FUZZY REGRESSION, CLUSTERING AND GENERALIZED MEASURES OF FUZZY INFORMATION

A THESIS
SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

MATHEMATICS

by

Rakesh Kumar

to the

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, H.P. - 173 215 (INDIA)

September 2008

CERTIFICATE

This is to certify that the thesis entitled *Fuzzy Regression, Clustering and Generalized Measures of Fuzzy Information* which is being submitted by *Rakesh Kumar* for the award of the degree of *Doctor of Philosophy* in *Mathematics* to the *Jaypee University of Information Technology, Waknaghat,* is a bonafide record of research work done under our guidance and supervision.

The thesis has reached the standard fulfilling the requirements of the regulations relating to the degree. The results obtained in the thesis have not been submitted to any other university or institute for the award of any degree or diploma.

(Dr. Harinder Singh)
Professor and Head
Department of Mathematics
JUIT, Waknaghat
PIN - 173 215 (H.P.) INDIA

(Dr. D.S. Hooda)
Professor and Head
Department of Mathematics
JIET, Guna
PIN - 473 226 (M.P.) INDIA

to the Almighty

Acknowledgement

I would like to express my sincere gratitude and indebtedness to my thesis supervisor, Prof. D.S. Hooda, Head, Department of Mathematics, Jaypee University of Engineering and Technology, Guna, M.P., INDIA, for his guidance, encouragement and invaluable suggestions that made this research possible.

I am grateful to Prof. H. Singh, Head, Department of Mathematics, Jaypee University of Information Technology, Waknaghat, H.P., INDIA, for his supervision and support throughout this work. I truly appreciate Jaipraksh Sewa Sansthan for providing me this opportunity for doing the research work. My special thanks are due to Dr. Y. Medury (COO, Jaypee Education System & Founder Vice Chancellor, JUIT) for his consistent encouragement and help. I also acknowledge my thanks to Prof. D.S. Chauhan (Vice Chancellor), Brig. (Retd.) Balbir Singh (Registrar), Prof. T.S. Lamba (Dean, A & R), Prof. R.M. Vasan (Dean) for their continuous motivation and encouragement.

I would like to pay my heartfelt thanks to my teachers Shri Sahdev Yadav, Dr. K.C. Sinha, Prof. Nand Lal and Dr. V. Sunderpandian who have always been a source of inspiration to me.

I express my sincere thanks and gratitude to my friend Dr. Gaurav Garg and Mr. Amit Srivastava for their help in writing and preparation of this thesis. I also thank Mr. Shri Ram (Librarian) and Shri Vipin Sharma, JUIT, Waknaghat for providing me necessary references and books.

Above all, I feel honored to remember the contribution of my Late grand father and grand mother, who taught me virtues of the life in my childhood. I pay my highest regard to my parents for their affection and encouragement during my life. I deeply acknowledge the love and patience of my wife during my outstation visits of research work.

(Rakesh Kumar)

Table of Contents

List of Figures **viii**

Abstract **ix**

1 Introduction **1**

1.1 Historical Background of Fuzzy Set 1

1.2 Basic Concepts and Notions of Fuzzy Set Theory 2

1.3 Operations and properties of Fuzzy Sets 8

1.4 Fuzzy Uncertainty 12

1.5 Probabilistic and Fuzzy Entropies 16

1.6 Fuzzy Directed Divergence Measure 17

1.7 Fuzzy Information Improvement Measure 19

1.8 Comparison of Probabilistic and Fuzzy Information Measures . . . 19

1.8.1 Similarities 19

1.8.2 Dissimilarities 20

1.9 Measures of 'Useful' Fuzzy Information and 'Useful' Fuzzy Directed Divergence 21

2 Restricted Fuzzy Linear Regression Model **23**

2.1 Introduction 23

2.2 The model 25

2.3 Estimation of Regression Coefficients 26

2.4 Simulation Study 30

2.5 Conclusion 33

3 Fuzzy Clustering Algorithm and its Application in Economics 35

3.1 Introduction 35

3.2 Formation of Economies and Their Contribution 36

3.3 Fuzzy Clustering - The Algorithm 40

3.4 Results of Cluster Convergence Analysis 42

3.5 Conclusion 46

4 Fuzzy Information Measures 49

4.1 Introduction 49

4.2 Generalized Measures of Fuzzy Directed Divergence 54

4.2.1 Fuzzy Directed Divergence of order α 54

4.2.2 Fuzzy Directed Divergence of order α and type β 58

4.3 Measures of Total Ambiguity 62

4.4 Generalized Fuzzy Information Improvement Measures 64

5 Monotonicity and Maximum Fuzziness of Generalized Fuzzy Information Measures 67

5.1 Introduction 67

5.2 Monotonicity of Generalized Measure of Fuzzy Information 69

5.3 Monotonicity of Fuzzy Directed Divergence Measure 75

5.4 Maximum Fuzziness of the Generalized Fuzzy Information and Directed Divergence Measures 79

6 *R*-norm Fuzzy Information Measures and their Generalizations 83

6.1 Introduction 83

6.2 A Generalized R-norm Fuzzy Information Measure 87

6.3 R-norm Fuzzy Directed Divergence Measure 89

6.4 Monotonicity of Fuzzy Information and Fuzzy Directed Divergence Measures 92

6.5 Measures of Total Ambiguity and Fuzzy Information Improvement 95

6.5.1 Total ambiguity 95

6.5.2 R-norm fuzzy information improvement measure 95

7 Measures of 'useful' Fuzzy Information 97

7.1 Introduction 97

7.2 'Useful' Fuzzy Information Measures 100

7.3 Total 'Useful' Fuzzy Information Measure 102

7.4 'Useful' Fuzzy Directed Divergence Measures 104

7.5 Constrained optimization of 'useful' Fuzzy Information Measure . 106

7.6 Constrained Optimization of 'useful' Fuzzy Directed Divergence . 108

Bibiliography 129

List of Figures

		page number
Figure 1.1:	Membership function of a crisp set	. . . 3
Figure 1.2:	Membership function of a fuzzy set	. . . 6
Figure 1.3:	Common membership functions	. . . 7
Figure 1.4:	Common t-norm (AND) operators	. . . 9
Figure 1.5:	Common t-conorm (OR) operators	. . . 10
Figure 1.6:	Fuzzy complement	. . . 10
Figure 1.7:	Symmetrical Triangular Fuzzy Number	. . . 14
Figure 1.8:	Representation of Real and Fuzzy Number	. . . 14
Figure 3.1:	GDP Growth Rate of World, BRIC and G6 Countries	. . . 48
Figure 3.2:	per capita GDP – BRIC Countries	. . . 55
Figure 3.3:	per capita GDP – G6 Countries	. . . 55
Figure 3.4:	Ratio of Fuzzy Cluster Centers (BRIC Countries) (c=2)	. . . 56
Figure 3.5:	Ratio of Fuzzy Cluster Centers (G6 Countries) (c=2)	. . . 56
Figure 3.6:	Ratio of Largest to Smallest Fuzzy Cluster Centers (BRIC Countries) (c=3)	. . . 57
Figure 3.7:	Ratio of Largest to Smallest Fuzzy Cluster Centers (G6 Countries)(c=3)	. . . 57
Figure 5.1:	Monotonic Nature of Fuzzy Entropy w.r.t. α	. . . 88
Figure 5.2:	Monotonic Nature of Fuzzy Entropy w.r.t. β	. . . 89
Figure 5.3:	Monotonic Nature of Fuzzy Directed Divergence w.r.t. α	. . . 93
Figure 5.4:	Monotonic Nature of Fuzzy Directed Divergence w.r.t. β	. . . 95
Figure 6.1:	Monotonic Nature of Generalized R-norm Fuzzy Entropy	. . . 108
Figure 6.2:	Monotonic Nature of R-norm Fuzzy Directed Divergence Measure	. . . 110

Abstract

The objective of this thesis entitled, "*Fuzzy Regression, Clustering and Generalized Measures of Fuzzy Information*", is to study fuzzy regression, fuzzy clustering with application, to characterize new measures of fuzzy information and to study their various generalizations. Fuzzy set theory has capability to describe the uncertain situations, containing ambiguity and vagueness. Fuzziness is found in our decision, in our thinking, in the way we process information, and particularly in our language.

As probabilistic "entropy" measures uncertain degree of the randomness in a probability distribution, Fuzzy entropy measures fuzziness of a set which arises from the intrinsic ambiguity or vagueness carried by the fuzzy set. The entropy of a fuzzy event is different from the classical Shannon entropy, as no probabilistic concept is needed in order to define it. We should note that fuzzy entropy deals with vague and ambiguous uncertainties, while Shannon entropy deals with probabilistic uncertainties. In literature, a number of measures of fuzzy entropy corresponding to the various probabilistic entropy measures have been proposed and studied.

The basic concepts and notions related to fuzzy sets and fuzzy information measures are defined and explained in chapter 1.

In chapter 2, a fuzzy linear regression model is developed using the symmetrical triangular fuzzy number (STFN) under assumption that the regression coefficients are subjected to some exact linear restrictions. The estimators of regression coefficients which satisfy the given restrictions under the model are derived. The dominance of the obtained estimators is demonstrated over the usual unrestricted estimator through simulation study in the sense of mean squared error and absolute bias.

In Chapter 3, we examine the projected convergence of per capita GDP of

the G6 and BRIC countries using a comparatively new technique - Fuzzy c-means Clustering Algorithm (FCM). In *fuzzy clustering*, the objects of the universe of discourse are not classified as belonging to one and only one cluster, but instead, they all possess a degree of membership with each of the clusters. In our application, we have two groups of countries – BRIC and G6, the former comprises of four countries while later of six countries. We have applied the algorithm (using computer programming in C) on the projected Goldman Sachs Report (2003) for both the groups individually by taking number of clusters as $c = 2$ and $c = 3$.

In Chapter 4, we define two new generalized measures of fuzzy directed divergence and prove their validity. Particular cases and computational structure of these directed divergence measures are also discussed. New measures of total ambiguity and generalized measures of fuzzy information improvement have been studied.

In Chapter 5, the monotonic nature of generalized fuzzy information measure with respect to the parameters is studied and verified by constructing the computed tables and plots on taking different fuzzy sets and different values of the parameters. Particular cases and comparison of monotonicity between the corresponding probabilistic measure and the generalized measures of fuzzy information are discussed. Similar kind of investigation has been carried out for the generalized measure of fuzzy directed divergence. Under a given constraint, the maximum fuzziness of the parametric generalized measures of fuzzy information and fuzzy directed divergence have also been discussed.

In Chapter 6, a new generalized R-norm fuzzy information measure is characterized and studied. A new R-norm fuzzy directed divergence measure has been proposed and proved its validity. The monotonic nature of the proposed R-norm fuzzy information measure and R-norm fuzzy directed divergence measure with respect to the parameters is studied. New generalized R-norm measures of total ambiguity and fuzzy information improvement have also been studied.

In Chapter 7, we introduce a new concept of 'useful' fuzzy information by attaching utility to the uncertainties of fuzziness and probabilities of randomness. A measure of total 'useful' fuzzy information is derived. We also define and prove the validity of a new measure of 'useful' fuzzy directed divergence of a fuzzy set from another fuzzy set. Under the given constraints, the optimization of 'useful' fuzzy information measure and 'useful' fuzzy directed divergence are studied.

Chapter 1

Introduction

"*We must exploit our tolerance for imprecision.*"- ***Lotfi A. Zadeh.***

1.1 Historical Background of Fuzzy Set

Binary logic deals with the variables which are either *true* or *false*, *black* or *white*, 1 or 0, *yes* or *no*, etc. The equivalent Fuzzy logic extends Boolean Logic to all values in the interval [0, 1] so that the variable may have a truth value that is neither completely true nor completely false. In other words, Fuzzy logic is basically a multivalued logic that allows intermediate values to be defined between conventional evaluations like *yes/no, true/false, black/white,* etc. Notions like *rather warm* or *pretty cold* can be formulated mathematically and processed by computers. In this way an attempt is made to apply a more human-like way of thinking in the programming of computers. Gradual transition from

- Traditional View to Modern View
- Certainty to Uncertainty
- Precision to Imprecision

- Specificity to Non-specificity
- Sharpness to Vagueness
- Consistency to Inconsistency

has made fuzzy logic a profitable tool for the controlling of subway systems and complex industrial processes, as well as for household and entertainment electronics, diagnosis systems and other expert systems.

Fuzzy Logic was initiated by Lotfi A. Zadeh (1965), professor of computer science at the University of California in Berkeley. However, the idea of an extended multivalued logic had been considered by physicists early in the 20^{th} century, but had not become a standard part of science (there was the concept of vague sets). Fuzzy system is an alternative to traditional concepts of set membership and logic that has its origins in ancient Greek philosophy. Fuzzy logic has ability to capture (mathematically) reasoning about the notions with inherited fuzziness, such as being tall, young, fat, hot etc. Similar to probabilistic logic, we have the real valued truth in fuzzy logic also i.e., the truth of certain statement can be any real number in the interval [0, 1].

Fuzziness is found in our decision, in our thinking, in the way we process information and particularly in our language. A Sunny day may contain some clouds (avoiding sharpness). Phrases like "*see you later*", "*a little more*" or "*I don't feel very well*" are fuzzy expressions. However, for most of the problems that we face, Zadeh (1973, 1975, 1984) suggests that we can do a better job in accepting some level of imprecision.

For any field S and any theory T can be fuzzified by replacing the concept of a crisp set in S and T by that of a fuzzy set. Fuzzification leads basic field such as arithmetic to fuzzy arithmetic, topology to fuzzy topology, graph theory to fuzzy graph theory, probability theory to fuzzy probability theory. Similarly, in application to applied fields such as neural network theory, stability theory, pattern recognition and mathematical programming, fuzzification leads to fuzzy neural network theory, fuzzy stability theory, fuzzy pattern recognition and fuzzy mathematical programming. Fuzzification gives greater generality, higher expressive power, an enhanced ability to model real world problems. Most importantly, it gives a methodology for exploiting the tolerance for imprecision i.e., a methodology which serves to achieve tractability, robustness and lower solution cost.

Zadeh's work had a profound influence on the thinking about uncertainty because it challenged not only probability theory as the sole representation for uncertainty, but the very foundations upon which probability theory was based - classical binary (two-valued) logic (refer to Klir and Yuan (1995)).

1.2 Basic Concepts and Notions of Fuzzy Set Theory

An ordinary or *crisp* set A in a universe of discourse U can be defined by listing all its members or by defining conditions to identify the elements $x \in A$ i.e., $A = \{x | x \text{ meets some condition}\}$. The characteristic function, generally called *membership function*, associated with A is a mapping $\mu_A : U \rightarrow \{0, 1\}$ such that for any element x of the universe, $\mu_A(x) = 1$, if x is a member of A and $\mu_A(x) = 0$, if x is not a member of A. Figure 1.1 shows the membership function characterizing the crisp set $A = \{x |\, 20 \leq x \leq 26\}$.

Fuzzy sets are generalization of crisp sets. A fuzzy set A defined on a universe of discourse U is characterized by a membership function $\mu_A(x)$ which takes values in the interval [0, 1] (i.e., $\mu_A : U \to [0,\, 1]$). The value $\mu_A(x)$ represents the *grade of membership* of $x \in U$ in A. This grade corresponds to the degree to which that element or individual is similar or compatible with the concept represented by the fuzzy set. Thus, the elements may belong in the fuzzy set to a greater or lesser degree as indicated by a larger or smaller membership grade. The membership function may be described as follows:

$$\mu_A(x) = \begin{cases} 0, & \text{if } x \notin A \text{ and there is no ambiguity,} \\ 1, & \text{if } x \in A \text{ and there is no ambiguity,} \\ 0.5, & \text{if there is maximum ambiguity whether } x \in A \text{ or } x \notin A \end{cases}$$

Semantically, the difference between the notions of probability and grade of membership lies in the fact - probability statements are about the likelihoods of outcomes i.e., an event either occurs or does not, and you can bet on it, but with fuzziness, one cannot say unequivocally whether an event occurred or not, and instead we try to model the extent to which an event occurred.

It may be noted that the term membership function makes more sense in the context of fuzzy sets as it stresses the idea that $\mu_A(x)$ denotes the degree to which x is a member of the set A. The fuzzy set A can be expressed as:

Vector: $A = \left\{\frac{\mu_A(x_i)}{x_i} \;:\; x_i \in U,\; i = 1,\, 2, \ldots n\right\}$.

Summation: $A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \ldots + \frac{\mu_A(x_n)}{x_n} = \sum_{i=1}^{n} \frac{\mu_A(x_i)}{x_i}$.

Ordered pairs: $A = \{(\mu_A(x_1), x_1), (\mu_A(x_2), x_2), \ldots, (\mu_A(x_n), x_n)\}$.

When the universe of discourse, U, is continuous and infinite, the fuzzy set is denoted by $A = \left\{\int \frac{\mu_A(x)}{x}\right\}$. Here, in these notations, the horizontal bar is not a quotient but rather a simple line. The summation is not the algebraic summation, but a theoretical aggregation operator or collection operator and similarly, the integral sign.

Example 1.1:

Let $U = \{0, 1, 2, 3, \ldots, 10\}$ and $A = \{x|\ x\,\text{is close to}\ 5,\ x \in U\}$. Using a triangular membership function, the set can be described in following forms:

$$A = \left\{\tfrac{0}{0}, \tfrac{0.2}{1}, \tfrac{0.4}{2}, \tfrac{0.6}{3}, \tfrac{0.8}{4}, \tfrac{1}{5}, \tfrac{0.8}{6}, \tfrac{0.6}{7}, \tfrac{0.4}{8}, \tfrac{0.2}{9}, \tfrac{0}{10}\right\};$$

$$A = \tfrac{0}{0} + \tfrac{0.2}{1} + \tfrac{0.4}{2} + \tfrac{0.6}{3} + \tfrac{0.8}{4} + \tfrac{1}{5} + \tfrac{0.8}{6} + \tfrac{0.6}{7} + \tfrac{0.4}{8} + \tfrac{0.2}{9} + \tfrac{0}{10};$$

$$A = \{(0,0), (0.2,1), (0.4,2), (0.6,3), (0.8,4), (1,5), (0.8,6), (0.6,7), (0.4,8), (0.2,9), (0,10)\}$$

The key difference between a crisp set and a fuzzy set is their membership function. A crisp set has unique membership function, where as a fuzzy set can have an infinite number of membership functions to represent it. For example, one can define a possible membership function for the set of real numbers close to 0 as follows:

$$\mu_A(x) = \frac{1}{1 + 10x^2};\ x \in \mathbb{R}.$$

Here the number 3 is assigned a grade of 0.01, the number 1 is assigned a grade of 0.09 and the number 0 is assigned a grade 1. For fuzzy sets uniqueness is sacrificed, but flexibility is gained because the membership function can be adjusted to maximize the utility/sensitivity for a particular application. It may be noted that elements in a fuzzy set, because the membership need not be complete, can also be member of other fuzzy sets on the same universe.

The operation that assigns a membership value $\mu(x)$ to a given value $x \in U$ is called *fuzzification*, e.g., Figure 1.2 shows the membership function of the fuzzy set $A = \{x|\,x$ is almost between 20 and 26$\}$ i.e., the fuzzy set representing approximately the same concept as that of the crisp set of Figure 1.1.

Membership functions might formally take any arbitrary form as they express only an element-wise membership condition. However, they usually exhibit smooth and monotonic shapes. This is due to the fact that membership functions are generally used to represent linguistic units described in the context of a coherent universe of discourse i.e., the closer the elements, the more similar the characteristics they represent, as in the case for variables with physical meaning. The most commonly used membership functions are triangular, trapezoidal, bell-shaped and Gaussian, which are explained below:

(a) Triangular membership function is specified by three parameters, defined by $trim\,f(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right)$ and is shown in Figure 1.3(a).

(b) Trapezoidal membership function is specified by four parameters, defined by $trapm\,f(x; a, b, c, d) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right)$ and is shown in Figure 1.3(b).

(c) A general bell-shaped membership function is defined by $gbellm\,f(x; a, b, c) = \frac{1}{1+\left|\frac{x-c}{b}\right|^{2b}}$ and is shown in Figure 1.3(c).

(d) A Gaussian membership function is specified by two parameters (m, δ) as

$$Gaussian(x : m, \delta) = \exp\left(-\frac{(x-m)^2}{\delta^2}\right)$$

where m and δ denote the center and width of the function respectively. We control the shape of the function by adjusting the parameter δ. A small δ

will generate a “thin” membership function, while a big δ will lead to a “flat” membership function.

A membership function can be designed in three ways:

- Interview those who are familiar with the underlying concepts and later adjust it based on a tuning strategy,
- Construct it automatically from data,
- Learn it based on feedback from the system performance.

The guidelines for designing membership function:

- Use parameterizable functions that can be defined by a small number of parameters. Parameterizable membership functions reduce the system design time and facilitate the automated tuning of the system.
- The parameterizable membership functions most commonly used in practice are the *triangular* and *trapezoidal* membership functions because of their simplicity.

- To learn the membership function using neural network learning techniques, we choose a differentiable (or even continuous differentiable) membership function, e.g., *Gaussian*.

However, constructing meaningful membership function in various contexts is a difficult task.

Fuzzy Logic is motivated by two objectives:

- First, it aims to alleviate difficulties in developing and analyzing complex systems encountered by conventional mathematical tools. This motivation requires fuzzy logic to work in quantitative and numeric domains.
- Second, it is motivated by observing that human reasoning can utilize concepts and knowledge that do not have well defined vague concepts. This motivation enables fuzzy logic to have a descriptive and qualitative form.

Components of Fuzzy Logic

- *Fuzzy Predicates*: tall, small, kind, expensive,...
- *Predicates modifiers* (hedges): very, quite, more or less, extremely, ...
- *Fuzzy truth values*: true, very true, fairly false, ...
- *Fuzzy quantifiers*: most, few, almost, usually, ...
- *Fuzzy probabilities*: likely, very likely, highly likely, ...

The discussion about fuzzy sets can be related according to need for a formal basis of fuzzy logic by taking advantage of the fact that “it is well established that propositional logic is isomorphic to set theory under the appropriate correspondence between components of these two mathematical systems”. Furthermore, both of these systems are isomorphic to a Boolean algebra. Some of the most important equivalences between these isomorphic domains are:

Sets	**Logic**	**Algebra**
Membership	Truth	Value
Member ($\in$)	True (T)	1
Non-member ($\notin$)	False (F)	0
Intersection ($\cap$)	AND ($\wedge$)	Product ($\times$)
Union ($\cup$)	OR ($\vee$)	Sum ($+$)
Complement ( )	NOT ($\sim$)	Complement ($'$)

1.3 Operations and properties of Fuzzy Sets

The operations on fuzzy sets are extension of the most commonly used crisp operations. This extension imposes a prime condition that all the fuzzy operations which are extensions of crisp concepts must reduce to their usual meaning when the fuzzy sets reduce themselves to crisp sets i.e., when they have only 1 and 0 as membership values. For the definitions of the following operations, we assume A and B are two fuzzy subsets of U; x denotes an arbitrary element of U:

Intersection/AND operation is defined as $\mu_{A\cap B}(x) = \mu_A(x) \wedge \mu_B(x)$. The most common operators (also known as t-norm operators), illustrated in **Figure 1.4** and are defined as follows:

minimum	:	$\min\{\mu_A(x),\ \mu_B(x)\}$
product	:	$\mu_A(x).\mu_B(x)$
bounded product	:	$\max\{0,\ \mu_A(x) + \mu_B(x) - 1\}$

Union/OR operation is defined as $\mu_{A\cup B}(x) = \mu_A(x) \vee \mu_B(x)$. The most common operators of this type (also known as t-conorm operators), illustrated in **Figure 1.5** and are defined as follows:

maximum	:	$\max\{\mu_A(x),\ \mu_B(x)\}$
probabilistic sum	:	$\mu_A(x) + \mu_B(x) - \mu_A(x).\mu_B(x)$
bounded sum	:	$\min\{1,\ \mu_A(x) + \mu_B(x)\}$

Complement/NOT operation is defined as $\mu_{\bar{A}}(x) = \mu_{\sim A}(x)$. This operator is also called *fuzzy complement* operator which is almost universally used in fuzzy inference systems. It is illustrated in **Figure 1.6** and defined as follows:

fuzzy complement : $1 - \mu_A(x)$

Example 1.2: Consider a fuzzy set of tall men described by

tall men $= \left\{\frac{0}{165}, \frac{0}{175}, \frac{0}{180}, \frac{0.25}{182.5}, \frac{0.5}{185}, \frac{0.75}{187.5}, \frac{1}{190}\right\} = A$ (say);

average men $= \left\{\frac{0}{165}, \frac{1}{175}, \frac{0.5}{180}, \frac{0.25}{182.5}, \frac{0}{185}, \frac{0}{187.5}, \frac{0}{190}\right\} = B$ (say); then

$A \cap B = \left\{\frac{0}{165}, \frac{0}{175}, \frac{0}{180}, \frac{0.25}{182.5}, \frac{0}{185}, \frac{0}{187.5}, \frac{0}{190}\right\} = \left\{\frac{0}{180}, \frac{0.25}{182.5}, \frac{0}{185}\right\}$.

$A \cup B = \left\{\frac{0}{165}, \frac{1}{175}, \frac{0.5}{180}, \frac{0.25}{182.5}, \frac{0.5}{185}, \frac{0.75}{187.5}, \frac{1}{190}\right\}$.

complement of tall men $= \bar{A} = \left\{\frac{1}{165}, \frac{1}{175}, \frac{1}{180}, \frac{0.75}{182.5}, \frac{0.5}{185}, \frac{0}{190}\right\}$.

Properties of Fuzzy sets:

(i) $A \cap A = A$, $A \cup A = A$, $\bar{\bar{A}} = A$.

(ii) Commutativity: $A \cap B = B \cap A$, $A \cup B = B \cup A$.

(iii) Associativity: $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$.

(iv) Distributivity: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

(v) DeMorgan's Laws: $(\overline{A \cap B}) = \bar{A} \cup \bar{B}$, $(\overline{A \cup B}) = \bar{A} \cap \bar{B}$.

(vi) Absorption: $A \cup (A \cap B) = A$, $A \cap (A \cup B) = A$.

(vii) Zero Law: $A \cup U = U$, $A \cap \phi = \phi$.

(viii) Identity Law: $A \cap U = A$, $A \cup \phi = A$.

Some important definitions related fuzzy set theory are given below:

Sigma Count of a Fuzzy set

For any fuzzy set A defined on a finite universal set U, we define scalar cardinality, called *sigma count* of A, denoted and defined by $|\mathrm{A}| := \sum_{x \in U} \mu_A(x) = \sum_{count}(A)$. It may be noted that $|A| \geq 0$.

Example 1.3: For $A = \{0.4/\mathrm{a}, 0.6/\mathrm{b}, 0.9/\mathrm{c}, 0.2/\mathrm{d}\}$ and $B = \{0.5/\mathrm{a}, 0.6/\mathrm{b}, 0.8/\mathrm{c}\}$, we compute the cardinalities or sigma count of A and B to be $|A| = 2.1$ and $|B| = 1.9$.

Subsethood

For any pair of fuzzy subsets defined on a finite universal set U, the degree of *subsethood* of A in B, denoted by $S(A, B)$ and is defined by

$$S(A,B) = \frac{1}{|A|}\left(|A| - \sum_{x \in U} \max\{0,\ \mu_A(x) - \mu_B(x)\}\right)$$
$$= 1 - \frac{\sum_{x \in U} \max\{0, \mu_A(x) - \mu_B(x)\}}{\sum_{count}(A)}$$

Kosko (1986) viewed a fuzzy set as a fuzzy message. He generalized the bit value representation of an ordinary set to fuzzy set. If $U = \{a,\ b,\ c,\ d\}$, then the subset $A = \{a, c\}$ can be represented by a bit vector $A = \{1,\ 0,\ 1,\ 0\}$. Similarly, a fuzzy subset of U can be represented as a fuzzy vector or fuzzy message, e.g., $A = \{0.2,\ 0.5,\ 0.3,\ 0.1\}$. The containment of a fuzzy message in another one has been expressed through a fuzzy message conditioning measure called subsethood (Kosko (1986),(1991)) of a fuzzy set in another one.

The $\sum$ term in the formula describes the sum of the degrees to which the subset inequality $\mu_A(x) \le \mu_B(x)$ is violated. |A| in the denominator is a normalizing factor for getting the range $0 \le S(A,B) \le 1$.

Also, $S(A, B) = \frac{|A \cap B|}{|A|} = \frac{\sum_{count}(A \cap B)}{\sum_{count}(A)}$; where the intersection is fuzzy intersection. Thus it can be concluded that

- Two fuzzy sets A and B are equal if and only if $S(A, B) = 1$.
- $S(A, B) = 0$ if and only if A and B have no points in common i.e., have no common points with fuzzy membership greater than 0.

For example, $A = (0.2,\ 0.7,\ 0.5,\ 0.4,\ 0.3)$ is a subset (submessage) of $B = (0.4,\ 0.8,\ 0.7,\ 0.5,\ 0.4)$, but $A' = (0.5,\ 0.7,\ 0.5,\ 0.4,\ 0.3)$ is not a subset of B because it violates $\mu_{A'} \le \mu_B$ for the first support. However, it is clear that A' has a high degree of containment in B.

Standard Fuzzy Sets

Fuzzy sets A and B are said to be *fuzzy-equivalent* if $\mu_B(x_i) =$ either $\mu_A(x_i)$ or $1 - \mu_A(x_i)$; $\forall x_i \in U$. From the fuzziness point of view there is no essential difference between fuzzy equivalent sets. A *standard fuzzy set* is that member of the class of fuzzy equivalent sets whose all membership values are less than or equal to 0.5.

Equality of Fuzzy Sets

Two fuzzy sets are said to be *equal* if and only if $\mu_A(x_i) = \mu_B(x_i)$; $\forall x_i \in U$.

Support of a Fuzzy Set

Given a fuzzy set A which is a subset of the universal set U, the *support* of A denoted by $supp(A)$, is an ordinary set defined as the set of elements whose degree of membership in A is greater than 0 i.e.,supp$(A) = \{x_i \in U | \, \mu_A(x_i) > 0\}$.

Fuzzy Number and Fuzzy Interval

A *fuzzy number* is a quantity whose value is imprecise rather than exact as is the case with "ordinary" (single-valued) numbers. Mathematically, a *fuzzy number* is a convex and normalized fuzzy set whose membership function is at least segmentally continuous having bounded support and has the functional value $\mu_A(x) = 1$ at precisely one element which is called modal value of fuzzy number.

A *symmetric triangular fuzzy number* (STFN) is defined by the membership function

$$\mu_{\tilde{A}_j(c)} = \begin{cases} 1 - \frac{|c-c_j|}{r_j} & \text{when } c_j - r_j \leq c \leq c_j + r_j \\ 0 & \text{otherwise,} \end{cases}$$

where c_j is known as the middle value for which $\tilde{A}_j(c_j) = 1$ and $r_j > 0$ is the spread of $\tilde{A}_j$. The fuzzy number $\tilde{A}_j$ expresses the linguistic terms approximately c_j or around c_j and it is denoted by $\tilde{A}_j = (c_j, r_j)$; $\forall j \in \{1, 2, \ldots, n\}$. The spread denotes the fuzziness of the function. This fuzzy number is shown by the Figure (1.7):

However, some fuzzy numbers have concave, irregular or even chaotic membership functions. There is no restriction on the shape of the membership curve as long as each value in the domain corresponds to one and only one grade in the range

A *fuzzy interval* is an uncertain set with a mean interval whose elements possess the membership function value $\mu_A(x) = 1$. As in fuzzy numbers, the membership function must be convex, normalized and at least segmentally continuous. Fuzzy numbers and fuzzy intervals (sometimes called trapezoidal fuzzy number) in contrast with crisp numbers and crisp intervals can be better viewed in Figure 1.8.

1.4 Fuzzy Uncertainty

In the real world the amount of information that is available is infinite (facts collected from observations or measurements) and at the same time there is lack of information (meaningful interpretation and correlation of data that allows one to make decisions). Infinite things change because one can go into greater and greater detail of description. The number of preconditions to the execution of any action is also infinite, as the number of things that can go wrong is infinite. Complexity in the world generally arises from uncertainty in the form of ambiguity. Uncertainty is produced when a lack of information exists i.e., the complexity also involves the degree of uncertainty. It may be noted that knowledge is information at a higher level of abstraction.

For example,

Amit is 10 years old. (fact)

Amit is not old. (knowledge)

According to Klir and Yuan (1995) " Uncertainty can be thought of in an epistemological sense as being the inverse of information. Information about a particular engineering or scientific problem may be incomplete, imprecise, fragmentary, vague, contradictory or deficient in some other way."

When we acquire more and more information about a problem, we become less and less uncertain about its formulation and solution. Problems that are characterized by very little information are said to be ill-posed, complex or not sufficiently known. These problems are charged with a high degree of uncertainty. Uncertainty can be manifested in many forms - it can be fuzzy (not sharp, unclear, imprecise, approximate), it can be vague (not specific, amorphous), it can be ambiguous (too many choices, contradictory), it can be of the form of ignorance (dissonant, not knowing something) or it can be a form due to natural variability (conflicting, random, chaotic, unpredictable). Many other linguistic labels have been applied to these various forms, but for now these shall suffice. Zadeh (2002) posed some simple examples of these forms in terms of a person's statements about when they shall return to a current place in time. The statement " I shall return soon" is vague, whereas the statement " I shall return in few minutes" is fuzzy; the former is not known to be associated with any unit of time (seconds, hours, days) and the latter is associated with an uncertainty that is at least known to be on the order of minutes. The phrase, "I shall return within 2 minutes of 6 PM." involves an uncertainty which has a quantifiable imprecision; probability theory could address this form.

Vagueness can be used to describe certain kinds of uncertainty associated with linguistic information or intuitive information. Examples of vague information are that the data quality is "good" or that the transparency of an optical element is "acceptable". Moreover, in terms of semantics, even the terms vague and fuzzy cannot be generally considered synonyms as explained by Zadeh (1995). Usually a vague proposition is fuzzy, but converse is not generally true (refer to Ross (2005)).

Since fuzzy set theory makes statements about one concrete object, therefore, it helps in modeling local vagueness. On the other hand, probability theory makes statements about a collection of objects from which one is selected, therefore, it helps in modeling global uncertainty.

The behavior of a fuzzy system is completely deterministic. Fuzzy logic differs from multivalued logic by introducing concepts such as linguistic variables and hedges to capture human linguistic reasoning. It can also be noticed that awareness of knowledge (what we know and what we do not know) and complexity goes together. The following are different types of information:

- *Uncertain information*: Information for which it is not possible to determine whether it is true or false, e.g., a person is “possibly 30 years old.”

- *Imprecise information*: Information which is not available as precise as it should be, e.g., a person is “around 30 years old.”
- *Vague information*: Information which is inherently vague, e.g., a person is “young.”
- *Inconsistent information*: Information which contains two or more assertions that cannot be true at the same time, e.g., two assertions are given: “Amit is 10.” and “Amit is older than 20.”
- *Incomplete information*: information for which data is missing or data is partially available, e.g., a person’s age is “not known” or a person is “between 25 and 32 years old.”
- Combination of the various types of such information may also exist, e.g., “possibly young”, “possibly around 30”, etc.

In order to explain the concept of fuzzy uncertainty in a more comprehensive way, we illustrate consider the following example:

- Suppose an editor of a magazine sends an article to n reviewers for their opinions.
- Each reviewer is asked to grade the article at some point in the scale 0, 0.1, 0.2, . . . ,0.9, 1.0.
- The grade 0 means that the article is completely useless i.e., there is no uncertainty in the mind of the reviewer about it.
- The grade 1 means that the article is completely useful and important and there is again no uncertainty in the mind of reviewer about it.

- Grade 0.1 means that the article is almost useless, but there is little content in the article which create some uncertainty in the mind of the reviewer.
- Similarly, grade 0.9 means that the article is almost useful and important, but there are some undesirable content which create uncertainty in the mind of reviewer about its utility.

As far as the editor of the magazine is concerned grades 0 and 1 give him clear indication, but grades 0.1 and 0.9 create a certain degree of uncertainty in his mind which is the same in both cases. Similarly, grades 0.2 and 0.8, grades 0.3 and 0.7, grades 0.4 and 0.6 represent the same degree of uncertainty for the editor. This type of uncertainty is called as fuzzy uncertainty which is different from probabilistic uncertainty. The fuzzy uncertainty is maximum when reviewer gives the grade 0.5 because the editor is completely uncertain whether to publish the article or reject it.

It is important to notice that if the grade x is 0 or 1, the fuzzy uncertainty is 0 and if the grade x is 0.5, the fuzzy uncertainty is maximum i.e., as x increases from 0 to 0.5, the fuzzy uncertainty increases from 0 to a certain maximum value and as x increases further from 0.5 to 1, the fuzzy uncertainty decreases from this maximum value to zero. Thus the fuzzy uncertainty is a function of x with 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 $f(x)$ is a continuous and differentiable function, but it is not necessary. Now if the n reviewers give independent grades $x_1, x_2, ..., x_n$ giving rise to fuzzy uncertainties $f(x_1)$, $f(x_2), ..., f(x_n)$, then the total fuzzy uncertainty is $f(x_1)+ f(x_2)+...+f(x_n)$.

Now suppose a die is thrown and it is asked to guess the top face. The uncertainty about the outcome is attributed to randomness. The best way to approach this question might be to describe the status of the die in terms of a probability distribution on the six faces. Uncertainty that arises due to chance is called Probabilistic Uncertainty (PU). Next, suppose it is asked to interpret the top face of the die say, HIGH (or LOW). In this case, we have other type of uncertainty which appears due to linguistic imprecision or vagueness, which is called Fuzzy Uncertainty (FU). Fuzzy uncertainty differs from probabilistic uncertainty because it deals with situations where set boundaries are not sharply defined. Probabilistic uncertainties are not due to ambiguity about set-boundaries, but rather about the belongingness of elements or events to crisp sets.

1.5 Probabilistic and Fuzzy Entropies

Let $X = (x_1, x_2, \ldots, x_n)$ is a discrete random variable with probability distribution $P = (p_1, p_2, \ldots, p_n)$ in an experiment. Shannon (1948) argued that the uncertainty associated with the probability distribution P should be a continuous function of $p_1, p_2, \ldots, p_n$ and permutationally symmetric. Shannon gave the mathematical formulation of information contained in the experiment as given below:

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i. \tag{1.5.1}$$

The expression (1.5.1) measures the uncertainty due to probabilistic nature of the phenomenon concerned. On the advice of the famous mathematician-physicist John Von Neumann, Shannon called the expression (1.5.1) as measure of entropy because it resembled the expression of entropy in thermodynamics. However, there was no real connection between the two entropies – thermodynamic entropy and information-theoretic entropy, but later some links were discovered and information-theoretic entropy was found useful in the study of thermodynamics.

If $x_1, x_2, ..., x_n$ are members of the universe of discourse U, then the vector $(\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n))$ is called fuzzy vector corresponding to the fuzzy set A; where $\mu_A(x_i)$ gives the *grade of membership* of $x_i \in U$ in A . Though all the membership values lie between 0 and 1, but these are not probabilities because their sum is not unity. However,

$$\Phi_A(x_i) = \frac{\mu_A(x_i)}{\sum\limits_{i=1}^{n} \mu_A(x_i)}, \ i = 1, 2, \ldots, n \tag{1.5.2}$$

is a probability distribution. Analogous to entropy (1.5.1) due to Shannon, De Luca and Termini (1972) suggested the following measure of fuzzy entropy:

$$H(A) = -\sum_{i=1}^{n} [\mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log (1 - \mu_A(x_i))]. \tag{1.5.3}$$

The fuzzy entropy given by expression (1.5.3) measures uncertainty due to vagueness and ambiguity, while probabilistic entropy measures uncertainty due to the information being available in terms of a probability distribution. It may be noted that fuzzy-equivalent sets have the same fuzzy entropy but two sets may have the same fuzzy entropy without being fuzzy equivalent.

On the other hand, Loo (1977) proposed a general mathematical form for measuring fuzziness as

$$H_L(A) = F\left[\sum_{i=1}^{n} c_i f_i(\mu_A(x_i))\right], \tag{1.5.4}$$

where $c_i \in \mathbb{R}^+$, f_i is a real valued function such that, $f_i(0) = f_i(1) = 0$ and $f_i(u) = f_i(1-u)$ for $u \in [0, 1]$. Here $f_i(.)$ is a strictly increasing function on [0,0.5].

It may be seen that the meaning of fuzzy entropy is different from the classical Shannon entropy because no probabilistic concept is needed in order to define it. This is due to the fact that fuzzy entropy contains vague and ambiguous uncertainties, while Shannon entropy contains the probabilistic uncertainties.

However, the measures of fuzzy entropy and probabilistic entropy have a great deal in common and the knowledge of probabilistic entropies is being used to enrich the literature on fuzzy measures. Fuzzy entropy has been studied and applied by many researchers in various fields like image processing, communication theory, pattern recognition, etc.

1.6 Fuzzy Directed Divergence Measure

Let $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$ be two probability distributions of a discrete random variable. The measure of directed divergence of P from Q is defined as a function $D(P : Q)$ satisfying the following conditions:

(i) $D(P : Q) \geq 0$

(ii) $D(P : Q) = 0$ if and only if $P = Q$

(iii) $D(P : U) = H(U) - H(P)$; where U is the uniform probability distribution and $H(P)$ is the measure of probabilistic entropy.

Kullback and Leibler (1951) defined the measure of directed divergence of probability distribution P from the probability distribution Q as

$$D(P : Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}. \tag{1.6.1}$$

(1.6.1) is also called distance measure as it measures how far the probability distribution P is from the probability distribution Q. Further, Kullback (1959) suggested the measure of symmetric divergence in the following manner:

$$J(P : Q) = D(P : Q) + D(Q : P) = \sum_{i=1}^{n} (p_i - q_i) \ln \frac{p_i}{q_i}. \tag{1.6.2}$$

The measures (1.6.1) and (1.6.2) have been generalized and studied by Taneja (1989, 1995).

Analogously, the measure of fuzzy directed divergence of fuzzy set A from fuzzy set B is defined as a function $I(A,B)$ which satisfies the following conditions:

(i) $I(A,B) \geq 0$

(ii) $I(A,B) = 0$, if and only if $A = B$

(iii) $I(A,A_F) = H(A_F) - H(A)$; where A_F is the most fuzzy set i.e., all of its membership values are $\frac{1}{2}$ and $H(A)$ is the fuzzy entropy of the set A.

Let $(\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n))$ and $(\mu_B(x_1), \mu_B(x_2), \ldots, \mu_B(x_n))$ are fuzzy vectors corresponding to fuzzy sets A and B respectively with same supporting points $x_1, x_2, \ldots, x_n$. The simplest measures of fuzzy directed divergence and symmetric divergence as suggested Bhandari and Pal (1993), are

$$I(A,B) = \sum_{i=1}^{n} \left[\mu_A(x_i) \log \frac{\mu_A(x_i)}{\mu_B(x_i)} + (1-\mu_A(x_i)) \log \frac{(1-\mu_A(x_i))}{(1-\mu_B(x_i))} \right] \tag{1.6.3}$$

and

$$J(A,B) = I(A,B) + I(B,A) = \sum_{i=1}^{n} [(\mu_A(x_i) - \mu_B(x_i)] \log \frac{\mu_A(x_i)(1-\mu_A(x_i))}{\mu_B(x_i)(1-\mu_B(x_i))}. \tag{1.6.4}$$

This measure of fuzzy symmetric divergence can discriminate between two fuzzy sets. It may be noted that $J(A,B)$ is symmetric with respect to μ_A and μ_B. It also satisfies the following properties:

(i) $J(A,B) \geq 0$, $J(A,B) = 0$ if and only if $A = B$,

(ii) $J(A,B) = J(B,A)$,

and $J(A,B)$ does not satisfy the triangle inequality property of a metric. Therefore, $J(A,B)$ can be called pseudometric. It may be noticed that if we take $B = A_F$ (the most fuzzy set) i.e., $\mu_B(x_i) = \frac{1}{2}; \forall\, i$, then from (1.6.3) and (1.5.3) we have

$$I(A, A_F) = n \ln 2 - \left[-\sum_{i=1}^{n} \mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log(1 - \mu_A(x_i)) \right]$$

i.e., $I(A, A_F) = H(A_F) - H(A)$.

This informative distance between A and A_F gives a measure of nonfuzziness in the set A.

1.7 Fuzzy Information Improvement Measure

The probabilistic measure of information improvement, suggested by Theil (1967), is given by

$$D(P:Q) - D(P:R) = \sum_{i=1}^{n} p_i \log \frac{r_i}{q_i} \tag{1.7.1}$$

where P and Q are observed and predicted probability distributions respectively of a random variable, and $R = (r_1, r_2, \ldots, r_n)$ is the revised probability distribution of Q.

Similarly, suppose the correct fuzzy set is A and originally our estimate for it was the fuzzy set B that was revised to set C, the original ambiguity was $I(A,B)$ and finally ambiguity is $I(A,C)$. Analogously, the reduction in ambiguity is given by

$$I(A,B) - I(A,C) = \sum_{i=1}^{n} \left[\mu_A(x_i) \log \frac{\mu_C(x_i)}{\mu_B(x_i)} + (1 - \mu_A(x_i)) \log \frac{(1 - \mu_C(x_i))}{(1 - \mu_B(x_i))} \right], \tag{1.7.2}$$

which may be called fuzzy information improvement measure.

1.8 Comparison of Probabilistic and Fuzzy Information Measures

Fuzziness is often confused with probability. A statement is probabilistic if it expresses a likelihood or degree of certainty or if it is the outcome of clearly defined but randomly occurring events i.e., probability measures the likelihood of a future event based on something known now. On the other hand, fuzziness describes the lack of distinction of an event, whereas chance describes the uncertainty in the occurrence of the event. In other words, probability is the theory of random events and is not capable of capturing uncertainty resulting from vagueness of linguistic terms.

Next, we enumerate the similarities and dissimilarities between the two types of measures - fuzzy entropy and probabilistic entropy.

1.8.1 Similarities

(i) For all probability distributions $0 \leq p_i \leq 1$ for each i and for every fuzzy set $0 \leq \mu_A(x_i) \leq 1$; for each i.

(ii) The probabilistic entropy measures the closeness of the probability distribution $P(p_1, p_2, \ldots, p_n)$ with uniform distribution $\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ and fuzzy entropy measures the closeness of fuzzy distribution with the most fuzzy vector distribution $\left(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}\right)$.

(iii) Probabilistic and fuzzy entropies are concave functions of $p_1, p_2, \ldots, p_n$ and $\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n)$ respectively. Starting with any values of $\mu_A(x_1)$, $\mu_A(x_2), \ldots, \mu_A(x_n)$ and approaching the vector $\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}$, the fuzzy entropy will increase. Also, starting with any probability vector $p_1, p_2, \ldots, p_n$ and approaching the vector $\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}$, the probabilistic entropy will increase.

(iv) Probabilistic directed divergence measures and fuzzy directed divergence measures are convex functions with minimum value zero. Further, analogous to each measure of probabilistic entropy and directed divergence, we have measure of fuzzy entropy and fuzzy directed divergence.

1.8.2 Dissimilarities

(i) While $\sum_{i=1}^{n} p_i = 1$ for all probability distributions, $\sum_{i=1}^{n} \mu_A(x_i)$ need not be equal to unity and it need not even be the same for all fuzzy sets i.e., the probabilities of $n-1$ outcomes will determine the probability of the n^{th} outcome, but the knowledge of fuzziness of $n-1$ elements gives no information about the fuzziness of the n^{th} element.

(ii) The probabilities p_i and $1-p_i$ make different contributions to probabilistic entropy. However, $\mu_A(x_i)$ gives the same degree of fuzziness as $1-\mu_A(x_i)$ because both are equidistant from $\frac{1}{2}$ and the crisp set values 0 and 1.

(iii) Most of the measures of probabilistic entropy are of the form $\sum_{i=1}^{n} f(p_i)$ while most measures of fuzzy entropy are of the form $\sum_{i=1}^{n} f(\mu_A(x_i)) + \sum_{i=1}^{n} f(1-\mu_A(x_i))$.

(iv) Similarly, for probability distributions $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$, most of the probabilistic directed divergence measures are of the form $\sum_{i=1}^{n} f(p_i, q_i)$ while for fuzzy sets A and B, the fuzzy directed divergence measures are of the form $\sum_{i=1}^{n} f(\mu_A(x_i),\, \mu_B(x_i)) + \sum_{i=1}^{n} f(1-\mu_A(x_i),\, 1-\mu_B(x_i))$.

1.9 Measures of 'Useful' Fuzzy Information and 'Useful' Fuzzy Directed Divergence

Though in many practical situations of probabilistic nature, subjective considerations play its own role, Shannon entropy does not take into account the effectiveness or importance of the events. Belis and Guisau (1968) considered qualitative aspect of information and attached a utility distribution $U = (u_1, u_2, ..., u_n)$, where $u_i > 0$ for each and is utility or importance of an event x_i whose probability of occurrence is p_i. In general, u_i is independent of p_i. They suggested that the occurrence of an event removes two types of uncertainty - the quantitative type related to its probability of occurrence and the qualitative type related to its utility (importance) for the fulfillment of some goal set by the experimenter.

Bhaker and Hooda (1993) obtained the generalized mean value characterization of the useful information measures for incomplete probability distributions:

$$H(P;U) = -\frac{\sum\limits_{i=1}^{n} u_i p_i \log p_i}{\sum\limits_{i=1}^{n} u_i p_i}, \tag{1.9.1}$$

and

$$H_\alpha(P;U) = \frac{1}{1-\alpha} \log \frac{\sum\limits_{i=1}^{n} u_i p_i^\alpha}{\sum\limits_{i=1}^{n} u_i p_i}; \quad \alpha \neq 1, \quad \alpha > 0. \tag{1.9.2}$$

Zadeh (1968) was the first to quantify the uncertainty associated with a fuzzy event in the context of a discrete probabilistic framework, who defined the (weighted) entropy of A with respect to (X, P) as

$$H(A,P) = -\sum_{i=1}^{n} \mu_A(x_i) p_i \log p_i, \tag{1.9.3}$$

where μ_A is the membership function of A and p_i is the probability of occurrence of x_i. It may be noted that the situation contains the different types of uncertainties, e.g., randomness, ambiguity and vagueness. $H(A, P)$ of a fuzzy event with respect to P is less than Shannon's entropy which is of P alone.

Let P and Q be two probability distributions of a random variable X having utility distribution U. Bhaker and Hooda (1993) characterized the following measure of 'useful' directed divergence:

$$D(P : Q : U) = \frac{\sum_{i=1}^{n} u_i p_i \log \frac{p_i}{q_i}}{\sum_{i=1}^{n} u_i p_i} \tag{1.9.4}$$

This measure (1.9.4) has been generalized by Hooda and Ram (2002).

Chapter 2

Restricted Fuzzy Linear Regression Model

2.1 Introduction

Regression analysis is an area of statistics that consists of finding a suitable relationship explaining the statistical dependence of a response variable, say Y, on a set of explanatory variables, say $X_1,\ X_2, \ldots, X_n$. The dependence is usually assumed to have a particular mathematical form with one or more parameters. The aim of regression analysis is to estimate the parameters on the basis of empirical data. In the crisp linear regression model, the parameters (regression coefficients are crisp) appear in a linear form i.e.,

$$Y = A_0 + A_1X_1 + A_2X_2 + \ldots + A_nX_n + \text{random error}. \qquad (2.1.1)$$

Once the coefficients $A_0,\ A_1,\ \ldots, A_n$ are evaluated from the observed data, the response variables can be estimated from any given set of $X_1,\ X_2, \ldots, X_n$ values.

Classical linear regression has many applications in problems which can occur in the following situations:

- Numbers of observations is inadequate (Small data set)

- Vagueness in the relationship between input and output variables

- Ambiguity of events or degree to which they occur

- Inaccuracy and distortion introduced by linearization

The motivation for developing fuzzy regression analysis results from the realization that sometimes the observations cannot be known or quantified exactly and we can only provide an approximate description of them or intervals to enclose them. For instance, "in measuring the influence of character size on the reading ability from a video display terminal, the reading ability of the subject depends on his/her eyesight, age, the environment, individual responses and even how tired is the individual. Some of these factors cannot be described accurately and these kinds of variables are described as fuzzy variables" (refer to Chang et al. (1996)). Since ambiguity and vagueness are found in human subjective appraisal or judgment, all response variable values or explanatory variables cannot be precisely measured in the actual state of things. Therefore, it may be sometimes inappropriate to use crisp values to represent different situations. Fuzzy set theory, developed by Zadeh (1965) has capability to describe the uncertain situations, containing ambiguity and vagueness. The concept of fuzzy regression was first introduced by Tanaka, Uejima and Asai (1980, 1982) and its general form is given by

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 X_1 + \ldots + \tilde{A}_n X_n, \tag{2.1.2}$$

where the value of the output variable defined by (2.1.2) is a fuzzy number $\tilde{Y}$, $\tilde{A}_0$, $\tilde{A}_1$, $\ldots, \tilde{A}_n$ are fuzzy regression coefficients and X_1, $X_2, \ldots, X_n$ are real valued input (explanatory) variables. We assume the regression coefficients to be symmetric triangular fuzzy numbers (refer to Dubois and Prade (1980) and Zimmerman (1985)) defined by the membership function

$$\mu_{\tilde{A}_j(c)} = \begin{cases} 1 - \frac{|c-c_j|}{r_j} & \text{when } c_j - r_j \le c \le c_j + r_j \\ 0 & \text{otherwise,} \end{cases} \tag{2.1.3}$$

where c_j is known as the middle value for which $\tilde{A}_j(c_j) = 1$ and $r_j > 0$ is the spread of $\tilde{A}_j$. The fuzzy number $\tilde{A}_j$ expresses c_j or around c_j in the linguistic terms and it is denoted by $\tilde{A}_j = (c_j, r_j)$; $\forall j \in \{1, 2, \ldots, n\}$. The spread denotes the fuzziness of the membership function.

The aim of fuzzy regression model is to find the regression coefficients $\tilde{A}_0$, $\tilde{A}_1$, $\ldots, \tilde{A}_n$ such that the fuzzy linear function (2.1.2) fits the given fuzzy data as best as possible. Tanaka et al. (1980, 1982) initiated the study of fuzzy linear regression and used linear programming technique to find fuzzy regression coefficients for the case of fuzzy dependent variable and crisp independent variables. Subsequently, several developments have been carried out by Chang and Lee (1994a-c, 1996), Diamond (1988, 1990), Kim et al. (1996), Peters (1994), Tanaka and Watada (1988). Based on various studies fuzzy regression may be roughly divided into two approaches - linear programming based methods and fuzzy least square methods. The fuzzy regression models have following categories depending on the nature of the variables:

(i) Both input and output variables are crisp.

(ii) Input variable is crisp but output variable is fuzzy .

(iii) Both input and output variables are fuzzy.

It may be noted that in classical linear regression, the difference between the observed values and the values estimated from the model is because of the random errors. Upper and lower bounds for the estimated values are established and the probability that the estimated values will be with in these two bounds represents the confidence of the estimate. In other words, classical regression analysis is probabilistic. However, in fuzzy regression, the difference between the observed and the estimated values is because of ambiguity inherently present in the system. The output for a specified input is assumed to be a range of possible values i.e., the output can take any of these possible values. Therefore, the fuzzy regression is possibilistic in nature. In a fuzzy regression there is a tendency that the greater the values of independent variables, the wider is the spread of dependent variables. The value of the center point of estimated fuzzy output may be either greater than the value of the right endpoint or smaller than the value of the left endpoint. A fuzzy least square approach directly uses information included in the input-output data set and considers the measure of best fitting based on distance under fuzzy consideration. Fuzzy least squares are fuzzy extensions of ordinary least squares. Here, we have used least square approach for estimators or parameters estimation for the doubly adaptive linear regression model (refer to D'Urso, P. and T. Gastaldi, (2000)).

In many situations, some prior information is available about unknown regression coefficients. Such information can arise from past experience of the experimenter or other sources. Incorporation of such information results more efficient estimators in probabilistic linear regression models. Such prior information can be expressed in different forms. We assume that such prior information can be expressed in the form of exact linear restrictions on regression coefficients. How to incorporate such prior information in the consistent estimation of regression coefficients in a doubly fuzzy linear regression model is the subject matter of present chapter.

In Section 2.2, we introduce the fuzzy linear regression model. In Section 2.3, the restricted estimators of regression coefficients are obtained. In order to compare the properties of restricted estimator with the usual unrestricted estimator, we conduct a simulation study. The conclusions of simulation study are presented in Section 2.4 followed by some concluding remarks in Section 2.5.

2.2 The model

We consider symmetrical triangular fuzzy number (STFN) in particular, and assume the membership function to be triangular in shape and completely identified by the two parameters c (center) and r (left and right spread). Suppose that the observed fuzzy number is $Y_i = (c_i, r_i),\ i \in \{1, 2, \ldots, n\}$, which is inherited by the corresponding theoretical observation $Y_i^* = (c_i^*, r_i^*)$.

The proposed fuzzy regression model, also referred as *doubly linear adaptive fuzzy regression model*, is defined as follows:

$$\begin{aligned} \boldsymbol{c} &= \boldsymbol{c^*} + \boldsymbol{\varepsilon_c} \qquad \text{where } \boldsymbol{c^*} = X\boldsymbol{\beta}, \\ \boldsymbol{r} &= \boldsymbol{r^*} + \boldsymbol{\varepsilon_r} \qquad \text{where } \boldsymbol{r^*} = \boldsymbol{c^*}b + \mathbf{1}d, \end{aligned} \tag{2.2.1}$$

where X is a $n \times p$ matrix containing the input variables (data matrix), $\boldsymbol{\beta}$ is a column p-vector containing the regression parameters of the first model (called as core regression model), $\boldsymbol{c}$ and $\boldsymbol{c^*}$ are the n-vector of observed centers and the vector of the interpolated centers respectively, $\boldsymbol{r}$ and $\boldsymbol{r^*}$ are the (n-vector of the assigned spreads and the vector of the interpolated spreads respectively, $\boldsymbol{\varepsilon_c}$ and $\boldsymbol{\varepsilon_r}$ are the interpolation error vectors, $\mathbf{1}$ is a n-vector of all 1's, b and d are regression parameters for the second regression model (called as *spread regression model*).

The model postulated above comprise of two linear models: one interpolates the centers of the fuzzy observations, second one yields the spreads by building another linear model over the first one. Observe that the explanatory variables X, through the observed centers, can explicitly be written as $\boldsymbol{r} = X\boldsymbol{\beta}b + \mathbf{1}d + \varepsilon_r$. The model has the capability to take into account possible linear relations between the size of the spreads and the magnitude of the estimated centers.

Moreover, we assume that the regression coefficients $\boldsymbol{\beta}$ are subjected to j ($j < p$) exact linear restrictions, which are given by

$$\boldsymbol{v} = V\boldsymbol{\beta}, \tag{2.2.2}$$

where $\boldsymbol{v}$ and V are known and the matrix V is of full row rank.

2.3 Estimation of Regression Coefficients

Let us define the Euclidean distance between two symmetrical fuzzy numbers $Y_i = (c_i, r_i)$ and $Y_i^* = (c_i^*, r_i^*)$ as follows:

$$\delta_i \equiv \delta(Y_i, Y_i^*) = \sqrt{(c_i - c_i^*)^2 + (r_i - r_i^*)^2} \tag{2.3.1}$$

In order to obtain the estimator of regression coefficient $\boldsymbol{\beta}$ under the model (2.2.1), we minimize the score function i.e., the following sum of square errors (using notation in matrix form):

$$\begin{aligned}
\phi(\boldsymbol{\beta},\, b,\, d) &= \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} (c_i - c_i^*)^2 + \sum_{i=1}^{n} (r_i - r_i^*)^2 \\
&= (\boldsymbol{c} - \boldsymbol{c^*})'(\boldsymbol{c} - \boldsymbol{c^*}) + (\boldsymbol{r} - \boldsymbol{r}^*)'(\boldsymbol{r} - \boldsymbol{r}^*) \\
&= \boldsymbol{c}'\boldsymbol{c} - 2\boldsymbol{c}'\boldsymbol{c^*} + \boldsymbol{c^*}'\boldsymbol{c^*} + \boldsymbol{r}'\boldsymbol{r} - 2\boldsymbol{r}'\boldsymbol{r}^* + \boldsymbol{r}^{*\prime}\boldsymbol{r}^* \\
&= \boldsymbol{c}'\boldsymbol{c} - 2\boldsymbol{c}'X\boldsymbol{\beta} + \boldsymbol{\beta}'X'X\boldsymbol{\beta} + \boldsymbol{r}'\boldsymbol{r} - 2\boldsymbol{r}'(X\boldsymbol{\beta}b + \mathbf{1}d) \\
&\quad + (X\boldsymbol{\beta}b + \mathbf{1}d)'(X\boldsymbol{\beta}b + \mathbf{1}d) \\
&= \boldsymbol{c}'\boldsymbol{c} - 2\boldsymbol{c}'X\boldsymbol{\beta} + \boldsymbol{\beta}'X'X\boldsymbol{\beta} + \boldsymbol{r}'\boldsymbol{r} - 2\boldsymbol{r}'X\boldsymbol{\beta}b - 2\boldsymbol{r}'\mathbf{1}d \\
&\quad + \boldsymbol{\beta}'X'X\boldsymbol{\beta}b^2 + 2\boldsymbol{\beta}'X'\mathbf{1}bd + nd^2 \\
\Rightarrow \phi(\boldsymbol{\beta},\, b,\, d) &= \boldsymbol{c}'\boldsymbol{c} - 2\boldsymbol{c}'X\boldsymbol{\beta} + \boldsymbol{\beta}'X'X\boldsymbol{\beta}(1 + b^2) + \boldsymbol{r}'\boldsymbol{r} - 2\boldsymbol{r}'X\boldsymbol{\beta}b \\
&\quad - 2\boldsymbol{r}'\mathbf{1}d + 2\boldsymbol{\beta}'X'\mathbf{1}bd + nd^2
\end{aligned} \tag{2.3.2}$$

Differentiating $\phi(\boldsymbol{\beta},\, b,\, d)$ partially with respect to $\boldsymbol{\beta}$ and equating it to zero, we get

$$\begin{aligned}
& \frac{\partial \phi(\boldsymbol{\beta},\, b,\, d)}{\partial \boldsymbol{\beta}} = 0 \\
\Rightarrow\quad & -2X'\boldsymbol{c} + 2(1 + b^2)X'X\boldsymbol{\beta} - 2bX'\boldsymbol{r} + 2bdX'\mathbf{1} = 0 \\
\Rightarrow\quad & (1 + b^2)X'X\boldsymbol{\beta} = X'\boldsymbol{c} + bX'\boldsymbol{r} - 2bdX'\mathbf{1} \\
\Rightarrow\quad & \hat{\boldsymbol{\beta}} = \frac{1}{(1 + b^2)} (X'X)^{-1} X'(\boldsymbol{c} + \boldsymbol{r}b - \mathbf{1}bd)
\end{aligned} \tag{2.3.3}$$

Similarly, differentiating (2.3.2) partially with respect to b and d, we get

$$\begin{aligned} &\frac{\partial \phi(\boldsymbol{\beta},\, b,\, d)}{\partial b} = 0 \\ \Rightarrow \quad & 2b\boldsymbol{\beta}'X'X\boldsymbol{\beta} - 2\boldsymbol{r}'X\boldsymbol{\beta} + 2d\boldsymbol{\beta}'X'\mathbf{1} = 0 \\ \Rightarrow \quad & \hat{b} = (\boldsymbol{\beta}'X'X\boldsymbol{\beta})^{-1}\left[\boldsymbol{r}'X\boldsymbol{\beta} - \boldsymbol{\beta}'X'\mathbf{1}d\right] \end{aligned} \tag{2.3.4}$$

and

$$\begin{aligned} &\frac{\partial \phi(\boldsymbol{\beta},\, b,\, d)}{\partial d} = 0 \\ \Rightarrow \quad & -2\boldsymbol{r}'\mathbf{1} + 2\boldsymbol{\beta}'X'\mathbf{1}b + 2nd = 0 \\ \Rightarrow \quad & \hat{d} = n^{-1}\left(\boldsymbol{r}'\mathbf{1} - \boldsymbol{\beta}'X'\mathbf{1}b\right) \end{aligned} \tag{2.3.5}$$

respectively (refer to D'Urso, P. and T. Gastaldi (2000)).

Next, we assume that the regression coefficients $\boldsymbol{\beta}$ are subject to the linear restrictions which are given by (2.2.2). It may be noted that the estimator $\hat{\boldsymbol{\beta}}$ obtained above in (2.3.3) does not satisfy the given restrictions (2.2.2). We aim to obtain the restricted estimator of $\boldsymbol{\beta}$ which satisfies the given restriction under the doubly linear adaptive fuzzy regression model (2.2.1). For this, we propose to minimize the following score function

$$\begin{aligned} S(\boldsymbol{\lambda},\, \boldsymbol{\beta},\, b,\, d) &= \phi(\boldsymbol{\beta},\, b,\, d) - 2\boldsymbol{\lambda}(V\boldsymbol{\beta} - \boldsymbol{v}) \\ &= \boldsymbol{c}'\boldsymbol{c} - 2\boldsymbol{c}'X\boldsymbol{\beta} + \boldsymbol{\beta}'X'X\boldsymbol{\beta}(1+b^2) + \boldsymbol{r}'\boldsymbol{r} - 2\boldsymbol{r}'X\boldsymbol{\beta}b - 2\boldsymbol{r}'\mathbf{1}d \\ &\qquad + 2\boldsymbol{\beta}'X'\mathbf{1}bd + nd^2 - 2\boldsymbol{\lambda}(V\boldsymbol{\beta} - \boldsymbol{v}), \end{aligned} \tag{2.3.6}$$

where $\mathbf{2}\boldsymbol{\lambda}$ is the vector of Lagrangian multiplier.

Differentiating $S(\boldsymbol{\lambda},\, \boldsymbol{\beta},\, b,\, d)$ partially with respect to $\boldsymbol{\beta}$ and equating it to zero, we get

$$\begin{aligned} &\frac{\partial S(\boldsymbol{\lambda},\, \boldsymbol{\beta},\, b,\, d)}{\partial \boldsymbol{\beta}} = 0 \\ \Rightarrow \quad & -2X'\boldsymbol{c} + 2(1+b^2)X'X\boldsymbol{\beta} - 2bX'\boldsymbol{r} + 2bdX'\mathbf{1} - 2\boldsymbol{\lambda}V = 0 \\ \Rightarrow \quad & \tilde{\boldsymbol{\beta}} = \frac{1}{(1+b^2)}\left(X'X\right)^{-1}X'(\boldsymbol{c} + \boldsymbol{r}b - \mathbf{1}bd) + \frac{1}{(1+b^2)}\left(X'X\right)^{-1}V'\boldsymbol{\lambda} \\ \Rightarrow \quad & \tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \frac{1}{(1+b^2)}\left(X'X\right)^{-1}V'\boldsymbol{\lambda}. \end{aligned} \tag{2.3.7}$$

Similarly, differentiating $S(\boldsymbol{\lambda},\, \boldsymbol{\beta},\, b,\, d)$ partially with respect to $\boldsymbol{\lambda}$, we get

$$\begin{aligned} & \frac{\partial S(\lambda,\, \beta,\, b,\, d)}{\partial \lambda} = 0 \\ \Rightarrow \quad & V\hat{\boldsymbol{\beta}} + \frac{1}{1+b^2} V\left(X'X\right)^{-1} V'\boldsymbol{\lambda} = \boldsymbol{v} \\ \Rightarrow \quad & \hat{\boldsymbol{\lambda}} = (1+b^2)\Big[V\left(X'X\right)^{-1} V'\Big]^{-1}\Big(\boldsymbol{v} - V\hat{\boldsymbol{\beta}}\Big). \end{aligned} \tag{2.3.8}$$

From equation (2.3.7) and (2.3.8), we have

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \left(X'X\right)^{-1} V'\Big[V\left(X'X\right)^{-1} V'\Big]^{-1}\Big(\boldsymbol{v} - V\hat{\boldsymbol{\beta}}\Big). \tag{2.3.9}$$

Also, differentiating (2.3.6) partially with respect to b and d, we get

$$\begin{aligned} & \frac{\partial S(\boldsymbol{\lambda},\, \boldsymbol{\beta},\, b,\, d)}{\partial b} = 0 \\ \Rightarrow \quad & \hat{b} = \left(\boldsymbol{\beta}'X'X\boldsymbol{\beta}\right)^{-1}\left[\boldsymbol{r}'X\boldsymbol{\beta} - \boldsymbol{\beta}'X'\mathbf{1}d\right] \end{aligned} \tag{2.3.10}$$

and

$$\begin{aligned} & \frac{\partial S(\boldsymbol{\lambda},\, \boldsymbol{\beta},\, b,\, d)}{\partial d} = 0 \\ \Rightarrow \quad & \hat{d} = n^{-1}\left(\boldsymbol{r}'\mathbf{1} - \boldsymbol{\beta}'X'\mathbf{1}b\right) \end{aligned} \tag{2.3.11}$$

respectively.

We have

$$\begin{aligned} & V\tilde{\boldsymbol{\beta}} = V\hat{\boldsymbol{\beta}} + V\left(X'X\right)^{-1} V'\Big[V\left(X'X\right)^{-1} V'\Big]^{-1}\Big(v - V\hat{\boldsymbol{\beta}}\Big) \\ \Rightarrow \quad & V\tilde{\boldsymbol{\beta}} = V\hat{\boldsymbol{\beta}} + \Big(\boldsymbol{v} - V\hat{\boldsymbol{\beta}}\Big) = \boldsymbol{v}. \end{aligned}$$

Therefore, the estimator $\tilde{\boldsymbol{\beta}}$ satisfies the given restrictions (2.2.2).

2.4 Simulation Study

In the previous section, we obtained the restricted estimator $\tilde{\boldsymbol{\beta}}$ which satisfies the given linear restrictions (2.2.2). Since some additional information is being used in obtaining this estimator, it is expected that the performance of this estimator is better than the unrestricted estimator $\hat{\boldsymbol{\beta}}$ in "some" sense of dominance. In order to study these dominance properties, we conducted simulation study. For this, we adopted various values of $\boldsymbol{\beta}$ and generated the sample Y_i^*, Y_i and X of sizes 22 and 48. Using these samples, we obtain the estimators $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ empirically. To study the properties of these estimators, we obtained mean squared error matrices (MSEM) and absolute bias (AB) of these estimators empirically for 50,000 repetitions. The absolute bias is defined as the positive square root of the some of squares of the elements of bias vector. In order to save space, we below present a few outcomes of simulation study.

(i) When $\boldsymbol{\beta} = \begin{pmatrix} 0.5 \\ 15.0 \\ -3.2 \\ 1.0 \end{pmatrix}$

- For $n = 22$

$$\text{MSEM}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} 0.1477 & 4.2766 & -0.9060 & 0.2900 \\ 4.2766 & 123.9044 & -26.2498 & 8.4013 \\ -0.9060 & -26.2498 & 5.5612 & -1.7799 \\ 0.2900 & 8.4013 & -1.7799 & 0.5697 \end{pmatrix},$$

$$\text{MSEM}(\tilde{\boldsymbol{\beta}}) = \begin{pmatrix} 5.0486 & -5.1027 & 7.3056 & -1.8279 \\ -5.1027 & 5.1574 & -7.3839 & 1.8475 \\ 7.3056 & -7.3839 & 10.5718 & -2.6450 \\ -1.8279 & 1.8475 & -2.6450 & 0.6618 \end{pmatrix},$$

$$\mathrm{AB}(\hat{\boldsymbol{\beta}}) = 11.4097 \text{ and } \mathrm{AB}(\tilde{\boldsymbol{\beta}}) = 3.8634.$$

- For $n = 48$

$$\mathrm{MSEM}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} 0.1475 & 4.2742 & -0.9056 & 0.2898 \\ 4.2742 & 123.9007 & -26.2508 & 8.3995 \\ -0.9056 & -26.2508 & 5.5618 & -1.7796 \\ 0.2898 & 8.3995 & -1.7796 & 0.5694 \end{pmatrix},$$

$$\mathrm{MSEM}(\tilde{\boldsymbol{\beta}}) = \begin{pmatrix} 4.1725 & -4.2172 & 6.0379 & -1.5107 \\ -4.2172 & 4.2624 & -6.1026 & 1.5269 \\ 6.0379 & -6.1026 & 8.7373 & -2.1861 \\ -1.5107 & 1.5269 & -2.1861 & 0.5469 \end{pmatrix},$$

$$\mathrm{AB}(\hat{\boldsymbol{\beta}}) = 11.4096 \text{ and } \mathrm{AB}(\tilde{\boldsymbol{\beta}}) = 3.8652.$$

(ii) When $\boldsymbol{\beta} = \begin{pmatrix} 3.0 \\ 3.5 \\ 4.0 \\ 4.5 \end{pmatrix}$

- For $n = 22$

$$\mathrm{MSEM}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} 5.0047 & 5.8338 & 6.6626 & 7.4918 \\ 5.8338 & 6.8004 & 7.7665 & 8.7331 \\ 6.6626 & 7.7665 & 8.8699 & 9.9738 \\ 7.4918 & 8.7331 & 9.9738 & 11.2151 \end{pmatrix},$$

$$\mathrm{MSEM}(\tilde{\boldsymbol{\beta}}) = \begin{pmatrix} 4.0907 & -4.1346 & 5.9196 & -1.4811 \\ -4.1346 & 4.1789 & -5.9830 & 1.4969 \\ 5.9196 & -5.9830 & 8.5660 & -2.1432 \\ -1.4811 & 1.4969 & -2.1432 & 0.5362 \end{pmatrix},$$

$$\text{AB}(\hat{\boldsymbol{\beta}}) = 5.6471 \text{ and } \text{AB}(\tilde{\boldsymbol{\beta}}) = 3.7744.$$

- For $n = 48$

$$\text{MSEM}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} 5.0041 & 5.8330 & 6.6619 & 7.4910 \\ 5.8330 & 6.7994 & 7.7655 & 8.7319 \\ 6.6619 & 7.7655 & 8.8690 & 9.9727 \\ 7.4910 & 8.7319 & 9.9727 & 11.2138 \end{pmatrix},$$

$$\text{MSEM}(\tilde{\boldsymbol{\beta}}) = \begin{pmatrix} 3.6623 & -3.7016 & 5.2997 & -1.3260 \\ -3.7016 & 3.7413 & -5.3565 & 1.3402 \\ 5.2997 & -5.3565 & 7.6690 & -1.9188 \\ -1.3260 & 1.3402 & -1.9188 & 0.4801 \end{pmatrix},$$

$$\text{AB}(\hat{\boldsymbol{\beta}}) = 5.6467 \text{ and } \text{AB}(\tilde{\boldsymbol{\beta}}) = 3.7723.$$

(iii) When $\boldsymbol{\beta} = \begin{pmatrix} -0.5 \\ -23.0 \\ -2.0 \\ -4.5 \end{pmatrix}$

- For $n = 22$

$$\text{MSEM}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} 0.1275 & 6.0813 & 0.5244 & 1.1859 \\ 6.0813 & 290.1406 & 25.0177 & 56.5803 \\ 0.5244 & 25.0177 & 2.1572 & 4.8787 \\ 1.1859 & 56.5803 & 4.8787 & 11.0338 \end{pmatrix},$$

$$\text{MSEM}(\tilde{\boldsymbol{\beta}}) = \begin{pmatrix} 5.7660 & -5.8278 & 8.3438 & -2.0876 \\ -5.8278 & 5.8903 & -8.4332 & 2.1100 \\ 8.3438 & -8.4332 & 12.0741 & -3.0209 \\ -2.0876 & 2.1100 & -3.0209 & 0.7558 \end{pmatrix},$$

$$\mathrm{AB}(\hat{\boldsymbol{\beta}}) = 17.4200 \text{ and } \mathrm{AB}(\tilde{\boldsymbol{\beta}}) = 1.5484.$$

- For $n = 48$

$$\mathrm{MSEM}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} 0.1276 & 6.0839 & 0.5247 & 1.1865 \\ 6.0839 & 290.1452 & 25.0214 & 56.5832 \\ 0.5247 & 25.0214 & 2.1578 & 4.8796 \\ 1.1865 & 56.5832 & 4.8796 & 11.0347 \end{pmatrix},$$

$$\mathrm{MSEM}(\tilde{\boldsymbol{\beta}}) = \begin{pmatrix} 2.8180 & -2.8482 & 4.0779 & -1.0203 \\ -2.8482 & 2.8788 & -4.1216 & 1.0312 \\ 4.0779 & -4.1216 & 5.9010 & -1.4764 \\ -1.0203 & 1.0312 & -1.4764 & 0.3694 \end{pmatrix},$$

$$\mathrm{AB}(\hat{\boldsymbol{\beta}}) = 17.4202 \text{ and } \mathrm{AB}(\tilde{\boldsymbol{\beta}}) = 1.5954.$$

On the basis of simulation outcomes, we make following conclusions:

(i) The absolute bias of unrestricted estimator $\hat{\boldsymbol{\beta}}$ and restricted estimator $\tilde{\boldsymbol{\beta}}$ are almost the same for the sample sizes 22 and 48. It seems that the bias of these estimators establish for the small sample size of 22.

(ii) The absolute bias of the restricted estimator $\tilde{\boldsymbol{\beta}}$ is much less than that of unrestricted estimator $\hat{\boldsymbol{\beta}}$.

(iii) Looking at the MSEMs of these estimators, it is clear that the mean squared errors (MSEs) (defined as the trace of MSEM) decrease as sample size increases. However, this decrement is very small.

(iv) The MSE of restricted estimator $\tilde{\boldsymbol{\beta}}$ is much less than that of unrestricted estimator $\hat{\boldsymbol{\beta}}$.

Thus, we conclude that the restricted estimator $\tilde{\boldsymbol{\beta}}$ is better than the unrestricted estimator $\hat{\boldsymbol{\beta}}$ with regard to absolute bias and mean squared error. Therefore, when some linear restrictions are available a priori on β, the use of the $\tilde{\boldsymbol{\beta}}$ is advisable over $\hat{\boldsymbol{\beta}}$ as an estimator of $\boldsymbol{\beta}$.

2.5 Conclusion

It is worth mentioning that under a doubly adaptive linear fuzzy regression model, where the regression coefficients are subjected to some exact linear restrictions, the obtained restricted estimator of regression coefficients are better than the usual unrestricted estimator with regard to mean squared error and absolute bias.

Chapter 3

Fuzzy Clustering Algorithm and its Application in Economics

3.1 Introduction

According to Goldman Sachs Report (2003), by the end of 2050, the BRIC (Brazil, Russia, India and China) countries will emerge as a very strong economy of the world and together they will be greater in absolute size than that of G6 (US, Japan, UK, Germany, France and Italy) countries. The report highlights key features of these economies and their growing contribution to world output and trade. It is also mentioned that, in absolute sense China will be number one economy, US will be number two and India number three, however, US will remain at the top on per capita basis. In this context, the projected convergence of per capita GDP of the G6 and BRIC countries is being studied using clustering analysis.

Clustering analysis is an important human activity. Clustering is basically a process of grouping a set of physical or abstract objects into classes of similar objects. The objective of such a cluster analysis is to partition the data set into a number of natural and homogeneous subsets, where the elements of each subset are as similar to each other as possible, and at the same time as different from those of the other sets as possible. A cluster of data objects can be treated collectively as one group in many applications. In other words, clustering involves the task of dividing data points into homogeneous classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible. Clustering can also be thought of a form of data compression, when a large number of samples are converted into a small number of representative prototypes or clusters.

Let $X \in \mathbb{R}^{m \times n}$ a set of data items representing a set of m points x_i in $\mathbb{R}^n$. The goal is to partition X into K groups C_k such that every data that belong to the same cluster are more "alike" than data in different groups. Each of K groups is called a cluster. The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis.

It is interesting to have the comparative study of projected per capita GDP convergence analysis of G6 and BRIC countries, based on the projected GDP and per capita GDP by Goldman Sachs Report (2003). To test the convergence, a comparatively new (clustering) technique which is becoming popular day by day among economists fraternity is used. This is called as fuzzy c-means algorithm. The algorithm is a least square function, when we have the number of data sets and the number of classes (partitions) into which the data sets are classified.

Fuzzy Clustering algorithms provide a fuzzy description of the discovered structure. The main advantage of this description is that it captures the imprecision encountered when describing real-life data. Thus, the user is provided with more information about the structure in the data compared to a crisp, non-fuzzy scheme. In fuzzy clustering, the objects of the universe of discourse are not classified as belonging to one and only one cluster, but instead, they all possess a degree of membership with each of the clusters. The most widely used fuzzy clustering algorithm is fuzzy c-means (FCM). The concept of fuzzy set was initiated by Lotfi A. Zadeh (1965) who was professor of computer science at the University of California in Berkeley. The idea of using fuzzy set theory for clustering is credited to Ruspini (1969, 1970). The first specific formulation of FCM is due to Dunn (1973), but its generalization and current framing is credited to Bezdek (1981).

3.2 Formation of Economies and Their Contribution

The historical review reveals that World War II hampered the world economy and its immediate consequences to the world were disturbed economic system, abated economic growth and poor international relations. That compelled the then economic powers to think constructively and to help the world to grow peacefully. This resulted in the formation of General Agreement On Tariffs And Trade (GATT) in 1948 and that showed their commitment towards the development of world economy. India was also one of the founder members of GATT having the same objective. United States emerged as one of the strongest economy and considered itself responsible for the economic growth of the entire world. During the oil crisis US initiated to form a group of six industrialized nations in 1973 and became successful to do so in 1975. The group formed was called as Group6, or simply G6 and member countries were – United States, United Kingdom, Japan, Germany, Italy and France. The group was economically very strong and together they constituted about 53 percent of world's GDP, at the time of creation. The economy grew continuously for more than two decades and its share in the world's GDP increased upto 65 percent in the year 1999. From the advent of new millennium, it has been found that their contribution is decreasing due to emergence of few other economies of the world considerably. In 2005 their share in world's GDP was 59 percent, which can be verified from the data compiled in Table 3.1.

At the time of formation of GATT the economic condition of India was not good and more than 50 percent of the population were living below poverty line. It was one of the least developing economies. After independence India adopted globalization and liberalization. Consequently its economy transformed entirely and is the third largest economy, after US and China on PPP (Purchasing Power Parity) basis. It is going to play a vital role in the world economy in the coming future.

The other economies, which showed phenomenal growth in the past few years and have potential to grow further, are mainly China, Canada, Mexico, Spain, Russia and Brazil. It has also been mentioned in Goldman Sachs Report (2003) that by the end of 2050, four countries of the world namely – Brazil, Russia, India and China, also called as BRIC countries, will grow very fast and emerge as dominant economies. If we look at the historical developments of the BRIC nations, we will find that these nations are not much developed. The World Bank has described India as "Low Income Country" while Brazil, Russia and China have been classified as "Lower Middle Income Countries".

The pattern of GDP growth rate of BRIC countries, G6 countries and World GDP can be analyzed from Table 3.1 and Figure 3.1. Table 3.1 shows the percentage contribution of BRIC and G6 economies into World GDP. It clearly shows that the contribution of BRIC economies decreased continuously from 1980 to 1999, but there is change in the pattern from 2000 onwards. This shows the gradual increase in the share of World GDP by BRIC economies. On the other hand, the contribution of G6 countries is appreciable from 1980 to 2000, but there is a significant change in the pattern from 2001 onwards. This shows the gradual decrease in the share of World GDP by G6 economies.

It may be noted that the sharp decline in contribution of BRIC economies in world's GDP in the year 1990 was due to the disintegration of erstwhile USSR into Russia.

Figure 3.1 clearly shows that from 2000 onwards the GDP growth rate of BRIC economies has surpassed the World's GDP growth rate and that fascinated the economists of the World to realize the potential of BRIC economies. This probably might be one of the reasons which inspired the Goldman Sachs to work upon BRIC economies.

Goldman Sachs Report (2003) is mainly based upon the model of capital accumulation and productivity growth. It has projected the figures of GDP growth, per capita income and currency movements in the BRIC economies until 2050. It has also predicted that by the end of 2050, the GDP of BRICs economies will be \$84,201 billion as compared to the GDP of G6 economies which will be \$54,433 billion. It further says that, in less than 40 years, the BRIC economies together could be larger than G6 in US dollar terms. From Table 3.1, we can see that in 2005 the total GDP of BRIC countries was about 16.7 percent of the GDP of G6. If we go by the report, the BRIC economies will be half of the size of G6 by 2025 and it will overtake G6 by 2040.

According to the above mentioned report, Brazil and Russia would dominate the world market in the supply of raw materials, while India and China would dominate in services and manufacturing. However, inspite of such a phenomenal growth, the BRIC countries would not be able to translate it into proportionate improvement in living standards, and per capita income of most of the BRIC countries (except Russia), would be below than that of G6 countries. After 1990, there is a very hot issue among the economists from all over the world whether or not per capita income across countries is converging. The neo-classical growth model tells "economies will converge towards their balanced growth paths where per capita growth is inversely related to the starting level of per capita income".

Early studies by Baumaol (1986), Barro (1991, 1992), Sala-i-Martin (1996), etc. gave their opinion in favor of convergence. They believed that convergence across countries will occur in the coming future with an average rate of 2 percent per year. However, Quah (1993, 1996) had doubt about the figure of 2 percent for convergence and said that convergence will occur in relatively homogenous convergence clubs i.e., relatively homogenous economies. McCoskey (2002) suggests that convergence clubs and relative homogeneity is probably unresolved with respect to less developed countries (LDCs) where geographic proximity and cross-national economic interdependence will cause group of LDCs to grow or falter as one. But according to Dobson and Ramlogan (2002), little is known about the convergence process among LDCs and a limited range of studies that have considered LDCs have proceeded at a highly aggregated level (Khan and Kumar 1993) or have focused on convergence within a particular country (Ferreira (2000), Nagraj et al. (2000), Choi and Li (2000)). Economic convergence among countries has also been studied by Giles (2001) and Holmes (2004).

In spite of so much development, the relation between trade openness and convergence is still not established as no theory has been proposed to get a clear relationship between them. Moreover, the objective of our work is not to link the trade openness and convergence among BRIC countries or G6 because inter-regional trades among these countries are very less. Next, we will have a comparative analysis in terms of the projected convergence of per capita GDP of BRIC countries and G6 till year 2050, based on the GDP and per capita GDP projections, by Goldman Sachs Report (2003).

3.3 Fuzzy Clustering - The Algorithm

The goal of traditional clustering is to assign each data point to one and only one cluster. In contrast, fuzzy clustering assigns different degrees of membership to each point. The membership of a point is thus shared among various clusters. This creates the concept of fuzzy boundaries which differs from the traditional concept of well-defined boundaries. Bezdek (1981) asserts that the well-defined boundary model usually does not reflect the description of real data. This assertion led him to develop a new family of clustering algorithms based on a fuzzy extension of the least-square error criterion.

Using Fuzzy c-means algorithm we determine the partitioning of the per capita GDP of BRIC and G6 countries into a number of clusters, year by year, as projected in Goldman Sachs Report (2003). These clusters have "fuzzy" boundaries, in the sense that each data value belongs to each cluster to some degree or other specified by a membership grade, and thus each point may belong to several clusters.

The FCM algorithm partitions a collection of "n" data points into c fuzzy clusters (where $c < n$), and simultaneously seeking the best possible locations of these clusters. This method uses distance concept in n-dimensional Euclidean space to determine the geometric closeness of data points by assigning them to various clusters or classes.

The mathematical notations used in developing FCM algorithm:

$x_k = k^{\text{th}}$ data point (possibly m-dimensional vector and $k = 1,\ 2,\ \ldots, n$).

$v_i =$ the center of the i^{th} fuzzy cluster ($i = 1,\ 2,\ \ldots, c$).

$d_{ik} = \|x_k - v_i\|_2 = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij})^2\right]^{1/2}$ is the distance between x_k and v_i.

The partitioned clusters are typically defined by a $c \times n$ matrix M, called the membership matrix, where each element μ_{ik} (the degree of membership of the k^{th}data point in the i^{th} cluster) in the range of [0, 1] i.e.,

$$M = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_{c1} & \mu_{c2} & \cdots & \mu_{cn} \end{bmatrix}, \quad \text{where} \quad \sum_{i=1}^{c} \mu_{ik} = 1. \tag{3.3.1}$$

Our objective is to minimize the functional

$$J(M,\, v_1, v_2, ..., v_c) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^q (d_{ik})^2, \tag{3.3.2}$$

where $q \in [\,1,\, \infty)$ is a weighting exponent parameter and it controls the extent of membership sharing between fuzzy clusters. For the case $q = 1$, FCM algorithm approaches to a hard c-means algorithm. In general, the larger q is, the fuzzier are the membership assignments of the clustering convergence of the algorithm which tends to be slower as the value of q increases. In practice, $q = 2$ is a preferred choice.

The following are two necessary conditions for J to reach a minimum:

$$v_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^q . x_k}{\sum_{k=1}^{n} (\mu_{ik})^q} \tag{3.3.3}$$

and

$$\mu_{ik} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{q-1}} \right]^{-1} \tag{3.3.4}$$

The fuzzy c-means algorithm due to Bezdek (1981) is simply iteration through the preceding two conditions and it determines the cluster centers v_i and the membership matrix M using the following steps:

Input: dataset D, number of clusters

Output: set of c clusters

Algorithm:

Initialize: initialize the membership matrix M with random values between 0 and 1 with in the constraints of equation (3.3.1).

Iteration: at each step k,
compute the c cluster centers using equation (3.3.3) and
compute the objective function according to equation (3.3.2).

Termination: the algorithm will terminate if objective function according to equation (3.3.2) is below a certain threshold level $\epsilon = 0.01$.

Once the centers of the fuzzy clusters have been determined, each of the "n" data points can be allocated to the cluster with the closest center. In our application, we have two groups of countries – BRIC and G6, the former comprises of four countries while later of six countries. We have applied the algorithm (using computer programming in C) for both the groups individually by taking number of clusters as $c = 2$ and $c = 3$.

3.4 Results of Cluster Convergence Analysis

Table 3.1 shows GDP (at market price, current price) of World, G6 and BRIC countries from 1976 to 2005 illustrating world's GDP, absolute GDP of G6 and BRIC countries as well as their share in world GDP. The figures clearly show that, initially the contribution of G6 countries in world GDP has increased, but in the later years, specially after 2002, it started decreasing; whereas the contribution of BRIC countries in world GDP shows just opposite pattern, it has decreased in the initial years but started increasing after 2002. Here, it may be noted that there is sudden decrease in GDP contribution of BRIC countries in 1990 due to the disintegration of erstwhile USSR.

The results of our projected convergence analysis are based on the fuzzy clustering algorithm and summarized in the form of figures and tables. Figure 3.2 and 3.3 shows the relative rise of per capita GDP of BRIC and G6 countries respectively. These figures have been plotted from Table 3.2. Figure 3.2 shows that per capita GDP of all the BRIC countries are increasing with respect to time, while for Brazil it shows the relative fall in initial period (2000 to 2004) and then rise in subsequent years. On the other hand, Figure 3.3 shows the per capita GDP of the G6 countries are increasing with respect to time, except for Japan, where the growth was slow in initial years and then increased subsequently.

Figures 3.4 and 3.5 show the ratio of the centers of the two fuzzy clusters ($c = 2$), year by year, for BRIC and G6 countries respectively. From Figure 3.4, we find that in the beginning, the ratio is high and fluctuating while it becomes smooth and tending towards one for later years. While for G6 countries, Figure 3.5, the ratio fluctuates in a very small range of 1.46 to 1.58, and in the later years it becomes almost constant. These unitless ratios are used because it helps in facilitating inter-year comparisons.

Figures 3.6 and 3.7 show ratio of the center of the highest cluster to that of the lowest ($c = 3$), year by year, for BRIC and G6 countries respectively, which confirms the results of Figures 3.4 and 3.5 and show the same pattern of convergence.

The case of classification into two clusters i.e., $c = 2$ is not shown by tables, while for $c = 3$, the classification is shown by Table 3.3 and Table 3.4. This is because, in the earlier case, there is not much variation in the membership of the countries in the clusters. In our analysis, we find that, for BRIC countries, from 2000 to 2025, India and China are in one cluster and Brazil and Russia in the other. Whereas from 2026 to 2050, Brazil, India and China are in one cluster and Russia remains isolated in the other cluster. On the other hand, for G6 countries, USA and Japan are in one cluster, and France, Germany, Italy and UK are in the other, throughout the years.

Tables 3.3 and 3.4 show the classification of BRIC and G6 countries into three clusters ($c = 3$) respectively. From Table 3.3, variation in the membership of countries in different clusters can be observed for initial years upto 2008 for BRIC countries; and after that Brazil and China are in one cluster, whereas Russia and India are isolated and remain in two different clusters. From Table 3.4, significant variation can be observed upto 2032 for G6 countries, and after that Japan and UK are in one cluster; France, Germany and Italy are in second; while USA remains isolated in the third cluster.

From our analysis and algorithm, in 2008, 2021 and 2032 Japan shows maximum ambiguity for its membership ($\mu_A(x) \approx 0.5$) which leads to shift its cluster from one to another.

Table 3.3: Composition of Fuzzy Clusters for BRIC Countries ($c = 3$)

Year	C1	C2	C3	Year	C1	C2	C3	Year	C1	C2	C3
2000	B, R	CH	I	2017	R	I	B,CH	2034	R	I	B,CH
2001	B, R	CH	I	2018	R	I	B,CH	2035	R	I	B,CH
2002	B, R	CH	I	2019	R	I	B,CH	2036	R	I	B,CH
2003	R	I, CH	B	2020	R	I	B,CH	2037	R	I	B,CH
2004	R	I, CH	B	2021	R	I	B,CH	2038	R	I	B,CH
2005	R	I, CH	B	2022	R	I	B,CH	2039	R	I	B,CH
2006	R	I, CH	B	2023	R	I	B,CH	2040	R	I	B,CH
2007	R	I, CH	B	2024	R	I	B,CH	2041	R	I	B,CH
2008	R	I, CH	B	2025	R	I	B,CH	2042	R	I	B,CH
2009	R	I	B,CH	2026	R	I	B,CH	2043	R	I	B,CH
2010	R	I	B,CH	2027	R	I	B,CH	2044	R	I	B,CH
2011	R	I	B,CH	2028	R	I	B,CH	2045	R	I	B,CH
2012	R	I	B,CH	2029	R	I	B,CH	2046	R	I	B,CH
2013	R	I	B,CH	2030	R	I	B,CH	2047	R	I	B,CH
2014	R	I	B,CH	2031	R	I	B,CH	2048	R	I	B,CH
2015	R	I	B,CH	2032	R	I	B,CH	2049	R	I	B,CH
2016	R	I	B,CH	2033	R	I	BCH	2050	R	I	BCH

Note: Ci denotes cluster "i" ; B = Brazil; R = Russia; I = India; CH = China.

Table 3.4: Composition of Fuzzy Clusters for G6 Countries ($c = 3$)

Year	C1	C2	C3	Year	C1	C2	C3	Year	C1	C2	C3
2000	J, U	F, G, K	I	2017	U	F, G, I	J, K	2034	U	J, K	F, G, I
2001	J, U	F, G, K	I	2018	U	F, G, I	J, K	2035	U	J, K	F, G, I
2002	J, U	F, G, K	I	2019	U	F, G, I	J, K	2036	U	J, K	F, G, I
2003	J, U	K	F, G, I	2020	U	F, G, I	J, K	2037	U	J, K	F, G, I
2004	J, U	K	F, G, I	2021	J, U	F, G, I	K	2038	U	J, K	F, G, I
2005	J, U	K	F, G, I	2022	J, U	F, G, I	K	2039	U	J, K	F, G, I
2006	J, U	F, G, I	K	2023	J, U	F, G, I	K	2040	U	J, K	F, G, I
2007	J, U	F, G, I	K	2024	J, U	K	F, G, I	2041	U	J, K	F, G, I
2008	J, U	F, G, I	K	2025	J, U	K	F, G, I	2042	U	J, K	F, G, I
2009	U	F, G, I	J, K	2026	J, U	K	F, G, I	2043	U	J, K	F, G, I
2010	U	F, G, I	J, K	2027	J, U	K	F, G, I	2044	U	J, K	F, G, I
2011	U	F, G, I	J, K	2028	J, U	K	F, G, I	2045	U	J, K	F, G, I
2012	U	F, G, I	J, K	2029	J, U	K	F, G, I	2046	U	J, K	F, G, I
2013	U	F, G, I	J, K	2030	J, U	K	F, G, I	2047	U	J, K	F, G, I
2014	U	F, G, I	J, K	2031	J, U	K	F, G, I	2048	U	J, K	F, G, I
2015	U	F, G, I	J, K	2032	J, U	K	F, G, I	2049	U	J, K	F, G, I
2016	U	F, G, I	J, K	2033	U	J, K	F, G, I	2050	U	J, K	F, G, I

Note: Ci denotes cluster "i" ; F = France; G = Germany; I = Italy; J = Japan; U = U.S.A; K = U.K.

3.5 Conclusion

The issue of projected convergence of per capita GDP among the G6 and BRIC countries using the Fuzzy c-means clustering algorithm is discussed. We find that by 2050, per capita GDP of all the four countries of BRIC will tend towards convergence. On the other hand, the degree of convergence is less among G6 countries but they show stable growth pattern. We have tested the convergence only by FCM algorithm, while the same can be tested by different classification techniques and the results may be compared. Moreover, the accuracy of this work depends on the accuracy of the data projected by Goldman Sachs Report (2003).

Chapter 4

Fuzzy Information Measures

4.1 Introduction

Uncertainty and fuzziness are the basic nature of human thinking and of many real world objectives. Fuzziness is found in our decision, in our language and in the way we process information. The main use of information is to remove uncertainty and fuzziness. In fact, we measure information supplied by the amount of probabilistic uncertainty removed in an experiment and the measure of uncertainty removed is also called as a measure of information while measure of fuzziness is the measure of vagueness and ambiguity of uncertainties.

Shannon (1948) used "entropy" to measure uncertain degree of the randomness in a probability distribution. Let X is a discrete random variable with probability distribution $P = (p_1, p_2, ..., p_n)$ in an experiment. The information contained in this experiment is given by

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i, \tag{4.1.1}$$

which is well known Shannon entropy.

The concept of entropy has been widely used in different areas, e.g., communication theory, statistical mechanics, finance, pattern recognition, and neural network etc. Fuzzy set theory developed by Lofti A. Zadeh (1965) has found wide applications in many areas of science and technology, e.g., clustering, image processing, decision making etc. because of its capability to model non-statistical imprecision or vague concepts.

Zadeh developed the concept of fuzzy set and defined the entropy of a fuzzy set, which is different from the classical Shannon entropy as no probabilistic concept is needed in order to define it. It may be noted that fuzzy entropy deals with vagueness and ambiguous uncertainties, while Shannon entropy deals with randomness (probabilistic) of uncertainties. Fuzzy entropy is a measure of fuzziness of a set which arises from the intrinsic ambiguity or vagueness carried by the fuzzy set.

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

$$\mu_A(x) = \begin{cases} 0, & \text{if } x \notin A \text{ and there is no ambiguity,} \\ 1, & \text{if } x \in A \text{ and there is no ambiguity,} \\ 0.5, & \text{if there is maximum ambiguity whether } x \in A \text{ or } x \notin A \end{cases}$$

In fact $\mu_A(x)$ associates with each $x \in U$, a grade of membership in the set A. When $\mu_A(x)$ is valued in $\{0, 1\}$, it is the characteristic function of a crisp (nonfuzzy) set.

Two fuzzy sets A and B are said to be *fuzzy-equivalent* if $\mu_B(x_i)$ = either $\mu_A(x_i)$ or $1 - \mu_A(x_i)$ for each value of i. It is clear that fuzzy-equivalent sets have the same entropy, but two sets may have the same fuzzy entropy without being fuzzy equivalent. From the fuzziness point of view there is no essential difference between fuzzy equivalent sets. A *standard fuzzy set* is that member of the class of fuzzy equivalent sets all of whose membership value are less than or equal to 0.5.

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

$$\mu_{A^*}(x_i) \leq \mu_A(x_i), \text{ if } \mu_A(x_i) \leq 0.5; \quad \forall i$$

and

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

It may be noted that if $x_1, x_2, ..., x_n$ are members of the universe of discourse, then all $\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n)$ lie between 0 and 1, but these are not probabilities because their sum is not unity. However,

$$\Phi_A(x_i) = \frac{\mu_A(x_i)}{\sum_{i=1}^{n} \mu_A(x_i)}; \quad i = 1, 2, \ldots, n, \tag{4.1.2}$$

is a probability distribution.

Kaufman (1980) defined entropy of a fuzzy set A having n support points by

$$H(A) = -\frac{1}{\log n} \sum_{i=1}^{n} \Phi_A(x_i) \log \Phi_A(x_i). \tag{4.1.3}$$

Since $\mu_A(x)$ and $1 - \mu_A(x)$ gives the same degree of fuzziness, therefore, analogous to the entropy due to Shannon (1948), De Luca and Termini (1972) suggested the following measure of fuzzy entropy:

$$H(A) = -\sum_{i=1}^{n} [\mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log (1 - \mu_A(x_i))]. \tag{4.1.4}$$

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

- P1 (*Sharpness*): $H(A)$ is minimum if and only if A is a crisp set i.e., $\mu_A(x) = 0$ or $1; \forall x$.

- P2 (*Maximality*): $H(A)$ is maximum if and only if A is most fuzzy set i.e., $\mu_A(x) = 0.5; \quad \forall x$.

- P3 (*Resolution*): $H(A) \geq H(A^*)$, where A^* is sharpened version of A.

- P4 (*Symmetry*): $H(A) = H(\bar{A})$, where $\bar{A}$ is the complement of A i.e., $\mu_{\bar{A}}(x_i) = 1 - \mu_A(x_i)$.

Later on Bhandari and Pal (1993) made a survey on information measures on fuzzy sets and gave some new measures of fuzzy entropy. Analogous to Rényi's (1961) entropy they have suggested the following measure:

$$H_\alpha(A) = \frac{1}{1-\alpha}\sum_{i=1}^{n} \log\left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\right]; \qquad \alpha \neq 1,\ \alpha > 0 \qquad (4.1.5)$$

and analogous to Pal and Pal's (1989) exponential entropy they introduced

$$H_e(A) = \frac{1}{n\sqrt{e}-1}\sum_{i=1}^{n} \log\left[\mu_A(x_i)e^{1-\mu_A(x_i)} + (1-\mu_A(x_i))\,e^{\mu_A(x_i)} - 1\right]. \qquad (4.1.6)$$

Kapur (1997) has given measure of fuzzy entropy analogous to Havrda and Charvát's (1967) entropy as

$$H^\alpha(A) = \frac{1}{1-\alpha}\sum_{i=1}^{n} \left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha - 1\right]. \qquad (4.1.7)$$

Further, Parkash (1998) proposed a measure of fuzzy entropy containing two real parameters α and β as

$$H_\alpha^\beta(A) = \frac{1}{[(1-\alpha)\,\beta]}\sum_{i=1}^{n} \left[\{\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\}^\beta - 1\right]; \qquad (4.1.8)$$

where $\alpha \neq 1,\ \alpha > 0,\ \beta \neq 0$. The fuzzy entropy given by (4.1.8) may be called as (α, β) fuzzy entropy. The fuzzy entropies (4.1.4) and (4.1.5) are limiting cases whereas (4.1.7) is a particular case of (4.1.8).

A large number of measures of entropy, directed divergence and symmetric divergence for probability distributions are known (refer to Aczél and Daróczy (1975), Taneja (2005)). Analogously, there is a set of variety of measures for fuzzy sets available. Cross-entropy measure is also known as the information discrepancy between two probability distributions. Derived from cross-entropy, fuzzy directed divergence measures the dissimilarity between two fuzzy sets.

Kullback and Leibler (1951) obtained the measure of directed divergence of probability distribution $P = (p_1, p_2, \ldots, p_n)$ from the probability distribution $Q = (q_1, q_2, \ldots, q_n)$ as

$$D(P : Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}. \tag{4.1.9}$$

Kullback (1959) suggested the measure of symmetric divergence as

$$J(P : Q) = \sum_{i=1}^{n} (p_i - q_i) \log \frac{p_i}{q_i}. \tag{4.1.10}$$

Let A and B be two standard fuzzy sets with same supporting points x_1, x_2, …,x_n and with fuzzy vectors $\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n)$ and $\mu_B(x_1), \mu_B(x_2)$, $\ldots, \mu_B(x_n)$. The simplest measure of fuzzy directed divergence as suggested by Bhandari and Pal (1993), is

$$I(A, B) = \sum_{i=1}^{n} \left[\mu_A(x_i) \log \frac{\mu_A(x_i)}{\mu_B(x_i)} + (1 - \mu_A(x_i)) \log \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))} \right] \tag{4.1.11}$$

and the analogous symmetric divergence measure by

$$J(A, B) = I(A, B) + I(B, A),$$

which on simplification gives

$$J(A, B) = \sum_{i=1}^{n} [(\mu_A(x_i) - \mu_B(x_i)] \log \frac{\mu_A(x_i)(1 - \mu_A(x_i))}{\mu_B(x_i)(1 - \mu_B(x_i))} \quad . \tag{4.1.12}$$

It is important to notice that if we take $B = A_F$ (the most fuzzy set) i.e. $\mu_B(x_i) = 0.5; \forall\, i$, then from (4.1.11) and (4.1.4) we have

$$I(A,A_F) = n\log 2 - \left[-\sum_{i=1}^{n} \mu_A(x_i)\log\mu_A(x_i) + (1-\mu_A(x_i))\log(1-\mu_A(x_i))\right]$$

$$\text{or} \qquad I(A,A_F) = n\log 2 - H(A). \tag{4.1.13}$$

Kapur (1997) suggested that (4.1.13) will hold whatever be the measure of directed divergence we use. Thus from every measure of directed divergence for fuzzy sets we can deduce a corresponding measure of entropy for a fuzzy set.

In literature, a number of measures of fuzzy entropy analogous to the various information measures have been proposed in order to combine the fuzzy set theory and its application to the entropy concept as fuzzy information measurements.

The uncertainty is the state of being uncertain (not certain to occur) which gives rise to fuzziness and ambiguity. Ambiguity can be viewed in non-specificity (indistinguishable alternatives) and conflict (distinguishable alternatives) while fuzziness can be viewed as lack of distinction between a set and its complement and vagueness is non-specific knowledge about lack of distinction. Thus the measure of total fuzzy ambiguity can be obtained by taking the sum of measure of fuzzy directed divergence and corresponding measure of fuzzy entropy. From (4.1.4) and (4.1.11), we have

$$\begin{aligned} TA &= -\sum_{i=1}^{n} \mu_A(x_i)\log\mu_A(x_i) - \sum_{i=1}^{n}(1-\mu_A(x_i))\log(1-\mu_A(x_i)) \\ &\quad +\sum_{i=1}^{n}\left[\mu_A(x_i)\log\frac{\mu_A(x_i)}{\mu_B(x_i)} + (1-\mu_A(x_i))\log\frac{(1-\mu_A(x_i))}{(1-\mu_B(x_i))}\right] \end{aligned}$$

or

$$TA = -\sum_{i=1}^{n} \mu_A(x_i)\log\mu_B(x_i) - \sum_{i=1}^{n}(1-\mu_A(x_i))\log(1-\mu_B(x_i)). \tag{4.1.14}$$

If we subtract directed divergence of fuzzy set A and fuzzy set C from the directed divergence of fuzzy set A and fuzzy set B, we get the reduction in ambiguity in revising B to C and obtain a measure, which can be called fuzzy information improvement measure.

In the present chapter, two new generalized measures of fuzzy directed divergence are introduced and studied in Section 4.2. Their particular cases have also been discussed. In Section 4.3, we have obtained the measures of total ambiguity. In Section 4.4, new generalized measures of fuzzy information improvement are defined and discussed.

4.2 Generalized Measures of Fuzzy Directed Divergence

Havrda and Charvát (1967) defined the directed divergence measure of a probability distribution $P = (p_1, p_2, \ldots, p_n)$ from another probability distribution $Q = (q_1, q_2, \ldots, q_n)$ as

$$D^\beta(P : Q) = \frac{1}{\beta - 1}\sum_{i=1}^{n}(p_i^\beta q_i^{1-\beta} - 1); \quad \beta > 0,\, \beta \neq 1, \tag{4.2.1}$$

which is called the generalized directed divergence of degree β.

The following measure of symmetric divergence was proposed by Kullback (1959):

$$J^\beta(P : Q) = D^\beta(P : Q) + D^\beta(Q : P) = \frac{1}{\beta - 1}\sum_{i=1}^{n}(p_i^\beta q_i^{1-\beta} + q_i^\beta p_i^{1-\beta} - 2), \tag{4.2.2}$$

which is also called a distance measure of degree β. The measures (4.2.1) and (4.2.2) have been further generalized and applied by Taneja (2005) in Markov chains, comparison of experiments, etc.

Analogous to the measure (4.2.1) and (4.2.2), Hooda (2004) suggested the following measures of fuzzy directed divergence:

$$I^\beta(A, B) = \frac{1}{\beta - 1}\sum_{i=1}^{n}\left[\mu_A^\beta(x_i)\mu_B^{1-\beta}(x_i) + (1 - \mu_A(x_i))^\beta(1 - \mu_B(x_i))^{1-\beta} - 1\right] \tag{4.2.3}$$

and

$$J^{\beta}(A,B) = I^{\beta}(A,B) + I^{\beta}(B,A) \tag{4.2.4}$$

respectively.

Further, it is proved that $I^{\beta}(A,B) \geq 0$ for all $\beta(\neq 1) > 0$ and it vanishes only when $A = B$, which shows that (4.2.3) is a valid measure of fuzzy directed divergence of fuzzy set A and B. Hence, $J^{\beta}(A,B)$ is a valid fuzzy symmetric divergence measure.

4.2.1 Fuzzy Directed Divergence of order α

Analogous to Rényi's measure (1961) of directed divergence

$$D_{\alpha}(P,Q) = \frac{1}{\alpha - 1} \log \left(\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} \right); \quad \alpha \neq 1,\ \alpha > 0, \tag{4.2.5}$$

we define the following measure of fuzzy directed divergence of fuzzy set A from fuzzy set B:

$$I_{\alpha}(A,B) = \frac{1}{\alpha - 1} \sum_{i=1}^{n} \log \left[\mu_A^{\alpha}(x_i) \mu_B^{1-\alpha}(x_i) + (1 - \mu_A(x_i))^{\alpha} (1 - \mu_B(x_i))^{1-\alpha} \right]; \tag{4.2.6}$$

where $\alpha \neq 1,\ \alpha > 0$ and measure of fuzzy symmetric divergence

$$J_{\alpha}(A,B) = I_{\alpha}(A,B) + I_{\alpha}(B,A). \tag{4.2.7}$$

The measures (4.2.6) and (4.2.7) are called the generalized measures of order α.

Next, we show that $I_{\alpha}(A,B)$ is a valid fuzzy directed divergence measure. $I_{\alpha}(A,B)$ is a valid measure only if it is non-negative. So it is proved that $I_{\alpha}(A,B) \geq 0$ with equality if $\mu_A(x_i) = \mu_B(x_i)$ for each $i = 1, 2, \ldots, n$.

Let $\sum_{i=1}^{n} \mu_A(x_i) = s, \quad \sum_{i=1}^{n} \mu_B(x_i) = t.$

Then

$$\frac{1}{\alpha-1}\left[\sum_{i=1}^{n}\left(\frac{\mu_A(x_i)}{s}\right)^{\alpha}\left(\frac{\mu_B(x_i)}{t}\right)^{1-\alpha}-1\right]\geq 0$$

or

$$\frac{1}{\alpha-1}\sum_{i=1}^{n}\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i) \quad \geq \quad \frac{1}{\alpha-1}s^{\alpha}t^{1-\alpha}. \tag{4.2.8}$$

Similarly,

$$\frac{1}{\alpha-1}\sum_{i=1}^{n}(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha} \quad \geq \quad \frac{1}{\alpha-1}(n-s)^{\alpha}(n-t)^{1-\alpha} \tag{4.2.9}$$

***Case 1*:** When $0<\alpha<1$,

i.e. $\frac{1}{\alpha-1}<0$ then from (4.2.8) and (4.2.9) we have

$$\sum_{i=1}^{n}\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha} \;\leq\; s^{\alpha}t^{1-\alpha}+(n-s)^{\alpha}(n-t)^{1-\alpha}, \tag{4.2.10}$$

where

$\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}<1$ but very close to 1,

$\forall i=1,2,\ldots,n$

and

$s^{\alpha}t^{1-\alpha}+(n-s)^{\alpha}(n-t)^{1-\alpha}<n$ but very close to n.

It implies

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

Or

$$\sum_{i=1}^{n}\log\left[\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right]\leq s^{\alpha}t^{1-\alpha}+(n-s)^{\alpha}(n-t)^{1-\alpha}-n. \tag{4.2.11}$$

Though left hand side and right hand side of (4.2.11) are negative but the absolute value of right hand side is less than the absolute value of left hand side. Therefore, multiplying both sides of (4.2.11) by $\frac{1}{\alpha-1}$, we get

$$I_\alpha(A,B) \geq \frac{1}{\alpha-1}[s^\alpha t^{1-\alpha} + (n-s)^\alpha(n-t)^{1-\alpha} - n]. \tag{4.2.12}$$

For the sake of simplicity and constructing Table 4.1, we denote

$a = s^\alpha t^{1-\alpha} + (n-s)^\alpha(n-t)^{1-\alpha} - n;$

$b_i = \mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i) + (1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha},\ i = 1,2,\ldots,n;$

$c = \sum_{i=1}^{n} \log(b_i)$ and

$d = I_\alpha(A,B).$

***Case 2*:** When $\alpha > 1$,

i.e. $\frac{1}{\alpha-1} > 0$, then from (4.2.8) and (4.2.9) we have

$$\sum_{i=1}^{n} b_i \geq s^\alpha t^{1-\alpha} + (n-s)^\alpha(n-t)^{1-\alpha}.$$

Similarly, it can be verified that

$$\sum_{i=1}^{n} \log(b_i) \geq [s^\alpha t^{1-\alpha} + (n-s)^\alpha(n-t)^{1-\alpha} - n]. \tag{4.2.13}$$

Multiplying both sides of (4.2.11) by $\frac{1}{\alpha-1}$, we get (4.2.12) in *Case 2* also. Further, let

$$\psi(s) = \frac{1}{\alpha-1}[s^\alpha t^{1-\alpha} + (n-s)^\alpha(n-t)^{1-\alpha} - n],$$

then

$$\psi'(s) = \frac{1}{\alpha-1}\left[\alpha\left(\frac{s}{t}\right)^{\alpha-1} - \alpha\left(\frac{n-s}{n-t}\right)^{\alpha-1}\right],$$

and

$$\psi''(s) = \left[\frac{\alpha}{t}\left(\frac{s}{t}\right)^{\alpha-2} + \frac{\alpha}{n-t}\left(\frac{n-s}{n-t}\right)^{\alpha-2}\right].$$

Clearly, $\psi''(s) > 0$ which shows that $\psi(s)$ is a convex function of s whose minimum value arises when $\frac{s}{t} = \frac{n-s}{n-t} = \frac{n}{n} = 1$. Now if $A = B$ i.e., $s = t$, then $\psi(s) = 0$. Hence, $\psi(s) > 0$ and vanishes only when $s = t$.

Thus for all $\alpha > 0$, $I_\alpha(A, B) \geq 0$ and vanishes only when $A = B$.

Therefore, $I_\alpha(A, B)$ is a valid measure of directed divergence of fuzzy sets A and B. Consequently, $J_\alpha(A, B)$ is a valid measure of symmetric divergence.

Particular Cases:

- $\lim\limits_{\alpha \to 1} I_\alpha(A, B) = I(A, B)$ and $\lim\limits_{\alpha \to 1} J_\alpha(A, B) = J(A, B)$,
  where $I(A, B)$ and $J(A, B)$ are the fuzzy directed divergence and symmetric divergence measures given by (4.1.11) and (4.1.12) respectively.

- Let $B = A_F$, the most fuzzy set, i.e. $\mu_B(x_i) = 0.5 \quad \forall x_i$, then

$$\begin{aligned} I_\alpha(A, A_F) &= \tfrac{1}{\alpha-1} \sum_{i=1}^{n} \log\left[\mu_A^\alpha(x_i)(0.5)^{1-\alpha} + (1-\mu_A(x_i))^\alpha (0.5)^{1-\alpha}\right] \\ &= \tfrac{1}{\alpha-1}\left[\sum_{i=1}^{n} \log(0.5)^{1-\alpha} + \sum_{i=1}^{n} \log\left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\right]\right] \\ &= \tfrac{1}{\alpha-1}\left[\sum_{i=1}^{n} (\alpha-1)\log 2 + \sum_{i=1}^{n} \log\left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\right]\right] \\ &= n\log 2 - \tfrac{1}{1-\alpha} \log \sum_{i=1}^{n} \left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\right]. \end{aligned}$$

  Thus $I_\alpha(A, A_F) = n\log 2 -$ (Entropy of the fuzzy set).

Example 4.1: Let $A = (0.1,\ 0.3,\ 0.4,\ 0.2,\ 0.1)$ and $B = (0.3,\ 0.5,\ 0.3,\ 0.1,\ 0.2)$ be two arbitrary standard fuzzy sets. *Case 1* and *Case 2* both can be verified from the computed Table 4.1 given below:

Table 4.1

α	a	b_1	b_2	b_3	b_4	b_5	c	d
0.1	-0.0044	0.9866	0.9922	0.9980	0.9966	0.9961	-0.0440	0.0489
0.2	-0.0078	0.9769	0.9862	0.9965	0.9939	0.9932	-0.0774	0.0967
0.5	-0.0120	0.9669	0.9789	0.9945	0.9899	0.9899	-0.1164	0.2327
0.7	-0.0100	0.9737	0.9825	0.9953	0.9912	0.9919	-0.0952	0.3175
0.9	-0.0042	0.9893	0.9926	0.9980	0.9961	0.9966	-0.0397	0.3976
1.2	0.0111	1.0267	1.0196	1.0055	1.0111	1.0085	0.1020	0.5100
1.5	0.0343	1.0782	1.0606	1.0174	1.0371	1.0253	0.3070	0.6140
2.0	0.0893	1.1905	1.1600	1.0476	1.1111	1.0625	0.7722	0.7722
5.0	0.7909	2.4606	2.7280	1.5881	3.6994	1.4479	5.8353	1.4588
10.0	3.1409	8.6406	14.4658	5.4772	102.6772	2.5981	17.4785	1.9420

4.2.2 Fuzzy Directed Divergence of order α and type β

Sharma and Mittal (1975) characterized non-additive entropy of discrete probability distribution given by

$$H_\alpha^\beta(P) = \frac{1}{2^{1-\beta}-1}\left[\left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}} - 1\right]; \tag{4.2.14}$$

where $\alpha \neq 1, \quad \alpha > 0, \ \beta > 0, \ \beta \neq 1.$

Hooda (2004) suggested the following fuzzy entropy analogous to (4.2.14):

$$H_\alpha^\beta(A) = \tfrac{1}{2^{1-\beta}-1}\sum_{i=1}^{n}\left[\left(\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\right)^{\frac{\beta-1}{\alpha-1}} - 1\right]; \tag{4.2.15}$$

where $\alpha \neq 1, \quad \alpha > 0, \ \beta > 0, \ \beta \neq 1.$

Sharma and Mittal (1977) also studied the following generalized measure of directed divergence:

$$\frac{1}{1-2^{1-\beta}}\left[\left(\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha}\right)^{\frac{\beta-1}{\alpha-1}}-1\right];\ \alpha\neq 1,\quad \alpha>0,\ \beta>0,\ \beta\neq 1. \qquad (4.2.16)$$

Analogous to (4.2.16), we define the following measures of fuzzy directed divergence and symmetric fuzzy directed divergence:

$$I_{\alpha}^{\beta}(A,B)=\frac{1}{2^{\beta-1}-1}\sum_{i=1}^{n}\left[\left(\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right)^{\frac{\beta-1}{\alpha-1}}-1\right]; \qquad (4.2.17)$$

where $\alpha\neq 1,\quad \alpha>0,\ \beta>0,\ \beta\neq 1$

and

$$J_{\alpha}^{\beta}(A,B)=I_{\alpha}^{\beta}(A,B)+I_{\alpha}^{\beta}(B,A) \qquad (4.2.18)$$

respectively. We may call (4.2.17) as fuzzy directed divergence of order α and type β.

$I_{\alpha}^{\beta}(A,B)$ is a valid measure only if it is non-negative. We claim that $I_{\alpha}^{\beta}(A,B)\geq 0$ with equality if $\mu_A(x_i)=\mu_B(x_i)$ for each $i=1,2,\ldots,n$. From (4.2.17) it is obvious that $I_{\alpha}^{\beta}(A,B)=0$ if $\mu_A(x_i)=\mu_B(x_i)$. Next, we prove that $I_{\alpha}^{\beta}(A,B)>0$ for two different fuzzy sets empirically.

For the sake of simplicity and constructing the Tables 4.2 to 4.5, we denote

$s=\sum_{i=1}^{n}\mu_A(x_i)\,;$

$t=\sum_{i=1}^{n}\mu_B(x_i)\,;$

$a=s^{\alpha}t^{1-\alpha}+(n-s)^{\alpha}(n-t)^{1-\alpha}-n;$

$$e_i=(b_i)^{\frac{\beta-1}{\alpha-1}}-1$$

$$=\left(\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right)^{\frac{\beta-1}{\alpha-1}}-1,\ \forall i=1,\,2,\ldots,n.$$

$e=\sum_{i=1}^{n}e_i$ and

$f=I_{\alpha}^{\beta}(A,B).$

Further taking all the possible cases of α and β with two above considered standard fuzzy sets in Example 4.1 of section 4.2. Next, we tabulate the values given in Table 4.2 to 4.5.

Case 1: $0 < \alpha < 1$ and $0 < \beta < 1$.

In this case $\frac{1}{2^{\beta-1}-1} < 0$, $\frac{\beta-1}{\alpha-1} > 0$ and the computed values are presnted in Table 4.2.

Table 4.2

α	β	e_1	e_2	e_3	e_4	e_5	e	$I_\alpha^\beta(A,B)$
0.2	0.6	-0.0116	-0.0069	-0.0017	-0.0030	-0.0034	-0.0267	0.110357
0.4	0.8	-0.0110	-0.0068	-0.0017	-0.0031	-0.0032	-0.0260	0.201301
0.1	0.5	-0.0074	-0.0043	-0.0010	-0.0018	-0.0021	-0.0169	0.057806
0.3	0.9	-0.0042	-0.0025	-0.0006	-0.0011	-0.0012	-0.0099	0.148134
0.6	0.4	-0.0459	-0.0300	-0.0079	-0.0147	-0.0141	-0.1128	0.331761
0.8	0.1	-0.0849	-0.0584	-0.016	-0.0303	-0.0270	-0.2168	0.467185
0.7	0.2	-0.0686	-0.0460	-0.0124	-0.0232	-0.0215	-0.1719	0.403862
0.5	0.7	-0.0199	-0.0127	-0.0033	-0.0060	-0.0060	-0.0480	0.256106
0.9	0.3	-0.0727	-0.0509	-0.0140	-0.0270	-0.0232	-0.1880	0.48921

Observation: In this case, $I_\alpha^\beta(A,B)$ is a monotonic increasing function with respect to α but not with respect to β.

Case 2: $0 < \alpha < 1$ and $\beta > 1$.

In this case $\frac{1}{2^{\beta-1}-1} > 0$, $\frac{\beta-1}{\alpha-1} < 0$ and computed values are presented in Table 4.3.

Table 4.3

α	β	e_1	e_2	e_3	e_4	e_5	e	$I_\alpha^\beta(A,B)$
0.2	1.5	0.0147	0.0086	0.0022	0.0038	0.0043	0.0337	0.08135
0.4	2	0.0568	0.0349	0.0088	0.0159	0.0166	0.1332	0.13323
0.1	2.5	0.0227	0.0131	0.0032	0.0056	0.0065	0.0513	0.028058
0.3	3	0.0890	0.0530	0.0132	0.0236	0.0255	0.2044	0.068157
0.6	3.5	0.2166	0.1356	0.0339	0.0637	0.0612	0.5111	0.109754
0.8	4	0.3445	0.2221	0.0552	0.1081	0.0958	0.8259	0.117996
0.7	4.5	0.3650	0.2290	0.0562	0.1083	0.0999	0.8586	0.083249
0.5	5	0.3086	0.1859	0.0452	0.0841	0.0841	0.7082	0.047215
0.9	5.5	0.6244	0.3993	0.0954	0.1930	0.1635	1.4758	0.068238

Observation: In this case, $I_\alpha^\beta(A,B)$ is not a monotonic function with respect to either α or with respect to β.

Case 3: $\alpha > 1$ and $0 < \beta < 1$.

In this case $\frac{1}{2^{\beta-1}-1} < 0$, $\frac{\beta-1}{\alpha-1} < 0$ and the computed values are presented in Table 4.4.

Table 4.4

α	β	e_1	e_2	e_3	e_4	e_5	e	$I_\alpha^\beta(A,B)$
1.5	0.1	-0.1268	-0.1005	-0.0305	-0.0635	-0.0440	-0.3653	0.787088
2	0.2	-0.1302	-0.1120	-0.0365	-0.0808	-0.0473	-0.4069	0.955849
2.5	0.3	-0.1250	-0.1149	-0.0402	-0.0942	-0.0472	-0.4216	1.096671
3	0.4	-0.1143	-0.1110	-0.0415	-0.1021	-0.0446	-0.4135	1.215378
3.5	0.5	-0.1000	-0.1014	-0.0404	-0.1036	-0.0399	-0.3853	1.315536
4	0.6	-0.0831	-0.0874	-0.0369	-0.0977	-0.0338	-0.3389	1.399456
4.5	0.7	-0.0643	-0.0697	-0.0310	-0.0843	-0.0265	-0.2758	1.468905
5	0.8	-0.0440	-0.0489	-0.0229	-0.0633	-0.0183	-0.1975	1.525482
5.5	0.9	-0.0225	-0.0256	-0.0125	-0.0351	-0.0095	-0.1052	1.570742

Observation: In this case, $I_\alpha^\beta(A,B)$ is a monotonic increasing function with respect to α and β together. In other words, if we do not take any one parameter in increasing order, then monotonicity may lost.

Case 4: $\alpha > 1$ and $\beta > 1$.

In this case $\frac{1}{2^{\beta-1}-1} > 0$, $\frac{\beta-1}{\alpha-1} > 0$ and the computed values are presented in Table 4.5.

Table 4.5

α	β	e_1	e_2	e_3	e_4	e_5	e	$I_\alpha^\beta(A,B)$
1.5	2.5	0.2536	0.1931	0.0530	0.1154	0.0779	0.6930	0.37903
2	3.5	0.5463	0.4493	0.1233	0.3013	0.1636	1.5839	0.34013
2.5	2	0.2102	0.1905	0.0604	0.1518	0.0716	0.6845	0.68450
3	4	0.8350	0.8005	0.2363	0.7138	0.2559	2.8416	0.40594
3.5	1.5	0.1111	0.1129	0.0421	0.1155	0.0416	0.4232	1.02168
4	5	1.3806	1.4957	0.4568	1.7954	0.4100	5.5384	0.36923
4.5	4.5	1.1711	1.3229	0.4446	1.7925	0.3680	5.0991	0.49440
5	5.5	1.7537	2.0926	0.6826	3.3566	0.5164	8.4020	0.38849
5.5	3	0.5776	0.6797	0.2855	1.0441	0.2093	2.7962	0.93208

Observation: In this case, $I_\alpha^\beta(A,B)$ is not a monotonic function with respect to either α or with respect to β.

In all the cases including the case of $\alpha = \beta$, we observe that $I_\alpha^\beta(A,B) > 0$. Hence, for all $\alpha, \beta > 0$, $I_\alpha^\beta(A,B) \geq 0$ and vanishes only when $A = B$. Thus $I_\alpha^\beta(A,B)$ is a valid measure of directed divergence of fuzzy sets A and B. Consequently, $J_\alpha^\beta(A,B)$ is a valid measure of symmetric divergence. It may be noted that similar tables can be constructed for any two standard fuzzy sets and we shall see the same results obtained above.

Particular Cases:

- $\lim\limits_{\beta\to 1} I_\alpha^\beta(A,B) = I_\alpha(A,B)$ and $\lim\limits_{\beta\to 1} J_\alpha^\beta(A,B) = J_\alpha(A,B)$.

- Let $B = A_F$, the most fuzzy set, i.e. $\mu_B(x_i) = 0.5 \quad \forall x_i$, then

$$
\begin{aligned}
I_\alpha^\beta(A, A_F) &= \tfrac{1}{2^{\beta-1}-1}\sum_{i=1}^{n}\left[(\mu_A^\alpha(x_i)(0.5)^{1-\alpha} + (1-\mu_A(x_i))^\alpha(0.5)^{1-\alpha})^{\frac{\beta-1}{\alpha-1}} - 1\right] \\
&= \tfrac{1}{2^{\beta-1}-1}\cdot\tfrac{1}{2^{1-\beta}}\sum_{i=1}^{n}\left[(\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha)^{\frac{\beta-1}{\alpha-1}} - 2^{1-\beta}\right] \\
&= \tfrac{1}{1-2^{1-\beta}}\sum_{i=1}^{n}\left[(\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha)^{\frac{\beta-1}{\alpha-1}} - 2^{1-\beta}\right] \\
&= \tfrac{-n.2^{1-\beta}}{1-2^{1-\beta}} + \tfrac{1}{1-2^{1-\beta}}\sum_{i=1}^{n}\left[(\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha)^{\frac{\beta-1}{\alpha-1}} - 1\right] + \tfrac{n}{1-2^{1-\beta}} \\
&= n - H_\alpha^\beta(A)
\end{aligned}
$$

Thus $I_\alpha^\beta(A, A_F) = n-$(Entropy of the fuzzy set).

4.3 Measures of Total Ambiguity

Let A and B be two fuzzy sets. The total ambiguity of the fuzzy set A about set B is the sum of two components:

- Fuzzy entropy present in the fuzzy set A.
- Fuzzy directed divergence of A from B measured by $I(A, B)$.

Using Havrda and Charvát measure, Kapur (1997) estimated the total ambiguity as

$$
\begin{aligned}
TA &= \frac{1}{1-\alpha}\sum_{i=1}^{n}[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha - 1] \\
&\quad + \frac{1}{\alpha-1}\left[\sum_{i=1}^{n}\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i) + \sum_{i=1}^{n}(1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha} - 1\right] \\
&= \frac{1}{1-\alpha}\left[\sum_{i=1}^{n}\mu_A^\alpha(x_i)(1-\mu_B^{1-\alpha}(x_i)) + \sum_{i=1}^{n}(1-\mu_A(x_i))^\alpha\left(1-(1-\mu_B(x_i))^{1-\alpha}\right)\right].
\end{aligned}
$$

Corresponding to fuzzy entropy (4.1.5) and fuzzy directed divergence (4.2.6), we have total ambiguity given by

$$\begin{aligned}TA &= \frac{1}{1-\alpha}\sum_{i=1}^{n}\log\left[\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right] \\ &\quad + \frac{1}{\alpha-1}\sum_{i=1}^{n}\log\left[\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right] \\ &= \frac{1}{\alpha-1}\log\left[\left(\frac{\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}}{\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}}\right)\right].\end{aligned}$$

Similarly, corresponding to the fuzzy entropy (4.2.15) and fuzzy directed divergence (4.2.17), we have following measure of total ambiguity:

$$\begin{aligned}TA &= \frac{1}{2^{1-\beta}-1}\sum_{i=1}^{n}\left[\left(\mu_A^{\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}}-1\right] \\ &\quad + \frac{1}{2^{\beta-1}-1}\sum_{i=1}^{n}\left[\left(\mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha}\right)^{\frac{\beta-1}{\alpha-1}}-1\right].\end{aligned}$$

Total ambiguity is a fuzzy measure of inaccuracy analogous to Kerridge (1961) inaccuracy and is related to two fuzzy sets. It is not symmetric as we get something different if we interchange the role of the fuzzy sets A and B.

4.4 Generalized Fuzzy Information Improvement Measures

Let P and Q be observed and predicted distributions respectively of a random variable. Let $R = (r_1, r_2, \ldots, r_n)$ be the revised probability distribution of Q, then

$$D(P:Q) - D(P:R) = \sum_{i=1}^{n} p_i \log\frac{r_i}{q_i} \qquad (4.4.1)$$

which is known as Theil's (1967) measure of information improvement and has found wide applications in economics, accounts and financial management.

Similarly, suppose the correct fuzzy set is A and originally our estimate for it was the fuzzy set B that was revised to set C, the original ambiguity was $I(A,B)$ and finally ambiguity is $I(A,C)$, so the reduction in ambiguity is

$$I(A,B)-I(A,C)=\sum_{i=1}^{n}\left[\mu_A(x_i)\log\frac{\mu_C(x_i)}{\mu_B(x_i)}+(1-\mu_A(x_i))\log\frac{(1-\mu_C(x_i))}{(1-\mu_B(x_i))}\right]. \tag{4.4.2}$$

(4.4.2) can be called fuzzy information improvement measure.

In case of fuzzy directed divergence given by (4.2.6), the reduction in ambiguity is given by

$$\begin{aligned} I_\alpha(A,B)-I_\alpha(A,C) &= I_\alpha(A,B,C) \\ &= \frac{1}{\alpha-1}\left[\sum_{i=1}^{n}\log\left[\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}\right]\right. \\ &\quad \left.-\sum_{i=1}^{n}\log\left[\mu_A^\alpha(x_i)\mu_C^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_C(x_i))^{1-\alpha}\right]\right] \\ &= \frac{1}{\alpha-1}\left[\sum_{i=1}^{n}\log\frac{\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}}{\mu_A^\alpha(x_i)\mu_C^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_C(x_i))^{1-\alpha}}\right], \end{aligned} \tag{4.4.3}$$

which can be called the generalized fuzzy information improvement measure of order α.

Corresponding to the fuzzy directed divergence (4.2.17), the reduction in ambiguity is given by

$$\begin{aligned} &I_\alpha^\beta(A,B)-I_\alpha^\beta(A,C)=I_\alpha^\beta(A,B,C) \\ &=\frac{1}{2^{\beta-1}-1}\sum_{i=1}^{n}\left\{\begin{array}{c}\left(\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}\right)^{\frac{\beta-1}{\alpha-1}} \\ -\left(\mu_A^\alpha(x_i)\mu_C^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_C(x_i))^{1-\alpha}\right)^{\frac{\beta-1}{\alpha-1}}\end{array}\right\}, \end{aligned}$$

which can be called the generalized measure of fuzzy information improvement of order α and type β.

Chapter 5

Monotonicity and Maximum Fuzziness of Generalized Fuzzy Information Measures

5.1 Introduction

A large number of measures of entropy, directed divergence, symmetric divergence and their properties for probability distribution are known (refer to Taneja (2005)). Analogously, there are various measures available for fuzzy sets with their properties.

Sharma and Mittal (1975) characterized the following non-additive entropies of discrete probability distribution:

$$H^{\beta}(P) = \frac{1}{2^{1-\beta}-1}\left[2^{(\beta-1)\sum_{i=1}^{n} p_i \log p_i} - 1\right]; \qquad (5.1.1)$$

where $\beta > 0$, $\beta \neq 1$ and

$$H_\alpha^\beta(P) = \frac{1}{2^{1-\beta}-1}\left[\left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}} - 1\right]; \tag{5.1.2}$$

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

In particular, when $\alpha \to 1$, (5.1.2) tends to (5.1.1).

Analogous to the entropies (5.1.1) and (5.1.2), Hooda (2004) suggested the following measures of fuzzy entropies respectively:

$$H^\beta(A) = \frac{1}{1-\beta}\left[2^{(\beta-1)\sum\limits_{i=1}^{n}\mu_A(x_i)\log\mu_A(x_i)+(1-\mu_A(x_i))\log(1-\mu_A(x_i))} - 1\right]; \tag{5.1.3}$$

where $\beta > 0$, $\beta \neq 1$ and

$$H_\alpha^\beta(A) = \frac{1}{1-\beta}\sum_{i=1}^{n}\left[\left(\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\right)^{\frac{\beta-1}{\alpha-1}} - 1\right]; \tag{5.1.4}$$

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

In particular, when $\alpha \to 1$, (5.1.4) tends to (5.1.3).

Sharma and Mittal (1977) also characterized the following generalized measure of directed divergence:

$$D_\alpha^\beta(P:Q) = \frac{1}{1-2^{1-\beta}}\left[\left(\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}\right)^{\frac{\beta-1}{\alpha-1}} - 1\right]; \tag{5.1.5}$$

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

Analogous to (5.1.5), we have defined a measure of fuzzy directed divergence in Chapter 4, which is given by

$$I_\alpha^\beta(A,B) = \frac{1}{2^{\beta-1}-1}\sum_{i=1}^{n}\left[\left(\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i) + (1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}\right)^{\frac{\beta-1}{\alpha-1}} - 1\right]; \tag{5.1.6}$$

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

Kapur (1987) has proved that for any given probability distribution the measure of entropy is monotonic decreasing function and corresponding directed divergence is monotonic increasing function of the parameter involved. Kapur (1997) also proved the monotonocity of Havrda and Charvát's measure of fuzzy entropy and fuzzy directed divergence.

Although, there is some parallelism between probability measures and fuzzy measures, the results for fuzzy measures are not very obvious since these are not identical. Let P and Q be two probability distributions of a discrete random variable and let A and B be two fuzzy sets. The essential differences between these two are:

- While $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} q_i = 1$, $\sum_{i=1}^{n} \mu_A(x_i)$ and $\sum_{i=1}^{n} \mu_B(x_i)$ are not necessarily unity.

- While for probability distributions only $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$ are involved, here in addition to $\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n)$ and $\mu_B(x_1), \mu_B(x_2)$, $\ldots$, $\mu_B(x_n)$ the expressions also involve $1-\mu_A(x_1)$, $1-\mu_A(x_2)$, $\ldots$, $1-\mu_A(x_n)$; $1-\mu_B(x_1), 1-\mu_B(x_2), \ldots, 1-\mu_B(x_n)$.

- In (5.1.2) we have power of summation but in (5.1.4) we have summation of powers.

Parkash (1998) proved that the (α, β) fuzzy entropy given by (4.1.8) is monotonically decreasing function of α i.e., for fixed value of β, (α, β) fuzzy entropy decreases from $\frac{n}{\beta}(2^{\beta}-1)$ to 0 as α goes from 0 to ∞. Also, (α, β) fuzzy entropy is an increasing function of β for $0 < \alpha < 1$ and is decreasing function of β for $\alpha > 1$.

In this chapter we investigate the monotonic property with respect to the parameters involved in case of generalized measures of fuzzy information and fuzzy directed divergence given by (5.1.4) and (5.1.6) in section 5.2 and section 5.3

respectively. Particular cases and comparison between probabilistic entropy and fuzzy entropy are also studied. Further, particular cases and comparison between probabilistic directed divergence and fuzzy directed divergence have also been discussed. In Section 5.4, the maximum fuzziness of the above mentioned generalized measures of fuzzy information and fuzzy directed divergence have been obtained.

5.2 Monotonicity of Generalized Measure of Fuzzy Information

In this section we show that (5.1.4) is a monotonic decreasing function of α. Differentiating (5.1.4) with respect to α, we get

$$\frac{d}{d\alpha}\left(H_\alpha^\beta(A)\right) = \frac{1}{1-\beta}\sum_{i=1}^{n}\left[\begin{array}{c}\left(\mu_A^\alpha(x_i)+(1-\mu_A(x_i))^\alpha\right)^{\frac{\beta-1}{\alpha-1}}. \\ .\left(\begin{array}{c}-\frac{\beta-1}{(\alpha-1)^2}.\log\left(\mu_A^\alpha(x_i)+(1-\mu_A(x_i))^\alpha\right)+ \\ \frac{\beta-1}{\alpha-1}\frac{\mu_A^\alpha(x_i).\log\mu_A(x_i)+(1-\mu_A(x_i))^\alpha\log(1-\mu_A(x_i))}{\mu_A^\alpha(x_i)+(1-\mu_A(x_i))^\alpha}\end{array}\right)\end{array}\right].$$

$$\Rightarrow \frac{d}{d\alpha}\left(H_\alpha^\beta(A)\right) = \frac{1}{(\alpha-1)^2}\sum_{i=1}^{n}\left(\mu_A^\alpha(x_i)+(1-\mu_A(x_i))^\alpha\right)^{\frac{\beta-1}{\alpha-1}}.[(1-\alpha)f(x_i)+g(x_i)], \tag{5.2.1}$$

where

$$(1-\alpha)f(x_i) = (1-\alpha)\frac{\mu_A^\alpha(x_i).\log\mu_A(x_i)+(1-\mu_A(x_i))^\alpha\log(1-\mu_A(x_i))}{\mu_A^\alpha(x_i)+(1-\mu_A(x_i))^\alpha} \tag{5.2.2}$$

and

$$g(x_i) = \log\left(\mu_A^\alpha(x_i)+(1-\mu_A(x_i))^\alpha\right). \tag{5.2.3}$$

Further, equation (5.2.1) can be written as

$$\frac{d}{d\alpha}\left(H_\alpha^\beta(A)\right) = \frac{1}{(\alpha-1)^2}\sum_{i=1}^{n}\left[\left(\mu_A^\alpha(x_i)+(1-\mu_A(x_i))^\alpha\right)^{\frac{\beta-1}{\alpha-1}}.\frac{M_i}{N_i}\right], \tag{5.2.4}$$

where

$$M_i = \mu_A^\alpha(x_i). \log \mu_A^{1-\alpha}(x_i) + (1-\mu_A(x_i))^\alpha \log(1-\mu_A(x_i))^{1-\alpha}$$
$$+[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha] \log [\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha] \tag{5.2.5}$$

and

$$N_i = \mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha. \tag{5.2.6}$$

Let

$$\phi(s) = s^\alpha \log s^{1-\alpha} + (1-s)^\alpha \log(1-s)^{1-\alpha} + [s^\alpha + (1-s)^\alpha] \log[s^\alpha + (1-s)^\alpha], \tag{5.2.7}$$

so that $\phi(s) = \phi(1-s), \quad \phi'(s) = -\phi'(1-s)$ and $\phi(0) = \phi(1) = 0$.

Now from (5.2.7), we have

$$\phi(s) = (\lambda s_1 + (1-\lambda)s_2) \log(\lambda s_1 + (1-\lambda)s_2) - \lambda s_1 \log s_1 - (1-\lambda)s_2 \log s_2 \tag{5.2.8}$$

where $\lambda = s, \quad 1-\lambda = 1-s, \quad s_1 = s^{\alpha-1}, \quad s_2 = (1-s)^{\alpha-1}$.

Since $s \log s$ is a convex function of s, therefore

$$\phi(s) \leq 0, \quad 0 \leq s < 1 \tag{5.2.9}$$

and

$$\phi(\mu_A(x_i)) \leq 0. \tag{5.2.10}$$

From (5.2.5) and (5.2.10), it implies that $M_i \leq 0$ and from (5.2.6) it is obvious that $N_i > 0$. Therefore, from (5.2.4) we conclude that $\frac{d}{d\alpha}\left(H_\alpha^\beta(A)\right) \leq 0$, which show that $H_\alpha^\beta(A)$ is a monotonic decreasing function of α.

Particular Cases:

- If A_F is the most fuzzy set i.e., $\mu_{A_F}(x_i) = \frac{1}{2} \quad \forall x_i$ then from (5.1.4) with simple calculations, we have $H_\alpha^\beta(A_F) = \frac{n(2^{1-\beta}-1)}{1-\beta}$; which is independent of α and from (5.2.4) we get $\frac{d}{d\alpha}\left(H_\alpha^\beta(A_F)\right) = 0$. However, for all other fuzzy sets, $H_\alpha^\beta(A)$ is a strictly monotonic decreasing function of α.

- For $\alpha = 0$, (5.1.4) gives $H_0^\beta(A) = \frac{n(2^{1-\beta}-1)}{1-\beta} \leq n \ \ \forall\beta \ (\beta \neq 1)$. In particular, equality holds if $\beta = 0$, i.e., $H_0^0(A) = n$.

- For $\alpha \to \infty$,

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

$$\Rightarrow H_\infty^\beta(A) = \tfrac{1}{1-\beta} \sum_{i=1}^{n} \left[\lim_{\alpha\to\infty} \left(\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\right)^{\frac{\beta-1}{\alpha-1}} - 1 \right],$$
$$= \tfrac{1}{1-\beta} \sum_{i=1}^{n} \left[\lim_{\alpha\to\infty} 2^{(1-\beta).\frac{\log\left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\right]}{1-\alpha}} - 1 \right],$$
$$= \tfrac{1}{1-\beta} \sum_{i=1}^{n} \left[\lim_{\alpha\to\infty} 2^{-(1-\beta).f(x_i)} - 1 \right];$$

where

$$f(x_i) = \frac{\mu_A^\alpha(x_i).\log\mu_A(x_i) + (1-\mu_A(x_i))^\alpha \log(1-\mu_A(x_i))}{\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha}.$$

Further,

$$H_\infty^\beta(A) = \frac{1}{1\text{ -}\beta} \sum_{i=1}^{n} \left[\ 2^{-(1-\beta).\max.[\log\mu_A(x_i);\log(1-\mu_A(x_i))]} - 1 \right]. \tag{5.2.11}$$

If each $\mu_A(x_i) = \frac{1}{2}$, then $H_\alpha^\beta(A) = \frac{n(2^{1-\beta}-1)}{1-\beta} \leq n \ \ \forall\beta \ (\beta \neq 1)$. Even if it is independent of α, then also it is less than n and equality holds when $\beta = 0$. Thus maximum value is always n, but the minimum value depends on set A and will be zero if and only if the set is crisp.

Let A=(0.3, 0.4, 0.4, 0.2, 0.5) be any standard fuzzy set (one can choose any fuzzy set). We have plotted the graph of $H_\alpha^\beta(A) \sim \alpha$ for three different values of β (but fixed), where $\alpha > 0$, $\alpha \neq 1$, shown in Figure 5.1 which is based on the computed Table 5.1. Figure 5.1 shows the monotonic decreasing nature of $H_\alpha^\beta(A)$ with respect to α i.e., the value of $H_\alpha^\beta(A)$ decreases as α increases for different values of β.

It may be observed that for $\beta = 0.5$, from (5.2.11) we calculate $H_\infty^{0.5}(A) =$ 2.618930 and the pattern in Table 5.1 agrees with this value. Similarly, for $\beta = 2$

and $\beta = 5$, we calculate the value of $H^{\beta}_{\infty}(A)$ from (5.2.11) which comes to be 1.8 and 1.00715 respectively, which shows convergence, refer to Table 5.1.

Similarly, it can be proved that $H^{\beta}_{\alpha}(A)$ is monotonically decreasing function with respect to β. In particular, for the above considered fuzzy set A=(0.3, 0.4, 0.4, 0.2, 0.5), the monotonic nature of $H^{\beta}_{\alpha}(A)$ w.r.t. β can be observed in the computed Table 5.2 and Figure 5.2.

Comparison Between $H^{\beta}_{\alpha}(P)$ and $H^{\beta}_{\alpha}(A)$:

- If P is the uniform distribution i.e., $\left(\frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ or P is a degenerate function then by (5.1.2), it is clear that $H^{\beta}_{\alpha}(P)$ is independent of α. Similarly, observe that if A is the fuzziest set or A is crisp set then $H^{\beta}_{\alpha}(A)$ is independent of α.
- Both are monotonic decreasing function of α.
- When $\alpha \to \infty$, $H^{\beta}_{\alpha}(P) \to \frac{1}{2^{1-\beta}-1}\sum_{i=1}^{n}\left[2^{-(1-\beta).\log p_{\max.}} - 1\right]$

  while $H^{\beta}_{\alpha}(A) \to \frac{1}{1-\beta}\sum_{i=1}^{n}\left[2^{-(1-\beta).\max.[\log \mu_A(x_i);\log(1-\mu_A(x_i))]} - 1\right]$.

Table 5.1: *Monotonic Nature of Fuzzy Entropy w.r.t.* α

α	$H_\alpha^{0.5}(A)$	$H_\alpha^{2}(A)$	$H_\alpha^{5}(A)$
0.1	4.092988	2.482408	1.16962
0.2	4.044955	2.464791	1.167248
0.4	3.952707	2.429751	1.162172
0.5	3.908683	2.412459	1.159484
0.8	3.785735	2.362192	1.150953
0.9	3.747885	2.346136	1.147997
1.01	3.708063	2.328952	1.144709
1.2	3.643668	2.300552	1.138993
1.3	3.611941	2.28629	1.135991
1.4	3.581647	2.272515	1.133009
1.5	3.552736	2.25923	1.130059
1.8	3.473731	2.222287	1.121488
2	3.426914	2.2	1.116072
5	3.042433	2.011328	1.06575
10	2.838211	1.91082	1.038378
20	2.724488	1.854173	1.022795
25	2.702424	1.842999	1.019641
30	2.687976	1.835641	1.017544
50	2.659731	1.821157	1.013368
100	2.639101	1.810494	1.010252
300	2.625604	1.80348	1.008182
500	2.622928	1.802086	1.007769
800	2.621427	1.801303	1.007537
1000	2.620927	1.801042	1.00746
$\to \infty$	2.618930	1.8	1.00715

Table 5.2: *Monotonic Nature of Fuzzy Entropy w.r.t. β*

β	$H^{\beta}_{0.5}(A)$	$H^{\beta}_{2}(A)$	$H^{\beta}_{5}(A)$	$H^{\beta}_{10}(A)$
0.1	4.508121	3.907157	3.439569	3.19135
0.2	4.347758	3.779167	3.334007	3.097658
0.3	4.194634	3.656635	3.232762	3.007682
0.4	4.04839	3.5393	3.135634	2.921253
0.5	3.908683	3.426914	3.042433	2.838211
0.6	3.775188	3.319241	2.952977	2.758403
0.7	3.647599	3.216059	2.867095	2.681685
0.8	3.525624	3.117156	2.784624	2.607917
0.9	3.408988	3.02233	2.705407	2.536968
1.01	3.286544	2.922502	2.621851	2.462032
1.2	3.088567	2.760453	2.485843	2.339819
1.3	2.990805	2.68012	2.418239	2.278958
1.4	2.897204	2.603001	2.35322	2.220348
1.5	2.807564	2.528949	2.290671	2.163892
2	2.412459	2.2	2.011328	1.91082
2.5	2.091664	1.929356	1.779387	1.699362
3	1.829423	1.7052	1.585537	1.521541
3.5	1.613548	1.518299	1.422459	1.371053
4	1.434581	1.361413	1.284369	1.242891
4.5	1.285149	1.228841	1.166684	1.133064
5	1.159484	1.116072	1.06575	1.038378
10	0.553984	0.550441	0.543279	0.538902
20	0.263157	0.263121	0.262881	0.262654
50	0.102041	0.102041	0.102041	0.102041
100	0.050505	0.050505	0.050505	0.050505
500	0.01002	0.01002	0.01002	0.01002
1000	0.005005	0.005005	0.005005	0.005005

5.3 Monotonicity of Fuzzy Directed Divergence Measure

Here we shall prove that $I_\alpha^\beta(A,B)$ given by (5.1.6) is a monotonic increasing function of α. Differentiating (5.1.6) with respect to α, we get

$$\frac{d}{d\alpha}I_\alpha^\beta(A,B)=\frac{1}{2^{\beta-1}-1}\sum_{i=1}^{n}\left[(X_i)^{\frac{\beta-1}{\alpha-1}}.\left\{\frac{\beta-1}{\alpha-1}.\frac{1}{X_i}\frac{dX_i}{d\alpha}-\frac{\beta-1}{(\alpha-1)^2}\log X_i\right\}\right], \tag{5.3.1}$$

where $X_i=\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}$

$$\Rightarrow \frac{d}{d\alpha}I_\alpha^\beta(A,B)=\frac{\beta-1}{2^{\beta-1}-1}.\frac{1}{(\alpha-1)^2}\sum_{i=1}^{n}\left[(X_i)^{\frac{\beta-1}{\alpha-1}}.\left\{(\alpha-1).\frac{1}{X_i}\frac{dX_i}{d\alpha}-\log X_i\right\}\right],$$

$$\Rightarrow \frac{d}{d\alpha}I_\alpha^\beta(A,B)=\frac{\beta-1}{2^{\beta-1}-1}.\frac{1}{(\alpha-1)^2}\sum_{i=1}^{n}\left[(X_i)^{\frac{\beta-1}{\alpha-1}}.\left\{(\alpha-1).\frac{Z_i}{X_i}-\log X_i\right\}\right], \tag{5.3.2}$$

where

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

Further, (5.3.2) can be expressed as

$$\frac{d}{d\alpha}I_\alpha^\beta(A,B)=\frac{\beta-1}{2^{\beta-1}-1}.\frac{1}{(\alpha-1)^2}\sum_{i=1}^{n}\left[(X_i)^{\frac{\beta-1}{\alpha-1}}.\frac{\phi(\mu_A(x_i),\mu_B(x_i))}{X_i}\right], \tag{5.3.3}$$

where

$$\begin{aligned}\phi(s,t)&=s^\alpha t^{1-\alpha}\log\left(\tfrac{s}{t}\right)^{\alpha-1}+(1-s)^\alpha(1-t)^{1-\alpha}\log\left(\tfrac{1-s}{1-t}\right)^{\alpha-1}\\&\quad-\left(s^\alpha t^{1-\alpha}+(1-s)^\alpha(1-t)^{1-\alpha}\right)\log\left(s^\alpha t^{1-\alpha}+(1-s)^\alpha(1-t)^{1-\alpha}\right).\end{aligned} \tag{5.3.4}$$

Now (5.3.4) can be written as

$$\phi(s,t)=\lambda\, t_1\log t_1+(1-\lambda)t_2\log t_2-(\lambda\, t_1+(1-\lambda)t_2)\log\left(\lambda\, t_1+(1-\lambda)t_2\right), \tag{5.3.5}$$

where

$$\lambda = s, \;\; t_1 = \left(\frac{s}{t}\right)^{\alpha-1} \log\left(\frac{s}{t}\right)^{\alpha-1}, \;\; t_2 = \left(\frac{1-s}{1-t}\right)^{\alpha-1} \log\left(\frac{1-s}{1-t}\right)^{\alpha-1}. \quad (5.3.6)$$

Because of convexity of $s \log s$, we have

$$\phi(s,t) \geq 0 \;\text{ and }\; \phi(\mu_A(x_i), \mu_B(x_i)) \geq 0. \quad (5.3.7)$$

Since

$$X_i = \mu_A^{\alpha}(x_i)\mu_B^{1-\alpha}(x_i) + (1-\mu_A(x_i))^{\alpha}(1-\mu_B(x_i))^{1-\alpha} > 0 \quad (5.3.8)$$

and

$$\frac{\beta-1}{2^{\beta-1}-1}.\frac{1}{(\alpha-1)^2} \geq 0, \quad (5.3.9)$$

therefore, from (5.3.3), (5.3.7), (5.3.8) and (5.3.9), we have $\frac{d}{d\alpha} I_{\alpha}^{\beta}(A,B) \geq 0$. Hence, $I_{\alpha}^{\beta}(A,B)$ is a monotonic increasing function of α. Let $A =$ (0.3, 0.4, 0.4, 0.2, 0.5) and $B =$ (0.4, 0.3, 0.3, 0.1, 0.2) be two standard fuzzy sets (One can choose any two fuzzy sets). We plot the graph of $I_{\alpha}^{\beta}(A,B) \sim \alpha$ for three different values of β (but fixed), where $\alpha > 0$, $\alpha \neq 1$, shown in Figure 5.3 which is based on the computed Table 5.3. Figure 5.3 shows the monotonic increasing nature of $I_{\alpha}^{\beta}(A,B)$ with respect to α i.e., the value of $I_{\alpha}^{\beta}(A,B)$ increases as α increases for different values of β.

For $\beta = 0.5$, from (5.4.5) we calculate the limiting case $I_{\infty}^{\beta}(A,B)$= 3.42297. The same convergence pattern can be observed in Table 5.3 and Figure 5.3 which shows that the value of $I_{\alpha}^{0.5}(A,B)$ increases from 0 (which is the minimum value when $\alpha = 0$) to 3.42297 as α increases from 0 to ∞. Similarly, for $\beta = 2$ and $\beta = 5$ from (5.4.5) we calculate $I_{\infty}^{\beta}(A,B)$ which comes out to be 3.33333 and 3.88240 respectively.

Particular case:

- When $\alpha = 0$, from (5.1.6) we see that $I_0^{\beta}(A,B)$ =0.

- When $\alpha \to \infty$, in the next section we have calculated that
$I_{\infty}^{\beta}(A,B) \to \frac{1}{2^{\beta-1}-1}\sum_{i=1}^{n}\left[2^{(\beta-1).\max\left\{\log\frac{\mu_A(x_i)}{\mu_B(x_i)},\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right\}}-1\right].$

Comparison Between $I_{\alpha}^{\beta}(P:Q)$ and $I_{\alpha}^{\beta}(A,B)$:

- $I_{\alpha}^{\beta}(P:Q)$ and $I_{\alpha}^{\beta}(A,B)$ are independent of α and have value zero when $P=Q$ and $A=B$ respectively. These both are also zero when $\alpha=0$.

- $I_{\alpha}^{\beta}(P:Q)$ and $I_{\alpha}^{\beta}(A,B)$ are monotonic increasing function of α for certain values of β. However, from the Table 5.4 and Figures 5.4(a) and 5.4(b), it can be inferred that $I_{\alpha}^{\beta}(A,B)$ is neither increasing nor decreasing function of β for certain values of α.

Table 5.3: *Monotonic Nature of Fuzzy Directed Divergence w.r.t.* α

α	$I_\alpha^{0.5}(A,B)$	$I_\alpha^{2}(A,B)$	$I_\alpha^{5}(A,B)$
0.1	0.051024212	0.030201517	0.008225
0.2	0.103246291	0.061780127	0.017209
0.3	0.156567194	0.094753833	0.027037
0.5	0.266016791	0.164889727	0.049603
0.8	0.434958171	0.280141748	0.092231
0.9	0.491861072	0.320997559	0.109186
1.1	0.605469	0.405785	0.147768
1.2	0.661823689	0.449466923	0.169539
1.5	0.827216388	0.583938941	0.245254
1.7	0.933177792	0.674921105	0.30435
1.8	0.984594032	0.720368479	0.33634
2	1.084026337	0.810515873	0.404736
3	1.511860336	1.224455495	0.799139
5	2.097418423	1.822465847	1.524991
10	2.751722691	2.525996006	2.49235
20	3.104707171	2.93413971	3.139426
50	3.300598347	3.175729179	3.573906
75	3.342090963	3.228571201	3.675045
100	3.362571557	3.254876672	3.726248
200	3.392966582	3.294191016	3.803849
300	3.403012452	3.307257536	3.829928
$\rightarrow \infty$	3.42297554	3.333331905	3.882405

Table 5.4

5.4 Maximum Fuzziness of the Generalized Fuzzy Information and Directed Divergence Measures

Here we study the maximum fuzziness for the generalized measures of fuzzy information and fuzzy directed divergence given by (5.1.4) and (5.1.6) respectively subject to $\sum_{i=1}^{n} \mu_A(x_i) = \alpha_0$.

When α_0 is equally distributed, (5.1.4) has maximum value which comes out to be

$$H_{\max.} = \frac{1}{1-\beta}\sum_{i=1}^{n}\left[\left(\left(\frac{\alpha_0}{n}\right)^{\alpha} + \left(1-\frac{\alpha_0}{n}\right)^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}} - 1\right],$$

$$\Rightarrow H_{\max.} = \frac{1}{2^{1-\beta}-1}n\left[\left(\left(\frac{\alpha_0}{n}\right)^{\alpha} + \left(1-\frac{\alpha_0}{n}\right)^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}} - 1\right]. \quad (5.4.1)$$

Differentiating (5.4.1) with respect to α_0, we get

$$\frac{\partial H_{\max.}}{\partial \alpha_0} = \frac{\alpha}{1-\alpha}n\left[\frac{1}{\alpha_0}\left(\frac{\alpha_0}{n}\right)^{\alpha} - \frac{1}{n}\left(1-\frac{\alpha_0}{n}\right)^{\alpha-1}\right]\left[\left(\frac{\alpha_0}{n}\right)^{\alpha} + \left(1-\frac{\alpha_0}{n}\right)^{\alpha-1}\right]^{\frac{\beta-1}{\alpha-1}}$$

Differentiating with respect to α_0 again, we get

$$\frac{\partial^2 H_{\max.}}{\partial \alpha_0^2} = -\frac{n\alpha^2}{(\alpha-1)^2}.(\beta-\alpha)[M] - \alpha.[N] \quad (5.4.2)$$

where

$$M = \left[\left(\frac{\alpha_0}{n}\right)^{\alpha} + \left(1-\frac{\alpha_0}{n}\right)^{\alpha}\right]^{\frac{1-2\alpha+\beta}{\alpha-1}}\left(\frac{\left(\frac{\alpha_0}{n}\right)^{\alpha}}{\alpha_0} - \frac{\left(1-\frac{\alpha_0}{n}\right)^{\alpha-1}}{n}\right)^2 > 0$$

and

$$N = \left[\left(\frac{\alpha_0}{n}\right)^{\alpha} + \left(1-\frac{\alpha_0}{n}\right)^{\alpha}\right]^{\frac{\beta-\alpha}{\alpha-1}}\left(\frac{\left(\frac{\alpha_0}{n}\right)^{\alpha}}{\alpha_0^2} + \frac{\left(1-\frac{\alpha_0}{n}\right)^{\alpha-2}}{n^2}\right) > 0.$$

On considering the co-efficient of M and N in (5.4.2) and using (5.2.11), we conclude that $\frac{\partial^2}{\partial \alpha_0^2}H_{\max.} < 0$, provided $\alpha < \beta$ and that shows $H_{\max.}$ is a concave function of α_0. Therefore, the value of $H_\alpha^\beta(A)$ is maximum when $\alpha_0 = n - \alpha_0$ or $\alpha_0 = \frac{n}{2}$ and the maximum value is

$H_{\max.} = \frac{n}{1-\beta}\left[\left(\left(\frac{1}{2}\right)^\alpha + \left(1-\frac{1}{2}\right)^\alpha\right)^{\frac{\beta-1}{\alpha-1}} - 1\right] = \frac{n(2^{1-\beta}-1)}{1-\beta}$; which was expected also.

Next, since $I_\alpha^\beta(A,B)$ is a monotonic increasing function of α, therefore, the maximum value of (5.1.6) is obtained when $\alpha \to \infty$ and is given by

$$I_\infty^\beta(A,B) = \frac{1}{2^{\beta-1}-1}\sum_{i=1}^n \left[\lim_{\alpha\to\infty} \left(\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i) + (1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}\right)^{\frac{\beta-1}{\alpha-1}} - 1\right]$$

$$\Rightarrow I_\infty^\beta(A,B) = \frac{1}{2^{\beta-1}-1}\sum_{i=1}^n \left[\lim_{\alpha\to\infty} 2^{\frac{\beta-1}{\alpha-1}\log\left(\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}\right)} - 1\right]$$

$$\Rightarrow I_\infty^\beta(A,B) = \frac{1}{2^{\beta-1}-1}\sum_{i=1}^n \left[2^{\beta-1 \lim_{\alpha\to\infty}\frac{\log\left(\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i)+(1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}\right)}{\alpha-1}} - 1\right]. \tag{5.4.3}$$

Considering the limit term in (5.4.3) separately, we have

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

$$= \lim_{\alpha\to\infty}\frac{\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i)\log\frac{\mu_A(x_i)}{\mu_B(x_i)} + (1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}}{\mu_A^\alpha(x_i)\mu_B^{1-\alpha}(x_i) + (1-\mu_A(x_i))^\alpha(1-\mu_B(x_i))^{1-\alpha}}$$

$$= \lim_{\alpha\to\infty}\frac{\mu_A(x_i)\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha-1}\log\frac{\mu_A(x_i)}{\mu_B(x_i)} + (1-\mu_A(x_i))\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha-1}\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}}{\mu_A(x_i)\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^{\alpha-1} + (1-\mu_A(x_i))\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^{\alpha-1}}. \tag{5.4.4}$$

Case 1: When $\mu_A(x_i) > \mu_B(x_i)$ or $1-\mu_A(x_i) < 1-\mu_B(x_i)$, limit is $\log\frac{\mu_A(x_i)}{\mu_B(x_i)}$.

Case 2: When $\mu_A(x_i) = \mu_B(x_i)$ or $1-\mu_A(x_i) = 1-\mu_B(x_i)$, limit is 0.

Case 3: When $\mu_A(x_i) < \mu_B(x_i)$ or $1 - \mu_A(x_i) > 1 - \mu_B(x_i)$, limit is $\log \frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}$.

In view of all the cases discussed above, from (5.4.3) and (5.4.4) we have

$$I_{\infty}^{\beta}(A,B) = \frac{1}{2^{\beta-1}-1} \sum_{i=1}^{n} \left[2^{(\beta-1).\max\left\{\log \frac{\mu_A(x_i)}{\mu_B(x_i)}, \log \frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right\}} - 1 \right]. \tag{5.4.5}$$

Under the constraint that $\sum_{i=1}^{n} \mu_A(x_i) = \alpha_0$ and $\sum_{i=1}^{n} \mu_B(x_i) = \beta_0$, the maximum fuzzy directed divergence is given by

$$I_{\max.}^{\beta}(A,B) = \frac{n}{2^{\beta-1}-1} \left[2^{(\beta-1).\max\left\{\log \frac{\alpha_0}{\beta_0}, \log \frac{n-\alpha_0}{n-\beta_0}\right\}} - 1 \right]. \tag{5.4.6}$$

Chapter 6

R-norm Fuzzy Information Measures and their Generalizations

6.1 Introduction

De Luca and Termini (1972) defined the following fuzzy information measure analogous to the Shannon's (1948) entropy:

$$H(A) = -\sum_{i=1}^{n} \left[\mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log (1 - \mu_A(x_i))\right] \tag{6.1.1}$$

On formulating the axioms P1 to P4 mentioned in Chapter 4, which became the essential properties required by the fuzzy information measure. However, we have other fuzzy information measures but (6.1.1) can be regarded as the first correct measure of ambiguity of a fuzzy set. In addition, Yager (1979) also defined an information measure of a fuzzy set based on the distance from the set to its complement set. Similarly, Kosko (1986, 1990) introduced another kind of fuzzy information measre by considering the distance from a set to its nearest nonfuzzy set and the distance from the set to its farthest nonfuzzy set. Another kind of fuzzy information measure with an exponential function was introduced by Pal and Pal (1989). Later on, they introduced the concept of higher r^{th} order entropy of a fuzzy set in their paper Pal and Pal (1992). Further, Bhandari and Pal (1993) made a survey on information measures on fuzzy sets and gave some new measures of fuzzy information. Kapur (1997) and Parkash (1998, 2001) also suggested new measures of fuzzy information and studied.

Let $\Delta_n = \{P = (p_1, p_2, \ldots, p_n),\ p_i \geq 0,\ i = 1,\ 2,\ \ldots, n \ \text{ and } \sum_{i=1}^{n} p_i = 1\}$ be the set of all probability distributions associated with a discrete random variable X taking finite values $x_1, x_2, \ldots, x_n$.

Boekee and Lubbe (1980) defined and studied R-norm information measure of the distribution P for $R \in \mathbb{R}^+$ as given by

$$H_R(P) = \frac{R}{R-1}\left[1 - \left(\sum_{i=1}^{n} p_i^R\right)^{\frac{1}{R}}\right] \quad ; \quad R > 0,\ R \neq 1. \tag{6.1.2}$$

The measure (6.1.2) is a real function from Δ_n to $\mathbb{R}^+$ and is called R-norm information measure. The most important property of this measure is that when $R \to 1$, it approaches to Shannon's entropy and in case $R \to \infty$, $H_R(P) \to (1 - \max p_i)$; $i = 1,\ 2, \ldots,\ n$.

Analogous to measure (6.1.2), Hooda (2004) proposed and characterized the following fuzzy information measure:

$$H_R(A) = \frac{R}{R-1}\left[\sum_{i=1}^{n} 1 - \left(\mu_A^R(x_i) + (1 - \mu_A(x_i))^R\right)^{\frac{1}{R}}\right]; R > 0,\ R \neq 1. \tag{6.1.3}$$

Further, Hooda and Ram (1998) gave a parametric generalization of (6.1.2) by

$$H_R^{\beta}(P) = \frac{R}{R+\beta-2}\left[1 - \left(\sum_{i=1}^{n} p_i^{\frac{R}{2-\beta}}\right)^{\frac{2-\beta}{R}}\right]; \tag{6.1.4}$$

where $0 < \beta \leq 1,\ R > 0$ and $R + \beta \neq 2$.

The measure (6.1.4) is called the generalized R-norm entropy of degree β and it reduces to (6.1.2), when $\beta \to 1$. In case $R \to 1$, (6.1.4) reduces to

$$H_1^{\beta}(P) = \frac{1}{\beta - 1}\left[1 - \left(\sum_{i=1}^{n} p_i^{1/(2-\beta)}\right)^{2-\beta}\right]; \tag{6.1.5}$$

where $0 < \beta \leq 1,\ R > 0$ and $R + \beta \neq 2$.

Setting $\theta = \frac{1}{2-\beta}$ in (6.1.5), we get

$$H^{\theta}(P) = \frac{\theta}{\theta - 1}\left[1 - \left(\sum_{i=1}^{n} p_i^{\theta}\right)^{\frac{1}{\theta}}\right]; \quad \frac{1}{2} < \theta \leq 1. \tag{6.1.6}$$

This is an information measure which has been mentioned by Arimoto (1971) as an example of a generalized class of information measures. It may also be noted that (6.1.6) approaches to Shannon's entropy when $\theta \to 1$.

Next, suppose that $P = (p_1, p_2, \ldots, p_n)$ and $Q = (q_1, q_2, \ldots, q_n)$ be two given probability distributions belonging to Δ_n. Kullback and Leibler (1951) obtained the measure of directed divergence of P from Q as

$$D(P : Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}. \tag{6.1.7}$$

Kullback (1959) suggested the measure of symmetric divergence as

$$J(P : Q) = D(P : Q) + D(Q : P) = \sum_{i=1}^{n} (p_i - q_i) \log \frac{p_i}{q_i}. \tag{6.1.8}$$

Motivated by Kullback and Leibler measure, Bhandari and Pal (1993) suggested the following fuzzy directed divergence measure of fuzzy set A from B:

$$I(A, B) = \sum_{i=1}^{n}\left[\mu_A(x_i) \log \frac{\mu_A(x_i)}{\mu_B(x_i)} + (1 - \mu_A(x_i)) \log \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))}\right], \tag{6.1.9}$$

and the analogous fuzzy symmetric divergence measure by

$$J(A, B) = I(A, B) + I(B, A),$$

which on simplification gives

$$J(A,B) = \sum_{i=1}^{n} [(\mu_A(x_i) - \mu_B(x_i)] \log \frac{\mu_A(x_i)(1-\mu_A(x_i))}{\mu_B(x_i)(1-\mu_B(x_i))}. \tag{6.1.10}$$

Further, analogous to Havrda and Charvát's (1967) measure of directed divergence given by

$$D^{\beta}(P:Q) = \frac{1}{\beta-1}\sum_{i=1}^{n} (p_i^{\beta} q_i^{1-\beta} - 1); \quad \beta > 0,\ \beta \neq 1, \tag{6.1.11}$$

Hooda (2004) suggested the following measures of fuzzy directed divergence and symmetric divergence measure:

$$I^{\beta}(A,B) = \frac{1}{\beta-1}\sum_{i=1}^{n} \left[\mu_A^{\beta}(x_i)\mu_B^{1-\beta}(x_i) + (1-\mu_A(x_i))^{\beta}(1-\mu_B(x_i))^{1-\beta} - 1\right]; \tag{6.1.12}$$

where $\beta > 0$, $\beta \neq 1$

and

$$J^{\beta}(A,B) = I^{\beta}(A,B) + I^{\beta}(B,A) \tag{6.1.13}$$

respectively.

A new generalized measure of fuzzy information analogous to (6.1.4) is proposed and its validity to be a fuzzy information measure is proved in Section 6.2. In Section 6.3, we propose a generalized fuzzy directed divergence measure analogous to a R-norm directed divergence and prove its validity. In Section 6.4, we investigate the monotonic nature of the generalized measure of fuzzy information and the R-norm fuzzy directed divergence. R-norm generalized measures of total ambiguity and Fuzzy information improvement are also studied in Section 6.5.

6.2 A Generalized *R*-norm Fuzzy Information Measure

Analogous to (6.1.4), we propose the following measure of fuzzy information:

$$H_R^\beta(A) = \frac{R}{R+\beta-2}\left[\sum_{i=1}^{n} 1 - \left[(\mu_A(x_i))^{\frac{R}{2-\beta}} + (1-\mu_A(x_i))^{\frac{R}{2-\beta}}\right]^{\frac{2-\beta}{R}}\right]; \qquad (6.2.1)$$

where $0 < \beta \leq 1,\ R > 0\ \ R + \beta \neq 2$ and prove its validity in the next theorem.

Theorem 6.1: The measure (6.2.1) is a valid measure of fuzzy information.

Proof: To prove that the measure (6.2.1) is a valid fuzzy information measure, we shall show that four properties (P1) to (P4) are satisfied.

The measure (6.2.1) can be written

$$H_R^\beta(A) = \lambda\left[\sum_{i=1}^{n} 1 - \left[(\mu_A(x_i))^{\nu} + (1-\mu_A(x_i))^{\nu}\right]^{\frac{1}{\nu}}\right]; \qquad (6.2.2)$$

where $\lambda = \frac{R}{R+\beta-2},\ \nu = \frac{R}{2-\beta},\ \nu > 0,\ \nu \neq 1.$

P1 (*Sharpness*):

If $H_R^\beta(A) = 0$, then

$$(\mu_A(x_i))^{\nu} + (1-\mu_A(x_i))^{\nu} \ = \ 1. \qquad (6.2.3)$$

Since $\nu(\neq 1) > 0$, therefore, (6.2.3) is satisfied in case $\mu_A(x_i)$= 0 or 1, $\forall i = 1,\ 2, \ldots,\ n$.

Conversely, if A be a non-fuzzy set, then either $\mu_A(x_i)$= 0 or $\mu_A(x_i)$= 1. It implies $(\mu_A(x_i))^{\nu} + (1-\mu_A(x_i))^{\nu} = 1$ for $\nu > 0,\ \nu \neq 1$, for which $H_R^\beta(A)$= 0. Hence, $H_R^\beta(A)$= 0 if and only if A is non-fuzzy set or crisp set.

P2 (*Maximality*):

Differentiating $H_R^\beta(A)$ with respect to $\mu_A(x_i)$, we have

$$\frac{\partial H_R^\beta}{\partial \mu_A(x_i)} = -\lambda[(\mu_A(x_i))^\nu + (1-\mu_A(x_i))^\nu]^{\frac{1-\nu}{\nu}} \left[(\mu_A(x_i))^{\nu-1} - (1-\mu_A(x_i))^{\nu-1}\right]. \tag{6.2.4}$$

Let $0 \leq \mu_A(x_i) < 0.5$, then two cases arise

Case 1: $R > 2-\beta$

In this case we have $\lambda > 0$, $\nu > 1$ and $(\mu_A(x_i))^{\nu-1} - (1-\mu_A(x_i))^{\nu-1} < 0$ which implies that $\frac{\partial H_R^\beta}{\partial \mu_A(x_i)} > 0$.

Case 2: $R < 2-\beta$

In this case we have $\lambda < 0$, $\nu < 1$ and $(\mu_A(x_i))^{\nu-1} - (1-\mu_A(x_i))^{\nu-1} > 0$ which implies that $\frac{\partial H_R^\beta}{\partial \mu_A(x_i)} > 0$.

Hence, $H_R^\beta(A)$ is an increasing function of $\mu_A(x_i)$ satisfying $0 \leq \mu_A(x_i) < 0.5$. Similarly, it can be proved that $H_R^\beta(A)$ is a decreasing function of $\mu_A(x_i)$ satisfying $0.5 < \mu_A(x_i) \leq 1$. It is evident that $\frac{\partial H_R^\beta}{\partial \mu_A(x_i)} = 0$, when $\mu_A(x_i) = 0.5$. Hence, $H_R^\beta(A)$ is a concave function and it has a global maximum at $\mu_A(x_i) = 0.5$. It shows that $H_R^\beta(A)$ is maximum if and only if A is the most fuzzy set.

P3 (*Resolution*):

Since $H_R^\beta(A)$ is an increasing function of $\mu_A(x_i)$ in [0, 0.5) and decreasing function in (0.5, 1], therefore

$$\mu_{A^*}(x_i) \leq \mu_A(x_i) \Rightarrow H_R^\beta(A^*) \leq H_R^\beta(A) \text{ in } [0, 0.5) \tag{6.2.5}$$

and

$$\mu_{A^*}(x_i) \geq \mu_A(x_i) \Rightarrow H_R^\beta(A^*) \leq H_R^\beta(A) \text{ in } (0.5, 1]. \tag{6.2.6}$$

Taking (6.2.5) and (6.2.6) together, we get $H_R^\beta(A^*) \leq H_R^\beta(A)$.

P4 (*Symmetry*):

Evidently, from the definition of $H_R^\beta(A)$ and with $\mu_{\bar{A}}(x_i) = 1 - \mu_A(x_i)$, we may conclude that $H_R^\beta(\overline{A}) = H_R^\beta(A)$.

Hence, $H_R^\beta(A)$ satisfies all the properties of fuzzy information measure and therefore, it is a valid measure of fuzzy information.

It may be noted that (6.2.1) reduces to (6.1.3), when $\beta = 1$ and reduces to (6.1.1) when $\beta = 1$ and $R \to 1$. In case $\beta = 1$ and $R \to \infty$, (6.2.1) reduces to $\sum_{i=1}^{n} \left[1 - \max\{\mu_A(x_i),\ 1-\mu_A(x_i)\}\right]$.

6.3 *R*-norm Fuzzy Directed Divergence Measure

Let $P(p_1, p_2, \ldots, p_n)$ and $Q(q_1, q_2, \ldots, q_n)$ be the posterior and prior probability distribution of a random variable respectively in an experiment. Recently, Hooda and Sharma (2007) defined the R-norm directed divergence given by

$$D_R(P:Q) = \frac{R}{R-1}\left[\left(\sum_{i=1}^{n} p_i^R q_i^{1-R}\right)^{\frac{1}{R}} - 1\right] \tag{6.3.1}$$

and R-norm measure of inaccuracy given by

$$D_R(P/Q) = D_R(P:Q) + H_R(P) = \frac{R}{R-1}\left[\left(\sum_{i=1}^{n} p_i^R q_i^{1-R}\right)^{\frac{1}{R}} - \left(\sum_{i=1}^{n} p_i^R\right)^{\frac{1}{R}}\right]. \tag{6.3.2}$$

It may be seen that when $R \to 1$, (6.3.1) reduces to

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

which is the measure of directed divergence due to Kullback and Leibler (1951) and (6.3.2) reduces to Kerridge inaccuracy due to Kerridge (1961)

$$D(P/Q) = -\sum_{i=1}^{n} p_i \log q_i.$$

Analogous to (6.3.1) we propose the following measure of fuzzy directed divergence of fuzzy set A from fuzzy set B:

$$I_R(A,B) = \frac{R}{R-1}\sum_{i=1}^{n}\left[\left(\mu_A^R(x_i)\mu_B^{1-R}(x_i) + (1-\mu_A(x_i))^R(1-\mu_B(x_i))^{1-R}\right)^{\frac{1}{R}} - 1\right], \tag{6.3.3}$$

where $R > 0$, $R \neq 1$ and measure of fuzzy symmetric divergence

$$J_R(A,B) = I_R(A,B) + I_R(B,A). \tag{6.3.4}$$

Next, we show that $I_R(A,B)$ is a valid measure i.e., $I_R(A,B) \geq 0$ with equality if $\mu_A(x_i) = \mu_B(x_i)$ for each $i = 1, 2, \ldots, n$.

Let $\sum\limits_{i=1}^{n} \mu_A(x_i) = s, \quad \sum\limits_{i=1}^{n} \mu_B(x_i) = t$, then

$$\sum_{i=1}^{n}\left(\frac{\mu_A(x_i)}{s}\right)^R\left(\frac{\mu_B(x_i)}{t}\right)^{1-R} - 1 \geq 0$$

or

$$\sum_{i=1}^{n} \mu_A^R(x_i)\mu_B^{1-R}(x_i) \geq \ s^R t^{1-R}. \tag{6.3.5}$$

Similarly, we can write

$$\sum_{i=1}^{n} (1-\mu_A(x_i))^R(1-\mu_B(x_i))^{1-R} \ \geq \ (n-s)^R(n-t)^{1-R}. \tag{6.3.6}$$

Adding (6.3.5) and (6.3.6), we get

$$\sum_{i=1}^{n} \mu_A^R(x_i)\mu_B^{1-R}(x_i) + (1-\mu_A(x_i))^R(1-\mu_B(x_i))^{1-R} \geq s^R t^{1-R} + (n-s)^R(n-t)^{1-R}. \tag{6.3.7}$$

Case 1: $0 < R < 1$

Let $\mu_A^R(x_i)\mu_B^{1-R}(x_i) + (1-\mu_A(x_i))^R(1-\mu_B(x_i))^{1-R} = x_i$, then $x_i < 1$ and $\frac{1}{R} > 1$ which implies $x_i - 1 > (x_i)^{1/R} - 1$. Since $\frac{R}{R-1} < 0$, therefore $\frac{R}{R-1}\sum\limits_{i=1}^{n}\left[(x_i)^{1/R} - 1\right] \geq \frac{R}{R-1}\sum\limits_{i=1}^{n}(x_i - 1)$. Thus, we have

$$I_R(A,B) \geq \frac{R}{R-1}\left[s^R t^{1-R} + (n-s)^R(n-t)^{1-R} - n\right].$$

Further let $\phi(s) = \frac{R}{R-1}\left[s^R t^{1-R} + (n-s)^R (n-t)^{1-R} - n\right]$, then

$\phi'(s) = \frac{R}{R-1}\left[R\left(\frac{s}{t}\right)^{R-1} - R\left(\frac{n-s}{n-t}\right)^{R-1}\right]$ and

$\phi''(s) = R^2\left[\frac{1}{t}\left(\frac{s}{t}\right)^{R-2} + \frac{1}{n-t}\left(\frac{n-s}{n-t}\right)^{R-2}\right] > 0.$

This shows that $\phi(s)$ is a convex function of s whose minimum value arises when $\frac{s}{t}\left(=\frac{n-s}{n-t}\right) = 1$ and is equal to zero. Hence, $\phi(s) > 0$ and vanishes only when $s = t$.

Case 2: $R > 1$

In this case (6.3.7) can be written

$$\left(\sum_{i=1}^{n} \mu_A^R(x_i)\mu_B^{1-R}(x_i) + (1-\mu_A(x_i))^R(1-\mu_B(x_i))^{1-R} - 1\right)^{1/R}$$
$$\geq \left(s^R t^{1-R} + (n-s)^R(n-t)^{1-R} - n\right)^{1/R}. \qquad (6.3.8)$$

Also, we have

$$\sum_{i=1}^{n}\left[\left(\mu_A^R(x_i)\mu_B^{1-R}(x_i) + (1-\mu_A(x_i))^R(1-\mu_B(x_i))^{1-R}\right)^{\frac{1}{R}} - 1\right]$$
$$\geq \left(\sum_{i=1}^{n} \mu_A^R(x_i)\mu_B^{1-R}(x_i) + (1-\mu_A(x_i))^R(1-\mu_B(x_i))^{1-R} - 1\right)^{1/R} \qquad (6.3.9)$$

Now (6.3.8) and (6.3.9) together implies that

$$I_R(A,B) \geq \frac{R}{R-1}\left[s^R t^{1-R} + (n-s)^R(n-t)^{1-R} - n\right]^{\frac{1}{R}}.$$

Let $\phi(s) = \frac{1}{R-1}\left[s^R t^{1-R} + (n-s)^R (n-t)^{1-R} - n\right]$, then

$\phi'(s) = \frac{1}{R-1}\left[R\left(\frac{s}{t}\right)^{R-1} - R\left(\frac{n-s}{n-t}\right)^{R-1}\right]$ and

$\phi''(s) = \left[\frac{R}{t}\left(\frac{s}{t}\right)^{R-2} + \frac{R}{n-t}\left(\frac{n-s}{n-t}\right)^{R-2}\right] > 0.$

This shows that $\phi(s)$ is a convex function of s whose minimum value arises when $\frac{s}{t}\left(=\frac{n-s}{n-t}\right) = 1$ and is equal to zero. Hence, $\phi(s) > 0$ and vanishes only when $s = t$ i.e., for all $R\,(\neq 1) > 0$, $I_R(A,B) \geq 0$ and vanishes only when $A = B$.

Thus $I_R(A, B)$ is a valid measure of directed divergence of fuzzy set A from fuzzy set B and consequently, the analogous measure of fuzzy symmetric divergence $J_R(A, B)$ is a valid measure.

It may be noted that $\lim_{R\to 1} I_R(A, B) = I(A, B)$ and $\lim_{R\to 1} J_R(A, B) = J(A, B)$, where $I(A, B)$ and $J(A, B)$ are the fuzzy directed divergence and symmetric divergence measures given by (6.1.9) and (6.1.10) respectively.

6.4 Monotonicity of Fuzzy Information and Fuzzy Directed Divergence Measures

Let A_1 = (0.2, 0.3, 0.4, 0.2, 0.3), A_2= (0.4, 0.3, 0.2, 0.2, 0.4), A_3= (0.3, 0.2, 0.3, 0.3, 0.3) be any three fuzzy sets in standard form. Consider four different values of R, say, 0.6, 1, 2, 3, and $0 < \beta \leq 1$. Using (6.2.1) we have constructed Table 6.1 listed below. Looking at Table 6.1, it is clear that the fuzzy information measure given by (6.2.1) is a monotonically decreasing function of β and R. This monotonic nature of the fuzzy information measure is shown in Figure 6.1 based on Table 6.1.

Table 6.1

β	$H^{\beta}_{0.6}(A)$			$H^{\beta}_{1}(A)$			$H^{\beta}_{2}(A)$			$H^{\beta}_{3}(A)$		
	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
0.001	7.76	7.86	7.84	4.41	4.47	4.46	2.89	2.96	2.94	2.45	2.52	2.50
0.002	7.75	7.85	7.84	4.41	4.47	4.46	2.89	2.96	2.94	2.45	2.52	2.50
0.005	7.74	7.83	7.82	4.40	4.47	4.46	2.89	2.95	2.94	2.45	2.52	2.50
0.01	7.71	7.80	7.79	4.39	4.46	4.45	2.89	2.95	2.94	2.45	2.52	2.49
0.02	7.65	7.75	7.73	4.38	4.44	4.43	2.88	2.94	2.93	2.44	2.51	2.49
0.05	7.49	7.58	7.56	4.32	4.39	4.38	2.86	2.92	2.91	2.43	2.50	2.48
0.1	7.22	7.31	7.29	4.24	4.30	4.29	2.83	2.89	2.88	2.40	2.48	2.45
0.2	6.71	6.80	6.79	4.06	4.13	4.12	2.76	2.83	2.81	2.36	2.43	2.40
0.3	6.25	6.33	6.32	3.90	3.96	3.95	2.70	2.76	2.75	2.31	2.39	2.36
0.4	5.82	5.90	5.89	3.74	3.80	3.79	2.63	2.70	2.68	2.26	2.34	2.31
0.5	5.43	5.50	5.49	3.59	3.65	3.64	2.56	2.63	2.61	2.21	2.29	2.26
0.6	5.06	5.13	5.12	3.45	3.50	3.49	2.50	2.57	2.54	2.16	2.25	2.21
0.7	4.73	4.79	4.78	3.30	3.36	3.35	2.43	2.50	2.48	2.11	2.20	2.15
0.8	4.41	4.48	4.47	3.16	3.22	3.21	2.36	2.43	2.40	2.06	2.15	2.10
0.9	4.12	4.18	4.17	3.03	3.09	3.08	2.29	2.36	2.33	2.01	2.10	2.04
1	3.85	3.91	3.90	2.89	2.95	2.94	2.21	2.29	2.26	1.95	2.05	1.99

It can be shown analytically that $\frac{d}{d\beta}\left(H^{\beta}_{R}(A)\right) \leq 0;\ \forall\, 0 < \beta \leq 1$ and $\frac{d}{dR}\left(H^{\beta}_{R}(A)\right) \leq 0;\ \forall\ R > 0$ and this implies monotonic decreasing nature of the fuzzy information measure with respect to β and R respectively. However, it is observed that for the most fuzzy set, the maximum value of $H^{\beta}_{R}(A)$ depends on the value of β and R, but it will be less than or equal to n. Similarly, monotonic nature of the fuzzy directed divergence measure given by (6.3.3) can be observed in the above computed Table 6.2 for three different sample pairs of fuzzy sets given by $A_1 = (0.3,\ 0.5,\ 0.3,\ 0.2,\ 0.1)$; $A_2 = (0.4,\ 0.3,\ 0.4,\ 0.2,\ 0.5)$;
$A_3 = (0.5,\ 0.2,\ 0.2,\ 0.3,\ 0.4)$; $B_1 = (0.2,\ 0.4,\ 0.4,\ 0.3,\ 0.2)$;
$B_2 = (0.2,\ 0.4,\ 0.4,\ 0.2,\ 0.2)$; and $B_3 = (0.3,\ 0.4,\ 0.4,\ 0.2,\ 0.3)$.

Table 6.2 shows that the fuzzy directed divergence given by (6.3.3) is monotonic increasing function of R. This monotonic increasing nature of the fuzzy directed divergence is also shown in Figure 6.2 based on Table 6.2. This monotonic nature of the fuzzy directed divergence can also be proved analytically by showing that $\frac{d}{dR}\left(I_R(A,B)\right) \geq 0;\ \ \forall R > 0$.

Table 6.2

R	$I_R(A_1, B_1)$	$I_R(A_2, B_2)$	$I_R(A_3, B_3)$
0.1	0.013827	0.029164	0.032484
0.2	0.027531	0.059652	0.065004
0.5	0.067891	0.159083	0.162301
0.8	0.107101	0.269945	0.258256
1	0.132602	0.349394	0.320959
1.2	0.157595	0.432339	0.382348
1.5	0.19415	0.561022	0.471532
2	0.252654	0.777309	0.611291
5	0.543459	1.674712	1.207667
10	0.817193	2.156136	1.640287
50	1.11155	2.563536	2.06191
75	1.135962	2.597848	2.096904
100	1.148134	2.615028	2.114372
150	1.160285	2.632223	2.131821
200	1.166353	2.640828	2.140539
300	1.172415	2.649436	2.149253
400	1.175444	2.653742	2.153608
420	1.175877	2.654357	2.15423
430	1.176078	2.654644	2.154519
440	1.17627	2.654917	2.154795

6.5 Measures of Total Ambiguity and Fuzzy Information Improvement

6.5.1 Total ambiguity

Let A and B be two fuzzy sets. The total ambiguity of the fuzzy set A about set B is the sum of two components:

- Fuzzy entropy present in the fuzzy set A, and
- fuzzy directed divergence of A from B measured by $I(A, B)$.

Using Havrda and Charvát's measure, Kapur (1997) estimated the total fuzzy ambiguity as

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

Corresponding to fuzzy information measure (6.1.3) and the proposed fuzzy directed divergence (6.3.3), total ambiguity is given by

$$\begin{aligned} TA &= \frac{R}{R-1}\Bigg[\sum_{i=1}^{n}\left(1-(\mu_A^R(x_i)+(1-\mu_A(x_i))^R\right)^{\frac{1}{R}} \\ &\qquad + \sum_{i=1}^{n}\left[\left\{\mu_A^R(x_i)\mu_B^{1-R}(x_i)+(1-\mu_A(x_i))^R(1-\mu_B(x_i))^{1-R}\right\}^{\frac{1}{R}}-1\right]\Bigg] \\ &= \frac{R}{R-1}\sum_{i=1}^{n}\Bigg[\left(\mu_B(x_i)\left(\frac{\mu_A(x_i)}{\mu_B(x_i)}\right)^R+(1-\mu_B(x_i))\left(\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right)^R\right)^{\frac{1}{R}} \\ &\qquad - \left(\mu_A^R(x_i)+(1-\mu_A(x_i))^R\right)^{\frac{1}{R}}\Bigg]. \end{aligned}$$

Total ambiguity is a fuzzy measure of inaccuracy analogous to Kerridge (1961) inaccuracy and is related to two fuzzy sets. It is not symmetric as we get something different if we interchange the role of the fuzzy sets A and B.

6.5.2 *R*-norm fuzzy information improvement measure

Let P and Q be observed and predicted distributions of a random variable respectively. Let $R = (r_1, r_2, \dots, r_n)$ be the revised probability distribution of Q, then

$$I(P:Q) - I(P:R) = \sum_{i=1}^{n} p_i \log \frac{r_i}{q_i}, \tag{6.5.1}$$

which is known as Theil's measure (1967) of information improvement and has found wide applications in economics, accounts and financial management. Similarly, suppose the correct fuzzy set is A and originally our estimate for it was the fuzzy set B and that was revised to fuzzy set C. The original ambiguity was $I(A, B)$ and final ambiguity is $I(A, C)$, so the reduction in ambiguity is

$$\begin{aligned} I(A,B,C) &= I(A,B) - I(A,C), \\ &= \sum_{i=1}^{n} \left[\mu_A(x_i) \log \frac{\mu_C(x_i)}{\mu_B(x_i)} + (1-\mu_A(x_i)) \log \frac{(1-\mu_C(x_i))}{(1-\mu_B(x_i))} \right]. \end{aligned} \tag{6.5.2}$$

The measure $I(A, B, C)$ given by (6.5.2) can be called fuzzy information improvement measure. Corresponding to fuzzy directed divergence given by (6.3.3), the reduction in ambiguity is given by

$$\begin{aligned} I_R(A,B,C) &= I_R(A,B) - I_R(A,C) \\ &= \frac{R}{R-1} \sum_{i=1}^{n} \Big[\left\{ \mu_A^R(x_i)\mu_B^{1-R}(x_i) + (1-\mu_A(x_i))^R (1-\mu_B(x_i))^{1-R} \right\}^{\frac{1}{R}} \\ &\quad - \left\{ \mu_A^R(x_i)\mu_C^{1-R}(x_i) + (1-\mu_A(x_i))^R (1-\mu_C(x_i))^{1-R} \right\}^{\frac{1}{R}} \Big], \end{aligned} \tag{6.5.3}$$

which can be called as R-norm fuzzy information improvement measure. It can also be proved that $I_R(A, B, C) \to I(A, B, C)$ i.e., (6.5.3) reduces to (6.5.2), when $R \to 1$.

Chapter 7

Measures of 'useful' Fuzzy Information

7.1 Introduction

Zadeh (1965) introduced the concept of fuzzy sets and developed the theory to measure the ambiguity of a fuzzy set. Fuzzy set theory makes use of entropy to measure the degree of fuzziness in a fuzzy set, which is also called fuzzy entropy or fuzzy information measure (Ebanks (1983), Pal (1994)). Fuzzy entropy is the measurement of fuzziness in a fuzzy set, and thus has especial important position in fuzzy systems such as fuzzy pattern recognition systems, fuzzy neural network systems, fuzzy knowledge base systems, fuzzy decision making systems, fuzzy control systems and fuzzy management information systems.

Let X is a discrete random variable with probability distribution $P = (p_1, p_2, \ldots, p_n)$ in an experiment. (X, P) is a *discrete probabilistic framework*. The information contained in this experiment is given by the well known Shannon's entropy (1948).

It has been already mentioned that the meaning of fuzzy entropy is different from the classical Shannon entropy because no probabilistic concept is needed in order to define it. Fuzzy entropy deals with vagueness and ambiguous uncertainties, while Shannon entropy deals with probabilistic uncertainties. De Luca and Termini (1972) characterized the fuzzy entropy with a set of postulates P1 to P4 and these have been used as a criterion for defining a new fuzzy entropy.

Analogous to entropy due to Shannon (1948), De Luca and Termini (1972) suggested the following measure of fuzzy entropy:

$$H(A) = -\sum_{i=1}^{n} [\mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log (1 - \mu_A(x_i))]. \qquad (7.1.1)$$

As (7.1.1) satisfies all four properties (P1) to (P4), so it is a valid measure of fuzzy entropy. Later on Bhandari and Pal (1993) made a survey on information measures on fuzzy sets and gave some measures of fuzzy entropy. Analogous to Rényi's (1961) entropy they have suggested the following measure:

$$H_\alpha(A) = \frac{1}{1-\alpha} \sum_{i=1}^{n} \log [\mu_A^\alpha(x_i) + (1 - \mu_A(x_i))^\alpha]; \qquad \alpha \neq 1,\ \alpha > 0 \qquad (7.1.2)$$

and analogous to Pal and Pal's (1989) exponential entropy they introduced

$$H_e(A) = \frac{1}{n\sqrt{e} - 1} \sum_{i=1}^{n} \log \left[\mu_A(x_i) e^{1-\mu_A(x_i)} + (1 - \mu_A(x_i))\, e^{\mu_A(x_i)} - 1\right]. \qquad (7.1.3)$$

It may be noted that Shannon entropy does not take into account the effectiveness or importance of the events, while in some practical situations of probabilistic nature subjective considerations also play their own role. Belis and Guisau (1968) considered qualitative aspect of information and attached a utility distribution $U = (u_1, u_2, ..., u_n)$, where $u_i > 0$ for each i and is utility or importance of an event x_i whose probability of occurrence is p_i. In general u_i is independent of p_i. They suggested that the occurrence of an event removes two types of uncertainty - the quantitative type related to its probability of occurrence and the qualitative type related to its utility (importance) for the fulfillment of some goal set by the experimenter. Bhaker and Hooda (1993) gave the generalized mean value characterization of the useful information measures for incomplete probability distributions:

$$H(P;U) = -\frac{\sum_{i=1}^{n} u_i p_i \log p_i}{\sum_{i=1}^{n} u_i p_i}, \tag{7.1.4}$$

and

$$H_\alpha(P;U) = \frac{1}{1-\alpha} \log \frac{\sum_{i=1}^{n} u_i p_i^\alpha}{\sum_{i=1}^{n} u_i p_i}; \;\; \alpha \neq 1, \;\; \alpha > 0. \tag{7.1.5}$$

The first attempt to quantify the uncertainty associated with a fuzzy event in the context of a discrete probabilistic framework appears to have been made by Zadeh (1968), who defined the (weighted) entropy of A with respect to (X, P) as

$$H(A, P) = -\sum_{i=1}^{n} \mu_A(x_i) p_i \log p_i, \tag{7.1.6}$$

where μ_A is the membership function of A and p_i is the probability of occurrence of x_i. One can notice that this situation contains the different types of uncertainties, e.g., randomness, ambiguity, and vagueness; i.e., randomness and fuzziness. This measure does not satisfy properties (P1) to (P4). $H(A, P)$ of a fuzzy event with respect to P is less than Shannon's entropy which is of P alone.

In Section 7.2 of the present chapter, we introduce a new concept of 'useful' fuzzy information measure by attaching utilities with uncertainties of fuzziness and probabilities of randomness. In Section 7.3, a new measure of total 'useful' fuzzy information, by considering the usefulness of an event along with fuzzy uncertainties and random uncertainties is introduced and studied. In Section 7.4, we define a measure of 'useful' fuzzy directed divergence of fuzzy set A from fuzzy set B and also prove its validity. The constrained optimization of 'useful' fuzzy information and 'useful' fuzzy directed divergence measures is discussed in Sections 7.5 and 7.6 respectively.

7.2 'Useful' Fuzzy Information Measures

Let $x_1, x_2, \ldots, x_n$ are members of the universe of discourse having probabilities of occurrence $p_1, p_2, \ldots, p_n$ and utility or importance $u_1, u_2, \ldots, u_n$ respectively. Let A be a fuzzy set. Then $\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n)$ are ambiguities or uncertainties which lie between 0 and 1, but these are not probabilities because their sum is not unity. However,

$$\Phi_A(x_i) = \frac{\mu_A(x_i)}{\sum_{i=1}^{n} \mu_A(x_i)}, \quad i = 1, 2, \ldots, n, \tag{7.2.1}$$

is a probability distribution.

On attaching utilities with the uncertainties of fuzziness and probabilities of randomness having utilities, we suggest the following measure of fuzziness of a fuzzy set is suggested analogous to De Luca and Termini's fuzzy entropy (7.1.1):

$$H(A;P;U) = -\frac{\sum_{i=1}^{n} u_i p_i [\mu_A(x_i) \log \mu_A(x_i) + (1-\mu_A(x_i)) \log(1-\mu_A(x_i))]}{\sum_{i=1}^{n} u_i p_i}, \tag{7.2.2}$$

which can be called as 'useful' fuzzy information measure of fuzzy set A.

Theorem 7.1: The measure (7.2.2) is a valid measure of fuzzy information.

***Proof*:**To prove that the measure (7.2.2) is a valid fuzzy information measure, we prove that four properties (P1) to (P4) hold.

P1 (*Sharpness*):

If $H(A;P;U) = 0$ then

$$\sum_{i=1}^{n} u_i p_i [\mu_A(x_i) \log \mu_A(x_i) + (1-\mu_A(x_i)) \log(1-\mu_A(x_i))] = 0,$$

$$\Rightarrow \mu_A(x_i) \log \mu_A(x_i) + (1-\mu_A(x_i)) \log(1-\mu_A(x_i)) = 0, \qquad [u_i, p_i > 0 \ \ \forall i]$$

$\Rightarrow$ either $\mu_A(x_i) = 0$ or 1 $\forall i = 1, 2, \ldots, n$,

$\Rightarrow$ A is a crisp set.

Conversely, let A be a crisp set, then either $\mu_A(x_i) = 0$ or $\mu_A(x_i) = 1 \quad \forall i$.

$\Rightarrow \mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log (1 - \mu_A(x_i)) = 0\,, \quad \forall i$

$\Rightarrow \sum_{i=1}^{n} u_i p_i [\mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log (1 - \mu_A(x_i))] = 0, \;\; [u_i, p_i > 0 \;\forall i]$

$\Rightarrow H(A; P; U) = 0.$

Hence, $H(A; P; U)$= 0 if and only if A is non-fuzzy set or crisp set.

P2 (*Maximality*):

Differentiating $H(A; P; U)$ with respect to $\mu_A(x_i)$, we have

$$\frac{\partial H(A; P; U)}{\partial \mu_A(x_i)} = \sum_{i=1}^{n} u_i p_i \log \frac{1 - \mu_A(x_i)}{\mu_A(x_i)} \quad . \tag{7.2.3}$$

Case 1: $0 < \mu_A(x_i) < 0.5$,

In this case, $\log \frac{1-\mu_A(x_i)}{\mu_A(x_i)} > 0 \Rightarrow \frac{\partial H(A;P;U)}{\partial \mu_A(x_i)} > 0$.

Thus $H(A; P; U)$ is an increasing function of $\mu_A(x_i)$ satisfying $0 < \mu_A(x_i) < 0.5$.

Case 2: $0.5 < \mu_A(x_i) < 1$,

In this case, $\log \frac{1-\mu_A(x_i)}{\mu_A(x_i)} < 0 \Rightarrow \frac{\partial H(A;P;U)}{\partial \mu_A(x_i)} < 0$.

Thus $H(A; P; U)$ is a decreasing function of $\mu_A(x_i)$ satisfying $0.5 < \mu_A(x_i) < 1$.

Also from (7.2.3) note that

$$\frac{\partial H(A; P; U)}{\partial \mu_A(x_i)} = 0; \quad \text{when} \quad \mu_A(x_i) = 0.5 \tag{7.2.4}$$

Hence, $H(A; P; U)$ is a concave function and it has a global maximum at $\mu_A(x_i) = 0.5$ which shows that $H(A; P; U)$ is maximum if and only if A is the most fuzzy set.

P3 (*Resolution*):

Since $H(A; P; U)$ is an increasing function of $\mu_A(x_i)$ in [0, 0.5) and decreasing function in (0.5, 1], therefore

$$\mu_{A^*}(x_i) \le \mu_A(x_i) \Rightarrow H(A^*; P; U) \le H(A; P; U) \quad \text{in } [0,\ 0.5) \tag{7.2.5}$$

and

$$\mu_{A^*}(x_i) \geq \mu_A(x_i) \Rightarrow H(A^*;P;U) \leq H(A;P;U) \quad \text{in } (0.5,\ 1] \tag{7.2.6}$$

Taking (7.2.5) and (7.2.6) together, we get $H(A^*;P;U) \leq H(A;P;U)$.

P4 (*Symmetry*):

Evidently, from the definition of $H(A;P;U)$ and with $\mu_{\overline{A}}(x_i) = 1 - \mu_A(x_i)$, we conclude that $H(\overline{A};P;U) = H(A;P;U)$. Hence, $H(A;P;U)$ satisfies all the properties of fuzzy information measure and therefore is a valid measure of 'useful' fuzzy information.

The measure (7.2.2) can further be generalized parametrically. Corresponding to (7.1.2) and (7.1.3), we have respectively

$$H_\alpha(A;P;U) = \frac{1}{1-\alpha} \frac{\sum_{i=1}^{n} u_i p_i \log\left[\mu_A^\alpha(x_i) + (1-\mu_A(x_i))^\alpha\right]}{\sum_{i=1}^{n} u_i p_i}; \quad \alpha(\neq 1) > 0 \tag{7.2.7}$$

and

$$H_e(A;P;U) = \frac{1}{n\sqrt{e}-1} \frac{\sum_{i=1}^{n} u_i p_i \log\left[\mu_A(x_i) e^{1-\mu_A(x_i)} + (1-\mu_A(x_i))\, e^{\mu_A(x_i)} - 1\right]}{\sum_{i=1}^{n} u_i p_i} \tag{7.2.8}$$

'useful' fuzzy information measures.

It may be noted that when $\alpha \to 1$, (7.2.7) reduces to (7.2.2). Thus we can define a 'useful' fuzzy information measure corresponding to the generalized measures of fuzzy entropies.

7.3 Total 'Useful' Fuzzy Information Measure

Total 'useful' fuzzy information measure is related to integration of fuzzy and probabilistic uncertainties with utilities. There have been several attempts to combine probabilistic and fuzzy uncertainties when (X, P) is a discrete probability framework. The entropy given by (7.1.6) is a measure of uncertainty associated with a fuzzy event, and was the first composite measure of probabilistic and fuzzy uncertainty.

De Luca and Termini (1972) introduced and studied the following total information measure of a fuzzy system which has been used by some authors as a measure of uncertainties of fuzziness and randomness of events in an experiment:

(a) The average uncertainty deduced from the "random" nature of the experiment is computed by Shannon's entropy given by

$$H(P) = -\sum_{i=1}^{n} p_i \log p_i \tag{7.3.1}$$

(b) The uncertainty that arises from the fuzziness of the fuzzy set relative to the ordinary set given by the fuzzy entropy (7.1.1).

(c) The statistical average, m, of the ambiguity of the whole set is given by

$$m(\mu_A, p_1, p_2, ..., p_n) = -\sum_{i=1}^{n} p_i[\mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log (1 - \mu_A(x_i))]. \tag{7.3.2}$$

(d) The total information measure is obtained by adding two kinds of uncertainties (7.3.1) and (7.3.2).

Analogously, we combine probabilistic and fuzzy uncertainties with utilities when $(A; P; U)$ is a discrete probabilistic framework with utilities and fuzzy set A which is characterized by membership function μ_A. If we also consider importance or usefulness of an event, then total 'useful' information measure is a measure of fuzzy uncertainties, random uncertainties and utilities of events and that is computed as follows:

(i) If we consider the uncertainties with 'usefulness' from "random" nature of an experiment, then average measure of these uncertainties introduced and characterized by Bhaker and Hooda (1993) is given by

$$H(P;U) = -\frac{\sum_{i=1}^{n} u_i p_i \log p_i}{\sum_{i=1}^{n} u_i p_i}. \tag{7.3.3}$$

(ii) If we consider the uncertainty that arises from fuzziness of the fuzzy set, then we can compute the amount of ambiguity by taking the proposed 'useful' fuzzy information measure (7.2.2) and can be written as

$$H(A;P;U) = -\frac{\sum_{i=1}^{n} u_i p_i H_i(A)}{\sum_{i=1}^{n} u_i p_i}, \tag{7.3.4}$$

where $H_i(A) = [\mu_A(x_i) \log \mu_A(x_i) + (1 - \mu_A(x_i)) \log (1 - \mu_A(x_i))]$.

(iii) Total 'useful' fuzzy information measure of fuzzy set A in a random experiment is obtained by adding (7.3.3) and (7.3.4):

$$H_{Total}(A; P : U) = H(P;U) + H(A;P;U). \tag{7.3.5}$$

7.4 'Useful' Fuzzy Directed Divergence Measures

Let A and B be two standard fuzzy sets with same supporting points $x_1, x_2, \ldots, x_n$ and with fuzzy vectors $\mu_A(x_1), \mu_A(x_2), \ldots, \mu_A(x_n)$ and$\mu_B(x_1), \mu_B(x_2), \ldots, \mu_B(x_n)$. The simplest measure of fuzzy directed divergence as suggested Bhandari and Pal (1993) is given by

$$I(A,B) = \sum_{i=1}^{n} \left[\mu_A(x_i) \log \frac{\mu_A(x_i)}{\mu_B(x_i)} + (1 - \mu_A(x_i)) \ln \frac{(1 - \mu_A(x_i))}{(1 - \mu_B(x_i))} \right]. \tag{7.4.1}$$

Next, attaching utilities with the uncertainties of fuzziness of fuzzy set A from fuzzy set B and probabilities of randomness we define the following 'useful' measure of fuzzy directed divergence of fuzzy set A from fuzzy set B corresponding to (7.4.1):

$$I(A,B;P;U) = \frac{\sum_{i=1}^{n} u_i p_i \left[\mu_A(x_i)\log\frac{\mu_A(x_i)}{\mu_B(x_i)} + (1-\mu_A(x_i))\log\frac{(1-\mu_A(x_i))}{(1-\mu_B(x_i))}\right]}{\sum_{i=1}^{n} u_i p_i} \tag{7.4.2}$$

and measure of 'useful' fuzzy symmetric divergence as

$$J(A,B;P;U) = I(A,B;P;U) + I(B,A;P;U). \tag{7.4.3}$$

Further, we show that $I(A,B;P;U)$ is a valid measure i.e., $I(A,B;P;U) \geq 0$ with equality if $\mu_A(x_i) = \mu_B(x_i)$ for each $i = 1, 2, \ldots, n$.

Let $\sum_{i=1}^{n} \mu_A(x_i) = s$, $\sum_{i=1}^{n} \mu_B(x_i) = t$ and $\sum_{i=1}^{n} u_i p_i = u$, then

$$\sum_{i=1}^{n} u_i p_i \left(\mu_A(x_i)\log\frac{\mu_A(x_i)}{\mu_B(x_i)}\right) \geq us\log\frac{s}{t}. \tag{7.4.4}$$

Similarly, we have

$$\sum_{i=1}^{n} u_i p_i \left((1-\mu_A(x_i))\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right) \geq u(n-s)\log\frac{n-s}{n-t}. \tag{7.4.5}$$

Adding (7.4.4) and (7.4.5), we get

$$\sum_{i=1}^{n} u_i p_i \left(\mu_A(x_i)\log\frac{\mu_A(x_i)}{\mu_B(x_i)} + (1-\mu_A(x_i))\log\frac{1-\mu_A(x_i)}{1-\mu_B(x_i)}\right) \geq u\left[s\log\frac{s}{t} + (n-s)\log\frac{n-s}{n-t}\right] \tag{7.4.6}$$

Let $f(s) = s\log\frac{s}{t} + (n-s)\log\frac{n-s}{n-t}$, then
$f'(s) = \left(\log\frac{s}{t} - \log\frac{n-s}{n-t}\right)$ and
$f''(s) = \frac{1}{s} + \frac{1}{n-s} = \frac{n}{s(n-s)} > 0.$

Hence, $f''(s) > 0$, which shows that $f(s)$ is a convex function of s and has its minimum value when $\frac{s}{t} = \frac{n-s}{n-t} = \frac{n}{n} = 1$. Now, if $A = B$ i.e., $s = t$, then

$f(s) = 0$. Hence, $f(s) > 0$ and vanishes only when $s = t$. As $\sum_{i=1}^{n} u_i p_i = u > 0$, so $I(A, B; P; U) \geq 0$ and vanishes only when $A = B$.

Thus $I(A, B; P; U)$ is a valid measure of 'useful' fuzzy directed divergence of fuzzy sets A and B; and consequently, $J(A, B; P; U)$ is also a valid measure of 'useful' fuzzy symmetric divergence.

It may be noted that if $B = A_F$, the most fuzzy set i.e., $\mu_B(x_i) = 0.5 \quad \forall x_i$, then $I(A, A_F; P, U) = n \log 2 - H(A; P; U)$.

The measure (7.4.2) can further be generalized parametrically. Corresponding to the following fuzzy directed divergence measure defined in Chapter 4:

$$I_\alpha(A, B) = \frac{1}{\alpha - 1} \sum_{i=1}^{n} \log \left[\mu_A^\alpha(x_i) \mu_B^{1-\alpha}(x_i) + (1 - \mu_A(x_i))^\alpha (1 - \mu_B(x_i))^{1-\alpha} \right], \tag{7.4.7}$$

we suggest a new measure of 'useful' fuzzy directed divergence of fuzzy set A from fuzzy set B which is given by

$$I_\alpha(A, B; P; U) = \frac{1}{\alpha - 1} \frac{\sum_{i=1}^{n} u_i p_i \log \left[\mu_A^\alpha(x_i) \mu_B^{1-\alpha}(x_i) + (1 - \mu_A(x_i))^\alpha (1 - \mu_B(x_i))^{1-\alpha} \right]}{\sum_{i=1}^{n} u_i p_i} \tag{7.4.8}$$

and measure of 'useful' fuzzy symmetric divergence is given by

$$J_\alpha(A, B; P; U) = I_\alpha(A, B; P; U) + I_\alpha(B, A; P; U). \tag{7.4.9}$$

Further, it can be proved that $I_\alpha(A, B; P; U)$ is a valid measure by showing $I_\alpha(A, B; P; U)$ with equality if $\mu_A(x_i) = \mu_B(x_i)$ for each $i = 1, 2, \ldots, n$.

It may be noted that if $B = A_F$, the most fuzzy set i.e., $\mu_B(x_i) = 0.5 \quad \forall x_i$, then we have

$$I_\alpha(A, A_F; P, U) = n \log 2 - H_\alpha(A; P; U).$$

7.5 Constrained optimization of 'useful' Fuzzy Information Measure

Consider the 'useful' fuzzy information measure given by (7.2.2) subject to $\sum_{i=1}^{n} \mu_A(x_i) = \alpha$. Using Lagrange's multiplier method, we have

$$u_i p_i \log \frac{\mu_A(x_i)}{1-\mu_A(x_i)} = m \tag{7.5.1}$$

It implies

$$\mu_A(x_i) = \frac{e^{\frac{m}{u_i p_i}}}{1+e^{\frac{m}{u_i p_i}}} = \frac{1}{1+e^{-\frac{m}{u_i p_i}}}, \tag{7.5.2}$$

where m is determined by

$$\phi(m) = \sum_{i=1}^{n} \frac{e^{\frac{m}{u_i p_i}}}{1+e^{\frac{m}{u_i p_i}}} - \alpha = 0 \tag{7.5.3}$$

and

$$\phi'(m) = \sum_{i=1}^{n} \frac{\frac{1}{u_i p_i} e^{\frac{m}{u_i p_i}}}{\left(1+e^{\frac{m}{u_i p_i}}\right)^2} \geq 0. \tag{7.5.4}$$

It may be seen that

$$\phi(-\infty) = -\alpha, \qquad \phi(0) = \frac{n}{2} - \alpha, \qquad \phi(\infty) = n - \alpha. \tag{7.5.5}$$

In view of (7.5.5), we see that equation (7.5.3) has a unique root which is negative if $\alpha < \frac{n}{2}$ and is positive if $\alpha > \frac{n}{2}$. It is observed that if m is negative i.e., if $\alpha < \frac{n}{2}$, then $\mu_A(x_i) \leq \frac{1}{2}$ $\forall i$, if m is positive i.e., if $\alpha > \frac{n}{2}$, then $\mu_A(x_i) \geq \frac{1}{2}$ $\forall i$ and if $m = 0$ i.e., if $\alpha = \frac{n}{2}$, then $\mu_A(x_i) = \frac{1}{2}$ $\forall i$. Thus maximization is possible:

(i) all $\mu_A(x_i)$ are less than or equal to $\frac{1}{2}$

(ii) all $\mu_A(x_i)$ are greater than or equal to $\frac{1}{2}$

(iii) all are equal to $\frac{1}{2}$.

When $\mu_A(x_i)$ is replaced by $1-\mu_A(x_i)$, the fuzziness of the i^{th} element does not change, but its contribution to the sum $\sum_{i=1}^{n} \mu_A(x_i)$ changes. So it will be better to take a fuzzy set in standard form i.e., $\mu_A(x_i) \leq \frac{1}{2}\ \forall i$.

Differentiating (7.2.2) with respect to $\mu_A(x_i)$, we get

$$\frac{dH}{d\mu_A(x_i)} = -\sum_{i=1}^{n} u_i p_i \log \frac{\mu_A(x_i)}{1-\mu_A(x_i)} > 0 \quad \text{since } \mu_A(x_i) \leq \frac{1}{2}. \tag{7.5.6}$$

Differentiating (7.5.2) with respect to m, we get

$$\frac{d\mu_A(x_i)}{dm} = \frac{\frac{1}{u_i p_i} e^{-\frac{m}{u_i p_i}}}{\left(1+e^{\frac{m}{u_i p_i}}\right)^2} > 0. \tag{7.5.7}$$

Differentiating (7.5.3) with respect to m, we get

$$\frac{d\alpha}{dm} = \sum_{i=1}^{n} \frac{\frac{1}{u_i p_i} e^{-\frac{m}{u_i p_i}}}{\left(1+e^{\frac{m}{u_i p_i}}\right)^2} > 0. \tag{7.5.8}$$

Taking (7.5.6), (7.5.7) and (7.5.8) together, we can conclude that $H(A;P;U)$ is an increasing function of α. Hence, $H(A;P;U)$ is maximum when $\alpha = \frac{n}{2}$ and consequently, $m = 0$ and $\mu_A(x_i) = \frac{1}{2}\ \forall i$. Thus $H_{\max}(A;P;U) = \log 2$ and $H_{\max}$ increases from 0 to $\log 2$ as α increases from 0 to $\frac{n}{2}$.

Minimum value of fuzzy information measure will occur when many of $\mu_A(x_i)$ values are 0 or 1 subject to the constraint being satisfied. Without loss of generality we can assume that $u_1p_1 \leq u_2p_2 \leq \ldots \leq u_np_n$. Next, we discuss the different cases for minimum values of fuzzy information measure as given below:

- When $\alpha = 0$, all $\mu_A(x_i)$ have to be zero and that gives $H_{\min}(A;P;U) = 0$.
- When $\alpha = \frac{1}{2}$, one of $\mu_A(x_i)$ values can be $\frac{1}{2}$ and rest have to be 0, so that $H_{\min}(A;P;U) = \frac{u_1 p_1 \log 2}{\sum_{i=1}^{n} u_i p_i}$.

- Similarly, when $\alpha = 1$, two of $\mu_A(x_i)$'s can be $\frac{1}{2}$ and rest have to be 0, so that $H_{\min}(A;P;U) = \frac{(u_1p_1+u_2p_2)\log 2}{\sum\limits_{i=1}^{n} u_ip_i}$.

- When $\alpha = \frac{n}{2}$, $H_{\min}(A;P;U) = \log 2$.

- When α lies between $\frac{m-1}{2}$ and $\frac{m}{2}$, $m-1$ of $\mu_A(x_i)$'s can be $\frac{1}{2}$, one can be a fraction η and rest can be 0, so that
  $H_{\min}(A;P;U) = \frac{(u_1p_1+u_2p_2+\ldots+u_{m-1}p_{m-1})\log 2 - u_mp_m(\eta\log\eta+(1-\eta)\log(1-\eta))}{\sum\limits_{i=1}^{n} u_ip_i}$.

Thus as α varies from $\frac{m-1}{2}$ to $\frac{m}{2}$, $H_{\min}$ varies from $\frac{(u_1p_1+u_2p_2+\ldots+u_{m-1}p_{m-1})\log 2}{\sum\limits_{i=1}^{n} u_ip_i}$ to $\log 2$. Therefore, $H_{\max}$ increases from 0 to $\log 2$ continuously while $H_{\min}$ also increases from 0 to the same value but in a piecewise continuous manner.

7.6 Constrained Optimization of 'useful' Fuzzy Directed Divergence

The maximum value of the 'useful' fuzzy directed divergence is obtained when many of $\mu_A(x_i)$ values are 0 or 1. Next, we discuss the different cases for maximum values of 'useful' fuzzy directed divergence measure as given below:

- When $\alpha = 0$, the maximum value is $-\sum\limits_{i=1}^{n} u_ip_i \log(1-\mu_B(x_i))$.

- When $\alpha = \frac{1}{2}$, the maximum value is
  $u_np_n \log\frac{1-\mu_B(x_n)}{\mu_B(x_n)} - \sum\limits_{i=1}^{n} u_ip_i \log(1-\mu_B(x_i))$
  provided $u_1p_1 \leq u_2p_2 \leq \ldots \leq u_np_n$.

- Similarly, when $\alpha = 1$, the maximum value is given by

$$u_np_n \log\frac{1-\mu_B(x_n)}{\mu_B(x_n)} + u_{n-1}p_{n-1}\log\frac{1-\mu_B(x_{n-1})}{\mu_B(x_{n-1})} - \sum_{i=1}^{n} u_ip_i \log(1-\mu_B(x_i)).$$

- Finally, when $\alpha = \frac{n}{2}$, the maximum value is $-\sum_{i=1}^{n} u_i p_i \log \mu_B(x_i)$.

It may be noted that the maximum value of 'useful' fuzzy directed divergence measure is a piecewise continuous function which increases from $-\sum_{i=1}^{n} u_i p_i \log (1 - \mu_B(x_i))$ to $-\sum_{i=1}^{n} u_i p_i \log \mu_B(x_i)$.

For minimization of the 'useful' fuzzy directed divergence measure of a standard fuzzy set A from a standard fuzzy set B, we consider (7.4.2) subject to $\sum_{i=1}^{n} \mu_A(x_i) = \alpha \leq \frac{n}{2}$.

On using method of Lagrange's multiplier, we get

$$u_i p_i \log \left(\frac{\mu_A(x_i)(1 - \mu_B(x_i))}{(1 - \mu_A(x_i))\mu_B(x_i)} \right) = m, \tag{7.6.1}$$

$$\Rightarrow \frac{\mu_A(x_i)}{1 - \mu_A(x_i)} = e^{\frac{m}{u_i p_i}} \frac{\mu_B(x_i)}{1 - \mu_B(x_i)} = e^{\frac{m}{u_i p_i}} \beta_i, \tag{7.6.2}$$

where m is determined by

$$\psi(m) = \sum_{i=1}^{n} \frac{e^{\frac{m}{u_i p_i}} \beta_i}{1 + e^{\frac{m}{u_i p_i}} \beta_i} - \alpha = 0. \tag{7.6.3}$$

and

$$\psi'(m) = \sum_{i=1}^{n} \frac{e^{\frac{m}{u_i p_i}} \frac{\beta_i}{u_i p_i}}{\left(1 + e^{\frac{m}{u_i p_i}} \beta_i\right)^2} \geq 0. \tag{7.6.4}$$

It can be seen that

$$\psi(-\infty) = -\alpha, \qquad \psi(0) = \sum_{i=1}^{n} \mu_B(x_i) - \sum_{i=1}^{n} \mu_A(x_i), \qquad \psi(\infty) = n - \alpha. \tag{7.6.5}$$

In view of (7.6.5), the equation (7.6.3) has a unique root which is positive if $\sum_{i=1}^{n} \mu_B(x_i) < \sum_{i=1}^{n} \mu_A(x_i)$, negative if $\sum_{i=1}^{n} \mu_B(x_i) > \sum_{i=1}^{n} \mu_A(x_i)$ and zero if $\sum_{i=1}^{n} \mu_B(x_i) = \sum_{i=1}^{n} \mu_A(x_i)$.

If $m \geq 0$, then $\mu_A(x_i) \geq \mu_B(x_i)$ and each minimizing value is less than the corresponding given value. Thus in every case we have three possibilities:
(i) $\mu_A(x_i) < \mu_B(x_i)$ (ii) $\mu_A(x_i) > \mu_B(x_i)$ (iii) $\mu_A(x_i) = \mu_B(x_i)$.

Bibliography

Aczél, J. and Z. Daróczy (1975), "*On Measures of Information and Their Generalizations*", Academic Press, New York.

Arimoto, S. (1971), "Information-Theoretic Considerations on Estimation Problems", *Information and Control*, **19**,181-190.

Barro, R.J. (1991), "Economic Growth in a Cross-section of Countries", *Quarterly Journal of Economics*, **CVI**, 407-445.

Barro, R.J. and X. Sala-i-Martin (1991), "Convergence Across States and Regions", *Brookings Papers on Economic Activity*, **1**, 107-158.

Barro, R.J. and X. Sala-i-Martin (1992), "Convergence", *Journal of Political Economy*, **100**, 223-251.

Baumol, W.J. (1986), "Productivity Growth, Convergence and Welfare: What the Long Run Data Show", *American Economic Review*, **76**, 1072-1085.

Belis, M. and S. Guiasu (1968), "Quantitative-Qualitative Measure of Information in Cybernetic Systems", *IEEE Trans. Inf. Th.*, **IT-14**, 591-592.

Bezdek, J.C. (1981), "*Pattern Recognition with Fuzzy Objective Function Algorithms*", Plenum Press, New York.

Bhaker, U.S. and D.S. Hooda (1993), "Mean Value Characterization of Useful Information Measures", *Tamkang Journal of Math.*, **24**, 383-394.

Bhandari, D. and N.R. Pal (1993), "Some New Information Measures for Fuzzy Sets", *Information Science*, **67**, 204-228.

Boekee D.E. and J.C.A. Lubbe (1980), "The *R*-norm Information Measures", *Information and Control*, **45**, 136-155.

Chang, P.T. and E.S. Lee (1994a),"Fuzzy Linear Regression with Spreads Unrestricted in Sign", *Comput. Math. Appl.*, **28**, 61-70.

Chang, P.T. and E.S. Lee (1994b), "Fuzzy Least Absolute Deviations Regression Based on the Ranking of Fuzzy Numbers", *IEEE Conference on Fuzzy Systems at 1994 World Congress on Computational Intelligence,* Orlando, FL, June 26-29.

Chang, P.T. and E.S. Lee (1994c), "Fuzzy Least Absolute Deviations Regression and the Conflicting Trends in Fuzzy Parameters", *Comput. Math. Appl.*, **28**, 89-101.

Chang, P.T. and E.S. Lee (1996), "A Generalized Fuzzy Weighted Least-Squares Regression", *Fuzzy Sets and Systems*, **82**, 289-298.

Choi, Hak and Hongyi Li (2000), "Economic Development and Growth in China", *Journal of International Trade and Economic Development*, **9**, 37-54.

De Luca, A. and S. Termini (1972), "A Definition of a Non-probabilistic Entropy in the Setting of fuzzy sets theory", *Information and Control*, **20**, 301-312.

Diamond, P. (1988), "Fuzzy Least Squares" *Information Sciences*, **46**(3), 141-157.

Diamond, P. (1990), "Higher Level Fuzzy Sets Arising in Linear Regression", *Fuzzy Sets and Systems*, **36**, 265-275.

Dobson, Stephen and C. Ramlogan (2002), "Economic Growth and Convergence in Latin America", *Journal of Development Studies*, **38**, 83-104.

Dubois, D. and H. Prade (1980), "*Fuzzy Sets and Systems: Theory and Application*", Academic Press, New York.

Dunn, J.C. (1973), "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *J. Cybernetics*, **3**(3), 32-57.

D'Urso, P. and T. Gastaldi (2000), "A Least Squares Approach to Fuzzy Linear Regression Analysis", *Computational Statistics and Data Analysis*, **34**, 427-440.

Ebanks, B.R. (1983), "On Measures of Fuzziness and Their Representations" *J. Math. Analysis and Applications*, **94**, 24-37.

Ferreira, Alfonso (2000), "Convergence in Brazil: Recent Trends and Long-Run Prospects", *Applied Economics* **32**, 79-90.

Giles, D.E.A (2001), "Output Convergence and International Trade: Time Series and Fuzzy Clustering Evidence for New Zealand and her Trading Partners, 1950 - 1992", working paper EWP0102, Deptt. of Economics, University of Victoria.

Goldman Sachs Report (2003), "Dreaming with BRICs: The Path to 2050", *Global Economics*, paper No.99.

Havrda, J. and F. Charvát (1967), "Quantification Method of Classification Processes: Concept of Structural α- entropy", *Kybernetika*, **3**, 30-35.

Holmes, J.M. (2004), "New Evidence on Long Run Output Convergence Among Latin American Countries", *Journal of Applied Economics*, **VIII**(2), 299-319.

Hooda D.S. and A. Ram (1998), "Characterization of a Generalized Measure of R-norm Entropy", *Caribb. J. Math. Comput. Sci.*, **8**,(1 & 2), 18-31.

Hooda D.S. and A. Ram (2002), "On Useful Relative Information and J-divergence", *Tamkang Journal of Mathematics*, **33**, 146-160.

Hooda, D.S. (2004), "On Generalized Measures of Fuzzy Entropy", *Mathematica Slovaca*, **54**, 315-325.

Hooda D.S. and D.K. Sharma (2007), "Generalized R-norm Information Measures", (Communicated).

Kapur, J.N.(1987), " Monotonicity and Concavity of Some Measures of Entropy, Directed Divergence and Related Functions", *Tamkang Journal Math.*, **18**, 21-40.

Kapur, J.N. (1997), "*Measures of Fuzzy Information*", Mathematical Science Trust Society, New Delhi.

Kaufman, A. (1980), "*Fuzzy subsets: Fundamental Theoretical Elements*" **Vol.3**, Academic Press, New York.

Kerridge, D.F. (1961), "Inaccuracy and Inference", *J. Royal Statist. Society*, **Ser. B, 23**, 184-194.

Khan, Mohsin S. and Manmohan S. Kumar (1993), "Public and Private Investment and the Convergence of per capita Incomes in Developing Countries", Working Paper **93/51**, Washington, DC, IMF.

Kim, K.J., Moskowitz, H., and M. Koksalan (1996), "Fuzzy versus Statistical Linear Regression", *European Journal of Operations Research*, **92**, 417-434.

Klir, G. and B. Yuan (1995), "*Fuzzy sets and Fuzzy Logic: Theory and Applications*", Prentice Hall, Upper Saddle River, New Jersey.

Kosko, B. (1986), "Fuzzy Entropy and Conditioning", *Information Sciences*, **40(2)**, 165-174.

Kosko, B. (1990), "Concepts of Fuzzy Information Measure on Continuous Domains", *International Journal of General Systems*, **17**, 211-240.

Kosko, B. (1991), "*Neural Networks and Fuzzy Systems: a Dynamical System Approach to Machine Intelligence*", Prentice Hall, New Jersey.

Kullback, S. and R.A. Leibler (1951), "On information and Sufficiency", *Ann. Math. Stat.*, **22**, 79-86.

Kullback, S. (1959), *"Information Theory and Statistics"*, Wiley and Sons, New Delhi.

Loo, S.G. (1977), "Measures of Fuzziness", *Cybernetica*, **20**, 201-210.

McCoskey, Suzanne (2002), "Convergence in Sub-Saharan Africa: A Non-stationary Panel Data Approach", *Applied Economics*, **34**, 819-829.

Nagaraj, Rayaprolu, Aristomene Varoudakis, and Marie-Ange Veganzones (2000), "Long-run Growth Trends and Convergence Across Indian States", *Journal of International Development*, **12**, 45-70.

Pal, N.R. and S.K. Pal (1989), "Object Background Segmentation Using New Definition of Entropy", *Proc. Inst. Elec. Eng.*, **136**, 284295.

Pal, N.R. and S.K. Pal (1992), "Higher Order Fuzzy Entropy and Hybrid Entropy of a Set", *Information Science*, **61** (3), 211231.

Pal, N.R. and J.C. Bezdek (1994), "Measuring Fuzzy Uncertainty", *IEEE Trans Fuzzy Systems*, **2**, 107-118.

Peters, G. (1994), "Fuzzy Linear Regression With Fuzzy Intervals", *Fuzzy Sets and Systems*, **63**, 45-55.

Parkash, O. (1998), "A New Parametric Measure of Fuzzy Entropy", *Proc. 7th International Conference IPMU'98*, **2**, 1732-1737.

Parkash, O. and P.K. Sharma (2001), "Noiseless Coding Theorems Corresponding to Fuzzy Entropies". 10^{mboxth} *IEEE International Conference on Fuzzy Systems*, **1**, 176-179.

Quah, D. (1993), "Empirical Cross-section Dynamics in Economic Growth", *European Economic Review*, **37**, 426-434.

Quah, D. (1996), "Regional Convergence Clusters Across Europe", *European economic review*, **40(3-5)**, 951-958.

Rényi, A. (1961), "On Measures of Entropy and Information", *Proc. 4^{th} Berkeley Symp. Math. Stat. Probab.*, University of California Press, **1**, 547-561.

Ross, J.T. (2005), "*Fuzzy Logic with Engineering Applications*", 2^{th} Edition, John Wiley & Sons Ltd, England.

Ruspini, E.H. (1969), "A New Approach to Clustering", *Information Control*, **15**(1), 22-32.

Ruspini, E.H. (1970), "Numerical Methods for Fuzzy Clustering", *Information Sciences*, **2**, 319-350.

Sala-i-Martin, X. (1996), "The Classical Approach to Convergence Analysis", *Economic Journal*,**106(437)**, 1019-1036.

Shannon, C.E. (1948), "A Mathematical Theory of Communication", *Bell Syst. Tech. Journal*, **27**, 379-423, 623-656.

Sharma, B.D. and D.P. Mittal (1975), "New Non-additive Measures of Entropy for Discrete Probability Distribution", *J. Math. Sci. (Calcutta)*, **10**, 28-40.

Sharma, B.D. and D.P. Mittal (1977), "New Non-additive Measures of Relative Information", *J. Comb. Inform. and Syst. Sci.*, **2**, 122-133.

Tanaka, H., Uejima, S. and K. Asai (1980), "Fuzzy Linear Regression Model", *IEEE Trans. on Systems, Man, and Cybernetics*, **10**, 2933-2938.

Tanaka, H., Uejima, S. and K. Asai (1982), "Linear Regression Analysis with Fuzzy Model", *IEEE Trans. on Systems, Man, and Cybernetics*, **12**, 903-907.

Tanaka, H. and J. Watada (1988), "Possibilistic Linear Systems and Their Application to the Linear Regression Model", *Fuzzy Sets and Systems*, **27**, 145-160.

Taneja, I.J. (1989), "On Generalized Information Measures and Their Applications", Chapter in:*Advances in Elect. and Elect. Physics*, Ed. P.W. Hawkes, **76**, 327-413.

Taneja, I.J. (1995), "New Developments in Generalized Information Measures", Chapter in: *Advances in Imaging and Electron Physics*, Ed. P.W. Hawkes, **91**, 37-135.

Taneja, I.J. (2005), "On Symmetric and Nonsymmetric Divergence Measures and Their Generalizations" , Chapter in: *Advances in Imaging and Electron Physics*, Ed. P.W. Hawkes, **138**, 177-250.

Theil, H. (1967), "*Economics and Information Theory*", North- Holland Publishing Co., Amesterdom.

Yager, R.R. (1979), "On the Measure of Fuzziness and Negation. Part I: Membership in the Unit Interval", *Inter. Jour. of General Systems*, **5**, 221-229.

Yazici, A. Dept. of Computer Engineering, Middle East Technical University, 06531, Ankara/Turkey [ppt slides]

Zadeh, L.A. (1965), "Fuzzy Sets", *Information and Control*, **8**, 338-353.

Zadeh, L.A. (1968), "Probability Measures of Fuzzy Events", *J. Math. Analysis and Applications*, **23**, 421-427.

Zadeh, L.A. (1973), "Outline of A New Approach to the Analysis of Complex Systems and Decision Processes", *IEEE Trans. on Systems, Man, and Cybernetics*, **3**, 28-44.

Zadeh, L.A. (1975), "The Concept of a Linguistic Variable and its Application to Approximate Reasoning-1", *Information Sciences*, **8**, 199-249.

Zadeh, L.A. (1984), "Making Computers Think Like People", *IEEE. Spectrum*, **8**, 26-32.

Zadeh L.A. (1995), "Discussion: Probability Theory and Fuzzy Logic are Complementary rather than Competitive," *Technometrics*, **37**, 271-276.

Zadeh L.A. (2002), "*Fuzzy Logic and Probability Applications: Bridging the Gap*", SIAM Publishers, Philadelphia PA.

Zimmermann, H.J. (1985), "*Fuzzy Sets Theory and its Applications*", Kluwer Publications, Boston, MA.

http://en.wikipedia.org/wiki/Fuzzy_set