THE DEVELOPMENT OF AFS THEORY UNDER PROBABILITY THEORY

XIAODONG LIU¹,²

Abstract. As moving further into the age of machine intelligence and automated decision-making, we have to deal with both the subjective imprecision of human perception-based information described in natural language and the objective uncertainty of randomness universally existing in the real world. A basic limitation of standard probability theory which cannot deal with information described in natural language becomes a serious problem. With its abilities to represent natural language, the notion of AFS (Axiomatic Fuzzy Set) theory has proven useful in the clustering, classifications, concept representations and decision trees. In this paper, we apply AFS theory and probability theory to propose a new interpretation of the membership functions taking both fuzziness (subjective imprecision) and randomness (objective uncertainty) into account. So that uncertainty of randomness and of imprecision can be treated in a unified and coherent manner under the AFS and probability framework. It opens a door to explore the deep mathematical analysis properties of fuzzy set theory and to a major enlargement of the role of natural languages in probability theory.

Key Words. AFS structures, AFS algebras, AFS fuzzy logic, Probability density, Parameter estimator

1. Introduction

The essence in the designs of intelligent systems is the ability to imitate remarkable human intelligent behavior, such as, in decision-making, pattern recognition and control. As Zadeh advocated since several years [52], information gathered by humans is perception-based with imprecision and uncertainty. Fuzzy sets were introduced by Zadeh (1965) based on the premise that an exact description of many real world situations is virtually impossible, and that imprecisely defined “classes” play an important role in human thinking and natural language. Examples are commonly used adjectives, such as “substantial,” “significant,” “accurate,” “approximate,” “small,” and “medium.” Furthermore, such activities as the communication of information, speech recognition, knowledge representation, medical diagnosis, and assessment of rare events suggest that the human brain often reasons with vague assertions, a fact that one needs to accept and to adjust to. All the same, computers that permeate our everyday lives do not reason as brains do, and it has been argued that the main distinction between human intelligence and

Received by the editors December 10, 2006 and, in revised form, January 22, 2007.
2000 Mathematics Subject Classification. 93A30.
This work is supported in parts by the National Science Foundation in China under Grant 60575039, 60534010 and by the National Key Basic Research Development Program of China under Grant 2002CB312200-06.
machine intelligence lies in the ability of humans to manipulate imprecise concepts and imprecise instructions. The notion of fuzzy sets strives to balance exactness and simplicity in such a way that complexity can be reduced without oversimplification. Imprecision can also arise in the context of real world decision making wherein the goals, the constraints, and the consequences of actions cannot be precisely specified. Whereas the human mind is able to cope with such impressions. Fuzzy sets provide a trade-off between the reality and the requirement. Indeed, it is in the context of control theory that the notion of fuzzy sets has had its biggest impact. For example, according to some estimates, in 1992 Japan produced about $2 billion worth of products with control mechanisms considered fuzzy [18].

There is a deep-seated tradition in science of dealing with uncertainty—whatever its form and nature—through the use of probability theory. What we see is that standard probability theory has many strengths and many limitations. But as we move further into the age of machine intelligence and automated decision-making, a basic limitation of standard probability theory becomes a serious problem. More specifically, in large measure, standard probability theory cannot deal with information described in natural language; that is, to put it simply, standard probability theory does not have natural language capability. A basic problem with standard probability theory is that it does not address partiality of truth. The principal limitation is that standard probability provides no tools for operating on information that is perception-based and is described in a natural language. This incapability is rooted in the fact that perceptions are intrinsically imprecise, reflecting the bounded ability of sensory organs, and ultimately the brain, to resolve detail and store information. As a consequence, there are no means in standard probability theory for understanding natural language. It does not have this capability principally because there is no mechanism in standard probability theory for (a) representing the meaning of perceptions and (b) computing and reasoning with representations of meaning. To deal with perception-based information, probability theory must be three-dimensional [54]; that is, it must provide tools with dealing with partiality of certainty, partiality of truth and partiality of possibility. This is a necessity because perceptions are described in a natural language, and natural languages are three-dimensional. Because it is possible that both uncertainty and imprecision can be present in the same problem [56], Zadeh has also claimed that “probability must be used in concert with fuzzy logic to enhance its effectiveness. In this perspective, probability theory and fuzzy logic are complementary rather than competitive.” It is this statement that has motivated the thesis of this article. Our aim here is to explore how fuzzy set theory and probability theory, call it FTP, can be made to work in concert, so that uncertainty of randomness and of imprecision can be treated in a unified and coherent manner. The capability of FTP to process perception-based information has an important implication. Specially, it opens the door to a major enlargement of the role of natural languages in probability theory.

The earliest attempt at making probability and fuzzy set theory work in concert was made by Loginov 1966 in [20], who interpreted the membership function as a frequentist conditional probability. Barrett and Woodall (1997),[1], in introducing a probabilistic alternative to fuzzy logic controllers, appeared to follow a similar line of thought. Zadeh in [56] dismissed Loginov’s interpretation on grounds that requiring each voter to classify an observed \( x \) in a fuzzy set \( A \) or \( A^c \) is unnatural, because fuzzy sets reject the law of the excluded middle. We concur with this criticism provided by Zadeh. Also, because membership functions are often specified by one individual based on subjective considerations, the consensus model involving an ensemble of
voters is untenable. Under their interpretation of the membership function that it is a likelihood function with $A$ taking the role of a fixed observation and the values of $x$ taking the role of the hypotheses, Singpurwalla and Booker in [44] develop a line of argument that demonstrates that probability theory has a sufficiently rich structure for incorporating fuzzy sets within its framework. Thus probabilities of fuzzy events can be logically induced. The philosophical underpinnings that make this happen are a subjectivistic interpretation of probability, an introduction of Laplace's famous genie, and the mathematics of encoding expert testimony. Singpurwalla and Booker in [44] provides a real advance in our understanding of fuzzy sets, by providing a sensible connection between membership functions and likelihood, and thereby probability. However, in [44] Singpurwalla and Booker focus on the interpretation of the probability measure of a fuzzy event with the membership function given in advance and have not discussed the problem of how to determine membership functions for fuzzy sets based on the theory they developed in [44].

Membership functions are usually given directly by the user’s subjectivity. But these membership functions cannot be used in the fuzzy observation model because they have no assurance to meet the restriction as the fuzzy event. In AFS theory [22]–[38], [49]–[58], the membership functions and logic operations of fuzzy concepts are determined by the distribution of original data and semantics through the AFS structures and AFS algebras. An AFS structure is a triple $(M, \tau, X)$ which is a special family of combinatorical objects [12], where $X$ is the universe of discourse, $M$ is a set of some simple (or elementary) concepts on $X$ (e.g., linguistic labels on a features such as “large”, “medium”, “small”) and $\tau$ is mathematical abstract of the complicated relations containing in the original data and semantics. An AFS algebra is a family of molecular lattices [47] (i.e., completely distributive lattices) generated by sets such as $X, M$. Using the AFS algebras and the AFS structures, a great large number of complex fuzzy concepts on $X$ and their logic operations can be expressed by $E_1$ algebra $E_M$, one kind of the AFS algebras. In [27], the complement operation of the fuzzy sets has been obtained, thus a fuzzy logic system, which is called AFS fuzzy logic, has been developed. Recently, AFS theory has been developed further and applied to fuzzy decision trees [35], fuzzy clustering analysis [7], [33], [43], concept representations [31], [32], fuzzy cognitive maps [28] and credit rating analysis [37], etc..

In this paper, first we analysis the membership function determining algorithms in AFS (Axiomatic Fuzzy set) theory under probability theory and show that the membership functions of AFS theory take both fuzziness (subjective imprecision) and randomness (objective uncertainty) into account. Then, we reinterpret the membership functions of fuzzy concepts in $E_M$ based on the probability measure of the fuzzy events in a probability space [44] and propose a new approach to make probability and fuzzy set theory work in concert. Finally, by the applications to real world examples, we show that uncertainty of randomness and of imprecision can be treated in a unified and coherent manner under the AFS and probability framework.

2. The preliminary of AFS algebras and AFS structures

In this section, we recall some notations and properties of AFS theory that were detailed in [22]–[38], [49]–[58] and will be used in the sequel.

2.1. AFS algebras. In [22], the author has defined a family of molecular lattices, the AFS algebras, i.e., the $E_1, E_{11}, ... , E_{1^n}$ algebras and applied AFS algebras to study the lattice valued representations for fuzzy concepts.
Definition 1. ([22]) Let $X_1, \ldots, X_n, M$ be $n + 1$ non-empty sets. Then the set $E X_1 \ldots X_n M^*$ is defined by

$$E X_1 \ldots X_n M^* = \{ \sum_{i \in I} (u_{i1}, \ldots, u_{in} A_i) | A_i \in 2^M, u_{ri} \in 2^X, r = 1, \ldots, n, i \in I, \}
$$

In the case $n = 0$,

$$E M^* = \{ \sum_{i \in I} A_i | A_i \in 2^M, i \in I, I \text{ is a non-empty indexing set}. \}
$$

where the element $\sum_{i \in I}(u_{i1}, \ldots, u_{in} A_i)$ is composed of terms $(u_{i1}, \ldots, u_{in}) A_i$’s, $i \in I$, separated by “+”.

Theorem 1. ([22]) Let $X_1, \ldots, X_n, M$ be $n + 1$ non-empty sets. A binary relation $R$ on $E X_1 \ldots X_n M^*$ is defined as follows. For $\sum_{i \in I}(u_{i1}, \ldots, u_{in} A_i), \sum_{j \in J}(v_{j1}, \ldots, v_{jn} B_j) \in E X_1 \ldots X_n M^*$,

$$\left[ \sum_{i \in I}(u_{i1}, \ldots, u_{in} A_i) \right] R \left[ \sum_{j \in J}(v_{j1}, \ldots, v_{jn} B_j) \right] \iff \begin{cases} (i) \forall (u_{i1}, \ldots, u_{in}) A_i \ (i \in I), \exists (v_{h1}, \ldots, v_{hn}) B_h \ (h \in J) \text{ such that } A_i \supseteq B_h, \ u_{ri} \subseteq v_{rh}, \ 1 \leq r \leq n; \\ (ii) \forall (v_{j1}, \ldots, v_{jn}) B_j \ (j \in J), \exists (u_{k1}, \ldots, u_{nk}) A_k \ (k \in I), \text{ such that } B_j \supseteq A_k, \ v_{rj} \subseteq u_{rk}, \ 1 \leq r \leq n. \end{cases}$$

It’s obvious that $R$ is an equivalence relation. The quotient set $E X_1 \ldots X_n M^*/R$ is denoted by $E X_1 \ldots X_n M$. By using notation $\sum_{i \in I}(u_{i1}, \ldots, u_{in} A_i) = \sum_{j \in J}(v_{j1}, \ldots, v_{jn} B_j)$ we mean that $\sum_{i \in I}(u_{i1}, \ldots, u_{in} A_i)$ and $\sum_{j \in J}(v_{j1}, \ldots, v_{jn} B_j)$ are equivalent under equivalence relation $R$.

Theorem 1. ([22]) Let $X_1, \ldots, X_n, M$ be $n + 1$ non-empty sets. Then $(E X_1 \ldots X_n M, \vee, \wedge)$ forms a completely distributive lattice under the binary compositions $\vee$ and $\wedge$ defined as follows. For any $\sum_{i \in I}(u_{i1}, \ldots, u_{in} A_i), \sum_{j \in J}(v_{j1}, \ldots, v_{jn} B_j) \in E X_1 \ldots X_n M$,

$$(1) \quad \sum_{i \in I}(u_{i1}, \ldots, u_{in} A_i) \vee \sum_{j \in J}(v_{j1}, \ldots, v_{jn} B_j) = \sum_{k \in I \cup J}(u_{i1}, \ldots, u_{in} C_k),$$

$$(2) \quad \sum_{i \in I}(u_{i1}, \ldots, u_{in} A_i) \wedge \sum_{j \in J}(v_{j1}, \ldots, v_{jn} B_j) = \sum_{k \in I \cup J}([u_{i1}, \ldots, u_{in} \cap v_{j1}, \ldots, v_{jn}](A_i \cup B_j)],$$

where for any $k \in I \cup J$ (the disjoint union of $I$ and $J$), $C_k = A_k, v_{rk} = u_{rk}$ if $k \in I$, and $C_k = B_k, w_{rk} = v_{rk}$ if $k \in J$, $r = 1, 2, \ldots, n$.

$(E X_1 \ldots X_n M, \vee, \wedge)$ is called the $EI^{n+1}$ (expanding $n + 1$ sets $X_1, \ldots, X_n, M$) algebra over $X_1, \ldots, X_n$ and $M$. For $\alpha = \sum_{i \in I}(u_{i1}, \ldots, u_{in} A_i), \beta = \sum_{j \in J}(v_{j1}, \ldots, v_{jn} B_j) \in E X_1 \ldots X_n M$, $\alpha \leq \beta \iff \alpha \vee \beta = \beta \iff \forall (u_{i1}, \ldots, u_{in}) A_i \ (i \in I), \exists (v_{h1}, \ldots, v_{hn}) B_h \ (h \in J) \text{ such that } A_i \supseteq B_h, \ u_{ri} \subseteq v_{rh}, \ 1 \leq r \leq n.$

We first explain the $EI$ algebras (in the case $n = 0$ for the $EI^{n+1}$ algebras) using the semantics represented by the elements of $EM$ shown in Example 1. This enables us to “understand” the abstract AFS algebras intuitively.

Example 1. Let $X = \{x_1, x_2, \ldots, x_{10}\}$ be a set of 10 persons and the features which are described by real numbers (i.e. age, height, weight, salary, estate), Boole values (i.e. male, female) and the order relations (hair black, hair white, hair yellow) are shown in Table 1.

| Table 1 - Descriptions of features |
Here the number $i$ in the hair color columns corresponding to $x \in X$ implies that the hair color of $x$ is numbered $i$ by comparing the hair color differences of the persons. For example, the numbers in the column “hair black” implies the order $x_7 > x_{10} > x_4 > x_8 > x_2 = x_9 > x_5 > x_6 = x_3 = x_1$ i.e., from right to left, the hair color of the persons look more and more black. In the order, $x_i > x_j$ (e.g., $x_7 > x_{10}$) means that the hair of $x_i$ looks blacker than $x_j$. $x_i = x_j$ (e.g., $x_4 = x_8$) means that the hair of $x_i$ looks as black as $x_j$. A concept on $X$ may associate to one or more than one features. For instance, the fuzzy concept “tall” associates a single feature “height” and the fuzzy concept “old white hair males” associates three features “age”, “haircolor_black” and “gender_male”. Many concepts may associate to a single feature. For instance, the fuzzy concepts “old”, “young” and “about 40 years old” all associate to feature “age”. Let $M = \{m_1, m_2, \ldots, m_{10}\}$ be the set of fuzzy or crisp concepts on $X$ and each $m \in M$ associate to a single feature. Where $m_1$ : “old persons”, $m_2$ : “tall persons”, $m_3$ : “heavy persons”, $m_4$ : “high salary”, $m_5$ : “more estate”, $m_6$ : “male”, $m_7$ : “female”, $m_8$ : “black hair persons”, $m_9$ : “white hair persons”, $m_{10}$ : “yellow hair persons”. The elements of $M$ are viewed as “elementary” (or “simple”) concepts. For the subset $A_i \subset M$ represents conjunction of the concepts in $A_i$ (e.g., $A_i = \{m_1, m_6\}$ representing a new fuzzy concept “old males”) and $\sum_{i \in I} A_i$ is the disjunction of the conjunctions represented by $A_i$’s (i.e., every element of $EM$ corresponds to the disjunctive normal form of a formula representing a concept). For example, $\gamma = \{m_1, m_6\} + \{m_1, m_3\} + \{m_2\} \in EM$ states that “old males” or “heavy old persons” or “tall persons”. Although $M$ may be a set of fuzzy or crisp concepts, every element of $EM$ has a well-defined meaning like the one we have discussed above. By Definition 2, we know that

$$\{m_3, m_8\} + \{m_4, m_4\} + \{m_1, m_6, m_7\} + \{m_1, m_4, m_8\} = \{m_3, m_8\} + \{m_1, m_4\} + \{m_1, m_6, m_7\}$$

This implies that the left side and right side are equivalent as two fuzzy concepts they represent. Considering the terms in left side, for any $x$, the degree of $x$ belonging to the fuzzy concept representing by $\{m_1, m_4, m_8\}$ is always less than or equal to the degree of $x$ belonging to the fuzzy concept representing by $\{m_1, m_4\}$. Therefore, term $\{m_1, m_4, m_8\}$ is redundant when forming the left side of the fuzzy concept. In the sequel, we call each element in $EM$ a fuzzy concept. By (1) and (2), we observe that the operations $\lor, \land$ of the elements of $EM$ correspond to the “or”, “and” of the corresponding fuzzy concepts respectively. For instance, $\alpha = \{m_1, m_4\} + \{m_2, m_5, m_6\}, \nu = \{m_5, m_6\} + \{m_5, m_8\} \in EM$, the semantic significations of
fuzzy concepts “α or ν” and “α and ν” can be simply expressed as α ∨ ν and α ∧ ν respectively. With (1), (2) and Definition 2, we have

\[ \alpha \lor \nu = \{m_1, m_4\} + \{m_2, m_5, m_6\} + \{m_5, m_6\} + \{m_5, m_8\} \]

\[ = \{m_5, m_6\} + \{m_5, m_8\} + \{m_1, m_4\}. \]

\[ \alpha \land \nu = \{m_1, m_4, m_5, m_6\} + \{m_2, m_5, m_6\} + \{m_1, m_4, m_5, m_8\} + \{m_2, m_5, m_6, m_8\} \]

\[ = \{m_1, m_4, m_5, m_6\} + \{m_2, m_5, m_6\} + \{m_1, m_4, m_5, m_8\}. \]

In [27], [29], the authors proved that the following operator “′” is an order-reversing involution of EI algebra EM, if for any \( \Sigma_{i \in I} A_i \in EM, \)

\[ (\sum_{i \in I} A_i)' = \land_{i \in I}(\lor_{a \in A_i}\{a'\}) = \land_{i \in I}(\sum_{a \in A_i} \{a'\}). \]

If a’ means the negation of the concept a \( \in M, \) then for any \( \zeta \in EM, \) \( \zeta' \) means the logical negation of \( \zeta. \) For instance,

\[ \gamma' = \{(m_1, m_6) + \{m_1, m_3\} + \{m_2\}\}' \]

\[ \quad = (\{m_1'\} + \{m_6\}'\} \land (\{m_1'\} + \{m_3\}'\} \land \{m_2\}'\} \]

\[ \quad = \{m_1', m_6'\} \land \{m_1', m_3'\} \land \{m_2\}'\} \]

\[ \gamma', \text{ which is the logical negation of } \gamma = \{m_1, m_6\} + \{m_1, m_3\} + \{m_2\}, \text{ states that} \]

“not old and not tall persons” or “not tall and not heavy females”.

For \( M \) a set of few fuzzy or crisp concepts, a great number of fuzzy concepts can be expressed by the elements of \( EM \) and the fuzzy logic operations can be implemented by the operations \( \lor, \land \) and ‘ in EI algebra system (\( EM, \lor, \land, ' \)). As long as we can determine the fuzzy logic operations of the few concepts in \( M, \) the fuzzy logic operations of all concepts in \( EM \) can also be determined. Thus, not only will the accuracy of the representations and the fuzzy logic operations of fuzzy concepts be improved in comparison with the fuzzy logic equipped with some \( t \)-norms and a negation operator, but also the complexity of determining membership functions and their logic operations for the complex fuzzy concepts will be alleviated. Let us stress that the complexity of human concepts is a direct result of the combinations of a few relatively simple concepts. It is obvious that the simpler the concepts in \( M, \) the more accurately and conveniently the membership functions and the fuzzy logic operations of the fuzzy concepts in \( EM \) will be determined. The few concepts in \( M \) serve like a basis in linear vector space. In what follows, we define the “simple concepts” which are suitable to serve as a basis.

**Definition 3.** Let \( \zeta \) be any concept on the universe of discourse \( X. \) \( R_\zeta \) is called a binary relation (i.e., \( R_\zeta \subseteq X \times X \)) of \( \zeta \) if \( R_\zeta \) satisfies: \( x, y \in X, (x, y) \in R_\zeta \Leftrightarrow x \) belongs to concept \( \zeta \) at some degree and the degree of \( x \) belonging to \( \zeta \) is larger than or equal to that of \( y, \) or \( x \) belongs to concept \( \zeta \) at some degree and \( y \) does not at all.

For instance, according to the value of each \( x \in X \) on the feature age shown in Table 1, we have the binary relations \( R_\zeta, R_\zeta', R_\gamma \) of the fuzzy concepts \( \zeta: \) “old”, \( \zeta': \) “not old”, \( \gamma: \) “the person whose age is about 40 years old” as follows:

\[ R_\zeta = \{(x, y) \mid (x, y) \in X \times X, \text{age}_x \geq \text{age}_y\} \]

\[ R_\zeta' = \{(x, y) \mid (x, y) \in X \times X, \text{age}_x \leq \text{age}_y\} \]

\[ R_\gamma = \{(x, y) \mid (x, y) \in X \times X, |\text{age}_x - 40| \leq |\text{age}_y - 40|\}, \]
where \( \text{age}_x \) is the age of \( x \). Note that \((x, x) \in R_{\text{m}} \) implies that \( x \) belongs to \( \eta \) at some degree and that \((x, x) \notin R_{\text{m}} \) implies that \( x \) does not belong to \( \zeta \) at all. For fuzzy concept \( m_8 : \text{“more estate”} \) in Example 1, by feature \text{“estate”} and Definition 3, we have \((x_5, x_5) \in R_{m_8} \) although the estate of \( x_5 \) is just 2, and \((x_2, x_2) \notin R_{m_8} \) because the estate of \( x_2 \) is 0. For a crisp concept \( \xi, (x, y) \in R_{\xi} \) implies that \( x \) belongs to concept \( \xi \). For instance, the crisp concept \( m_0 : \text{“male”} \) in Example 1, by the feature \text{“male”} and Definition 3, we have \((x_1, y), (x_3, y), (x_5, y), (x_7, y), (x_8, y) \in R_{m_0} \) and \((x_2, y), (x_3, y), (x_6, y), (x_9, y), (x_{10}, y) \notin R_{m_0} \) for any \( y \in X \).

In real world applications, the comparison of the degrees of a pair \( x \) and \( y \) belonging to a concept can be obtained through the feature values or human intuitions without representing the degrees by \([0, 1]\) or a lattice in advance. For instance, we can obtain the binary relation \( R_{m_8} \) for fuzzy concept \( m_8 : \text{“black hair persons”} \) in Example 1, just by comparing the black color differences of each pair of persons’ hair through our intuitions. Based on Table 1 and our intuitions, we can obtain the binary relation \( R_{\text{m}} \) of each concept \( m \in M \) in Example 1.

**Definition 4.** Let \( X \) be a set and \( R \) be a binary relation on \( X \). \( R \) is called a sub-preference relation on \( X \) if for \( x, y, z \in X, x \neq y, R \) satisfies the following conditions:

1. If \((x, y) \in R \), then \((x, x) \in R\);
2. If \((x, x) \in R \) and \((y, y) \notin R \), then \((x, y) \in R\);
3. If \((x, y), (y, z) \in R \), then \((x, z) \in R\);
4. If \((x, x) \in R \) and \((y, y) \in R \), then either \((x, y) \in R \) or \((y, x) \in R \).

A concept \( \zeta \) is called a simple concept on \( X \) if \( R_{\zeta} \) is a sub-preference relation. Otherwise \( \zeta \) is called a complex concept on \( X \). Where \( R_{\zeta} \) is the binary relation of \( \zeta \) defined by Definition 3.

Many concepts associating to more than one feature (attribute) are complex concepts. For example, let \( X \) be a set of persons and cars. If \( x, y \in X \), \( x \) is a person and \( y \) is a car and we consider concept \text{“beautiful”} , then the degrees of \( x, y \) belonging to \text{“beautiful”} are incomparable although both \( x \) and \( y \) may belong to \text{“beautiful”} at some degree, i.e., \((x, x), (y, y) \in R_{\text{beautiful}}, (x, y) \notin R_{\text{beautiful}} \). This implies that 4 of Definition 4 is not satisfied and \text{“beautiful”} is a complex concept on \( X \). By Table 1 and Definition 4, one can verify that each concept \( m \in M \) in Example 1 is a simple concept. Let fuzzy concept \( \beta = \{m_1, m_2\} \) meaning “tall old persons”. One can verify that \((x_5, x_8), (x_4, x_4) \in R_{\beta} \) but neither \((x_8, x_4) \) nor \((x_4, x_8) \) in \( R_{\beta} \). This implies that 4 of Definition 4 is not satisfied by \( R_{\beta} \) and \( \beta \) is also a complex concept. The fuzzy concept \( \gamma = \sum_{i \in I} A_i \in EM \) may be complex concept provided that \((x, y) \in R_{\gamma} \Leftrightarrow \exists k \in I \text{ such that } (x, y) \in R_{A_k} \) (i.e., \( \forall a \in A_k, (x, y) \in R_{a} \)) for \((x, y) \in X \). For instance, the fuzzy concept \( \gamma = \{m_1\} + \{m_2\} \in EM \) states that \text{“old persons”} or \text{“tall persons”} in Example 1. By the data shown in Table 1, i.e., \( x_8 : \text{age}=70, \text{height}=1.6; x_1 : \text{age}=20, \text{height}=1.9; x_4 : \text{age}=80, \text{height}=1.8 \), we have \((x_8, x_1) \in R_{\gamma} \) because \( x_8 \) is older than \( x_1 \) and \((x_1, x_4) \in R_{\gamma} \) because \( x_1 \) is taller than \( x_4 \). But \( x_8 \) is neither older nor taller than \( x_4 \), i.e., \((x_8, x_4) \notin R_{\gamma} \). Thus the binary relation \( R_{\gamma} \) does not satisfy 3 of Definition 4 and concept \( \gamma \) is a complex concept.

2.2. AFS structure. An AFS structure, a triple \((M, \tau, X)\), gives rise to the lattice representations of the membership degrees and fuzzy logic operations of the concepts in \( EM \).
In this section, we recall the concept, we have \( \tau : X \times X \to 2^M \). \((M, \tau, X)\) is called an AFS structure if \( \tau \) satisfies the following conditions:

1. \( \forall (x_1, x_2) \in X \times X, \tau(x_1, x_2) \subseteq \tau(x_1, x_1) \);  (AX1)
2. \( \forall (x_1, x_2), (x_2, x_3) \in X \times X, \tau(x_1, x_2) \cap \tau(x_2, x_3) \subseteq \tau(x_1, x_3) \). (AX2)

\( X \) is called universe of discourse, \( M \) is called a concept set and \( \tau \) is called a structure.

Let \( X \) be a set of objects and \( M \) be a set of simple concepts on \( X \). If \( \tau : X \times X \to 2^M \) is defined as follows: for any \( (x, y) \in X \times X \)

(4) \[ \tau(x, y) = \{m \in M | (x, y) \in R_m \} \in 2^M, \]

where \( R_m \) is the binary relation of simple concept \( m \in M \) (refer to Definition 3). Then \((M, \tau, X)\) is an AFS structure. Now we prove it. For any \((x_1, x_2) \in X \times X\), if \( m \in \tau(x_1, x_2) \), then by (4) we know \((x_1, x_2) \in R_m \). Because each \( m \in M \) is a simple concept, we have \((x_1, x_1) \in R_m \) by Definition 4, i.e., \( m \in \tau(x_1, x_1) \). This implies that \( \tau(x_1, x_2) \subseteq \tau(x_1, x_1) \) and AX1 of Definition 5 holds. For \((x_1, x_2), (x_2, x_3) \in X \times X\), if \( m \in \tau(x_1, x_2) \cap \tau(x_2, x_3) \), then \((x_1, x_2), (x_2, x_3) \in R_m \). Since \( m \) is a simple concept, so by Definition 4, we have \((x_1, x_3) \in R_m \), i.e., \( m \in \tau(x_1, x_3) \). This implies \( \tau(x_1, x_2) \cap \tau(x_2, x_3) \subseteq \tau(x_1, x_3) \) and AX2 of Definition 5 holds. Therefore \((M, \tau, X)\) is an AFS structure. By the above discussion, an AFS structure based on a data set can be established by (4), as long as each concept in \( M \) is a simple concept on \( X \).

Let us continue to study Example 1, in which \( X = \{x_1, x_2, ..., x_{10}\} \) is the set of 10 persons and their feature descriptions are shown in Table 1. \( M = \{m_1, m_2, ..., m_{10}\} \) is the set of simple concepts shown in Example 1. By Table 1 and Definition 4, one can verify that each concept \( m \in M \) is a simple concept. Thus for any \( x, y \in X, \tau(x, y) \) is well-defined by (4). For instance, we have

\[
\begin{align*}
\tau(x_4, x_4) &= \{m_1, m_2, m_3, m_4, m_5, m_6, m_8, m_9, m_{10}\} \\
\tau(x_4, x_7) &= \{m_1, m_2, m_6, m_9, m_{10}\}
\end{align*}
\]

by comparing the attribute values of \( x_4, x_7 \) shown in Table 1 as follows:

<table>
<thead>
<tr>
<th>age</th>
<th>height</th>
<th>weight</th>
<th>salary</th>
<th>estate</th>
<th>male</th>
<th>female</th>
<th>black</th>
<th>white</th>
<th>yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_4 )</td>
<td>80</td>
<td>1.8</td>
<td>73</td>
<td>20</td>
<td>80</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>( x_7 )</td>
<td>45</td>
<td>1.7</td>
<td>78</td>
<td>268</td>
<td>90</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

Similarly, we can obtain \( \tau(x, y) \) for other \( x, y \in X \). Finally we have the AFS structure \((M, \tau, X)\) corresponding to Table 1.

### 2.3. The representations of fuzzy concepts in \( EM \)

In this section, we recall some representations of fuzzy concepts in the framework of AFS theory which will be used in the sequel.

**Theorem 2.** ([22]) Let \((M, \tau, X)\) be an AFS structure. For \( x \in X, A \subseteq M \), we define the symbol

(5) \[ A_\tau(x) = \{ y | y \in X, \tau(x, y) \supseteq A \}. \]

For any given \( x \in X \), if we define \( \phi_x : EM \to EM \) as follows: For any \( \sum_{i \in I} A_i \in EM \),

\[ \phi_x(\sum_{i \in I} A_i) = \sum_{i \in I} A_i(\{x\})A_i \in EM, \]

then \( \phi_x \) is a homomorphism from the lattice \((EM, \vee, \wedge)\) to the lattice \((EM, \vee, \wedge)\).
Theorem 2 implies that for any given concept \( \eta = \sum_{i \in I} A_i \in EM \), we get a map 
\[
\eta(x) = \sum_{i \in I} A_i(\{x\})A_i \in EXM.
\]

Since \((EXM, \vee, \wedge)\) is a lattice, hence map \(\sum_{i \in I} A_i\) is a L-fuzzy set with membership degrees valued by \(EI^2\) algebra \(EXM\). For \(\alpha, \beta \in EM\), \(\alpha \vee \beta\) and \(\alpha \wedge \beta\) are “\(\alpha\) and \(\beta\)”. \(\neg\) is the negation of the L-fuzzy sets in \(EM\). Thus we have an \(EI^2\) algebra representations and logic operations of fuzzy concepts in \(EM\), and \((EM, \vee, \wedge', \neg)\) can be viewed as a fuzzy logic system.

**Definition 6.** ([31]) Discrete case: Let \(X\) be a finite set and \(S \subseteq 2^X\) be a \(\sigma\)-algebra over \(X\). \(\rho : X \rightarrow R^+ = [0, \infty)\) with \(0 < \sum_{x \in X} \rho(x) < \infty\). For any \(A \in S\), a measure \(m_\rho\) over \(\sigma\)-algebra \(S\) is defined as follows:

\[
m_\rho(A) = \frac{\sum_{x \in A} \rho(x)}{\sum_{x \in X} \rho(x)}.
\]

**Definition 7.** Let \(\nu\) be a simple concept on \(X\), \(\rho_\nu : X \rightarrow R^+ = [0, \infty)\). \(\rho_\nu\) is called a weight function of simple concept \(\nu\) if \(\rho_\nu\) satisfies the following conditions:
1. \(\rho_\nu(x) = 0 \iff (x, x) \notin R_\nu, x \in X\),
2. \(\rho_\nu(x) \geq \rho_\nu(y) \iff (x, y) \in R_\nu, x, y \in X\),

where \(R_\nu\) is the binary relation of concept \(\nu\) (refer to Definition 3).

**Definition 8.** Let \(XM\) be non-empty sets. Let \(M\) be a finite set of simple concepts, \(EXM\) be \(EI^2\) algebra over \(X, M\), and \(S\) be a \(\sigma\)-algebra over \(X\). For every simple concept \(\zeta \in M\), \(m_\zeta\) is the measure defined by Definition 6 for \(\rho_\zeta\). If \(\zeta = \sum_{i \in I} a_i A_i \in EXM\) satisfies \(a_i \in S, \forall i \in I\), then \(||\zeta||\), the norm of \(\zeta\), is defined as follows:

\[
|| \sum_{i \in I} (a_i A_i) || = \sup_{i \in I} \prod_{\gamma \in A_i} m_{\rho_\gamma}(a_i).
\]

For \(a_k A_k\), \(k \in I\), if \(A_k = \emptyset\), we define

\[
\prod_{\gamma \in A_k} m_{\rho_\gamma}(a_k) = \max_{\gamma \in A_k} \{m_{\rho_\gamma}(a_k)\}.
\]

Let \((M, \tau, X)\) be an AFS structure and \(||.||\) be a norm defined by Definition 8. For any fuzzy concept \(\eta = \sum_{i \in I} A_i \in EM\), the membership function of \(\eta\) is defined as follows.

\[
\mu_\eta(x) = ||\eta(x)|| \in [0, 1], \ \forall x \in X,
\]

where \(\eta(x)\) is defined by (6).

### 3. Probability and AFS theory in concert

The key point to note is that Zadeh (1965) introduced fuzzy set theory as a mathematical construct in set theory with no intention of using it to enhance, complement, or replace probability theory. The sometimes-held perception that fuzzy set theory is a substitute for probability theory is not correct! Indeed, Zadehs article titled “Probability Measures of Fuzzy Events” [50] suggested that, his intent was to simply expand the scope of applicability of probability theory to include fuzzy sets. In essence, the AFS framework supports studies on how to convert the information in the training examples and databases to the membership functions and fuzzy logic operations. In this section we will discuss how the AFS framework take both fuzziness (subjectivity and imprecision) and randomness (objectivity and
uncertain) into account, so that uncertainty of randomness and of imprecision can be treated in a unified and coherent manner.

3.1. Probability and fuzzy set theory in concert: previous work. In [53], Zadeh used the following simple example to illustrate the deference between the subjective imprecision and objective uncertain. Suppose that Robert is three-quarters German and one-quarter French. If he were characterized as German, the characterization would be imprecise but not uncertain. Equivalently, if Robert stated that he is German, his statement would be partially true; more specially, its truth value would be 0.75. Again, 0.75 has no relation to probability. Another basic, and perhaps more serious, limitation is rooted in the fact that, in general, our assessment of probabilities is based on information which is a mixture of measurements and perceptions [46,3]. Reflecting the bounded human ability to resolve detail and store information, perceptions are intrinsically imprecise. More specially, perceptions are f-granular [51], that is: (a) perceptions are fuzzy in the sense that perceived values of variables are not sharply defined and (b) perceptions are granular in the sense that perceived values of variables are grouped into granules, with a granule being a clump of points drawn together by indistinguishability, similarity, proximity or functionality.

The earliest attempt at making probability and fuzzy set theory work in concert was made by Loginnov 1966 in [20], who interpreted the membership function as a frequentist conditional probability. Specifically, if an experiment $\varepsilon$ were to yield an outcome $x$, then $\mu_A(x)$ would be the probability that $x$ is classified as a member of $A$, that is,

$$\mu_A(x) = P(x \in A | \varepsilon \text{ yields } x)$$

For a frequentist interpretation of this probability, Loginov conceptualized an infinite-sized ensemble of membership function specifiers, each of whom has to vote on $x \in A$ or $x \in A^c$. Zadeh in [56] dismissed Loginov’s interpretation on grounds that requiring each voter to classify an observed $x$ in $A$ or $A^c$ is unnatural, because fuzzy sets reject the law of the excluded middle. We concur with this criticism provided by Zadeh. Also, because membership functions are often specified by one individual based on subjective considerations, the consensus model involving an ensemble of voters is untenable.

The second attempt at making probability theory and fuzzy set theory work in concert was made by Zadeh in [50], titled “Probability Measures of Fuzzy Events.” His construction proceeds along the following lines. Let $(\Omega, \mathcal{F}, P)$ be a “probability measure space.” Recall that $x$, an outcome of $\varepsilon$, is a member of $A$, and assume for now that $A$ is a countable crisp set, where $A \in \mathcal{F}$, and let $I_A(x)$ be the characteristic function of $A$, i.e., $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise. Then it is easy to see that

$$\mathcal{P}(A) = \sum_{x \in \Omega} I_A(x)P(x)$$

where $P(x)$ is the probability of $x$. An analog of the foregoing result when $A$ is not countable is a relationship of the form

$$\mathcal{P}(A) = \int_{\Omega} I_A(x)d\mathcal{P}(x)$$
Motivated by this (well-known) result, Zadeh has declared that the probability measure of a fuzzy subset $A$ of $\Omega$, which he calls a fuzzy event, is

$$\Pi(A) = \int_{\Omega} \mu_A(x)d(P(x)) = E(\mu_A(x))$$

where $\mu_A(x)$ is the membership function of $A$ and $E$ denotes expectation. The point to be emphasized here is that the expectation is taken with respect to the initial probability measure $P$ that has been defined on the (crisp) sets of $\Omega$. Having defined $\Pi(A)$ as before, Zadeh proceeded to show that

$$A \subseteq B \Rightarrow \Pi(A) \leq \Pi(B)$$

$$\Pi(A \cup B) = \Pi(A) + \Pi(B) - \Pi(A \cap B)$$

$$\Pi(A + B) = \Pi(A) + \Pi(B) - \Pi(A \cdot B)$$

where $A \cdot B$ is the product (not the intersection) of $A$ and $B$. Finally, $A$ and $B$ are declared to be independent if

$$\Pi(A \cdot B) = \Pi(A) \cdot \Pi(B),$$

and the conditional probability of $A$, were $B$ to occur, denoted by $\Pi(A|B)$, is defined as

$$\Pi(A|B) = \frac{\Pi(A \cdot B)}{\Pi(B)}$$

Thus when $A$ and $B$ are independent,

$$\Pi(A|B) = \Pi(A)$$

Whereas the definition (11) has the virtue that when $A$ is a crisp set, $\Pi(A) = P(A)$, so that the measure $\Pi$ can be seen as a generalization of the measure $P$, the question still remains as to whether $\Pi$ is a probability measure. Properties (13), (14), and (15) seem to suggest that $\Pi$ could indeed be viewed as a probability measure. But in [44], Singpurwalla and Booker shown that such a conclusion would be premature by the following arguments:

a) With property (14), the evaluation of $\Pi(A)$ and $\Pi(B)$ is enough to evaluate $\Pi(A \cap B)$, whereas with probability, the evaluation of $P(A)$ and $P(B)$ is not sufficient to evaluate $P(A \cap B)$, unless $A$ and $B$ are independent. Note that because

$$\Pi(A \cap B) = E(\mu_{A \cap B}(x)) = E(min(\mu_A(x), \mu_B(x))),$$

it can be easily seen that for $A = \{x : \mu_A(x) \leq \mu_B(x)\}$

$$\Pi(A \cup B) = \int_{x \in A} \mu_B dP(x) + \int_{x \in A^c} \mu_A dP(x)$$

b) Property (15) has no analog in probability, because the notions of $(A + B)$ and $(A \cdot B)$ are not part of classical set theory. More important, conditional probability has only been defined in terms of $(A \cdot B)$.

Based on the above discussions, we agree with Singpurwalla and Booker’s view that Zadeh’s [50] attempt at making fuzzy set theory and probability theory work in concert has also been unsuccessful. In what follows, we present Singpurwalla and Booker’s line of argument [44] that is able to achieve Zadeh’s goal “probability measures of fuzzy events” provided that the membership functions are given in advance. Let

$$P_D(A) = P_D(X \in A)$$
where the generic $X$ denotes the uncertain outcome of an experiment $\varepsilon$ and the subscript $D$ denotes the fact that what is being assessed is $D$’s personal probability, that is $D$’s willingness to bet. To incorporate the role of membership functions in the assessment of a probability measure of a fuzzy set $A$, Singpurwalla and Booker introduced a new character into the analysis – namely an expert, say $Z$ (in honor of Zadeh), whose expertise lies in specifying a membership function $\mu_A(x)$ for all $x \in \Omega$, and a fuzzy set $A$. Singpurwalla and Booker assume that $D$ has no access to any membership function of $A$ or a membership function $\mu_A(x)$ is given by $Z$. With the fuzzy set $A$ entering the picture, $D$ is confronted with both the imprecision and the uncertainty, i.e., about the membership of $x$ in $A$ and the other about the outcome $X = x$. If $A$ is a crisp set (as is normally the case in standard probability theory), then $D$ would be confronted with only the uncertainty, namely the uncertainty that $X = x$. As a subjectivist, $D$ views the imprecision as simply another uncertainty, and to $D$ all uncertainties can be quantified only by probability. Thus $D$ specifies two probabilities:

(a) $P_D(x)$, which is $D$’s prior probability that an outcome of $\varepsilon$ will be $x$, and 
(b) $P_D(x \in A)$, which is $D$’s prior probability that an outcome $x$ belongs to $A$.

Whereas the specification of $P_D(x)$ is a operation in standard probability theory, the assessment of $P_D(x \in A)$ raises an issue. Specifically, because $P_D(x \in A)$ is $D$’s personal probability that $x$ is classified in $A$, the question arises as to who is doing the classification and on what basis such classification is done.

Central to Singpurwalla and Booker’s development is the notion that nature is able to classify any $x$ in any set $A$ or $A^c$, with precision, so that to nature there is no such a thing as a fuzzy set; all sets are crisp. That is, to nature the membership function for any fuzzy set $A$ is of the form $\mu_A(x) = 1$ or 0. Thus fuzzy sets are only a manifestation of our uncertainty about the boundaries of sharp sets. Consequently, $P_D(x \in A)$ is merely a reflection of $D$’s uncertainty (or partial knowledge) of the boundaries of a crisp set. Nature will never reveal these boundaries, so $D$’s uncertainty of classification is impossible to ever resolve. The foregoing idealization is based on Laplace’s famous genie (see Gigerenzer et al. [11], p. 11), who knows it all and rarely tells it all. The genie is able to classify $x$ with precision, but $D$ is unsure of this classification. All the same, $D$ has partial knowledge of the genies actions, and this is encapsulated in $D$’s $P_D(x \in A)$. Laplace invoked the genie in his interpretation of probability. Like Newton, who preceded him, Laplace was a “determinist.” Thus to Laplace, probability was merely a reflection of a humans partial knowledge about the behavior of nature that is fully deterministic. To Laplace genie, there is no such thing as probability.

No matter how one chooses to interpret $P_D(x \in A)$ (In AFS theory, we interpret it as $\rho(x)$ the weight function of fuzzy concept $A$, refer to Definition 7), an assessment of this measure is essential for developing a normative approach for assessing probability measures of fuzzy sets. In introducing $P_D(x \in A)$, Singpurwalla and Booker have in fact reaffirmed Lindley’s [21] claim that probability is able to handle any situation that fuzzy logic can.

With the foregoing arguments in place, once $P_D(x)$ and $P_D(x \in A)$ have been specified by $D$ for all $x \in \Omega$, $D$ will use the law of total probability to write

$$P_D(A) = P_D(X \in A) = \sum_x P_D(X \in A|X = x)P_D(x) = \sum_x P_D(x \in A)P_D(x)$$

(16)
which is the expected value of $\mathcal{D}$’s classification probability with respect to $\mathcal{D}$’s prior probability of $X$. Thus Singpurwalla and Booker give a probability measure for a fuzzy set $A$ that can be justified on the basis of personal (i.e., subjective) probabilities and the notion that probability is a reflection of one’s partial knowledge about an event of interest. Equation (16) is based on $\mathcal{D}$’s inputs alone. The membership function $\mu_A(x)$, which is the mainstay of fuzzy set theory, has yet to play a role.

\[
P_D(X \in A|\mu_A(x)) = \sum_x P_D(x \in A|\mu_A(x), X = x)P_D(X = x|\mu_A(x)).
\]

In writing out the foregoing, $\mathcal{D}$ treats $\mu_A(x)$ as an unknown quantity, the supposition being that when $\mathcal{D}$ is contemplating his or her prior probabilities, $Z$’s response is unknown. If $\mathcal{D}$ assumes (and reasonably so) that the process by which $X$ is generated is independent of the manner by which $Z$ specifies $\mu_A(x)$; then (17) becomes

\[
P_D(X \in A|\mu_A(x)) = \sum_x P_D(x \in A|\mu_A(x))P_D(x)
\]

In writing out the foregoing, Singpurwalla and Booker have assumed that in $\mu_A(x)$, only the value at $x$ affects $P_D(x \in A|\mu_A(x))$. To proceed further, $\mathcal{D}$ must evaluate $P_D(x \in A|\mu_A(x))$, and for this $\mathcal{D}$ appeals to Bayes law; specifically,

\[
P_D(x \in A|\mu_A(x)) \propto P_D(\mu_A(x)|x \in A)P_D(x \in A)
\]

But in actuality, $\mu_A(x)$ is known to $\mathcal{D}$, as $Z$’s expert testimony. Thus the middle term becomes a likelihood, and (19) gets written as

\[
P_D(x \in A; \mu_A(x)) \propto L_D(x \in A; \mu_A(x))P_D(x \in A)
\]

Here $L_D(x \in A; \mu_A(x))$ is $\mathcal{D}$’s likelihood that $Z$ specifies $\mu_A(x)$, were nature to classify $x$ as belonging to $A$. In writing the foregoing, Singpurwalla and Booker have used the convention that $\mathcal{D}$ knows all terms on the right side of the semicolon at the time of making the relevant assessments.

If $\mathcal{D}$ wishes to adopt $Z$’s input without any modification, then $\mathcal{D}$ would set,

\[
\forall x \in \Omega,
\]

\[
L_D(x \in A; \mu_A(x)) = \mu_A(x)
\]

and then (18), rewritten to account for the fact that $\mu_A(x)$ is known to $\mathcal{D}$, becomes

\[
P_D(x \in A|\mu_A(x)) \propto \sum_x \mu_A(x)P_D(x \in A)P_D(x)
\]

This is $\mathcal{D}$’s nonnormalized probability measure of the fuzzy set $A$ given by Singpurwalla and Booker in [44].

In above, Singpurwalla and Booker discuss the sensible connection between membership functions and likelihood, and thereby probability and it is an important contribution to a better understanding of the probability measure of the fuzzy events whose membership functions are given. However, Singpurwalla and Booker have not touched the problem of how to determine membership functions for fuzzy sets based on the theory they developed in [44].
3.2. A concerted representations of imprecision and uncertainty. One of the main objectives of fuzzy logic is not only to describe how people reason, but also to develop applied systems that would make automated decisions under both imprecision and uncertainty. It is well known that we humans are not perfect: we make mistakes, we make logical errors, and our reasoning under imprecision and uncertainty is not always flawless. But we humans can balance exactness and simplicity in such a way that complexity can be reduced without oversimplification. Fuzzy sets provide a trade-off between the reality and the requirement. In this section, we apply the probability measure of the fuzzy set [44] and the AFS theory to determine membership functions and to explore how fuzzy set theory and probability theory can be made to work in concert, so that uncertainty of randomness and of imprecision can be treated in a unified and coherent manner. Compared probability theory with fuzzy set theory, the first point to note is that, like $P(A)$, the probability of a set $A$, fuzzy set theory does not tell us how to specify $\mu_A(x)$, the membership function of a fuzzy set $A$. The second point to note is that whereas there is a logical requirement that $P(A) \in [0, 1]$, the fact that $\mu_A(x) \in [0, 1]$ is simply a convenience of scaling. The third point to note is that whereas $P(A)$ can be interpreted as a two-sided bet (which in principle can be settled when $A$ reveals itself), $\mu_A(x)$ reflects an individuals view of the extent to which $x \in A$; thus $\mu_A(x)$ cannot be made operational in the same sense as $P(A)$. Finally, it is not a requirement that $\sum_{x \in X} \mu_A(x) = 1$, and thus $\mu_A(x)$ as a function of $x$ cannot be interpreted as a probability or, for that matter, as a conditional probability, as was done by Logino in [20] and also by Barrett and Woodall in [1]. How then can we interpret the membership function $\mu_A(x)$? Because $\mu_A(x)$, as a function of $x$, reflects the extent to which $x \in A$ (i.e., $\mu_A(x)$ is an indicator of how likely it is that $x \in A$), we may view $\mu_A(x)$ as the likelihood of $x$ for a fixed (i.e., specified) $A$. Recall that even though the interpretation of a likelihood is almost always derived from a probability model, the likelihood is not a probability (in particular, it does not obey the addition rule) and in statistical inference, the likelihood function reflects the relative degrees of support that a fixed observation provides to several hypotheses. Furthermore, the specification of a likelihood is subjective. Thus our interpretation of the membership function is that it is a likelihood function with $A$ taking the role of a fixed observation and the values of $x$ taking the role of the hypotheses.

Here a natural question arises: how to determine the membership functions of the fuzzy sets. Membership functions are usually given directly by the user’s subjectivity. But these membership functions cannot be used in the fuzzy observation model because they do not include the uncertainty information of the distribution of the data. In the framework of AFS theory, the membership functions are constructed by taking both fuzziness (subjective imprecision) which is represented by the weight functions of the simple fuzzy concepts and randomness (objective uncertainty) which is represented by the ratio of the data associated with the fuzzy concepts into account. The ordinary identification of a membership function is given by the direct subjectivity of specialists or decision makers depending upon the respective application examples, while the determining methods discussed in AFS theory aim to obtain a set of membership functions which coincide with the relation between the underlying distribution of the original data and the semantics of the fuzzy sets obtained from it.

When a person constructs a membership function subjectively he or she probably rely, consciously or unconsciously, on his/her knowledge base in the brain, which can
be considered as the result of a kind of statistical processing of his/her experience and learning in the past. Then we imagine that when he/she determines a grade of membership he/she gives a high value if he/she finds in his/her knowledge base many materials (the more the higher) supporting that $x$ belongs to $A$.

Throughout this section we consider the following setting for the representations of subjective imprecision and the objective randomness. There is a “probability measure space”, $(\Omega, \mathcal{F}, \mathcal{P})$ of possible instances $X$ or the observed samples over which various representations of fuzzy concepts may be defined. We assume that different instances in $X$ may be encountered with different frequencies. A convenient way to model this is to assume the probability distribution $\mathcal{P}$ defines the probability of encountering each instance in $X$ (e.g., $\mathcal{P}$ might assign a higher probability to encountering 19-year-old people than 109-year-old people). Notice $\mathcal{P}$ says nothing about the degree of $x$ belonging to a concept $\zeta$; $\mathcal{P}(x)$ only determines the probability that $x$ will be encountered. Let $\Omega$ be the universe of discourse and $F = \{f_1, f_2, \ldots, f_s\}$ be the set of all features on the objects in $\Omega$. For any $x = (v_1, v_2, \ldots, v_s) \in \Omega$, $1 \leq i \leq s$, $v_j = f_j(x)$ is the value of $x$ on the feature $f_j$. Let $M = \{m_{ij} \mid 1 \leq i \leq s, 1 \leq j \leq k_i\}$ be the set of simple concepts on $\Omega$ associating to the features. Where $m_{i1}, m_{i2}, \ldots, m_{ik_i}$ are the simple concepts, such as “small”, “medium”, “no-medium”, “large” associating to the feature $f_i$ and there are $k_i$ simple concepts on the feature $f_i$. In practice, the observed data $X$ is a finite subset of $\Omega$. Let $X = \{x_1, x_2, \ldots, x_n\}$ and $N_x$ be the number of times that $x \in X$ is observed as a sample. Thus for any $x \in X$, we have

$$\mathcal{P}(x) \approx \frac{N_x}{|X|},$$

when $|X|$ is sufficiently great. For example, if $m_{i1}, m_{i2}, m_{i3}, m_{i4}$ are the fuzzy concepts, “small”, “medium”, “no-medium”, “large” associating to the feature $f_i$ respectively, the weight functions of them can be defined according to the observed data $X$ and their semantics as follows:

$$\rho_{m_{i1}}(x_i) = \frac{h_{j1} - f_j(x_i)}{h_{j1} - h_{j2}}$$

$$\rho_{m_{i2}}(x_i) = \frac{h_{j4} - f_j(x_i) - h_{j3}}{h_{j4} - h_{j5}}$$

$$\rho_{m_{i3}}(x_i) = \frac{|f_j(x_i) - h_{j3} - h_{j5}}{h_{j4} - h_{j5}}$$

$$\rho_{m_{i4}}(x_i) = \frac{f_j(x_i) - h_{j2}}{h_{j1} - h_{j2}}$$

where $j = 1, 2, \ldots, s$,

$$h_{j1} = \max\{f_j(x_1), f_j(x_2), \ldots, f_j(x_n)\},$$

$$h_{j2} = \min\{f_j(x_1), f_j(x_2), \ldots, f_j(x_n)\},$$

$$h_{j3} = \frac{f_j(x_1) + f_j(x_2) + \cdots + f_j(x_n)}{n},$$

$$h_{j4} = \max\{|f_j(x_k) - h_{j3}| \mid k = 1, 2, \ldots, n\},$$

$$h_{j5} = \min\{|f_j(x_k) - h_{j3}| \mid k = 1, 2, \ldots, n\}.$$
simple concept $\nu$. It is obvious that for a given simple concept $\gamma$ we can define many different functions $\rho_x : X \to [0, +\infty)$ such that satisfies the weight function conditions shown in Definition 7. In order to provide a tool kit for representing and managing an infinitely complex reality, the weight functions for simple concepts are mental constructs with the subjective imprecisions (i.e., subjectively selecting one from the functions satisfying Definition 7 for the concerned simple concepts). But the constructs of the weight function for a simple concept $\gamma$ have to observe the objectivity in nature, which is the sub-preference relation $R$, objectively determined by the observed data $X$ and the semantics of $\gamma$. The multi-options of the weight functions just reflect the subjective uncertainties of the perceptions of the observed data. For each $x \in X$, $\rho_x(x)$ weights the essentiality of the inclusion of $x$ to form the extent of simple concept $\nu$. For a simple concept $A$ such as “small”, “medium”, “no-medium”, “large”, $\rho_A(x) = \mathcal{P}_D(x \in A), \forall x \in X$ (refer to (16)), is a weight function, because if $(x, y) \in R_A$, i.e., the degree of $x$ belonging to $A$ is greater or equal to that of $y$, then $\mathcal{P}_D(x \in A) \geq \mathcal{P}_D(y \in A)$ and if $(x, y) \not\in R_A$, i.e., $x$ does not belong to $A$ at all, then $\mathcal{P}_D(x \in A) = 0$. In other words, we can regard $\mathcal{P}_D(x \in A)$ in (16) as an interpretation of weight function $\rho_x(x)$. For each simple concept $\nu = \{m\} \in EM$, let

$$\mathcal{P}_D(x \in \nu) = \rho_x(x), \forall x \in X.$$  

Thus the probability measure of fuzzy simple concept $\nu$ defined by (16) is shown as

$$\mathcal{P}_D(\nu) = \sum_{x \in X} \mathcal{P}_D(x \in \nu) \mathcal{P}_D(x) = \sum_{x \in X} \rho_x(x) \mathcal{P}_D(x) = \frac{1}{|X|} \sum_{x \in X} \rho_x(x) N_x \quad \text{(because of (23))}$$

(28)

It is natural to define the probability measure of fuzzy simple concept $\nu$ on $W \subseteq X$, a subset of $X$ as follows:

$$\mathcal{P}_D(\nu : W) = \sum_{x \in W} \mathcal{P}_D(x \in \nu) \mathcal{P}_D(x) = \sum_{x \in W} \rho_x(x) \mathcal{P}_D(x) = \frac{1}{|X|} \sum_{x \in W} \rho_x(x) N_x$$

(29)

Suppose that every $m \in M$ is a simple concept on $X$. If $\tau : X \times X \to 2^M$ is defined as follows:

$$\tau(x, y) = \{m | m \in M, (x, y) \in R_m\} \in 2^M,$$

then $(M, \tau, X)$ is an AFS structure (refer to (4)). For any $x \in X$ and $A \subseteq M$, $\mathcal{A}(|x\}) = \{y \mid y \in X, \tau(x, y) \supseteq A\} \subseteq X$ (refer to (5)) is the set of all elements in $X$ whose degree of belonging to fuzzy concept $A$ is less than or equal to that of $x$. It is obvious that the greater the probability measure of $A$ on $\mathcal{A}(|x\})$ defined by (29), the greater the degree of $x$ belonging to fuzzy concept $A$, thus we can define the membership function of a molecular fuzzy concept $A \subseteq M$ as follows: for any
If each \( x \in X \) is observed just one time, i.e., \( N_x = 1 \), then (30) is
\[
\mu_A(x) = \prod_{\gamma \in A} \frac{P_D(\gamma : A_i(x))}{P_D(\gamma)} = \prod_{\gamma \in A} \frac{\sum_{x \in A_i(x)} \rho_i(x) N_x}{\sum_{x \in X} \rho_i(x) N_x} \quad (Because \ of \ (29) \ and \ (28))
\]
where \( m_{\rho_i} \) is the measure for the weight function \( \rho_i \) (refer to (7)). Thus the membership function of the molecular fuzzy concept \( A \) defined by (30) equal to that defined by (10). Furthermore, for any fuzzy concept \( \xi \in EM \), the membership function of \( \xi \) based on the probability measure of fuzzy simple concepts in \( M \) is defined as follows.

**Definition 9.** Let \((\Omega, \mathcal{F}, \mathcal{P})\) be a probability measure space and \( M \) be the set of some simple concepts on \( \Omega \). Let \( \rho_i \) be the weight function of the simple concept \( \gamma \in M \) (refer to Definition 7). \( X \subseteq \Omega \), \( X \) is a finite set of the observed samples from the probability space \((\Omega, \mathcal{F}, \mathcal{P})\). Let \((\Omega, \tau, X)\) and \((M, \tau, X)\) be the AFS structures defined as (4). If for any \( A \subseteq M \) and any \( x \in \Omega \), \( A_i(x) \in \mathcal{F} \), then for any fuzzy concept \( \xi = \sum_{i \in I} A_i \in EM \), the membership functions of \( \xi \) on the observed data \( X \) and the total space \( \Omega \) are defined as follows:

\[
\mu_{\xi}(x) = \sup_{i \in I} \prod_{\gamma \in A_i} \frac{\sum_{x \in A_i(x)} \rho_i(x) N_x}{\sum_{x \in X} \rho_i(x) N_x}, \quad \forall x \in X,
\]
\[
\mu_{\xi}(x) = \sup_{i \in I} \prod_{\gamma \in A_i} \frac{P(\gamma : A_i(x))}{P(\gamma)} = \sup_{i \in I} \prod_{\gamma \in A_i} \frac{\int_{A_i(x)} \rho_i(t) d\mathcal{P}(t)}{\int_{A_i(x)} \rho_i(t) d\mathcal{P}(t)}, \quad \forall x \in \Omega,
\]
where \( N_x \) is the number of times that \( x \) is observed as a sample.

The large number law in probability theory ensures that the membership function defined by (32) converges to the membership function defined by (33) for all \( x \in \Omega \) with the increase of the number of the observed sample set \( X \). Definition 9 not only provides a concerted representations of imprecision and uncertainty which takes both fuzziness (subjective imprecision) and randomness (objective uncertainty) into account and treats the uncertainty of randomness and of imprecision in a unified and coherent manner, but also gives a practical algorithm of determining membership functions of fuzzy concepts according to observed data or statistical technology. Specially, it opens the door to explore the deep mathematical analysis property fuzzy set theory and to a major enlargement of the role of natural languages in probability theory.

### 3.3. The demonstrations of the real world examples
In this section, by the applications of Definition 9 to real world examples, we show how the AFS and probability framework take both fuzziness (subjective imprecision) and randomness (objective uncertainty) into account. The well-known Iris dataset is provided by Fisher in 1936 [42]. The Iris data has 150 \( \times \) 4 matrix \( W = (w_{ij})_{150 \times 4} \) evenly distributed in three classes: \( C_1 \) iris-setosa, \( C_2 \) iris-versicolor, and \( C_3 \) iris-virginica. Vector of sample \( i, (w_{i1}, w_{i2}, w_{i3}, w_{i4}) \) has four features: sepal length and width,
and petal length and width (all given in centimeters). Let \( X = \{x_1, x_2, ..., x_{150}\} \) be the set of 150 observed samples form the probability space \((\Omega, \mathcal{F}, \mathcal{P})\), and \( F = \{f_1, f_2, f_3, f_4\} \) be the set of features. \( x_i = (w_{i1}, w_{i2}, w_{i3}, w_{i4}) \), \( i = 1, 2, ..., 150 \), where \( w_{i1} = f_1(x_i) \) is the sepal-length of \( x_i \), \( w_{i2} = f_2(x_i) \) is the sepal-width of \( x_i \), \( w_{i3} = f_3(x_i) \) is the petal-length of \( x_i \), \( w_{i4} = f_4(x_i) \) is the petal-width of \( x_i \). Let \( c_{ij} \) and \( \sigma_{ij} \) be the mean and the standard variance of the values of the samples in \( C_j \) on the feature \( f_i \), \( i = 1, 2, 3, 4 \), \( j = 1, 2, 3 \). By the 150 observed data matrix \( W \), we have \( c_{11} = 5.0060, c_{21} = 3.4180, c_{31} = 1.4640, c_{41} = 0.2440; c_{12} = 5.9360, c_{22} = 2.7700, c_{32} = 4.2600, c_{42} = 1.3260; c_{13} = 6.5880, c_{23} = 2.9740, c_{33} = 5.5520, c_{43} = 2.0960; \sigma_{11} = 0.3525, \sigma_{12} = 0.5162, \sigma_{13} = 0.6359, \sigma_{21} = 0.3810, \sigma_{22} = 0.3138, \sigma_{23} = 0.3225, \sigma_{31} = 0.1735, \sigma_{32} = 0.4699, \sigma_{33} = 0.5519, \sigma_{41} = 0.1072, \sigma_{42} = 0.1978, \sigma_{43} = 0.2747 \). Let \( M = \{m_{ij} | 1 \leq i \leq 4, 1 \leq j \leq 6\} \) be the set of simple concepts on the features. The semantics of the simple concepts \( m \in M \) are shown as follows:

\( m_{11} \)：“the sepal length is about \( c_{11} \)”, \( m_{1,2} \) is the negation of simple concept \( m_{11} \);  
\( m_{1,3} \)：“the sepal length is about \( c_{12} \)”, \( m_{1,4} \) is the negation of simple concept \( m_{13} \);  
\( m_{1,5} \)：“the sepal length is about \( c_{13} \)”, \( m_{1,6} \) is the negation of simple concept \( m_{15} \);  
\( m_{21} \)：“the sepal width is about \( c_{21} \)”, \( m_{2,2} \) is the negation of simple concept \( m_{21} \);  
\( m_{2,3} \)：“the sepal width is about \( c_{22} \)”, \( m_{2,4} \) is the negation of simple concept \( m_{23} \);  
\( m_{2,5} \)：“the sepal width is about \( c_{23} \)”, \( m_{2,6} \) is the negation of simple concept \( m_{25} \);  
\( m_{31} \)：“the petal length is about \( c_{31} \)”, \( m_{3,2} \) is the negation of simple concept \( m_{31} \);  
\( m_{3,3} \)：“the petal length is about \( c_{32} \)”, \( m_{3,4} \) is the negation of simple concept \( m_{33} \);  
\( m_{3,5} \)：“the petal length is about \( c_{33} \)”, \( m_{3,6} \) is the negation of simple concept \( m_{35} \);  
\( m_{41} \)：“the petal width is about \( c_{41} \)”, \( m_{4,2} \) is the negation of simple concept \( m_{41} \);  
\( m_{4,3} \)：“the petal width is about \( c_{42} \)”, \( m_{4,4} \) is the negation of simple concept \( m_{43} \);  
\( m_{4,5} \)：“the petal width is about \( c_{43} \)”, \( m_{4,6} \) is the negation of simple concept \( m_{45} \).

By Definition 4, one can verify that each \( m \in M \) is a simple concept. For any \( x, y \in X \), if \( \tau \) is defined by (4) as follows

\[ \tau(x, y) = \{m | m \in M, (x, y) \in R_m\}, \]

then \((M, \tau, X)\) is an AFS structure. Let the \( \sigma \)-algebra on \( X \) be \( S = 2^X \). For each \( m_{ij} \in M \), if

\begin{align*}
\rho_{m_{ij}}(x) &= e^{-\frac{(f_i(x)-c_{ik})^2}{2\sigma_{ik}^2}}, i = 1, 2, 3, 4, j = 2k - 1, k = 1, 2, 3, \forall x \in \Omega, \\
\rho_{m_{ij}}(x) &= 1 - \rho_{m_{i(j-1)}}(x), i = 1, 2, 3, 4, j = 2k, k = 1, 2, 3, \forall x \in \Omega,
\end{align*}

then by the semantics of each \( m \in M \) and Definition 7 we can verify that \( \rho_{m_{ij}} \) is a weight function of the simple concept \( m_{ij} \). We assume that distribution \( \mathcal{P} \) in the probability space \((\Omega, \mathcal{F}, \mathcal{P})\) is a multi-normal distribution. By the statistical technology in [4] and the observed 150 samples, the density function \( p(x) \) of \( \mathcal{P} \) is estimated as follows:

\[ p(x) = \frac{1}{3} \sum_{i=1}^{3} \frac{1}{\sqrt{(2\pi)^3 |\Sigma_i|}} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)} \]
of the above examples. Each individual person's knowledge and intelligence is based on the observed data \( X \), which is uncertainty for different individuals, and his or her perception information on \( X \) imprecisely represented by natural language. Since

\[
\mu_i = \frac{1}{50} \sum_{j=50(i-1)+1}^{50i} (w_{j1}, w_{j2}, w_{j3}, w_{j4})', \quad i = 1, 2, 3,
\]

\[
\Sigma_i = \frac{1}{50} W_i' H W_i, \quad i = 1, 2, 3,
\]

\( W_1 \) is the sub-block matrix of \( W \) selecting from 1th to 50th rows, \( W_2 \) is the sub-block matrix of \( W \) selecting from 51th to 100th rows, \( W_3 \) is the sub-block matrix of \( W \) selecting from 101th to 150th rows, \( H = I - \frac{1}{50} J \), \( J \) is a \( 50 \times 50 \) matrix whose entries are all 1. We show them detail as follows:

\[
\mu_1 = (5.0060, 3.4180, 1.4640, 0.2440)', \quad \mu_2 = (5.9360, 2.7700, 4.2600, 1.3260)', \quad \mu_3 = (6.5880, 2.9740, 5.5520, 0.2600)'.
\]

\[
\Sigma_1 = \begin{bmatrix}
0.1218 & 0.0983 & 0.0158 & 0.0103 \\
0.0983 & 0.1423 & 0.0114 & 0.0112 \\
0.0158 & 0.0114 & 0.0295 & 0.0056 \\
0.0103 & 0.0112 & 0.0056 & 0.0113
\end{bmatrix},
\]

\[
\Sigma_2 = \begin{bmatrix}
0.2611 & 0.0835 & 0.1792 & 0.0547 \\
0.0835 & 0.0965 & 0.0810 & 0.0404 \\
0.1792 & 0.0810 & 0.2164 & 0.0716 \\
0.0547 & 0.0404 & 0.0716 & 0.0383
\end{bmatrix},
\]

\[
\Sigma_3 = \begin{bmatrix}
0.3963 & 0.0919 & 0.2972 & 0.0481 \\
0.0919 & 0.1019 & 0.0700 & 0.0467 \\
0.2972 & 0.0700 & 0.2985 & 0.0478 \\
0.0481 & 0.0467 & 0.0478 & 0.0739
\end{bmatrix}
\]

Thus applying the weight functions defined by (34) and (35) and the density function given by (36) to formulas (32) and (33) in Definition 9, we can obtain the membership functions of any fuzzy concept \( x \in EM \) on both the observed data \( X \) and the total space \( \Omega \). The following Figure 1, Figure 2 and Figure 3 show the membership functions of fuzzy concepts \( \{m_{1,1}\}, \{m_{2,1}\} \) and \( \{m_{1,1}, m_{2,1}\} \) for any \( x \in X \) which are obtained by (32) and (33) respectively, denoted as observed-memb-fun (obtained by (32)) and total-memb-fun (obtained by (33)).

The about examples verify our conclusion that the membership function defined by (32) converges to the membership function defined by (33) for all \( x \in \Omega \) with increase of the number of the observed sample set \( X \), which is ensured by the large number law in probability theory. By [27], we know that \((EM, \lor, \land, ^\prime)\) is a logic system, i.e., the fuzzy concepts in \( EM \) are closed under the fuzzy logic operations \( \lor, \land, ^\prime \). Thus the membership functions and their fuzzy logic operations of the fuzzy concepts in \( EM \) are totally determined by (32) or (33) in which the weight functions \( \rho, \gamma \in M \) represent the subjective imprecisions of human perception associating to the semantics of simple concepts \( \gamma \) in natural language and the probability distribution \( \mathcal{P} \) represents the objective uncertainty of random observations of the samples. This implies that the membership function defined by Definition 9 is a concerted representations of imprecision and uncertainty which takes both fuzziness (subjective imprecision) and randomness (objective uncertainty) into account. Thus we can treat the uncertainty of randomness and of imprecision in a unified and coherent manner. In practice, the distribution \( \mathcal{P} \) in the probability space \((\Omega, \mathcal{F}, \mathcal{P})\) is unknown, but it can be estimated by the observed data \( X \subseteq \Omega \) like what we show in the above examples. Each individual person’s knowledge and intelligence is based on the observed data \( X \), which is uncertainty for different individuals, and his or her perception information on \( X \) imprecisely represented by natural language. Since
the large number law in probability theory ensures that the membership functions and their fuzzy logic operations of the fuzzy concepts in $EM$ defined by (32) based on $X$, the partially observation of total probability space $\Omega$, can approximate to those defined by (33) on the total space $\Omega$, hence the knowledge and rules discovered from the observed data $X$ and represented by the fuzzy sets defined by (32) can predict and analyze the system behavior on the total space $\Omega$.

4. Conclusions

In this paper, we propose a practical algorithm of determining membership functions and their fuzzy logic operations of fuzzy concepts according to the observed data and the statistical technology. Specially, it opens the door to explore the deep mathematical analysis properties of fuzzy set theory and to a major enlargement
The membership degree of the role of natural languages in probability theory. Probability theory ensures that the membership functions defined by (32) can approximate to the membership function defined by (33) for all $x \in \Omega$ with the increase of the number of the observed sample set $X$. Definition 9 not only provides a concerted representations of imprecision and uncertainty which takes both fuzziness (subjective imprecision) and randomness (objective and uncertainty) into account and treats the uncertainty of randomness and of imprecision in a unified and coherent manner, but also gives a practical algorithm of determining membership functions of fuzzy concepts according to the observed data and the statistical technology. Along this approach direction, more systematic studies may be carried out in view of an organic integration of the mentioned aspects within a general framework for statistical analysis based on a wider notion of Information/Uncertainty including fuzziness and its statistical treatment.

References


1 Research Center of Information and Control, Dalian University of Technology, Dalian 116024, P. R. China
2 Department of Mathematics, Dalian Maritime University, Dalian 116024, P. R. China
E-mail: xdluiros@hotmail.com