of portfolio analysis, Nocetti [16] has explored the possibility that how investors allocate mental effort to learn about the mean return of a number of assets and analyzed how this allocation changes the portfolio selection problem. Bera and Park [1] remarked that Markowitz's [13] mean-variance efficient portfolio selection is one of the most widely used approaches in solving portfolio diversification problem. However, contrary to the notion of diversification, mean-variance approach often leads to portfolios highly concentrated on a few assets. In their communication, Bera and Park [1] proposed to use cross entropy measure as the objective function with side conditions coming from the mean and variance-covariance matrix of the resampled asset returns and illustrated their procedure with an application to the international equity indexes. Now, since risk is associated with the concept of uncertainty, we should be able to develop measures of risk based on the concepts of divergence or cross-entropy. We can develop such measures of divergence and then show how we can develop efficient frontiers for maximizing expected returns and simultaneously minimize measures of risk.

In the literature, there exist many well known measures of divergence which find their tremendous applications to various disciplines of mathematical sciences and the most appropriate found to be useful in many real life situations is due to Kullback-Leibler's [12], usually known as measure of distance or cross entropy. This measure is given by

$$D(P;Q)=\sum_{i=1}^{n} p_i \log\frac{p_i}{q_i}. \tag{5.1.1}$$

In Chapter-I, we have introduced a new non-parametric measure of divergence, given by

$$D_1(P;Q)=\sum_{i=1}^{n}\left(\frac{p_i^2}{q_i}+\frac{q_i^2}{p_i}-2q_i\right). \tag{5.1.2}$$

Shore and Gray [21] provided the applications of cross entropy measure to engineering problems related with pattern classification and cluster analysis. The authors have delivered an approach which is based on the minimization of cross-entropy and can be viewed as a refinement of a general classification method due to Kullback [11]. The refinement exploits special properties of cross-entropy that hold when the probability densities involved happen to be minimum cross-entropy densities. The authors have commented that "The special properties of cross-entropy that hold for minimum cross-entropy densities result in a pattern classification method with several advantages. It is optimal in a well defined, information-theoretic sense; it is computationally attractive and it includes a self-consistent, simple method of computing the set of cluster centroids in terms of which the classification is made. Further, this method can be used successfully in a variety of other disciplines".

Some applications of minimum cross entropy principle have been provided by Shore and Johnson [22] where the authors have remarked that the principle of minimum cross-entropy is a general method of inference about an unknown probability density when there exists a prior estimate of the density and new information in the form of constraints on expected values. The authors have extended cross-entropy's well-known properties as an information measure and strengthened when one of the densities involved is the result of cross-entropy minimization. Moreover, the authors have pointed out the interplay between properties of cross-entropy minimization as an inference procedure and properties of cross-entropy as an information measure by providing examples and general analytic and computational methods of finding minimum cross-entropy probability densities. Carlson and Clements [5] provided the applications of cross entropy in another field of engineering known as signal processing. The authors have remarked that "an important component in the application and design of speech processing systems is the comparision of two signals in the form of a similarity measure and a suitable measure for such comparision comes from information and detection theory in the form of the directed divergence which measures the discriminating information between two signal classes".

It has been observed that generalized measures of cross entropy should be introduced because upon optimization, these measures lead to useful probability distributions and mathematical models in various disciplines and also introduce flexibility in the system. Moreover, there exist a variety of mathematical models in science and engineering and a large number of models applicable in economics, social sciences, biology and even in physical sciences, for each of which, a single measure of cross-entropy cannot be adequate. Thus, we need a variety of generalized parametric measures of cross-entropy to extend the scope of their applications. For example, in ecology, we have to measure the difference in the relative frequencies of different spices. In signal processing, one has to find the difference between the spectral density functions. Bhattacharya's [3] measure of divergence and Kullback-Liebler's [12] measure of cross-entropy have extensively been used in the field of pattern recognition. Similarly, each parametric and non-parametric measure of cross-entropy having its own merits, demerits and limitations can successfully be employed to various disciplines of mathematical sciences. Theil [26] developed a measure of cross entropy also known as information improvement, after revising the original probability distribution and applied this measure in economics for measuring the equality of income. Thus, keeping in view the need of generalized divergence measures, Parkash and Mukesh [17,18] have introduced some new parametric and non-parametric measures of divergence. In Chapter-I, we have introduced a new measure of divergence given by

$$D_\alpha(P;Q)=\frac{\sum_{i=1}^{n} p_i \alpha^{\log\frac{p_i}{q_i}}-1}{\alpha-1}, \alpha>\frac{1}{2}, \alpha\neq 1 \tag{5.1.3}$$

where α is a real parameter.

Zheng and Chen [27] proposed a new coherent risk measure called iso-entropic risk measure, which is based on relative entropy and pointed out that this measure is just the negative expectation of the risk portfolio position under the probability measure. Hanna, Gutter and Fan [7] remarked that prior subjective risk tolerance measures have lacked a rigorous connection to economic theory. In their study, the authors presented an improved measurement of subjective risk tolerance based on economic theory. Bertsimas, Lauprete and Samarov [2] introduced a risk measure, called it shortfall and examined its properties and discussed its relation to such commonly used risk measures as standard deviation. Moreover, the authors have shown that the mean-shortfall optimization problem can be solved efficiently as a convex optimization problem. Some other contributors who extended the theory for the applications towards equity and stock market are Siegel [23, 24], Siegel and Thaler [25] etc.

In the present chapter, we make use of parametric and non-parametric divergence measures (5.1.2) and (5.1.3) for the measurement of risk in portfolio analysis. Before developing these measures, we give a brief introduction to the concept of mean-variance efficient frontier due to Markowitz [13].

Markowitz [13] Mean-Variance Efficient Frontier: Let π_j be the probability of the jth outcome for $j=1,2,\ldots,m$ and the r_{ij} be the return on the ith security for $i=1,2,\ldots,n$ when the jth outcome occurs. Then the expected return on the ith security is given by

$$\overline{r_i}=\sum_{j=1}^{m}\pi_j r_{ij}\ ,\quad i=1,2,\ldots,n \tag{5.1.4}$$

Also, variances and covariances of returns are given by

$$\sigma_i^2=\sum_{j=1}^{m}\pi_j\left(r_{ij}-\overline{r_i}\right)^2,\quad i=1,2,\ldots,n \tag{5.1.5}$$

and

$$\rho_{ik}\sigma_i\sigma_k=\sum_{j=1}^{m}\pi_j\left(r_{ij}-\overline{r_i}\right)\left(r_{kj}-\overline{r_k}\right),\quad i,k=1,2,\ldots,n;\ i\neq k\ . \tag{5.1.6}$$

Let a person decide to invest proportions $x_1,x_2,\ldots,x_n$ of his capital in n securities. Assume that $x_i\geq 0$ for all i, and that

$$\sum_{i=1}^{n}x_i=1\,. \tag{5.1.7}$$

Then, the expected return and variance of the return are given by

$$E=\sum_{i=1}^{n}x_i\overline{r_i}\ , \tag{5.1.8}$$

and

$$V=\sum_{i=1}^{n}x_i^2\sigma_i^2+2\sum_{k=1}^{n}\sum_{i<k}x_ix_k\rho_{ik}\sigma_i\sigma_k\ . \tag{5.1.9}$$

Markowitz [13] suggested that $x_1,x_2,\ldots,x_n$ be chosen so as to maximize E and to minimize V or alternatively, to minimize V when E is kept at a fixed value. Now

$$V=\sum_{j=1}^{m}\pi_j\left(x_1r_{1j}+x_2r_{2j}+\cdots+x_nr_{nj}-x_1\overline{r_1}-x_2\overline{r_2}-\cdots-x_n\overline{r_n}\right)^2$$

$$=\sum_{j=1}^{m}\pi_j\left(R_j-\overline{R}\right)^2 \tag{5.1.10}$$

where $R_j=\sum_{i=1}^{n}x_ir_{ij}$ and $\overline{R}=\sum_{i=1}^{n}x_i\overline{r_i}$ (5.1.11)

that is, R_j is the return on investment when the jth outcome arises and $\overline{R}$ is the mean return on investment.

In the next section, we develop an optimizational principle using our divergence measure (5.1.2).

5.2 DEVELOPMENT OF NEW OPTIMIZATION PRINCIPLE

Markowitz's [13] criterion for a choice from $x_1,x_2,\ldots,x_n$ was to minimize the variance, that is, to make $R_1,R_2,\ldots,R_m$ as equal as possible among themselves. Any departure of $R_1,R_2,\ldots,R_m$ from equality was considered a measure of risk. The same purpose can be accomplished if we choose $x_1,x_2,\ldots,x_n$ so as to minimize the directed divergence measure given by (5.1.2) of the distribution

$P=\left(\frac{\pi_1R_1}{\sum_{j=1}^{m}\pi_jR_j},\frac{\pi_2R_2}{\sum_{j=1}^{m}\pi_jR_j},\cdots,\frac{\pi_mR_m}{\sum_{j=1}^{m}\pi_jR_j}\right)$ from $\pi=\left(\pi_1,\pi_2,\ldots,\pi_m\right)$, that is, we choose $x_1,x_2,\ldots,x_n$ so as to minimize the following divergence measure:

$$D_1(P;\pi)=\sum_{j=1}^{m}\left(\frac{P_j^2}{\pi_j}+\frac{\pi_j^2}{P_j}-2\pi_j\right)$$

$$=\sum_{j=1}^{m}\left(\frac{\pi_j^2R_j^2}{\overline{R}^2\pi_j}+\frac{\pi_j^2\overline{R}}{\pi_jR_j}-2\pi_j\right)$$

$$=\frac{1}{\overline{R}^2}\sum_{j=1}^{m}\pi_jR_j^2+\overline{R}\sum_{j=1}^{m}\frac{\pi_j}{R_j}-2 \tag{5.2.1}$$

where $\sum_{j=1}^{m}\pi_jR_j=\sum_{j=1}^{m}\pi_j\sum_{i=1}^{n}x_ir_{ij}$

$$=\sum_{i=1}^{n}x_i\overline{r}_i=\overline{R} \tag{5.2.2}$$

Thus, we can formulate our optimization principle as follows:

Choose $x_1,x_2,\ldots,x_n$ so as to minimize

$$\sum_{j=1}^{m}\pi_j\left(x_1r_{1j}+x_2r_{2j}+\cdots+x_nr_{nj}\right)^2+\sum_{j=1}^{m}\frac{\pi_j}{\left(x_1r_{1j}+x_2r_{2j}+\cdots+x_nr_{nj}\right)} \tag{5.2.3}$$

subject to the following constraints:

$$\sum_{j=1}^{m}\pi_j\left(x_1r_{1j}+x_2r_{2j}+\cdots+x_nr_{nj}\right)=\text{Constant} \tag{5.2.4}$$

$$x_1+x_2+\cdots x_n=1 \tag{5.2.5}$$

and $x_1\geq 0, x_2\geq 0,\ldots, x_n\geq 0$.

Implementation of the principle

We may use the Lagrange's method for implementing the above extremum problem. But, in this process, the x_j's will not be guaranteed to be positive, although there is a guarantee that the R_j's will turn out to be positive. If we find that some of the x_j's are negative, we set these x_j's equal to zero and solve the new optimization problem and proceed as before, repeating the process if necessary. Alternatively, one can use a nonlinear optimization program that considers the inequality constraints on the probabilities x_j's.

The above principle has been explained with the help of a numerical example as follows:

Numerical: Consider the two securities, each with eight possible outcomes and with probabilities and returns as shown in the following Table-5.2.1:

Table-5.2.1

Probability	Return-I	Return-II
0.05	0.10	0.20
0.10	0.05	0.15
0.15	0.15	0.10
0.20	0.20	0.05
0.10	0.10	0.15
0.15	0.05	0.10
0.10	0.20	0.15
0.15	0.20	0.20

We have to find the optimum values of investment proportions x_1 and x_2, when the mean return on the investment is 0.13.

Thus, our problem can be reformulated as follows:

Minimize $\sum_{j=1}^{8} \pi_j \left(x_1 r_{1j} + x_2 r_{2j}\right)^2 + \sum_{j=1}^{8} \frac{\pi_j}{\left(x_1 r_{1j} + x_2 r_{2j}\right)}$,

subject to the set of constraints

$$\sum_{j=1}^{8} \pi_j \left(x_1 r_{1j} + x_2 r_{2j}\right) = 0.13 \tag{5.2.6}$$

$$x_1 + x_2 = 1 \tag{5.2.7}$$

and

$x_1 \geq 0, x_2 \geq 0$

where

$\pi_1 = 0.05$, $\pi_2 = 0.10$, $\pi_3 = 0.15$, $\pi_4 = 0.20$,

$\pi_5 = 0.10$, $\pi_6 = 0.15$, $\pi_7 = 0.10$, $\pi_8 = 0.15$

and

$r_{11} = 0.10$, $r_{12} = 0.05$, $r_{13} = 0.15$, $r_{14} = 0.20$,

$r_{15} = 0.10$, $r_{16} = 0.05$, $r_{17} = 0.20$, $r_{18} = 0.20$,

$r_{21} = 0.20$, $r_{22} = 0.15$, $r_{23} = 0.10$, $r_{24} = 0.05$,

$r_{25} = 0.15$, $r_{26} = 0.10$, $r_{27} = 0.15$, $r_{28} = 0.20$

Consider the Lagrange function

$$L = \sum_{j=1}^{8} \pi_j \left(x_1 r_{1j} + x_2 r_{2j}\right)^2 + \sum_{j=1}^{8} \frac{\pi_j}{\left(x_1 r_{1j} + x_2 r_{2j}\right)} - \lambda \left[\sum_{j=1}^{8} \pi_j \left(x_1 r_{1j} + x_2 r_{2j}\right) - 0.13\right] - \mu\left(x_1 + x_2 - 1\right)$$

Differentiating the above equation with respect to x_1 and x_2, and equating to zero, we get

$$2\sum_{j=1}^{8} \pi_j r_{1j} \left(x_1 r_{1j} + x_2 r_{2j}\right) - \sum_{j=1}^{8} \frac{\pi_j r_{1j}}{\left(x_1 r_{1j} + x_2 r_{2j}\right)^2} - \lambda \sum_{j=1}^{8} \pi_j r_{1j} - \mu = 0 \tag{5.2.8}$$

$$2\sum_{j=1}^{8} \pi_j r_{2j} \left(x_1 r_{1j} + x_2 r_{2j}\right) - \sum_{j=1}^{8} \frac{\pi_j r_{2j}}{\left(x_1 r_{1j} + x_2 r_{2j}\right)^2} - \lambda \sum_{j=1}^{8} \pi_j r_{2j} - \mu = 0 \tag{5.2.9}$$

Using the values of π_j ’s and r_{ij} ’s in (5.2.6), (5.2.7), (5.2.8), (5.2.9) and solving, we get

$x_1 = 0.3333$, $x_2 = 0.6667$.

5.3 MEASURING RISK IN PORTFOLIO ANALYSIS USING PARAMETRIC MEASURES OF DIVERGENCE

In this section, we consider the following cases for measuring risk in portfolio analysis by using parametric measures of divergence.

I. If we use one parametric measure of divergence (5.1.3), we get a measure of risk in accordance with the optimization principle developed in the above section. This measure is developed as follows:

$$R_1 = \frac{\sum_{j=1}^{m} P_j \alpha^{\log \frac{P_j}{\pi_j}} - 1}{\alpha - 1} = \frac{\sum_{j=1}^{m} \frac{\pi_j R_j}{\overline{R}} \alpha^{\log \frac{\pi_j R_j}{\overline{R}\pi_j}} - 1}{\alpha - 1}$$

$$= \frac{\frac{1}{\overline{R}} \sum_{j=1}^{m} \pi_j \left[R_j \alpha^{\log \frac{R_j}{\overline{R}}} \right] - 1}{\alpha - 1}$$

$$= \frac{1}{\alpha - 1} \left[\frac{1}{\overline{R}} E\left(R \alpha^{\log \frac{R}{\overline{R}}} \right) - 1 \right]. \tag{5.3.1}$$

If $\frac{1}{2} < \alpha < 1$, minimizing the measure (5.3.1), we mean the maximization of expected utility of a person whose utility function is given by $u(x) = x\alpha^{\log \frac{x}{\overline{R}}}$. In this case the person is risk-averse. If $\alpha > 1$, minimizing the measure (5.3.1), we mean minimization of the expected utility of a person whose utility function is given by $u(x) = x\alpha^{\log \frac{x}{\overline{R}}}$. In this case the person is risk-prone.

Thus, minimizing this measure implies the minimization of the expected utility of a risk-prone person and maximization of the expected utility of a risk-averse person.

II. Now, we know that Renyi's [20] measure of directed divergence is given by

$$_{\alpha}D(P\,;Q) = \frac{1}{\alpha - 1} \log \sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha}, \; \alpha \neq 1, \; \alpha > 0 \tag{5.3.2}$$

If we use Renyi's [20] measure, we get a measure of risk in accordance with the above mentioned optimization principle. This measure is developed as follows:

$$R_2 = \frac{\log \sum_{j=1}^{m} P_j^{\alpha} \pi_j^{1-\alpha}}{\alpha - 1} = \frac{\log \sum_{j=1}^{m} \left(\frac{\pi_j R_j}{\overline{R}} \right)^{\alpha} \pi_j^{1-\alpha}}{\alpha - 1}$$

$$= \frac{\log \sum_{j=1}^{m} \pi_j \left(\frac{R_j^{\alpha}}{\overline{R}^{\alpha}} \right)}{\alpha - 1}$$

$$= \frac{1}{\alpha - 1} \left[\log \frac{1}{\overline{R}^{\alpha}} E\left(R^{\alpha}\right) \right]. \tag{5.3.3}$$

If $\alpha < 1$, minimizing the measure (5.3.3), we mean the maximization of expected utility of a person whose utility function is given by $u(x) = x^{\alpha}$. In this case the person is risk-averse. If $\alpha > 1$, minimizing the measure (5.3.3), we mean minimization of the expected utility of a person whose utility function is given by $u(x) = x^{\alpha}$. In this case the person is risk-prone.

III. Now, we use Havrada-Charvat's [8] extended measure of directed divergence given by

$$D^{\alpha}(P;Q) = \frac{\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} - 1}{\alpha(\alpha - 1)}, \alpha > 0, \alpha \neq 0,1 \tag{5.3.4}$$

Thus, again we get a measure of risk in accordance with the above mentioned optimization principle. This measure is developed as follows:

$$R_3 = \frac{1}{\alpha(\alpha - 1)} \left[\sum_{j=1}^{m} P_j^{\alpha} \pi_j^{1-\alpha} - 1 \right] = \frac{1}{\alpha(\alpha - 1)} \left[\sum_{j=1}^{m} \left(\frac{\pi_j R_j}{\overline{R}} \right)^{\alpha} \pi_j^{1-\alpha} - 1 \right]$$

$$= \frac{1}{\alpha(\alpha - 1)} \left[\sum_{j=1}^{m} \pi_j \left(\frac{R_j^{\alpha}}{\overline{R}^{\alpha}} \right) - 1 \right]$$

$$= \frac{1}{\alpha(\alpha - 1)} \left[\frac{1}{\overline{R}^{\alpha}} E\left(R^{\alpha}\right) - 1 \right] \tag{5.3.5}$$

If $0 < \alpha < 1$, minimizing the measure (5.3.5), implies the maximization of expected utility function $u(x) = x^{\alpha}$. In this case the person is risk-averse. If $\alpha > 1$, minimizing the measure (5.3.5), we mean minimization of the expected utility $u(x) = x^{\alpha}$ and in this case the person is risk-prone.

Thus, we conclude that minimizing this measure implies the minimization of the expected utility of the risk-prone person and maximization of the expected utility of a risk-averse person.
Proceeding on similar lines, many new optimization principles can be developed by using different measures of directed divergence.

5.4 APPLICATIONS OF INFORMATION MEASURES TO THE FIELD OF QUEUEING THEORY

In dealing with the queueing theory, we suppose that the queueing system is in a steady-state condition and this steady-state is reached if the state of the system becomes essentially independent of the initial state and the elapsed time. In fact, the usual analysis of the main queueing system is based on the birth-death process, according to which, for any state n of the queueing system, the mean rate at which the entering incidents occur must equal the mean rate at which leaving incidents occur. Queueing theory must assume some kind of stability for obtaining a probabilistic model of the system's evaluation and the basic formulae obtained are reliable to the extent to which the conditions of the alleged birth-death process are satisfied. The principle of maximum entropy is another method, to construct the most uncertain probability distribution subject to some constraints expressed by mean values. By using this principle, we cannot only reobtain some known formulas from queueing theory, but also give an analytical expression to the solution of some more queueing systems.

If the real probability distribution of the possible states of the queueing system is known, the corresponding entropy is a number, which may be effectively computed for measuring the amount of uncertainty about the real state of the system. But generally, we do not know this real probability distribution. The available information is summarized in mean values, mean arrival rates, mean service rates of the mean number of customers in the system. Suppose that we know the mean number of customers in the system, and then we can have several probability distributions on the possible states of the system subject to the given number of the customers. From this set of feasible probability distributions, let us take the unique probability distribution for which the corresponding entropy is maximum and such a probability distribution gives the largest probabilistic model, treating the possible states of the system as uniform as possible.

It was Guiasu [6], who obtained a probabilistic model when the queueing system is in the maximum entropy condition. For applying the entropic approach, the only information required is represented by mean values. For some one-server queueing systems, when the expected number of customers is given, the maximum entropy condition gives the same probability distribution of the possible states of the system as the birth-death process applied to $M/M/1$ system in a steady state condition. For other queueing systems, as for $M/G/1$ instance, the entropic approach gives a simple probability distribution of possible states, while no close expression for such a probability distribution is known in the general framework of a birth-death process.

In simple birth-death process of queueing theory, let $p_n(t)$ denotes the probability of there being n persons in the population at time t and let n_0 denote the number of persons at time $t = 0$, then Medhi [15] has obtained an expression for $p_n(t)$. In fact, if we define the probability generating function by

$$\phi(s,t)=\sum_{n=o}^{\infty} p_n(t)s^n , \tag{5.4.1}$$

we get the following result

$$\phi(s,t)=\left[\frac{(\lambda-\mu)s+\mu(x-1)}{(\lambda-\lambda x)s+(\lambda x-1)}\right]^{n_0} , \lambda \neq \mu \tag{5.4.2}$$

$$= \left[\frac{\lambda t-(\lambda t-1)s}{1-\lambda t-\lambda ts}\right]^{n_0} , \lambda = \mu \tag{5.4.3}$$

where $x = \exp(\lambda-\mu)t$ (5.4.4)

By expanding $\phi(s,t)$ in power series of s, we can find $p_n(t)$. In a queuing system, let λ and μ denote arrival and service rates in the steady state case, then we have

$$p_n = (1-\rho)\rho^n , \; n = 0,1,2.3,....; \; \rho = \frac{\lambda}{\mu} \tag{5.4.5}$$

At any time t, the number of persons in the system can be 0, 1, 2,…, so that there is uncertainty about the number of persons in the system. We want to develop a measure of this uncertainty, which shows how this uncertainty varies with λ, μ and t. Kapur [10] has studied such types of variations by considering well known measures of uncertainty. Prabhakar and Gallager [19] have undertaken the study of queues which deal with two single server discrete-time queues. These are

(i) first come first served queue with independent and identically distributed service times and

(ii) a preemptive resume last come first served queue with non-negative valued independent and identically distributed service times. For these systems, the authors have shown that when units arrive according to an arbitrary ergodic stationary arrival process, the corresponding departure process has an entropy rate no less than the entropy rate of the arrival process. Using this approach from the entropy standpoint, the authors have established connections with the time capacity of queues.

In the sequel, we have obtained the measures of uncertainty for a probability of the population size at any time in a birth-death process as functions of birth rate and death rate. It has been shown that in case of steady state queuing process, as the arrival rate increases relatively to service rate, the uncertainty increases whereas in the case of non-steady birth-death process, the uncertainty measure first increases and attains its maximum value and then with the passage of time, it decreases and attains its minimum value. These results have been presented in the following cases:

Case-I: Applications of information measures for the study of variations of entropy in the steady state queueing process.

In this case, we have studied the variations of our own measure of entropy in the steady state queueing processes. For this purpose, we have considered the following measure of entropy introduced in Chapter-I. This measure is given by

$$H_\beta\left(\lambda,\mu\right)=-\frac{\log_D \sum_{i=0}^{\infty} p_i \beta^{\log_D p_i}}{\log_D \beta}, \beta>1, \beta\neq 1 \tag{5.4.6}$$

$$=-\frac{\log_D\left(p_0\beta^{\log_D p_0}+p_1\beta^{\log_D p_1}+p_2\beta^{\log_D p_2}+\ldots+p_n\beta^{\log_D p_n}+\ldots\right)}{\log_D \beta}$$

$$=-\frac{\log_D\begin{bmatrix}(1-\rho)\beta^{\log_D(1-\rho)}+\rho(1-\rho)\beta^{\log_D(1-\rho)\rho}+\rho^2(1-\rho)\beta^{\log_D \rho^2(1-\rho)}+\\ \ldots+\rho^n(1-\rho)\beta^{\log_D \rho^n(1-\rho)}+\ldots\end{bmatrix}}{\log_D \beta}$$

$$=-\frac{\log_D\left[(1-\rho)\left(\beta^{\log_D(1-\rho)}+\rho\beta^{\log_D(1-\rho)\rho}+\rho^2\beta^{\log_D\rho^2(1-\rho)}+\ldots+\rho^n\beta^{\log_D\rho^n(1-\rho)}+\ldots\right)\right]}{\log_D \beta}$$

$$=-\frac{\log_D\left[(1-\rho)\beta^{\log_D(1-\rho)}\left(1+\rho\beta^{\log_D\rho}+\rho^2\beta^{\log_D\rho^2}+\ldots+\rho^n\beta^{\log_D\rho^n}+\ldots\right)\right]}{\log_D \beta}$$

$$=-\frac{\log_D\left[(1-\rho)\beta^{\log_D(1-\rho)}\left(1+\rho\beta^{\log_D\rho}+\rho^2\beta^{2\log_D\rho}+\ldots+\rho^n\beta^{n\log_D\rho}+\ldots\right)\right]}{\log_D \beta}$$

$$= -\frac{\log_D\left(\left(1-\rho\right)\beta^{\log_D(1-\rho)}\left(\frac{1}{1-\rho\beta^{\log_D \rho}}\right)\right)}{\log_D \beta}$$

$$= -\frac{\log_D\left(1-\rho\right)+\log_D\left(\beta^{\log_D(1-\rho)}\right)+\log_D\left(\frac{1}{1-\rho\beta^{\log_D \rho}}\right)}{\log_D \beta}$$

$$= -\frac{\log_D\left(1-\rho\right)+\log_D\left(1-\rho\right)\log_D \beta-\log_D\left(1-\rho\beta^{\log_D \rho}\right)}{\log_D \beta} \tag{5.4.7}$$

Letting $\beta \to 1$ in (5.4.7), we get $\frac{0}{0}$ form. So applying L-Hospital's rule, we get

$$\lim_{\beta\to 1} H_\beta\left(\lambda,\mu\right) = -\lim_{\beta\to 1}\frac{\frac{\log_D\left(1-\rho\right)}{\beta}+\frac{\rho\beta^{\log_D \rho-1}\log_D \rho}{1-\rho\beta^{\log_D \rho}}}{\frac{1}{\beta}}$$

Thus, we have

$$H_1\left(\lambda,\mu\right) = -\left[\frac{\left(1-\rho\right)\log\left(1-\rho\right)+\rho\log\rho}{1-\rho}\right] \tag{5.4.8}$$

a result studied by Kapur [10].

Now, differentiating (5.4.7) w.r.t ρ, we get

$$\frac{\partial H_\beta\left(\lambda,\mu\right)}{\partial\rho} = -\frac{1}{\log_D \beta}\left(-\frac{1}{\left(1-\rho\right)}-\frac{\log_D \beta}{1-\rho}+\frac{\beta^{\log_D \rho}+\beta^{\log_D \rho}\log_D \beta}{\left(1-\rho\beta^{\log_D \rho}\right)}\right)$$

Letting $\beta \to 1$, we get

$$\lim_{\beta\to 1}\frac{\partial H_\beta\left(\lambda,\mu\right)}{\partial\rho} = \lim_{\beta\to 1}\frac{1}{\left(1-\rho\right)}$$

$$-\lim_{\beta\to 1}\beta\left(\frac{\left(1-\rho\beta^{\log_D\rho}\right)\left(\beta^{\log_D\rho-1}\log_D\rho+\beta^{\log_D\rho-1}\log_D\rho\log_D\beta+\beta^{\log_D\rho-1}\right)-\left(\beta^{\log_D\rho}+\beta^{\log_D\rho}\log_D\beta\right)\left(-\rho\beta^{\log_D\rho-1}\log_D\rho\right)}{\left(1-\rho\beta^{\log_D\rho}\right)^2}\right)$$

$$=-\frac{\log\rho}{(1-\rho)^2}>0$$

Thus, we see that in this case, the uncertainty increases monotonically from 0 to ∞ as ρ increases from 0 to unity.

Case-II: Applications of information measures for the variations of entropy in the non-steady state queueing process.

In this case, we have studied the variations of our own measure of entropy in the non-steady state queueing processes. For this purpose, we have again considered the measure of entropy (5.4.6) and for this study, we first of all develop the following results:

Equation (5.4.3) gives

$$\sum_{n=0}^{\infty}p_n(t)s^n=\frac{\lambda t}{1+\lambda t}\left(1-\frac{\lambda t-1}{\lambda t}s\right)\left(1-\frac{\lambda t}{1+\lambda t}s\right)^{-1}$$

$$=\frac{\lambda t}{1+\lambda t}\left(1-\frac{\lambda t-1}{\lambda t}s\right)\sum_{n=0}^{\infty}\left(\frac{\lambda t}{1+\lambda t}\right)^n s^n$$

so that

$$p_n(t)=\frac{\lambda t}{1+\lambda t}\left(\frac{\lambda t}{1+\lambda t}\right)^n-\left(\frac{\lambda t-1}{\lambda t}\right)\left(\frac{\lambda t}{1+\lambda t}\right)^{n-1}$$

$$=\frac{(\lambda t)^{n-1}}{(1+\lambda t)^{n+1}},\ n\geq 1$$

Also

$$p_0(t)=\frac{\lambda t}{1+\lambda t}$$

Now, we study the different variations of the entropy (5.4.6) as follows:

$$H_\beta(\lambda,\mu) = -\frac{\log_D\left(p_0\beta^{\log_D p_0} + p_1\beta^{\log_D p_1} + p_2\beta^{\log_D p_2} + \ldots + p_n\beta^{\log_D p_n} + \ldots\right)}{\log_D \beta} \tag{5.4.9}$$

Taking $p_0(t) = \frac{\lambda t}{1+\lambda t}$ and $p_n(t) = \frac{(\lambda t)^{n-1}}{(1+\lambda t)^{n+1}}$, equation (5.4.9) becomes

$$H_\beta(\lambda,\mu) = -\frac{\log_D\left(\begin{array}{l}\left(\frac{\lambda t}{1+\lambda t}\right)\beta^{\log_D\left(\frac{\lambda t}{1+\lambda t}\right)} + \frac{1}{(1+\lambda t)^2}\beta^{\log_D\frac{1}{(1+\lambda t)^2}} + \frac{\lambda t}{(1+\lambda t)^3}\beta^{\log_D\frac{\lambda t}{(1+\lambda t)^3}} + \\ \ldots + \frac{(\lambda t)^{n-1}}{(1+\lambda t)^{n+1}}\beta^{\log_D\frac{(\lambda t)^{n-1}}{(1+\lambda t)^{n+1}}} + \ldots\end{array}\right)}{\log_D \beta}$$

$$= -\frac{\log_D\left(\left(\frac{\lambda t}{1+\lambda t}\right)\beta^{\log_D\left(\frac{\lambda t}{1+\lambda t}\right)} + \frac{1}{(1+\lambda t)^2}\beta^{\log_D\frac{1}{(1+\lambda t)^2}}\left(\begin{array}{l}1+\frac{\lambda t}{1+\lambda t}\beta^{\log_D\frac{\lambda t}{1+\lambda t}} + \left(\frac{\lambda t}{1+\lambda t}\right)^2\beta^{2\log_D\frac{\lambda t}{1+\lambda t}} \\ +\left(\frac{\lambda t}{1+\lambda t}\right)^3\beta^{3\log_D\frac{\lambda t}{1+\lambda t}}\ldots+\left(\frac{\lambda t}{1+\lambda t}\right)^n\beta^{n\log_D\frac{\lambda t}{1+\lambda t}}+\ldots\end{array}\right)\right)}{\log_D \beta}$$

$$= -\frac{\log_D\left(\left(\frac{\lambda t}{1+\lambda t}\right)\beta^{\log_D\left(\frac{\lambda t}{1+\lambda t}\right)} + \frac{1}{(1+\lambda t)^2}\beta^{\log_D\frac{1}{(1+\lambda t)^2}}\left(\frac{1}{1-\frac{\lambda t}{1+\lambda t}\beta^{\log_D\frac{\lambda t}{1+\lambda t}}}\right)\right)}{\log_D \beta}$$

$$= -\frac{\log_D\left(\frac{\beta^{\log_D\left(\frac{1}{1+\lambda t}\right)}}{(1+\lambda t)}\left(\lambda t\beta^{\log_D \lambda t} + \frac{\beta^{\log_D\left(\frac{1}{1+\lambda t}\right)}}{1+\lambda t-\lambda t\beta^{\log_D\frac{\lambda t}{1+\lambda t}}}\right)\right)}{\log_D \beta}$$

$$= \frac{-\log_D\left(\frac{\beta^{\log_D\left(\frac{1}{1+\lambda t}\right)}}{(1+\lambda t)}\right) - \log_D\left(\lambda t\beta^{\log_D \lambda t} + \frac{\beta^{\log_D\left(\frac{1}{1+\lambda t}\right)}}{1+\lambda t-\lambda t\beta^{\log_D\frac{\lambda t}{1+\lambda t}}}\right)}{\log_D \beta}$$

$$=\frac{\log_D(1+\lambda t)+\log_D\beta\log_D(1+\lambda t)-\log_D\left(\lambda t\beta^{\log_D\lambda t}+\dfrac{\beta^{\log_D\left(\frac{1}{1+\lambda t}\right)}}{1+\lambda t-\lambda t\beta^{\log_D\frac{\lambda t}{1+\lambda t}}}\right)}{\log_D\beta} \qquad (5.4.10)$$

Letting $\beta\to 1$ in (5.4.10), we get $\dfrac{0}{0}$ form and applying L Hospital's rule, we get

$$\lim_{\beta\to 1}H_\beta(P)=\lim_{\beta\to 1}\beta\left\{\frac{\log_D(1+\lambda t)}{\beta}-\frac{1}{\left(\lambda t\beta^{\log_D\lambda t}+\dfrac{\beta^{\log_D\left(\frac{1}{1+\lambda t}\right)}}{1+\lambda t-\lambda t\beta^{\log_D\frac{\lambda t}{1+\lambda t}}}\right)}\left\{\lambda t\log_D\lambda t\beta^{\log_D\lambda t-1}\right\}\right.$$

$$\left.-\lim_{\beta\to 1}\beta\frac{\left(1+\lambda t-\lambda t\beta^{\log_D\frac{\lambda t}{1+\lambda t}}\right)\left(-\log_D(1+\lambda t)\beta^{-\log_D(1+\lambda t)-1}\right)-\beta^{-\log_D(1+\lambda t)}\left(-\lambda t\log_D\dfrac{\lambda t}{1+\lambda t}\beta^{\log_D\frac{\lambda t}{1+\lambda t}-1}\right)}{\left(\lambda t\beta^{\log_D\lambda t}+\dfrac{\beta^{\log_D\left(\frac{1}{1+\lambda t}\right)}}{1+\lambda t-\lambda t\beta^{\log_D\frac{\lambda t}{1+\lambda t}}}\right)\left(1+\lambda t-\lambda t\beta^{\log_D\frac{\lambda t}{1+\lambda t}}\right)^2}\right\}$$

$$=\frac{2\left((1+\lambda t)\log(1+\lambda t)-\lambda t\log\lambda t\right)}{1+\lambda t} \qquad (5.4.11)$$

Differentiating (5.4.10) w.r.t λt, we get

$$\frac{\partial H_\beta(\lambda,\mu)}{\partial(\lambda t)}=\frac{1}{\log D\log_D\beta}\left(\frac{(1+\log_D\beta)}{(1+\lambda t)}\right)$$

$$-\frac{1}{\log D\log_D\beta}\left(\frac{1}{\left(\lambda t\beta^{\log_D\lambda t}+\dfrac{\beta^{\log_D\left(\frac{1}{1+\lambda t}\right)}}{1+\lambda t-\lambda t\beta^{\log_D\frac{\lambda t}{1+\lambda t}}}\right)}\times\left(\beta^{\log_D\lambda t}(1+\log_D\beta)+\frac{(1+\log_D\beta)\left(-\beta^{\log_D\left(\frac{1}{1+\lambda t}\right)}+\beta^{\log_D\frac{\lambda t}{(1+\lambda t)^2}}\right)}{\left(1+\lambda t-\lambda t\beta^{\log_D\frac{\lambda t}{1+\lambda t}}\right)^2}\right)\right)$$

Letting $\beta\to 1$, we get

$$\frac{\partial H_1(\lambda,\mu)}{\partial(\lambda t)} = -\frac{2\log\lambda t}{(1+\lambda t)^2}$$

This shows that uncertainty increases so long as $\lambda t < 1$ and decreases after this and maximum entropy occurs when $\lambda t = 1$, and

$$\text{Max } H_\beta(\lambda,\mu) = 2\log 2$$

When $t = 0$, the uncertainty is zero and when $t \to \infty$

$$\lim_{t\to\infty} \frac{2\left[(1+\lambda t)\log(1+\lambda t) - \lambda t \log\lambda t\right]}{1+\lambda t} = \lim_{x\to\infty} \frac{2\left[(1+x)\log(1+x) - x\log x\right]}{1+x}$$

$$= \lim_{x\to\infty} \frac{2\log(1+x)/x}{1} = 0$$

Thus, in this case, we observe that the uncertainty starts with zero value at time $t = 0$ and ends with zero value as $t \to \infty$ and in between it attains its maximum value 2 *log*2 when $t = 1/\lambda$.

Concluding Remarks: The findings of our study reveal that by using parametric measure of cross-entropy, we can talk of maximizing the expected utility of risk-averse persons and of minimizing the expected utility of risk-prone persons. Such a study can be made available by the use of a variety of other generalized parametric and non-parametric measures of cross entropy. Moreover, we have provided the applications of our own entropy measure developed in Chapter-I to the field of queueing theory for the study of different variations in different states of the queueing processes. By making use of other well known existing entropy measures, such variations can be studied and some interesting conclusions can be made.

REFERENCES

[1] Bera, A. K., & Park, S. Y. (2008). Optimal portfolio diversification using maximum entropy principle. *Econometric Reviews, 27*, 484-512.

[2] Bertsimas, D., Lauprete, G. J., & Samarov, A. (2004). Shortfall as a risk measure: properties, optimization and applications. *Journal of Economics Dynamics & Control, 28*, 1353-1381.

[3] Bhattacharya, A. (1943). On a measure of divergence between two statistically populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society, 35,* 99-109.

[4] Bugár, G., & Uzsoki, M. (2011). Portfolio optimization strategies: Performance evaluation with respect to different objectives. *Journal of Transnational Management, 16*, 135-148.

[5] Carlson, B. A., & Clements, M. A. (1991). A computationally compact divergence measure for speech processing. *IEEE Transaction on Pattern Analysis and Machine Intelligence, 13*(12), 1255-1260.

[6] Guiasu, S. (1986). Maximum entropy condition in queueing theory. *Journal of the Operational Research Society, 37*, 293-301.

[7] Hanna, S. D., Gutter M. S., & Fan, J. X. (2001). A measure of risk tolerance based on economic theory. *Financial Counseling and Planning, 12*(2), 53-60.

[8] Havrada, J. H., & Charvat, F. (1967). Quantification methods of classification process: Concept of structural α-entropy. *Kybernetika, 3*, 30-35.

[9] Jianshe, O. (2005). Theory of portfolio and risk based on incremental entropy. *Journal of Risk Finance, 6*(1), 31-39.

[10] Kapur, J. N. (1995). *Measures of information and their applications*. New York: Wiley Eastern.

[11] Kullback, S. (1959). *Information theory and statistics*. New York: Wiley.

[12] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*, 79-86.

[13] Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77-91.

[14] Markowitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments*. New York: John Wiley and Sons, Inc.

[15] Medhi, J.M.. (1982). *Stochastic Processes*. New Delhi: Wiley Eastern.

[16] Nocetti, D. (2006). Portfolio selection with endogenous estimation risk. *Economics Bulletin, 7*(6), 1-9.

[17] Parkash, O., & Mukesh (2011). Two new symmetric divergence measures and information inequalities. *International Journal of Mathematics and Applications, 4*(2), 165-179.

[18] Parkash, O., & Mukesh (2012). New generalized parametric measure of entropy and cross entropy. *American Journal of Mathematics and Sciences, 1*(1), 91-96.

[19] Prabhakar, B., & Gallager, R. (2003). Entropy and the timing capacity of discrete queues. *IEEE Transactions on Information Theory, 49*, 357-370.

[20] Renyi, A. (1961). On measures of entropy and information. *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability* (pp.547-561).

[21] Shore, J. E. and Gray, R. M. (1982). Minimum cross-entropy pattern classification and cluster analysis. *IEEE Transactions on Pattern Analysis and Machine intelligence, 4*(1), 11-17.

[22] Shore, J. E., & Johnson, R. W. (1981). Properties of cross entropy minimization. *IEEE Transactions on Information Theory, 27*(4), 472-482.

[23] Siegel, J. J. (1992a). The equity premium: Stock and bond returns since 1802. *Financial Analyst Journal, 48*, 28-38.

[24] Siegel, J. J. (1992b). The real rate of interest from 1800 to 1990: A study of the U.S. and U.K. *Journal of Monetary Economics*, *29*, 227-252.

[25] Siegel, J. J., & Thaler, R. H. (1997). The equity premium puzzle. *Journal of Economic Perspectives, 11*(1), 191-200.

[26] Theil, H. (1967). *Information theory and economics*. North Holland: Amastaerdam.

[27] Zheng, C., & Chen, Y. (2012). Coherent risk measure based on relative entropy. *Applied Mathematics & Information Sciences, 6*(2), 233-238.

CHAPTER-VI

NEW MEAN CODEWORD LENGTHS AND THEIR CORRESPONDENCE WITH INFORMATION MEASURES

ABSTRACT

The objective of the present chapter is to provide a deep study of the problem of correspondence between well known measures of entropy and the mean codeword lengths. With the help of some standard measures of entropy, we have illustrated such a correspondence. In the literature, we usually come across many inequalities which are frequently used in information theory. Keeping this idea in mind, we have developed such inequalities by the use of coding theory and divergence measures. Moreover, we have made use of the concept of weighted entropy and generated some new possible generalized measures of weighted entropy through coding theorems. Such theorems also provide a correspondence between standard measures of entropy and their upper bounds which are neither mean codeword lengths themselves nor some monotonic increasing functions of mean codeword lengths. These findings have been discussed in the last section of the chapter.

***Keywords:* Codeword, Code alphabet, Best 1-1 code, Uniquely decipherable code, Mean codeword length, Entropy, Directed divergence.**

6.1 INTRODUCTION

In early days, computers were less reliable as compared to the computers of today, the consequences of which could result in the failure of entire calculation work. The engineers of the day devised ways to detect faulty relays so that these could be replaced. It was R.W. Hamming working for Bell Labs who thought that if the machine was capable of knowing it was in error, wasn't it also possible for the machine to correct that error. Setting to work on this problem, Hamming devised a way of encoding information so that if an error was detected, it could also be corrected. Based in part on this work, Claude Shannon developed the theoretical framework and set the origin of coding theory. The basic problem of coding theory is that of communication over an unreliable channel that results in errors in the transmitted message. It is worthwhile noting that all communication channels have errors, and thus codes are widely used. In fact, these codes are not just used for network communication, USB channels, satellite communication and so on, but also in disks and other physical media which are also prone to errors. In addition to their practical application, coding theory has many applications in computer science and as such it is a topic that is of interest to both practitioners and theoreticians.

The last few years have witnessed an impressive convergence of interests between disciplines which are a priori well separated: coding and information theory. The underlying reason for this convergence is the importance of probabilistic models in each of these domains. This has long been obvious in information theory and now it has become apparent in coding theory. More accurately, the two basic concepts viz, entropy and coding are strongly related to each other and a very important and basic relation between the two concepts was first obtained by Shannon [22], where he first introduced the concept of entropy, to measure uncertainty of the probability distribution $P=\left(p_1, p_2, p_3, \ldots, p_n\right)$. In the field of coding theory, we usually come across the problem of efficient coding of messages to be sent over a noiseless channel, that is, our only concern is to maximize the number of messages that can be sent over the channel in a given time.

Let us assume that the messages to be transmitted are generated by a random variable X and each value $x_i, (i=1,2,\ldots,n)$ of X must be represented by a finite sequence of symbols chosen from the set $\{a_1, a_2, \ldots\ldots, a_D\}$. This set is called code alphabet or set of code characters and sequence assigned to each $x_i, (i=1,2,\ldots,n)$ is called code word. Let l_i $(i=1,2,\ldots,n)$ be the length of code word associated with x_i satisfying Kraft's [13] inequality given by the following expression:

$$D^{-l_1}+D^{-l_2}+\ldots+D^{-l_n} \leq 1 \tag{6.1.1}$$

where D is the size of alphabet.

In calculating the long run efficiency of communications, we choose codes to minimize average code word length, given by

$$L=\sum_{i=1}^{n} p_i l_i \tag{6.1.2}$$

where p_i is the probability for the occurrence of x_i. For uniquely decipherable codes, Shannon's [22] noiseless coding theorem which states that

$$\frac{H(P)}{\log D} \leq L < \frac{H(P)}{\log D}+1 \tag{6.1.3}$$

determines the lower and upper bounds on L in terms of Shannon's [22] entropy $H(P)$.

It is to be observed that the arithmetic mean is not the only mean codeword length for lengths $l_1, l_2, \ldots, l_n$. In fact, we can consider infinity of means of the form $f^{-1}\left(\sum_{i=1}^{n} p_i f\left(l_i\right)\right)$ where f is a one-

one function with inverse f^{-1}. As a special case, we can have power means by taking $f(x)=x^r$ to get infinity of power means $\left(\sum_{i=1}^{n} p_i l_i^r\right)^{\frac{1}{r}}$ for all real values of r. This mean includes the popular arithmetic, geometric and harmonic means.

We can also have infinity of exponentiated means by taking $f(x)=D^{kx}$ where D the size of the alphabet used. We can also consider a large number of functions depending upon the probability distribution, which satisfy the essential properties of a measure of entropy. Though, usually we apply Kraft's [13] inequality as unique decipherability constraint for providing relation between entropy and codeword length, we may consider other relevant and useful constraints as well. Thus, while Shannon [22] considered only one mean, one constraint, that is, Kraft's inequality and his own measure of entropy, we can considerably enlarge the scope of the subject under consideration with the help of a variety of generalized parametric and non-parametric measures of entropy. We, thus have three entities viz, means, constraints and measures of entropy and we get the following problems to consider:

(i) Given a specific mean and a specific constraint, one has to find the minimum value of the mean subject to the given constraint.

(ii) Given an entropy measure and a constraint, one has to find the suitable mean codeword length for which the given entropy measure will give the minimum value for the given constraint.

(iii) Given the mean codeword length and the well known measure of entropy, one has to find a suitable constraint for which the measure of entropy will be the minimum value for the given mean codeword length.

A suitable code is defined as a code in which lengths $l_1, l_2, l_3, ..., l_n$ of the codewords satisfy a suitable relation in such a way that the minimum value of a specified mean codeword length for all codes satisfying the given relation lies between two specified values. Kapur [10] has well explained the following criteria while providing correspondence with measures of information, suitable codeword lengths and the construction of suitable constraints:

1. If we want the lower bound of arithmetic mean of codeword length to lie between $H(P)$ and $H(P)+1$ where $H(P)$ is Shannon's [22] entropy, then we must choose codeword lengths $l_1, l_2, l_3, ..., l_n$ to satisfy Kraft's [13] inequality.

2. Similarly, if we want the lower bound of exponentiated mean codeword length of order α to lie between $R_\alpha(P)$ and $R_\alpha(P)+1$, where $R_\alpha(P)$ is Renyi's [21] entropy of order α, then again we should choose $l_1, l_2, l_3, ..., l_n$ to satisfy Kraft's inequality. The exponentiated mean here may be Campbell's [4] mean L_α.

3. If we want the lower bound for $\sum_{i=1}^{n} p_i l_i$ to lie between $K(P:Q)$ and $K(P:Q)+1$, where $K(P:Q)$ is Kerridge's [11] measure of inaccuracy, then we have to choose $l_1, l_2, l_3, ..., l_n$ to satisfy the modified Kraft's inequality, given by

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

4. If we want the lower bound to lie between $D(P:Q)$ and $D(P:Q)+1$, where $D(P:Q)$ is Kullback- Leibler [14] measure, then the constraint should be

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

5. If we want the lower bound of exponentiated mean codeword length to lie between $\dfrac{\frac{1}{1-\alpha}\log_D \sum_{i=1}^{n} p_i^r q_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^r}$ and $\left(\dfrac{\frac{1}{1-\alpha}\log_D \sum_{i=1}^{n} p_i^r q_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^r}\right)+1$, then the appropriate constraints should be the following one:

$$\frac{\sum_{i=1}^{n} p_i^r q_i^{-1} D^{-l_i}}{\sum_{i=1}^{n} p_i^r} \leq 1$$

While dealing with coding theory, we usually come across the problem of unique decipherability. To tackle this problem, we would like to have the following definition.

"A code is uniquely decipherable if every finite sequence of code characters corresponds to at most one message".

One way to insure unique decipherability is to require that no code word should be a "prefix" of another code word. A code having the property that no code word is a prefix of another code word is said to be instantaneous. Before turning to the problem of characterizing uniquely decipherable codes, we note that every instantaneous code is uniquely decipherable, but not conversely.

The implicit assumption that works behind the restriction of encoding of X to uniquely decodable codes is that a sequence of random variables $x_1, x_2, ..., x_n$ is being encoded and that the individual codeword will be concatenated together. So, the code must be uniquely decodable in order to recover $x_1, x_2, ..., x_n$. Since, there are communication situations where a single random variable is being transmitted instead of sequence of random variables, Leung-Yan-Cheong and Cover [16] considered 1:1 codes, that is, codes which assign a distinct binary codeword to each outcome of the random variable, without regard to the constraint that concatenation of these descriptions must be uniquely decodable. Moreover, if $L_{1:1}$ is the mean length for the best 1-1 binary code, then Leung-Yan-Cheong and Cover [16] proved the following inequalities in terms of Shannon's [22] entropy $H(P)$ and the 1-1 codes:

$$L_{1:1} \geq H(P) - a(1 + \log(1 + H(P)) - \log(2^a - 1)/(2^a + 1), a > 1$$

$$L_{1:1} \geq H(P) - 2\log(1 + H(P)) \tag{6.1.4}$$

$$L_{1:1} \geq H(P) - \log(1 + H(P)) - \log(\log H(P) + 1) + ...$$

It is to be noted that uniquely decipherable codes (U.D.) are constrained codes because their lengths have to satisfy Kraft's inequality whereas 1-1 codes are less constrained. The mean length of the best U.D. code is greater than Shannon's [22] entropy $H(P)$. The mean length of the best 1-1 code $L_{1:1}$ is also a measure of entropy and the two measures of entropy need not be related by equality or inequality relations. It was Campbell [4], who for the first time introduced the idea of exponentiated mean code word length for uniquely decipherable codes and proved a noiseless coding theorem. The author considered a special exponentiated mean of order α given by

$$L_\alpha = \frac{\alpha}{1-\alpha} \log_D \left[\sum_{i=1}^{n} p_i \, D^{(1-\alpha) n_i/\alpha} \right] \tag{6.1.5}$$

and showed that its lower bound lies between $R_\alpha(P)$ and $R_\alpha(P) + 1$ where

$$R_\alpha(P) = (1-\alpha)^{-1} \log_D \left[\sum_{i=1}^{n} p_i^{\alpha} \right]; \alpha > 0, \alpha \neq 1 \tag{6.1.6}$$

is Renyi's [21] measure of entropy of order α.

As $\alpha \to 1$, it can easily be shown that $L_\alpha \to L$ and $R_\alpha(P)$ approaches $H(P)$.

After the introduction of Campbell's [4] exponentiated mean of order α, many researchers proved different coding theorems and consequently, developed their exponentiated means. Some of the pioneer towards such contributions are Kumar and Kumar [15], Ramamoorthy [20], Parkash and Kakkar [18], Parkash and Priyanka [19] etc.

Of course, Kapur [10] defined the following mean:

$$L^{\alpha} = \frac{1}{1-\alpha} \log_D \left[\sum_{i=1}^{n} p_i^{\alpha} \, D^{-(1-\alpha) l_i} / \sum_{i=1}^{n} \, p_i^{\alpha} \right] \tag{6.1.7}$$

and gave its proper definition as follows:

L^{α} is said to be exponentiated mean if it satisfies the following properties:

(i) If $l_1 = l_2 = l_3 = \ldots = l_n = l$, then $L^{\alpha} = l$

(ii) L^{α} must lie between minimum and maximum values of $l_1, l_2, l_3, \ldots, l_n$

(iii) $\underset{\alpha \to 1}{lt} \, L^{\alpha} = L$ where $L = \sum_{i=1}^{n} p_i l_i$

In the literature of coding theory, for the given inequality, say, Kraft's [13] inequality, we have a pair of problems. For a given mean codeword length, we can find its lower bounds for all uniquely decipherable codes. In the inverse problem, we can find the mean value for the given pair of lower bounds. The direct problem has a unique answer, though it may not always be easy to find an analytical expression for it. However, the inverse problem has no unique answer in the sense that the same lower bounds may arise for a number of means. The challenge is to find as many of the means as possible, which have the given pair of values as lower bounds.

The object of the present chapter is to go deeper into the problem of correspondence between well known measures of entropy and mean codeword lengths. We state the following fundamental results in a broader framework:

(a) To every mean codeword length, there corresponds a measure of entropy or a monotonic increasing function of a measure of entropy.

(b) To every measure of entropy, there corresponds a mean codeword length or a monotonic increasing function of the mean codeword length.

For many purposes, especially for maximization of entropy, every monotonic increasing function of a measure of entropy is as good as a measure of entropy and for such purposes; all such functions should be regarded as equivalent. A monotonic increasing function of mean codeword lengths is not the same as a mean codeword length, but minimizing a monotonic increasing function of a mean codeword length gives the same results as minimizing the mean codeword length itself. Thus, there is no harm to use monotonic increasing functions of entropy and mean codeword lengths. Below, we illustrate the correspondence between standard measures of entropy and the codeword lengths.

6.2 DEVELOPMENT OF EXPONENTIATED CODEWORD LENGTHS THROUGH MEASURES OF DIVERGENCE

In this section, we prove certain coding theorems, the special cases of which provide exponentiated mean codeword lengths already existing in the literature of coding theory.

Theorem-I: If $l_1, l_2, l_3, ..., l_n$ are the lengths of a uniquely decipherable code, then

$$L_{\alpha,\beta,k} \geq \left[H_\beta^\alpha(P)\right]_k - \frac{k(1-\beta)-(1-\alpha)}{\beta-\alpha}\log_D \sum_{i=1}^{n} D^{-l_i} \tag{6.2.1}$$

where $L_{\alpha,\beta,k} = \dfrac{1}{\alpha-\beta}\left[(\alpha-1)L^\alpha - k(\beta-1)L^\beta\right]$, k is some real constant, α, β are real parameters, $\left[H_\beta^\alpha(P)\right]_k$ is well known measure of entropy due to Kapur [7] and

$$L^\alpha = \frac{1}{\alpha-1}\log_D\left(\frac{\sum_{i=1}^{n} p_i^\alpha D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} p_i^\alpha}\right), L^\beta = \frac{1}{\beta-1}\log_D\left(\frac{\sum_{i=1}^{n} p_i^\beta D^{-l_i(1-\beta)}}{\sum_{i=1}^{n} p_i^\beta}\right)$$

are Kapur's [10] exponentiated mean codeword lengths.

Proof. We know that Kapur's [9] measure of directed divergence is given by

$$K(P:Q) = \frac{1}{\alpha-\beta}\log_D \frac{\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}}{\left(\sum_{i=1}^{n} p_i^\beta q_i^{1-\beta}\right)^k}; \alpha \neq 1, \beta \neq 1, \alpha, \beta > 0,\ k > 0,\ (\alpha-1)\, and \left(\beta-1\right)$$

have opposite signs.

Since $K(P:Q) \geq 0$, letting $q_i = \dfrac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}$, the above expression gives

$$\frac{1}{\alpha-\beta}\log_D \frac{\sum_{i=1}^{n} p_i^{\alpha}\left(\frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}\right)^{1-\alpha}}{\left(\sum_{i=1}^{n} p_i^{\beta}\left(\frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}\right)^{1-\beta}\right)^{k}} \geq 0$$

$$\text{or}\ \frac{1}{\alpha-\beta}\left[\log_D \frac{\sum_{i=1}^{n} p_i^{\alpha} D^{-l_i(1-\alpha)}}{\left(\sum_{i=1}^{n} p_i^{\beta} D^{-l_i(1-\beta)}\right)^{k}} + \log_D \left(\sum_{i=1}^{n} D^{-l_i}\right)^{k(1-\beta)-(1-\alpha)}\right] \geq 0$$

$$\text{or}\ \frac{1}{\alpha-\beta}\left[\log_D \frac{\frac{\sum_{i=1}^{n} p_i^{\alpha} D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} p_i^{\alpha}}}{\left(\frac{\sum_{i=1}^{n} p_i^{\beta} D^{-l_i(1-\beta)}}{\sum_{i=1}^{n} p_i^{\beta}}\right)^{k}} + \log_D \left(\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\left(\sum_{i=1}^{n} p_i^{\beta}\right)^{k}}\right) + \log_D \left(\sum_{i=1}^{n} D^{-l_i}\right)^{k(1-\beta)-(1-\alpha)}\right] \geq 0$$

$$\text{or}\ \frac{1}{\alpha-\beta}\left[(\alpha-1)\overline{L_\alpha} - k(\beta-1)\overline{L_\beta}\right] \geq \frac{1}{\beta-\alpha}\log_D \left(\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\left(\sum_{i=1}^{n} p_i^{\beta}\right)^{k}}\right) - \frac{k(1-\beta)-(1-\alpha)}{\alpha-\beta}\log_D \sum_{i=1}^{n} D^{-l_i} \qquad (6.2.2)$$

The equation (6.2.2) further gives

$$L_{\alpha,\beta,k} \geq \left[H_\beta^\alpha(P)\right]_k - \frac{k(1-\beta)-(1-\alpha)}{\alpha-\beta}\log_D \sum_{i=1}^{n} D^{-l_i}$$

where

$$\left[H_{\beta}^{\alpha}(P)\right]_k = \frac{1}{\beta-\alpha}\log_D\left(\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\left(\sum_{i=1}^{n} p_i^{\beta}\right)^k}\right); \alpha \neq 1, \beta \neq 1, \alpha > 1, \beta < 1 \ or\ \alpha < 1, \beta > 1$$ is Kapur's [8] additive measure of entropy. This proves the theorem.

Special cases

1. For $k=1$, (6.2.2) becomes

$$\frac{1}{\alpha-\beta}\left[(\alpha-1)L^{\alpha}-(\beta-1)L^{\beta}\right] \geq \frac{1}{\beta-\alpha}\log_D\left(\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i^{\beta}}\right) - \log_D \sum_{i=1}^{n} D^{-l_i},$$

that is,

$$L_{\alpha,\beta} \geq H_{\beta}^{\alpha}(P) - \log_D \sum_{i=1}^{n} D^{-l_i} \tag{6.2.3}$$

where

$$L_{\alpha,\beta} = \frac{1}{\alpha-\beta}\left[(\alpha-1)L^{\alpha}-(\beta-1)L^{\beta}\right]$$

is the exponentiated mean of order α and type β and $H_{\beta}^{\alpha}(P)$ is Kapur's [8] entropy of order α and type β.

Now, since $\sum_{i=1}^{n} D^{-l_i}$ always lies between D^{-1} and 1, equation (6.2.3), the lower bound for $L_{\alpha,\beta}$ lies between $H_{\beta}^{\alpha}(P)$ and $H_{\beta}^{\alpha}(P)+1$.

2. For $k=1, \beta=1$ (6.2.2) becomes

$$L^{\alpha} \geq \frac{1}{1-\alpha}\log_D \sum_{i=1}^{n} p_i^{\alpha} - \log_D \sum_{i=1}^{n} D^{-l_i}$$

that is,

$$L^{\alpha} \geq R_{\alpha}(P) - \log_D \sum_{i=1}^{n} D^{-l_i}$$

where L^{α} is the exponentiated mean of order α, $R_{\alpha}(P)$ is Renyi's [21] entropy of order α.

So, the lower bound for L^{α} lies between $R_{\alpha}(P)$ and $R_{\alpha}(P)+1$.

3. For $k=1, \beta=1$ and $\alpha \to 1$, equation (6.2.2) becomes

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

that is, $L \geq H(P) - \log_D \sum_{i=1}^{n} D^{-l_i}$

where $L = \sum_{i=1}^{n} p_i l_i$ is the mean codeword length and $H(P)$ is Shannon's [22] measure of entropy.

Thus, the lower bound for L lies between $H(P)$ and $H(P)+1$.

Theorem-II: If $l_1, l_2, l_3, ..., l_n$ are the lengths of a uniquely decipherable code, then

$$L^{\alpha} \geq R_{\alpha}(P) - \log_D \sum_{i=1}^{n} D^{-l_i} \tag{6.2.4}$$

where $R_{\alpha}(P)$ is well known Renyi's [21] measure of entropy and

$$L^{\alpha} = \frac{1}{\alpha - 1} \log_D \frac{\sum_{i=1}^{n} p_i^{\alpha} D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} p_i^{\alpha}}$$

is Kapur's [10] exponentiated mean codeword length of order α.

Proof. We know that Kapur's [9] measure of directed divergence is given by

$$K(P:Q) = \frac{1}{\alpha - 1}\left[tan^{-1}\left(\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} \right) - \frac{\pi}{4} \right]; \ \alpha > 1$$

Thus, we must have

$$\frac{1}{\alpha - 1}\left[tan^{-1}\left(\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} \right) - \frac{\pi}{4} \right] \geq 0 \tag{6.2.5}$$

Letting $q_i = \frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}$ in (6.2.5), we have

$$tan^{-1}\left(\sum_{i=1}^{n} p_i^{\alpha} \frac{D^{-l_i(1-\alpha)}}{\left(\sum_{i=1}^{n} D^{-l_i} \right)^{1-\alpha}} \right) \geq \frac{\pi}{4}$$

$$\text{or} \sum_{i=1}^{n} p_i^{\alpha} \frac{D^{-l_i(1-\alpha)}}{\left(\sum_{i=1}^{n} D^{-l_i}\right)^{1-\alpha}} \geq 1$$

Taking logarithms both sides, the above equation gives

$$\log_D \sum_{i=1}^{n} p_i^{\alpha} D^{-l_i(1-\alpha)} \geq (1-\alpha) \log_D \sum_{i=1}^{n} D^{-l_i}$$

that is,

$$\log_D \frac{\sum_{i=1}^{n} p_i^{\alpha} D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} p_i^{\alpha}} + \log_D \sum_{i=1}^{n} p_i^{\alpha} \geq (1-\alpha) \log_D \sum_{i=1}^{n} D^{-l_i}$$

that is,

$$\frac{1}{\alpha-1} \log_D \frac{\sum_{i=1}^{n} p_i^{\alpha} D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} p_i^{\alpha}} + \frac{1}{\alpha-1} \log_D \sum_{i=1}^{n} p_i^{\alpha} \geq -\log_D \sum_{i=1}^{n} D^{-l_i}$$

that is,

$$\frac{1}{\alpha-1} \log_D \frac{\sum_{i=1}^{n} p_i^{\alpha} D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} p_i^{\alpha}} \geq \frac{1}{1-\alpha} \log_D \sum_{i=1}^{n} p_i^{\alpha} - \log_D \sum_{i=1}^{n} D^{-l_i}$$

that is,

$$L^{\alpha} \geq R_{\alpha}(P) - \log_D \sum_{i=1}^{n} D^{-l_i}$$

where L^{α} is Kapur's [10] exponentiated mean codeword length of order α and $R_{\alpha}(P)$ is Renyi's [21] measure of entropy.

This proves the theorem.

Note: Since $\sum_{i=1}^{n} D^{-l_i}$ lies between D^{-1} and 1, the lower bound for L^{α} lies between $R_{\alpha}(P)$ and $R_{\alpha}(P)+1$.

In the next section, we have provided the derivations of the mean codeword lengths already existing in the literature of coding theory.

6.3 DERIVATIONS OF WELL KNOWN EXISTING CODEWORD LENGTHS

In this section, we provide the applications of Holder's inequality to prove a new coding theorem, the special cases of which provide the existing mean codeword lengths due to Campbell [4] and Shannon [22].

Theorem-I: If $l_1, l_2, l_3, ..., l_n$ are the lengths of a uniquely decipherable code, then we have

$$-\frac{1+\log_D \beta}{\log_D \beta}\log_D \sum_{i=1}^{n}\left(p_i^{\frac{1}{1+\log_D \beta}} \beta^{\frac{\log_D p_i}{1+\log_D \beta}} D^{-l_i\left(\frac{\log_D \beta}{1+\log_D \beta}\right)}\right) \geq -\frac{1}{\log_D \beta}\log_D \sum_{i=1}^{n} p_i \beta^{\log_D p_i} \tag{6.3.1}$$

where $\beta > 0, \beta \neq 1$.

Proof. Applying Holder's inequality, we have

$$\sum_{i=1}^{n} x_i y_i \geq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}} \text{ where } \frac{1}{p}+\frac{1}{q}=1,\ p \text{ or } q<1 \tag{6.3.2}$$

Substituting $x_i = p_i^{-\frac{1}{\log_D \beta}} \beta^{-\frac{\log_D p_i}{\log_D \beta}}, y_i = p_i^{\frac{1}{\log_D \beta}} \beta^{\frac{\log_D p_i}{\log_D \beta}} D^{-l_i}, \frac{1}{p} = -\frac{1}{\log_D \beta}, \frac{1}{q} = \frac{1+\log_D \beta}{\log_D \beta}$ in equation (6.3.2), we get

$$\sum_{i=1}^{n} D^{-l_i} \geq \left[\sum_{i=1}^{n}\left(p_i^{\frac{-1}{\log_D \beta}} \beta^{\frac{-\log_D p_i}{\log_D \beta}}\right)^{-\log_D \beta}\right]^{\frac{-1}{\log_D \beta}} \left[\sum_{i=1}^{n}\left(p_i^{\frac{1}{\log_D \beta}} \beta^{\frac{\log_D p_i}{\log_D \beta}} D^{-l_i}\right)^{\frac{\log_D \beta}{1+\log_{D\beta}}}\right]^{\frac{1+\log_D \beta}{\log_D \beta}}$$

that is, $$\sum_{i=1}^{n} D^{-l_i} \geq \left[\sum_{i=1}^{n} p_i \beta^{\log_D p_i}\right]^{\frac{-1}{\log_D \beta}} \left[\sum_{i=1}^{n}\left(p_i^{\frac{1}{1+\log_D \beta}} \beta^{\frac{\log_D p_i}{1+\log_D \beta}} D^{-l_i\left(\frac{\log_D \beta}{1+\log_D \beta}\right)}\right)\right]^{\frac{1+\log_D \beta}{\log_D \beta}}$$

Taking logarithms both sides, the above equation gives

$$0 \geq -\frac{1}{\log_D \beta}\log_D \sum_{i=1}^{n} p_i \beta^{\log_D p_i} + \frac{1+\log_D \beta}{\log_D \beta}\log_D \sum_{i=1}^{n}\left(p_i^{\frac{1}{1+\log_D \beta}} \beta^{\frac{\log_D p_i}{1+\log_D \beta}} D^{-l_i\left(\frac{\log_D \beta}{1+\log_D \beta}\right)}\right)$$

that is,

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

which proves the theorem.

This is to be noted that the R.H.S of (6.3.1) is a measure of entropy but the L.H.S is not a mean codeword length. Next, we discuss the particular cases:

Case-I: Taking $\beta = D$, (6.3.1) becomes

$$-2\log_D \sum_{i=1}^{n}\left(p_i^{\frac{1}{2}} D^{\frac{\log_D p_i}{2}} D^{-l_i\left(\frac{1}{2}\right)} \right) \geq -\log_D \sum_{i=1}^{n} p_i D^{\log_D p_i}$$

that is,

$$-2\log_D \sum_{i=1}^{n}\left(p_i D^{-l_i\left(\frac{1}{2}\right)} \right) \geq -\log_D \sum_{i=1}^{n} p_i^2 \qquad (6.3.3)$$

where L.H.S of (6.3.3) is Campbell's [4] exponentiated mean codeword length given by

$$L_\alpha = \frac{\alpha}{1-\alpha}\log_D\left(\sum_{i=1}^{n} p_i D^{l_i(1-\alpha)/\alpha} \right) \text{for } \alpha = 2$$

and the R.H.S of (6.3.3) is Renyi's [21] entropy of order 2.

Case-II: If we take $\beta \to 1$ in (6.3.1), we get $\frac{0}{0}$ form. Applying L Hospital's rule, we get

$$\lim_{\beta\to 1}\left(-\log_D \sum_{i=1}^{n}\left(p_i^{\frac{1}{1+\log_D\beta}} \beta^{\frac{\log_D p_i}{1+\log_D\beta}} D^{-l_i\left(\frac{\log_D\beta}{1+\log_D\beta}\right)} \right) - \frac{(1+\log_D\beta)}{\sum_{i=1}^{n}\left(p_i^{\frac{1}{1+\log_D\beta}} \beta^{\frac{\log_D p_i}{1+\log_D\beta}} D^{-l_i\left(\frac{\log_D\beta}{1+\log_D\beta}\right)} \right)} \right)$$

$$\lim_{\beta\to 1}\sum_{i=1}^{n}\left(-\frac{p_i^{\frac{1}{1+\log_D\beta}} \log_D p_i . \beta^{\frac{\log_D p_i}{1+\log_D\beta}} D^{-l_i\left(\frac{\log_D\beta}{1+\log_D\beta}\right)}}{(1+\log_D\beta)^2} \right)$$

$$\lim_{\beta\to 1}\sum_{i=1}^{n}\left(\frac{p_i^{\frac{1}{1+\log_D\beta}} \beta^{\frac{\log_D p_i}{1+\log_D\beta}} \log_D p_i D^{-l_i\left(\frac{\log_D\beta}{1+\log_D\beta}\right)}}{(1+\log_D\beta)^2} - \frac{l_i p_i^{\frac{1}{1+\log_D\beta}} \beta^{\frac{\log_D p_i}{1+\log_D\beta}} D^{-l_i\left(\frac{\log_D\beta}{1+\log_D\beta}\right)}}{(1+\log_D\beta)^2} \right)$$

$$\geq \lim_{\beta\to 1} -\frac{\frac{1}{\sum_{i=1}^{n} p_i \beta^{\log_D p_i}}\left(\sum_{i=1}^{n} p_i \log_D p_i \beta^{\log_D p_i - 1} \right)}{\frac{1}{\beta}}$$

Upon simplification, the above equation gives

$$\sum_{i=1}^{n} p_i l_i \geq -\sum_{i=1}^{n} p_i \log_D p_i \tag{6.3.4}$$

The R.H.S of equation (6.3.4) is Shannon's [22] measure of entropy and the L.H.S is arithmetic mean codeword length.

In the sequel, we shall provide the derivations of the existing mean codeword lengths by the use of weighted information, a concept introduced by Belis and Guiasu [3]. The authors enriched the usual description of the information source by endowing each source letter with an additional parameter, for example, the profit or advantage the user gets when the event corresponding to that letter occurs. They remarked that the additional parameter should be subjective in nature and not an objective character. The authors called the additional parameter as utility or weight and considered the following model for a finite random experiment:

$$\begin{bmatrix} A \\ \\ P \\ \\ W \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ \\ p_1 & p_2 & \cdots & p_n \\ \\ w_1 & w_1 & \cdots & w_n \end{bmatrix} \tag{6.3.5}$$

where A is the alphabet, P is the probability distribution and W is the utility or weighted distribution.

Taking in view the weighted scheme (6.3.5), Guiasu and Picard [5] considered the problem of encoding the letter output by means of a single-letter prefix code, whose code words have lengths $l_1, l_2, \ldots, l_n$ satisfying the constraint

$$\sum_{i=1}^{n} D^{-l_i} = 1 \tag{6.3.6}$$

where D is the size of the code alphabet. The authors defined the quantity

$$L(W) = \frac{\sum_{i=1}^{n} l_i w_i p_i}{\sum_{i=1}^{n} w_i p_i} \tag{6.3.7}$$

and called it the weighted mean codeword length of the code and derived a lower bound for it. Longo [17] provided the interpretation for the weighted mean codeword length by considering the following transmission problem:

Suppose an information source generates the letters $\{a_1, a_2, \ldots, a_n\}$ with probabilities $\{p_1, p_2, \ldots, p_n\}$ and these letters are to be encoded by means of prefix code with code words of length $\{l_1, l_2, \ldots, l_n\}$. Since the different letters are of the different importance to the receiver, their code words should be handled differently, for example, these code words should be transmitted through the channel with different conditions or these should go through different channels or these should be recorded using different storage devices. Under the present situation, this will yield a different per letter cost for each codeword, w_i (say). Thus, the cost of transmitting the letter a_i is proportional to $l_i w_i$, that is,

$$C_i = l_i w_i \tag{6.3.8}$$

Thus, if one has to minimize the average cost

$$C = \sum_{i=1}^{n} C_i p_i \, , \tag{6.3.9}$$

then one can try to minimize the quantity

$$C = \sum_{i=1}^{n} l_i w_i p_i \tag{6.3.10}$$

This is equivalent to minimizing the codeword length given in (6.3.7).

In the next theorem, we make use of well known source coding theorem due to Gurdial and Pessoa [6] proved under the weighted distribution to derive the existing codeword length, that is, we shall prove that Kapur's [8] measure of entropy is the lower bound of the well known mean codeword length already existing in the literature of coding theory. The purpose of proving this theorem is to make use of the weighted code word length explained above.

Theorem-II: If $l_1, l_2, l_3, \ldots, l_n$ are the lengths of a uniquely decipherable code, then we have

$$\frac{1}{\beta-\alpha}\log_D\left\{\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i^{\beta}}\right\} \le \frac{1}{\beta-\alpha}\log_D \frac{\left(\sum_{i=1}^{n} p_i D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\alpha}}{\left(\sum_{i=1}^{n} p_i D^{-l_i\left(\frac{\beta-1}{\beta}\right)}\right)^{\beta}} \tag{6.3.11}$$

Proof. To prove this theorem, we consider source coding theorem proved by Gurdial and Pessoa's [6]. According to this theorem, if $l_1, l_2, l_3, \ldots, l_n$ are the lengths of a uniquely decipherable code, then the following inequality hold:

$$\frac{\alpha}{1-\alpha}\log_D\left(\sum_{i=1}^{n} p_i \left\{\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right\}^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right) \geq \frac{1}{1-\alpha}\log_D\left\{\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i}\right\} \tag{6.3.12}$$

where

$$H_{\alpha}^{\beta}(P,W) = \frac{1}{1-\alpha}\log_D\left\{\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i}\right\}; \alpha \neq 1, \alpha > 0 \tag{6.3.13}$$

is well known Gurdial and Pessoa's [6] weighted entropy and

$$L_{\alpha}^{\beta}(W) = \frac{\alpha}{1-\alpha}\log_D\left(\sum_{i=1}^{n} p_i \left\{\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right\}^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right); \alpha \neq 1, \alpha > 0 \tag{6.3.14}$$

is the parametric weighted mean codeword length introduced by Gurdial and Pessoa [6].

Now, we apply the above inequality (6.3.12) to derive the existing codeword length as follows:

When $\alpha < 1$, from inequality (6.3.12), we have

$$\log_D\left\{\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i}\right\} \leq \log_D\left(\sum_{i=1}^{n} p_i \left\{\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right\}^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\alpha} \tag{6.3.15}$$

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

$$-\log_D\left\{\frac{\sum_{i=1}^{n} w_i p_i^{\beta}}{\sum_{i=1}^{n} w_i p_i}\right\} \leq -\log_D\left(\sum_{i=1}^{n} p_i \left\{\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right\}^{\frac{1}{\beta}} D^{-l_i\left(\frac{\beta-1}{\beta}\right)}\right)^{\beta} \tag{6.3.16}$$

Adding equations (6.3.15) and (6.3.16), we get the following equation:

$$\log_D\left\{\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i}\right\}-\log_D\left\{\frac{\sum_{i=1}^{n} w_i p_i^{\beta}}{\sum_{i=1}^{n} w_i p_i}\right\}$$

$$\leq \log_D\left(\sum_{i=1}^{n} p_i\left\{\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right\}^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\alpha}-\log_D\left(\sum_{i=1}^{n} p_i\left\{\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right\}^{\frac{1}{\beta}} D^{-l_i\left(\frac{\beta-1}{\beta}\right)}\right)^{\beta} \tag{6.3.17}$$

Upon ignoring weights, equation (6.3.17) gives

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

Thus, for $\alpha<1, \beta>1$, we have

$$\frac{1}{\beta-\alpha}\log_D\left\{\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i^{\beta}}\right\}\leq \frac{1}{\beta-\alpha}\log_D\frac{\left(\sum_{i=1}^{n} p_i D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\alpha}}{\left(\sum_{i=1}^{n} p_i D^{-l_i\left(\frac{\beta-1}{\beta}\right)}\right)^{\beta}} \tag{6.3.18}$$

Proceeding on similar lines, the result (6.3.18) can be proved for $\alpha>1$ and $\beta<1$.

The L.H.S. of equation (6.3.18) is Kapur's [8] measure of entropy of order α and type β and R.H.S. is Kapur's [10] mean codeword length. Hence the theorem.

Proceeding as above, many existing codeword lengths can be derived via weighted code word lengths or by the use of Holder's inequality.

6.4 DEVELOPMENT OF INFORMATION THEORETIC INEQUALITIES VIA CODING THEORY AND MEASURES OF DIVERGENCE

In this section, we introduce some new inequalities usually applicable in the field of information theory. These inequalities have been developed by the use of coding theory and divergence measures. Below, we discuss the method for obtaining these inequalities:

I. Firstly, we consider the expression of divergence measure taken by Kapur [9]. This measure is also known as Jensen-Burg measure of directed divergence and is given by

$$D_\lambda\left(P:Q\right)=\log_D\prod_{i=1}^{n}\frac{\lambda p_i+(1-\lambda)q_i}{p_i^{\lambda}q_i^{1-\lambda}};\ 0\le\lambda\le1,\ 0\le p_i\le1 \tag{6.4.1}$$

Now $\log_D\prod_{i=1}^{n}\frac{\lambda p_i+(1-\lambda)q_i}{p_i^{\lambda}q_i^{1-\lambda}}\ge0$

$$\Rightarrow\log_D\prod_{i=1}^{n}(\lambda p_i+(1-\lambda)q_i)\ge\log_D\prod_{i=1}^{n}p_i^{\lambda}q_i^{1-\lambda} \tag{6.4.2}$$

Putting $q_i=\frac{D^{-l_i}}{\sum_{i=1}^{n}D^{-l_i}}$ in (6.4.2), we get

$$\log_D\prod_{i=1}^{n}p_i^{\lambda}\left(\frac{D^{-l_i}}{\sum_{i=1}^{n}D^{-l_i}}\right)^{1-\lambda}\le\log_D\prod_{i=1}^{n}\left(\lambda p_i+(1-\lambda)\frac{D^{-l_i}}{\sum_{i=1}^{n}D^{-l_i}}\right)$$

$$\Rightarrow\sum_{i=1}^{n}\log_D p_i^{\lambda}\left(\frac{D^{-l_i}}{\sum_{i=1}^{n}D^{-l_i}}\right)^{1-\lambda}\le\sum_{i=1}^{n}\log_D\left(\lambda p_i+(1-\lambda)\frac{D^{-l_i}}{\sum_{i=1}^{n}D^{-l_i}}\right)$$

Using the inequality $\log x\le x-1$, the above equation becomes

$$\sum_{i=1}^{n}\log_D p_i^{\lambda}\left(\frac{D^{-l_i}}{\sum_{i=1}^{n}D^{-l_i}}\right)^{1-\lambda}\le\sum_{i=1}^{n}\left(\lambda p_i+(1-\lambda)\frac{D^{-l_i}}{\sum_{i=1}^{n}D^{-l_i}}-1\right) \tag{6.4.3}$$

Letting $l_i=l_2=\ldots=l_n=l$, equation (6.4.3) gives

$$\sum_{i=1}^{n}\log_D p_i^{\lambda}\left(\frac{1}{n}\right)^{1-\lambda}\le\sum_{i=1}^{n}\left(\lambda p_i+(1-\lambda)\frac{1}{n}-1\right)$$

$$\Rightarrow\lambda\sum_{i=1}^{n}\log_D p_i+(\lambda-1)n\log_D n\le1-n$$

$$\Rightarrow\lambda\sum_{i=1}^{n}\log_D p_i\le1-n+(1-\lambda)n\log_D n \tag{6.4.4}$$

For $\lambda=1$, equation (6.4.4) becomes

$$\Rightarrow \sum_{i=1}^{n} \log_D p_i \le 1-n$$

which is a new information theoretic inequality involving probabilities.

II. Again, we consider a generalized divergence measure due to Kapur [9], given by

$$D_\alpha^\beta\left(P,Q\right)=\frac{1}{\alpha-\beta}\log_D\left(\frac{\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}}{\sum_{i=1}^{n} p_i^\beta q_i^{1-\beta}}\right);\ \alpha\ge 1,\beta\le 1;\ \alpha\le 1,\beta\ge 1 \tag{6.4.5}$$

Using the condition of non-negativity of divergence measure and taking $q_i=\frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}$ in (6.4.5), we get

$$\frac{1}{\alpha-\beta}\log_D\left\{\sum_{i=1}^{n} p_i^\alpha\left\{\frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}\right\}^{1-\alpha}\right\}-\frac{1}{\alpha-\beta}\log_D\left\{\sum_{i=1}^{n} p_i^\beta\left\{\frac{D^{-l_i}}{\sum_{i=1}^{n} D^{-l_i}}\right\}^{1-\beta}\right\}\ge 0$$

that is,

$$\frac{1}{\alpha-\beta}\left\{\log_D\left\{\sum_{i=1}^{n} D^{-l_i}\right\}^{\alpha-\beta}+\log_D\frac{\sum_{i=1}^{n} p_i^\alpha D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} p_i^\beta D^{-l_i(1-\beta)}}\right\}\ge 0 \tag{6.4.6}$$

If we take $l_1=l_2=...=l_n=l$ in (6.4.6), we get

$$\log_D\left\{\sum_{i=1}^{n} D^{-l}\right\}+\frac{1}{\alpha-\beta}\left\{\log_D D^{l(\alpha-\beta)}+\log_D\frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i^\beta}\right\}\ge 0$$

that is, $\log_D\left\{\sum_{i=1}^{n} D^{-l}\right\}+l-\frac{1}{\beta-\alpha}\log_D\frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i^\beta}\ge 0$

that is, $\log_D n\ge\frac{1}{\beta-\alpha}\log_D\frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i^\beta}$

that is, $H_\alpha^\beta(P)\le\log_D n$ (6.4.7) where $H_\alpha^\beta(P)=\frac{1}{\beta-\alpha}\log_D \frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\beta}$ is well known Kapur's [8] measure of entropy.

Thus, (6.4.7) is another information inequality.

Proceeding as above, many new information theoretic inequalities can be developed either via coding theory approach or with the help of divergence measures.

6.5 GENERATING POSSIBLE GENERALIZED MEASURES OF WEIGHTED ENTROPY VIA CODING THEORY

It is to be observed that in coding theory, while proving coding theorems, one should know

(i) the well defined mean codeword lengths

(ii) the existing measures of entropy

With the help of first approach, we can generate new measures of entropy whereas the second approach gives us a method of developing new mean codeword lengths. In this section, we have made use of the concept of weighted entropy introduced by Belis and Guiasu [3] and generated some new possible generalized measures of weighted entropy by using the first approach. These measures have been developed with the help of following theorems:

Theorem-I: If $l_1,l_2,l_3,...,l_n$ are the lengths of a uniquely decipherable code, then we have

$$\frac{\alpha\beta}{1-\alpha}\log_D \sum_{i=1}^n\left(p_i\left(\frac{w_i}{\sum_{i=1}^n w_i p_i}\right)^{\frac{1}{a}} D^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}\right) \geq \frac{\alpha(1-\beta)-1}{\alpha-1}\log_D\left(\sum_{i=1}^n p_i^{\frac{\alpha\beta}{1-\alpha(1-\beta)}}\left(\frac{w_i}{\sum_{i=1}^n w_i p_i}\right)^{\frac{\beta}{1-a(1-\beta)}}\right) \quad (6.5.1)$$

$;\alpha \neq 0, \alpha>1,\ \beta>1$

Proof. By Holder's inequality, we have

$$\sum_{i=1}^n x_i y_i \geq \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^n y_i^q\right)^{\frac{1}{q}} \text{ where } \frac{1}{p}+\frac{1}{q}=1,\ p\, or\, q<1 \quad (6.5.2)$$

Substituting $x_i = p_i^{\frac{\alpha\beta}{\alpha-1}}\left(\frac{w_i}{\sum_{i=1}^n w_i p_i}\right)^{\frac{\beta}{\alpha-1}} D^{-l_i}$, $y_i = p_i^{\frac{\alpha\beta}{1-\alpha}}\left(\frac{w_i}{\sum_{i=1}^n w_i p_i}\right)^{\frac{\beta}{1-\alpha}}$, $p=\frac{\alpha-1}{\alpha\beta}$, $q=\frac{\alpha-1}{\alpha(1-\beta)-1}$

in equation (6.5.2), we get

$$\sum_{i=1}^{n} D^{-l_i} \geq \left[\sum_{i=1}^{n}\left(p_i^{\frac{\alpha\beta}{\alpha-1}}\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{\beta}{\alpha-1}} D^{-l_i}\right)^{\frac{\alpha-1}{\alpha\beta}}\right]^{\frac{\alpha\beta}{\alpha-1}}\left[\sum_{i=1}^{n}\left(p_i^{\frac{\alpha\beta}{1-\alpha}}\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{\beta}{1-\alpha}}\right)^{\frac{\alpha-1}{\alpha(1-\beta)-1}}\right]^{\frac{\alpha(1-\beta)-1}{\alpha-1}}$$

that is,

$$\sum_{i=1}^{n} D^{-l_i} \geq \left[\sum_{i=1}^{n} p_i\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}\right]^{\frac{\alpha\beta}{\alpha-1}}\left[\sum_{i=1}^{n}\left(p_i^{\frac{\alpha\beta}{1-\alpha(1-\beta)}}\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{\beta}{1-\alpha(1-\beta)}}\right)\right]^{\frac{\alpha(1-\beta)-1}{\alpha-1}}$$

Taking logarithms both sides, the above equation gives

$$0 \geq \frac{\alpha\beta}{\alpha-1}\log_D\left[\sum_{i=1}^{n} p_i\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{1}{\alpha}} D^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}\right]$$

$$+\frac{\alpha(1-\beta)-1}{\alpha-1}\log_D\left[\sum_{i=1}^{n}\left(p_i^{\frac{\alpha\beta}{1-\alpha(1-\beta)}}\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{\beta}{1-\alpha(1-\beta)}}\right)\right]$$

that is, $$\frac{\alpha\beta}{1-\alpha}\log_D\sum_{i=1}^{n}\left(p_i\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{1}{a}} D^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}\right) \geq \frac{\alpha(1-\beta)-1}{\alpha-1}\log_D\left(\sum_{i=1}^{n} p_i^{\frac{\alpha\beta}{1-\alpha(1-\beta)}}\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{\beta}{1-a(1-\beta)}}\right)$$

which proves the theorem.

From the above theorem, we have developed a new weighted mean codeword length as given below:

$$ {}_{\alpha}^{\beta}L(W)=\frac{\alpha\beta}{1-\alpha}\log_D\sum_{i=1}^{n}\left(p_i\left(\frac{w_i}{\sum_{i=1}^{n}w_ip_i}\right)^{\frac{1}{a}}D^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}\right);\ \alpha\neq 1,\ \alpha>1,\beta>1 \tag{6.5.3}$$

Ignoring weights equation (6.5.3) gives

$$ {}_{\alpha}^{\beta}L=\frac{\alpha\beta}{1-\alpha}\log_D\sum_{i=1}^{n}\left(p_iD^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}\right) \tag{6.5.4}$$

For $\beta=1$, (6.5.4) becomes

$$ {}_{\alpha}^{1}L=\frac{\alpha}{1-\alpha}\log_D\sum_{i=1}^{n}\left(p_iD^{-l_i(\alpha-1)}\right)$$

which is Campbell's [4] exponentiated mean codeword length.

Moreover, as $\alpha\to 1$ in equation (6.5.3), the nature of ${}_{\alpha}^{\beta}L(W)$ is of the $\frac{0}{0}$ form and thus, applying L Hospital's rule, we get

$$ \underset{\alpha\to 1}{lt}\ {}_{\alpha}^{\beta}L(W)=-\alpha\beta\left[\frac{\sum_{i=1}^{n}p_i\left(-\left(\frac{w_i}{\sum_{i=1}^{n}w_ip_i}\right)^{\frac{1}{a}}\frac{D^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}l_i}{\alpha^2\beta}-\left(\frac{w_i}{\sum_{i=1}^{n}w_ip_i}\right)^{\frac{1}{a}}\frac{D^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}}{\alpha^2}\log_D\left(\frac{w_i}{\sum_{i=1}^{n}w_ip_i}\right)\right)}{\left(p_i\left(\frac{w_i}{\sum_{i=1}^{n}w_ip_i}\right)^{\frac{1}{a}}D^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}\right)}\right]$$

$$ -\beta\log_D\left[\sum_{i=1}^{n}p_i\left(\frac{w_i}{\sum_{i=1}^{n}w_ip_i}\right)^{\frac{1}{a}}D^{-l_i\left(\frac{\alpha-1}{\alpha\beta}\right)}\right]$$

that is,

$$\underset{\alpha\to 1}{lt}\ {}_{\alpha}^{\beta}L(W)=\sum_{i=1}^{n}\frac{p_i w_i l_i}{\sum_{i=1}^{n} w_i p_i}-\beta\frac{\sum_{i=1}^{n} w_i p_i \log_D\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)}{\sum_{i=1}^{n} w_i p_i}$$

For $\beta=0$, we have

$$\underset{\alpha\to 1}{lt}\ {}_{\alpha}L(W)=\frac{\sum_{i=1}^{n} p_i l_i w_i}{\sum_{i=1}^{n} w_i p_i}$$

which is Longo's [17] weighted mean codeword length.

Thus, we conclude that the weighted mean codeword length introduced in (6.5.3) is a valid generalized weighted mean codeword length.

New possible measure of weighted entropy

Equation (6.5.1) provides the following new possible generalized weighted measure of entropy:

$${}_{\alpha}^{\beta}H(P,W)=\frac{\alpha(1-\beta)-1}{\alpha-1}\log_D\left(\sum_{i=1}^{n} p_i^{\frac{\alpha\beta}{1-\alpha(1-\beta)}}\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{\beta}{1-\alpha(1-\beta)}}\right);\alpha>1,\ \beta>1 \tag{6.5.5}$$

Note: If $w_i=1$, for all i, then (6.5.5) becomes

$${}_{\alpha}^{\beta}H(P)=\frac{\alpha(1-\beta)-1}{\alpha-1}\log_D\left(\sum_{i=1}^{n} p_i^{\frac{\alpha\beta}{1-\alpha(1-\beta)}}\right) \tag{6.5.6}$$

Letting $\alpha\to 1$ in (6.5.6), we get

$$\underset{\alpha\to 1}{lt}\ {}_{\alpha}^{\beta}H(P)=\underset{\alpha\to 1}{lt}\left[\begin{array}{l}\left(\alpha(1-\beta)-1\right)\dfrac{1}{\sum_{i=1}^{n} p_i^{\frac{\alpha\beta}{1-\alpha(1-\beta)}}}\sum_{i=1}^{n} p_i^{\frac{\alpha\beta}{1-\alpha(1-\beta)}}\log_D p_i\,\dfrac{\beta}{\left(1-\alpha(1-\beta)\right)^2}\\ +(1-\beta)\log_D\left(\sum_{i=1}^{n} p_i^{\frac{\alpha\beta}{1-\alpha(1-\beta)}}\right)\end{array}\right]$$

$$=\left[\beta\frac{1}{\sum_{i=1}^{n}p_i}\sum_{i=1}^{n}p_i\log_D p_i\frac{1}{\beta}+(1-\beta)\log_D\left(\sum_{i=1}^{n}p_i\right)\right]$$

Thus, we have

$\underset{\alpha\to1}{lt}\ {}_{\alpha}^{\beta}H(P)=-\sum_{i=1}^{n}p_i\log_D p_i$ which is Shannon's [22] measure of entropy.

Hence, we conclude that by proving the above theorem, we have generated a new possible generalized measure of weighted entropy given in (6.5.5).

Theorem-II: If $l_1,l_2,l_3,...,l_n$ are the lengths of a uniquely decipherable code, then we have

$$L^{\alpha}(W)\geq\frac{\alpha}{\alpha-1}\log_D\left(\sum_{i=1}^{n}\frac{w_i^{\frac{1}{\alpha}}p_i}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)\tag{6.5.7}$$

where

$$L^{\alpha}(W)=\frac{1}{1-\alpha}\log_D\left(\frac{\sum_{i=1}^{n}w_ip_i^{\alpha}}{\sum_{i=1}^{n}w_ip_i^{\alpha}D^{l_i(\alpha-1)}}\right)$$

is a new weighted mean codeword length.

Proof. Applying Holder's inequality (6.5.2) and substituting

$$x_i=\frac{w_i^{\frac{1}{1-\alpha}}p_i^{\frac{\alpha}{1-\alpha}}D^{-l_i}}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{1-\alpha}}},\ y_i=\frac{w_i^{\frac{1}{\alpha-1}}p_i^{\frac{\alpha}{\alpha-1}}}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{\alpha-1}}},\quad p=1-\alpha,\quad q=\frac{\alpha-1}{\alpha}$$, we have

$$\sum_{i=1}^{n}D^{-l_i}\geq\left[\sum_{i=1}^{n}\left(\frac{w_i^{\frac{1}{1-\alpha}}p_i^{\frac{\alpha}{1-\alpha}}D^{-l_i}}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{1-\alpha}}}\right)^{1-\alpha}\right]^{\frac{1}{1-\alpha}}\left[\sum_{i=1}^{n}\left(\frac{w_i^{\frac{1}{\alpha-1}}p_i^{\frac{\alpha}{\alpha-1}}}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{\alpha-1}}}\right)^{\frac{\alpha-1}{\alpha}}\right]^{\frac{\alpha}{\alpha-1}}$$

that is,

$$\sum_{i=1}^{n} D^{-l_i} \ge \left(\sum_{i=1}^{n} \frac{w_i p_i^{\alpha} D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} w_i p_i^{\alpha}} \right)^{\frac{1}{1-\alpha}} \left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha} \right)^{\frac{1}{\alpha}}} \right)^{\frac{\alpha}{\alpha-1}}$$

Taking logarithms both sides, the above equation gives

$$0 \ge \frac{1}{1-\alpha} \log_D \left(\sum_{i=1}^{n} \frac{w_i p_i^{\alpha} D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} w_i p_i^{\alpha}} \right) + \frac{\alpha}{\alpha-1} \log_D \left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha} \right)^{\frac{1}{\alpha}}} \right)$$

that is,

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

or $$\frac{1}{1-\alpha} \log_D \left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}} \right) \ge \frac{\alpha}{\alpha-1} \log_D \left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha} \right)^{\frac{1}{\alpha}}} \right)$$

which proves the theorem.

From the above theorem, we have developed a new weighted mean codeword length as provided by the following expression:

$$L^{\alpha}(W) = \frac{1}{1-\alpha} \log_D \frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}; \ \alpha \ne 1, \alpha > 1 \tag{6.5.8}$$

Ignoring weights, equation (6.5.8) gives

$$L^{\alpha} = \frac{1}{1-\alpha} \log_D \frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i^{\alpha} D^{l_i(\alpha-1)}}$$

which is Kapur's [10] exponentiated mean codeword length.

Letting $\alpha \to 1$ in (6.5.8), we get

$$\underset{\alpha\to 1}{lt}\, L^{\alpha}(W) = -\underset{\alpha\to 1}{lt}\frac{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}{\sum_{i=1}^{n} w_i p_i^{\alpha}}\left(\frac{\left(\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}\right)\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\log_D p_i\right)-\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)\left(\sum_{i=1}^{n} w_i\left(p_i^{\alpha}\log_D p_i D^{l_i(\alpha-1)} + p_i^{\alpha} D^{l_i(\alpha-1)} l_i\right)\right)}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}\right)^2}\right)$$

$$= -\left(\frac{\left(\sum_{i=1}^{n} w_i p_i\right)\left(\sum_{i=1}^{n} w_i p_i \log_D p_i\right)-\left(\sum_{i=1}^{n} w_i p_i\right)\left(\sum_{i=1}^{n} w_i\left(p_i\log_D p_i + p_i l_i\right)\right)}{\left(\sum_{i=1}^{n} w_i p_i\right)^2}\right)$$

$$= \frac{\sum_{i=1}^{n} w_i p_i l_i}{\sum_{i=1}^{n} w_i p_i}$$

which is a Longo's [17] weighted mean codeword length.

Thus, we conclude that the weighted mean codeword length introduced in (6.5.8) is a valid generalized weighted mean codeword length.

New possible measure of weighted entropy

Equation (6.5.7) provides the following new possible generalized weighted measure of entropy:

$$H^{\alpha}(P,W) = \frac{\alpha}{\alpha-1}\log_D\left(\sum_{i=1}^{n}\frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right) \qquad (6.5.9)$$

<u>**Note:**</u> If $w_i = 1$, for all i, then (6.5.9) becomes

$$H^{\alpha}(P) = \frac{1}{1-\alpha}\log_D \sum_{i=1}^{n} p_i^{\alpha} \qquad (6.5.10)$$

which is Renyi's [21] measure of entropy.

Thus, we conclude that the mathematical expression (6.5.9) provides a new possible generalized measure of weighted entropy.

Theorem-III: If $l_1, l_2, l_3, ..., l_n$ are the lengths of a uniquely decipherable code, then we have

$$ {}_{\alpha}^{\beta}L(W) \geq \frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\sum_{i=1}^{n}\frac{w_i^{\frac{1}{\alpha}}p_i}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha\beta}\right] \tag{6.5.11} $$

where

$$ {}_{\alpha}^{\beta}L(W) = \frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\frac{\sum_{i=1}^{n}w_ip_i^{\alpha}}{\sum_{i=1}^{n}w_ip_i^{\alpha}D^{l_i(\alpha-1)}}\right)^{\beta}\right] $$

is a new weighted mean codeword length.

Proof. Applying Holder's inequality (6.5.2) and substituting

$$ x_i = \frac{w_i^{\frac{1}{1-\alpha}}p_i^{\frac{\alpha}{1-\alpha}}D^{-l_i}}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{1-\alpha}}}, y_i = \frac{w_i^{\frac{1}{\alpha-1}}p_i^{\frac{\alpha}{\alpha-1}}}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{\alpha-1}}}, \quad p = 1-\alpha, \quad q = \frac{\alpha-1}{\alpha} \ , $$

we get the following inequality

$$ \sum_{i=1}^{n}D^{-l_i} \geq \left[\sum_{i=1}^{n}\left(\frac{w_i^{\frac{1}{1-\alpha}}p_i^{\frac{\alpha}{1-\alpha}}D^{-l_i}}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{1-\alpha}}}\right)^{1-\alpha}\right]^{\frac{1}{1-\alpha}}\left[\sum_{i=1}^{n}\left(\frac{w_i^{\frac{1}{\alpha-1}}p_i^{\frac{\alpha}{\alpha-1}}}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{\alpha-1}}}\right)^{\frac{\alpha-1}{\alpha}}\right]^{\frac{\alpha}{\alpha-1}} $$

or

$$ \sum_{i=1}^{n}D^{-l_i} \geq \left(\sum_{i=1}^{n}\frac{w_ip_i^{\alpha}D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n}w_ip_i^{\alpha}}\right)^{\frac{1}{1-\alpha}}\left(\sum_{i=1}^{n}\frac{w_i^{\frac{1}{\alpha}}p_i}{\left(\sum_{i=1}^{n}w_ip_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{\frac{\alpha}{\alpha-1}} $$

Taking logarithms both sides, the above equation gives

$$0 \geq \frac{1}{1-\alpha}\log_D\left(\sum_{i=1}^{n}\frac{w_i p_i^{\alpha} D^{-l_i(1-\alpha)}}{\sum_{i=1}^{n} w_i p_i^{\alpha}}\right) + \frac{\alpha}{\alpha-1}\log_D\left(\sum_{i=1}^{n}\frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)$$

that is,

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

or $$\frac{1}{1-\alpha}\log_D\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right) \geq \frac{\alpha}{\alpha-1}\log_D\left(\sum_{i=1}^{n}\frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)$$

or

$$\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\frac{1}{1-\alpha}} \geq \left(\sum_{i=1}^{n}\frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{\frac{\alpha}{\alpha-1}} \qquad (6.5.12)$$

<u>Case-I:</u> For $\alpha > 1$, $1 - D^{(1-\alpha)\beta} > 0$.

Thus, equation (6.5.12) becomes

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

that is, $$\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta} \le \left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha\beta}$$

or $$1-\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta} \ge 1-\left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha\beta}$$

or $$\frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta}\right] \ge \frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha\beta}\right]$$

<u>Case-II:</u> For $0<\alpha<1$, $1-D^{(1-\alpha)\beta}<0$

Thus, equation (6.5.12) becomes

$$\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right) \ge \left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha}$$

that is, $$\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta} \ge \left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha\beta}$$

or $$1-\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta} \le 1-\left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha\beta}$$

or $$\frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta}\right] \geq \frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\sum_{i=1}^{n}\frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha\beta}\right]$$

which proves the theorem.

From the above theorem, we have developed a new weighted mean codeword length as given below:

$$ {}_{\alpha}^{\beta}L(W)=\frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta}\right];\ \alpha \neq 1,\ \alpha>0, \beta \neq 0 . \tag{6.5.13}$$

Letting $\alpha \to 1$ in (6.5.13), we get

$$\underset{\alpha\to 1}{lt}\ L(W)=\underset{\alpha\to 1}{lt}\ \frac{-\beta\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta-1}\left(\frac{\left(\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}\right)\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\log_D p_i\right)-\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)\left(\sum_{i=1}^{n} w_i\left(p_i^{\alpha}\log_D p_i D^{l_i(\alpha-1)}+p_i^{\alpha} D^{l_i(\alpha-1)} l_i\right)\right)}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}\right)^2}\right)}{\beta D^{(1-\alpha)\beta}}$$

$$=-\frac{\left(\sum_{i=1}^{n} w_i p_i\right)\left(\sum_{i=1}^{n} w_i p_i \log_D p_i\right)-\left(\sum_{i=1}^{n} w_i p_i\right)\left(\sum_{i=1}^{n} w_i\left(p_i\log_D p_i+p_i l_i\right)\right)}{\left(\sum_{i=1}^{n} w_i p_i\right)^2}$$

$$=\frac{\sum_{i=1}^{n} w_i p_i l_i}{\sum_{i=1}^{n} w_i p_i}$$ which is a Longo’s [17] weighted mean codeword length.

Letting $\beta \to 0$ in (6.5.13), we get

$$\underset{\beta\to 0}{lt}\, L(W) = \underset{\beta\to 0}{lt} \frac{\left(\dfrac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta} \log_D \left(\dfrac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)}{(1-\alpha) D^{(1-\alpha)\beta}}$$

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

which is mean codeword length already defined in (6.5.8).

New possible measure of weighted entropy

Equation (6.5.11) provides the following new possible generalized weighted measure of entropy:

$${}_{\alpha}^{\beta}H(P,W) = \frac{1}{1-D^{(1-\alpha)\beta}} \left[1 - \left(\sum_{i=1}^{n} \frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha\beta}\right] \tag{6.5.14}$$

Upon ignoring weights, equation (6.5.14) gives

$${}_{\alpha}^{\beta}H(P) = \frac{1}{1-D^{(1-\alpha)\beta}} \left[1 - \left(\sum_{i=1}^{n} p_i \left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{-\frac{1}{\alpha}}\right)^{-\alpha\beta}\right] \tag{6.5.15}$$

Letting $\alpha \to 1$ in (6.5.15), we get

$\underset{\alpha\to 1}{lt}\, {}_{\alpha}^{\beta}H(P) = -\sum_{i=1}^{n} p_i \log_D p_i$ which is Shannon's [22] measure of entropy.

Thus, we conclude that (6.5.9) provides a new possible generalized measure of weighted entropy.

Proceeding on similar way, we can generate many new generalized parametric measures of weighted entropy and consequently, conclude that one can develop new generalized measures of entropy via coding theory approach.

6.6 MEASURES OF ENTROPY AS POSSIBLE LOWER BOUNDS

In the existing literature of coding theory, it has been proved that many well known measures of entropy including those of Shannon's [22] entropy, Renyi's [21] entropy of order α, Kapur's [7, 8] entropies of order α and type β etc., provide lower bounds for different mean codeword lengths, while some other measures including those of Arimoto's [1] and Behara and Chawla's [2] measures of entropy provide lower bounds for some monotonic increasing functions of the mean codeword lengths but not for mean codeword lengths themselves. Such a correspondence can be discussed by taking into consideration the measures of weighted entropy.

Below, we have discussed such a correspondence between standard measures of entropy and their possible lower bounds:

I. Sharma and Taneja's [23] entropy as a possible lower bound

To show that Sharma and Taneja's [23] entropy is a possible lower bound, we make use of Khan and Autar's [12] source coding theorem. According to this theorem, if $l_1, l_2, l_3, ..., l_n$ are the lengths of a uniquely decipherable code, then we have

$$\frac{1}{1-D^{(1-\alpha)/\alpha}}\left[1-\sum_{i=1}^{n} p_i\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{1}{a}} D^{-l_i(\alpha-1)/\alpha}\right] \geq \frac{1}{1-D^{(1-\alpha)/\alpha}}\left[1-\left\{\sum_{i=1}^{n}\frac{w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i}\right\}^{\frac{1}{a}}\right]; \alpha \neq 1, \alpha > 0 \qquad (6.6.1)$$

where the L.H.S. of (6.6.1) is a weighted codeword length and the R.H.S. is Khan and Autar's [12] weighted entropy.

Now, when $\alpha < 1$, $\frac{1}{1-D^{(1-\alpha)/\alpha}} < 0$, we have

$$\left[1-\sum_{i=1}^{n} p_i\left(\frac{w_i}{\sum_{i=1}^{n} w_i p_i}\right)^{\frac{1}{a}} D^{-l_i(\alpha-1)/\alpha}\right] \leq \left[1-\left\{\sum_{i=1}^{n}\frac{w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i}\right\}^{\frac{1}{a}}\right]$$

that is,

$$\left\{\sum_{i=1}^{n}\frac{w_i p_i^{\alpha}}{\sum_{i=1}^{n}w_i p_i}\right\}\leq\left[\sum_{i=1}^{n}p_i\left(\frac{w_i}{\sum_{i=1}^{n}w_i p_i}\right)^{\frac{1}{a}}D^{-l_i(\alpha-1)/\alpha}\right]^{\alpha} \tag{6.6.2}$$

Similarly, for $\beta>1$, $\frac{1}{1-D^{(1-\beta)/\beta}}>0$, we have

$$-\left\{\sum_{i=1}^{n}\frac{w_i p_i^{\beta}}{\sum_{i=1}^{n}w_i p_i}\right\}\leq-\left[\sum_{i=1}^{n}p_i\left(\frac{w_i}{\sum_{i=1}^{n}w_i p_i}\right)^{\frac{1}{\beta}}D^{-l_i(\beta-1)/\beta}\right]^{\beta} \tag{6.6.3}$$

Adding equations (6.6.2) and (6.6.3) and then dividing by $\beta-\alpha>0$, we get the following equation:

$$\frac{1}{\beta-\alpha}\left\{\frac{\sum_{i=1}^{n}w_i p_i^{\alpha}}{\sum_{i=1}^{n}w_i p_i}-\frac{\sum_{i=1}^{n}w_i p_i^{\beta}}{\sum_{i=1}^{n}w_i p_i}\right\}$$

$$\leq\frac{1}{\beta-\alpha}\left(\left[\sum_{i=1}^{n}p_i\left(\frac{w_i}{\sum_{i=1}^{n}w_i p_i}\right)^{\frac{1}{a}}D^{-l_i(\alpha-1)/\alpha}\right]^{\alpha}-\left[\sum_{i=1}^{n}p_i\left(\frac{w_i}{\sum_{i=1}^{n}w_i p_i}\right)^{\frac{1}{\beta}}D^{-l_i(\beta-1)/\beta}\right]^{\beta}\right) \tag{6.6.4}$$

Upon ignoring weights, equation (6.6.4) gives

$$\frac{1}{\beta-\alpha}\left\{\sum_{i=1}^{n}p_i^{\alpha}-\sum_{i=1}^{n}p_i^{\beta}\right\}\leq\frac{1}{\beta-\alpha}\left(\left(\sum_{i=1}^{n}p_i D^{-l_i\left(\frac{\alpha-1}{\alpha}\right)}\right)^{\alpha}+\left(\sum_{i=1}^{n}p_i D^{-l_i\left(\frac{\beta-1}{\beta}\right)}\right)^{\beta}\right) \tag{6.6.5}$$

Proceeding on similar lines, the result (6.6.5) can be proved for $\alpha<1$ and $\beta<1$.

This is to be noted that the L.H.S. of equation (6.6.5) is Sharma and Taneja's [23] measure of entropy of order α and type β but R.H.S. is neither a mean codeword length nor a monotonic increasing function of mean codeword length.

II. Arimoto's [1] entropy as a possible lower bound

To prove that Arimoto's [1] entropy can be a possible lower bound, we use Theorem-III proved in above section. We have shown that if $l_1, l_2, l_3, ..., l_n$ are the lengths of a uniquely decipherable code, then

$$\frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right)^{\beta}\right] \geq \frac{1}{1-D^{(1-\alpha)\beta}}\left[1-\left(\sum_{i=1}^{n}\frac{w_i^{\frac{1}{\alpha}} p_i}{\left(\sum_{i=1}^{n} w_i p_i^{\alpha}\right)^{\frac{1}{\alpha}}}\right)^{-\alpha\beta}\right]$$

Now, when $\alpha < 1, \beta > 0$, then $\frac{1}{1-D^{(1-\alpha)\beta}} < 0$. Thus, the above equation gives

$$\left\{\sum_{i=1}^{n}\frac{w_i^{\frac{1}{a}} p_i}{\left\{\sum_{i=1}^{n} w_i p_i^{\alpha}\right\}^{\frac{1}{a}}}\right\}^{-\alpha} \leq \left[\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right] \tag{6.6.6}$$

$$\text{or}\left\{\frac{\left\{\sum_{i=1}^{n} w_i p_i^{\alpha}\right\}}{\left\{\sum_{i=1}^{n} w_i^{\frac{1}{a}} p_i\right\}^{\alpha}}\right\} \leq \left[\frac{\sum_{i=1}^{n} w_i p_i^{\alpha}}{\sum_{i=1}^{n} w_i p_i^{\alpha} D^{l_i(\alpha-1)}}\right] \tag{6.6.7}$$

The equation (6.6.7) upon ignoring weights, gives

$$\left\{\sum_{i=1}^{n} p_i^{\alpha}\right\}^{\frac{1}{a}} \leq \left[\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i^{\alpha} D^{l_i(\alpha-1)}}\right]^{\frac{1}{\alpha}}$$

$$\text{or}\, 1-\left\{\sum_{i=1}^{n} p_i^{\alpha}\right\}^{\frac{1}{a}} \geq 1-\left[\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i^{\alpha} D^{l_i(\alpha-1)}}\right]^{\frac{1}{\alpha}}$$

Thus, for $\alpha < 1$, we have

$$\frac{\alpha}{\alpha-1}\left\{1-\left\{\sum_{i=1}^{n} p_i^{\alpha}\right\}^{\frac{1}{\alpha}}\right\} \leq \frac{\alpha}{\alpha-1}\left\{1-\left[\frac{\sum_{i=1}^{n} p_i^{\alpha}}{\sum_{i=1}^{n} p_i^{\alpha} D^{l_i(\alpha-1)}}\right]^{\frac{1}{\alpha}}\right\} \tag{6.6.8}$$

The equation (6.6.8) can similarly be proved for $\alpha > 1$.

This is again to be noted that the L.H.S. of (6.6.8) is Arimoto's [1] measure of entropy but R.H.S. is neither a mean codeword length nor a monotonic increasing function of mean codeword length. Proceeding on similar lines, many other existing measures of entropy can be shown as lower bounds for the functions which are neither mean codeword lengths themselves, nor the monotonic increasing functions of mean codeword lengths.

Concluding Remarks: Keeping in view a huge variety of information measures in the existing literature of information theory, one can develop many new codeword lengths with desirable properties. It is to be observed that Kraft's inequality plays an important role in proving a noiseless coding theorem and is uniquely determined by the condition for unique decipherability. It cannot be dependent on the probabilities or utilities and certainly cannot be modified in an arbitrary manner, motivated by the desire to prove noiseless coding theorem. If we modify this inequality, we shall get codes with a different structure other than satisfying the condition of unique decipherability. In this Chapter, we have made use of Kraft's inequality in original to prove some noiseless coding theorems and consequently, developed some new mean codeword lengths. We have used the concept of weighted information and verified some existing mean codeword lengths with this concept. We have also tried to introduce some new possible weighted measures of entropy via coding theorems. Keeping in mind the application areas, this work can be extended to generate a variety of new information measures and their correspondence with coding theory can be discussed.

REFERENCES

[1] Arimoto, S. (1971). Information theoretical consideration on estimation problems. *Information and Control, 19*, 181-194.

[2] Behara, M. and Chawla, J. S. (1974). Generalized γ-entropy. *Selecta Statistica Canadiana,* **2**, 15-38.

[3] Belis, M., & Guiasu, S. (1968). A quantitative-qualitative measure of information in cybernetic systems. *IEEE Transactions on Information Theory, 14*, 593-594.

[4] Campbell, L.L. (1965). A coding theorem and Renyi's entropy. *Information and control*, 8, 423-429.

[5] Guiasu, S., & Picard, C. F. (1971). Borne in ferictur de la longuerur utile de certains codes. *Comptes Rendus Mathematique Academic des Sciences Paris, 273*, 248-251.

[6] Gurdial, & Pessoa, F. (1977). On useful information of order α. *Journal of Combanatrics Information and System Sciences, 2*, 158-162.

[7] Kapur, J. N. (1967). Generalized entropy of order α and type β. *Mathematics Seminar, 4*, 79-84.

[8] Kapur, J. N. (1986). Four families of measures of entropy. *Indian Journal of Pure and Applied Mathematics, 17*, 429-449.

[9] Kapur, J. N. (1995). *Measures of information and their applications*. New York: Wiley Eastern.

[10] Kapur, J. N. (1998). *Entropy and coding*. New Delhi: Mathematical Sciences Trust Society.

[11] Kerridge, D. F. (1961). Inaccuracy and inference. *Journal of Royal Statistical Society Series B,* 23, 184-194**.**

[12] Khan, A.B. and Autar, R. (1984). On generalized useful information for incomplete distribution. *Journal of Combanatrics Information and System Sciences* **14**: 187-191.

[13] Kraft, L. G. (1949). *A device for quantizing grouping and coding amplitude modulated pulses*", M.S. Thesis, Electrical Engineering Department, MIT.

[14] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*, 79-86.

[15] Kumar, S., & Kumar, R. (2011). Some noiseless coding theorem connected with Havrda and Charvat and Tsallis's entropy. *Kragujevac Journal of Mathematics, 35*(1), 111-117.

[16] Leung-Yan-Cheong, S.K. and Cover, T.M. (1978): Some equivalences between Shannon entropy and Kolmogrov complexity. *IEEE Transactions on Information Theory*, 24, 331-338.

[17] Longo, G. (1972). *Quantitative-qualitative measures of information*. New York: Springer-Verlag.

[18] Parkash, O., & Kakkar, P. (2012). Development of two new mean codeword lengths. *Information Sciences, 207*, 90-97.

[19] Parkash, O., & Priyanka (2011). Contribution of Kraft's inequality to coding theory. *Pacific Asian Journal of Mathematics, 5*(1), 45-53.

[20] Ramamoorthy, A. (2011). Minimum cost distributed source coding over a network. *IEEE Transactions on Information Theory, 57*(1), 461-475.

[21] Renyi, A. (1961). On measures of entropy and information. *Proceedings 4th Berkeley Symposium on Mathematical Statistics and Probability* (pp.547-561).

[22] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-659.

[23] Sharma, B. D., & Taneja, I. J. (1975). Entropies of type (α,β) and other generalized measures of information theory. *Metrika, 22*, 202-215.

CHAPTER-VII

A STUDY OF MAXIMUM ENTROPY PRINCIPLE FOR DISCRETE PROBABILITY DISTRIBUTIONS

ABSTRACT

The present chapter deals with the detailed study of maximum entropy principle applied to discrete probability distributions and for this study, we have made use of Lagrange's method of undetermined multipliers to maximize different measures of entropy under a set of one or more specified constraints. To obtain maximum entropy probability distribution (MEPD), the method has been explained with the help of numerical examples. We have also described a method for approximating a given probability distribution by using maximum entropy principle and deduced some interesting and desirable results.

Keywords: ***Entropy, Uncertainty, Directed divergence, Probability distribution, Non-linear programming, Maximum entropy principle.***

7.1 INTRODUCTION

When the classical entropy theory is combined with mathematical optimization, the resulting entropy optimization models are generated which can prove to be very useful and find successful applications in areas such as pattern recognition, statistical inference, queuing theory, statistical mechanics, transportation planning, urban and regional planning, input-output analysis, portfolio investment, and linear and nonlinear programming. We know out of all probability distributions, the uncertainty is maximum when the distribution is uniform, that is, it is supposed to be the uniform distribution which contains the largest amount of uncertainty, but this is just Laplace's principle of insufficient reasoning, according to which if there is no reason to discriminate between two or several events, the best strategy is to consider them equally likely.

Jaynes [6] provided a very natural criterion of choice by introducing the principle of maximum entropy and stressed that from the set of all probability distributions compatible with one or several mean values of one or several random variables, choose the only distribution that maximizes Shannon's [13] measure of entropy. Jaynes [6] also argued that the entropy in statistical mechanics, and the entropy in information theory, are principally the same and consequently, statistical mechanics should be seen just as a particular application of a general tool of information theory. The principle of maximum entropy is a method for analyzing available qualitative information in order to determine a unique probability distribution and it states that the least biased distribution that encodes certain given information is the one which maximizes the information entropy.

In real life situations, we are very frequently concerned with the problems of constrained optimization, for example, a company may wish to maximize its production or profit under certain constraints because of limited resources at its disposal. A very natural question then arises that why do we maximize entropy? The obvious answer is that we like reduction in uncertainty by obtaining more and more information. We explain this concept as follows:

Suppose that, for the probability distribution $P=(p_1, p_2, ..., p_n)$, we have no information other than

$$P=\left\{p_1, p_2, ..., p_n; p_1 \ge 0, p_2 \ge 0, ..., p_n \ge 0, \sum_{i=1}^{n} p_i = 1\right\} \tag{7.1.1}$$

Then there may be infinite number of probability distributions consistent with (7.1.1). Out of these, one distribution, viz., the uniform distribution $(\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$ has the maximum entropy, viz., $\log n$. There may also be n degenerate distributions, viz., $(1, 0, ..., 0)$, $(0, 1, ..., 0)$, …, $(0, 0, ..., 1)$ each of which has the minimum entropy, viz., zero. Now, suppose we are given an additional piece of information in the form of following constraint:

$$g_{11}p_1 + g_{12}p_2 + ... + g_{1n}p_n = a_1 \tag{7.1.2}$$

The probability distributions which were consistent with (7.1.1) may not be consistent with (7.1.2). Consequently, maximum value of entropy of all distributions consistent with both (7.1.1) and (7.1.2) will be $\le \log n$ and the minimum value of this entropy will be ≥ 0.

Similarly, each additional piece of information in the form of a relation among $p_1, p_2, ..., p_n$ will decrease and at least will not increase the maximum entropy $S_{\max}$ and similarly, each additional piece of information will increase and at least will not decrease the minimum entropy $S_{\min}$. Here $S_{\max}$ is a monotonic decreasing function and $S_{\min}$ is a monotonic increasing function of the number of constraints imposed in succession. After n independent and consistent constraints are imposed, we are likely to have a unique probability distribution, and at this stage $S_{\max}$ and $S_{\min}$ will coincide as shown in the following Fig.-7.1.1.

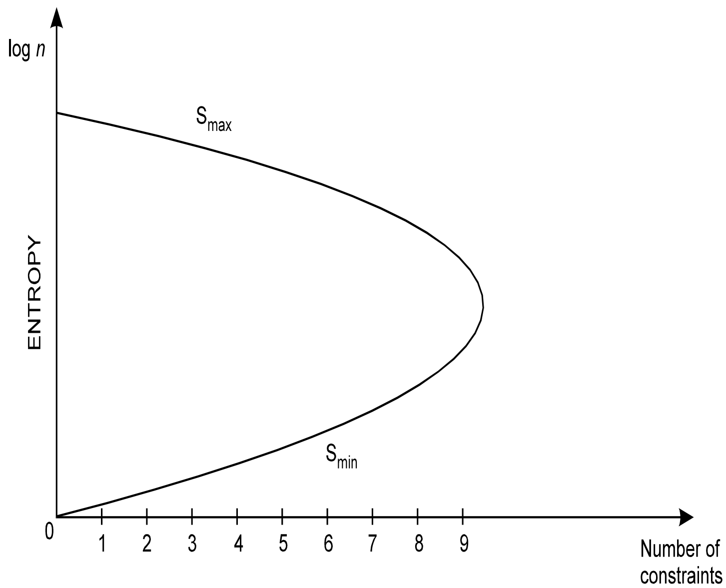


Fig.-7.1.1 Representation of $S_{\max}$ and $S_{\min}$

Thus, the relation between uncertainty as given by a specific entropy measure and information can be expressed as follows:

With any given amount of information, expressed in the form of constraints on the p_i, we do not have a unique value of uncertainty, rather, we have a range of values of uncertainty. At any stage, one cannot definitely say that one's uncertainty is precisely so much it lies between two given uncertainty limits represented by $S_{\max}$ and $S_{\min}$. Any additional piece of information does not reduce uncertainty; it only reduces this range of uncertainty. As we proceed with more and more independent pieces of information, the range of uncertainty goes on becoming narrower and narrower till, hopefully, it becomes zero. At this stage only, we can speak of a unique value of uncertainty. The final probability distribution may be non-degenerate or degenerate, and the final value of uncertainty may be non-zero or zero. If the final probability distribution is degenerate and if the final entropy is zero, the phenomenon concerned is deterministic, otherwise, it is stochastic.

For example, in the problem of identifying a number, if we have 512 integers from 1 to 512, the maximum entropy is $\log 512$ while the minimum entropy is zero. After we are given the information whether the number lies between 1 and 256 or between 257 and 512, the maximum entropy is reduced to $\log 256$ and the minimum entropy remains zero. In this process the maximum entropy continues to decrease to the values $\log 128, \log 64, \log 32, \log 16, \log 8, \log 4, \log 2$ *and* $\log 1 = 0$ and the minimum entropy continues to remain at zero and such a type of phenomenon is deterministic in nature. On the other hand, let us suppose that we have a die and we do not even know the number of faces the die has. In this case, we have a great deal of uncertainty and we may have many probability distributions of the form of equation (7.1.1).

Further, we are given the information that the die has six faces, and then by this information, the uncertainty is reduced. We are now only limited to probability distribution $\{p_1, p_2, ..., p_6\}$, with the constraint (7.1.1) being satisfied. If, in addition, we are also given the mean number of points on the die, that is, we are given that

$$1p_1 + 2p_2 + 3p_3 + 4p_4 + 5p_5 + 6p_6 = 4.5, \qquad (7.1.3)$$

our choice of distribution is now restricted to those satisfying (7.1.1) and (7.1.3), and the uncertainty has been further reduced.

Further, if we are given more information that

$$1^2 p_1 + 2^2 p_2 + 3^2 p_3 + 4^2 p_4 + 5^2 p_5 + 6^2 p_6 = 15, \qquad (7.1.4)$$

then the choice of distribution is further restricted and uncertainty is further reduced. We may go on getting more and more information and, accordingly, the amount of uncertainty goes on decreasing. If we get in three stages three more independent linear constraints consistent with (7.1.1), (7.1.3) and (7.1.4), we may get a unique set of values of p_1 through p_6 and the uncertainty about these values is completely removed. The phenomenon concerned is stochastic in nature. At any stage, we may have infinity of probability distributions consistent with the given constraints, say,

$$\sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i g_{ri} = a_r, \ r = 1,2,...,m, \ m+1 < n, \ p_i \geq 0 \qquad (7.1.5)$$

Out of these distributions, one has maximum uncertainty $S_{\max}$ and the uncertainty S of any other distribution is less than $S_{\max}$. Now, since the uncertainty can be reduced by the given additional information, the use of any distribution other than the maximum uncertainty distribution implies the use of some information in addition to that given by equation (7.1.5). Kapur and Kesavan [8] remarked that one should use the given information and should not use any information which is not given. According to the first part of this statement, we should use only those distributions which are consistent with the constraints (7.1.5), but there may be infinity of such distributions. The second part of the above statement now enables us to choose one out of these distributions and we choose the distribution having the maximum uncertainty, $S_{\max}$. Thus, the above statement of the principle of maximum uncertainty or maximum entropy can be stated as follows:

"Out of all probability distributions consistent with a given set of constraints, choose the only distribution that has maximum uncertainty".

Jaynes [6] MaxEnt aims at maximizing uncertainty subject to given constraints and it restricts its use only to Shannon's [13] measure of uncertainty. But, the concept of uncertainty is too deep and complex and as such it may be difficult to be measured by a single measure under all conditions. Thus, there is a definite need for generalized measures of entropy just to extend the scope of their applications for the study of different optimization principles. On the other hand, Kullback's [9] principle of minimum cross-entropy aims to choose that probability distribution, out of all those that satisfy the given constraints, which is closest to a given a priori probability distribution. However, the principle confines itself to only one measure of directed divergence, given by Kullback and Leibler [10]. Again, directed divergence is too deep a concept to be represented by one measure only and in the literature of information theory, a large number of measures of directed divergence are available to measure distances between two probability distributions, some of these find tremendous applications in various fields of science and engineering and hence can be applied for the study of different optimization principles.

Jaynes [6] maximum entropy principle and Kullback's [9] MinxEnt have found wide applications, yet their scope can be considerably enhanced by using the generalized measures of entropy and cross entropy. According to these principles, we choose the probability distribution P, out of all the distributions satisfying the given constraints, which minimizes a specified generalized measure of divergence. The rationale for this arises from the following entropy optimization postulate:

"Every probability distribution, theoretical or observed, is an entropy optimization distribution, that is, it can be obtained by either maximizing an appropriate entropy measure or by minimizing a cross entropy measure with respect to an appropriate priori distribution, subject to its satisfying appropriate constraints".

From the above statement, we conclude an important inference that entropy optimization problems and entropy optimization principles are alike. The entropy optimization postulate establishes a link among the measure of entropy or cross entropy, set of moment constraints, a priori distribution and posterior distribution, which leads to the generalization of entropy optimization principles. It can also be stated that if given any three of the above four entities, the fourth should be chosen such that either the entropy of the probability distribution P is maximum or the cross-entropy of P relative to Q is minimum.

It is to be observed that the generalized measures of information contain one or more non-stochastic parameters and thus, represent a family of measures which includes Shannon's [13] and Kullback and Leibler's [10] measures as particular cases. The physical meaning and interpretations for these parameters depend on the data generating phenomenon as well as on the knowledge and experience of the subject expert. Reliability theory, marketing, measurement of risk in portfolio analysis, quality control, log linear models etc. are the areas where generalized optimization principles have successfully been applied. Similarly, various other mathematical models can better be explained if we use generalized parametric measures of entropy and divergence.

Herniter [4] used Shannon's [13] measure of entropy in studying the market behavior and obtained interesting results whereas some work related with the study of maximum entropy verses risk factor has recently been presented by Topsoe [14]. A large number of applications of maximum entropy principle in science and engineering have been provided by Kapur and Kesavan [8], Kapur, Baciu and Kesavan [7], whereas some optimization principles in the field of agriculture, that is, towards the study of crop area have been discussed and investigated by Hooda and Kapur [5]. Guiasu and Shenitzer [2] have well explained the principle of maximum entropy under a set of constraints. All these principles deal with the discrete or continuous probability distributions only but we have tried to extend the applications of the maximum entropy principle for the fuzzy distributions also and for this purpose, we have used the concept of weighted fuzzy entropy and provided the applications of optimization principle for unequal constraints, the findings of which have been investigated by Parkash, Sharma and Mahajan [12].

7.2 MAXIMUM ENTROPY PRINCIPLE FOR DISCRETE PROBABILITY DISTRIBUTIONS

It is usually observed that in many practical situations, often only partial information is available about the probability distributions and thus, the maximum entropy principle can be applied. The partial information may be in the form of some moments and in such cases, we use only the available information and do not use any unknown information and consequently, apply MaxEnt to estimate the corresponding maximum entropy probability distribution (MEPD). Here we consider a discrete random variable taking a finite number of values $1, 2, 3, \ldots, n$ and use Lagrange's method of undetermined multipliers to maximize different measures of entropy under a set of one or more specified constraints.

For this purpose, we consider the following measures:

I. Burg's [1] measure of entropy, given by

$$B(P) = \sum_{i=1}^{n} \log p_i \tag{7.2.1}$$

To elaborate the MaxEnt principle for this entropy measure, we consider the following cases:

Case-I. When no constraint is given except the natural constraint

In this case our problem becomes as follows:

Maximize (7.2.1) subject to the natural constraint $\sum_{i=1}^{n} p_i = 1$

The corresponding Lagrangian is given by

$$L = \sum_{i=1}^{n} \log p_i - \lambda \left\{ \sum_{i=1}^{n} p_i - 1 \right\}$$

Now $\frac{\partial L}{\partial p_i} = 0$ gives $p_i = \frac{1}{\lambda}\ \forall i$.

Thus, $p_1 = p_2 = p_3 = ... = p_n$

Since $\sum_{i=1}^{n} p_i = 1$, we get $p_i = \frac{1}{n}\ \forall i$ which is a uniform distribution.

Case-II. When arithmetic mean is prescribed

Let us now suppose that we have the knowledge of the arithmetic mean $m(1 < m < n)$ of the distribution. In this case our problem becomes as follows:

Maximize (7.2.1) under the following conditions:

$$\sum_{i=1}^{n} p_i = 1 \tag{7.2.2}$$

and

$$\sum_{i=1}^{n} i p_i = m \tag{7.2.3}$$

The corresponding Lagrangian is given by

$$L = \sum_{i=1}^{n} \log p_i - \lambda \left\{ \sum_{i=1}^{n} p_i - 1 \right\} - \mu \left\{ \sum_{i=1}^{n} i p_i - m \right\}$$

Now $\frac{\partial L}{\partial p_i} = 0$ gives

$$p_i = \frac{1}{\lambda + i\mu} \tag{7.2.4}$$

Applying condition (7.2.2), we have

$$\sum_{i=1}^{n} \frac{1}{\lambda + i\mu} = 1 \tag{7.2.5}$$

Also by condition (7.2.3), we have

$$\sum_{i=1}^{n} \frac{i}{\lambda + i\mu} = m$$

or $\frac{1}{\mu}\left[\sum_{i=1}^{n}\left\{1 - \frac{\lambda}{\lambda + i\mu}\right\}\right] = m$

or $n - \sum_{i=1}^{n} \frac{\lambda}{\lambda + i\mu} = \mu m$

Applying equation (7.2.5), we get

$$\lambda = n - \mu m \tag{7.2.6}$$

Equation (7.2.6) gives λ in terms of μ.

Thus

$$\lambda + i\mu = n - \mu(m - i) \tag{7.2.7}$$

Consequently, $p_i = \frac{1}{n - \mu(m-i)}$ (7.2.8)

where μ has to be determined from the following equation:

$$\sum_{i=1}^{n} \frac{1}{n - \mu(m-i)} = 1 \tag{7.2.9}$$

which gives μ for known values of m and n.

Now, the maximum value of the Burg's [1] measure of entropy under the constraints (7.2.2) and (7.2.3) is given by

$$\left[B(P)\right]_{\max} = -\sum_{i=1}^{n} \log\left\{n - \mu(m-i)\right\} \tag{7.2.10}$$

Differentiating equation (7.2.10) with respect to m, we get

$$\frac{d}{dm}\left[B(P)\right]_{\max} = \sum_{i=1}^{n} \frac{1}{n-\mu\left(m-i\right)}\left\{\mu+(m-i)\frac{d\mu}{dm}\right\}$$

$$= \sum_{i=1}^{n} p_i\left[\mu+\left(m-i\right)\frac{d\mu}{dm}\right]$$

$$= \sum_{i=1}^{n} \mu p_i + \frac{d\mu}{dm}\sum_{i=1}^{n} p_i\left(m-i\right) = \mu$$

Thus, we have

$$\frac{d^2\left[B(P)\right]_{\max}}{dm^2} = \frac{d\mu}{dm} \tag{7.2.11}$$

Now, from equation (7.2.8), we have

$$\left\{n-\mu\left(m-i\right)\right\} p_i^2 = p_i \tag{7.2.12}$$

Taking summation both sides of equation (7.2.12) and differentiating with respect to m, we get

$$\frac{d\mu}{dm} = \frac{\mu\sum_{i=1}^{n} p_i^2}{\sum_{i=1}^{n} p_i^2\left\{i-m\right\}} \tag{7.2.13}$$

Using equation (7.2.13), equation (7.2.11) gives

$$\frac{d^2\left[B(P)\right]_{\max}}{dm^2} = \frac{\mu\sum_{i=1}^{n} p_i^2}{\sum_{i=1}^{n} p_i^2\left\{i-m\right\}}$$

Now since $\left[B(P)\right]_{\max}$ has to be a concave function, we must have

$\frac{d^2\left[B(P)\right]_{\max}}{dm^2} < 0$ which is possible only if $\frac{d\mu}{dm} < 0$, that is, if either

(i) $\mu > 0, \sum_{i=1}^{n} p_i^2\left\{i-m\right\} < 0$

or

(ii) $\mu < 0, \sum_{i=1}^{n} p_i^2\left\{i-m\right\} > 0$

Now, from equation (7.2.4), the probabilities are decreasing if $\mu > 0$ and increasing if $\mu < 0$.

If $\mu = 0$, then $p_1 = p_2 = p_3 = ... = p_n$ and using the condition (7.2.2), we get $p_i = \frac{1}{n} \ \forall i$

Thus, the condition (7.2.3) implies that $m = \frac{n+1}{2}$

Hence, we have the following observations:

(i) If $\mu > 0,$ the probabilities are decreasing and $m < \frac{n+1}{2}$

(ii) If $\mu < 0,$ the probabilities are increasing and $m > \frac{n+1}{2}$

Thus, we consider the following cases:

Case-I: When $m < \frac{n+1}{2}$, we have

$$\sum_{i=1}^{n} p_i^2 \{i-m\} = \sum_{i=1}^{n} ip_i^2 - \left\{\sum_{i=1}^{n} p_i^2\right\}\left\{\sum_{i=1}^{n} ip_i\right\}$$

$$= \left\{\sum_{i=1}^{n} p_i^2\right\}\left[\frac{\sum_{i=1}^{n} ip_i^2}{\sum_{i=1}^{n} p_i^2} - \frac{\sum_{i=1}^{n} ip_i}{\sum_{i=1}^{n} p_i}\right] \tag{7.2.14}$$

Now, the R.H.S. of equation (7.2.14) gives the difference between two terms representing weighted averages of the numbers $1, 2, 3, ..., n$. In the first term, the weights are proportional to $p_1^2, p_2^2, p_3^2, ..., p_n^2$ whereas in the second term, the weights are proportional to $p_1, p_2, p_3, ..., p_n$. Now, since in this case, the probabilities are decreasing, the weights in the first case decrease faster relative to weights in the second case. This shows that the first term in the square bracket of equation (7.2.14) will be less than that of second term. Thus, we must have

$$\sum_{i=1}^{n} p_i^2 \{i-m\} < 0$$

Case-II: When $m > \frac{n+1}{2}$

Proceeding on similar lines as discussed in case-I, we must have

$$\sum_{i=1}^{n} p_i^2 \{i-m\} > 0$$

Thus, in both the cases, we have proved that $\left[B(P)\right]_{\max}$ is a concave function of m. Hence, in each case, the maximizing probabilities can be calculated and the MEPD can be obtained.

The above method has been illustrated with the help of following numerical:

Numerical Example

We maximize the entropy (7.2.1) under the set of constraints (7.2.2) and (7.2.3) for $n=8$ and for different values of m.

For $m=1.5$ and $n=8$, equation (7.2.9) gives the following expression:

$$\frac{1}{8-0.5\mu}+\frac{1}{8+0.5\mu}+\frac{1}{8+1.5\mu}+\frac{1}{8+2.5\mu}+\frac{1}{8+3.5\mu}+\frac{1}{8+4.5\mu}+\frac{1}{8+5.5\mu}+\frac{1}{8+6.5\mu}=1 \tag{7.2.15}$$

Case-I: When $m<\dfrac{n+1}{2}$

Upon simplification, equation (7.2.15) gives the best possible value $\mu=13.5554$

Thus, equation (7.2.6) gives $\lambda=-12.3331$

With these values of λ and μ, we get the following set of probabilities:

$p_1=0.81813$, $p_2=0.06767$,

$p_3=0.03529$, $p_4=0.02387$,

$p_5=0.01804$, $p_6=0.01449$,

$p_7=0.01211$, $p_8=0.01040$.

Obviously,

$$\sum_{i=1}^{8} p_i=1$$

The above procedure is repeated for different values of m when $n=8$. This is to be noted that in certain cases, we got negative probabilities. To tackle this problem, we ignored these probabilities and reformulated the above problem for the remaining probabilities and again solved it by using Lagrange's method. The results of the computations are shown in Table-7.2.1.

Table-7.2.1

m	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
1.5	0.81813	0.06767	0.03529	0.02387	0.01804	0.01449	0.01211	0.01040
2.5	0.48824	0.16622	0.10016	0.07167	0.05580	0.04569	0.03868	0.03353
3.5	0.25011	0.17861	0.13890	0.11363	0.09614	0.08332	0.07352	0.06578
4.5	0.00000	0.20933	0.17648	0.15255	0.13433	0.11999	0.10843	0.09889
5.5	0.00000	0.00000	0.00000	0.32811	0.22992	0.17697	0.14384	0.12116
6.5	0.00000	0.00000	0.00000	0.00000	0.00000	0.63715	0.22571	0.13715
7.5	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.50000	0.50000

Note: The Table-7.2.1 clearly shows that when $\mu > 0$, the probabilities are decreasing for the values of $m < \frac{n+1}{2} = 4.5$. Thus, our theoretical results provided above coincide with the experimental results.

Case-II: When $m > \frac{n+1}{2}$

Proceeding as above, we have obtained the set of probabilities as shown in Table-7.2.2.

Table-7.2.2

m	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
1.5	0.50000	0.50000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2.5	0.13715	0.22571	0.63715	0.00000	0.00000	0.00000	0.00000	0.00000
3.5	0.12116	0.14384	0.17697	0.22992	0.32811	0.00000	0.00000	0.00000
4.5	0.09889	0.10843	0.11999	0.13433	0.15255	0.17648	0.20933	0.00000
5.5	0.06578	0.07352	0.08332	0.09614	0.11363	0.13890	0.17861	0.25011
6.5	0.03353	0.03868	0.04569	0.05580	0.07167	0.10016	0.16622	0.48824
7.5	0.01040	0.01211	0.01449	0.01804	0.02387	0.03529	0.06767	0.81813

Note: The above Table-7.2.2 clearly shows that when $\mu < 0$, the probabilities are increasing for the values of $m > \frac{n+1}{2} = 4.5$. Thus, our theoretical results provided above, find total compatibility with the experimental findings.

Case-III. When geometric mean is prescribed

Let us now suppose that we have the knowledge of the geometric mean $g(1 < g < n)$ of the distribution. In this case our problem becomes as follows:

Maximize (7.2.1) under the natural constraint (7.2.2) and

$$\sum_{i=1}^{n} p_i \log i = \log g \tag{7.2.16}$$

The corresponding Lagrangian is given by

$$L = \sum_{i=1}^{n} \log p_i - \lambda\left\{\sum_{i=1}^{n} p_i - 1\right\} - \mu\left\{\sum_{i=1}^{n} p_i \log i - \log g\right\} \tag{7.2.17}$$

Now $\frac{\partial L}{\partial p_i} = 0$ gives

$$p_i = \frac{1}{\lambda + \mu \log i} \tag{7.2.18}$$

Applying conditions (7.2.2) and (7.2.16) and proceeding as above, we get

$$p_i = \frac{1}{n + \mu\left(\log \frac{i}{g}\right)} \tag{7.2.19}$$

where μ has to be determined from the following equation:

$$\sum_{i=1}^{n} \frac{1}{n + \mu\left(\log \frac{i}{g}\right)} = 1 \tag{7.2.20}$$

which gives μ for known values of g and n.

$$\text{Now}\left[B(P)\right]_{\max} = -\sum_{i=1}^{n} \log\left\{n + \mu\left(\log \frac{i}{g}\right)\right\} \tag{7.2.21}$$

Differentiating equation (7.2.21) with respect to g, we get

$$\frac{d}{dg}\left[B(P)\right]_{\max} = \mu$$

Thus, we have

$$\frac{d^2\left[B(P)\right]_{\max}}{dg^2}=\frac{d\mu}{dg} \tag{7.2.22}$$

From equation (7.2.19), we have

$$\frac{d\mu}{dg}=\frac{\mu\sum_{i=1}^{n}p_i^2}{g\sum_{i=1}^{n}p_i^2\left\{\log i-\log g\right\}} \tag{7.2.23}$$

This further gives

$$\frac{d^2\left[B(P)\right]_{\max}}{dg^2}=\frac{\mu\sum_{i=1}^{n}p_i^2}{g\sum_{i=1}^{n}p_i^2\left\{\log i-\log g\right\}} \tag{7.2.24}$$

Now since $\left[B(P)\right]_{\max}$ has to be a concave function, we must have

$\frac{d^2\left[B(P)\right]_{\max}}{dg^2}<0$ which is possible only if $\frac{d\mu}{dg}<0$, that is, if either

(i) $\mu>0,\ g\sum_{i=1}^{n}p_i^2\left\{\log i-\log g\right\}<0$

or (ii) $\mu<0,\ g\sum_{i=1}^{n}p_i^2\left\{\log i-\log g\right\}>0$

Proceeding as discussed in the above article, we observe the following:

(i) If $\mu>0,$ the probabilities are decreasing and $\log g<\frac{\log\lfloor n}{n}$

(ii) If $\mu<0,$ the probabilities are increasing and $\log g>\frac{\log\lfloor n}{n}$

Case-I: When $\log g<\frac{\log\lfloor n}{n}$, we have

$$g\sum_{i=1}^{n}p_i^2\left\{\log i-\log g\right\}=g\left\{\sum_{i=1}^{n}p_i^2\right\}\left[\frac{\sum_{i=1}^{n}p_i^2\log i}{\sum_{i=1}^{n}p_i^2}-\frac{\sum_{i=1}^{n}p_i\log i}{\sum_{i=1}^{n}p_i}\right] \tag{7.2.25}$$

which shows that

$$g\sum_{i=1}^{n}p_i^2\left\{\log i-\log g\right\}<0$$

Case-II: When $\log g > \frac{\log\lfloor n}{n}$, we have

$$g\sum_{i=1}^{n} p_i^2 \left\{\log i - \log g\right\} > 0$$

Thus, in both the cases, we have proved that $\left[B(P)\right]_{\max}$ is a concave function of g. Hence, in each case, the maximizing probabilities can be calculated and the MEPD can be obtained.
The above method has been illustrated with the help of following numerical:

Numerical Example

We maximize the entropy (7.2.1) under the set of constraints (7.2.2) and (7.2.16) for $n = 8$ and for different values of g.

For $g = 1.5$ and $n = 8$, equation (7.2.20) gives the following expression:

$$\frac{1}{8+\mu\log\frac{1}{1.5}} + \frac{1}{8+\mu\log\frac{2}{1.5}} + \frac{1}{8+\mu\log\frac{3}{1.5}} + \frac{1}{8+\mu\log\frac{4}{1.5}} + \frac{1}{8+\mu\log\frac{5}{1.5}}$$
$$+\frac{1}{8+\mu\log\frac{6}{1.5}} + \frac{1}{8+\mu\log\frac{7}{1.5}} + \frac{1}{8+\mu\log\frac{8}{1.5}} = 1 \qquad (7.2.26)$$

Case-I: When $\log g < \frac{\log\lfloor n}{n}$

Upon simplification, equation (7.2.26) gives the best possible value $\mu = 11.231$

The corresponding value of λ is given by $\lambda = 1.430286$

With these values of λ and μ, we get the following set of probabilities:

$p_1 = 0.69916$, $p_2 = 0.07898$, $p_3 = 0.05200$,

$p_4 = 0.04185$, $p_5 = 0.03635$, $p_6 = 0.03283$,

$p_7 = 0.03034$, $p_8 = 0.02847$

Obviously, we have

$$\sum_{i=1}^{8} p_i = 0.99999 \cong 1$$

The above procedure is repeated for different values of g when $n = 8$ and the results of the computations are shown in Table-7.2.3.

Table-7.2.3

g	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
1.5	0.69916	0.07898	0.05200	0.04185	0.03635	0.03283	0.03034	0.02847
2.5	0.34247	0.14787	0.11098	0.09429	0.08444	0.07780	0.07295	0.06921
3.5	0.15510	0.13686	0.12806	0.12247	0.11846	0.11537	0.11288	0.11081
4.5	0.00000	0.14910	0.14591	0.14373	0.14208	0.14077	0.13967	0.13874
5.5	0.00000	0.00000	0.00000	0.28352	0.21934	0.18510	0.16352	0.14852
6.5	0.00000	0.00000	0.00000	0.00000	0.00000	0.61299	0.23435	0.15266
7.5	0.00000	0.00000	0.00000	0.00000	0.00000	0.0000	0.50000	0.50000

Case-II: When $\log g > \dfrac{\log\lfloor n}{n}$

Proceeding as above, we have obtained the set of probabilities as shown in Table-7.2.4.

Table-7.2.4

g	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
1.5	0.41504	0.58496	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2.5	0.16968	0.22416	0.27599	0.33016	0.00000	0.00000	0.00000	0.00000
3.5	0.12831	0.13597	0.14089	0.14460	0.14762	0.15018	0.15242	0.00000
4.5	0.06973	0.08758	0.10300	0.11770	0.13235	0.14733	0.16293	0.17938
5.5	0.03338	0.04755	0.06326	0.08263	0.10837	0.14537	0.20435	0.31510
6.5	0.01568	0.02318	0.03221	0.04450	0.06321	0.09628	0.17268	0.55226
7.5	0.00453	0.00677	0.00954	0.01343	0.01966	0.03166	0.06539	0.84903

Note: The Table-7.2.3 and Table-7.2.3clearly show that when $\mu < 0$, the probabilities are decreasing for the values of $\log g < \frac{\log \lfloor n}{n} = 3.76435$ and when $\mu > 0$, the probabilities are increasing for the values of $\log g > 3.76435$. These results are desirable and find total compatibility with the theory.

II. Onicescu's [11] measure of entropy, given by

$$H(P) = \sum_{i=1}^{n} p_i^2 \tag{7.2.27}$$

To elaborate this MaxEnt principle, we consider the following cases:

Case-I. When no constraint is given except the natural constraint

In this case our problem becomes as follows:

Maximize (7.2.27) subject to the natural constraint $\sum_{i=1}^{n} p_i = 1$

The corresponding Lagrangian is given by

$$L = \sum_{i=1}^{n} p_i^2 - \lambda \left\{ \sum_{i=1}^{n} p_i - 1 \right\}$$

Now $\frac{\partial L}{\partial p_i} = 0$ gives $p_i = \frac{\lambda}{2} \; \forall i$

Thus, $p_1 = p_2 = p_3 = \ldots = p_n$

Since $\sum_{i=1}^{n} p_i = 1$, we get $p_i = \frac{1}{n} \; \forall i$ which is a uniform distribution.

Case-II. When arithmetic mean is prescribed

Let us now suppose that we have the knowledge of the arithmetic mean $m (1 \le m \le n)$ of the distribution. In this case our problem becomes as follows:

Maximize (7.2.27) subject to the constraints given in equations (7.2.2) and (7.2.3).

The corresponding Lagrangian is given by

$$L = \sum_{i=1}^{n} p_i^2 - 2\lambda \left\{ \sum_{i=1}^{n} p_i - 1 \right\} - 2\mu \left\{ \sum_{i=1}^{n} i p_i - m \right\}$$

Thus $\frac{\partial L}{\partial p_i} = 0$ gives

$$p_i = \lambda + i\mu \tag{7.2.28}$$

Next, applying condition (7.2.2), we get

$$(\lambda+\mu)+(\lambda+2\mu)+...+(\lambda+n\mu)=1$$

This gives the following expression for λ:

$$\lambda=\frac{1}{n}\left[1-\frac{n(n+1)}{2}\mu\right] \tag{7.2.29}$$

Applying condition (7.2.3), we have

$$(\lambda+\mu)+2(\lambda+2\mu)+3(\lambda+3\mu)+...+n(\lambda+n\mu)=m$$

or $\lambda(1+2+3+...+n)+\mu\left(1^2+2^2+3^2+...+n^2\right)=m$

or $$\lambda\frac{n(n+1)}{2}+\mu\frac{n(n+1)(2n+1)}{6}=m \tag{7.2.30}$$

Using (7.2.29) in (7.2.30), we get

$$\mu=6\left\{\frac{2m-n-1}{n(n+1)(7n+5)}\right\} \tag{7.2.31}$$

With this value of μ, λ can be calculated from equation (7.2.29) and consequently, probabilities can be obtained from equation (7.2.28).

The above method has been illustrated with the help of following numerical:

Numerical Example

We maximize the entropy (7.2.27) under the set of constraints (7.2.2) and (7.2.3) for $n=8$ and for different values of m.

For $m=1$, we have μ = -0.009563, λ = 0.168033

With these values of λ and μ, we get the following set of probabilities:

$p_1=0.158470$, $p_2=0.148907$, $p_3=0.139344$,

$p_4=0.129781$, $p_5=0.120219$, $p_6=0.110656$,

$p_7=0.101093$, $p_8=0.091530$.

Obviously, $\sum_{i=1}^{8} p_i=1$

The above procedure is repeated for different values of the mean m for the given number of observations, that is, $n=8$. The various results of the computations so obtained have been shown in the following Table-7.2.5:

Table-7.2.5

m	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
1.0	0.158470	0.148907	0.139344	0.129781	0.120219	0.110656	0.101093	0.091530
1.5	0.153689	0.145492	0.137295	0.129098	0.120902	0.112705	0.104508	0.096311
2.0	0.148907	0.142077	0.135246	0.128415	0.121585	0.114754	0.107923	0.101093
2.5	0.144126	0.138661	0.133197	0.127732	0.122268	0.116803	0.111339	0.105874
3.0	0.139344	0.135246	0.131148	0.127049	0.122951	0.118852	0.114754	0.110656
3.5	0.134563	0.131831	0.129098	0.126366	0.123634	0.120902	0.118169	0.115437
4.0	0.129781	0.128415	0.127049	0.125683	0.124317	0.122951	0.121585	0.120219
4.5	0.125000	0.125000	0.125000	0.125000	0.125000	0.125000	0.125000	0.125000
5.0	0.120219	0.121585	0.122951	0.124317	0.125683	0.127049	0.128415	0.129781
5.5	0.115437	0.118169	0.120902	0.123634	0.126366	0.129098	0.131831	0.134563
6.0	0.110656	0.114754	0.118852	0.122951	0.127049	0.131148	0.135246	0.139344
6.5	0.105874	0.111339	0.116803	0.122268	0.127732	0.133197	0.138661	0.144126
7.0	0.101093	0.107923	0.114754	0.121585	0.128415	0.135246	0.142077	0.148907
7.5	0.096311	0.104508	0.112705	0.120902	0.129098	0.137295	0.145492	0.153689
8.0	0.091530	0.101093	0.110656	0.120219	0.129781	0.139344	0.148907	0.158470

Note: In this case, we observe that as m approaches to $\frac{n+1}{2}$, the probability distribution approaches uniform distribution and at $m = \frac{n+1}{2}$, we get exact uniform distribution.

Case-III. When geometric mean is prescribed

Let us now suppose that we have the knowledge of the geometric mean $g(1 < g < n)$ of the distribution.

In this case our problem becomes as follows:

Maximize (7.2.27) under the constraints (7.2.2) and (7.2.16).Upon simplification, we get

$$p_i = \lambda + \mu \log i \tag{7.2.32}$$

where $\lambda = \dfrac{1-\mu\log\lfloor n}{n}$ and $\mu = \dfrac{n\log g - \log\lfloor n}{n\left\{(\log 1)^2 + (\log 2)^2 + (\log 3)^2 + ...(\log n)^2\right\} - \left\{\log\lfloor n\right\}^2}$

Numerical Example

We maximize (7.2.27) under the constraints (7.2.2) and (7.2.16) for $n = 8$ and for different values of g.

For $g = 1.5$ and $n = 8$, we get $\mu = -0.24872$ and $\lambda = 0.535097$. With these values, we get the following probability distribution:

$p_1 = 0.535097$, $p_2 = 0.286374$, $p_3 = 0.14088$, $p_4 = 0.03765$,

$p_5 = 0.00000$, $p_6 = 0.00000$, $p_7 = 0.00000$, $p_8 = 0.00000$ and $\sum_{i=1}^{8} p_i = 1$

The above procedure is repeated for different values of g when $n = 8$ and the results of the computations are shown in Table-7.2.6.

Table-7.2.6

g	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
1.5	0.535097	0.286373	0.140880	0.037650	0.000000	0.000000	0.000000	0.000000
2.0	0.369970	0.240711	0.165099	0.111451	0.069839	0.035839	0.007092	0.000000
2.5	0.281651	0.199738	0.151822	0.117824	0.091454	0.069908	0.051691	0.035911
3.0	0.211869	0.166445	0.139874	0.121021	0.106398	0.09445	0.084348	0.075597
3.5	0.152868	0.138296	0.129772	0.123723	0.119032	0.115199	0.111958	0.109151
4.0	0.101760	0.113912	0.121021	0.126065	0.129977	0.133173	0.135876	0.138217
4.5	0.056679	0.092404	0.113302	0.128129	0.139630	0.149027	0.156972	0.163854
5.0	0.016353	0.073165	0.106398	0.129977	0.148266	0.163209	0.175844	0.186788
5.5	0.000000	0.035667	0.088553	0.126077	0.155182	0.178963	0.199070	0.216487
6.0	0.000000	0.000000	0.050704	0.110998	0.157765	0.195977	0.228285	0.256271
6.5	0.000000	0.000000	0.000000	0.062951	0.144226	0.210632	0.266778	0.315413
7.0	0.000000	0.000000	0.000000	0.000000	0.070479	0.202884	0.314832	0.411805
7.5	0.000000	0.000000	0.000000	0.000000	0.000000	0.063780	0.345913	0.590307

In the next section, we have described a method for approximating a given probability distribution by using maximum entropy principle.

7.3 MAXIMUM ENTROPY PRINCIPLE FOR APPROXIMATING A GIVEN PROBABILITY DISTRIBUTION

It is known fact that while dealing with various disciplines of operations research and statistics, we come across many practical problems when we do not get simple expressions for the probability distributions. In all such cases, it becomes very difficult to apply these complicated expressions for further mathematical treatment in the manipulation of new results. Thus, it becomes the desirability to approximate these probability distributions. The approximating probability distributions should have some common properties with the given distribution and it is observed that the simplest property is of having some common moments.

There may be an infinite number of distributions with the same first moment as the given distribution but we are interested with only that probability distribution which is most unbiased and from the theory of maximum entropy principle, we accept only that distribution which has the maximum entropy and this fundamental principle will provide our first approximation to the given distribution. This result is based upon the postulate that most probability distributions are either maximum entropy distributions or very nearly so. To find a better approximation, we try to find that maximum entropy probability distribution which has two moments in common with the given probability distribution. As the number of moments goes on increasing, we get much better and better approximations and as a result of this procedure, we obtain the desired approximation of the given probability distribution under study. We illustrate the above mentioned principle by considering the following numerical example:

Numerical Example

Let us consider the theoretical probability distribution P , given by

i :	0	1	2	3	4
p_i :	0.4	0.3	0.2	0.07	0.03

Our problem is to find MEPD with

(I) same mean;

(II) same mean and p_o ;

(III) same mean, p_o and p_1 ; and

(IV) same first two moments

Our purpose is also to find that maximum entropy probability distribution which is closest to the given probability distribution P .

To solve the above problem, we make use of maximum entropy principle by using Havrada and Charvat's [3] extended entropy of order 2 and our problem becomes:

(I) Maximize Havrada and Charvat's [3] entropy of order 2 given by

$$H(P)=\frac{1}{2}\left\{1-\sum_{i=0}^{4} p_i^2\right\} \tag{7.3.1}$$

subject to the following set of constraints

(i) $\sum_{i=0}^{4} p_i =1$ (7.3.2)

(ii) $\sum_{i=0}^{4} ip_i =1.03$ (7.3.3)

The corresponding Lagrangian is given by

$$L=\frac{1}{2}\left\{1-\sum_{i=0}^{4} p_i^2\right\}-\lambda\left\{\sum_{i=0}^{4} p_i -1\right\}-\mu\left\{\sum_{i=0}^{4} ip_i -1.03\right\}$$

Hence $\frac{\partial L}{\partial p_i}=0$ gives

$$p_i=-(\lambda+i\mu) \tag{7.3.4}$$

Applying (7.3.2), we get

$$\lambda+2\mu=-0.2 \tag{7.3.5}$$

Applying (7.3.3), we get

$$\lambda+3\mu=-0.103 \tag{7.3.6}$$

From (7.3.5) and (7.3.6), we have

$\lambda=-0.394$ and $\mu=0.097$

With these values of λ and μ , equation (7.3.4) gives the following set of probability distribution:

$p_0=0.3940$

$p_1=0.2970$

$p_2=0.2000$

$p_3=0.1030$

$p_4=0.0006$

Obviously, $\sum_{i=0}^{4} p_i = 0.9946 \cong 1$

Thus, the first MEPD P_1 is given by

$P_1 = \{0.3940, 0.2970, 0.2000, 0.1030, 0.0006\}$

(II) In this case, our problem is to maximize Havrada and Charvat's [3] entropy (7.3.1) under the set of constraints (7.3.2), (7.3.3) and $p_o = 0.4$.

Now $\sum_{i=0}^{4} p_i = 1$ gives that

$p_o + \sum_{i=1}^{4} p_i = 1$

or $\sum_{i=1}^{4} (\lambda + i\mu) = -0.6$

or $4\lambda + 10\mu = -0.6$ (7.3.7)

Applying (7.3.3), we get

$10\lambda + 30\mu = -1.03$ (7.3.8)

Equations (7.3.7) and (7.3.8) together give

$\lambda = -0.385$ and $\mu = 0.094$

With these values of λ and μ, equation (7.3.4) gives the following set of probability distribution:

$p_0 = 0.4000$

$p_1 = 0.2910$

$p_2 = 0.1970$

$p_3 = 0.1030$

$p_4 = 0.0090$

Obviously, $\sum_{i=0}^{4} p_i = 1$

Thus, the second MEPD P_2 is given by

$P_2 = \{0.4000, 0.2910, 0.1970, 0.1030, 0.0090\}$

(III) In this case, our problem is to maximize Havrada and Charvat's [3] entropy (7.3.1) under the set of constraints (7.3.2), (7.3.3), $p_o = 0.4$ and $p_1 = 0.3$.

Thus, we have

$$p_o + p_1 + \sum_{i=2}^{4} p_i = 1$$

This gives $\sum_{i=2}^{4} p_i = 0.3$

Applying (7.3.4), we get

$$\lambda + 3\mu = -0.1 \tag{7.3.9}$$

Also (7.3.3) gives

$$9\lambda + 29\mu = -0.73 \tag{7.3.10}$$

Equations (7.3.9) and (7.3.10) together give

$\lambda = -0.355$ and $\mu = 0.085$

With these values of λ and μ, equation (7.3.4) gives the following set of probability distribution:

$p_0 = 0.4000$

$p_1 = 0.3000$

$p_2 = 0.1850$

$p_3 = 0.1000$

$p_4 = 0.0150$

Obviously, $\sum_{i=0}^{4} p_i = 1$

Thus, the third MEPD P_3 is given by

$P_3 = \{0.4000, 0.3000, 0.1850, 0.1000, 0.0150\}$

(IV) In this case, our problem is to maximize Havrada and Charvat's [3] entropy of order 2 subject to the set of constraints (7.3.2) and (7.3.3) along with the additional constraint given by

$$\sum_{i=0}^{4} i^2 p_i = 2.21 \tag{7.3.11}$$

The corresponding Lagrangian is given by

$$L = \frac{1}{2}\left\{1 - \sum_{i=0}^{4} p_i^2\right\} - \lambda\left\{\sum_{i=0}^{4} p_i - 1\right\} - \mu\left\{\sum_{i=0}^{4} ip_i - 1.03\right\} - w\left\{\sum_{i=0}^{4} i^2 p_i - 2.21\right\}$$

Hence $\frac{\partial L}{\partial p_i} = 0$ gives

$$p_i = -(\lambda + i\mu + i^2 w) \tag{7.3.12}$$

Applying (7.3.2), equation (7.3.12) gives

$$5\lambda+10\mu+30w=-1 \tag{7.3.13}$$

Applying (7.3.3), equation (7.3.12) gives

$$10\lambda+30\mu+100w=-1.03 \tag{7.3.14}$$

Applying (7.3.11), equation (7.3.12) gives

$$30\lambda+100\mu+354w=-2.21 \tag{7.3.15}$$

After solving equations (7.3.13), (7.3.14) and (7.3.15), we get

$w=-0.0064$, $\mu=0.1226$ and $\lambda=-0.4068$

With these values of w, λ and μ equation (7.3.12) gives the following set of probability distribution:

$p_0=0.4068$

$p_1=0.2906$

$p_2=0.1872$

$p_3=0.0966$

$p_4=0.0188$

Obviously, $\sum_{i=0}^{4} p_i=1$

Thus, the fourth MEPD P_4 is given by $P_4=\{0.4068, 0.2906, 0.1872, 0.0966, 0.0188\}$

Our next aim is to find that maximum entropy probability distribution which is closest to the given probability distribution P.

For this purpose, we use Havrada and Charvat's [3] extended directed divergence of order 2 to determine the approximity of the distributions to P. We know that Havrada and Charvat's [3] directed divergence is given by

$$D^{\alpha}(P:Q)=\frac{1}{\alpha(\alpha-1)}\left[\sum_{i=0}^{4} p_i^{\alpha}\, q_i^{1-\alpha}-1\right], \alpha\neq 1, \alpha>0$$

Thus for $\alpha=2$, we have

$$D(P:Q)=\sum_{i=0}^{4}\frac{p_i^{2}}{q_i}-1 \tag{7.3.16}$$

Using (7.3.16), we get the following results:

$D(P_1:P)=0.017439$, $D(P_2:P)=0.015286$

$D(P_3:P)=0.010741$, $D(P_4:P)=0.007759$

Hence, we observe that MEPD P_4 is closest to the given probability distribution P .

Important observations: We also make out the following observations:

(a) Since, P_2 is based upon information about mean and p_0 whereas P_1 is based upon information about mean only, we must expect that $D(P_2 : P) < D(P_1 : P)$

In our case, it is found to be true.

(b) Since, P_3 is based upon information about mean, p_0 and p_1 ,we must expect that

$D(P_3 : P) < D(P_2 : P) < D(P_1 : P)$

In our case, it is found to be true.

(c) Since, P_4 is based upon information about the first two moments, whereas P_1 is based upon information about mean only, we must expect that $D(P_4 : P) < D(P_1 : P)$

In our case, it is found to be true.

(d) Since, P_2 is based upon information about mean and p_0 ,whereas P_4 is based upon mean and second moment and $D(P_4 : P) < D(P_2 : P)$

Thus, we conclude that p_0 gives less information than the second moment.

Concluding Remarks: The maximum entropy principle plays a very important and significant role in optimization problems and for its study; we usually apply Lagrange's method of undetermined multipliers to maximize different measures of entropy under a set of one or more constraints. This method has a weakness that sometimes it gives negative probabilities upon optimization. To remove this problem, we ignore the negative probabilities and reformulate our optimization problem for the remaining probabilities and again solve it by using Lagrange's method. The procedure is repeated until we get all maximizing probabilities to be non-negative. In the present chapter, we have made use of constraints in terms of arithmetic and geometric means only but the study can be extended by taking more constraints in terms of existing means and moments. Moreover, the procedure adopted in this chapter can be extended to the study of minimum cross entropy principle. Furthermore, such a study of both types of principles can be extended to discrete and continuous type of fuzzy distributions.

REFERENCES

[1] Burg, J. P. (1972). The relationship between maximum entropy spectra and maximum likelihood spectra. In Childrers, D. G. (Ed.), *Modern Spectral Analysis* (pp. 130-131).

[2] Guiasu, S., & Shenitzer, A. (1975). The principle of maximum entropy. *Mathematical Intelligencer, 7*, 42-48.

[3] Havrada, J. H., & Charvat, F. (1967). Quantification methods of classification process: Concept of structural α-entropy. *Kybernetika, 3*, 30-35.

[4] Herniter, J. D. (1973). An entropy model of brand purchase behaviour. *Journal Marker Review*, 11, 20-29.

[5] Hooda, D.S. and Kapur, J. N. (2001). Crop area distributions for optimal yield. *Operations Research Society of India,* 38, 407-418.

[6] Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Review, 106,* 620-630.

[7] Kapur, J.N., Baciu, G. and Kesavan, H.K. (1994). The minimax information measure. *International Journal of System Sciences,* 26, 1-10.

[8] Kapur, J. N., & Kesavan, H. K. (1987). *The generalized maximum entropy Principle*. Canada: Standard Educational Press, University of Waterloo.

[9] Kullback, S. (1959). *Information theory and statistics*. New York: Wiley.

[10] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*, 79-86.

[11] Onicescu, O. (1966). Energie Informationelle. *Comptes Rendus Mathematique. Academie des Sciences Paris* 841-842.

[12] Parkash, O., Sharma, P. K., & Mahajan, R. (2010). Optimization principle for weighted fuzzy entropy using unequal constraints. *Southeast Asian Bulletin of Mathematics, 34*, 155-161.

[13] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-659.

[14] Topsoe, F. (2002). Maximum entropy versus risk and applications of some classical – discrete distributions. *IEEE Transactions on Information Theory, 48*, 2368-2376.

C

Cai, H. 19
Campbell, L. L. 118
Carlson, B. A. 96
Characteristic function 29
Charvat, F. 5, 51, 64, 74, 77, 103, 173
Chawla, J. S. 145
Chemistry 6
Chen, X. 20
Chen, Y. 97
Clements, M. A. 96
Cluster analysis 96
Code alphabet 115
Code characters 115
Codes 114
Codeword 115
Coding theory 73, 114
Coefficient of determination 17
Computer science 114
Concave function 7, 12, 32, 61, 77, 164
Concavity 7, 32
Conditional entropy 29
Convex function 7, 23, 54
Convex optimization problem 98
Covariance 98
Covariance matrix 95
Cover, T.M. 118
Crispness 50
Crisp set 29, 42, 61
Cross entropy 23, 68, 95

D

Daroczy, Z. 5

A

Abdel-Fadel, A. M. A.
Aczel, J. 5
Additivity property 9
Arimoto, S. 145
Arithmetic mean 16, 116,157, 167, 176
Arrival process 105
Autar, R. 145

B

Baciu, G. 156
Behara, M. 145
Belis, M. 10, 52, 68, 127, 133
Bera, A. K. 95
Bertsimas, D. 97
Best 1-1 code 118
Bezdek, J. C. 31, 52
Bhandari, D. 30, 51, 60, 70, 74
Bhattacharya, A. 97
Biology 6, 97
Birth-death process 105
Birth rate 106
Boltzmann-Shannon entropy 20
Booker, J.M. 29
Brissaud, J. B. 6
Bugár, G. 95
Burg, J. P. 157
Burgin, M. 6

Death rate 106
Decreasing function 31, 61, 77
Degenerate distributions 8, 11, 152
Dehmer, M. 6
De Luca, A. 30, 63, 74
Deterministic 153
Directed divergence 50, 65, 133, 155
Dispersion 16
Diversity 16
Dubois, D. 50, 75
Duckstein, L. 51

E

Ebanks, B. R. 31, 75
Ecology 97
Economics 97
Eguchi, S. 20
El-Sersy, K. F.
Encoding 114
Entropic approach 105
Entropy 104, 114, 151
Expected utility 94, 111
Exponentiated mean 116

F

Fan, J. L. 50, 75
Fan, J. X. 97
Ferreri, C. 20, 51, 76
Folger, T. A. 31
Fuzziness gap 39
Fuzzy cross entropy 82
Fuzzy distributions 43, 50, 71, 156, 176
Fuzzy divergence
Fuzzy entropy 50, 73
Fuzzy numbers 51
Fuzzy set 28, 73
Fuzzy uncertainty 29, 60, 65
Fuzzy vector 29, 66, 82

G

Gallager, R. 105
Garrido, A. 6
Generating function 28, 42, 50, 68
Geometric mean 16, 163, 169, 176
Gibbs entropy 2
Global minimum 18
Golomb, S. 42
Gray, R. M. 96
Guiasu, S. 10, 52, 68, 76, 104, 127, 156
Guo, X. Z. 31, 75
Gurdial 128
Gutter M. S. 97

H

Hanna, S. D. 97
Harmonic mean 16, 116
Havrada, J. H. 5, 52, 64, 74, 77, 103, 173
Herniter, J. D.156
Hessian matrix 22
Hillion, E. 6
Holder’s inequality 125, 130
Hooda, D. S. 156

Hu, Q. 30

I

Ibraheem, N. M. A. 31
Impossible event 8
Inaccuracy 117
Increasing function 9, 31, 77, 114, 152
Information gain 17
Information improvement 97
Input-output analysis 151
Instantaneous code 117

J

Jaynes, E.T. 3, 151, 155
Jensen Burg measure 131
Jianshe, O. 95
Johnson, R. W. 96
Joint entropy 29
Joint probability distribution 9

K

Kakkar, P. 119
Kapur, J. N. 6, 29, 51, 74, 105, 120, 156
Kar, S. 20
Kato, S. 20
Kaufmann, A. 50
Kerridge, D. F. 117
Kesavan, H. K. 76, 156
Khan, A. B. 145
Klir, G. J. 31
Kosko, B. 50
Kraft, L. G. 116
Kraft's inequality 116
Kulkarni, S. 19
Kullback, S. 17, 50, 68, 75, 96, 117, 155
Kumar, R. 119
Kumar, S. 119

L

Lagrange's method 8, 73, 151
Laplace's principle 151
Lauprete, G. J. 97
Leibler, R. A. 50, 68, 75, 96, 117, 155
Leung-Yan-Cheong, S. K. 118
Li, X. 31
Linear programming 73, 151
Linguistics 6
Liu, B. 31
Local minimum 18
Log linear models 156
Longo, G. 127, 138
Lowen, R. 52

M

Mahajan, R. 156
Majumder, D. D. 52, 74
Marketing 156
Markowitz, H. M. 94, 98
Maximizing probabilities 11
Maximum cross entropy principle 76
Maximum entropy principle 31, 76

Ma, Y. L. 50, 75
Mean arrival rates 104
Mean codeword length 114
Mean number of customers 104
Mean return 95
Mean service rate 104
Mean–Variance efficient frontier 95, 98
Measures of risk 95
Medhi, J. M. 105
Medical diagnosis 30
Membership function 29
MEPD 151, 156, 165, 176
Metric spaces 18
Minimum cross entropy 155
M/M/1 system 105
Moments 171
Monotonicity 37
Most fuzzy set 65
Mowshowitz, A. 6
Mukesh 17, 97
Mutual entropy 29

N

Nocetti, D. 95
Noiseless coding theorem 115, 145
Non-degenerate 153
Nonlinear optimization program 100
Nonlinear programming 151
Non steady birth death process 94
Non steady state 108
Non-symmetric measure 17
Normalized measures 28, 46

O

Onicescu, O. 167
Operations research 73, 171
Optimization principle 76, 99, 155
Optimization problems 73
Osman, M. S. A. 31

P

Pal, N. R. 30, 51, 60, 70, 75
Parkash, O. 6, 10, 17, 30, 52, 61, 74, 76, 97, 119,156
Park, S. Y. 95
Partial information 28, 38, 156
Pattern classification 9
Pattern recognition 30, 151
Permutationally symmetric 7, 11, 77
Pessoa, F. 128
Physical sciences 97
Picard, C. F. 127
Piecewise convex 86
Portfolio analysis 73, 94, 156
Portfolio investment 151
Posterior distribution 155
Power mean 16
Prabhakar, B. 105
Prade, H. 50, 52, 75
Prefix code 127
Priori distribution 155
Priyanka 119

Probability distribution 17, 68, 104, 116, 151, 169
Probability generating function 105
Probability spaces 18

Q

Qualitative-quantitative measure 10
Quality control 156
Queueing theory 94, 104, 151

R

Ralescu, D. A. 20
Ramamoorthy, A. 119
Randomness 17
Regional planning 151
Reisher, C. 68
Relative entropy 17, 29
Relative information 68
Reliability theory 156
Renyi, A. 30, 51, 68, 74, 102, 117, 145
Risk-averse 102, 111
Risk prone 102, 111
Rosenfeld, A. 50, 75
Rudas, I. J. 30

S

Samarov, A. 97
Satellite communication 114
Shannon, C. E. 1, 28, 73, 118, 151, 156
Sharma, B. D. 5, 52, 76, 145
Sharma, P. K. 31, 51, 62, 70, 74, 156
Shenitzer, A. 76, 156
Shore, J. E. 96
Siegel, J. J. 98
Signal processing 96
Singh, R. P. 31
Singh, R. S. 10
Singh, Y. B. 6
Singpurwalla, N. D. 29
Social sciences 97
Sociology 6
Source coding theorem 128, 145
Spectral density functions 97
Statistica 171
Statistical inference 151
Statistical mechanics 151
Statistical physics 20
Steady state 105
Steady state queueing process 94
Stochastic 153
Suitable code 116
Sunita 6
Symmetric directed divergence 21

T

Taneja, H. C. 10
Taneja, I. J. 5, 52, 76, 145
Termini, S. 30, 63, 74
Thaler, R. H. 98
Theil, H. 97
Thermodynamic entropy 3
Thukral, A. K. 17

Tomar, V. P. 31
Topsoe, F. 156
Tran, L. 51
Transportation planning 151
Triangle inequality 18
Tuli, R. K. 52

U

Uncertainty 1, 8, 20, 28, 60, 94, 104, 151
Uniform distribution 8, 151, 157, 167
Uniquely decipherable code 115, 145
Uzsoki, M. 95

V

Variance 17, 31, 94, 98
Verdu, S. 19

W

Weighted averages 160
Weighted distribution 1, 127
Weighted divergence measure 59
Weighted entropy 10, 114, 129
Weighted fuzzy divergence 52
Weighted fuzzy information 68, 73
Weighted mean codeword length 127

X

Xie, W. X. 50, 75
Xin, X. L. 30, 75

Y

Yager, R. R. 30, 50
Yang, Y. W. 30
Yu, D. 30

Z

Zadeh, L. A. 2, 29, 73, 74
Zheng, C. 97
Zyczkowski, K. 5

Om Parkash is a Professor at the Guru Nanak Dev University, Department of Mathematics in Amritsar, India.

Having 32 years of teaching and research experience, Dr. Om Parkash is a Professor of Mathematics with Guru Nanak Dev University, Amritsar, India. He has edited 5 books and published 95 research papers in national and international journals of high repute. Under his supervision, 9 scholars have been awarded Ph.D. degrees, and many more are in the pipeline.

Besides, being a member of several mathematical societies and on the panel of experts and editorial boards of several national and international journals, he has presented his research work and delivered invited talks in several national and international conferences in India and in abroad including those held at Hamburg, Germany and Tskuba, Japan. He has accomplished many research project in the field of information theory, organized many conferences sponsored by various Indian agencies and chaired many sessions at several national and international conferences. He was bestowed award of excellence by the Indian Society of Information Theory and Applications for his contribution to the field of information theory.

The present book comprising of various mathematical models, basically deals with three main scientific contributions, viz, entropy, distance and coding which are closely related to each other and will be useful to all those interested in the development of information measures, and using entropy optimization problems in a variety of disciplines. It will be of interest to statisticians, engineers, life-scientists, economists and operational researchers interested in applying the powerful methodology based on maximum entropy principle in their respective disciplines. It will also be a source of inspiration and help to information theoreticians who have been using fuzzy information in their research work and will serve as a source of ready reference for the scientific community.

978-1-61804-371-9