On Testing Hypothesis of Fuzzy Sample Mean

Berlin WU* and Shu-Kwang CHANG†

Department of Mathematical Sciences
  National Chengchi University, Taiwan
  E-mail: berlin@nccu.edu.tw
†*Department of Mathematical Sciences*
  National Chengchi University, Taiwan

In many expositions of fuzzy methods, fuzzy techniques are described as an alternative to a more traditional statistical approach. In this paper, we present a class of fuzzy statistical decision process in which testing hypothesis can be naturally reformulated in terms of interval-valued statistics. We provide the definitions of fuzzy mean, fuzzy distance as well as investigation of their related properties. We also give some empirical examples to illustrate the techniques and to analyze fuzzy data. Empirical studies show that fuzzy hypothesis testing with soft computing for interval data are more realistic and reasonable in the social science research. Finally certain comments are suggested for the further studies. We hope that this reformation will make the corresponding fuzzy techniques more acceptable to researchers whose only experience is in using traditional statistical methods.

Key words: membership function, fuzzy sampling survey, fuzzy mean, human thought

1. Introduction

A statistical test is general conducted by means of a hypothesis testing for which the probability distribution is determined by the assumption that the null hypothesis H_0 is true. Under the significant level α, a critical region is compute such that if the observed statistics falls in the critical reason, we reject the null hypothesis. The statistical test provides information from which we can decide the significance of the increase (or decrease) in any experimental result.

However there are some vague information, formulated by terms from natural language, is not easy to describe in statistical terms. To handle this information and knowledge, it is natural to use intelligent computing techniques. Fortunately, in many expositions of intelligent computing methods, fuzzy techniques are described as an alternative to a more traditional statistical approach. Such a description makes fuzzy techniques difficult to understand and difficult to accept for researchers who are accustomed to statistical methods.

In this paper we will introduce the concept of fuzzy statistics via discrete fuzzy sample and continuous fuzzy sample. The definitions of fuzzy mean for two kinds of fuzzy data are proposed. Using these definitions we are able to set up the fuzzy testing hypothesis such as fuzzy equals to and the fuzzy belongs to. These testing processes are very useful tools for decision make in a fuzzy system.

Although the statistical methods used to testing fuzzy sample mean are based on the traditional decision theory, extended Neyman–Peason's lemma about most

powerful test are not austerely investigated for two reasons: (i) soft-computing for critical region with fuzzy number is still not identified. (ii) distribution for fuzzy population is vague, incomplete or unknown.

Since our main objectives is to promote the understanding of these two classes of techniques—statistics and fuzzy—to researchers who may only know well one of theses techniques, we go into some detail explaining the basics of techniques. A reader who is well familiar with one or both of these techniques is advised to at least browse through or exposition of their basics.

In reality, more complex situations are possible, in which an expert is not 100% sure whether a given estimate x_i is possible; in this case, we can no longer use polling to get numerical characteristics of the expert knowledge. Fuzzy set methodology can handle such more complex situation as well. The above polling method of eliciting the values of the membership functions is only one of the many known elicitation techniques; see, e.g., [1, 3, 4, 6, 8].

In this paper, we present a class of fuzzy statistical decision process in which testing hypothesis can be naturally reformulated in terms of interval-valued statistics. To describe these situations we will start with a brief motivation of traditional statistical techniques, and then give a brief motivation of the corresponding fuzzy methods, and then describes the relation between these two classes of techniques. We provide the definitions of fuzzy mean, fuzzy distance as well as investigation of their related properties. We also give some empirical examples to illustrate the techniques and to analyze fuzzy data. Empirical studies show that fuzzy hypothesis testing with soft computing for interval data are more realistic and reasonable in the social science research. Finally certain comments are suggested for the further studies.

Our result is in good agreement with a general result from [7], [9], according to which an arbitrary fuzzy set can be interpreted in statistical terms: namely, as a random *set*. For the latest developments in this area, see, e.g., [2], [5]. We hope that this reformation will make the corresponding fuzzy techniques more acceptable to researchers whose only experience is in using traditional statical methods.

2. Fuzzy data with soft computing

2.1. Membership function

In the traditional statistical approach, we assume that each expert provides the exact estimate of the desired quantity x; so, after polling the experts, we get, for each possible value x_i of this quantity, the probability $p(x_i)$ that a randomly chosen expert selected this value as his/her estimate.

In the fuzzy case, we take into consideration that expert often cannot provide definite estimates. The simplest case is when an expert has in mind several possible values of the estimated quantity. In this case, instead of asking each expert for a definite estimate, we can ask each experts which of ν values they consider possible and which not. As a result, for each value x_i, we can count the total number $N(x_i)$ of experts who consider this value possible.

Similar to the statistical approach, we do not want to ask all N experts, so we would like to get a representative sample of $n \ll N$ experts, and ask only these n experts. In doing this, we hope that for each x_i the portion $\mu(x_i) = n(x_i)/n$ of experts in this sample who consider x_i to be a possible estimate is approximately the same as the portion $N(x_i)/N$ corresponding to all N experts. Therefore, we must consider the ratios $\mu(x_i) = n(x_i)/n$.

Similarly to probabilities, these numbers belong to the interval $[0, 1]$. However, we can no longer call these values probabilities, because they do not necessarily add up to 1. For example, if every expert considers all the values possible, then $\mu(x_i) = 1$ for all i, so the sum of these values is which no longer equal to 0. The values $\mu(x_i)$ are called *degrees of possibility*, and the function μ which maps each value x_i into the corresponding degree $\mu(x_i)$ is called a *membership function*, or a *fuzzy set*.

In considering the question related with fuzzy property, we consider the information itself has the uncertainty and fuzzy property. Hence, let's firstly give an easy and precise explanation about fuzzy numbers.

DEFINITION 2.1 (Fuzzy Number (Nguyen and Wu 2006 [5], p. 111)). *Let U be an universal set, $A = \{A_1, A_2, \ldots, A_n\}$ be the subset of discussion factors in U. For any term or statement X on U, it's membership corresponding to $\{A_1, A_2, \ldots, A_n\}$ is $\{\mu_1(X), \mu_2(X), \ldots, \mu_n(X)\}$, here $\mu: U \to [0, 1]$ is a real function. Then the fuzzy number of X can be written as the following:*

$$\mu_U(X) = \sum_{i=1}^{n} \mu_i(X) I_{A_i}(X), \qquad (2.1)$$

where $I_{A_i}(x) = 1$, if $x \in A_i$; $I_{A_i}(x) = 0$, if $x \notin A_i$.

If the domain of the universal set is continuous, then the fuzzy number can be written as: $\mu_U(X) = \int_{A_i \subseteq A} \mu_i(X) I_{A_i}(X)$.

In the research of social science, the sampling survey is always used to evaluate and understand public opinion on certain issues. The traditional survey forces people to choose one answer from the survey, but it ignores the uncertainty of human thinking. For instance, when people need to choose the answer from the survey which lists five choices including "Very satisfactory," "Satisfactory," "Normal," "Unsatisfactory," "Very unsatisfactory," traditional survey become quite exclusive.

The advantages of evaluation with fuzzy number include: (i) Evaluation process becomes robust and consistent by reducing the degree of subjectivity of the evaluator. (ii) Self-potentiality is highlighted by indicating individual distinctions. (iii) Provide the evaluators with an encouraging, stimulating, self-reliant guide that emphasizes on individual characteristics. While the drawback is that the calculating process will be a little complex than the conventional one.

2.2. Fuzzy mean

In the traditional statistical approach, we start with a collection of real numbers, i.e., in more precise terms, we used *number-valued* statistics. In an interval

situation, we start with a collection of intervals instead of a collection of numbers. So, if we use statistical methods to process this collection, it is natural to call these statistical methods *interval-valued*.

Let us see what statistical characteristics we can naturally extract from this collection. In traditional statistical techniques, each expert presents a single number, and from this collection, we can extract the probabilities $p(x_i)$ and cumulative probabilities $F(x_i)$. In the interval case, each expert presents *two* numbers x^- and x^+. So instead of a single collection of numbers, we have two collections: a collection of the lower endpoints x^-, and the collection of upper endpoints x^+. It is therefore natural to apply the standard statistical procedure to each of these collections.

Fuzzy techniques were developed to describe and analyze the situation when an expert is not sure about the value of the estimated quantity x, and may consider several different values to be possible.

We start with the set $X = \{x_1, \ldots, x_v\}$ of all possible values of x. In principle, the values which an expert considers possible can form an arbitrary subset of this set. Usually, however, not all subset occur. Typically, if an expert believes that two values $x' < x''$ are possible, than all intermediate values x (i.e., values for which $x' < x < x''$) are possible as well. In such a situation, to describe the set of all the values x which an expert consider possible, it is sufficient to describe the smallest x^- and the largest x^+ of these values; then all the values between x^- and x^+ are possible as well. In other words, in such situation, the set of all values x_i which an expert considers possible forms an *interval* $[x^-, x^+] = \{x^- \leq x \leq x^+\}$.

Therefore, all we have to collect from the experts is these interval, i.e., to be more precise, their endpoints. As a result, we have a collection of intervals.

DEFINITION 2.2 (Fuzzy mean of data with interval values). *Let U be the universe set, and $\{FS_i = [a_i, b_i], \ a_i, b_i \in \mathbf{R}, \ i = 1, 2, \ldots, n\}$ be a sequence of random fuzzy sample on U. Then the fuzzy expected value is defined as $\overline{FX} = \left[\frac{1}{n}\sum_{i=1}^{n} a_i, \frac{1}{n}\sum_{i=1}^{n} b_i\right].$*

DEFINITION 2.3 (Fuzzy mean for unbounded sample). *Let U be the universe set, and $\{FS_i = [a_i, \infty), \ a_i \in \mathbf{R}, \ i = 1, 2, \ldots, n\}$ be a sequence of random fuzzy sample on U. Then the fuzzy expected value is defined as $\overline{FX} = \left[\frac{a_1 + a_2 + \cdots + a_n}{n}, 8\right)$. If $X_1 = (8, b_1], \ldots, X_n = (8, b_n]$, then the fuzzy expected value is defined as $\overline{FX} = \left(8, \frac{b_1 + b_2 + \cdots + a_n}{n}\right].$*

Example 2.1. In a survey with the starting salary for the new undergraduate students' salary, we find the following 5 data as follows: $[1000, 2000]$, $[2000, 2500]$, $[3000, 4000]$, $[1500, 2000]$, $[1000, 1500]$. Then according to the Definition 2.2 the fuzzy mean becomes:

$$\overline{FX} = \left[\frac{1 + 2 + 3 + 1.5 + 1}{5}, \frac{2 + 2.5 + 4 + 2 + 1.5}{5}\right] = [1.7, 2.4] \quad \text{(unit: thousand)}.$$

DEFINITION 2.4 (Defuzzyfication for discrete fuzzy data). *Let D is a fuzzy sample on universe domain U with ordered linguistic variable $\{L_i; i = 1, \ldots, k\}$. $\mu_D(L_i) = m_i$ is the membership with respect to L_i, $\sum_{i=1}^{k} \mu_D(L_i) = 1$. We say $D_f = \sum_{i=1}^{k} m_i L_i$ is the defuzzyfication value for discrete fuzzy data D.*

DEFINITION 2.5 (Defuzzyfication for interval fuzzy data). *Let C is a fuzzy sample on universe domain U with support on $[a, b]$, that is, the membership of $\mu_C(x) = f(x)$, if $x \in [a, b]$, $0 \leq f(x) \leq 1$ is a convex function. The $C_f = \frac{\int_a^b x f(x)\,dx}{\int_a^b f(x)\,dx}$ is called the defuzzyfication value for interval fuzzy data.*

3. Some properties and soft computing about fuzzy data

It is a new research topic about the hypothesis testing of fuzzy mean with interval values. First of all, we will give a brief definition about the defuzzification. Then under the fuzzy significant level δ, we make a one side or two side testing. These methods are a little different from traditional significant level α. In order to get the robustic characteristic properties, we will set up the rejection area level F_δ, according to the fuzzy population.

3.1. Fuzzy equal and fuzzy belongs to for fuzzy data

DEFINITION 3.1 (Fuzzy equal for discrete data). *Let U be a universe domain, $L = \{L_1, L_2, \ldots, L_k\}$ be sequence of rank ordering of linguistic variables on U. $\left\{X_i = \frac{m_{1l}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1, 2\right\}$, $\sum_{j=1}^{k} m_{ij} = 1$ are two random samples from U. If $m_{1j} = m_{2j}$ $(j = 1, 2, \ldots, k)$. Then we say that X_1 fuzzy equals to X_2, denoted by $X_1 \approx_F X_2$.*

DEFINITION 3.2 (Fuzzy index equal for discrete data). *Let U be a universe domain, $L = \{L_1, L_2, \ldots, L_k\}$ be sequence of rank ordering of linguistic variables on U. $\left\{X_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1, 2\right\}$, $\sum_{j=1}^{k} m_{ij} = 1$ are two random samples from U. Then the center of fuzzy number for discrete type is $CX_i = \sum_{j=1}^{k} m_{ij} L_j$. If $CX_1 = CX_2$, we say that X_1 fuzzy index equals to X_2, denoted by $X_1 \approx_I X_2$.*

DEFINITION 3.3 (Fuzzy equal for interval data). *Let A, B be two fuzzy data with membership functions $\mu_A \colon [a, b] \to [0, 1]$, $\mu_B \colon [c, d] \to [0, 1]$. If A, B have the same support and f, g are all convex functions then we say A is fuzzy equal to B, written as $A =_F B$, or briefly $A =_F [c, d]$.*
For left unbounded or right unbounded, the definitions are similar.

DEFINITION 3.4 (Fuzzy belongs to for interval data). *Let A, B be two fuzzy data with membership functions $\mu_A \colon [a, b] \to [0, 1]$, $\mu_B \colon [c, d] \to [0, 1]$. If the support of A is contained in the support B, f, g are all convex functions, then we say A is fuzzy belongs to B, written as $A \in_F B$, or briefly $A \in_F [c, d]$.*
For left unbounded or right unbounded, the definitions are similar.

In order to set up an appropriate testing hypothesis on the fuzzy data, it is necessary for us to give definitions about measurement of distance of fuzzy set. In the following, we set up firstly, the definition of fuzzy distance with fuzzy interval data. The definition is different from the traditional interval operations. Our consideration is concentrated on the statistical point of view.

DEFINITION 3.5 (Distance of fuzzy interval set). *Let A, B be two fuzzy data with membership functions $\mu_A \colon [a,b] \to [0,1]$, $\mu_B \colon [c,d] \to [0,1]$. We give three definitions of distance:*

$$d_1(A,B) = \inf\{|x-y| \colon x \in A,\ y \in B\},$$

$$d_2(A,B) = \sup\{|x-y| \colon x \in A,\ y \in B\},$$

$$d_3(A,B) = \inf\{\varepsilon_1, \varepsilon_2\},$$

$$\varepsilon_1 = \inf\{\varepsilon \colon [c,d] \subset [a-\varepsilon, b+\varepsilon]\}, \quad \varepsilon_2 = \inf\{\varepsilon \colon [a,b] \subset [c-\varepsilon, d+\varepsilon]\},$$

$$d_4(A,B) = \sup\{\varepsilon_1, \varepsilon_2\}.$$

Example 3.1.   Let A, B be two fuzzy data with support $[1,3]$, $[2,5]$. Then

$$d_1(A,B) = \inf\{|x-y| \colon x \in A,\ y \in B\} = 0,$$

$$d_2(A,B) = \sup\{|x-y| \colon x \in A,\ y \in B\} = 4,$$

$$\varepsilon_1 = \inf\{\varepsilon \colon [c,d] \subset [a-\varepsilon, b+\varepsilon]\} = 2, \quad \varepsilon_2 = \inf\{\varepsilon \colon [a,b] \subset [c-\varepsilon, d+\varepsilon]\} = 1,$$

$$d_3(A,B) = \inf\{\varepsilon_1, \varepsilon_2\} = 1,$$

$$d_4(A,B) = \sup\{\varepsilon_1, \varepsilon_2\} = 2.$$

3.2.   Some properties about Fuzzy data

PROPERTY 3.1.   *Let A, B be two fuzzy data with membership functions $\mu_A \colon [a,b] \to [0,1]$, $\mu_B \colon [c,d] \to [0,1]$. The fuzzy equals implies fuzzy belongs to. The inverse is not true.*

Proof.   If $A =_F [c,d]$, since $[a,b] = [c,d]$, then $[a,b] \subset [c,d]$. Hence $A \in_F [c,d]$. $\square$

PROPERTY 3.2.
(i)   *For any fuzzy set C with support $[m,n]$ which has no intersection with the support of A and B, if $b < m$, $d < m$ and $d_1(A,C) = d_1(B,C)$, $d_2(A,C) = d_2(B,C)$, then $A =_F B$.*
(ii)   *For any fuzzy set C with support $[m,n]$ which have no intersection with the support of A, B, if $a > n$, $c > d$ and $d_1(A,C) = d_1(B,C)$, $d_2(A,C) = d_2(B,C)$, then $A =_F B$.*

Proof.   (i)   If $d_1(A,C) = d_1(B,C)$, since $\inf\{|x-z| \colon x \in A,\ z \in C\} = \inf\{|y-z| \colon y \in B,\ z \in C\}$, then $m-b = m-d$, we get $b = d$. If $d_2(A,C) = d_2(B,C)$, since $\sup\{|x-z| \colon x \in A,\ z \in C\} = \sup\{|y-z| \colon y \in B,\ z \in C\}$, then $n-a = n-c$, we get $a = c$. Hence $A =_F B$.

(ii) is proved similarly. ☐

PROPERTY 3.3.   *For any fuzzy set C with support $[m,n]$ which has no intersection with the support of A and B, if $A \in_F B$, then $d_1(A,C) \geq d_1(B,C)$ and $d_2(A,C) \leq d_2(B,C)$. The inverse is not true.*

Proof.   We only prove $d < m$ case. The other cases are similar.

Since $A \in_F B$, $a \geq c$, $b \leq d$, $m - b \geq m - d$, and $n - a \leq n - c$. So we have $d_1(A,C) \geq d_1(B,C)$ and $d_2(A,C) \leq d_2(B,C)$.

On the other hand, if we choose $[a,b] = [1,3]$, $[c,d] = [15,18]$, $[m,n] = [8,10]$, then $d_1(A,C) \geq d_1(B,C)$ and $d_2(A,C) \leq d_2(B,C)$, but $A \notin_F B$.   ☐

PROPERTY 3.4.   *Let $[a,b] \cap [c,d] = \emptyset$ and $d - c \geq b - a$. Then $d_3(A,B) = c - a$ if $b < c$; $d_3(A,B) = b - d$, if $a > d$.*

Proof.   Considering $b < c$ (the other case are similar)

$$\varepsilon_1 = \inf\{\varepsilon\colon [c,d] \subset [a - \varepsilon, b + \varepsilon]\} = d - b, \quad \varepsilon_2 = \inf\{\varepsilon\colon [a,b] \subset [c - \varepsilon, d + \varepsilon]\} = c - a.$$

Since $\varepsilon_1 - \varepsilon_2 = (d - c) - (b - a) \geq 0$, $d_3(A,B) = \inf\{\varepsilon_1, \varepsilon_2\} = c - a$.   ☐

PROPERTY 3.5.   *Let $[a,b] \cap [c,d] \neq \emptyset$ and $[a,b] \not\subset [c,d]$ and $d - c \geq b - a$. Then $d_3(A,B) = c - a$ if $a < c$; $d_3(A,B) = b - d$ if $b > d$.*

Proof.   Since $d - c \geq b - a$, we have $d - b \geq c - a$. The rest of proof is similar to that of Property 3.4.   ☐

PROPERTY 3.6.   *If $[a,b] \subset [c,d]$ then $d_3(A,B) = 0$.*

Proof. $\varepsilon_1 = \inf\{\varepsilon\colon [c,d] \subset [a - \varepsilon, b + \varepsilon]\} > 0$, $\varepsilon_2 = \inf\{\varepsilon\colon [a,b] \subset [c - \varepsilon, d + \varepsilon]\} = 0$, hence $d_3(A,B) = 0$.   ☐

PROPERTY 3.7.   *For any fuzzy set C with support $[m,n]$,*
(i)   *if $n - m \geq d - c$ and $A \in_F B$, then $d_3(A,C) \leq d_3(B,C)$;*
(ii)   *if $n - m \geq d - c$ and $A =_F B$, then $d_3(A,C) = d_3(B,C)$.*

Proof.   (i)   If $A \in_F B$ then $a \geq c$, $b \leq d$.
CASE 1: $d < m$, then $d_3(A,C) = m - a \leq m - c = d_3(B,C)$.
CASE 2: $c < n$, then $d_3(A,C) = b - n \leq d - n = d_3(B,C)$.
CASE 3: $c < m$, $a < m$, $d \geq m$ then $d_3(A,C) = m - a \leq m - c = d_3(B,C)$.
CASE 4: $c < m$, $a \geq m$, then $d_3(A,C) = 0 < m - c = d_3(B,C)$.
CASE 5: $c \geq n$, $d \leq n$, then $d_3(A,C) = 0 = d_3(B,C)$.
CASE 6: $c \leq n$, $b > n$, $d > n$, then $d_3(A,C) = b - n \leq d - n = d_3(B,C)$.
CASE 7: $c \leq n$, $b \leq n$, $d > n$, then $d_3(A,C) = 0 < d - n = d_3(B,C)$.
The proof is completed.
(ii) is proved similarly.   ☐

4. Testing hypothesis with fuzzy data

It is a new research topic about the hypothesis testing of fuzzy mean with interval values. First of all, we will give a brief definition about the defuzzification. Then under the fuzzy significant level δ, we make a one side or two side testing. These methods are a little different from traditional significant level α. In order to get the robustic properties, we will set up the rejection area level F_δ, according to the fuzzy population.

4.1. Testing hypothesis for fuzzy equal

Let U be the universal set (a discussion domain), $L = \{L_1, L_2, \ldots, L_k\}$ a set of k-linguistic variables on U, and $A = \{A_1, A_2, \ldots, A_m\}$ and $B = \{B_1, B_2, \ldots, B_n\}$ be two sets of fuzzy sample drawn from categorical populations with numbers on U. For each sample we assign a linguistic variable L_j and a normalized membership m_{ij} ($\sum_{j=1}^{k} m_{ij} = 1$), and let $Fn_{A_j} = \sum_{i=1}^{m} Ln_{ij}$, $Fn_{B_j} = \sum_{i=1}^{n} Ln_{ij}$ $j = 1, 2, \ldots, k$ be the total memberships for each data set. The following statements are process for testing hypothesis

Testing hypothesis of fuzzy equal for discrete fuzzy mean

Consider a k-cell multinomial vector $n = \{n_1, n_2, \ldots, n_k\}$ with $\sum_i n_i = n$. The *Pearson Chi-squared test* ($\chi^2 = \sum_i \sum_j \frac{n_{ij} - e_{ij}}{e_{ij}}$) is a well known statistical test for investigating the significance of the differences between observed data arranged in k classes and the theoretically expected frequencies in the k classes. It is clear that the large discrepancies between the observed data and expected cell counts will result in larger values of χ^2.

However, a somewhat ambiguous question is whether (quantitative) discrete data can be considered categorical data, for which the traditional χ^2-*test* can be used. For example, suppose a child is asked the following question: *"how much do you love your sister?"* If the responses is a fuzzy number (say, 70 % of the time), it is certainly inappropriate to use the traditional χ^2-test for the analysis. We will present a χ^2-test for fuzzy data as follows:

Procedures for testing hypothesis of fuzzy equal for discrete fuzzy mean
1. *Hypothesis: two populations have the same distribution ratio.*
2. $\chi^2 = \sum_{i \in A, B} \sum_{j=1}^{c} \frac{([Fn_{ij}] - e_{ij})^2}{e_{ij}}$. *(In order to preform the Chi-square test for fuzzy data, we transfer the decimal fractions of Fn_{ij} in each cell of fuzzy category into the integer $[Fn_{ij}]$ by counting 0.5 or higher fractions as 1 and discard the rest.)*
3. *Decision rule: under significance level α, if $\chi^2 > \chi_\alpha^2(k-1)$, then we reject H_0.*

Testing hypothesis of fuzzy index equal for discrete fuzzy mean

Let $\overline{FX}$ be the fuzzy sample mean, $\bar{X}_f$ be the *defuzzyfication* of $\overline{FX}$. Under the fuzzy significant level F_δ, and the corresponding critical value F_δ, we want to test $H_0 \colon \overline{FX} = F\mu$, where F_μ is the fuzzy mean of the underlying population. Let μ is the defuzzyfication value of $F\mu$, then the above hypothesis becomes $H_0 \colon \mu = \mu_0$.
1. *Hypothesis: $H_0 \colon F\mu = F\mu_0$ vs. $H_1 \colon F\mu \neq F\mu_0$.*

2. *Statistics: find $\overline{FX}$ from a random sample $\{S_i,\ i = 1, \ldots, n\}$.*
3. *Decision rule: under the fuzzy significant level F_δ, if $|\bar{X}_f - \mu_0| > \delta$, then reject H_0.*

Note: for left side test H_0: $\mu \le \mu_0$ vs. H_1: $\mu > \mu_0$ under the fuzzy significant level F_δ, if $\mu_0 - \bar{X}_f > \delta$, we reject H_0. The right hand side testing is similar.

Testing hypothesis with continuous fuzzy mean

1. *Hypothesis: H_0: $F\mu =_F [a, b]$ vs. H1: $F\mu \ne_F [a, b]$.*
2. *Statistics: find $\overline{FX} = [x_l, x_u]$ from a random sample $\{S_i,\ i = 1, \ldots, n\}$.*
3. *Decision rule: under the significant level F_δ, find $k = dr$ (where $r = b - a$), if $|x_l - a| > k$ or $|x_u - b| > k$ then reject H_0.*

4.2. Testing hypothesis for fuzzy belongs to

Testing of fuzzy belongs to with bounded sample

1. *Hypothesis: H_0: $F_\mu \in_F [a, b]$ vs. H_1: $F\mu \notin_F [a, b]$.*
2. *Statistics: find $\overline{FX} = [x_l, x_u]$ from a random sample $\{S_i,\ i = 1, \ldots, n\}$.*
3. *Decision rule: under the significant level F_δ, find $k = dr$ (where $r = b - a$), if $x_l < a - k$ or $x_u > b + k$, then reject H_0.*

Testing of fuzzy belongs to with unbounded below sample

1. *Hypothesis: H_0: $F\mu \in_F (-\infty, b]$ vs. H_1: $F\mu \in_F (-\infty, b]$.*
2. *Statistics: find $\overline{FX} = (\infty, x_u]$ from a random sample $\{S_i,\ i = 1, \ldots, n\}$.*
3. *Decision rule: under the significant level F_δ, find $k = dr$ (where r is a constant), if $x_u > b + k$, then reject H_0.*

Testing of fuzzy belongs to with unbounded above sample

1. *Hypothesis: H_0: $F\mu \in_F [a, \infty)$ vs. H_1: $F\mu \notin_F [a, \infty)$.*
2. *Statistics: find $\overline{FX} = [x_l, 8)$ from a random sample $\{S_i,\ i = 1, \ldots, n\}$.*
3. *Decision rule: under the significant level F_δ, find $k = dr$ (where r is a constant), if $x_l < a - k$, then reject H_0.*

5. Empirical studies

Example 5.1. How do Chinese and English-speaking children's conditional reasoning and expressions develop over time? Is language difference, such as English versus Chinese, related to children's understanding of conditionals? Among the testing stimuli, six conditional questions with different degrees of hypotheticality were asked based on a picture book to the two groups of children in their native language respectively. The questions are listed below.

The test for fuzzy equals is shown in the following table.

From Table 5.2 we find that, there was no difference between these two groups. The fuzzy testing hypothesis of fuzzy equal for discrete fuzzy data, uses more differentiated categories and tends to reflect a more truthful picture of the data.

Example 5.2. A farmer wants to adapt a new cooking style of fry chicken from traditional techniques. He invites 5 experts to join the evaluating experiment.

Table 5.1.   Types of conditional questions and examples.

Conditional questions	Examples
1. Future open conditionals	If you ask your Mom whether she loves you, what will she say?
2. Present open conditionals	If somebody bites you, does it hurt?
3. Past open conditionals	There are lions in the zoo. If I have been to the zoo, would I see the lions?
4. Imaginative conditionals	Which animal would you like to be if you were the piglet? Why would you want to be a _____?
5. Present counterfactuals	The mother pig is afraid after the piglet becomes a lion because the lion might bite her with its sharp teeth. What if the lion didn't have sharp teeth?
6. Past counterfactuals	The piglet was a lion before. But he changed back to be a piglet again at the end. What if the piglet had not changed back to himself, what would the mommy pig have done then?

Table 5.2.   A comparison of the traditional and fuzzy statistical analyses on Chinese- and English-speaking children's responses to 6 conditional questions.

H_0: Language difference does not affect children's understanding of conditionals, i.e., H_0: $F\mu_{\text{Chinese}} =_F F\mu_{\text{English}}$		Con.[a]	Ind.[b]	Non. Con.[c]	Chi-Square Test[e]
1. Future open conditionals	Chinese	29.8	21.1	5	$\chi^2 = 3.71$, $p = 0.16$
	English	16.8	4.4	0.8	Accept H_0
2. Present open conditionals	Chinese	20.6	22.4	12.6	$\chi^2 = 4.68$, $p = 0.096$
	English	14.2	5.6	2.2	Reject H_0
3. Past open conditionals	Chinese	22.8	23.2	10	$\chi^2 = 0.55$, $p = 0.76$
	English	11.4	8	2.6	Accept H_0
4. Imaginative conditionals	Chinese	42.2	12.4	1.4	$\chi^2 = 0.41$[d]
	English	17	4.8	0.2	
5. Present counterfactuals	Chinese	25	9.2	1.8	$\chi^2 = 0.34$, $p = 0.84$
	English	14.4	6	1.6	Accept H_0
6. Past counterfactuals	Chinese	25.6	22.4	8	$\chi^2 = 2.76$, $p = 0.25$
	English	10.8	5.2	6	Accept H_0

Note. [a]Conditional response, [b]Indeterminate response, [c]Non-conditional response, [d]More than two cells numbers are ≤ 1, Chi-Square test is invalid. [e]In order to perform the Chi-square test for fuzzy data, we transfer the decimal fractions in each cell of fuzzy category into the integer by counting 5 and higher fractions as 1 and discard the rest.

After they tested the new fry chicken, they are asked to give a fuzzy grading with: very unsatisfactory $= 1$, unsatisfactory $= 2$, no difference $= 3$, satisfactory $= 4$, very satisfactory $= 5$. Table 5.3 shows the result of the 5 experts' evaluation.

Table 5.3.   Evaluation result for 5 experts.

Expert	Very un-satisfactory $= 1$	Unsatisfactory $= 2$	No difference $= 3$	Satisfactory $= 4$	Very satis-factory $= 5$
A	0	0	0	0.7	0.3
B	0	0	0	0	1.0
C	0	0.4	0.6	0	0
D	0	0	0	0.8	0.2
E	0.1	0.9	0	0	0

Let's set up the hypothesis testing for fuzzy index equal: H_0: $\bar{X}_f = 3$ vs. H_1: $\bar{X}_f \neq 3$. Under the significant level $\delta = 0.1$, since $\bar{X}_f = 2.4$ and $\mu_0 - \bar{X}_f = 3 - 2.4 = 0.6 > 0.1$. Hence we reject H_0. And since the sample fuzzy index $\bar{X}_f$ is less than 3, the manager will not apply this new cooking style.

Example 5.3.   A company administrator wants to control the time of turning on air-condition base on the energy saving reason. He feel that the temperature over 28°C will be hot and is the time to turn on. However, he wants to know how the other staff feeling. So, he asks for five staffs at random to investigate and then gets five data $[27, \infty)$, $[26, \infty)$, $[29, \infty)$, $[24, \infty)$, $[26, \infty)$.

Hypothesis H_0: $\mu = [28, \infty)$ *vs.* H_1: $\mu \neq [28, \infty)$.

After simple computation, we have $\overline{FX} = [26.4, \infty)$. Under the significant level $\delta = 0.2$, since $28 - 26.4 > 0.2$, we reject H_0 and suggest that turn on the air condition when the temperature is below 28°C.

Example 5.4.   The human resource department announced that 20 to 26 years old people request their salary between 20 thousands and 40 thousands with deviation 5 thousands. The manger asks the statistical department to check it up. Suppose they find 10 young man between 20 and 26 years old, survey their request salary, the sample data are: $[3, 4]$, 1.8, $[2, 3]$, $[4, 6]$, $[1.5, 2]$, $[3, 4]$, 2, $[2, 3]$, $[3, 5]$, $[2.5, 4]$ (unit: 10 thousands).

Hypothesis H_0: $F\mu \in_F [2, 4]$ *vs.* H_1: $F\mu \notin_F [2, 4]$.

We treat 1.8 as $[1.8, 1.8]$. After simple computation, we get

$$\overline{FX}$$

$$= \left[\frac{3 + 1.8 + 2 + 4 + 1.5 + 3 + 2 + 2 + 3 + 2.5}{10}, \frac{4 + 1.8 + 3 + 6 + 2 + 4 + 2 + 3 + 5 + 4}{10} \right]$$

$$= [2.38, 3.48].$$

Under the significant level $\delta = 0.5$, since $2.38 > 2 - 0.5$ and $3.48 < 4 + 0.5$ and $2.38 < 2 + 0.5$ but $3.48 < 4 - 0.5$, we do not accept what human resource department say. We accept H_0: $\overline{FX} \in_F [2, 4]$.

Example 5.5.　Suppose a salesman wonder how the living standard will influence the sales of volume for two communities X and Y. They want to find out which has higher income level to make the sales strategy. He chooses 100 families at random from each community and gets data.

Hypothesis H_0: Two communities have same income level, H_1: Two communities have different income level.

After simple computation, we get $\overline{FX} = [4.3, 5.5]$ and $\overline{FY} = [6.7, 7.7]$ (unit: 10 thousands). Under the significant level $F_\delta = 1$, $\overline{FX}_\delta = 4.9$, $\overline{FY}_\delta = 7.2$. Since $7.2 - 4.9 > 1$, we say community Y has higher income level than that of X.

6. Conclusion

Fuzzy statistical analysis grows as a new discipline from the necessity to deal with vague samples and imprecise information caused by human thought in certain experimental environments. In this paper, we made an attempt to link the gap between the binary logic based on multiple choice survey with a more complicated yet precise fuzzy membership function assessment, such as fuzzy mode, fuzzy median, fuzzy weight and α-cut etc. We carefully revealed how theses factors can be properly and easily utilized in various fields to reveal the contradictory characteristics of human concepts. Through these processes, human ideas are no longer presented as discrete but as a natural and continuous flow. There are illustrated examples demonstrated to explain how to find the fuzzy mode and fuzzy median, and how to use the results to help people reaching their decisions.

However, there are still some problems we need to investigate in the future:

1. We can further research on data simulation so that we can understand features of the fuzzy linguistic, multi-facet assessment, and the balance of the moving consensus. Moreover, the choice of different significant α-cut will influence the statistical result. An appropriate criterion for selecting significant α-cut should be investigated in order to reach the best common agreement of human beings.

2. There are other types of membership functions we could explore in the future. For the fuzzy mode of continuous type, we can extend the uniform and triangular types of membership functions to non-symmetric or multiple peaks types.

References

[1]　D. Dubois and H. Prade, Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions. Fuzzy Sets and Systems, **40** (1991), 143–202.

[2]　J. Goutsias, R.P.S. Mahler and H.T. Nguyen (eds.), Random Sets: Theory and Applications. Springer-Verlag, N.Y., 1997.

[3]　G.S. Liang and M.J. Wang, A fuzzy multicriteria decision making method for facility site selection. International Journal of Production Research, **29** (1991), 2313–2330.

[4]　H.T. Nguyen and B. Wu, Fuzzy Mathematics and Statistical Applications. Hua-Tai Book Company, Taipei, 2000.

[5]　H.T. Nguyen and B. Wu, Fundamentals of Statistics with Fuzzy Data. Springer-Verlag, Heidelberg, 2006.

[6]   M. Stojakovic, Fuzzy random variables, expectation, and martingales. Journal of Mathematical Analysis and Applications, **184** (1994), 594–606.

[7]   B. Wu and C. Sun, Interval-valued statistics, fuzzy logic, and their use in computational semantics. Journal of Intelligent and Fuzzy Systems, **11** (2001), 1–7.

[8]   B. Wu and W. Yang, Application of fuzzy statistics in the sampling survey. Development and Application for the Quantity Methods of Social Science. Academic Sinica, Taiwan, 1998, 289–316.

[9]   H.J. Zimmermann, Fuzzy Set Theory and Its Applications, Kluwer Academic, Boston, 1991.