# INTERPRETATION AND ESTIMATION OF MEMBERSHIP FUNCTIONS

by

Chow Kan Shing

A

Thesis

Submitted to

(Division of Statistics)

the Graduate School

of

The Chinese University of Hong Kong

In Partial Fulfilment

of the Requirements for the Degree of

Master of Philosophy

May, 1993

THE CHINESE UNIVERSITY OF HONG KONG

GRADUATE SCHOOL

The undersigned certify that we have read a thesis, entitled "Interpretation and Estimation of Membership Functions" submitted to the Graduate School by Chow Kan Shing ( 周勤城 ) in partial fulfilment of the requirement for the degree of Master of Philosophy in Statistics. We recommend that it is accepted.

Kim-Hung Li

Dr. K. H. Li,
Supervisor

N.N. Chan

Dr. N.N. Chan

Leung Pui Lam

Dr. P.L. Leung

Dr. S.K. Tse
External Examiner

DECLARATION

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

# ACKNOWLEDGEMENT

The author is deeply indebted to many people. First, I have to thank my supervisor, Dr. Li Kim Hung, who has given many helpful comments and suggestions during these two years. Second, I have to thank my classmates, who have given many supports during these two years.

ABSTRACT

Membership function is important in fuzzy set theory. In this thesis we will discuss the interpretation and the estimation methods of the membership functions. Focus is placed on the estimation part. We will view the usual estimation methods of the membership functions in a statistical way. Moreover, suggestions on the estimation of the membership function are given.

Keywords: fuzzy set theory, membership function, direct rating, polling, set-valued statistics, reverse rating.

CONTENTS

Page

# Chapter 1. Introduction.

In classical set theory, an element can either belong to a set or not belong to a set. Thus a classical set B can be written in the form:

$B = \{(x, \alpha_B(x)): x \in X\}$

where X is the population of the elements and $\alpha_B$ is a function, called the characteristic function, which maps X to $\{0,1\}$. When $\alpha_B(x) = 0$, x is called not belonging to B; when $\alpha_B(x) = 1$, x is called belonging to B. But there are many examples in real life that classical set theory cannot explain. For example, instead of considering a "set of people with height greater than 170 cm" which is a classical set, we may want to consider the "set of tall people". Then in this case classical set theory fails. It is this phenomenon that led to the inception of "fuzzy set theory" in 1965 (Zadeh, 1965). Fuzzy set theory is an extension of classical set theory. A set is fuzzy when an element can belong partly to it, rather than having to belong to it completely or not at all. Although we know what is a fuzzy set, there is not a consensus on the definition of fuzziness yet. Zadeh and many other researchers conceive fuzziness as a nonprobability kind of vagueness. Klir (1985) associates fuzziness with the difficulties of making sharp or precise distinctions in real life. Hisdal (1986) believes that Klir's interpretation is insufficient and suggests a broader definition of fuzziness. She lists eleven sources of fuzziness and Klir's interpretation can be viewed as a special case of her interpretation. For example, suppose we discuss the set of "wealthy man". This set is a fuzzy one because everyone has his own interpretation of "wealthiness". But, suppose everybody agree that a man is "wealthy" if and only if he has more than 100 million dollars. In Klir's interpretation of fuzziness, the set of "wealthy man" will no longer be a fuzzy one. In Hisdal's interpretation, this set is still a fuzzy one because of other sources of fuzziness. For example, the subject will make error when he estimates the amount of money that the man has and this error makes the set of "wealthy man" a fuzzy one.

In this thesis, we will consider the fuzziness that due to different interpretation of concept and also that due to incomplete information about the objects.

Zadeh (1965) proposed two important items in fuzzy set theory, namely the membership function and the many valued operators working in this function. In these 30 years, most papers on fuzzy set theory relate to the many valued operators. The interpretation and estimation of the membership function have not received enough attention. Most researchers assume that such function exists and do not try to find out what it is. In this thesis, we will discuss the membership function only. In Chapter 2, we will give a brief review on the fuzzy set theory. Then, in Chapter 3, we will summarize the views on the membership function by previous researchers and discuss their characteristics. In Chapter 4, we will summarize the estimation methods of the membership function used by previous researchers and discuss each method's characteristics. Then we will go over a survey carried out by us and give suggestions to those who want to estimate the membership function through a survey.

# Chapter 2. A Brief Review on the Fuzzy Set Theory.

In classical set theory, an element either belongs to a set or does not belong to a set. However, life is not so simple. We can find many phenomena in real life that the classical set theory cannot explain, e.g. the set of old men, the set of long lines, the set of beautiful girls, etc. In many cases, it is difficult, or even impossible, to determine whether an object belongs to a set or not. Nevertheless, it is possible to say to what degree an element belongs to a set. It was the above phenomenon that led to the inception of the fuzzy set theory in 1965. In the following sections, we will introduce the fundamental concepts of fuzzy set theory for the preparation of the following chapters.

## 2.1. The Concept of Fuzzy Set Theory.

Definition 2.1. Let X denote an universe of discourse. A fuzzy set A in X is a set of ordered pairs

$$A = \{ ( x, \mu_A(x) ) : x \in X \} \tag{2.1}$$

where $\mu_A$ is called "the membership function of A" and $\mu_A(x)$ is called "the membership value of x in A".

From the above definition, we know that each fuzzy set is defined by an universe of discourse and a membership function. Thus two fuzzy sets are different if either their universes of discourse or their membership functions are different.

The range of membership function can take a general structure such as a lattice but throughout this thesis, we will just assume that the range is the closed interval [0,1] as most researchers do. If $\mu_A(x) = 0$, x does not belong to A completely. If $\mu_A(x) = 1$, x belongs to A completely. Any value between 0 and 1 indicates the degree of belonging to A. It is clear that the difference between classical set theory and fuzzy set theory is a gradual instead of an abrupt transition from membership to nonmembership and we can say fuzzy set theory is an extension of classical set theory. We will now consider an example of fuzzy set.

Example 2.1 : Suppose X is the set of population in Hong Kong and A is the set of tall man in Hong Kong. Then we may define $\mu_A(x)$ as

$$\mu_A(x) = \begin{cases} 0 & \text{if the height of } x < 150 \text{ cm} \\ \dfrac{\text{height of } x - 150}{30} & \text{if the height of } x \geq 150 \text{ cm but} \leq 180 \text{ cm} \\ 1 & \text{if the height of } x > 180 \text{ cm} \end{cases}$$

### 2.2. Fundamental Operations on Fuzzy Sets

In Section 3.1, we would like to compare the different interpretations of membership value. One of the criteria is that: the interpretation should be able to apply in the case of connective, e.g. the interpretation still holds for the membership value of fuzzy set $A \cap B$. We would like to introduce certain basic operations such as union, intersection and negation for the preparation of Section 3.1.

Definition 2.2. Let A and B be fuzzy sets of X. The membership function of the union of A and B, $(A \cup B)$, is defined as

$$\mu_{A \cup B}(x) = \text{Max}[\ \mu_A(x),\ \mu_B(x)\ ],\ \forall\ x \in X.$$

Definition 2.3. Let A and B be fuzzy sets of X. The membership function of the intersection of A and B, $(A \cap B)$, is defined as

$$\mu_{A \cap B}(x) = \text{Min}\ [\ \mu_A(x),\ \mu_B(x)\ ],\ \forall\ x \in X.$$

Definition 2.4. Let A be a fuzzy set of X. The membership function of negation of A, $(\bar{A})$, is defined as

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x),\ \forall\ x \in X.$$

The above definitions for union, intersection and negation in fuzzy set theory are not unique. Definitions 2.2 - 2.4 were proposed by Zadeh (1965) and are the ones most frequently used by the researchers. This characteristic of fuzzy set theory posed a lot of criticism (see French (1984) and Hisdal (1988)). Some researchers argue that minimum in the case of intersection may not explain the matter well. Actually many researchers use product instead of minimum in the case of intersection (definition 2.3). In Thole et al. (1979), the suitability of minimum and product operators for the intersection of fuzzy sets is discussed. The authors conclude that neither the product nor the minimum fits the data sufficiently well, but the latter is preferred.

With the above definitions, we note that when the range of the membership function becomes {0,1} instead of [0,1], the membership function in fuzzy set becomes the characteristic function in classical set theory and definitions 2.2 - 2.4 become the union, intersection and negation in classical set theory.

## 2.3. Two Approaches to Investigate Fuzzy Set Theory.

The analysis of fuzzy set theory can be classified into two approaches: the "axiomatic" approach and the "semantic" approach. In the "axiomatic" approach, the researchers set up a set of "axioms" or a formal structure and then theorems are derived from these axioms. Most of the papers take this approach. In the "semantic" approach, researchers first give an interpretation of membership values. Then "axioms" and "laws" are derived from the interpretation and further analysis are based on these derived "axioms" and "laws". In probability theory, there is also such a classification. In the "axiomatic" approach, axioms are used to define a mathematical structure involving concepts such as $\sigma$-algebra. In the "semantic" approach, the definition and properties of probability are derived from practical concepts such as relative frequency. These two approaches are different in nature. Sharing the same view with Hisdal (1988), we believe that the second approach is more realistic. One of the purposes of fuzzy set theory is to deal with different interpretations of concepts in human's mind. So it is better to have a clear interpretation of membership values. In this thesis, we will discuss those papers that use the second approach: those try to interpret the membership function rather than derive it from certain set of "axioms".

# Chapter 3. The Interpretation of Membership Function.

In this chapter, we will explore the meaning of the membership value in subsections 3.1.1 - 3.1.6. In each subsection, we would review and discuss a view of membership value. In Section 3.2, we would give our views on some items relating to membership value. Before the discussion, we would like to emphasize that for every fuzzy set, there are many membership functions. In addition to the membership function for each subject, there is a "true" membership function. In this thesis we define the "true" membership function as the average of all the subjects' membership functions. In the following subsections, the term "membership function" is used to stand for "subject's membership function".

## 3.1. Review and Comparison of the Interpretation of Membership Value.

The interpretation of membership values has not received enough attention. When compared to the total number of papers on fuzzy set theory, there are very few published papers which focus on the interpretation of membership values. Among these limited number of papers , we find the following six interpretations of membership value and we will discuss and compare them in the following subsections. In order to make the comparison unified, we set up criteria to decide whether an interpretation is a good one or not:

1) Generality:

The area of application of the interpretation should be large, i.e. the interpretation should be able to apply to all fuzzy sets. Moreover the interpretation should be able to apply in the case of connective, e.g. the interpretation still holds for the membership value of fuzzy set $A \cap B$.

2) Estimability:

The interpretation should suggest a corresponding estimation procedure. We will show in Section 4.1. the four most frequently used methods in the estimation of membership function. It is of interest to know whether the interpretation will favor one of the methods and whether the interpretation provides other estimation procedures.

3) Testiability:

The interpretation's assumptions, if exist, should be natural and testable.

4) Intuition:

As fuzzy set theory is used to analyze fuzzy information in real life, the interpretation should be intuitive. Moreover the operations to find out the membership value should be simple enough so that it is reasonable for a human to carry out.

Besides the above criteria, we are also interested in the scale of the membership value as we want to know what operations can be applied to the membership value.

### 3.1.1. Interpretation in terms of Betting.

Giles (1982) considered the following interpretation of subjective probability:

> "...to assign a probability p to a proposition A means that one regards the following as a "fair bet":
>
> If you pay me \$p then I will agree to pay you \$1
>
> if A is the case (and nothing otherwise). "

Here "fair" means fair to both parties. The expected gain of both sides of the bet are zero. Then he followed this approach and gave a corresponding interpretation of membership value. For any fuzzy concept, he linked it with a test-procedure for the object which has two possible outcomes: passes the test or not. For example, for the membership value of certain glass to the set "unbreakable glass", one possible test-procedure is

> "Drop the glass from a height of 5 ft onto a wooden floor; if it breaks the outcome is no and the outcome is yes if it does not break."

After the test-procedure is defined, the meaning of the membership value $\mu_A(x)$ is that the subject agrees that the following is a "fair bet":

> "If you pay me \$ $\mu_A(x)$ then I will agree to pay you \$ 1 if the test-procedure gives outcome "yes"".

Thus the assignment of membership values becomes simply a special case of the assignment of a subjective probability.

After the introduction of this interpretation, we would like to discuss how this interpretation related to the above four criteria:

1) This interpretation applies to most of the fuzzy set. As for any fuzzy set, we can find a suitable test and find the corresponding "fair bet". But there are some cases that this interpretation fails. Suppose we ask a subject to assign his degree of agreement with an attitudinal statement. It is difficult to imagine that the subject bet with oneself on his own degree of agreement with the statement in a "fair bet". In the case of connective, this interpretation leads to the following problem:

   Under this interpretation, $A \cap A \neq A$. For example, when the set is the set of "breakable glass" and the test procedure is the one shown above. When we consider $A \cap A$, we need to do the test twice and we assign a positive outcome only if the glass breaks on neither test. Obviously it is equivalent to a harder test than that of A. In other words, $A \cap A \neq A$ under this interpretation. This phenomenon raises the following question: "Which identity in the classical set theory will remain to be valid in fuzzy set theory ?" As stated above, one of the purpose of fuzzy set theory is to deal with different interpretations of concepts in human's mind. So we may ask ourselves whether $A \cap A = A$ should be true. For example, if John said, "Mary is beautiful." twice, should we think that it carries different meaning from that if John said "Mary is beautiful" only once.

2) The assignment of membership values is simply a special case of the assignment of a subjective probability. By this subjective nature of

the interpretation, direct rating (see section 4.1.1. and 4.2.1.1.) is the most suitable estimation method for the membership function.

3) This interpretation has assumed that for any fuzzy set, there is a corresponding test. This assumption is very weak and so it will hold in most of the cases.
4) Using betting to describe membership values is intuitive as membership values resemble subjective probability. As mentioned above, Giles uses betting in his interpretation to avoid clarifying how the membership value is actually computed.
5) As the membership value is interpreted as a kind of subjective probability, it is on ratio scale.

According to this interpretation, each subject has his own test. It introduces the fuzziness. If the test is specified and all subjects know the test, then it will be a problem of subjective probability. Thus we can view fuzzy set theory as an extension of subjective probability.

**3.1.2. Interpretation in terms of Payoff Function.**

Giles (1988) stated that "property" and "set" are equivalent and so "fuzzy set" can be explained by "fuzzy property". Then membership value can be viewed as a measure of the "degree of possession" of the corresponding property. Then he related this concept with the "degree of truth" of a fuzzy sentence. For example, the membership value of John to the set "tall man" is equal to the "degree of truth" of the fuzzy sentence "John is a tall man". As the principal function of a fuzzy sentence is to transmit a "degree of belief" and a "degree of belief" is simply a subjective form of a "degree of truth", discussion of "degree of truth" is equivalent to discussion of "degree of belief". Moreover as the practical function of a sentence is to be asserted, Giles consider an assertion rather than a sentence. For a fuzzy sentence, subjects may have various degrees of willingness to assert it; indeed, the degree of willingness may be related to the "degree of belief" of the sentence. Assert a fuzzy sentence or not is a general problem in decision theory. Thus Giles introduced payoff function in decision theory to interpret the membership value.

**Definition:** A payoff function $u(B,\omega)$ gives the utility of the outcome determined by any act B and world state $\omega$.

If the decision problem is a trivial one that has only one act, the payoff function will be a function of the state only. In this case, we denote the payoff function as $a(\omega)$ where $\omega$ is the world state. The relation between assertion and payoff function is :

> "If we define the meaning of an assertion as the information necessary and sufficient, in conjunction with an agent's beliefs about the world state, to allow the agent to decide whether or not to assert the sentence, then the meaning of an assertion is given by its payoff function."

**Definition:** For each assertion a, let $[a_{min}, a_{max}]$ be the smallest closed interval that contains every payoff value attained by a. A simple assertion is an assertion whose payoff $a(\omega)$ takes only two values as $\omega$ ranges over the set $\Omega$ of all (pure) world states.

Of course, the two payoff values for a simple assertion are $a_{max}$ and $a_{min}$. We say that the assertion holds or is true in the state $\omega$ if $a(\omega) = a_{max}$ and does not hold or is false if $a(\omega) = a_{min}$. Although the payoff of a simple assertion for a pure state is either $a_{min}$ or $a_{max}$, it may take intermediate values in the case of a mixed state.

If $a_{min} = 0$ and $a_{max} = 1$, then a simple assertion is called a standardized assertion. The relation between the truth value of a sentence and the payoff for a standardized assertion of this sentence can be summarized as follows:

> "In any world state (pure or mixed) the truth value of a sentence (fuzzy or not) coincides with the payoff for a standardized assertion of this sentence."

He related the membership value and the above terms as follow:

> "For any object x in X the membership value of x is the truth value of the fuzzy sentence P(x), "x has property P"; or, equivalently, it is the payoff for a standardized assertion of this sentence."

After the introduction of this interpretation, we would like to discuss how this interpretation is related to the above four criteria:

1) This interpretation applies to all of the fuzzy set. For any fuzzy set, we can find the corresponding assertion. Then we transform the assertion to a standardized assertion. The payoff function of this standardized assertion is the membership value. In the case of connective, this interpretation faces a problem:

Giles interprets membership values as truth value of fuzzy sentences but the max-min laws of disjunction and conjunction for truth functions cannot hold generally. For example, the truth values of "it will rain today" and "it will not rain today" may be 0.3 and 0.7 respectively, but the truth value of the disjunction of these sentences is surely 1. Giles does not view this inconsistency with the max-min laws as a problem. It is because the interpretation is in terms of a payoff function, whose values are real numbers. But characteristic function only has values zero and one. Thus there is no question of a connective for assertions "reducing in the classical case to one of the connective of classical set", and therefore there is no need to seek generalizations of the connective in classical set theory. Giles (1988) discussed how the assertions can be combined and concluded that linear connective can be applied in this kind of membership values.

2) He has suggested a method to estimate the payoff function of the subject for the assertion "John is tall". He first asked the subject to decide the least height he would describe as tall. Then he asked the subject to decide other negative / positive payoff values by offering bribe / penalty.

3) This interpretation has assumed that for any fuzzy set, there is a corresponding assertion and payoff function. This assumption is very weak and so it will hold in most of the cases.
4) This interpretation differs from the first interpretation that it uses a payoff function rather than a bet to deal with the numeric representation of membership. As mention above, Giles uses payoff function in his interpretation and so what he need to do is the estimation of the payoff function. He has given suggestions on the determination of the payoff function of the assertion "John is tall".
5) The interpretation is in terms of the payoff function of an standardized assertion. In this case, a natural zero is defined and so it is on the ratio scale.

In this interpretation, fuzziness is due to the difference of payoff function between subjects.

**3.1.3. Interpretation in terms of Amount of Relevant Attribute.**

Orlovski (1990) viewed a fuzzy set as a collection of objects showing a common property. This property itself is defined as decomposable and is represented by a set of elementary properties that objects may have or not have. In order to quantify the relative importance of different collections of elementary properties with respect to the decomposable property in question, he has used a set function which he refers to as "pseudomeasure". Then the membership value of an object can be interpreted as relative powers (or relative importance) of collections of elementary properties that the object has.

Mabuchi (1992) supposed that we can gather a set of all the reference items u(x) or grounds, referred as a ground set $G_A(x)$, relevant to the judgment of whether the given element x belongs to the fuzzy set A or not. Moreover there is a set of affirmative evidence grounds $A_e$ supporting $x \in A$. He supposed that the ground sets are all finite. For example, when the universe of discourse is all human beings, A is the set of "old people", then the ground set may be a set of attributes that characterize "old". With the ground set, the membership value is determined by relative weight, or simply the relative size, of $A_e$ with respect to that of $G_A(x)$. In order to facilitate the discussion of later sections, especially the section of set operators, Mabuchi interprets the membership function as a conditional probability $P(A|x)$, a conditional probability of A given x.

After the introduction of this interpretation, we would like to discuss how this interpretation related to the above four criteria:

1) This interpretation is general enough that it can be applied in all fuzzy set. As Mabuchi (1992) interprets membership value of x to the set A as a conditional probability $P(A|x)$, the max-min laws for disjunction and conjunction generally fail. Mabuchi has derived the corresponding connectives under his interpretation which is very similar to the connectives in probability theory.

2) This interpretation provides a corresponding estimation method. Mabuchi has given suggestion on how to determine the ground set. As the elements in the ground set may not carry the same weight in the determination of membership value, Mabuchi proposes to use a weight function to solve this problem. But he has not given any suggestion or guideline on how to choose this weight function and so the estimation procedure is not complete.
3) This interpretation assumes that for any fuzzy set, we can find a corresponding ground set.
4) This interpretation is intuitive and the computation of the membership value is easy with given ground set and weight function.
5) As it is in terms of conditional probability, the membership value is on the ratio scale.

In Mabuchi's interpretation, the ground set and the weight of elements in the ground set are both unknown to the subjects. Each subject has his estimated ones. It is the source of fuzziness.

### 3.1.4. Interpretation in terms of the TEE Model.

Hisdal (1986(a), 1986(b), 1988) introduced a "TEE-model" which tries to establish clear meaning of the membership value concept. She has listed fourteen sources of fuzziness in the above papers. She concerned the case when only the first source of fuzziness exists and all the other ones do not exist. The first source of fuzziness is due to the subject's recognition that under non-exact conditions of observation, a person may make errors in the estimation of the attribute values of objects. In this model, three different types of experiments are mentioned, namely:

1) LB (or labelling) experiments

2) YN (or yes-no) experiments

3) MU (or grade of membership) experiments.

The LB experiment corresponds to the assignment of certain label to an object, e.g., John is "tall"; and the YN experiment corresponds to the answering of a question by "Yes" or "No", e.g. "Is John tall?". In a MU experiment, the subjects are required to give a membership value $\mu_{\lambda}(x) \in [0,1]$ to an object concerning label $\lambda$. Then she proposed three assumptions on the TEE models:

**First Assumption of the TEE model.**

A subject who performs a LB, YN or MU experiment estimate the object's attribute value $u^{ex}$, and use the estimated attribute value u to estimate the membership value. u is equal to $u^{ex}$ in an exact experiment.

**Second Assumption of the TEE model.**

When a subject performs a YN experiment under exact or nonexact conditions of observation, for each label $\lambda \in \Lambda$ he sets a lower and upper threshold, $u_{\lambda l}$ and $u_{\lambda u}$ respectively in the universe of attribute values. He gives a Yes answer to the object concerning the label $\lambda$ when the object's u value falls in between the two thresholds and gives a No answer when u falls outside the two thresholds. For exact conditions $u = u^{ex}$.

**Third Assumption of the TEE model.**

When a subject performs a MU experiment under exact conditions of observation, he puts himself into the situation of an observation under nonexact conditions. Thus he knows that he will make measurement error of the object's attribute. Then he takes account of this knowledge by constructing an estimated error curve of the attribute. Then his grade of membership curve is his estimate of the modification of his nonfuzzy LB or YN threshold curve by the error curve.

For example, suppose that for the fuzzy set "tall man", a subject has his nonfuzzy lower threshold $u_{\lambda l}$, say 170 cm, i.e. he will consider a man with height greater than 170 cm as a "tall" man. Suppose there is a man with height 175 cm. Then with the corresponding estimated error curve of height, the membership value of this man to the set "tall" man is the probability that the estimated height will be greater than the lower threshold 170 cm.

Therefore, this model interprets a membership value $\mu_{\lambda}(u^{ex})$ assigned by a subject to an object of attribute value $u^{ex}$ as his estimate of the probability that the label $\lambda$ would be assigned to that object in an LB or YN experiment; e.g. by himself under nonexact conditions of observation; or by another subject.

After the introduction of this interpretation, we would like to discuss how this interpretation related to the above four criteria:

1) This interpretation can apply to most of the fuzzy set as it is in terms of some conditional probability. As the interpretation is in terms of probability, the max-min laws for connective do not satisfy in this case. Hisdal has derived the corresponding connective for her model. For example, the membership value of the conjunction of two fuzzy set A and B is given by their product instead of their minimum $\mu_{A\cap B}(u^{ex}) = \mu_{A}(u^{ex})\mu_{B}(u^{ex})$.

2) If we assume that the only source of fuzziness is measurement error, then we can find a corresponding estimation procedure. We may ask the subjects questions to estimate the threshold curve and the error curve for any fuzzy set A.

3) This interpretation needs the three assumptions stated above. These three assumptions are reasonable.

4) This interpretation is intuitive as it is in terms of some conditional probability.

5) As it is in terms of conditional probability, the membership value is on the ratio scale.

### 3.1.5. Interpretation in terms of a Measurement Model.

Norwich and Turksen (1984) and Turksen (1991) proposed a measurement model for fuzziness when the universe of discourse has an associated physical continuum. They used the result in Krantz et al. (1971) to give an interpretation of membership value. They defined the following terms which are necessary for the Representation Theorem and the Uniqueness Theorem:

**Definition:** Let X be a set and $\geq_A$ be a binary relation on X, i.e. $\geq_A$ is a subset of X x X. Then $\geq_A$ is called a weak order in X if

1) for any $x_i, x_j \in X$, either $x_i \geq_A x_j$ or $x_j \geq_A x_i$; and

2) for any $x_i, x_j, x_k \in X$, $x_i \geq_A x_j$ and $x_j \geq_A x_k$ will imply $x_i \geq_A x_k$.

**Definition:** Given a universe of discourse X, define a weak order $\geq_A$ in X, so that

$$x_1 \geq_A x_2 \quad \text{for } x_1, x_2 \in X$$

if an observer judges that " $x_1$ is at least as A as $x_2$ is" or "it is at least as true that $x_1$ is A as it is that $x_2$ is A" or " $x_1$ is at least as large as $x_2$ with respect to being A" or " $x_1$ is A-er than $x_2$". The system $\{X, \geq_A\}$ defined above will be called a "multivalued membership structure".

**Definition:** A multivalued membership structure $\{X, \geq_A\}$ is called "bounded" if there exist elements $x_M$ and $x_m$ in X such that $x_M \geq_A x$ and $x \geq_A x_m$ for all $x \in X$.

The above definitions are suitable for comparison of x's in X. In order to strength the scale of the membership, Krantz et al. (1971) consider that the subjects are able, in general, to compare any pair of intervals specified by a total of four points in X in order to establish a weak-order of intervals in X. For example, suppose a subject states that $x_2$ is at least as A as $x_1$ (i.e. $x_2$ has at least as much membership in the set A as $x_1$ does) and that $x_4$ is at least as A as $x_3$ (i.e. $x_4$ has at least as much membership in the set A as $x_3$ does). Furthermore, suppose he states that the increase in being A from $x_1$ to $x_2$ is at least as large as that from $x_3$ to $x_4$, i.e. $x_2$ is A-er than $x_1$ by at least as much as $x_4$ is A-er than $x_3$ or, equivalently, the gain in membership in the set A from $x_1$ to $x_2$ is at least as large as that from $x_3$ to $x_4$ or, equivalently, $x_2$'s membership exceeds $x_1$'s by at least as much as $x_4$'s exceeds $x_3$'s. The above comparison considers about intervals in X, and may be denoted by:

$$x_2x_1 \geq'_A x_4x_3 \quad \forall\, x_1, x_2, x_3, x_4 \in X.$$

**Definition:** For a bounded, multivalued membership structure $\{X, \geq_A\}$, if there exists an ordering $\geq'_A$ on X x X such that

$$x_2x_1 \geq'_A x_4x_3 \quad \forall\, x_1, x_2, x_3, x_4 \in X$$

means $x_2 \geq_A x_1$ by at least as much as $x_4 \geq_A x_3$. Then this bounded, multivalued membership structure $\{X, \geq_A\}$ will be called "difference-comparable" and denoted by $\{X \times X, \geq'_A\}$.

From Krantz et al. (1971), we know that if a difference-comparable, bounded, multivalued membership structure satisfies the following five axioms, then it forms an "algebraic-difference structure":

Axiom 1: $\{X \times X, \geq'_A\}$ is a weak-order.

Axiom 2: If $x_2x_1 \geq'_A x_4x_3$, then $x_3x_4 \geq'_A x_1x_2$.

Axiom 3: If $x_2x_1 \geq'_A x_5x_4$ and $x_3x_2 \geq'_A x_6x_5$, then $x_3x_1 \geq'_A x_6x_4$.

Axiom 4: If $x_2x_1 \geq'_A x_4x_3 \geq'_A x_1x_5$, then there exist $x_6, x_7 \in X$

such that $x_6x_1 \succeq'_A x_4x_3 \succeq'_A x_2x_7$.

Axiom 5: Suppose $x_ix_j \sim'_A x_kx_l$ whenever $x_ix_j \succeq'_A x_kx_l$ and $x_kx_l \succeq'_A x_ix_j$ and $x_ix_j \succ'_A x_kx_l$ whenever $x_ix_j \succeq'_A x_kx_l$ and not $x_ix_j \sim'_A x_kx_l$. If $x_1, x_2, \ldots, x_i, \ldots$ is a strictly bounded standard sequence (i.e. $x_{i+1}x_i \sim'_A x_2x_1$ for every $x_i, x_{i+1}$ in the sequence; not $x_2x_1 \sim'_A x_1x_1$; and there exist $x'_i$ and $x''_i \in X$ such that $x''_ix'_i \succ'_A x_ix_1 \succ'_A x'_ix''_i$ for all $x_i$ in the sequence), then this sequence is finite.

**Representation Theorem.** Let X be a set which is order-dense with {X, $\succeq_A$} a difference-comparable, bounded multi-valued membership structure such that {X × X, $\succeq'_A$} is an algebraic-difference structure. Then there exists a bounded, real-valued function denoted by $\mu_A$ on X such that, for all $x_1, x_2, x_3, x_4 \in X$:

$$x_2 \succeq_A x_1 \Leftrightarrow \mu_A(x_2) \geq \mu_A(x_1) \tag{3.1}$$

and $$x_2x_1 \succeq'_A x_4x_3 \Leftrightarrow \mu_A(x_2) - \mu_A(x_1) \geq \mu_A(x_4) - \mu_A(x_3). \tag{3.2}$$

Note that, in this case, the weak ordering $\succeq_A$ is implicit in the weak ordering $\succeq'_A$. For $x_2x_1 \succeq'_A x_3x_3$, then (3.2) implies that

$$\mu_A(x_2) - \mu_A(x_1) \geq \mu_A(x_3) - \mu_A(x_3)$$

so that $\mu_A(x_2) \geq \mu_A(x_1)$ which, by (3.1), means that $x_2 \succeq_A x_1$.

**Uniqueness Theorem.** With the same conditions in the above theorem, if $\mu^*_A(x)$ is another function satisfying (3.1) and (3.2), then $\mu^*_A(x) = c_1\mu_A(x) + c_2$, where $c_1 > 0$.

Thus the interpretation of the membership value is as follow:
For any fuzzy set, if we can find a corresponding ordering, we can determine the membership function up to a linear transformation.

After the introduction of this interpretation, we would like to discuss how this interpretation related to the above four criteria:

1) This interpretation is not general enough because it requires the membership structure to be "bounded" and this restricts the application of this interpretation. Norwich and Turksen (1984) used "scale invariance" to decide whether any operation on one or more membership values to be meaningful. For example, they showed that in the case of conjunction, minimum operator will be scale invariance while product operator will not.
2) Norwich and Turksen has not provided an estimation procedure corresponding to his interpretation. But we have found an interesting characteristic of this interpretation. Let us consider the set of "tall people". For this set, "$x_2 \geq_{tall} x_1$" is equivalent to "$x_2$ is taller than $x_1$" and which is equivalent to "height of $x_1$ ≥ height of $x_1$". Moreover "$x_2 x_1 \geq'_{tall} x_4 x_3$" is equivalent to "$x_2$ is taller than $x_1$ by at least as much as $x_4$ is taller than $x_3$" and which is equivalent to "height of $x_2$ - height of $x_1$ ≥ height of $x_4$ - height of $x_3$". Without loss of generality, we assume that the height of people is a bounded concept. Moreover if we define $\mu_{tall}(x)$ = height of x, then we note that $\mu_{tall}(x)$ is a bounded real-valued function and

satisfies:

$$x_2 \succeq_{tall} x_1 \iff \mu_{tall}(x_2) \geq \mu_{tall}(x_1) \text{ and}$$

$$x_2x_1 \succeq'_{tall} x_4x_3 \iff \mu_{tall}(x_2) - \mu_{tall}(x_1) \geq \mu_{tall}(x_4) - \mu_{tall}(x_3).$$

Thus by the Uniqueness theorem, the membership function $\mu^*_{tall}(x)$ should relate to $\mu_{tall}(x)$ by

$$\mu^*_{tall}(x) = c_1\mu_{tall}(x) + c_2, \text{ where } c_1 > 0$$

Therefore the membership function is a linear function of height in this case.

3) The assumption of boundedness is natural but it reduces the generality of this interpretation. Moreover, this assumption can be tested easily. We can ask the subjects to check whether the attribute is bounded or not. Besides the assumption of boundedness, the algebraic-difference structure needs to satisfy five axioms in order for the Representation Theorem holds. The test of the axioms is simple (see Wallsten et al. (1986)).
4) This interpretation is based on mathematical theory, so that this criteria may not be applicable to this interpretation.
5) From the Uniqueness Theorem, it is obvious that the membership value is at least on an interval scale.

This interpretation implies that the source of fuzziness is that every subject has his ordering.

### 3.1.6. Interpretation in terms of Prototype Theory.

Osherson and Smith (1981, 1982) and Smithson (1986) discussed a theory called "prototype theory". This theory can be used to interpret the membership value in fuzzy set theory. According to prototype theory, many concepts can be represented in the form <A, d, p, c> where

A is a set of readily envisionable objects (real or imagined) called a conceptual domain;

d is a function from A × A into the positive real numbers, called a distance metric;

p is a member of A, called the concept's prototype; and

c is a function from A into [0,1], called the concept's characteristics function.

Moreover we need the following two conditions hold:

Condition 1 <A,d> is a metric space, i.e.,

$(\forall x \in A)(\forall y \in A)$

$$d(x,y) = 0 \text{ if and only if } x = y \tag{3.3}$$

$$d(x,y) = d(y,x) \tag{3.4}$$

$$d(x,y) + d(y,z) \geq d(x,z) \tag{3.5}$$

Condition 2 $(\forall x \in A)(\forall y \in A)$

$$d(x,p) \leq d(y,p) \Rightarrow c(y) \leq c(x). \tag{3.6}$$

The second condition requires that the closer an object is to its prototype, the more characteristics it is of the concept.

It is obvious that the prototype theory relates to the fuzzy set theory with characteristic function be viewed as the membership function. In this case the membership function is interpreted as the function which measure the difference between the object and the prototype.

After the introduction of this interpretation, we would like to discuss how this interpretation related to the above four criteria:

1) The interpretation is general enough that we can apply it to most fuzzy sets. But there are some problems in this interpretation. Osherson and Smith (1981) have shown that the prototype theory in conjunction with max-min laws for connective in fuzzy set theory contradicts strong intuitions that we have about concepts. They discuss this problem in three areas:

   a) when an object is more prototypical of a conjunction than of its constituents, the minimum operator for conjunction will lead to a contradiction,

   b) when we consider logically empty or logically universal concepts, the maximum operator for disjunction and the minimum operator for conjunction will lead to a contradiction,

   c) the maximum operator for disjunction will lead to a contradiction.

   Besides the max-min laws, we may define the conjunction and disjunction as:

   $$c_{A\cap B}(x) = c_A(x)\ c_B(x)$$

   $$c_{A\cup B}(x) = c_A(x) + c_B(x) - c_A(x)\ c_B(x).$$

   We can show that the above contradictions, except the third one, still exist.

2) This interpretation does not correspond to any estimation procedure

as there is no specification of the determination of the functions c and d.

3) The only assumption in this interpretation is that the existence of the distance metric and the characteristic function. Note that there is no need to assume the existence of the prototype as the prototype can be imaginary. Moreover, if we interpret fuzzy set theory in terms of prototype theory, the restrictions on the function d may be relaxed. For example, d needs not be a metric and it only needs to be a function measuring the "distance" from the prototype.

4) This interpretation is intuitive but we do not know membership value as the functions c and d are not determined. The determination of c and d can be arbitrary when we discuss one fuzzy set. But when we discuss two fuzzy sets A and B in case of connective, the determination of the corresponding $c_1$, $c_2$, $d_1$ and $d_2$ may not be so arbitrary. For example, suppose we use minimum operator for the conjunction connective. For $C = A \cap B$, the corresponding c of C will be equal to

$$c(x) = \mathrm{Min}\ [c_1(x),\ c_2(x)].$$

Moreover, we need $c_1(x) = g_1(d_1(x))$ and $c_2(x) = g_2(d_2(x))$ where $g_1$ and $g_2$ are some strictly decreasing functions, $d_1$ and $d_2$ are some functions measuring the "distance" of x from the corresponding prototype. If $g_1 = g_2$, then $c(x) = \mathrm{Min}\ [g_1(d_1(x)),\ g_1(d_2(x))] = g_1(d_3(x))$ where $d_3(x) = \mathrm{Max}\ [d_1(x),\ d_2(x)]$. In such definition, $d_3$ may not be a distance metric because we can show that (3.5) may not hold in some cases. Thus, in case of minimum, the determination of $d_1$

and $d_2$ is not arbitrary. If we want $d_3$ to be a distance metric, $d_1$ and $d_2$ must satisfy:

$d_1(x,y) \geq d_2(x,y) \quad \forall\ x,y$ or

$d_1(x,y) \leq d_2(x,y) \quad \forall\ x,y.$

The sources of fuzziness of this interpretation are that subjects may have different prototypes, different distance metric and different characteristic functions.

In the above sections, we have summarized the views on the membership values by different researchers and made a discussion. From the above review, we note the following point about interpretations:

"some of the interpretations are in terms of probability (usually subjective probability)."

It is natural as membership function resembles a lot as probability: both are between 0 and 1, both of them are used to quantify some belief, etc. Actually the question "whether membership values are probabilities" is raised since the appearance of fuzzy set theory and we will discuss this problem in Section 3.2.

**3.2. Discussion about Membership Function.**

We will discuss some questions about membership value, namely

1) Is grade of membership value probabilities?

2) What is the scale of membership value?

1) Ever since the inception of fuzzy set theory, a debate exists on whether fuzziness is equivalent to randomness and whether membership function is equivalent to probability function. The researchers, little by little, distinguished fuzziness from randomness conceptually and mathematically. The difference between probability theory and fuzzy set theory became clear when Zadeh (1978) proposed possibility theory:

"For any fuzzy set A, there will be a concept S correspond to A. (sometimes A is equal to S) S may be considered as a variable on a set T. For example, for the fuzzy set "tall man" A, there is a concept "tallness" S which may be considered as a variable on the real line T. Every number in the real line (corresponds to the height) may be assigned a membership value in S. Those membership values act as a fuzzy restriction on T which may be included in S. In this case, we says S induces a possibility distribution on T such that for each $x \in T$, the statement "x is S" has a possibility value $p_A(x)$."

In many cases, there are utterances which seem to involve possibility but can be interpreted in a probability approach. For example, "It is quite possible that John will be promoted". But there are also some uncertainties that can be interpreted in terms of possibility but not probability. For example, the possibility that John will take a job at Happy Factory is 0 if Happy Factory does not exist. The possibility may increase slightly if there is such a Happy

Factory and might increase to 1 if John is made an offer of a job at Happy Factory. But we do not know the probability that John will take a job at this factory when we allow John to choose it freely. Thus we know that possibility and probability are different. However many researchers have raised the questions on the similarities between probability theory and possibility theory. Natvig (1983) showed that a possibility distribution can be, at least in some applications, interpreted as a family of probabilities (usually subjective). Civanlar and Trussell (1986) gave a guideline to construct the membership functions for fuzzy sets whose elements have an attribute with a known probability density function (pdf) in the universe of discourse. Bordley (1989) gave three probabilistic models in which probability functions act like membership functions. So it is natural to ask what is the relation between fuzzy set theory and probability theory, even if they are different.

We know that the analysis of probability can be roughly divided in two approaches: one with a frequentist basis for estimating probabilities of events and the other one uses subjective probability. When we viewed randomness in the first approach, we can easily show the difference between fuzziness and randomness. For example, the outcome of tossing a coin is a problem of randomness, and there is no fuzziness related to this case. But the difference between fuzziness and subjective probability is not so obvious. Membership function measures degree of belonging and subjective probability measures degree of belief. These two quantities resemble a lot. Actually some researchers interpret membership value through a betting procedure (Giles, 1982), which is a usual practice in subjective probability. However, when a subject is asked to assign his or her degree of agreement with an attitudinal statement, it is not natural to consider it as a betting procedure. Moreover, with the reason in section 3.1.1, sometimes we can view fuzzy set theory as an extension of subjective probability. Thus subjective probability and fuzziness probably are not the same thing. However, as the probability theory and possibility theory resemble a lot, perhaps we would use the results in probability theory to develop results in fuzzy set theory.

2) For the scale of the membership value, we believe that it depends on the interpretations of membership value. For example, Turksen (1991) claimed interval scale for membership value and Saaty (1974, 1986) claimed ratio scale. It is because their interpretations of membership value are different.

**Chapter 4. Estimation of the Membership Function.**

As stated above, each person has his own membership function, and we have assumed that there is a "true" membership function, which is the average of all subjects' the membership functions. In a survey, only a part of the population is chosen. Thus their membership functions can only be used to estimate the "true" membership function. In this chapter, we would like to discuss the estimation of the "true" membership function based on the membership functions of m subjects. This chapter will be divided into five sections. In Section 4.1, we summarize the four data collection methods that are currently used in the estimation of the "true" membership value. We will investigate what information we can get from these methods in order to estimate the "true" membership value. In Section 4.2, we discuss the estimation procedures for the "true" membership values based on the data obtained by the four data collection methods and discuss their characteristics. In Section 4.3, we would like to point out the connections between the four data collection methods. In Section 4.4, we introduce some other estimation procedures besides those in Section 4.2. In Section 4.5, we discuss a survey carried out by us and some advice would be given to the researchers about what they need to take care in conducting their survey or experiment.

## 4.1. The Data Collection Methods for the Estimation of Membership Function.

Suppose we have a universe of discourse X and a fuzzy set A. We would like to estimate $\mu_A(x)$, the "true" membership value of x to the fuzzy set A based on the membership values of x of m subjects $\mu_{Ai}(x)$, i = 1,...,m. In current literatures, there are four data collection methods for estimation of the membership value, namely direct rating, polling, set-valued statistics and reverse rating.

### 4.1.1. Direct Rating.

The researcher will show a $x \in X$ to each of the m subjects, and the subjects are required to give the membership value of this x to the fuzzy set A. For example, the researcher may ask the subject "What is the membership value of this man to the set "tall man"?". Thus what we get is the estimate of the membership value of each subject in the survey, $\hat{\mu}_{Ai}(x)$, i=1,...,m.

### 4.1.2. Polling.

The researcher will show a $x \in X$ to the subjects and ask each of them to decide whether this x is a member of the set A or not. Each subject is required to give a "yes" or "no" answer. In this case, we get a "yes" or "no" answer from each subject. But how does this answer relate with the membership value $\mu_{Ai}(x)$? We make the assumption that the $i^{th}$ subject will give a "yes" answer if, and only if $\mu_{Ai}(x) \geq 0.5$. This assumption is reasonable when there are only two possible classifications, A or not

A, to describe the fuzzy set. Thus, under this assumption, we know which $\mu_{Ai}(x)$ is greater than 0.5 in the survey by the polling method.

### 4.1.3. Set-valued Statistics.

The researcher asks each of the m subjects to give a subset $A_{*i}$ of X which he think corresponds to the fuzzy set A. Then for any $x \in X$, whether x is in $A_{*i}$ is precise. But how do this answer relates with the membership value $\mu_{Ai}(x)$? We think that the relation between these $A_{*i}$'s and the subject's membership values is that:

$A_{*i} = \{x \in X: \mu_{Ai}(x) \geq \alpha_i,\ \alpha_i \in [0,1]\}$, i.e. $A_{*i}$ is the $\alpha_i$-level set. Thus, what we get are the subjects's $\alpha$-level sets, with $\alpha$ unknown. However, when there are only two possible classifications, A or not A, to describe the set A, it is reasonable to think that $\alpha \approx 0.5$. Thus what we obtain from the m subjects are m subsets of X. Each of these m subsets contains $x \in X$ such that the corresponding membership value is larger than 0.5. It seems that the information obtained by set-valued statistics is more than that from polling.

### 4.1.4. Reverse Rating.

The researcher will select a membership value $y \in [0,1]$ and ask each of the m subjects to identify which $x \in X$ will have this membership value in the fuzzy set A, i.e. find $x \in X$ such that $\mu_{Ai}(x) = y$ for a given $y \in [0,1]$. Normally the subject is randomly presented the same y a reasonable number of times in between the other random presentation of $y \in [0,1]$ in order to avoid memorization. Then his answers are averaged and used as the estimate of x. Thus in this case, we get an estimate of x, $\hat{x}$, such that $\mu_{A_i}(\hat{x}) = y$. Thus what we obtain is m subsets of X from the m subjects, each of these m subsets contains $x \in X$ such that the corresponding membership value is equal to y. It seems that the information from reverse rating is more than that from set-valued statistics in most cases.

## 4.2. Estimation Procedures for the Membership Function and Their Characteristics.

In order to estimate the "true" membership function, we use the data obtained from either of the above four data collection methods. There are two classes of estimation procedures to estimate the "true" membership function, namely "nonparametric" estimation procedures and "parametric" estimation procedures. For the first class, researchers use the data from either of the above four data collection methods to estimate the "true" membership values without making any assumptions on the "true" membership function. In the second class, we have to make certain assumptions on the "true" membership function. We will discuss two types of assumptions: the first type is that we assume a mathematical form of the membership function and the second type is that we assume the "true" membership values follow a certain statistical distribution. In the first type, most of the mathematical form depends on several parameters and what researchers need to do is the estimation of the parameters based on the data. In the second type, we will need to estimate the parameters in the distribution in order to estimate the "true" membership value. Moreover, we will summarize the work of Cai (1993) who estimates membership value based on assumptions different from the above two types.

### 4.2.1. Nonparametric Estimation Procedures.

After we have collected data from each of the above four methods, we can estimate the "true" membership values as follows:

#### 4.2.1.1. Direct Rating.

For any fixed $x \in X$, we have a random sample $\mu_{A1}(x), \ldots, \mu_{Am}(x)$ which are the membership values of x to the fuzzy set A assigned by the m subjects. This sample is a random sample from a certain distribution with $E(\mu_{Ai}(x)) = \mu_A(x)$. Thus a usual and reasonable estimator of $\mu_A(x)$ is the mean of this sample.

#### 4.2.1.2. Polling.

For any $x \in X$, we have a random sample of $y_1, \ldots y_m$ where $y_i$'s are random variables with value either zero or one. $Y_i = 1$ means that the $i^{th}$ subject classifies x to the fuzzy set A; while $Y_i = 0$ means that x is not classified to A. In usual practice, we use the average of these $Y_i$'s as an estimate of $\mu_A(x)$.

#### 4.2.1.3. Set-valued Statistics.

For any fuzzy set A, we have a random sample $A_1, \ldots, A_m$ where each of the $A_i$'s is a classical subset of X such that the $i^{th}$ subject think that it corresponds to A. For any $x \in X$, the estimated "true" membership value of x to the fuzzy set A is defined as the frequency of X in $A_i$, i.e.

$$\hat{\mu}_A(x) \equiv \frac{\text{the number of times of } `x \in A_i'}{m}.$$

For example, in order to estimate the "true" membership function of the fuzzy set "young man", Zhang (1981) asked 129 subjects to give an interval of age that they believe correspond to the set "young man". For any age $x \in X$, the estimate of the membership value of x to the set "young man" is equal to (total number of intervals that include x)/129.

#### 4.2.1.4. Reverse Rating.

For any fixed $y \in [0,1]$, we have a random sample of $x_1,\ldots,x_m$ which are the values such that $\mu_{Ai}(x_i) = y$ where $\mu_{Ai}$ is the membership function of the $i^{th}$ subject, $i = 1,\ldots,m$. The usual practice is to take the mean of these $x_i$'s, $\bar{x}$, to be an estimate of the value x, such that $\mu_A(x) = y$.

### 4.2.2. The Characteristics of the Non-parametric Estimation Procedures.

#### 4.2.2.1. Direct Rating.

This method is the most intuitive one when we believe that the determination of membership value is subjective in nature. Besides this advantage, this method needs no computation to obtain the membership value except taking average. However taking average is easily affected by unusual value. Among these four nonparametric estimation procedure, it is reasonable to believe that this estimation procedure is the most reliable one.

#### 4.2.2.2. Polling.

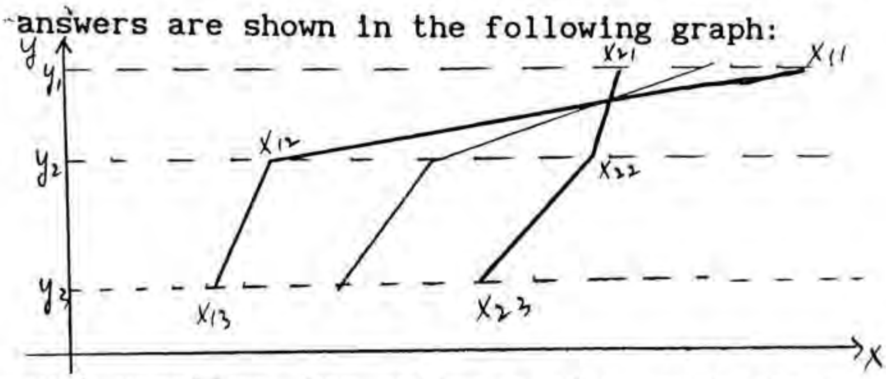
The subjects can make the decision easily for this estimation procedure. Moreover the computation of the membership values is easy. But a disadvantage is that, we need a rather large number of subjects so that the resulting membership values are more stable. Besides this, from past result, we know that the fuzzy region of the membership function by polling is smaller from that by direct rating. It seems that these two methods are not estimating the same thing. We will try to give an explanation in section 4.3.1.

#### 4.2.2.3. Set-valued Statistics.

The above two methods can only estimate the membership value of certain $x \in X$. However this method, with the reasons stated in section 4.3.2, is able to estimate the membership values of several x's at the same time. However this method need a large number of subjects so that the resulting membership values are more stable.

#### 4.2.2.4. Reverse Rating.

This method needs no computation to obtain the membership value except taking average. However taking average is easily affected by unusual values. Moreover it seems that taking average may not be a proper way to estimate the membership values. For example, suppose we have two subjects and we have given $y_1$, $y_2$ and $y_3 \in [0,1]$ to them and asked them to give the estimate of the corresponding $x_j$ such that $\mu_A(x_j) = y_j$. Let $x_{i1}$, $x_{i2}$ and $x_{i3}$ be the answers of the $i^{th}$ subject. Without loss of generality, we assume that $y_1 > y_2 > y_3$ and $x_{i1} > x_{i2} > x_{i3}$. Suppose their answers are shown in the following graph:




Note that in usual practice, the estimate of $x_j$ such that $\mu_A(x_j) = y_j$ is the mean of $x_{1j}$ and $x_{2j}$. However from the above graph, as the "distance" between $x_{11}$ and $x_{12}$ is much larger than that between $x_{21}$ and $x_{22}$, we would probably believe that the estimate of $x_2$ should be near $x_{22}$ than $x_{12}$, rather than $x_2$ is equidistant from $x_{12}$ and $x_{22}$. Thus taking average is inappropriate in this case.

In order to estimate the "true" membership function, we would first estimate the subject's membership function by connecting the points by a piecewise straight line. Then the average of these straight lines can be used as an estimate of the "true" membership function. Note that a piecewise straight line is inadequate in most cases. In order to solve this problem, we may fit the given set of points into a spline function for each subject. Then the average of these spline functions can be taken as an estimate of the "true" membership function. For details of the calculation of the spline function, please refer to Sastry et al. (1993).

### 4.2.3. Parametric Estimation Procedures.

In this procedure, we will make some assumptions on the "true" membership function. In the first type of assumptions, we will assume a mathematical form of the membership function. Usually this mathematical form consists of some parameters. What researchers need to do is to estimate these parameters using the data from either of the above four data collection methods. In current literatures, most of the researchers estimate the parameters based on the data from direct rating. For example, we propose a membership function for the fuzzy set "short people" as $\mu_A(x) = e^{-kx^2}$ where x is the height of people, $k > 0$ is a parameter to be determined. For a given set of height, we ask the subjects to determine the corresponding membership values by, say, direct rating method. Then we use these heights $x_i$ and the corresponding membership values to estimate k. In section 4.2.3.1., we will investigate how we can estimate the "true" membership function based on the first type assumption. Moreover we will summarize the mathematical forms used in the past. The forms will be divided into four classes according to the shape of the corresponding membership functions: decreasing, increasing, bell-shaped and inverted bell-shaped. In section 4.2.3.2., we will give details on how Dombi (1990) decides his mathematical form of membership function as an example. In the second type of assumptions, we will assume the "true" membership value has certain distribution or it is in some mathematical form with the parameters having some distribution. What we need to do is the estimation of the parameters in the distribution. In section 4.2.3.3., we try to estimate the "true" membership function if each of the individual membership function is a straight line and the end-points are uniform random variables. Besides the above two assumptions, Cai (1993) provides a third assumption on the membership function. He assumed that the membership function is in a certain mathematical form and he used maximum scale likelihood method to estimate the parameters in the mathematical form. We will discuss his work in section 4.2.3.4.. In section 4.2.3.5., we will discuss the characteristics of the parametric estimation procedures.

### 4.2.3.1. Parametric Estimation Procedures based on the First Type of Assumptions.

As there are many different estimation procedures, we will use the most popular approach, the least-square approach, to do the estimation. Suppose there are m subjects. We show $x_1, \ldots, x_n$ to each of them and ask them to give the corresponding membership values $\mu_{Aj}(x_i)$, i=1,...,n and j = 1,..., m. Thus for each subject, we base on the n membership values to estimate the parameters in the mathematical form. After we have found the m membership functions, we estimate the "true" membership function by the average of these m membership functions. We will show how to estimate the parameters by the least-square approach in the following example.

**Example** Suppose X, the universe of discourse, is the population of human being and A, the fuzzy set on X, is 'x is short' where $x \in X$. Moreover we assume that $\mu_A(x) = e^{-kx^2}$, where k is a positive constant. Suppose we have m subjects. In order to estimate the membership function of each subject, we showed each of them n x's, namely $x_1, \ldots, x_n$ and asked them to decide the corresponding membership value $\mu_A(x_1), \ldots, \mu_A(x_n)$. Then we can estimate k by finding the one which minimizes $\sum_{i=1}^{n} \left(\mu_A(x_i) - e^{-kx_i^2}\right)^2$ and which is the solution of the equation

$$\sum_{i=1}^{n} \left(\mu_A(x_i) - e^{-kx_i^2}\right) e^{-kx_i^2} x_i^2 = 0.$$

After we have estimated the k for each subject, then the 'true' membership value of a certain x is estimated by $\left(\sum_{i=1}^{m} e^{-\hat{k}_i x^2}\right)/m$. Note that, in usual practice, we will linearlized the membership function of each subject as $\ln \mu_A(x) = -kx^2$. Then we will estimate k by finding the one which minimizes $\sum_{i=1}^{n} \left(\ln \mu_A(x_i) + kx_i^2\right)^2$. The estimate is equal to

$$-\frac{\sum_{i=1}^{n} x_i^2 \ln \mu_A(x_i)}{\sum_{i=1}^{n} x_i^4}.$$

Programs for linear regression can be used in this approach but we do not think that it is a reasonable one, as we believe that each subject observes $e^{-kx^2}$ with certain additive error, rather than observing $-kx^2$ with certain additive error. Thus we will not linearize the membership function before we do the estimation.

In this section, we will summarize some mathematical forms of the membership functions that correspond to fuzzy set 'x is small', 'x is large', '$|x|$ is small' and '$|x|$ is large'.

**Table 1. Membership Function Corresponds to "x is small".**

| Domain | Function |
|---|---|
| 1) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 1, & 0 \le x \le a, \\ 0 & \text{otherwise.} \end{cases}$ |
| 2) $\mathbb{R}^+$ | $\mu_A(x) = e^{-kx}, \ k > 0.$ |
| 3) $\mathbb{R}^+$ | $\mu_A(x) = e^{-kx^2}, \ k > 0.$ |
| 4) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 1 & 0 \le x \le a_1, \\ \dfrac{a_2 - x}{a_2 - a_1} & a_1 \le x \le a_2, \\ 0 & a_2 \le x. \end{cases}$ |
| 5) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 1 - ax^k, & 0 \le x \le a^{-\frac{1}{k}}, \\ 0 & , \text{otherwise.} \end{cases}$ |
| 6) $\mathbb{R}^+$ | $\mu_A(x) = \dfrac{1}{1 + kx^2}, \ k > 1.$ |
| 7) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 0, & 0 \le x \le a, \\ \dfrac{1}{2} - \dfrac{1}{2} \sin \dfrac{\pi}{b - a} \left( x - \dfrac{a + b}{2} \right), & a \le x \le b, \\ 1, & b \le x. \end{cases}$ |

Table 2. Membership Function Corresponds to "x is large".

| Domain | Function |
|---|---|
| 1) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 0, & 0 \le x \le a, \\ 1, & a \le x. \end{cases}$ |
| 2) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 0, & 0 \le x \le \alpha, \\ 1 - e^{-k(x - \alpha)}, & \alpha \le x. \end{cases} \quad k > 0.$ |
| 3) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 0, & 0 \le x \le \alpha, \\ 1 - e^{-k(x - \alpha)^2}, & \alpha \le x. \end{cases} \quad k > 0.$ |
| 4) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 1 & 0 \le x \le a_1, \\ \dfrac{x - a_1}{a_2 - a_1} & a_1 \le x \le a_2, \\ 0 & a_2 \le x. \end{cases}$ |
| 5) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 0, & 0 \le x \le \alpha, \\ a(x - \alpha)^k, & \alpha \le x \le \alpha + a^{-\frac{1}{k}}, \\ 1, & \alpha + a^{-\frac{1}{k}}. \end{cases}$ |
| 6) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 0, & 0 \le x \le \alpha, \\ \dfrac{k(x - \alpha)^2}{1 + k(x - \alpha)^2}, & \alpha \le x \le \infty. \end{cases}$ |
| 7) $\mathbb{R}^+$ | $\mu_A(x) = \begin{cases} 0, & 0 \le x \le a \\ \dfrac{1}{2} + \dfrac{1}{2} \sin \dfrac{\pi}{b - a} \left( x - \dfrac{a + b}{2} \right), & a \le x \le b \\ 1, & b \le x. \end{cases}$ |

Table 3. Membership Function Corresponds to "$|x|$ is small".

| Domain | Function |
|---|---|
| 1) $\mathbb{R}$ | $\mu_A(x) = \begin{cases} 0, & -\infty \le x \le -a, \\ 1, & -a \le x \le a, \\ 0, & a \le x \le \infty. \end{cases}$ |
| 2) $\mathbb{R}$ | $\mu_A(x) = \begin{cases} e^{kx}, & -\infty < x \le 0, \\ e^{-kx}, & 0 \le x < \infty. \end{cases} \quad k > 1.$ |
| 3) $\mathbb{R}$ | $\mu_A(x) = e^{-kx^2}, \quad k > 0.$ |
| 4) $\mathbb{R}$ | $\mu_A(x) \begin{cases} = 0, & -\infty < x \le -a_2, \\ = \dfrac{a_2 + x}{a_2 - a_1}, & -a_2 \le x \le -a_1, \\ = 1, & -a_1 \le x \le a_1, \\ = \dfrac{a_2 - x}{a_2 - a_1}, & a_1 \le x \le a_2. \\ = 0, & a_2 \le x. \end{cases}$ |
| 5) $\mathbb{R}$ | $\mu_A(x) \begin{cases} = 0, & -\infty < x \le a^{-1/k}, \\ = 1 - a(-x)^k, & a^{-1/k} \le x \le 0, \\ = 1 - a(x)^k, & 0 \le x \le a^{-1/k}, \\ = 0, & a^{-1/k} \le x < \infty. \end{cases}$ |
| 6) $\mathbb{R}$ | $\mu_A(x) = \dfrac{1}{1 + kx^2}, \quad k > 1.$ |
| 7) $\mathbb{R}$ | $\mu_A(x) \begin{cases} = 0, & -\infty < x \le -b, \\ = \dfrac{1}{2} + \dfrac{1}{2} \sin \dfrac{\pi}{b - a} \left( x + \dfrac{a + b}{2} \right), & -b \le x \le -a, \\ = 1, & -a \le x \le a, \\ = \dfrac{1}{2} - \dfrac{1}{2} \sin \dfrac{\pi}{b - a} \left( x - \dfrac{a + b}{2} \right), & a \le x \le b, \\ = 0, & b \le x. \end{cases}$ |

**Table 4. Membership Function Corresponds to "$|x|$ is large".**

| Domain | Function |
|---|---|
| 1) $\mathbb{R}$ | $\mu_A(x) = \begin{cases} 1, & -\infty \le x \le -a, \\ 0, & -a \le x \le a, \\ 1, & a \le x \le \infty. \end{cases}$ |
| 2) $\mathbb{R}$ | $\mu_A(x) = \begin{cases} 1 - e^{kx}, & -\infty < x \le 0, \\ & \qquad k > 1. \\ 1 - e^{-kx}, & 0 \le x < \infty. \end{cases}$ |
| 3) $\mathbb{R}$ | $\mu_A(x) = 1 - e^{-kx^2}, \qquad k > 0.$ |
| 4) $\mathbb{R}$ | $= 1, \qquad -\infty < x \le -a_2,$ <br> $= -\dfrac{a_1 + x}{a_2 - a_1}, \qquad -a_2 \le x \le -a_1,$ <br> $\mu_A(x) = 1, \qquad -a_1 \le x \le a_1,$ <br> $= \dfrac{x - a_1}{a_2 - a_1}, \qquad a_1 \le x \le a_2,$ <br> $= 1, \qquad a \le x.$ |
| 5) $\mathbb{R}$ | $= 1, \qquad -\infty < x \le a^{-1/k},$ <br> $= a(-x)^k, \qquad a^{-1/k} \le x \le 0,$ <br> $\mu_A(x)$ <br> $= a(x)^k, \qquad 0 \le x \le a^{-1/k},$ <br> $= 1, \qquad a^{-1/k} \le x < \infty.$ |
| 6) $\mathbb{R}$ | $\mu_A(x) = \dfrac{kx^2}{1 + kx^2}, \quad k > 1.$ |
| 7) $\mathbb{R}$ | $= 1, \qquad -\infty < x \le -b,$ <br> $= \dfrac{1}{2} - \dfrac{1}{2} \sin \dfrac{\pi}{b - a} \left( x + \dfrac{a + b}{2} \right), \qquad -b \le x \le -a,$ <br> $\mu_A(x) = 0, \qquad -a \le x \le a,$ <br> $= \dfrac{1}{2} - \dfrac{1}{2} \sin \dfrac{\pi}{b - a} \left( x - \dfrac{a + b}{2} \right), \qquad a \le x \le b,$ <br> $= 0, \qquad b \le x.$ |

We will show below the estimates of the parameters in the above mathematical forms for each of the m subjects. Then the "true" membership function will be estimated by the average of these m membership functions. As the forms in the four tables are very similar, we will only show those in table 1 as examples. Suppose that we have $x_1,\ldots,x_n$ and the corresponding membership values $y_1 = \mu_A(x_1),\ldots, y_n = \mu_A(x_n)$.

1) The largest $x_i$ such that $\mu_A(x_i) = 1$ is an estimator of a.

2) Based on the least-square approach, we estimate k by finding the one which is the solution of the equation $\sum_{i=1}^{n}\left(y_i - e^{-kx_i}\right)x_i e^{-kx_i} = 0.$

3) Based on the least-square approach, we estimate k by finding the one which is the solution of the equation $\sum_{i=1}^{n}\left(\mu_A(x_i) - e^{-kx_i^2}\right) x_i^2\, e^{-kx_i^2} = 0.$

4) In order to estimate $a_1$ and $a_2$, we will do the following constrained minimization problem:

$$\underset{a_1,a_2}{\mathrm{Min}} \sum_{i=1}^{n}\left[ y_i - \mu_A(x_i) \right]^2 \quad \text{subject to}$$

$$\mu_A(x) = \begin{cases} 1 & 0 \le x \le a_1, \\ \dfrac{a_2 - x}{a_2 - a_1} & a_1 \le x \le a_2, \\ 0 & a_2 \le x. \end{cases}$$

5) In order to estimate $a_1$ and $a_2$, we will do the following constrained minimization problem:

$$\min_k \sum_{i=1}^{n} \left( y_i - \mu_A(x_i) \right)^2 \quad \text{subject to}$$

$$\mu_A(x) = \begin{cases} 1 - ax^k, & 0 \le x \le a^{-\frac{1}{k}}, \\ 0 & , \text{otherwise.} \end{cases}$$

6) Based on the least-square approach, we estimate k by finding the one which is the solution of the equation $\sum\left(y_i - \dfrac{1}{1 + kx_i^2}\right)x_i^2 = 0.$

7) In order to estimate $a_1$ and $a_2$, we will do the following constrained minimization problem:

$$\min_k \sum_{i=1}^{n} \left( y_i - \mu_A(x_i) \right)^2 \quad \text{subject to}$$

$$\mu_A(x) = \begin{cases} 0, & 0 \le x \le a, \\ \dfrac{1}{2} - \dfrac{1}{2} \sin \dfrac{\pi}{b - a} \left( x - \dfrac{a + b}{2} \right), & a \le x \le b, \\ 1, & b \le x. \end{cases}$$

#### 4.2.3.2. Dombi's Mathematical Forms of Membership Functions.

Dombi (1990) summarized the mathematical forms of membership functions that had appeared in the journal "Fuzzy Set and System". After the review, he pointed out some properties that all of these membership functions have:

1) all membership functions are continuous,

2) $\mu_A(x)$ : $[a,b] \longrightarrow [0,1]$,

3) $\mu_A(x)$ is (a) either increasing, or (b) decreasing, or (c) could be divided into an increasing and decreasing part,

4) the monotonous membership function in the whole interval is (a) either convex functions or (b) concave functions, or (c) $\exists$ a point $c \in [a,b]$ such that the function is convex in $[a,c]$ and concave in $[c,b]$,

5) if the function is increasing, then $\mu_A(a)=0$ and $\mu_A(b)=1$; if the function is decreasing, then $\mu_A(a)=1$ and $\mu_A(b)=0$,

6) the membership function can be easily linearlized.

Basing on the above findings, he set up some criteria that a membership function from [0,1] to [0,1] should satisfy:

1) $\mu_A(x)$ is a continuous increasing function : $[0,1] \longrightarrow [0,1]$,

2) $\mu_A(0) = 0$, $\mu_A(1) = 1$,

3) $\mu_A'(0) = 0$, $\mu_A'(1) = 0$,

4) $\mu_A(x) = \dfrac{a_0x^n+...+a_n}{A_0x^m+...+A_m}$,

5) for such kind of $\mu_A(x)$ in (4) n+m need to be minimum.

Basing on the above five criteria, he derived the form of such membership function from [a,b] to [0,1] :

For increasing $\mu_A(x)$

$$\mu_A(x) = \frac{(1-\nu)^{\lambda-1}(x-a)^{\lambda}}{(1-\nu)^{\lambda-1}(x-a)^{\lambda} + \nu^{\lambda-1}(b-x)^{\lambda}} \qquad x \in [a,b] .$$

For decreasing $\mu_A(x)$

$$\mu_A(x) = \frac{(1-\nu)^{\lambda-1}(b-x)^{\lambda}}{(1-\nu)^{\lambda-1}(b-x)^{\lambda} + \nu^{\lambda-1}(x-a)^{\lambda}} \qquad x \in [a,b] ,$$

where $\nu$ is the intersection of $y = \mu_A(x)$ and $y = x$ (and so $\nu$ determines the shape of $\mu_A(x)$ ) and $\mu_A'(\nu) = \lambda$ (and so $\lambda$ determines the sharpness of $\mu_A(x)$ )

**4.2.3.3. Parametric Estimation based on the Second Type of Assumptions.**

In this section, we will assume that the membership function of each of the m subjects is a straight line, with end-points $(a_i,0)$ and $(b_i,1)$. This assumption is reasonable as the average of straight lines is a s-shaped curve. Moreover we assume that the random variables $(a_i,b_i)$ have a joint probability density function $f(a_i,b_i) = \frac{1}{\theta_1(\theta_3 - \theta_2)}$ for $0 < a_i < \theta_1 < x < \theta_2 < b_i < \theta_3$ where $x \in X$. We would like to find the maximum likelihood estimates of $\theta_1$, $\theta_2$ and $\theta_3$ based on the data from either of the above four data collection methods.

**4.2.3.3.1. Direct Rating.**

As we assume $\theta_1 < x < \theta_2$, $Pr(0 < y_i < 1) = 1$. From direct rating, we observe $y_1,\ldots, y_m$ where $y_i = \frac{x - a_i}{b_i - a_i}$. Let $z_i = x - a_i$. Then the joint probability density function of $y_i$ and $z_i$ is:

$$f(y_1, z_1) = \frac{z_1}{y_1^2\ \theta_1\ (\theta_3 - \theta_2)}$$

$$\text{where} \begin{cases} \dfrac{x - \theta_1}{\theta_3 - \theta_1} < y_1 < \dfrac{x}{\theta_2} \quad \text{and} \\ \text{Max}\left( x - \theta_1, \dfrac{(\theta_2 - x)y_1}{1 - y_1} \right) < z_1 < \text{Min}\left(x, \dfrac{(\theta_3 - x)y_1}{1 - y_1}\right). \end{cases}$$

The probability density function of $y_1$ is

$$f(y_1) = \frac{g(y_1)}{2y_1^2\ \theta_1\ (\theta_3 - \theta_2)} \left\{ \text{Min}\left(x, \frac{(\theta_3 - x)y_1}{1 - y_1}\right)^2 - \text{Max}\left(x - \theta_1, \frac{(\theta_2 - x)y_1}{1 - y_1}\right)^2 \right\}$$

where $g(y_1)$ is an indicator function and $g(y_1) = 1$ if $\dfrac{x - \theta_1}{\theta_3 - \theta_1} < y_1 < \dfrac{x}{\theta_2}$ and $= 0$ otherwise.

In this case, when we observe $y_1, \dots, y_m$, the loglikelihood equation is:

$L(\theta_1, \theta_2, \theta_3) =$

$$\sum_{i=1}^{m} \log \frac{g(y_i)}{2y_i^2\ \theta_1\ (\theta_3 - \theta_2)} \left\{ \text{Min}\left(x, \frac{(\theta_3 - x)y_i}{1 - y_i}\right)^2 - \text{Max}\left(x - \theta_1, \frac{(\theta_2 - x)y_i}{1 - y_i}\right)^2 \right\}.$$

Thus by maximizing $L(\theta_1, \theta_2, \theta_3)$ with respect to $\theta_1$, $\theta_2$ and $\theta_3$, we get the maximum likelihood estimates of $\theta_1$, $\theta_2$ and $\theta_3$. Then for any given x, we estimate $\mu_A(x)$ by finding $\int y\ f_1(y)\ dy$ where $f_1(y)$ is equal to $f(y)$ with $\theta_1$, $\theta_2$ and $\theta_3$ replaced by their maximum likelihood estimates. Note that the most important advantage of this method is that after we have observed m membership values of one x, we can estimate the "true" membership values of other x's.

#### 4.2.3.3.2. Polling.

Under this method, we observe $v_1, \ldots, v_m$ where $v_i = \begin{cases} 1 & \text{if } y_i \geq 0.5 \\ 0 & \text{if } y_i < 0.5. \end{cases}$

Then the likelihood equation is

$$L(\theta_1, \theta_2, \theta_3) = \left( \Pr(Y_1 \geq 0.5)^{\sum v_i} \Pr(Y_1 < 0.5)^{m - \sum v_i} \right).$$

Note that if $\dfrac{x - \theta_1}{\theta_3 - \theta_1} \geq 0.5$, then $\Pr(Y_1 \geq 0.5) = 1$ and if $\dfrac{x}{\theta_2} \leq 0.5$, then $\Pr(Y_1 \geq 0.5) = 0$. Suppose we observe some $v_i = 1$ and some $= 0$. Then we have $\dfrac{x - \theta_1}{\theta_3 - \theta_1} \leq 0.5 \leq \dfrac{x}{\theta_2}$ and $\Pr(Y_1 \geq 0.5) = \displaystyle\int_{0.5}^{\frac{x}{\theta_2}} f(y_1)\, dy_1$ where $f(y_1)$ is the one in section 4.2.3.3.1. Thus, in this case, the likelihood equation is :

$$L(\theta_1, \theta_2, \theta_3) =$$

$$\left( \int_{0.5}^{\frac{x}{\theta_2}} f(y_1)\, dy_1 \right)^{\sum v_i} \left( 1 - \int_{0.5}^{\frac{x}{\theta_2}} f(y_1)\, dy_1 \right)^{n - \sum v_i}.$$

Note that if $0.5 < \dfrac{x - \theta_1}{\theta_3 - \theta_1}$, then $\displaystyle\int_{0.5}^{\frac{x}{\theta_2}} f(y_1)\, dy_1 = 1$.

If $\dfrac{x - \theta_1}{\theta_3 - \theta_1} < 0.5 < \dfrac{x - \theta_1}{\theta_2 - \theta_1}$, then $\displaystyle\int_{0.5}^{\frac{x}{\theta_2}} f(y_1)\, dy_1 =$

$$\frac{\left(2x - \theta_2\right)\left[x\left(\theta_3 - x\right)^2 - \left(\theta_2 - x\right)\left(x - \theta_1\right)^2\right]}{2\, x\, \theta_1 \left(\theta_2 - x\right)\left(\theta_3 - \theta_2\right)}.$$

If $\dfrac{x - \theta_1}{\theta_2 - \theta_1} < 0.5 < \dfrac{x}{\theta_3}$, then $\displaystyle\int_{0.5}^{\frac{x}{\theta_2}} f(y_1)\, dy_1 =$

$$\frac{\left(2x - \theta_2\right)\left[\left(\theta_3 - x\right)^2 - \left(\theta_2 - x\right)^2\right]}{2\, \theta_1 \left(\theta_2 - x\right)\left(\theta_3 - \theta_2\right)}.$$

If $\dfrac{x}{\theta_3} < 0.5 < \dfrac{x}{\theta_2}$, then $\displaystyle\int_{0.5}^{\frac{x}{\theta_2}} f(y_1)\, dy_1 = \frac{2\, x \left(x - \theta_2\right)}{\theta_1 \left(\theta_3 - \theta_2\right)}.$

If $\dfrac{x}{\theta_2} < 0.5$, then $\displaystyle\int_{0.5}^{\frac{x}{\theta_2}} f(y_1)\, dy_1 = 0.$

Thus by maximizing $L(\theta_1, \theta_2, \theta_3)$ with respect to $\theta_1$, $\theta_2$ and $\theta_3$, we get the maximum likelihood estimates of $\theta_1$, $\theta_2$ and $\theta_3$. Note that for each x, we observe $\sum v_i$. In order to find the MLE's of $\theta_1$, $\theta_2$ and $\theta_3$, we need three x's and the corresponding three $\sum v_i$.

#### 4.2.3.3.3. Set-valued Statistics and Reverse Rating.

Suppose we observe $\beta(b_1 - a_1) + a_1, \ldots, \beta(b_m - a_m) + a_m$ where $\beta \in (0,1)$. Note that in the case of set-valued statistics, $\beta = 0.5$. We would like to find the probability density function of $c_i = \beta(b_i - a_i) + a_i$. Then the joint probability density function of $(c_i, a_i)$ is,

$$f(c_i, a_i) = \frac{1}{\beta\, \theta_1 (\theta_3 - \theta_2)}$$

$$\text{if} \begin{cases} \beta\, \theta_2 < c_i < (1-\beta)\, \theta_1 \text{ and} \\ \mathrm{Max}\left(0, \dfrac{c_i - \beta\, \theta_3}{1-\beta}\right) < d_i < \mathrm{Min}\left(\theta_1, \dfrac{c_i - \beta\, \theta_2}{1-\beta}\right). \end{cases}$$

Therefore,

$$f(c_i) = \frac{g(c_i)}{\beta\, \theta_1 (\theta_3 - \theta_2)} \left\{ \mathrm{Min}\left(\theta_1, \frac{c_i - \beta\, \theta_2}{1-\beta}\right) - \mathrm{Max}\left(0, \frac{c_i - \beta\, \theta_3}{1-\beta}\right) \right\}$$

where $g(c_i)$ is an indicator function which is equal to one if $\beta\theta_2 < c_i < (1-\beta)\theta_1$ and is equal to zero otherwise. Thus when we observed $c_1, \ldots, c_m$, the loglikelihood equation is:

$L(\theta_1, \theta_2, \theta_3) =$

$$\sum_{i=1}^{m} \log\left[ \frac{g(c_i)}{\beta\, \theta_1 (\theta_3 - \theta_2)} \left\{ \mathrm{Min}\left(\theta_1, \frac{c_i - \beta\, \theta_2}{1-\beta}\right) - \mathrm{Max}\left(0, \frac{c_i - \beta\, \theta_3}{1-\beta}\right) \right\} \right].$$

Thus by maximizing $L(\theta_1, \theta_2, \theta_3)$ with respect to $\theta_1$, $\theta_2$ and $\theta_3$, we get the maximum likelihood estimates of $\theta_1$, $\theta_2$ and $\theta_3$. Then for any given x, we estimate $\mu_A(x)$ by following the same method used in Section 4.2.3.3.1.

### 4.2.3.4. Parametric Estimation based on the Third Type of Assumptions.

Cai (1993) assumes that the membership function of a certain fuzzy set is $\mu_A(x) = \exp\left[-\left(\frac{x-a}{b}\right)^2\right]$ where $a \in \mathbb{R}$ and $b > 0$. In order to estimate the parameters a and b, he has used a method called maximum scale likelihood method. This method is very similar to maximum likelihood method in probability theory, except that it uses possibility instead of probability and it uses a concept "unrelated" instead of "independent" in probability theory. In maximum likelihood method, our objective is to maximize a function called likelihood function which is the product of n probability density functions. In maximum scale likelihood method, our objective is to maximize a function called scale likelihood function which is the minimum of n possibility distribution functions. In the paper, he has discussed parameter estimation in three distinct situations: estimation of a with b known, estimation of b with a known and estimation of a and b when both unknown.

Case 1 Estimation of a with b known

Suppose we have observed n $x_i$'s such that $\mu_A(x_i) = 1$. By Point Estimation Method (PEM), the estimate of a is determined by the average of these $x_i$'s. By Maximum Scale Likelihood Method (MSLM), the estimate of a is determined by

$$\hat{a} = 0.5\left(\underset{1 \le i \le n}{\text{Max}}\ x_i + \underset{1 \le i \le n}{\text{Min}}\ x_i\right).$$

Case 2 Estimation of b with a known

Suppose we have observed n $x_i$'s such that $\mu_A(x_i) = 1$. The estimate of b is determined by

$$\hat{b} = \frac{\max\limits_{1\le i\le n} |x_i - a|}{\varepsilon_\alpha}$$ where $\varepsilon_\alpha$ is a predetermined constant which determined the accuracy of the estimate.

Case 3 Estimation of a and b when both unknown

Suppose we have observed n $x_i$'s such that $\mu_A(x_i) = 1$. The estimates of a and b are determined by

$$\hat{a} = 0.5\left(\max_{1\le i\le n} x_i + \min_{1\le i\le n} x_i\right) \text{ and } \hat{b} = \frac{\max\limits_{1\le i\le n} x_i - \min\limits_{1\le i\le n} x_i}{2\varepsilon_\alpha}$$ where $\varepsilon_\alpha$ is a predetermined constant which determined the accuracy of the estimate.

**4.2.3.5. The Characteristics of the Parametric Estimation Procedure.**

We summarize the advantages and disadvantages of this estimation procedure as:

Advantage

1) Given the functional form of the membership function, the analysis of fuzzy set theory will be easier. For example, Bobrowicz et al. specify the membership function in some special functional form so that they can use this function's properties to carry out further analysis and consider possible applications.

Disadvantages

1) Usually the authors do not give the reasons why the membership function is in such form even though the functional forms they proposed are reasonable.
2) This approach can be used only in the case that the attribute can be mapped to the real line. For those cases where the attribute

cannot be mapped to the real line, the above approach cannot be used. For example, for the set of "beautiful girl", obviously we cannot follow the above approach and write down the membership function in term of several parameters.

3) In most cases, we believe that each fuzzy set corresponds to some attributes and the membership value of certain element x to certain fuzzy set A depends on the quantity of certain attribute that x owns. For example, when we consider the set of "young people", we will believe that their age is the only factor to determine their membership values to this fuzzy set. But from past experience, we know that beside the age, subject will consider other factors, such as appearance, character, dressing and even career when they consider the set of "young people". So in this case, this approach may not be applicable. Even if it is applicable, the problem will be very complicated.

After discussing the characteristics of these four methods, we would like to see if there are any connection between them.

### 4.3. Connections between the Four Data Collection Methods in Section 4.1

In Section 4.1, we have introduced the four methods that are commonly used in the past to collect data for the estimation of the membership value. Now we would like to see if they are related or not.

### 4.3.1. Relation between Direct Rating and Polling.

For the polling method, past result showed that the fuzzy region (the region of x such that the membership values lie strictly between zero and one) is smaller than that obtained by direct rating method. We would like to explain this phenomenon in this section. We believe that the reason of this phenomenon is that, by polling method, what we are estimating is not the same as that by direct rating. When researchers asked the subjects to answer a "yes" "no" question, the subjects will set a bound on the attribute of x (or $\mu_A(x)$) such that when the given element y's attribute (or $\mu_A(y)$) exceeds that of x, the subject will give an "yes" answer. For example, when a subject is asked to decide whether a man is a tall man or not, the subject may probable decide a certain height of man in his mind such that if the given man's height exceeds this height, he will give a "yes" answer to the question. Equivalently, the subject may decide a membership value such that if the given man's membership value to the set of "tall man" exceeds this membership value, then he will give a "yes" answer. Now let us consider the latter case. For a subject at certain place and time, the membership value of x to the fuzzy set A is different. So we can view $\mu_A(x)$ as a random variable for each subject. Suppose f(z) is the probability density function of $z=\mu_A(x)$ and F(z) is the corresponding distribution function. Then

$$\int_0^1 F(z)\, dz = \Big[z\, F(z)\Big]_0^1 - \int_0^1 z\, f(z)\, dz$$

$$= 1 - \mu$$

where $\mu = E(z) = \int_0^1 zf(z)\,dz$. We assume that for $\mu_A(x) \geq c$, the subject will answer "yes" where c is a constant between zero and one. Let $p = F(c) = \int_0^c f(z)\,dz$, so that the probability to answer "yes" is equal to

$$Pr(\mu_A(x) \geq c) = \int_c^1 f(z)\,dz = 1 - p.$$

Under our interpretation, polling method will give an estimate of $1 - p$, rather than $\mu$. We would like to find out the relation between $1 - p$ and $\mu$. In the following theorem, we will consider f(z) as a unimodal function with peak (d,f(d)) and c = 1/2. We will prove that $1 - p \geq \mu$ when $\mu \geq 3/4$.

**Theorem 1.** Let $z=\mu_A(x)$ be a random variable in the region [0,1] with probability density function f(z) and distribution function F(z). Suppose f(z) is an unimodal function with peak (d,f(d)). Then

$Pr(z \geq 0.5) \geq E(z)$ when $E(z) \geq 3/4$.

In order to prove this theorem, we prove the following lemma which consider the relation between p and $\mu$ when f(z) is nondecreasing.

**Lemma 1:**

Let $z = \mu_A(x)$ be a random variable in the region [0,1] with probability density function (pdf) f(z) and f(z) is nondecreasing in [0,1]. Let $\mu = E(z)$, c be a point in [0,1] and $p = \int_0^c f(z)\,dz$. Then we will have the following relations between c, p and $\mu$:

a) If $0 \le c \le \frac{1}{2}$ and $1 - 2\mu + c \ge 0$,

$2c(1 - \mu) \ge p \ge 1 - 2\mu + c$.

b) If $0 \le c \le \frac{1}{2}$ and $1 - 2\mu + c \le 0$,

$2c(1 - \mu) \ge p \ge 0$.

c) If $\frac{1}{2} \le c \le 1$, $c \le \mu$ and $1 - 2\mu + c \ge 0$,

$\frac{1 - \mu}{2(1-c)} \ge p \ge 1 - 2\mu + c$.

d) If $\frac{1}{2} \le c \le 1$, $c \le \mu$ and $1 - 2\mu + c \le 0$,

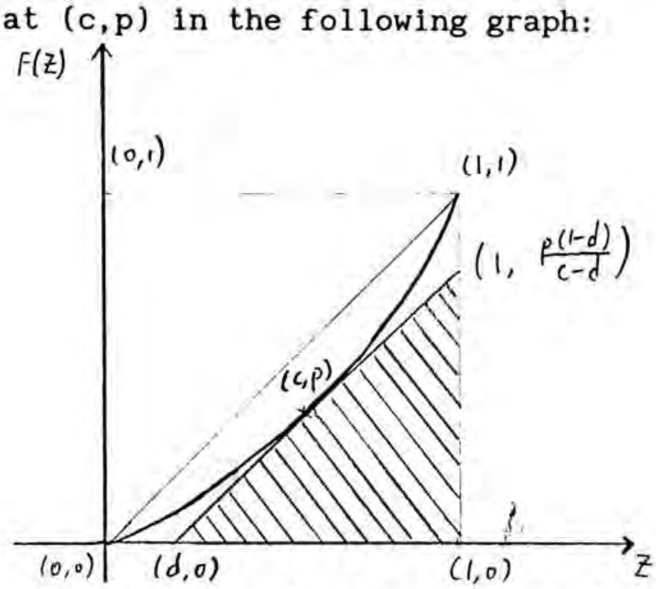
$\frac{1 - \mu}{2(1-c)} \ge p \ge 0$.

e) If $\frac{1}{2} \le c \le 1$, $c \ge \mu$ and $1 - 2\mu + c \ge 0$,

$\frac{1 - 2\mu + c}{2(1 - \mu)} \ge p \ge 1 - 2\mu + c$.

**Proof of lemma 1:**

Suppose F(z) is the distribution function of z. When f(z) is nondecreasing in [0,1], F(z) is a convex function in [0,1]. We draw a tangent line at (c,p) in the following graph:




Then $\int_0^1 F(z)\,dz \ge$ (minimum area of the shaded triangle). Suppose the tangent line cuts the x-axis at (d,0). The equation of this tangent line is:

$$y = [p/(c-d)]z - pd/(c-d).$$

As for z=1, $y \le 1$, therefore we have $\frac{p(1-d)}{c-d} \le 1 \Rightarrow d \le \frac{c-p}{1-p}$.

As for z=0, $y \le 0$, therefore we have $-pd \le 0 \Rightarrow d \ge 0$.

We conclude that the d need to satisfy

$$0 \le d \le \frac{c-p}{1-p} \le c < 1 \tag{4.1}$$

Note that for y = 0, z will equal to d and for z = 1, y will equal to $\frac{p(1-d)}{c-d}$, so the area of the shaded triangle say g(d) is $\frac{p(1-d)^2}{2(c-d)}$. So $g'(d) = \frac{p(1-d)(1-2c+d)}{2(c-d)^2}$.

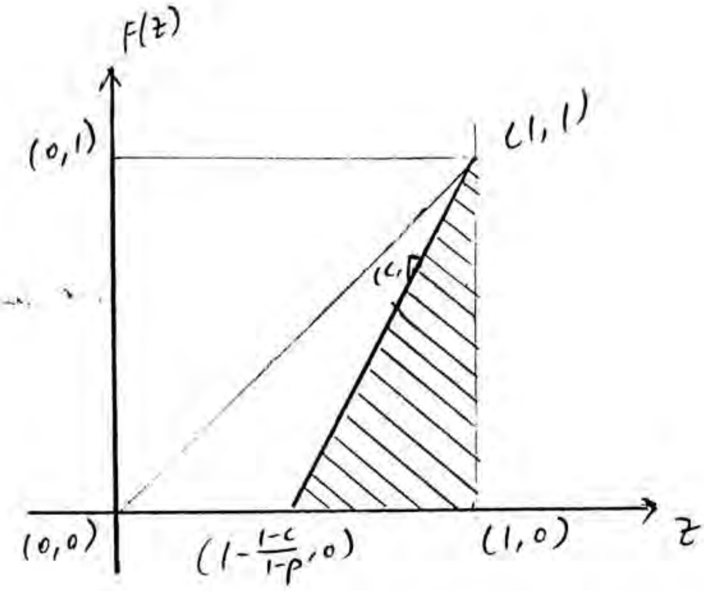
Therefore $g'(d) = 0$ if and only if $d = 1$ or $d = 2c - 1$. So in order to find the minimum area of the shaded triangle we need to take care the case when d = 1 and the case when d = 2c - 1. But d = 1 does not satisfy (4.1). So we only need to take care of d = 2c-1. From (4.1) and when $d = 2c - 1$, $0 \le d \le \frac{c-p}{1-p} \Rightarrow 0 \le 2c - 1 \le \frac{c-p}{1-p}$. Note that $0 \le 2c - 1 \Rightarrow 1/2 \le c$ and $2c - 1 \le \frac{c-p}{1-p} \Rightarrow \frac{(1-c)(1-2p)}{1-p} \ge 0 \Rightarrow p \le 1/2$.

Therefore for $1/2 \le c \le 1$ and $p \le 1/2$, minimum area of the shaded triangle is $g(2c-1) = 2p(1-c)$. So $1 - \mu \ge 2p(1-c) \Rightarrow p \le \frac{1-\mu}{2(1-c)}$.

That is, for $1/2 \le c \le 1$ and $p \le 1/2$, $p \le \frac{1-\mu}{2(1-c)}$. **(4.2)**

For $1/2 \le c \le 1$ and $p \ge 1/2$, it is easy to conclude, from the following graph, that the minimum area of the shaded triangle = $\frac{1-c}{2(1-p)}$. Therefore $1 - \mu \ge \frac{1-c}{2(1-p)}$ which implies $p \le \frac{1-2\mu+c}{2(1-\mu)}$ and so $1/2 \le p \le \frac{1-2\mu+c}{2(1-\mu)}$. And $1/2 \le \frac{1-2\mu+c}{2(1-\mu)} \Rightarrow \mu \le c$.

That is, for $1/2 \le c \le 1$ and $p \ge 1/2$, we have $1/2 \le p \le \frac{1-2\mu+c}{2(1-\mu)}$ and $\mu \le c$. **(4.3)**




In the above, (4.2), (4.3), we have found the relation between c, $\mu$ when p is in some region. However we would like to find the lower and upper bound of p in term of $\mu$ and c. So let us divide the above two cases into the following four cases :

a) If $1/2 \le c \le 1$, $c < \mu$ and $p \ge 1/2$, from (4.3)

$$1/2 \le p \le \frac{1-2\mu+c}{2(1-\mu)} \Rightarrow \mu \le c \text{ which is impossible.}$$

b) If $1/2 \le c \le 1$, $c < \mu$ and $p \le 1/2$, from (4.2)

$$p \le \frac{1-\mu}{2(1-c)}.$$

As $c < \mu \Leftrightarrow \frac{1-\mu}{2(1-c)} < \frac{1}{2}$, we have $\frac{1-\mu}{2(1-c)} \ge p \Rightarrow \frac{1}{2} > p$.

c) If $1/2 \le c \le 1$, $c \ge \mu$ and $p \ge 1/2$, from (4.3)

$$\frac{1}{2} \le p \le \frac{1-2\mu+c}{2(1-\mu)}.$$

d) If $1/2 \le c \le 1$, $c \ge \mu$ and $p \le 1/2$, from (4.2)

$$p \le \frac{1-\mu}{2(1-c)}.$$

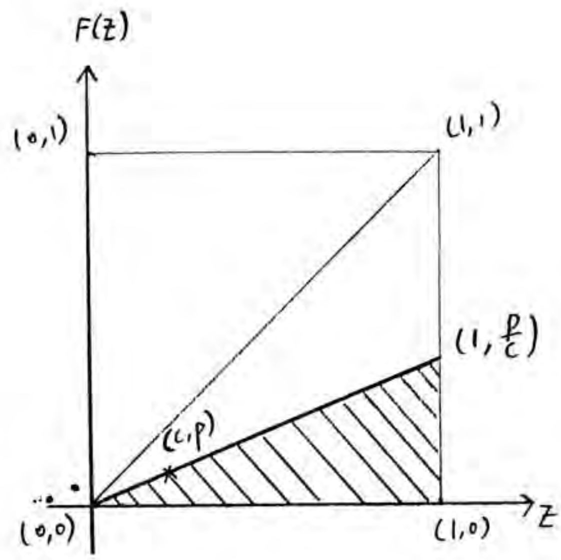
Case (c) and case (d) can be combined as:

If $\frac{1}{2} \le c \le 1$ and $c \ge \mu$, then $\frac{1-2\mu+c}{2(1-\mu)} \ge p$ (note that $\frac{1-2\mu+c}{2(1-\mu)} < \frac{1-\mu}{2(1-c)}$ ) and case (b) is equivalently to

If $\frac{1}{2} \le c \le 1$ and $c < \mu$, then $\frac{1-\mu}{2(1-c)} \ge p$. (note that $\frac{1-\mu}{2(1-c)} < \frac{1}{2}$)

Case (b) to case (d) are for $\frac{1}{2} \le c \le 1$. Let us consider the case when $0 \le c \le \frac{1}{2}$.

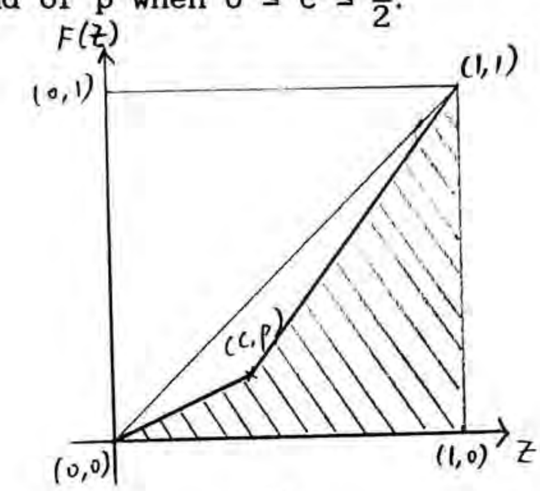
From the graph below, when $0 \le c \le \frac{1}{2}$, minimum area of the shaded triangle $= \frac{p}{2c}$.




Therefore $1 - \mu \ge \frac{p}{2c} \Rightarrow p \le 2c(1-\mu)$.

That is, for $0 \le c \le \frac{1}{2}$, $p \le 2c(1-\mu)$. **(4.4)**

We have found the upper bound of p. From the graph below, we can write out the lower bound of p when $0 \le c \le \frac{1}{2}$.



$1 - \mu \le \frac{pc}{2} + \frac{(1-c)(1+p)}{2}$ which is equivalently to $p \ge 1 - 2\mu + c$.

That is, for $0 \le c \le \frac{1}{2}$, $p \ge 1 - 2\mu + c$. **(4.5)**

We may conclude the above result as below:

a) If $0 \le c \le \frac{1}{2}$ and $1 - 2\mu + c \ge 0$,

$$2c(1 - \mu) \ge p \ge 1 - 2\mu + c. \tag{4.6}$$

b) If $0 \le c \le \frac{1}{2}$ and $1 - 2\mu + c \le 0$,

$$2c(1 - \mu) \ge p \ge 0. \tag{4.7}$$

c) If $\frac{1}{2} \le c \le 1$, $c \le \mu$ and $1 - 2\mu + c \ge 0$,

$$\frac{1 - \mu}{2(1-c)} \ge p \ge 1 - 2\mu + c. \tag{4.8}$$

d) If $\frac{1}{2} \le c \le 1$, $c \le \mu$ and $1 - 2\mu + c \le 0$,

$$\frac{1 - \mu}{2(1-c)} \ge p \ge 0. \tag{4.9}$$

e) If $\frac{1}{2} \le c \le 1$, $c \ge \mu$ and $1 - 2\mu + c \ge 0$,

$$\frac{1 - 2\mu + c}{2(1 - \mu)} \ge p \ge 1 - 2\mu + c. \tag{4.10}$$

f) If $\frac{1}{2} \le c \le 1$, $c \ge \mu$ and $1 - 2\mu + c \le 0$,

$$\frac{1 - 2\mu + c}{2(1 - \mu)} \ge p \ge 0. \tag{4.11}$$

for (f), $1 + c \le 2\mu \le 2c \Rightarrow 1 \le c$ which is impossible.

This completes the proof of lemma 1.

**Proof of the theorem:**

Assume $f(z)$ is a unimodal function in $[0,1]$. Without loss of generality we assume that it is increasing for $z \in [0,d]$ and is nonincreasing for $z \in (d,1]$. Let $q$ be the probability that $z \in [0,d]$ and so $1 - q$ is the probability that $z \in (d,1]$. Therefore we can view $z$ as a mixture of two distributions, which is symbolically expressed as $z = qS + (1-q)T$ where

S, T are random variables with S corresponds to the increasing part and T corresponds to the nonincreasing part. Moreover we let $\mu_1 = E(S)$ and $\mu_2 = E(T)$ and we assume that $\mu_1 \neq \mu_2$. Thus we have

$$\mu = q\,\mu_1 + (1-q)\,\mu_2 \quad \text{or } q = \frac{\mu_2 - \mu}{\mu_2 - \mu_1}.$$

We will use the result from lemma 1 with c = 1/2. (Note that it is reasonable that the subjects use 0.5 as a cutting point when there are only two adjectives to describe the fuzzy set, e.g. the set of "tall" and "not tall" man, the set of "long" and "not long" line.) Now we will split the above case into two cases:

<u>Case 1</u> $d > c$

$$\Pr(z \le 1/2) = q \Pr(S \le 1/2)$$
$$= q \Pr\left(\frac{S}{d} \le \frac{1}{2d}\right).$$

As F(S) is convex in [0,d], $F\left(\frac{S}{d}\right)$ is convex in [0,1]. Therefore we can use the result in lemma 1 with c replaced by $\frac{1}{2d}$ and $\mu$ replaced by $\frac{\mu_1}{d}$. From (4.6) to (4.10), we will have (4.12) to (4.16) as follow:

a) If $0 \le \frac{1}{2d} \le 1/2$ and $1 - \frac{2\mu_1}{d} + \frac{1}{2d} \ge 0$,

$$2\,q\,\frac{1}{2d}\left(1 - \frac{\mu_1}{d}\right) \ge p \ge \left(1 - 2\,\frac{\mu_1}{d} + \frac{1}{2d}\right) q. \tag{4.12}$$

b) If $0 \le \frac{1}{2d} \le 1/2$ and $1 - \frac{2\mu_1}{d} + \frac{1}{2d} \le 0$,

$$2q\,\frac{1}{2d}\left(1 - \frac{\mu_1}{d}\right) \ge p \ge 0. \tag{4.13}$$

c) If $1/2 \le \frac{1}{2d} \le 1$, $1 - \frac{2\mu_1}{d} + \frac{1}{2d} \ge 0$ and $\frac{1}{2d} \le \frac{\mu_1}{d}$,

$$q\,\frac{1 - \frac{\mu_1}{d}}{2\left(1 - \frac{1}{2d}\right)} \ge p \ge \left(1 - 2\,\frac{\mu_1}{d} + \frac{1}{2d}\right) q. \tag{4.14}$$

d) If $1/2 \le \frac{1}{2d} \le 1$, $1 - \frac{2\mu_1}{d} + \frac{1}{2d} \le 0$ and $\frac{1}{2d} \le \frac{\mu_1}{d}$,

$$q \frac{1 - \frac{\mu_1}{d}}{2(1 - \frac{1}{2d})} \ge p \ge 0. \tag{4.15}$$

e) If $1/2 \le \frac{1}{2d} \le 1$, $1 - \frac{2\mu_1}{d} + \frac{1}{2d} \ge 0$ and $\frac{1}{2d} \ge \frac{\mu_1}{d}$,

$$q \frac{1 - \frac{2\mu_1}{d} + \frac{1}{2d}}{2 ( 1 - \frac{\mu_1}{d} )} \ge p \ge ( 1 - 2 \frac{\mu_1}{d} + \frac{1}{2d} ) q. \tag{4.16}$$

Note that for (4.12) and (4.13), $\frac{1}{2d} \le 1/2 \Rightarrow d \ge 1$, which is impossible.

As we are only interested in the upper bound of p, we only need to take care the upper bound of p in (4.14) to (4.16)

(4.14) Suppose $1/2 \le d \le 1$, $\mu_1 \ge 1/2$ and $1 - \frac{2\mu_1}{d} + \frac{1}{2d} \ge 0$.

(4.15) Suppose $1/2 \le d \le 1$, $\mu_1 \ge 1/2$ and $1 - \frac{2\mu_1}{d} + \frac{1}{2d} \le 0$.

Both of them has an upper bound of p as

$$q \frac{1 - \frac{\mu_1}{d}}{2 ( 1 - \frac{1}{2d} )} \ge p.$$

As for (4.14) and (4.15), we want to maximize

$$q \frac{1 - \frac{\mu_1}{d}}{2 ( 1 - \frac{1}{2d} )}$$

which was equal to $\dfrac{\mu_2 - \mu}{\mu_2 - \mu_1} \dfrac{1 - \frac{\mu_1}{d}}{2 ( 1 - \frac{1}{2d} )}$

$$= \frac{\mu_2 - \mu}{2(d - 0.5)} \frac{d - \mu_1}{\mu_2 - \mu_1}$$

$$= \frac{\mu_2 - \mu}{2( d - 0.5)} \frac{\mu_2 - \mu_1 - ( \mu_2 - d)}{\mu_2 - \mu_1}$$

As $\mu_2 \geq \mu_1$ and $\mu_2 \geq d$, the quantity is maximized if $\mu_1$ is minimized to 1/2 and the corresponding quantity is $\frac{\mu_2 - \mu}{2( \mu_2 - 0.5)}$. This quantity is equal to $1/2\left(1 + \frac{0.5 - \mu}{\mu_2 - 0.5}\right)$ and this one is maximized if $\mu_2$ is equal to 1. Therefore

$$\text{Max} \left\{ q \frac{1 - \frac{\mu_1}{d}}{2( 1 - \frac{1}{2d})} \right\} = 1 - \mu \geq p.$$

For (4.16), we want to maximize $q \frac{1 - \frac{2 \mu_1}{d} + \frac{1}{2d}}{2( 1 - \frac{\mu_1}{d} )}$ which was equal to

$$\frac{\mu_2 - \mu}{\mu_2 - \mu_1} \frac{1 - \frac{2\mu_1}{d} + \frac{1}{2d}}{2( 1 - \frac{\mu_1}{d})}.$$

We can show that in order to maximize the given quantity, we need the largest $\mu_1$ ( = 1/2), the largest $\mu_2$ ( = 1.0 ), the largest d ( = 1.0) and $\mu \geq 3/4$. In this case the corresponding quantity is $1 - \mu$.

Summarizing the result in (4.14) to (4.16), we conclude that for $1/2 \leq d$, we have $1 - \mu \geq p$ when $\mu \geq 3/4$.

<u>Case 2</u> $d \leq c$

As now $\Pr (z \leq c) = q + (1 - q) \Pr (T \leq c)$

$$= q + (1 - q)\left[1 - \Pr\left(\frac{1 - T}{1 - d} \le \frac{1 - c}{1 - d}\right)\right].$$

As F(T) is concave in [d,1], $F\left(\frac{1 - T}{1 - d}\right)$ is convex in [0,1]. Therefore we can use the result in lemma 1 with $c \Rightarrow \frac{1 - c}{1 - d}$ and $\mu \Rightarrow \frac{1 - \mu_2}{1 - d}$. As $\mu_2 \ge 1/2$, $0 \le d \le 1/2$, we can show that only case (e) in lemma 1 is possible. Therefore $p \le q + (1 - q)\left[1 - 1 + 2\,\frac{1 - \mu_2}{1 - d} - \frac{1 - c}{1 - d}\right]$

$$= \frac{\mu_2 - \mu}{\mu_2 - \mu_1} + \frac{\mu - \mu_1}{\mu_2 - \mu_1}\left[\frac{3 - 4\,\mu_2}{2(1 - d)}\right] \equiv f(\mu_1, \mu_2, d)$$

Obviously if we want to maximize f with respect to d, it occurs at $d = 1/2$ and $f(\mu_1, \mu_2, 1/2) = (3 - 4\,\mu_2) + \frac{\mu_2 - \mu}{\mu_2 - \mu_1}(4\mu_2 - 2)$

As $\mu_1 \le 1/2$, therefore with respect to $\mu_1$, maximum of $f(\mu_1, \mu_2, 1/2)$ occurs at $\mu_1 = 1/2$ and $f(1/2, \mu_2, 1/2) = (3 - 4\,\mu)$.

As $3 - 4\,\mu \le 1 - \mu \Leftrightarrow 2/3 \le \mu$, therefore $p \le 1 - \mu$ if $\mu \ge 2/3$ in this case.

From both case 1 and case 2, we conclude that $1 - \mu \ge p$ when $\mu \ge 3/4$.

**Q.E.D.**

**Theorem 2.** Let $z = \mu_A(x)$ be a random variable in the region [0,1] with probability density function f(z) and distribution function F(z). Suppose f(z) is an unimodal function with peak (d,f(d)). Then

$\Pr(z \ge 0.5) \le E(z)$ when $E(z) \le 1/4$.

**Proof of theorem 2.**

The proof is the same as that of theorem 1 except that we consider $v = 1 - z$ in this case.

**Q.E.D.**

Therefore if subjects use a threshold as 1/2 in membership value to decide a "yes" "no" answer, then we have shown why polling will overestimate the membership value when the membership value is greater than 0.5 and underestimate it when the membership value is smaller than 0.5.

#### 4.3.2. Relation between Polling and Reverse Rating.

If we have only two classification A and not A, it is reasonable to assume that membership value 0.5 is a cutting point for the polling method. Under this assumption, we can show that polling and reverse rating are closely related. Suppose the researcher has asked 100 subjects to give a $y \in X$ such that $\mu_A(y) = 0.5$. Then in this case, the researcher will get 100 such y's and he can use them to estimate the membership function obtained by polling. For any $x \in X$, the researcher counts number of y's such that $x \geq y$ in some sense and let us call this number $N_{x,y}$. Then $(N_{x,y})/100$ is an estimate of the membership value by polling method. So reverse rating obtains more information than polling in this case as it can estimate the membership value by polling method but not vice verse.

#### 4.3.3. Relation between Reverse Rating and Set-valued Statistics.

In the above sense, reverse rating and set-valued statistics are closely related. Both of them estimate the membership value by the information in the threshold of membership value. But the reverse rating requires the subjects to estimate the threshold and set-valued statistics only need the subjects to estimate an subset which contains the threshold.

The above are the estimation procedures based on the data from the four data collection methods. In current literatures, there are some estimation procedures that are based on the other form of data. We will mention two of these as an example.

### 4.4. Other Estimation Procedures.

#### 4.4.1. Procedure based on Saaty's Matrix.

Saaty (1974, 1986) proposed the following data collection method to estimate the membership values:
Suppose we have n objects, $x_1,\ldots,x_n$ that we want to find their membership values to the fuzzy set A. Then we ask the subject to give an estimate of $m_{ij} = \frac{\mu_A(x_i)}{\mu_A(x_j)}$, i, j = 1,...,n and i ≠ j. In order to be unified and simplify the matter, the subject can choose $m_{ij}$ only from the set $\Omega = \{9,8,\ldots,2,1, 1/2, 1/3, 1/4,\ldots,1/9\}$ to determine the ratio. Each number has its meaning as below :

**Table 5. Scale of numerical values and qualitative judgements**

| Intensity of Importance | Definition | Explanation |
|---|---|---|
| 1 | Equal importance | Two alternatives contribute equally to the objective |
| 3 | Weak importance of one over another | Experience and judgement slightly favor one alternative over another |
| 5 | Essential or strong importance | Experience and judgement strongly favor one alternative over another |
| 7 | Demonstrated importance | An alternative is strongly favored and its dominance is demonstrated in practice |
| 9 | Absolute importance | The evidence favoring one alternative over another is of the highest possible order of affirmation |
| 2,4,6,8 | Intermediate values between the two adjacent judgement | When compromise is needed |
| Reciprocals of above numbers | If alternative i has one of the above numbers assigned to it when compared with alternative j, then j has the reciprocal value compared with i. | |

After we have formed the matrix M, we have the following estimation procedures to estimate the membership values:

1) Saaty's (1974) method:

For the n × n matrix $M = (m_{ij})$ with $m_{ij} = \dfrac{\mu_A(x_i)}{\mu_A(x_j)}$. We find that $Mw = nw$ where $w' = (w_1, w_2, \ldots, w_n)$ and $w_i = \mu_A(x_i)$. In this case, w is an eigenvector of M with the eigenvalue n. So in order to estimate the membership values of $x_i$'s, the subject is needed to give an estimate of M. After the matrix M is formed, the researchers find the largest eigenvalue of this matrix and the corresponding eigenvector. The eigenvalue needs to be close to n and the corresponding standardized membership values are obtained by dividing the elements of the eigenvector by their sum.

2) First method of Chu et al.:

Chu et al. (1979) solve the following constrained minimization problem to estimate the membership values:

$$\text{Minimize } S = \sum_{i=1}^{n}\sum_{j=1}^{n}(m_{ij}w_j - w_i)^2$$

$$\text{subject to } \sum_{i=1}^{n} w_i = 1 \text{ and } w_i > 0 \text{ for } i = 1,\ldots,n.$$

3) Second method of Chu et al.:

Chu et al. (1979) also suggest another constrained minimization problem to estimate the membership values:

$$\text{Minimize } S = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(m_{ij} - \frac{w_j}{w_i}\right)^2$$

$$\text{subject to } \sum_{i=1}^{n} w_i = 1 \text{ and } w_i > 0 \text{ for } i = 1,\ldots,n.$$

Advantage

1) This estimation procedure can apply to all kinds of fuzzy set.

Disadvantages

1) In order to estimate the membership values of n objects, the subject needs to make $\frac{n(n-1)}{2}$ comparisons. Thus the number of questions needed is much larger than that required by other procedures.
2) This approach assumes that the membership value is on the ratio scale.
3) In order to be unified and simplify the estimation procedure, Saaty restricts the choice of $m_{ij}$ to the set of nineteen elements. But it is not natural to do so.
4) These estimation procedures are not intuitive. It is very difficult to believe an human can find the eigenvalue and eigenvector of a matrix in his mind or find the solution of a constrained minimization problem.
5) We need to do some calculations in order to obtain the estimated membership values.
6) Saaty (1974) recommends normalizing each component of the eigenvector with respect to the sum of all components. Thus, results are affected by the number of objects and no membership value could ever be 0 or 1.
7) Saaty has not proved that whether the eigenvalues will be greater than or equal to 0.

### 4.4.2. Procedure based on Mabuchi's Interpretation of the Membership Function.

Based on the interpretation of Mabuchi (1992) on membership function, we can find the following data collection method to estimate the membership function:

For any fuzzy set A, we ask the subjects to decide the corresponding ground set $G_A(x)$ which is a set of attribute that these subjects think that an object should have in order to belong to A.

Suppose for a fuzzy set A, we have determined the corresponding $G_A(x)$. Then for any object x, we check that whether it satisfies the attributes, the elements of $G_A(x)$ to find what percentage of them, with weighted valuation if necessary, support that x belongs to A. The obtained percentage gives an estimate of $\mu_A(x)$.

Mabuchi has not given any suggestion on the choice of the weight function of the attribute. A possible way is to count the number of subjects that think the attribute should be included. Suppose we have m subjects. Each of them is required to give a list of attributes. The weight of each attributes can be determined by the number of these m subjects who include this attribute in his list. This weight function is reasonable but we have to assume that we can know whether two attributes are distinct or not. For example, suppose we discuss the set "old man". If a subject thinks that the attribute "character" should be included while the other subject thinks that the attribute "behavior" should be included. Then we need to make a difficult decision that whether these two attributes are different.

In section 4.3.1 to 4.3.3, we have shown the relations between the methods in Section 4.1. This relations depend heavily on the polling methods but this method always gives biased estimate of the membership values when compared with direct rating method. So we would like to see what we can do to reduce this "bias" from polling data. We have conducted an survey to acquire the membership values of given lines to the set "long line". Both direct rating data and polling data are obtained. In Section 4.5, we will briefly introduce this survey. Then we will summarize the result of the survey. We will suggest a method to reduce the "bias" of the polling data. At last we will summarize the problems that we had faced during the survey and we will give some advices to researchers about what they need to take care during their surveys.

## 4.5. The Survey.

### 4.5.1. Introduction of the Survey.

Among the published papers in fuzzy set in these 30 years, some of them are about acquisition of the membership function through an experiment. In order to know more about the membership function, we have conducted a survey. In the survey, we asked the subjects to answer a questionnaire, which is included in the appendix. The whole questionnaire is divided into three parts: the first part is about the membership values of some given lines to the set "long line", the second part is about the membership values of certain ages to the set "young people in Hong Kong" and the third part is about the subject's personnel information. We choose the set "long line" because we are interested in the membership values when the range of the attribute is fixed. We choose the set of "young people in Hong Kong" because we are interested in the membership values when the range of the attribute is finite but unknown. The subjects of this survey are thirteen graduate students of the Statistics Department of The Chinese University of Hong Kong. All of them are volunteers and are not paid.

In the beginning of the first part, we have stated that the lengths of the lines are between 0 cm to 10 cm. The reason for this is that the subjects can see the line on the questionnaire. Moreover we have stated that there are only two classifications for the lines, namely "a long line" and "not a long line" so that the subject cannot classify a line as a "medium line". In question 1, we asked the subjects what factors would they consider, besides the line's length, when they are asked to classify a line to either "a long line" or "not a long line". This question is raised because we want to know what are the factors affecting the subjects to decide the membership values. In question 2, we asked what is the minimum length of a line for them to consider it as a "long" line. This question is raised because we want to know whether the mid-value of the range of the attribute will be a "cutting point" when the label set has "two elements". In question 3 to question 6, we present 10 lines to the subjects in each question. All the lines are drawn with the exact length. The lines are ordered so that the length of the lines are decreasing from top to bottom. Their lengths are not explicitly given but grid lines are added so that the subjects can estimate their lengths. The subjects are required to classify the lines to either "a long line". or "not a long line". Besides this, they are required to give the membership value of each lines to the set of "long line". These four questions are raised because we want to answer the following questions :

Does the membership function resemble a straight line?

Does having membership value larger than or equal to 0.5 imply that the line will be classified to a "long line"?

Are the membership values affected by the length of other lines?

In question 7, thirty lines are shown to the subjects. They are required to answer the same questions as in question 3 for eleven lines out of thirty. The eleven lines are specified. This question is raised because we suspect that the number of lines in question (3) to (6) is too small. And so we increase the number of lines and want to know the effects on membership values.

In the beginning of the second part, we have stated that we assume that we can classify a Hong Kong people to either "a young man" or "not a young man". In question 1, we asked the subjects what factors would they consider, besides people's ages, when they are asked to classify a people to either "a young people" or "not a young people". This question is raised due to the same reason as in part one. In question 2 to 4, we ask the subjects information (the maximum, the mean and the median) about the distribution of the age of Hong Kong people. We asked the maximum because we want to know whether the range and the mid-value have any meanings in the membership values. We asked the mean and the median because we want to know whether these ages could affect the determination of membership values. In question 5, we asked the subject what is minimum ages of a Hong Kong people that he would consider the people as "not a young people". The purpose of this question is the same as that of question 2 in part 1. In question 6, twelve ages are given to the subjects. They are required to do the following three things :

give the membership values for each age,

classify a people with this age as "a young people" or "not a young people",

classify a people with this age as "an old people" or "not an old people".

The purpose of this question is the same as that of question (3) to (7) in part 1. We require the subjects to do the second and the third thing because we want to know whether "young" is equal to "not old" and "not young" is equal to "old".

In the third part, we asked the subjects about their personnel information such as sex and age.

**4.5.2. The Result of the Survey.**

The result of the experiment is summarized as follow:

Part 1

1.1) Some of the subjects replied that they would consider other factors when they are required to do such a classification. The factors include the length of other lines, the thickness of the lines. It is reasonable for the first factor but it is interesting that thickness is also a factor. Actually it is difficult to believe that a line can has any thickness.

1.2) The distribution of the minimum length of line for the subjects to consider it as a "long line" :

**Table 6. Distribution of the minimum length.**

| Length in cm | count |
|---|---|
| 5 | 4 |
| 6 | 2 |
| 6.5 | 1 |
| 7 | 3 |
| 7.5 | 1 |
| 8 | 2 |

with mean = 6.38 cm and standard deviation = 1.14 cm.

From the above result, we found that in most cases the mid-value of length ( 5 cm ) is not the "cutting point". The "cutting point" is greater than the mid-value.

1.3) The answer to this question is summarized as follow:

**Table 7. Estimated membership values of ten lines.**

| length in cm | membership value obtained by | |
|---|---|---|
| | polling | direct rating (the mean of 13 answers) |
| 8.1 | 1 | 0.878 |
| 7.8 | 0.923 | 0.831 |
| 6.9 | 0.923 | 0.759 |
| 6.5 | 0.615 | 0.671 |
| 5.8 | 0.385 | 0.542 |
| 3.7 | 0 | 0.300 |
| 3.1 | 0 | 0.215 |
| 2.9 | 0 | 0.180 |
| 2.5 | 0 | 0.138 |
| 2.2 | 0 | 0.116. |

1.4) The answer to this question is summarized as follow:

**Table 8. Estimated membership values of ten lines.**

| length in cm | membership value obtained by | |
|---|---|---|
| | polling | direct rating (the mean of 13 answers) |
| 9.03 | 1 | 0.942 |
| 7.39 | 0.923 | 0.812 |
| 6.05 | 0.615 | 0.599 |
| 4.95 | 0.154 | 0.435 |
| 4.06 | 0 | 0.333 |
| 3.32 | 0 | 0.276 |
| 2.72 | 0 | 0.220 |
| 2.22 | 0 | 0.172 |
| 1.82 | 0 | 0.129 |
| 1.49 | 0 | 0.081. |

1.5)The answer to this question is summarized as follow:

**Table 9. Estimated membership values of ten lines.**

| length in cm | membership value obtained by | |
|---|---|---|
| | polling | direct rating (the mean of 13 answers) |
| 10 | 1 | 0.994 |
| 9 | 1 | 0.925 |
| 8 | 1 | 0.860 |
| 7 | 0.923 | 0.753 |
| 6 | 0.615 | 0.617 |
| 5 | 0.154 | 0.450 |
| 4 | 0 | 0.335 |
| 3 | 0 | 0.248 |
| 2 | 0 | 0.155 |
| 1 | 0 | 0.061. |

1.6) The answer to this question is summarized as follow:

**Table 10. Estimated membership values of ten lines.**

| length in cm | membership value obtained by | |
|---|---|---|
| | polling | direct rating (the mean of 13 answers) |
| 10 | 1 | 0.987 |
| 9.8 | 1 | 0.954 |
| 9.5 | 1 | 0.936 |
| 9.3 | 1 | 0.919 |
| 9.0 | 1 | 0.903 |
| 8.8 | 1 | 0.886 |
| 8.5 | 1 | 0.871 |
| 8.3 | 1 | 0.858 |
| 8.0 | 1 | 0.840 |
| 1.0 | 0 | 0.063. |

1.7) The answer to this question is summarized as follow:

**Table 11. Estimated membership values of eleven lines.**

| length in cm | membership value obtained by | |
|---|---|---|
| | polling | direct rating (the mean of 13 answers) |
| 9.73 | 1 | 0.959 |
| 8.23 | 1 | 0.872 |
| 6.94 | 0.769 | 0.717 |
| 5.42 | 0.308 | 0.493 |
| 4.91 | 0 | 0.405 |
| 4.41 | 0 | 0.347 |
| 2.88 | 0 | 0.229 |
| 2.12 | 0 | 0.168 |
| 1.03 | 0 | 0.091 |
| 0.36 | 0 | 0.046 |
| 0.05 | 0 | 0.011. |

From the results in question (3) to (7), we have the following findings:

1) Polling overestimates the membership value when the membership value is large than 0.5 (true for question 4, 6 and 7) and underestimates it when it is small than 0.5 (true for question 3 to 7) when compared with the membership value obtained by direct rating.

2) Among the answers of the thirteen subjects, only two of them show that "membership value from direct rating ≥ 0.5" may not imply "the

answer is "yes" to the question "Is this a long line?"". The answers of the remaining eleven subjects show the above relation. All the subjects will give a "yes" answer only if the direct rating membership value ≥ 0.5. None of the subjects decide a line as a long line with the membership value < 0.5. This result seems to contradict our assumption. The reason of this contradiction may be :

a) the two subjects may not only use the two labels to describe the lines, e.g. they may classify a line as a "medium line". Then in this case, it is reasonable to find such contradiction.

b) even they use the two labels to describe the lines, they may not interpret these two labels as we do. For example, for the label "not long", the subjects may perceive it as the same meaning of "medium or short".

3) We are interested in whether the membership function is a straight line. Actually we have carried out a statistical test for it. Suppose we have 10 lines of length $x_i$, $i = 1,\ldots,10$ and $x_1 \geq x_2 \geq \ldots \geq x_{10}$. The corresponding membership values are $\mu_A(x_i)$, $i = 1,\ldots,10$ and $1 > \mu_A(x_1) \geq \ldots \geq \mu_A(x_{10}) > 0$. Then we consider

$$y_i = \frac{\mu_A(x_{i+1}) - \mu_A(x_i)}{x_{i+1} - x_i} \quad \text{for } i = 1,\ldots,9.$$

If the assumption of a straight line is correct, then we will have $y_1 = \ldots = y_9$. Let $z_i = y_i - y_{i+1}$ $i=1,\ldots,8$. We assume that these $z_i$'s come from a normal distribution and use the Hotelling T-test to test the hypothesis

$H_o$ : $E(Z_1) = E(Z_2) = \ldots = E(Z_8) = 0$ vs $H_1$: not all $E(Z_i)$'s are equal to zero.

Suppose the $\alpha$-level is 0.05. The corresponding $T_0^2 = 92.544$ and we reject $H_0$ if $T^2 > T_0^2$. The results for question (3) to (7) are as follow:

question 3) $T^2 = 166.43$ and so $H_0$ is rejected

question 4) $T^2 = 30.08$ and so $H_0$ is not rejected

question 5) $T^2 = 19.47$ and so $H_0$ is not rejected

question 6) $T^2 = 68.38$ and so $H_0$ is not rejected

question 7) $T^2 = 246.49$ and so $H_0$ is rejected.

It seems that we cannot reject the idea that each subject's membership function is a straight line.

4) We found that lines of same length may not have the same membership values when the other lines are different. For example, the mean membership value of the line with length 5.8 cm is 0.542 in question 3, the mean membership value of line with length 6.05 cm is 0.599 in question 4 and the mean membership value of line with length 6 cm is 0.617 in question 5. This may be explained by the fact that in the determination of membership value, the subjects will consider not only the length of the line but also the length of the other lines.

5) We may find that the membership values is frequently equal to zero by polling method. This may be because the number of subjects in this experiment is too small.

6) When the results of question 3 to question 6 are compared with that of question 7, we could not find a large difference. Thus the number of lines does not seem to affect the membership value.

Part 2

2.1)The subjects will consider other factors when they are asked to do such a classification, e.g. appearance, dressing, manner, career, character, behavior, marital status, interest, life style, etc. It is interesting to find that there are so many factors to consider when they are asked to do such a classification.

2.2) to (2.5) The mean and the standard deviation for these questions are summarized as follow:

**Table 12. Summarized results of question 2.2 - 2.6**

| | Mean | Standard Deviation |
|---|---|---|
| 2.2) | 95.9 | 8.72 |
| 2.3) | 39.8 | 6.18 |
| 2.4) | 40.4 | 5.76 |
| 2.5) | 33.7 | 3.61 |

Note that 95.9/2 = 48, 39.8 and 40.4 are much greater than 33.7. Thus it seems that the range, the mid-value, the mean and the median may not affect the determination of membership values.

2.6)The answers of this question may be summarized as follow:

1)To most of the subjects, "young" ≠ "not old" and "old" ≠ "not young".

2)One of the subjects classifies a people with age 5 as a "not young people", quite different from our intuitive sense. The reason of the subject is that he does not think that a baby is a young people. Thus, to some people, the membership function of this set may not be nondecreasing when we consider the whole population.

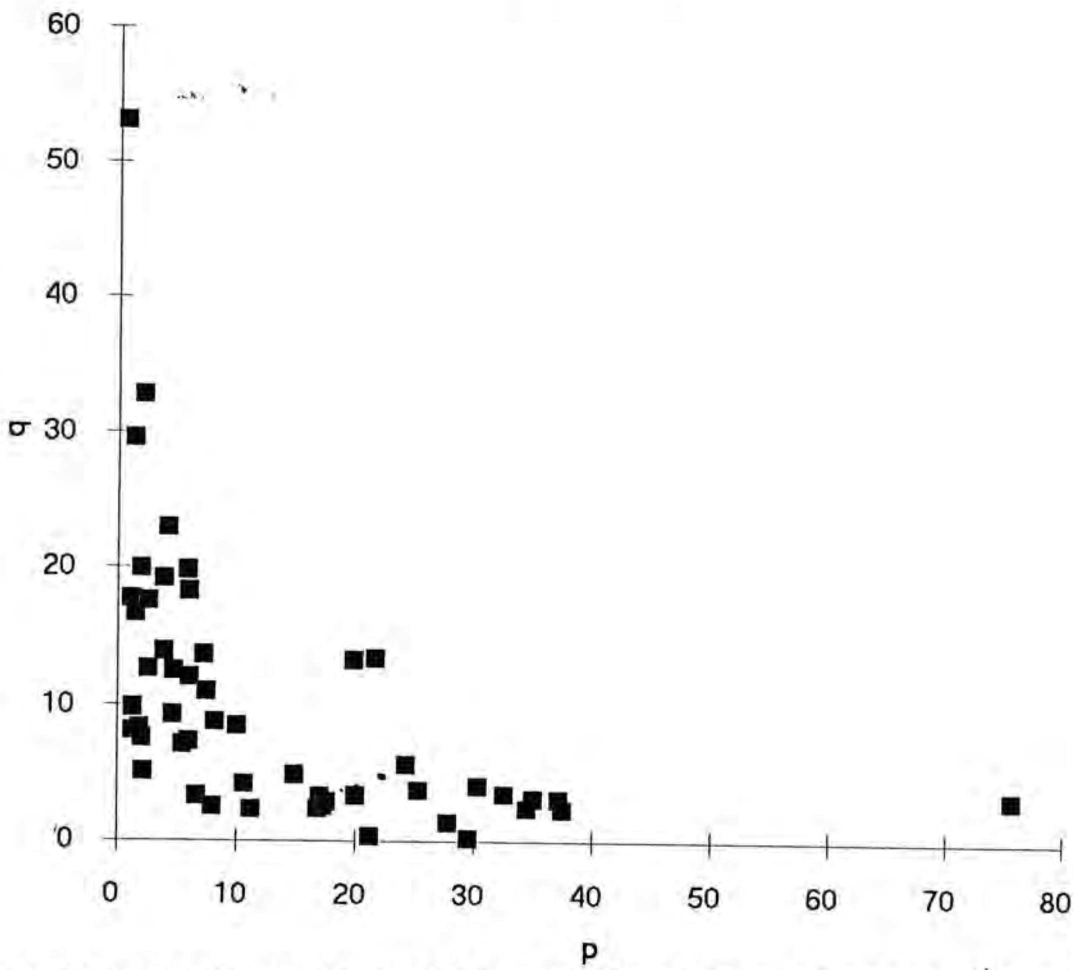
3)The other result resembles that of part 1 and will not be summarized here.

### 4.5.3. An Approach to Reduce the 'bias' in Polling.

From our survey, we found that in most cases polling will overestimate the membership values when they are larger than 0.5 and underestimate the membership values when they are smaller than 0.5 when compared with direct rating. In section 4.3.1., we have tried to explain this phenomenon under some assumptions on the assignment of the membership values. In this section, we would provide a method to reduce the "bias", i.e. what we can do if we have estimates of membership values by polling method and we want to find out the corresponding estimates by direct rating method.

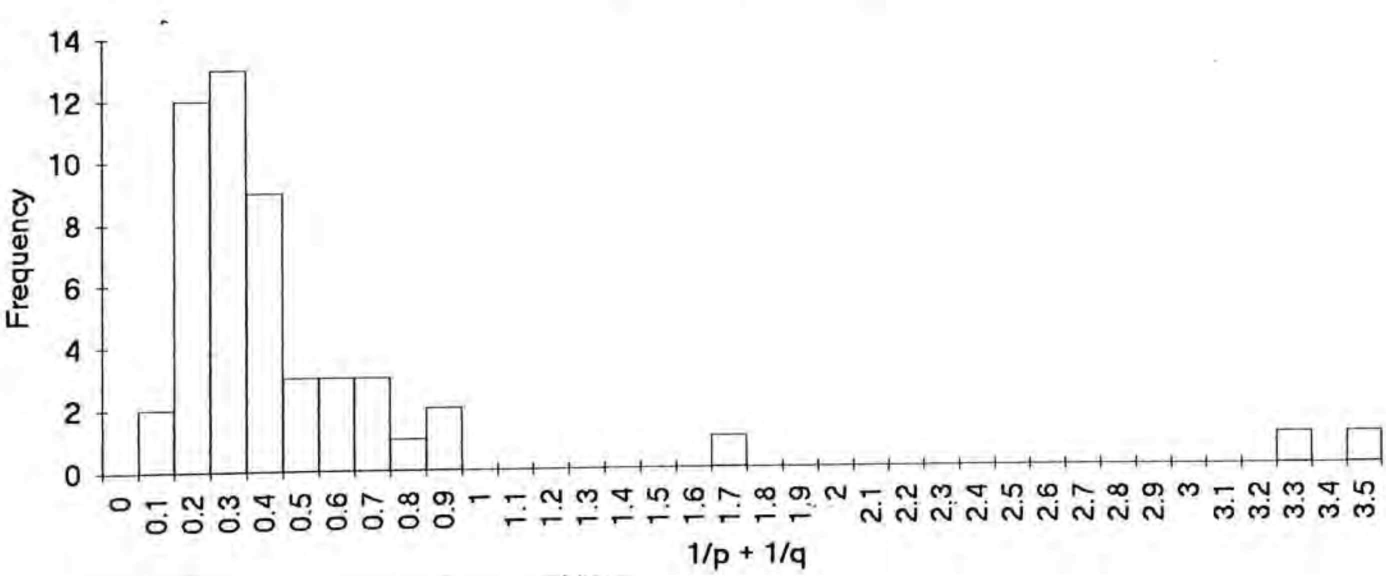
In Turksen (1991) and Leung (1981), the membership value of each objects is assumed to be a Beta random variable. It is reasonable as the range of a Beta random variable is [0,1]. Assume that the distribution of the membership value is a Beta distribution with two parameters p and q. As we have two parameters to be determined and we have only one equation about p and q, we need to find one more equation for p and q. For the direct rating data in part 1 of our survey, we found the mean and the variance of each membership values. Note that if z is a Beta [p,q] random variable, then $E(z) = \frac{p}{p+q}$ and $Var(z) = \frac{pq}{(p+q)^2(p+q+1)}$. From question 3 to question 7, we have totally fifty-one observations of sample mean and sample variance. For each observation, we treat the sample values as their corresponding population values and find the corresponding p and q by solving the above two equations and plot these (p,q)'s on the graph as below:

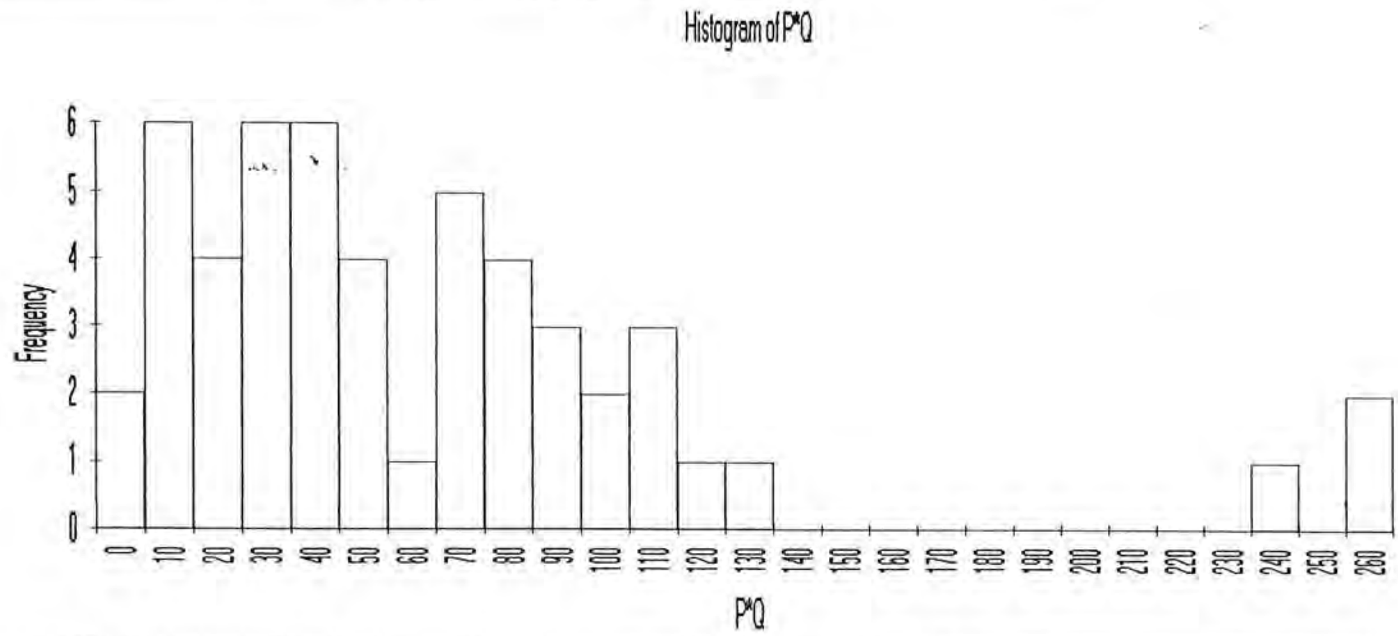
The Graph of q vs p




From the graph, we know that usually when p is large, then q is small and vice versa, Thus we suspect that the relation between p and q can be formalized as either pq is a constant or $\frac{1}{p} + \frac{1}{q}$ is a constant. From the histograms of these pq and $\frac{1}{p} + \frac{1}{q}$, the latter one appears to be more plausible.

Histogram of 1/p + 1/q






Assume $\frac{1}{p} + \frac{1}{q} = c$ say. We estimate c by the mean of the fifty-one $\frac{1}{p} + \frac{1}{q}$. It gives c = 0.5.

We have assumed that $\mu_{A,1}(x) = \Pr(\mathrm{Beta}(p,q) \geq 0.5)$, where $\mu_{A,1}(x)$ is the membership value obtained by polling method. Therefore for any given polling value, we can find (p,q) such that $\frac{1}{p} + \frac{1}{q} = 0.5$ and $\mu_{A,1}(x) = \Pr(\mathrm{Beta}(p,q) \geq 0.5)$. Then the improved estimate of the membership value will be $\frac{p}{(p + q)}$. For example, with $\mu_{A,1}(x) = 0.615$, we can find out that we need p = 4.434, q = 3.643 and so the improved estimate will be $\frac{p}{(p + q)} = 0.549$. From the result in part 1 in our survey, we have the following estimated membership values by polling, $\mu_{A,1}(x)$, 0.923, 0.769, 0.615, 0.385, 0.308 and 0.154. We have discarded the membership values of zero and one as this approach fails in this cases. The following is the summary of this approach :

**Table 13. Summarized results of improved estimate for 1/p + 1/q = 0.5**

| Direct Rating | Polling | Improved estimate |
|---|---|---|
| 0.831<br>0.812<br>0.759<br>0.753 | 0.923 | 0.714 |
| 0.717 | 0.769 | 0.620 |
| 0.671<br>0.617<br>0.599 | 0.615 | 0.549 |
| 0.542 | 0.385 | 0.451 |
| 0.493 | 0.308 | 0.417 |
| 0.450<br>0.435 | 0.154 | 0.339 |

The first column is the estimated membership values obtained by direct rating. The second column is the estimated membership values obtained by polling while the last column is the estimate of the membership value from the above approach. You may note that the above result may not be so good but it may be due to the small number of subjects involved in the survey. With the above result, we have done the following in order to check whether this poor performance is due to this approach or due to the small number of subjects:

Suppose for estimated membership value by direct rating, $\mu_{A,2}(x)$ = 0.450 and the corresponding estimated membership value by polling, $\mu_{A,1}(x)$ = 0.154. We have the improved estimated membership value, $\mu_{A,3}(x) = 0.339$. We will assume $\frac{1}{p} + \frac{1}{q} = 0.5$ in the following.

For $\frac{p}{p+q} = 0.45$ and $\frac{1}{p} + \frac{1}{q} = 0.5$, we find that p = 3.64 and q =4.44. Let v = Pr [Beta(3.64, 4.44) $\geq$ 0.5] = 0.3822. Then $1 - v^{13} - (1 - v)^{13}$ =

0.998086 ≡ D. Let Y be the number of subjects classified the line as a long line. Excluding the cases when Y = 0 or 13, we find the distribution of Y as follow:

**Table 14. Distribution of p'**

| Y | corresponding polling value, v2 | corresponding p' such that Pr[Beta(p',q' ≥ 0.5)] = v2. | Corresponding probability |
|---|---|---|---|
| 1 | Y ÷ 13 = 0.07692 | 2.801 | $\frac{{}_{13}C_1 v^1 (1 - v)^{12}}{D}$ = 0.015390 |
| 2 | Y ÷ 13 = 0.15385 | 3.025 | $\frac{{}_{13}C_2 v^2 (1 - v)^{11}}{D}$ = 0.057126 |
| 3 | 0.23077 | 3.228 | 0.129583 |
| 4 | 0.30769 | 3.430 | 0.200416 |
| 5 | 0.38462 | 3.643 | 0.223176 |
| 6 | 0.46154 | 3.875 | 0.184089 |
| 7 | 0.53846 | 4.134 | 0.113886 |
| 8 | 0.61538 | 4.434 | 0.052842 |
| 9 | 0.69231 | 4.796 | 0.018161 |
| 10 | 0.76923 | 5.257 | 0.004494 |
| 11 | 0.84615 | 5.901 | 0.000758 |
| 12 | 0.92308 | 6.993 | 0.000078 |

From the above table, we can find the expected value of p', E(p') = 3.6690, $E(\frac{1}{p'})$ = 0.27597 and $E(\mu) = 1 - 2E(\frac{1}{p'}) = 0.44806$. Moreover we can find $Pr(\mu \le E(\mu)) = 0.403$ ( $= Pr\left( p' \le \frac{1}{E(\frac{1}{p'})} \right) = Pr( p' \le 3.624) =$ 0.01539 + 0.05713 + 0.12958 + 0.20042 ). We will summarize the result as follow:

**Table 15. Characteristics of $\mu$ for $\frac{1}{p} + \frac{1}{q} = 0.5$**

| Direct | Polling | $\frac{p}{(p+q)}$ | $E(\mu)$ | $Bias(\mu)$ | $Pr[\mu \le E(\mu)]$ |
|---|---|---|---|---|---|
| 0.831 | 0.923 | 0.714 | 0.713 | -0.118 | 0.020 |
| 0.812 | | | 0.711 | -0.101 | 0.044 |
| 0.759 | | | 0.703 | -0.056 | 0.191 |
| 0.753 | | | 0.701 | -0.052 | 0.216 |
| 0.717 | 0.769 | 0.620 | 0.688 | -0.029 | 0.392 |
| 0.671 | 0.615 | 0.549 | 0.662 | -0.009 | 0.645 |
| 0.617 | | | 0.618 | 0.001 | 0.379 |
| 0.599 | | | 0.601 | 0.002 | 0.501 |
| 0.542 | 0.385 | 0.451 | 0.544 | 0.002 | 0.428 |
| 0.493 | 0.308 | 0.417 | 0.493 | 0.000 | 0.546 |
| 0.450 | 0.153 | 0.339 | 0.448 | -0.002 | 0.403 |
| 0.435 | | | 0.433 | -0.002 | 0.505 |

Thus we found that the poor performance of this approach may be due to the small number of subjects involved.

#### 4.5.4. Advice to Researchers.

In the past, few researchers had conducted surveys or experiments to estimate the membership values of certain fuzzy set. In this section, we would like to point out points that are important in the survey.

Before the survey

1) The researchers must decide the fuzzy set according to their interests. For example, when a researcher is interested in a fuzzy set about classification of lines, he must decide whether the fuzzy set is the set of "long line" or the set of "short line" or something else.
2) After the researchers have decided the fuzzy set, he must define the universe of discourse clearly so that every subject understands what the set is. For example, in considering the set of "young man", it is better for the researchers to clarify what this set refer to. Does this set include only male but not female? Does this set include people from all country, or just include people in one country? The researchers must define the universe of discourse clearly so that there will be no ambiguity.
3) Whenever a researcher uses questionnaire or shows some figures to the subjects, it is better for the researcher to introduce elementary fuzzy set theory to the subjects. Moreover, the wordings need to be careful. The subjects may find it difficult to understand the researcher's instructions in these kind of surveys.

4) The researchers have to choose their subjects carefully. The choice of the subjects highly depends on the aim of survey. If the researchers want to estimate the membership values of certain fuzzy set of general public, it is better for them to choose people with different background. It is reasonable to believe that the membership values of fuzzy set are nearly the same if their ages, educational level and life styles are the same.

5) The researchers have to choose the fuzzy set carefully. In some fuzzy set, the attribute is bounded but it may not be so for other fuzzy set. The attribute is bounded or not is important as it is an assumption for certain interpretation of membership values (see Turksen (1991) and Norwich and Turksen (1984)).

During the survey

1) The duration of the survey must not be too long. It must be ended before the subjects feel boring.

2) In the case of questionnaire, it is better for the researchers to be beside the subjects so that they can answer the subject's questions.

Fuzzy set theory is a new theory, that has an history of only thirty years. Thus it is reasonable that many parts of this theory are not well-developed yet. Many parts raise the interest of researchers and so many papers are published each year on fuzzy set theory. In this thesis, we have tried to discuss the interpretation and estimation of the membership function. We want to view the aspects of the membership function in a statistical sense, an approach that is seldomly found in current literatures on fuzzy set theory.

For the interpretation of the membership function, we found that many interpretations exist. All of them are reasonable but there is not an interpretation that is accepted by all researchers. Each of the interpretation is suitable in some circumstances but not in all. However, we found that many interpretations are related to the interpretation of probability theory. It is reasonable as both theory are used in case of impreciseness. Perhaps we can foresee the development of fuzzy set theory according to that of probability theory in the past. The operations on probability theory may be used as a guide on the operations on fuzzy set theory. For example, when we consider the union or intersection of two fuzzy events, perhaps we need to know not only the "individual distribution" of each of them but also their "joint distribution". Here, we have no idea on the determination of "joint distribution" of two fuzzy events; perhaps we need to take care of the relation of the two fuzzy events, but we believe that using "joint distribution" of two fuzzy events is more reasonable to using just maximum and minimum in case of union and intersection of two fuzzy events. For the interpretation of the membership function, perhaps in the future it will be divided into many classes, just like the phenomenon that the interpretation of probability theory is divided into many classes.

For the estimation of the membership function, we found that there is no estimation procedure that is universally accepted by all researchers. It is reasonable as estimation procedure is closely related to the interpretation. We cannot expect to have an universally accepted estimation procedure when there is no universally accepted interpretation.

Among the currently used estimation procedures, direct rating is the most frequently used one. It is reasonable as we believe that the membership function is subjective in nature and so the most intuitive way to estimate it is to ask the subject to tell us the membership value. Polling is the second most frequently used one due to its ease of computation. Past researchers have already noted that the result of these two procedures are different but they had not given any reason. We have given one justification of the difference under certain reasonable assumptions. Set-valued statistics and reverse rating are the other two frequently used estimation procedures. We have connected polling and set-valued statistics and so we expect that the result of set-valued statistics will be different from that of direct rating, similar to that of polling. For the reverse rating, most researchers think that it can be used as a check of the result only. However, we think that the information from reverse rating is more than that from set-valued statistics and so it can be used as an estimation procedure by itself. We believe that each of the four methods obtains different information about the membership function. So we need to know what information does each method get before we do the estimation.

For the survey, we think that the most interesting result is that subjects do not consider the length and the age only when discussing the membership function of the set "long line" and "young people". This phenomenon increases the difficulties in the estimation of the membership function. Moreover the result that "young" ≠ "not old" and "not young" ≠ "old" also advices us to choose the fuzzy set clearly in a survey or experiment.

At last we want to conclude that the interpretation and estimation procedure of the membership function are not well-developed. We need to develop the interpretation first. We have not done enough on this part in this thesis, but we believe that we can borrow the idea from probability theory in the future.

**References**

Bobrowicz, O., Choulet, C., Haurat, A., Sandz, F., Tebba, M., A Method to Build Membership Functions: applications to numerical/symbolic interface building.

Bordley, R.F. (1989), Fuzzy Set Theory, Observer Bias and Probability Theory. *Fuzzy Set and System*, **33**, 347-354.

Cai, K.Y. (1993), Parameter Estimations of Normal Fuzzy Variables. *Fuzzy Sets and Systems*, **55**, 179-185.

Chu, A.T.W., Kalaba, R.E., and Spingarn, K. (1979), A Comparison of Two Methods for Determining the Weights of Belonging to Fuzzy Sets, *Journal of Optimization Theory and Applications*, **27**, 4, 531-538.

Civanlar, M.R., Trussell, H.J. (1986), Constructing Membership Function Using Statistical Data. *Fuzzy Set and System*, **18**, 1-13.

Dombi, J. (1990), Membership Function as an Evaluation. *Fuzzy Set and System*, **35**, 1-21.

French, S. (1984), Fuzzy Decision Analysis : some Criticisms. *TIMS/studies in the Management Sciences*, **20**, 29-44.

Giles, R. (1982), Semantics for Fuzzy Reasoning. *Int. J. Man.-Machine Studies*, **17**, 401-415.

Giles, R. (1988), The Concept of Grade of Membership. *Fuzzy Set and System*, **25**, 297-323.

Hisdal, E. (1986a), Infinite-valued Logic based on Two-valued Logic and Probability. Part 1.1. Difficulties with Present-day Fuzzy-set Theory and their Resolution in the TEE Model. *Int. J. Man-Machine Studies*, **25**, 89-111.

Hisdal, E. (1986b), Infinite-valued Logic based on Two-valued Logic and Probability. Part 1.2. Different Sources of Fuzziness. *Int. J. Man-Machine Studies*, **25**, 113-138.

Hisdal, E. (1988a), Are Grades of Membership Probabilities? *Fuzzy Set and System*, **25**, 325-348.

Hisdal, E. (1988b), The Philosophical Issues Raised by Fuzzy Set Theory. *Fuzzy Set and System*, **25**, 349-356.

Klir, G. J. (1985), Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the Like? *Abstracts of the First I.F.S.A. Congress*, University of Palma de Mallorca, Spain. Vol. 1, Parallel Sessions 1.

Krantz, D.H., Luce, R.D., Suppes, P. and Tversky, A. (1971), Foundations of Measurement, Vol. 1 (Academic Press, New York).

Leung, Y. (1981), On tha Exactness of Membership Functions in Fuzzy Set Theory. *Proceedings of the International Conference on Policy analysis and Information Systems*, Taiwan: Tamkang University, 765-775.

Mabuchi, S. (1992), An Interpretation of Membership Functions and the Properties of General Probabilistic Operators as Fuzzy Set Operators - Part I: Case of Type I Fuzzy Sets. *Fuzzy Set and System*, **49**, 271-283.

Natvig, B. (1983), Possibility versus Probability. *Fuzzy Set and System*, **10**, 31-36.

Norwich, A.M., Turksen, I.B. (1982), The Construction of Membership Functions. in: Yager, R.R., Ed., *Fuzzy Set and Possibility Theory*,

(Pergamon Press, New York.)

Norwich, A.M., Turksen, I.B. (1984), A Model for the Measurement of Membership and the Consequence of its Empirical Implementation. *Fuzzy Set and System*, **12**, 1-25.

Orlovski, S.A. (1990), Decomposition of Properties and Definition of Fuzzy Set. *Fuzzy Set and System*, **37**, 53-63.

Osherson, D.N., Smith, E.E. (1981), On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition*, **9**, 35-58.

Osherson, D.N., Smith, E.E. (1982), Gradedness and Conceptual Combination. *Cognition*, **12**, 299-318.

Saaty, T.L. (1974), Measuring the Fuzziness of Sets. *Journal of Cybernetics*, **4**, 4, 53-61.

Saaty, T.L. (1986), Scaling the Membership Function. *European Journal of Operational Research*, **25**, 320-329.

Sastry, V.N. (1993), Spline Membership Function and its Application in Multiple Objective Fuzzy Control Problem. *Fuzzy Sets and Systems*, **55**, 143-156.

Smithson, M. (1986), Fuzzy Set Analysis for Behavioral and Social Sciences. (Springer-Verlag, New York.)

Thole, U., Zimmermann, H.J., Zysno, P. (1979), On the Suitability of Minimum and Product Operators for the Intersection of Fuzzy Sets. *Fuzzy Set and System*, **2**, 167-180.

Turksen, I.B. (1991), Measurement of Membership Functions and Their Acquisition. *Fuzzy Set and System*, **40**, 5-38.

Wallsten, T.S., Budescu, D.V., Rapoport, A., Zwick R. and Forsyth B.

(1986), Measuring the Vague Meanings of Probability Terms. *Journal of Experimental Psychology: General*, **115**, 4, 348-365.

Zadeh, L.A. (1965), Fuzzy Sets. *Information and Control*, **8**, 338-353.

Zadeh, L.A. (1978), Fuzzy Sets as an Basis for a Theory of Possibility. *Fuzzy Set and System*, **1**, 3-28.

Zhang, N.L. (1981), The Subordination and Probability Feature of Random Phenomena. *Academic Journal of Wuhan Building Material College 2.*

APPENDIX

在古典集合理論中，一元素只可以屬於或不屬於一集合中。但是在日常生活中有很多現象都是古典集合理論所不能解釋，於是便有模糊集合理論的產生。在這理論中，一元素是可以部份地屬於一集合中，而屬於的程度便是由隸屬函數所代表。通常如果一元素的隸屬函數的值是零的話，那便代表它不屬於這集合；如果一元素的隸屬函數的值是一的話，那便代表它完全屬於這集合；如果一元素的隸屬函數的值是零至一的任何數字的話，那數字便代表這元素屬於這集合的程度。

這份問卷是希望探索當一個人遇到一些模糊集合（例如：長線，短線，年輕的人，年老的人等）時，他是會如何決定這些集合的隸屬函數。

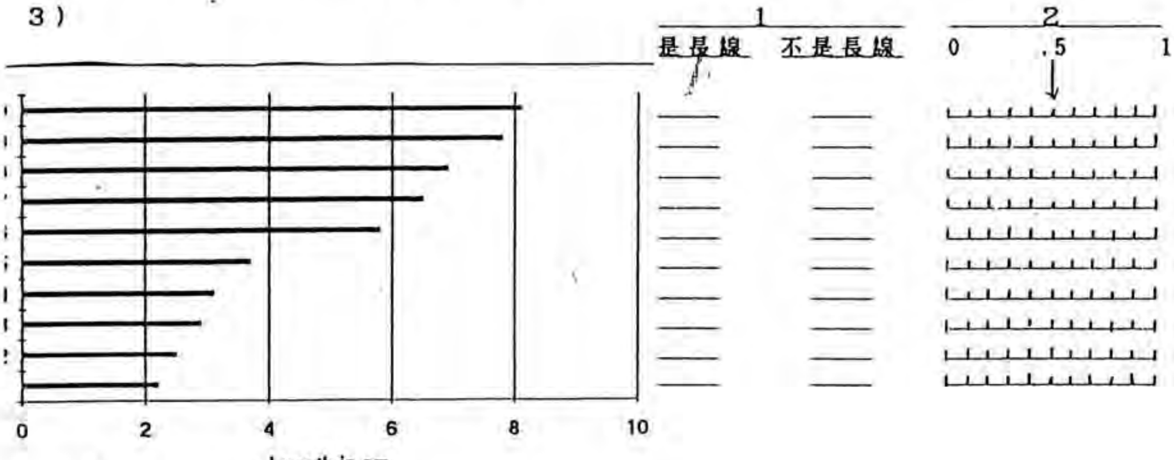
第一部份：這部份是希望探索在一組給定直線中，我們是如何決定這些線對於「長線」這模糊集合的隸屬程度。所有給定之線的長度是在0cm至10cm，而且我們規定一條線只可以被評為一條長線或一條不長的線。

1）如果你要決定一條線是不是一條長線的話，除了它的長度外，你會不會考慮其它因素？（請以✓表示你的選擇）
　會 ____ （請列出）________________
不會 ____

2）請問你認為最少多長的線才算是一條長線而不是一條不長的線？ ________ cm
（請記著現在線的長度是由0cm至10cm）

（3）－（6）題都會列出十條線，請你
1）以✓表示你認為他們那條在這十條線當中算是一條長線，那條不算是一條長線。
2）在右面適當位置上劃下一個x符號以表示你認為這一條線在所列十條線中是一條長線的程度。
0表示你認為這一條線絕對不是一條長線，
1表示你認為這一條線絕對是一條長線，
0－1之間的位置表示這一條線是一條長線的程度。

3）



(3) - (6) 題都會列出十條線，請你

1) 以 ✓ 表示你認為他們那條在這十條線當中算是一條長線，那條不算是一條長線。

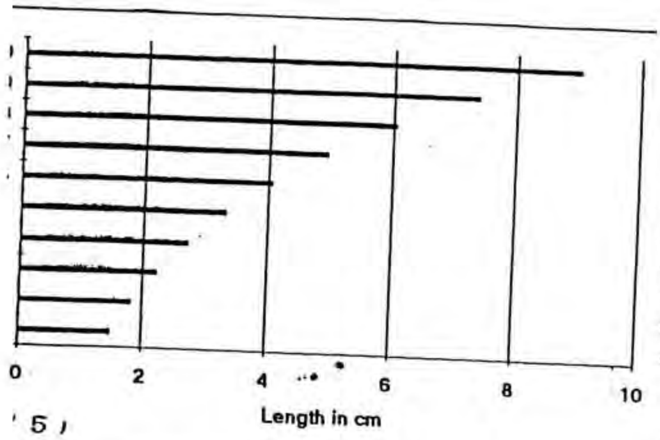
2) 在右面適當位置上劃下一個 x 符號以表示你認為這一條線在所列十條線中是一條長線的程度。

0 表示你認為這一條線絕對不是一條長線，

1 表示你認為這一條線絕對是一條長線，

0 - 1 之間的位置表示這一條線是一條長線的程度。

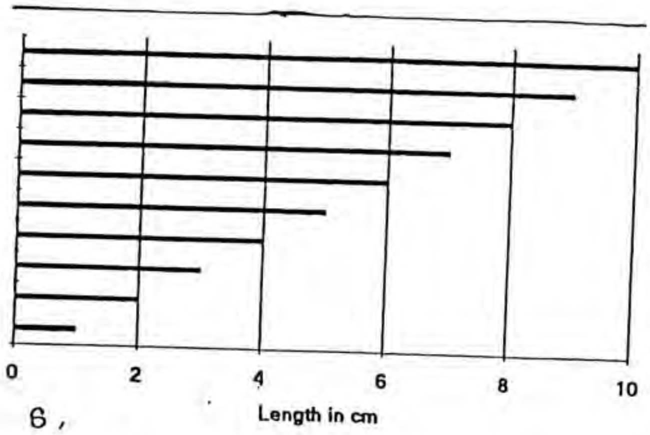
4)




1: 是長線　不是長線

2: 0　.5　1

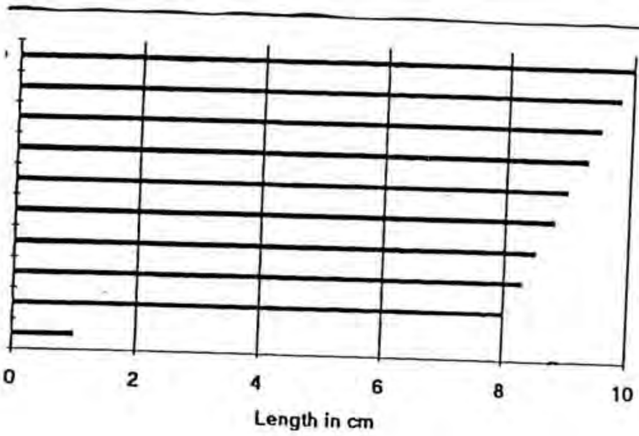
5)




1: 是長線　不是長線

2: 0　.5　1

6)




1: 是長線　不是長線

2: 0　.5　1

7）在下列三十條線當中，請你在指定的十一條線（1,4,7,10,13,16,19,22,25,28,30）

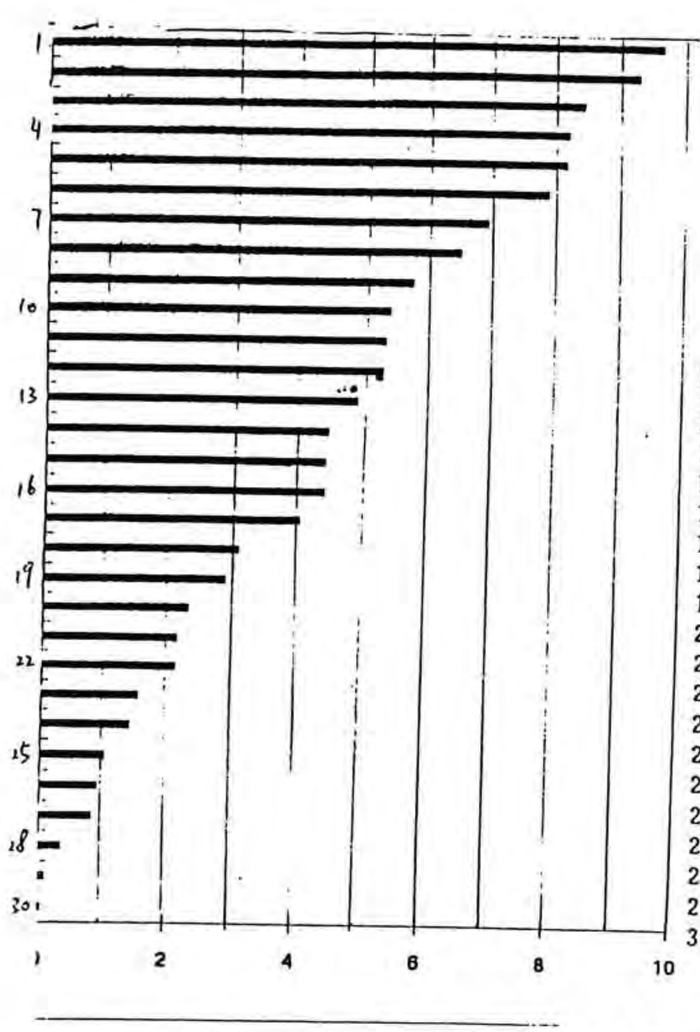
1）以 ✓ 表示你認為他們那條在這三十條線當中算是一條長線，那條不算是一條長線。

2）在右面適當位置上劃下一個 x 符號以表示你認為這一條線在所列三十條線中是一條長線的程度。

0 表示你認為這一條線絕對不是一條長線，

1 表示你認為這一條線絕對是一條長線，

0－1之間的位置表示這一條線是一條長線的程度。




| | 1 | | 2 |
|---|---|---|---|
| | 是長線 | 不是長線 | 0 .5 1 |
| 1 | ____ | ____ | |
| 2 | | | |
| 3 | | | |
| 4 | ____ | ____ | |
| 5 | | | |
| 6 | | | |
| 7 | ____ | ____ | |
| 8 | | | |
| 9 | | | |
| 10 | ____ | ____ | |
| 11 | | | |
| 12 | | | |
| 13 | ____ | ____ | |
| 14 | | | |
| 15 | | | |
| 16 | ____ | ____ | |
| 17 | | | |
| 18 | | | |
| 19 | ____ | ____ | |
| 20 | | | |
| 21 | | | |
| 22 | ____ | ____ | |
| 23 | | | |
| 24 | | | |
| 25 | ____ | ____ | |
| 26 | | | |
| 27 | | | |
| 28 | ____ | ____ | |
| 29 | | | |
| 30 | ____ | ____ | |

—— 第一部份完 ——

第二部份：<u>這部份是希望探索當我們面對只有兩個選擇時（年輕人，非年輕人），我們如何介定一個香港人是一個年輕人抑或是非年輕人。</u>

1）如果你要決定一個香港人是不是一個年輕人，除了他的年齡外，你會不會考慮其他因素？（請以✓表示你的選擇）
會 ____（請列出）________________________
不會 ____

2）請問在你心目中香港人的年齡分佈是由0歲至幾歲？ ____歲

3）請問在你心目中香港人的中位歲數（median age）是幾歲？ ____歲

4）請問在你心目中香港人的平均歲數（mean age）是幾歲？ ____歲

5）請問最低限度多少歲數的人你才會認為他是一個非年輕人而不是一個年輕人？ ____歲

6）下列是十二個香港人的歲數，請你在下面適當的位置上
1）劃下一個x符號以表示你認為這一個人是年輕人的程度。
0表示你認為這一個人絕對不是一個年輕人，
1表示你認為這一個人絕對是一個年輕人，
0－1之間的位置表示這個人是一個年輕人的程度。
2）劃下一個✓以表示你認為這一個人是一個年輕人抑或是非年輕人。
3）劃下一個✓以表示你認為這一個人是一個老年人抑或是非老年人。

| | 1 | 2 | | 3 | |
|---|---|---|---|---|---|
| 年齡 | 0 .5 1 | 年輕人 | 非年輕人 | 老年人 | 非老年人 |
| 5 | | | | | |
| 15 | | | | | |
| 25 | | | | | |
| 35 | | | | | |
| 45 | | | | | |
| 55 | | | | | |
| 65 | | | | | |
| 75 | | | | | |
| 85 | | | | | |
| 95 | | | | | |
| 你心目中的中位歲數 | | | | | |
| 你心目中的平均歲數 | | | | | |

——第二部份完——

第三部份：<u>個人資料</u>

1）姓別：男／女
2）年齡：____歲

——第三部份完——

<u>全卷完 多謝合作</u>