How to Fully Represent Expert Information about Imprecise Properties in a Computer System – Random Sets, Fuzzy Sets, and Beyond: An Overview

Hung T. Nguyen[a,b] and **Vladik Kreinovich**[c,*]

[a]Dept. Mathematical Sciences, New Mexico State Univ., Las Cruces, NM 88003, USA [b]Faculty of Economics, Chiang Mai University, Thailand [c]Department of Computer Science, University of Texas, El Paso, TX 79968, USA

Abstract

To help computers make better decisions, it is desirable to describe all our knowledge in computer-understandable terms. This is easy for knowledge described in terms on numerical values: we simply store the corresponding numbers in the computer. This is also easy for knowledge about precise (well-defined) properties which are either true or false for each object: we simply store the corresponding "true" and "false" values in the computer. The challenge is how to store information about *imprecise* properties. In this paper, we overview different ways to fully store the expert information about imprecise properties. We show that in the simplest case, when the only source of imprecision is disagreement between different experts, a natural way to store all the expert information is to use *random sets*; we also show how *fuzzy sets* naturally appear in such random-set representation. We then show how the random-set representation can be extended to the general ("fuzzy") case when, in addition to disagreements, experts are also unsure whether some objects satisfy certain properties or not.

Keywords

random sets; fuzzy sets; imprecise properties

1. Introduction

Need to describe properties in computer-understandable terms

In the modern world, we use computers in many important activities – to help us make decisions, to help us control different systems, etc. To make computers as helpful as possible, it is desirable to make them understand and use – as much as possible – our knowledge about different objects.

How do we describe objects? To describe an object, usually:

*Corresponding author. vladik@utep.edu.

- we list the properties that this object satisfies, and

- we describe numerical values of different quantities characterizing this property.

For example, we can describe a person as blond (*property*), tall (*property*), with blue eyes (*property*), and weighing 80 kg (*numerical value*).

Thus, to make computers understand our knowledge, we must describe properties and numerical values in computer-understandable terms. It is easy to represent numerical values: computers were originally designed to represent and process numbers. So, the remaining challenge is to represent properties.

Precise and imprecise properties: a challenge

At first glance, representing properties also seems easy:

- if an object satisfies a property, we store "true" in the corresponding place in the computer memory ("true", in the computer, is usually stored as 1); and

- if an object does not satisfy a property, we store "false" in the corresponding place in the computer memory ("false", in the computer, is usually stored as 0).

In the computer, all the information is represented as 0s and 1s, so this is also a very computer-friendly representation.

For each property P, we thus get a sequence of 0–1 values $\chi_P(x)$ describing which objects x satisfy this property and which do not. In mathematical terms, this sequence of values $\chi_P(x)$ is known as the *characteristic function* of the *set* of all the objects which satisfy the property P.

The problem with the 0–1 (set) representation is that this representation is only possible for *precise* (well-defined) properties; e.g., for a property to be taller than 180 cm. For such properties:

- once a person gets all needed information about the object, this person can uniquely decide whether this object satisfies the given property or not; and

- different people make the same decision about this property, i.e., if one person comes to the conclusion that the object satisfies the given property, then other people come to the same conclusion.

In practice, many properties such as "tall" (and even "blond") are *not* precise, in the sense that at least one of the above meaning of precision is not satisfied:

- It could be that for some objects, some people – even when given all the information – are not 100% sure whether this object satisfies the given property. For example, to most people, someone 2 m high is clearly tall, while someone 1.6 m high is clearly not tall; however, in the borderline cases, e.g., of 1.8 m, a person may be not sure whether this height indicates tallness.

- It could also be that different people have different opinions about who is tall and who is not tall. For example, to most people, 1.85 m is tall, but to a 2-m-tall basketball player, 1.85 m may not feel tall.

How can we describe such imprecise properties in a computer – in such a way that we preserve all the information that human experts have?

What we do in this paper

In this paper, we describe different ways how expert information about imprecise properties can be fully represented in a computer.

We start with a simple case when disagreement between people is the only source of imprecision. In other words, we start with the situation in which:

- each person has a definite opinion on which objects satisfy the property and which do not, and

- the only problem is that different people may have different opinions about the same property.

We will show that in this case, a full description of the expert knowledge corresponds to the mathematical notion of a *random set*. We also also show that this description naturally leads to *fuzzy sets* as first approximations to random sets.

Then, we extend our analysis to a general case when not only people disagree, they may also be unsure which objects satisfy the property and which do not.

Comment: The main objective of this overview is to provide, to a general reader, detailed convincing motivations for the use of random sets and for their relation to fuzzy sets. To the best of our knowledge, this is the first paper where all these motivations come together.

As appropriate for an overview paper, while some results are new, many of the results have already appeared earlier; for example, many definitions were already introduced in Goodman (1982).

2. Case When Every Person Is Certain, But Different People May Disagree: Enter Random Sets

Description of the case

As planned, we start with the case when for each object *x* and each property *P*, an expert is:

- either absolutely sure that the object *x* satisfies the property *P*,

- or absolutely sure that the object *x does not* satisfy the property *P*.

How to fully represent expert information in this case: first idea

Let us fix a property *P* and analyze how to represent all the information corresponding to this property. Each expert is absolutely sure which objects satisfy the property *P* and which objects do not satisfy this property. Thus, for each expert, his/her opinion about *P* can be

represented as the corresponding sequence of 0s and 1s, i.e., as the set of all objects which – according to this expert – satisfy the property P.

So, in principle, to fully store the knowledge of all the experts corresponding to the property P, it is sufficient to store the sets $S_1, \ldots, S_N$ corresponding to all N experts.

For example, in case of tallness, Expert 1 may believe that everyone of height 170 cm and above is tall, so $S_1 = [170, \infty)$. Similarly, we may have $S_2 = S_3 = [180, \infty)$, $S_4 = [200, \infty)$, etc.

Case when experts are equal

In some situations, we have expert of varying degree of expertise. In such cases, it is important of keep track of which opinion set corresponds to which expert. In such situations, the above representation – of a sequence of sets labeled by names of different experts – is a perfect full representation of the expert knowledge about the property P.

In many other cases, however, we do not have any reason to believe that different people have different degree of expertise. One needs to be an expert to classify a tumor as malignant or benign, but no one is an expert in deciding who is tall or who is young. For such commonsense terms, everyone is equally an "expert".

Since everyone has the same level of expertise, it does not matter which expert corresponds to the set S_1, and which expert corresponds to the set S_2. So, in this case, it does not matter in what order we list the sets – permutation of two sets S_i and S_j does not change the resulting knowledge.

How to deal with equally important sets

It is quite possible that two or more experts fully agree about which objects have the property P and which do not. In this case, their set S_i will coincide. (People agree with each other rather frequently, so such situations are common.)

In this case, instead of storing several copies of the same set S_i, it is easier to simply store the number of experts whose beliefs about the property P are described by this set. In other words, instead of a list of all (possibly repeating) sets, we can store a list of *non-repeating* sets S, together with the number of repetitions $n(S)$ of each set S from this list.

How to make this representation intuitively clearer

From the mathematical viewpoint, the above representation is sufficient. However, as we will see, this representation is not very intuitive. Since we want to have a computer representation that captures expert knowledge, it is a good idea to make this presentation as intuitive as possible.

The representation is not very intuitive because, e.g., if we learn that 99 experts supported some opinion set S, this information does not tell us much; is it 99 out of 100? out of 10,000? In these two different cases, the same number 99 has different meanings:

- if 99 out of 100 experts agree on the set S, this means that we, in effect, have a consensus, so it is probably safe to use the set S as a commonly agreed description of the imprecise property P;

- on the other hand, if 99 out of 10,000 expert agree, it may be the case that all other 9,901 experts agree on some other set $S' \quad S$, so this different set S' should be taken as a consensus representation of the property P.

To make the above description more intuitively clear, it is therefore reasonable to replace the original values $n(S)$ with the corresponding *frequencies* $p(S) = \dfrac{n(S)}{N}$. Thus, we arrive at the following representation of the expert knowledge about the property P being satisfied or not for objects from the set X:

- we have a list of sets $S \subseteq X$;

- to each set from this class, we assign a value $p(S) \quad 0$.

Since each expert produces some set, we have $\sum_S n(S) = N$ and hence, $\sum_S p(S) = 1$.

The above construction corresponds to the mathematical notion of a random set

The above construction is known in mathematics; to be more precise, this construction corresponds to a notion which is reasonably well known inside math (and barely known outside): the notion of a random set; see, e.g., Nguyen (2006). To understand the notion of a random set, let us first recall a more general idea of a random object, and its most well-known case – a random variable.

A *random object* means that we have several possible objects, and we have a probability assigned to each object – so that the total probability is equal to 1. For example, we may have the set of three objects: an apple, an orange, and a tomato, and we assign probabilities 0.2, 0.5, and 0.3 to these objects. In practical terms, this means that in a repeated experiment, in 20% of the cases we get an apple, in 50% of the cases we get an orange, and in the remaining 30% of the cases we get a tomato.

The usual example of random objects is a *random number*, where we have different numbers with different probabilities. For example:

- in a uniform distribution on the interval [0, 1] we get all numbers from this interval with equal probability;

- for a normal distribution with 0 mean and standard deviation 1, we get numbers between –2 and 2 with probability $\approx 90\%$, numbers from the interval [–3, 3] with the probability $\approx 99.9\%$, etc.

A slightly more complicated example is a *random vector* – which corresponds to a joint distribution of several (in general, correlated) random variables.

A *random set* is when we get different sets with different probabilities (adding up to 1) – and this is exactly what we came up with. For example, if:

- two experts out of ten think that "tall" means taller than or equal to 170 cm,

- five experts think that "tall" means taller than or equal to 180 cm, and

- the remaining three experts think that "tall" means taller than or equal to 200 cm,

then we have a random set, in which we have three possible sets $[170, \infty)$, $[180, \infty)$, and $[200, \infty)$ with probabilities, correspondingly, 0.2, 0.5, and 0.3.

Definition 1—Let X be a set. By a random set we mean a probability measure p on the set 2^X of all subsets of the set X.

Comments

- Please note that in knowledge representation, a random set is also known as a *body of evidence*, when we assign, to different sets, non-negative "masses" which add up to 1; see, e.g., Dempster (1967), Shafer (1976), Yager, Kacprzyk, and Pedrizzi (1994), Yager and Liu (2008).

- In the above example, we have random sets for which there are finitely many possible sets $S_1, \ldots, S_N$. Sometimes, it is convenient to use a *continuous* random set instead, in which we have continually many different sets. For example, if different experts select, for the property "old", thresholds which are uniformly distributed between 50 and 60, then, instead of storing all these randomly selected thresholds, it makes sense just to store the information that these the corresponding thresholds are uniformly distributed. (The corresponding description of continuous fuzzy sets is given, e.g., in Nguyen, Kreinovich, and Xiang (2008).)

3. Fuzzy Sets as a Natural First Approximation to Random Sets

We need approximations

Ideally, to describe the expert knowledge, we list all possible sets and their probabilities. In practice, we may have too many experts who have different opinions of what "tall" (or "young") means. In this case, we will have too many sets to represent. It is therefore desirable to come up with an *approximation* to a random set, an approximation which would enable us to capture the main essence of the expert opinions without having to store all the information.

How can we do that?

Let us use the general experience of approximate representation of random objects

Random sets are objects of a relatively new area of study, and there is not much experience with their various approximations. However, since random sets are a particular case of random objects, we can use the general experience of approximating random *objects*.

As we have mentioned earlier, a set $S \subseteq X$ can be described as a sequence of 0–1 truth values $\chi_S(x)$ of the statements $x \in S$ corresponding to different elements $x \in X$. In other words, a set can be viewed as a vector $(\chi_S(x_1), \chi_S(x_2), \ldots)$ of truth (0–1) values

corresponding to different elements $x_i \in X$. A random set, in this representation, is a random vector, and thus, describes a joint distribution of several 0–1 random variables $\chi_S(x_i)$.

How do we usually represent the joint distribution of several random variables v_1, v_2, …?

- In the first approximation, we provide a representation of each of the random variables, and we ignore dependence between them. In other words, we only represent *marginal* distributions of each v_i.

- In the next approximation, we take into account possible relation between pairs of random variables. In this case, we describe joint distributions of all pairs (v_i, v_j).

- In the third approximation, we describe all possible distributions of triples, etc.

The larger the size of the tuples that we take into account, the more accurate our representation. For finite sets X, when this size coincide with the size of the universal set X, we get the exact representation of the joint distribution. For infinite sets, the larger the sample size, the more accurate our representation – so that the actual distribution can be viewed as a limit of representations corresponding to tuples of different size.

Let us see how this general idea can be applied to the case of random sets.

First approximation to random sets: analysis

In the case of random sets, we have variables $v_i = \chi_S(x_i)$ corresponding to different elements $x_i \in X$. Thus, to get a natural first approximation to a random set, we describe the (marginal) probability distribution of each of these random variables.

Each variable $v_i = \chi_S(x_i)$ attains only two possible values: 0 and 1. Thus, to describe the probability distribution of this variable, it is sufficient to describe the probability $\mu_P(x) \stackrel{\text{def}}{=} p(x \in S)$ that this variable attains the value 1. Since all experts are considered equal, this probability is equal to the proportion of experts who believe that the value x satisfies the given property.

In other words, as a natural first approximation to a random set describing a property P, we can use a function which assigns, to each value $x \in X$, a proportion $\mu_P(x) \in [0, 1]$ of experts who believe that this value x satisfies the property P.

This is exactly a fuzzy set

A reader who is familiar with fuzzy sets will immediately recognize that this is one of the most widely ways to determine the values of the membership function $\mu_P(x)$ corresponding to a property P: we ask N experts, and if M of them think that the value x satisfies the property P, we take $\mu_P(x) = \dfrac{M}{N}$; see, e.g., Klir and Yuan (1995), Nguyen and Walker (2006). (This method of eliciting membership values from experts is known as *polling*.)

Thus, we arrive at a conclusion that *a natural first approximation to a random set is a fuzzy set*.

Historical comment: The above relation between random sets and fuzzy sets was first described in Goodman (1982); in this paper, the function $\mu_P(x)$ is called a *one-point coverage function*.

Definition 2—Let *P* be a random set on the set *X*. By a fuzzy set corresponding to *P*, we mean a function $\mu_P(x) \overset{\text{def}}{=} p(x \in S)$.

Mathematical comment: In precise terms, $p(x \in S)$ is the probability measure $p(\{S: x \in S\})$ of the set $\{S: x \in S\}$ of all the sets *S* that contain the value *x*.

Comment: The above argument shows that fuzzy sets can be viewed as a first approximation to random sets, but it *does not* mean that every fuzzy set is such an approximation. There are other methods of eliciting membership degrees, methods which are especially useful when we only have a single expert: e.g., we can ask the expert to mark, on a scale from 0 to some *n*, to what extent this expert agrees that *x* satisfies the property *P*. If an expert marks *m* on a scale from 0 to *n*, we take $\mu_P(x) = m/n$. For example, if an expert marks 4 on a scale from 0 to 5, we take $\mu_P(x) = 4/5 = 0.8$.

Such fuzzy sets do not come from polling and thus, cannot be naturally interpreted in terms of random sets. In one of the following sections, we describe how to take this type of fuzziness into account when describing knowledge of multiple experts.

An alternative approximation scheme

An alternative approximation scheme can be obtained by using another analogy: between a random variable and a random set. For a random variable *X*, one of the most widely used descriptions is in terms of a *cumulative distribution function* (cdf) $F(x) \overset{\text{def}}{=} p(X \le x)$. It is known that if we are given the cdf, then we can uniquely reconstruct the original probability distribution.

Once we know the probabilities $F(x) = p(X \quad x)$ of the events $X \quad x$, we can also compute the probabilities of opposite events as $G(x) \overset{\text{def}}{=} p(X \nleq x) = 1 - F(x)$. Since the real numbers are linearly ordered – i.e., for every two real numbers *x* and *y*, we have $x \quad y$ or $y \quad x$ – the "negative" condition $X \nleq x$ can be described in the equivalent "positive" form, as $X > x$. Thus, $G(x) = p(X > x)$.

The definition of the cdf is based on the fact that on the set of real numbers, there is a natural order $a \quad b$. On the class of all sets, there is also a natural ordering relation: the subset relation $A \subseteq B$. Thus, it makes sense to define, for each set *A*, the value $F(A) \overset{\text{def}}{=} p(S \subseteq A)$. In the Dempster-Shafer approach, the corresponding value $F(A)$ is known as *belief* and usually denoted by Bel(*A*). It is known that if we are given all the values Bel(*A*), then we can uniquely reconstruct the original probability distribution on the class of all sets.

Similarly to the case of a random variable, we can also define the probability of the opposite event $p(S \nsubseteq A) = 1 - \text{Bel}(A)$. In contrast to the ordering of real numbers – which is linear – the subset relation is not a linear order, we can have two sets *A* and *B* for which $A \nsubseteq B$ and *B*

$\not\subseteq A$. Thus, the description of the "negative" condition $S \not\subseteq A$ in an equivalent "positive" form is somewhat more complicated than in the case of real numbers – but still possible. Namely, one can easily check that S is not a subset of A ($S \not\subseteq A$) if and only if S has common elements with the complement $B \overset{\text{def}}{=} -A$ of the set A: $S \cap -A \quad \varnothing$. Thus, $p(S \cap A \quad \varnothing) = 1 -$ Bel(A). Such a function is also considered in the Dempster-Shafer approach. For historical reasons, it is associated with the complement $-A$, not with the set itself: a *plausibility* Pl(B) of any set B is defined as $p(S \cap B \quad \varnothing)$. Due to the above formula for $1 -$ Bel(A), we conclude that Pl(B) = $1 -$ Bel($-B$).

In this representation, to get a full representation of a random set, we need to store either the values Bel(A) for all sets $A \subseteq X$ or the values Pl(B) for all sets $B \subseteq X$. From this viewpoint, to get an approximation, we should keep only the values corresponding to the simplest sets A and B. The simplest possible nonempty sets are 1-element sets, next in complexity are 2-element sets, etc. So, the first approximation corresponds to the use of 1-point sets, the second approximation to 2-points sets, etc.

For most properties, the set S is infinite, so S cannot be a subset of a finite set A. Since the condition $S \subseteq A$ is impossible (for such finite sets A), the probability Bel(A) = $p(S \subseteq A)$ is equal to 0. Thus, the only non-zero parts of the corresponding approximation are the values Pl(B). In particular, in the first approximation, we keep the values Pl(B) = $p(S \cap B \quad \varnothing)$ corresponding to 1-element sets $B = \{x\}$. For such sets, the condition $B \cap \{x\} \quad \varnothing$ is equivalent to $x \in S$. So, the corresponding values Pl($\{x\}$) simply coincide with the probability $\mu_P(x) = p(x \in S)$.

4. From a Traditional Fuzzy Set to More Accurate Descriptions of a Random Set

Fuzzy set, what next?

As we have mentioned, it is desirable to get a representation of the expert knowledge which is as accurate as possible. In the previous section, we showed that a fuzzy set is a natural first approximation to a random set. What is the next approximation?

Why do we need to go beyond the first approximation?

While the first approximation often leads to intuitively reasonable results, sometimes the results are far from intuitive. For example, in the traditional fuzzy logic, we use an "and" - operation (t-norm) $f_\&(a, b)$ to combine the expert's degree of belief $d(A)$ and $d(B)$ in statements A and B into an estimate $f_\&(d(A), d(B))$ for the degree of belief $d(A \& B)$ that both A and B are true. Most frequently, we use $f_\&(a, b) = \min(a, b)$ or $f_\&(a, b) = a \cdot b$.

Similarly, to we use a negation operation $f_\&(a)$ (usually, $f_\&(a) = 1 - a$) to estimate the degree to which the negation $\neg A$ is true.

In many practical application, these estimates lead to reasonable results. However, in some cases, they lead to counterintuitive conclusions. For example, suppose that:

- the expert's degree of belief that a 50-year-old is old is 0.1, and

- the expert's degree of belief that a 60-year-old is old is 0.8.

What is the expert's degree of belief that 50 is old but 60 is not old? The above procedure leads to:

- d(60 is not old) $= 1 - d$(60 is old) $= 1 - 0.8 = 0.2$, and thus, to

- d((50 is old) & (60 is not old)) $= f_\&(0.1, 0.2)$.

Whether we use $f_\&(a, b) = \min(a, b)$ or $f_\&(a, b) = a \cdot b$, we get a positive degree – which makes no sense, since if an expert considers 50-year-olds to be old, then of course this expert should also consider 60-years-olds to be old.

The reason for this counterintuitive result is that the traditional fuzzy logic does not consider possible dependence between the statements. This problem does not appear if we use the full random set description: indeed, in this case,

- none of the sets S_i describing "old" would contain 50 but not 60, and so,

- the proportion p((50 is old) & (60 is not old)) of those sets for which 50 is old but 60 is not will be 0.

So, in principle, one way to avoid this problem is to use full random sets. However, as we have mentioned, this often means storing and processing too much information. Let us show that to avoid the above counter-intuitive result, we do not need to consider full random sets: it is sufficient to go one step beyond the *first* approximation (which corresponds to traditional fuzzy sets) and consider a natural *second* approximation to random sets.

What is a natural second approximation: analysis of the problem

The above analysis of random sets as a joint distribution of random variables $\chi_S(x)$ shows that a natural second approximation emerges when we consider *joint* distributions of pairs $(v, v') = (\chi_S(x), \chi_S(x'))$ of these random variables.

For each pair of objects (x, x'), since each of the two variables $\chi_S(x)$ and $\chi_S(x')$ takes only values 0 and 1, the corresponding pair $(v, v') = (\chi_S(x), \chi_S(x'))$ has *four* possible values: (0, 0), (0, 1), (1, 0), and (1, 1). So, to get a full representation of this probability distribution, we need to know four probabilities which add up to 1, i.e., we need *three* independent probabilities.

In addition to $\mu_P(x) = p(x \in S)$ and $\mu_P(x') = p(x' \in S)$, we can, e.g., consider an additional probability $\mu_{PP}(x, x') \stackrel{\text{def}}{=} p((x \in S)\&(x' \in S))$. Once we know these three probabilities, we can determine all four probabilities of different values of the pair (v, v'):

$$p((x \in S)\&(x' \notin S)) = p(x \in S) - p((x \in S)\&(x' \in S)) =$$
$$\mu_P(x) - \mu_{PP}(x, x');$$
$$p((x \notin S)\&(x' \in S)) = p(x' \in S) - p((x \in S)\&(x' \in S)) =$$
$$\mu_P(x') - \mu_{PP}(x, x');$$
$$p((x \notin S)\&(x' \notin S)) = 1 - p(x \in S) - p((x \notin S)\&(x' \in S)) =$$
$$1 - \mu_P(x) - (\mu_P(x') - \mu_{PP}(x, x')) = 1 - \mu_P(x) - \mu_P(x') + \mu_{PP}(x, x').$$

How to describe this second approximation in fuzzy terms

Using this second approximation means that:

- in addition to determining, for each $x \in X$, the proportion of experts who believe that x satisfies the property P,

- we should also determine, for all pairs of elements $x \in X$ and $x' \in X$, the proportion $\mu_{PP}(x, y)$ of experts who believe that both x and x' satisfy the desired property P.

These additional values enable us to describe, in fuzzy terms, the dependence between the membership degrees at x and x'.

Towards a formal description

In the traditional fuzzy approach, a property is described by a single function $\mu_P: X \to [0, 1]$. In the new approach, to describe a property, we need *two* functions:

- a function $\mu_P: X \to [0, 1]$, and

- a function $\mu_{PP}: X \times X \to [0, 1]$.

Since we now need two functions to describe a property, it is natural to call such pairs of functions (μ_P, μ_{PP}) *double fuzzy sets*.

Since the probability $p((x \in S) \& (x' \in S))$ does not depend on the order in which we list x and x', we should have $\mu_{PP}(x, x') = \mu_{PP}(x', x)$.

Since $p((x \in S) \& (x' \in S)) \quad p(x \in S)$, we should have $\mu_{PP}(x, x') \quad \mu_P(x)$. Thus, we arrive at the following definition.

Definition 3—Let X be a set. By a double fuzzy set, we mean a pair of functions $\mu_P: X \to [0, 1]$ and $\mu_{PP}: X \times X \to [0, 1]$ for which, for all $x, x' \in X$, we have $\mu_{PP}(x, x') = \mu_{PP}(x', x)$ and $\mu_{PP}(x, x') \quad \mu_P(x)$.

Definition 4—Let p be a random set over the set X. By a double fuzzy set corresponding to the random set p, we mean a pair consisting of the functions $\mu_P(x) = p(x \in S)$ and $\mu_{PP}(x, x') = p((x \in S) \& (x' \in S))$.

Comment: This definition also goes back to Goodman (1982), where the function $\mu_{PP}(x, x')$ is called a *two-point coverage function*.

Discussion

In the traditional fuzzy approach, we only use membership degrees $\mu_P(x)$ and $\mu_P(x')$, we do not have any additional information about the relation between x and x'. Thus, to estimate our degree of belief that x satisfies the property P *and* x' satisfies the property P, we can only use the degree $\mu_P(x)$ that x satisfies the property P and the degree $\mu_P(x')$ that x' satisfies the property P.

An operation $f_\&(a, b)$ that transforms our degrees of belief $d(A)$ and $d(B)$ in statements A and B into an estimated degree of belief $d(A \& B) \approx f_\&(d(A), d(B))$ for a composite statement A

& *B* is called an *and-operation* or a *t-norm*; see, e.g., Klir and Yuan (1995), Nguyen and Walker (2006). Thus, in the traditional fuzzy approach, the degree of belief that both *x* and *x*′ satisfy the property *P* would be estimated as $f_\&(\mu_P(x), \mu_P(x'))$.

In the new approach, instead of using this approximate description, we actually measure how many experts believe that both *x* and *x*′ satisfy the property *P*. This does not mean that we no longer have to use t-norms (or similar operations). For example:

- while, in this second approximation, we explicitly ask experts about the *pairs* of values (*x*, *x*′),

- we do not ask the experts about the *triples* (*x*, *x*′, *x*″).

Thus, to get an estimate for the expert's degree of belief that all three given elements *x*, *x*′, and *x*″ satisfy the property *P*, we still need to use a t-norm (or a similar operation): e.g., we can apply a t-norm to the values $\mu_{PP}(x, x')$ and $\mu_P(x'')$. (A more detailed analysis is given in the special section on such operations.)

What we gain by using the double fuzzy descriptions

As we have mentioned, for the property "tall", the intuitive meaning is that if a person A is tall, and B is taller than A, then B is tall too. In other words, if a value *x* satisfies this property and $x < x'$, then *x*′ also satisfies this property. Thus, for this property, each set *S* describing an expert's opinion should contain, with each element *x*, all larger elements *x*′. In mathematical terms, this means that if $x < x'$ then $\chi_S(x) \quad \chi_S(x')$, i.e., that the set characteristic function of the set *S* is *monotonic* (specifically, *non-decreasing*).

Definition 5

- We say that the set S is monotonic if $x < x'$ and $x \in S$ implies that $x' \in S$, i.e., if it is impossible to have $x \in S$, $x' \notin S$, and $x < x'$. In other words, a set is monotonic if for $x < x'$ it is impossible to have $x \in S$ and $x' \notin S$.

- We say that a random set p is monotonic if for every pair $x < x'$, the probability $p((x \in S) \& (x' \notin S))$ is equal to 0.

Comment: If for a random set, all possible sets are monotonic, then this random set is clearly monotonic as well.

The monotonicity property cannot be described in terms of the traditional fuzzy sets

Let us show that the non-decreasing property cannot be described in terms of the traditional fuzzy set.

Proposition 1—There exist random sets *p* and *p*′ such that:

- the random set p is monotonic,

- the random set p′ is not monotonic, and

- the same fuzzy set $\mu_P(x)$ corresponds to both random sets p and p′.

Proof—As p, we take a random set in which we get the set of all real numbers and the empty set with equal probability 0.5. For both these sets S, if x belongs to S and $x < x'$, then $x' \in S$. Since both these sets are monotonic, the random set is monotonic.

As p', we take a random set in which we have $(-\infty, 0)$ and $[0, \infty)$ with equal probability 0.5. In this case, for $x = -1 < x' = 0$, with probability $0.5 > 0$, we have $-1 \in S$ and $0 \notin S$. Thus, the random set p' is not monotonic.

One can easily check that both random sets lead to the same membership function $\mu_P(x) = 0.5$ for all x.

We can describe monotonicity in terms of double fuzzy sets

With double fuzzy sets, we can describe the monotonicity property in terms of the functions $\mu_P(x)$ and $\mu_{PP}(x)$ as follows:

Proposition 2—For a random set p, the following two conditions are equivalent to each other:

- the random set p is monotonic;

- if $x < x'$, then $\mu_{PP}(x, x') = \mu_P(x)$.

Proof—Let $x < x'$. We already know that $p((x \in S) \& (x' \notin S)) = \mu_P(x) - \mu_{PP}(x, x')$. Thus, if the random set p is monotonic, this implies that

$$p((x \in S) \& (x' \notin S)) = \mu_P(x) - \mu_{PP}(x, x') = 0$$

and so, that $\mu_{PP}(x, x') = \mu_P(x)$.

Vice versa, if $\mu_{PP}(x, x') = \mu_P(x)$, then $p((x \in S) \& (x' \notin S)) = \mu_P(x) - \mu_{PP}(x, x') = 0$. Since this is true for all pairs $x < x'$, this means that the random set p is monotonic.

An alternative approach to second approximation

In the above-described alternative approach, a second approximation to a fuzzy set is provided by the plausibility values $\text{Pl}(B) = p(S \cap B \neq \varnothing)$ corresponding to 1- and 2-element sets B. We already know that the plausibility values for 1-element sets $B = \{x\}$ are simply values $\mu_P(x)$. For a two-element set $B = \{x, x'\}$, the condition $S \cap \{x, x'\} \neq \varnothing$ is equivalent to $x \in S$ or $x' \in S$. Thus, what we have, in this approximation, are values $p((x \in S) \lor (x' \in S))$.

It turns out that this approximation carries exactly the same information as doubly fuzzy sets, since we always have $p(A \lor B) + p(A \& B) = p(A) = p(B)$. In particular, for $x \in S$ and $x' \in S$, we have

$$p((x \in S) \lor (x' \in S)) + p((x \in S) \& (x' \in S)) = p(x \in S) + p(x' \in S).$$

In both approximations, we know $\mu_P(x) = p(x \in S)$ and $\mu_P(x') = p(x' \in S)$. Thus:

- If we know $\mu_{PP}(x, x') = p((x \in S) \& (x' \in S))$, then we can reconstruct $p((x \in S) \vee (x' \in S))$ as $\mu_P(x) + \mu_P(x') - \mu_{PP}(x, x')$.

- If we know $p((x \in S) \vee (x' \in S))$, then we can reconstruct $\mu_{PP}(x, x')$ as $\mu_P(x) + \mu_P(x') - p((x \in S) \vee (x' \in S))$.

Is second approximation sufficient?

On the example of the property "old", the use of double fuzzy sets instead of the traditional fuzzy sets helped make our estimates more intuitive. However, as we will see, there are other cases when double fuzzy sets are not sufficient: e.g., when we want to describe a property "medium", for which if $x < x' < x''$ and x and x'' satisfy this property, then the intermediate value x' should also satisfy the same property.

Definition 6

- A subset S of the set of real numbers is called *convex* if whenever $x < x' < x''$ and both x and x'' belong to the set S, then x' also belongs to the set S. In other words, a set is convex if for $x < x' < x''$ it is impossible to have $x \in S$, $x'' \in S$, and $x' \notin S$.

- A random set p is called *convex* if for each $x < x' < x''$, the probability $p((x \in S) \& (x' \notin S) \& (x'' \in S))$ is equal to 0.

Comment: If all sets S corresponding to a random set are convex, then, as one can easily see, this random set is convex as well.

Examples—A property like "close to 0" should be convex: if $x < x' < x''$ and both x and x'' are close to 0, then it is reasonable to conclude that the intermediate value x' is also close to 0. The same is true for properties like "small".

Proposition 3—There exist random sets p and p' such that:

- the random set p is convex;

- the random set p' is not convex, and

- the same double fuzzy set $\mu_P(x)$ and $\mu_{PP}(x)$ corresponds to both random sets p and p'.

Proof—Here:

- As p, we take a random set in which we have fours sets with probability 0.25 each: $S_1 = [0, 3)$, $S_2 = [0, 1)$, $S_3 = [1, 2)$, and $S_4 = [2, 3)$. All fours sets are convex, so the random set p is also convex.

- As p', we take a random set in which we have fours sets with probability 0.25 each: $S'_1 = [0, 1) \cup [2, 3)$, $S'_2 = [0, 2)$, $S'_3 = [1, 3)$, and $S'_4 = \varnothing$. Here, for $x = 0 < x' = 1 < x'' = 2$, the corresponding probability is positive:

$$p((x \in S) \& (x^{'} \notin S) \& (x^{''} \in S)) = 0.25 > 0$$

(since this property is true for $S_1^{'}$), and thus, the random set p' is not convex.

If we divide the domain $X = [0, 3)$ into three zones $[0, 1)$, $[1, 2)$, and $[2, 3)$, then we conclude that for both random sets p and p':

- $\mu_P(x) = 0.5$ for all x and

- $\mu_{PP}(x, x') = 0.5$ when x and x' belong to the same zone and $\mu_{PP}(x, x') = 0.25$ if x and x' are in different zones.

From double to triple fuzzy sets

The above example shows that we need to go beyond double fuzzy sets. A natural next step is to consider the probabilities $\mu_{PPP}(x, x^{'}, x^{''}) \overset{\text{def}}{=} p((x \in S) \& (x^{'} \in S) \& (x^{''} \in S))$. Similarly to the case of the double sets, we arrive at the following definitions.

Definition 7—Let X be a set. By a *triple fuzzy set*, we mean a triple of functions $\mu_P: X \to [0, 1]$, $\mu_{PP}: X \times X \to [0, 1]$, and $\mu_{PPP}: X^3 \to [0, 1]$ for which, for all $x, x', x'' \in X$, we have:

- $\mu_{PP}(x, x') = \mu_{PP}(x', x)$,

- $\mu_{PPP}(x, x', x'') = \mu_{PPP}(x, x'', x') = \mu_{PPP}(x', x'', x)$, and

- $\mu_{PPP}(x, x', x'') \quad \mu_{PP}(x, x') \quad \mu_P(x)$.

Definition 8—Let p be a random set over the set X. By a *triple fuzzy set* corresponding to p, we mean a triple consisting of the functions $\mu_P(x) = p(x \in S)$, $\mu_{PP}(x, x') = p((x \in S) \& (x' \in S))$, and

$$\mu_{PPP}(x, x^{'}, x^{''}) = p((x \in S) \& (x^{'} \in S) \& (x^{''} \in S)).$$

Triple fuzzy sets can do what neither traditional not double fuzzy sets can do: namely, detect convexity

Let us show that triples fuzzy sets can do what previous approximations cannot do: detect convexity.

Proposition 4—For a random set p, the following two conditions are equivalent to each other:

- the random set p is convex;

- if $x < x' < x''$, then $\mu_{PPP}(x, x', x'') = \mu_{PP}(x, x'')$.

Proof—Let $x < x' < x''$. By definition, convexity of a random set means that $p((x \in S) \& (x' \notin S) \& (x'' \in S)) = 0$. It is easy to see that

$$p((x \in S)\&(x' \notin S)\&(x'' \in S))=p((x \in S)\&(x'' \in S))-$$
$$p((x \in S)\&(x' \in S)\&(x'' \in S))=\mu_{PP}(x,x'')-\mu_{PPP}(x,x',x'').$$

So, if the random set is convex, then $p((x \in S) \& (x' \notin S) \& (x'' \in S)) = 0$ implies that $\mu_{PP}(x, x'') - \mu_{PPP}(x, x', x'') = 0$, i.e., that $\mu_{PPP}(x, x', x'') = \mu_{PP}(x, x'')$.

Vice versa, if $\mu_{PPP}(x, x', x'') = \mu_{PP}(x, x'')$, then

$$p((x \in S)\&(x' \notin S)\&(x'' \in S))=0.$$

Since this is true for all $x < x' < x''$, this means that the random set p is indeed convex.

Tuples of arbitrary size

The traditional fuzzy logic corresponds to points x. Double and triple fuzzy logic correspond to pairs and triples. In general, for tuples of size k, we need to consider joint distributions of k random boolean (0–1) variables $v_i = \chi_S(x_i)$ corresponding to a set of k elements $s_k \stackrel{\text{def}}{=} \{x_1, \ldots, x_k\}$.

Each of these k variables v_i takes two possible values 0 and 1. Thus, the tuple $(v_1, \ldots, v_k)$ takes 2^k possible values $(0, \ldots, 0)$, $(0, \ldots, 0, 1)$, $\ldots$, $(1, \ldots, 1)$. To describe a probability distribution on such tuples, we therefore need to list the 2^k probabilities of such values – probabilities that should add up to 1. Similarly to the case of pairs, it is sufficient, for each subset $s \subseteq \{1, \ldots, k\}$, to describe the probability that $v_i = 1$ for all i from this subset.

Proposition 5—For each tuple $(\varepsilon_1, \ldots, \varepsilon_k)$ of 0s and 1s, the probability

$$p((v_1=\varepsilon_1)\&\ldots\&(v_k=\varepsilon_k))$$

if we know, for each sets $s \subseteq \{1, \ldots, k\}$, the probability $p(v_i = 1$ for all $i \in s)$.

Proof—We will prove this result by induction, by reducing, for each $z > 0$, probabilities corresponding to tuples with z zero ("false") values of ε_i to probabilities of tuples with $z - 1$ zero values. By using this reduction, we can eventually get to the probabilities of tuples in which all values ε_i are "true" – and for all these tuples, probabilities are given.

Specifically, if $\varepsilon_i = 0$, then

$$p((v_1=\varepsilon_1)\&\ldots\&(v_{i-1}=\varepsilon_{i-1})\&(v_i=0)\&(v_{i+1}=\varepsilon_{i+1})\&\ldots\&(v_k=\varepsilon_k))=$$
$$p((v_1=\varepsilon_1)\&\ldots\&(v_{i-1}=\varepsilon_{i-1})\&(v_{i+1}=\varepsilon_{i+1})\&\ldots\&(v_k=\varepsilon_k))-$$
$$p((v_1=\varepsilon_1)\&\ldots\&(v_{i-1}=\varepsilon_{i-1})\&(v_i=1)\&(v_{i+1}=\varepsilon_{i+1})\&\ldots\&(v_k=\varepsilon_k)).$$

Example—Let us show how we can use the above construction to find the probability $p((v_1 = 0)$ & $(v_1 = 0)$ & $(v_3 = 0))$. In the corresponding tuple $(0, 0, 0)$, all three values ε_i are equal to 0. Let us first use the above construction with $i = 1$, to reduce this probability to the cases when two values ε_i are equal to zero; we get

$$p((v_1=0)\&(v_2=0)\&(v_3=0))=p((v_2=0)\&(v_3=0))-p((v_1=1)\&(v_2=0)\&(v_3=0)).$$

For each of the two terms, we apply the same construction with $i = 2$, we reduce the problem to cases when only one value ε_i is equal to 0:

$$p((v_2=0)\&(v_3=0))=p(v_3=0)-p((v_2=1)\&(v_3=0));$$
$$p((v_1=1)\&(v_2=0)\&(v_3=0))=$$
$$p((v_1=1)\&(v_3=0))-p((v_1=1)\&(v_2=1)\&(v_3=0)).$$

Now, we have reduced the problem to computing four probabilities $p(v_3 = 0)$, $p((v_2 = 1)$ & $(v_3 = 0))$, $p((v_1 = 1)$ & $(v_3 = 0))$, and

$$p((v_1=1)\&(v_2=1)\&(v_3=0)).$$

For each of these four probabilities, we apply the above construction to $i = 3$ and reduce these probabilities to the known probabilities – corresponding to the cases when all the values ε_i are "true": $p(v_3 = 0) = 1 - p(v_3 = 1)$;

$$p((v_2=1)\&(v_3=0))=p(v_2=1)-p((v_2=1)\&(v_3=1));$$
$$p((v_1=1)\&(v_3=0))=p(v_1=1)-p((v_1=1)\&(v_3=1));$$
$$p((v_1=1)\&(v_2=1)\&(v_3=0))=$$
$$p((v_1=1)\&(v_2=1))-p((v_1=1)\&(v_2=1)\&(v_3=1)).$$

Substituting these expressions into the formulas for $p((v_2 = 0)$ & $(v_3 = 0))$ and $p((v_1 = 1)$ & $(v_2 = 0)$ & $(v_3 = 0))$, we conclude that

$$p((v_2=0)\&(v_3=0))=1-p(v_2=1)-p(v_3=1)+p((v_2=1)\&(v_3=1));$$
$$p((v_1=1)\&(v_2=0)\&(v_3=0))=p(v_1=1)-p((v_1=1)\&(v_2=1))-$$
$$p((v_1=1)\&(v_3=1))+p((v_1=1)\&(v_2=1)\&(v_3=1)).$$

Finally, substituting these expressions into the formula for

$$p((v_1=0)\&(v_2=0)\&(v_3=0)),$$

we get

$$p((v_1=0)\&(v_2=0)\&(v_3=0))=1-p(v_1=1)-p(v_2=1)-p(v_3=1)+$$
$$p((v_1=1)\&(v_2=1))+p((v_1=1)\&(v_3=1))+p((v_2=1)\&(v_2=1))-$$
$$p((v_1=1)\&(v_2=1)\&(v_3=1)).$$

Definition 9—Let X be a set, and let $k > 0$ be an integer. By a k-ary fuzzy set, we mean a tuple of functions $\mu_P: X \to [0, 1]$, $\mu_{PP}: X \times X \to [0, 1]$, …, $\mu_{P\ldots P} X^k \to [0, 1]$ for which, for all $x, \ldots, x' \in X$, we have:

- $\mu_{P\ldots P} (\ldots, x, \ldots, x', \ldots) = \mu_{P\ldots P} (\ldots, x', \ldots, x, \ldots)$, *and*

- $\mu_{P\ldots PP} (x, \ldots, x', x'') \quad \mu_{P\ldots P} (x, \ldots, x')$.

Definition 10—Let p be a random set over the set X and let $k > 0$ be an integer. By a k-ary fuzzy set corresponding to P, we mean a tuple consisting of the functions $\mu_{P\ldots P}(x, \ldots, x') = p((x \in S)\& \ldots \&(x' \in S))$ corresponding to $i = 1, 2, \ldots, k$ inputs.

<u>Historical comment:</u> This notion was first introduced in Goodman (1982) as *many-point coverage functions*.

<u>Comment:</u> The larger k, the more information we retain about the original random set. In the limit $k \to \infty$, we get a complete description of the random set.

Alternative approach

In the alternative approach, we store values $\text{Pl}(B) = p(S \cap B \quad \varnothing)$ corresponding to sets B with k elements. For each such set $B = \{x_1, \ldots, x_j\}$ with $j \quad k$ elements, the corresponding plausibility is equal to $p((x_1 \in S) \vee \ldots \vee (x_j \in S))$. Similarly to the case of a double fuzzy sets, one can show that this approximation is equivalent to the k-ary fuzzy sets in the sense that:

- once we know all the probabilities $p((x_1 \in S) \vee \ldots \vee (x_j \in S))$, we can uniquely reconstruct probabilities $\mu_{P\ldots P}(x_1, \ldots, x_j) = p((x_1 \in S)\& \ldots \&(x_j \in S))$;

- vice versa, once we know the probabilities

$$\mu_{P\ldots P}(x_1, \ldots, x_j) = p((x_1 \in S)\& \ldots \&(x_j \in S)),$$

we can uniquely reconstruct all the probabilities $p((x_1 \in S) \vee \ldots \vee (x_j \in S))$.

5. Case of Several Properties

Idea

What happens if we have several different properties $P^{(1)}, \ldots, P^{(\ell)}$? In situations in which each expert has a clear opinion on when each property is satisfied, each expert has a set describing each of these properties. In other words, the opinions of each expert can be described by a tuple of sets $(S^{(1)}, \ldots, S^{(\ell)})$ corresponding to different properties. In this case, to describe the opinion of several experts, we need to describe which tuples appear with what frequency, i.e., we need to describe a probability measure on the set of such tuples.

Definition 11—Let X be a set, and let $\ell > 0$ be a positive integer. By a random tuple of sets we mean a probability measure on the set $(2^X)^\ell$ of all ℓ-tuples of subsets of the set X.

First and second approximations

To describe a tuple, we need to describe boolean variables $x \in S^{(i)}$ corresponding to different elements $x \in X$ and to different properties $i = 1, \ldots, \ell$. As a first approximation, we can therefore consider the probabilities $p(x \in S^{(i)})$ of all these boolean variables. In other words, in the first approximation, we consider ℓ fuzzy sets $\mu_{P^{(i)}}(x) = p(x \in S^{(i)})$ corresponding to different properties.

In the second approximation, we need to consider probabilities of pairs of such variables. This means that, in addition to the original membership functions $\mu_{P^{(i)}}(x)$ and double membership functions $\mu_{P^{(i)}P^{(i)}}(x, x')$, we also need to consider "mixed" membership functions $\mu_{P^{(i)}P^{(j)}}(x, x') = p((x \in S^{(i)}) \& (x' \in S^{(j)}))$. These functions describe dependence between different properties.

In general, we get the following definitions.

Definition 12—Let X be a set, and let $k > 0$ and $\ell > 0$ be integers. By a k-ary fuzzy description of ℓ properties, we mean a collections of functions $\mu_{P^{(i_1)}\ldots P^{(i_s)}} : X^s \to [0, 1]$ corresponding to all possible combinations of indices $i_j \leq \ell$ with $s \leq k$ for which, for all $x, \ldots, x' \in X$, we have:

- $\mu_{\ldots P \ldots P \ldots}(\ldots, x, \ldots, x', \ldots) = \mu_{\ldots P \ldots P \ldots}(\ldots, x', \ldots, x, \ldots)$, and

- $\mu_{P^{(i_1)}\ldots P^{(i_{j-1})}P^{(i_j)}}(x, \ldots, x', x'') \leq \mu_{P^{(i_1)}\ldots P^{(i_{j-1})}}(x, \ldots, x')$.

Definition 13—Let P be a random set over the set X and let $k > 0$ and $\ell > 0$ be integers. By a k-ary fuzzy description of ℓ properties corresponding to P, we mean a collection of functions

$$\mu_{P^{(i_1)}\ldots P^{(i_j)}}(x_1, \ldots, x_j) = p((x_1 \in S^{(i_1)}) \& \ldots \& (x_j \in S^{(i_j)})).$$

Alternative approximation

For tuples of sets $(A, A', \ldots)$ and $(B, B', \ldots)$ a natural order is component-wise inclusion $(A \subseteq B) \& (A' \subseteq B') \& \ldots$ Thus, a natural analog of a cdf is the probability

$$\mathrm{Bel}(A^{(1)}, \ldots, A^{(\ell)}) \stackrel{\mathrm{def}}{=} p((S^{(1)} \subseteq A^{(1)}) \& \ldots \& (S^{(\ell)} \subseteq A^{(\ell)})).$$

The corresponding positive reformulation leads to the probability

$\mathrm{Pl}(B^{(1)}, \ldots, B^{(\ell)}) \stackrel{\mathrm{def}}{=} p((S^{(1)} \cap A^{(1)} \neq \varnothing) \vee \ldots \vee (S^{(\ell)} \cap A^{(\ell)} \neq \varnothing))$. To approximate the random set, we consider values corresponding to finite sets $A^{(i)}$ and $B^{(i)}$ in which, e.g., the total number of elements in all the sets $B^{(i)}$ does not exceed k. For such sets, the belief values

are equal to 0, and plausibility values $\text{Pl}(B^{(1)}, \ldots, B^{(\ell)})$ are equal to the probabilities $p((x_1 \in S^{(i_1)}) \,\&\, \ldots \,\&\, (x_j \in S^{(ij)}))$.

Similarly to the case of a single property, we can show that this representation is equivalent to a k-ary fuzzy description; namely:

- if we know all the probabilities

$$\mu_{P^{(i_1)}\ldots P^{(i_j)}}(x_1, \ldots, x_j) = p((x_1 \in S^{(i_1)}) \,\&\, \ldots \,\&\, (x_j \in S^{(i_j)})),$$

  then we can uniquely reconstruct the plausibility values

$$p((x_1 \in S^{(i_1)}) \vee \ldots \vee (x_j \in S^{(i_j)}));$$

- vice versa, if we know all the plausibility values

$$p((x_1 \in S^{(i_1)}) \vee \ldots \vee (x_j \in S^{(i_j)})),$$

  then we can uniquely reconstruct all the probabilities $\mu_{P^{(i_1)}\ldots P^{(ij)}}(x_1, \ldots, x_j) = p((x_1 \in S^{(i_1)}) \,\&\, \ldots \,\&\, (x_j \in S^{(ij)}))$.

6. "And"- and "Or"-Operations for Double and Triple Fuzzy Sets

"And"- and "or"-operations for traditional fuzzy sets: reminder

As we have mentioned, often, we only know the probabilities (degrees of belief) $d_1 = p(s_1)$ and $d_2 = p(s_2)$ of two statements s_1 and s_2, we have no information about their correlation, and we need to estimate the probability of $s_1 \,\&\, s_2$. In this case, we use some algorithm to transform d_1 and d_2 into an estimate for $p(s_1 \,\&\, s_2)$. Let $f_\&(a, b)$ denote the function computed by this algorithm; then, the resulting estimate for $p(A \,\&\, B)$ takes the form $f_\&(a, b) = f_\&(p(A), p(B))$. In fuzzy logic, this function $f_\&(a, b)$ is known as an "*and*"-*operation* or, alternatively, as a *t-norm*.

Similarly, when we want to estimate $d(A \vee B)$, we use an appropriate "or"- operation $f_\vee(a, b)$; such "or"-operations are also known as t-conorms.

We need similar operations for double, triple, etc. fuzzy sets

In the case of a double fuzzy set, we do not need to estimate the degree to which both x and x' satisfy a given property (or properties), since these probabilities are also given. However, we do need to estimate the triple probabilities – since such triple probabilities are not given. In general, we thus arrive at the following problem: for three statements s_1, s_2, and s_3:

- we know the probabilities $d_1 = p(s_1)$, $d_2 = p(s_2)$, $d_3 = p(s_3)$, $d_{12} = p(s_1 \,\&\, s_2)$, $d_{13} = p(s_1 \,\&\, s_3)$, and $d_{23} = p(s_2 \,\&\, s_3)$; and

- we want to estimate $x \overset{\text{def}}{=} p(s_1 \,\&\, s_2 \,\&\, s_3)$ based on the known probabilities.

For k-ary tuples, we get a similar problem of estimating the joint probability of $k + 1$ events.

Let us use the experience of the usual t-norms

To solve the above estimation problem, let us use the experience of the usual "and"-operations (t-norms).

There are many different t-norms. Since we are dealing with the probabilistic situation, in this paper, we focus of two probability-related techniques of producing t-norms: the inequality (linear programming) approach and the Maximum Entropy approach. Let us describe both approaches in detail.

Inequalities (linear programming) approach

To get a full description of the joint probability distribution on the set of two statements s_1 and s_2, we need to know the probabilities of all basic combinations $s_1 \& s_2$, $s_1 \& \neg s_2$, $\neg s_1 \& s_2$, and $\neg s_1 \& \neg s_2$. We have already shown that, once we know the probabilities $d_1 = p(s_1)$ and $d_2 = p(s_2)$ and the probability $x \stackrel{\text{def}}{=} p(s_1 \& s_2)$, we can uniquely determine all the remaining probabilities: $p(s_1 \& \neg s_2) = p(s_1) - p(s_1 \& s_2) = d_1 - x$, $p(\neg s_1 \& s_2) = p(s_2) - p(s_1 \& s_2) = d_2 - x$, and $p(\neg s_1 \& \neg s_2) = 1 - p(s_1) - p(s_2) + p(s_1 \& s_2) = 1 - d_1 - d_2 + x$.

For which values x do these formulas lead to a probability distribution? In a probability distribution, all the basic probabilities are non-negative and add up to 1. It is easy to check that the values x, $d_1 - x$, $d_2 - x$, and $1 - d_1 - d_2 + x$ always add up to 1. Thus, to make sure that the value x describes a probability distribution, it is sufficient to make sure that all fours resulting values of basic probabilities are non-negative, i.e., that the following four inequalities hold: $x \geq 0$, $d_1 - x \geq 0$, $d_2 - x \geq 0$, and $1 - d_1 - d_2 + x \geq 0$. In general, several possible value x satisfy these inequalities. It is reasonable to find the *range* of such values x, i.e., to find the smallest and the largest value x for which the above four expressions form a probability distribution.

From the mathematical viewpoint, we thus need to find the maximum and the minimum of x under the above four linear inequalities. The problem of optimizing a linear function under linear equalities and/or inequalities is known as *linear programming*; there exist efficient algorithms for solving such problems; see, e.g., Ceberio et al. (2007), Chopra (2008), Cormen et al. (2009), Nilsson (1986). In view of this relation, the above approach is also known as the linear programming approach.

For the above inequalities, we can find an explicit solution if we move x to one of the sides of each inequality and all the other terms to the other side. As a result, we get the following system of four inequalities: $x \geq 0$, $x \leq d_1$, $x \leq d_2$, and $x \geq d_1 + d_2 - 1$. The inequalities $x \leq d_1$ and $x \leq d_2$ can be described as $x \leq \min(d_1, d_2)$. Similarly, the inequalities $x \geq 0$ and $x \geq d_1 + d_2 - 1$ can be described as $x \geq \max(d_1 + d_2 - 1, 0)$. Thus, the value x determines a probability distribution if and only if $\max(d_1 + d_2 - 1, 0) \leq x \leq \min(d_1, d_2)$. We have thus found the desired range; its lower endpoint is the value $\max(d_1 + d_2 - 1, 0)$, its upper endpoint is the value $\min(d_1, d_2)$.

Both endpoints serve as possible t-norms:

- $f_\&(a, b) = \max(a + b - 1, 0)$ is the smallest possible t-norm; while

- $f_\&(a, b) = \min(a, b)$ is the largest possible t-norm; it is actually one of the most widely used t-norms.

Maximum Entropy approach

In applications of probability theory, we often encounter situations when we do not know the exact probability distribution, i.e., when several different distributions are consistent with our knowledge. Some of these distributions have smaller uncertainty, some have larger uncertainty. In this case, a reasonable idea is not to hide the possible uncertainty, i.e., to select a distribution with the largest uncertainty. There are reasonable arguments that uncertainty of a probability distribution is best described by its entropy $S = -\Sigma p_i \cdot \ln(p_i)$; as a result, we usually select a distribution with the largest entropy; see, e.g., Chokr and Kreinovich (1994), Jaynes (2003).

In the above case, we have four probabilities x, $d_1 - x$, $d_2 - x$, and $1 - d_1 - d_2 + x$, so the entropy takes the form

$$S = -x \cdot \ln(x) - (d_1 - x) \cdot \ln(d_1 - x) - (d_2 - x) \cdot \ln(d_2 - x) - (1 - d_1 - d_2 + x) \cdot \ln(1 - d_1 - d_2 + x).$$

To find the value x for which entropy is the largest, we differentiate this expression with respect to x and equate the derivative to 0. As a result, we get

$$-\ln(x) + \ln(d_1 - x) + \ln(d_2 - x) - \ln(1 - d_1 - d_2 + x) = 0.$$

Moving all negative terms to the right-hand side, we get

$$\ln(d_1 - x) + \ln(d_2 - x) = \ln(x) + \ln(1 - d_1 - d_2 + x).$$

Raising e to the power of both sides, and taking into account that $e^{a+b} = e^a \cdot e^b$ and that $e^{\ln(z)} = z$, we conclude that $(d_1 - x) \cdot (d_2 - x) = x \cdot (1 - d_1 - d_2 + x)$. Opening parentheses, we get $d_1 \cdot d_2 - x \cdot (d_1 + d_2) + x^2 = x - x \cdot (d_1 + d_2) + x^2$. Canceling similar terms in both sides, we get $x = d_1 \cdot d_2$.

The corresponding "and"-operation is indeed one of the most widely used in fuzzy logic.

"And"-operations for double fuzzy sets: inequalities approach

Let us apply the above approaches to estimate $x = p(s_1 \& s_2 \& s_3)$ for double fuzzy sets. Let us start with the inequalities approach. To fully describe the probability distribution for the case of three statements, we need to find the probabilities of all eight possible basic combinations: $p(s_1 \& s_2 \& s_3)$, $p(s_1 \& s_2 \& \neg s_3)$, $p(s_1 \& \neg s_2 \& s_3)$, $p(s_1 \& \neg s_2 \& \neg s_3)$, $p(\neg s_1 \& s_2 \& s_3)$, $p(\neg s_1 \& s_2 \& \neg s_3)$, $p(\neg s_1 \& \neg s_2 \& s_3)$, and $p(\neg s_1 \& \neg s_2 \& \neg s_3)$.

As we have shown earlier, if we know the values $d_1 = p(s_1)$, $d_2 = p(s_2)$, $d_3 = p(s_3)$, $d_{12} = p(s_1$ & $s_2)$, $d_{13} = p(s_1$ & $s_3)$, $d_{23} = p(s_2$ & $s_3)$, and $x = p(s_1$ & s_2 & $s_3)$, then we can uniquely reconstruct all remaining seven probabilities:

$$p(s_1 \& s_2 \& \neg s_3) = p(s_1 \& s_2) - p(s_1 \& s_2 \& s_3) = d_{12} - x;$$
$$p(s_1 \& \neg s_2 \& s_3) = p(s_1 \& s_3) - p(s_1 \& s_2 \& s_3) = d_{13} - x;$$
$$p(\neg s_1 \& s_2 \& s_3) = p(s_2 \& s_3) - p(s_1 \& s_2 \& s_3) = d_{23} - x;$$
$$p(s_1 \& \neg s_2 \& \neg s_3) = p(s_1) - p(s_1 \& s_2) - p(s_1 \& s_3) + p(s_1 \& s_2 \& s_3) =$$
$$d_1 - d_{12} - d_{13} + x;$$
$$p(\neg s_1 \& s_2 \& \neg s_3) = p(s_2) - p(s_1 \& s_2) - p(s_2 \& s_3) + p(s_1 \& s_2 \& s_3) =$$
$$d_2 - d_{12} - d_{23} + x;$$
$$p(\neg s_1 \& \neg s_2 \& s_3) = p(s_3) - p(s_1 \& s_3) - p(s_2 \& s_3) + p(s_1 \& s_2 \& s_3) =$$
$$d_3 - d_{13} - d_{23} + x;$$
$$p(\neg s_1 \& \neg s_2 \& \neg s_3) =$$
$$1 - p(s_1) - p(s_2) - p(s_3) + p(s_1 \& s_2) + p(s_1 \& s_3) + p(s_2 \& s_3) - p(s_1 \& s_2 \& s_3) =$$
$$1 - d_1 - d_2 - d_3 + d_{12} + d_{13} + d_{23} - x.$$

Similar to the case of two statements, these eight probabilities add up to 1, so the only requirement is that all these eight expressions are non-negative: $x \geq 0$, $d_{12} - x \geq 0$, $d_{13} - x \geq 0$, $d_{23} - x \geq 0$, $d_1 - d_{12} - d_{13} + x \geq 0$, $d_2 - d_{12} - d_{23} + x \geq 0$, $d_3 - d_{12} - d_{23} + x \geq 0$, $1 - d_1 - d_2 - d_3 + d_{12} + d_{23} + d_{13} - x \geq 0$. By moving x to one side and all other terms to another side, we get an equivalent set of inequalities: $x \geq 0$, $x \leq d_{12}$, $x \leq d_{13}$, $x \leq d_{23}$, $x \geq d_{12} + d_{13} - d_1$, $x \geq d_{12} + d_{23} - d_2$, $x \geq d_{13} + d_{23} - d_3$, and $x \leq 1 - d_1 - d_2 - d_3 + d_{12} + d_{13} + d_{23}$. These inequalities provide several lower and upper bounds for x. The value x is larger than or equal to several lower bounds if and only it is larger than or equal than the largest of these lower bounds. Similarly, the value x is smaller than or equal to several upper bounds if and only if it is smaller than or equal than the smallest of these upper bounds. Thus, the above eight inequalities are equivalent to the following inequality:

$$\max(d_{12} + d_{13} - d_1, d_{12} + d_{23} - d_2, d_{13} + d_{23} - d_3, 0) \leq x \leq$$
$$\min(d_{12}, d_{13}, d_{23}, 1 - d_1 - d_2 - d_3 + d_{12} + d_{13} + d_{23}).$$

So, we get the formulas for the lower and upper estimations for $p(s_1$ & s_2 & $s_3)$:

- we can take $\max(d_{12} + d_{13} - d_1, d_{12} + d_{23} - d_2, d_{13} + d_{23} - d_3, 0)$ as the lower estimate, and

- we can take $\min(d_{12}, d_{13}, d_{23}, 1 - d_1 - d_2 - d_3 + d_{12} + d_{13} + d_{23})$ as the upper estimate.

"And"-operations for double fuzzy sets: Maximum Entropy approach

For each value x form the corresponding range, we get a probability distribution with probabilities x, $d_{12} - x$, $d_{13} - x$, $d_{23} - x$, $d_1 - d_{12} - d_{13} + x$, $d_2 - d_{12} - d_{23} + x$, $d_3 - d_{12} - d_{23} + x$, and $1 - d_1 - d_2 - d_3 + d_{12} + d_{23} + d_{13} - x$. The entropy of this distribution is equal to

$$S = -x \cdot \ln(x) - (d_{12}-x) \cdot \ln(d_{12}-x) - (d_{13}-x) \cdot \ln(d_{13}-x) - (d_{23}-x) \cdot \ln(d_{23}-x) - $$
$$(d_1-d_{12}-d_{13}+x) \cdot \ln(d_1-d_{12}-d_{13}+x) - (d_2-d_{12}-d_{23}+x) \cdot \ln(d_2-d_{12}-d_{23}+x) - $$
$$(d_3-d_{13}-d_{23}+x) \cdot \ln(d_3-d_{13}-d_{23}+x) - $$
$$(1-d_1-d_2-d_3+d_{12}+d_{23}+d_{13}-x) \cdot \ln(1-d_1-d_2-d_3+d_{12}+d_{23}+d_{13}-x).$$

Differentiating this expression with respect to x and equating the derivative to 0, we conclude that

$$-\ln(x)+\ln(d_{12}-x)+\ln(d_{13}-x)+\ln(d_{23}-x)-$$
$$\ln(d_1-d_{12}-d_{13}+x)-\ln(d_2-d_{12}-d_{23}+x)-\ln(d_3-d_{13}-d_{23}+x)+$$
$$\ln(1-d_1-d_2-d_3+d_{12}+d_{23}+d_{13}-x)=0.$$

If we raise e to the power of both sides, we get a 4-th order equation (actually 3rd order since terms x^4 cancel out). In this case, however, we do not have a solution in a nice close form, we need to use numerical methods to solve this equation.

7. General Case, When Experts Are Not Necessarily 100% Certain About Their Statements

General case: reminder

In the previous sections, we considered the case when each expert is absolute sure which object satisfies the given property and which object does not (e.g., who is tall and who is not tall), and the only uncertainty comes from the fact that different experts may have different opinions. In practice, experts are often not absolutely certain about their judgments. So, we need to take into account their degree of certainty.

Individual degrees of certainty

To describe the corresponding degree of certainty, we can, e.g., ask the experts to mark their certainty by selecting a mark on a scale from 0 to some positive integer n, so that n corresponds to full certainty and 0 corresponds to no certainty. If an expert marks m on a scale from 0 to n, we take the ratio m/n as the expert's degree of certainty in the given statement.

An alternative idea: subjective probabilities

Some people can easily mark their uncertainty on a scale, but for other people, this is a difficult task. To get the information about the degree of certainty of these people, we can use subjective probabilities.

To get the main idea behind such probabilities, it is important to recall why we are storing this imprecise knowledge like "small" or "old" – because we want to help computer emulate human decisions, and humans describe their decision making by using such imprecise terms. So, a natural way to describe to what extend, e.g., a 50-year-old is old is to elicit, from an expert, the subjective probability that, e.g., a medical treatment which is efficient for old

people will be successful for a 50-year-old. This subjective probability can be obtained, e.g., by asking the expert to select between the following two situations:

- the situation L_0, in which the expert wins a certain some of money (e.g., \$100) if a medical treatment which is efficient for old people succeeds for a randomly selected 50-year-old patient; and

- the situation $L(p)$, in which the expert wins the same some of money with probability p.

Clearly, for an expert, the alternative $L(0)$ in which he or she never gets any money is worse than the alternative L_0 in which an expert has a chance to win some money; we will denote this by $L(0) < L_0$. Similarly, the alternative $L(1)$ in which the expert unconditionally gets the sum is preferable to the alternative L_0 in which there is a chance that the expert will get nothing: $L_0 < L(1)$. A *subjective probability* is defined as the probability p for which, to the expert, the corresponding alternative $L(p)$ is equivalent to L_0: $L(p) \sim L_0$.

This value p can be found by *bisection*. At each stage of the bisection procedure, we maintain an interval $[\underline{p}, \overline{p}]$ that contains the desired probability p, i.e., for which $L(\underline{p}) < L_0 < L(\overline{p})$. In the beginning, we take $\underline{p} = 0$ and $\overline{p} = 1$. On each iteration step, we compute the midpoint $\tilde{p} = (\underline{p} + \overline{p})/2$ and ask the expert to compare L_0 with the alternative $L(\tilde{p})$ corresponding to this midpoint. Depending on the result of this comparison, we do the following:

- if $L(\tilde{p}) \sim L_0$, we have found the desired subjective probability, it is $\tilde{p}$;

- if $L_0 < L(\tilde{p})$, we can take $\tilde{p}$ as the new value of the upper bound $\overline{p}$;

- if $L(\tilde{p}) < L_0$, we can take $\tilde{p}$ as the new value of the lower bound $\underline{p}$.

In all these cases, we either find the value of the subjective probability, or divide the width of the interval $[\underline{p}, \overline{p}]$ in half. We started with an interval of width 1. Thus, in m steps, we get an interval of width 2^{-m} which contains the desired value of the subjective probability – and so, e.g., the midpoint of this interval approximates the subjective probability with accuracy $2^{-(m+1)}$. For example:

- to find the subjective probability $d(x)$ with accuracy 10%, it is sufficient to make three iterations, i.e., to ask the expert to make three comparisons;

- to get the accuracy of 1%, it is sufficient to perform six iterations, i.e., to ask the expert to make six comparisons, etc.

Resulting representation of expert knowledge

For each property P, for each expert, and for each possible value x of the corresponding quantity, the expert has a certain degree of certainty $d(x)$ that the value x satisfies the given property P. Thus, the opinion of an *individual* expert about the property P can be described by a function which assigns, to each x, the corresponding degree $d(x)$.

In fuzzy set theory, this function is called a *membership function*. In these terms, to describe the opinions of *all* the experts, we need to store the membership functions corresponding to *all* the experts.

What if experts are equal

Some experts agree between themselves; as a result, their membership functions coincide. When the experts are equal, there is no need to store several identical copies of the same membership function, it is sufficient to store different membership functions d, together with the frequency $p(d)$ with which different membership functions occur.

When we have different numerical values with different probabilities, this is called a random variable. When we have different sets with different probabilities, this is called a random set. In our case, we have different membership functions (fuzzy sets) with different probabilities, this is a *random function* (also known a a *stochastic process*) or, a *random fuzzy set*.

How to approximate expert knowledge

In the above simple case, when each expert is absolutely sure whether each object satisfies each property or not, we have mentioned that it often takes too much space to store (and too much time to process) all the truth value $\chi_P(x)$ corresponding to all experts and to all values x. So, instead of the exact representation, we need an approximate representation of random sets.

In the general case, for each expert and for each value x, we need to store not just one bit ("true"-"false", 0–1 value), we need to store the entire real number $d(x) \in [0, 1]$. In the general case, we therefore also need to use some approximations.

Similar to the simple case, we will approximate the general probability measure by marginal distributions, i.e., in this case, by:

- distributions of $d(x)$ corresponding to each x,

- distributions of pairs $(d(x), d(x'))$ corresponding to pairs (x, x'),

- distributions of triples $(d(x), d(x'), d(x''))$ corresponding to triples (x, x', x''),

- in general, distributions of k-tuples $(d(x_1), \ldots, d(x_k))$ corresponding to k-tuples $(x_1, \ldots, x_k)$.

In the simple case, this approximation was sufficient, since, e.g., for each x, to get a full description of the probabilities of different values of $\chi_P(x)$, it is sufficient to provide a single probability $\mu_P(x) = p(\chi_P(x) = 1)$. In the general case, even for a single variable $d(x)$, we fully describe its probability distribution, we need to describe, e.g., its cumulative distribution function $F(z) = p(d(x) \quad z)$ – and to represent this function exactly, we need to describe the values $F(z)$ corresponding to many values z. Thus, in the general case, we need to approximate each such marginal distribution as well.

A natural way to describe a probability distribution is to describe its moments. In the first approximation, we represent the expert knowledge by storing all first moments; in the

second approximation, we also store all second moments, etc. Let us describe this idea in more detail.

First approximation

In the first approximation, we represent only the first moments $\mu_P(x) \overset{\text{def}}{=} E[d(x)]$, i.e., the values of the membership function averaged over all the experts. In this representation, we ignore the variations between the opinions of different experts, and only use the averages.

This representation corresponds to the traditional fuzzy logic.

Second approximation

In the second approximation, in addition to the means $\mu_P(x) = E[d(x)]$, we also store the second moments $E[d(x) \cdot d(x')]$. It is known that describing the second moments is equivalent to describing:

- the variance $V(x) = \sigma^2(x) \overset{\text{def}}{=} E[(d(x) - \mu_P(x))^2]$ that gauges the difference in expert's opinions, and

- the covariance $C(x, x') \overset{\text{def}}{=} E[(d(x) - \mu_P(x)) \cdot (d(x') - \mu_P(x'))]$ that describes the dependence between the expert's opinions about different values $x, x' \in X$.

The idea behind the standard deviation $\sigma(x)$ is similar to the idea of type-2 fuzzy logic (see, e.g., Mendel (2001), Mendel and Wu (2006)), which also takes into account how different are opinions of different experts. The covariance, however, captures the dependence which is not captured by the type-2 fuzzy set approach.

Case of several properties

A similar description can be used when we have several properties $P^{(1)}, \ldots, P(\ell)$. In this case, for each property $P^{(i)}$, we get, from each expert, the corresponding individual membership function $d^{(i)}(x)$. In the first approximation, we use only the first moments, i.e., we take the average membership functions $\mu^{P^{(i)}}(x) = E[d^{(i)}(x)]$.

In the second approximation, in addition to these averages, and to variances $V_P(x) = (\sigma_P(x))^2 = E[(d_P(x) - \mu_P(x))^2]$ and covariances $C_{PP}(x, x') = E[(d_P(x) - \mu_P(x)) \cdot (d_P(x') - \mu_P(x'))]$ corresponding to each property P, we also store covariances describing dependence between different properties: $C_{PP'}(x, x') \overset{\text{def}}{=} E[(d_P(x) - \mu_P(x)) \cdot (d_{P'}(x') - \mu_{P'}(x'))]$.

Example: "and"- and "or"-operations in the first and second approximations

Let us assume that we use the algebraic product $f_\&(a, b) = a \cdot b$ as an "and"-operation. In this case, if we know the exact expert's degree of certainty a and b in statements A and B, then we estimate the expert's degree of certainty in $A \& B$ as $a \cdot b$.

In the first approximation, instead of the exact degrees a and b, we know the means $\mu(A) = E[a]$ and $\mu(B) = E[b]$ of both degrees. We want to estimate $\mu(A \& B) = E[a \cdot b]$. Strictly speaking, we do not have enough information to get an exact estimate for this quantity, since

the exact computation would require that, in addition to the means $E[a]$ and $E[b]$, we also know the covariance $C = E[(a − \mu(A)) \cdot (b − \mu(B)] = E[a \cdot b] − \mu(A) \cdot \mu(B)$. If we knew the covariance C, then we would be able to get the exact value $E[a \cdot b] = \mu(A) \cdot \mu(B)+C$. This covariance corresponds to the second approximation, so, in the first approximation, it can be safely ignored. Thus, in the first approximation, we estimate $\mu(A \& B) = E[a \cdot b]$ as $\mu(A) \cdot \mu(B)$.

Similarly, if we use the "or"-operation $f_\vee(a, b) = a + b − a \cdot b$, we estimate $\mu(A \vee B) = E[a + b − a \cdot b] = E[a] + E[b] − E[a \cdot b]$ as $\mu(A) + \mu(B) − \mu(A) \cdot \mu(B)$. In other words, in the first approximation, "and"- and "or"-operations are the same as in the traditional fuzzy logic.

In the second approximation, the situation differs since we already know the covariance $C = E[(a − \mu(A)) \cdot (b − \mu(B))] = E[a \cdot b] − \mu(A) \cdot \mu(B)$. In this case, we get an exact value of $\mu(A \& B) = E[a \cdot b] = \mu(A) \cdot \mu(B) + C$ and, correspondingly, the exact value of $\mu(A \vee B) = E[a] +E[b] − E[a \cdot b] = \mu(A)+\mu(B) − \mu(A) \cdot \mu(B) − C$.

However, since we are in the second approximation, it is now not enough to estimate the values $\mu(A \& B)$ and $\mu(A \vee B)$, we also need to estimate the corresponding standard deviations $\sigma[A \& B]$ and $\sigma[A \vee B]$. Here, $\sigma^2[A \& B] = E[(a \cdot b)^2] − \mu^2(A \& B)$. The first term is equal to $E[a^2] \cdot E[b^2] + c$, where c is the covariance between a^2 and b^2. This covariance is a fourth-order term, so in the second approximation, it can be ignored. Since $E[a^2] = \mu^2(A) +\sigma^2(A)$ and $E[b^2] = \mu^2(B)+\sigma^2[B]$, we conclude that $\sigma^2[A \& B] = (\mu^2(A)+\sigma^2(A)) \cdot (\mu^2(B) +\sigma^2(B)) − (\mu(A) \cdot \mu(B)+C)^2$. Opening parentheses and canceling the terms $\mu^2(A) \cdot \mu^2(B)$ and $− \mu^2(A) \cdot \mu^2(B)$, we get

$$\sigma^2[A\&B]=\mu^2(A)\cdot\sigma^2(B)+\mu^2(B)\cdot\sigma^2(A)+\sigma^2(A)\cdot\sigma^2(B)-2\cdot\mu(A)\cdot\mu(B)\cdot C-C^2.$$

For $A \vee B$, from the fact that $a + b − a \cdot b = 1 − (1 − a) \cdot (1 − b)$ and that $E[1 − a] = 1 − E[a]$ and $E[1 − b] = 1 − E[b]$, we conclude that $\sigma^2[A \vee B] = ((1 − \mu(A))^2 + \sigma^2(A)) \cdot ((1 − \mu(B))^2 + \sigma^2[B]) − ((1 − \mu(A)) \cdot (1 − \mu(B)) + C)^2$, and

$$\sigma^2[A \vee B]=(1-\mu(A))^2 \cdot \sigma^2(B)+(1-\mu(B))^2 \cdot \sigma^2(A)+\sigma^2(A) \cdot \sigma^2(B)- \\ 2 \cdot (1-\mu(A)) \cdot (1-\mu(B)) \cdot C-C^2.$$

Third and higher order approximations

In the third and higher order approximations, in addition to the first and second moments, we also store third and higher order moments.

How to take into account that an expert is often uncertain about his or her degrees of belief

In the beginning of this section, we assumed that the expert can always meaningfully describe his or her degree of belief by a number from 0 to n. Sometimes, however, an expert is not sure about his or her degree of belief. For example, instead of selecting a single value

(such as 6, 7, or 8) on a scale from 0 to 10, the expert selects the whole interval [6, 8]. In such situations, for each property P, for each expert, and for each possible value x, instead of a single value $d(x)$, we have an interval [$\underline{d}(x)$, $\overline{d}(x)$] which describes the expert's degree of certainty that the value x satisfies the property P.

Thus, for each individual expert, we have an *interval-valued membership function* (see, e.g., Mendel (2001), Mendel and Wu (2006)). To describe the opinion of all the experts, we need to describe a probability measure on the set of all such functions, i.e., we need to describe a *random interval-valued fuzzy set*. We can approximate this general description by storing moments corresponding to $\underline{d}(x)$ and $\overline{d}(x)$.

We can similarly more complex descriptions of the individual expert's uncertainty: e.g., in addition to marking an interval, the expert can also describe, for each point from this interval, he or she is sure that this mark reflects his/her degree of confidence. In this case, each individual membership function is itself a type-2 membership function: to each possible value $d = m/n$ of expert's degree of confidence, we assign a value $d_2(d)$ describing the degree to which d is a possible value.

8. Conclusion

To adequately represent and process expert knowledge, we need, in particular, to represent expert information about imprecise properties. In this paper, we show that the need to represent such information naturally leads to *random sets*.

Representing a general random set is, however, computationally taxing, so we need to use computationally efficient approximations to general random sets. We show that a natural first approximation is equivalent to a *fuzzy set*.

We also describe reasonable second, third, etc. approximations – which correspond to "double", "triple", etc. fuzzy sets. We show how "and" and "or"-operations (t-norms and t-conorms) can be naturally extended from the usual fuzzy sets to such "double", "triple", etc. fuzzy sets.

We also show how the relation between random sets and fuzzy sets can be extended to interval-valued and more general type-2 fuzzy sets.

Acknowledgments

References

Ceberio M, Kreinovich V, Chopra S, Longpré L, Nguyen HT, Ludäscher B, Baral C. Interval-type and affine arithmetic-type techniques for handling uncertainty in expert systems. Journal of Computational and Applied Mathematics. 2007; 199(2):403–410.

Chokr, B.; Kreinovich, V. How far are we from complete knowledge: complexity of knowledge acquisition in Dempster-Shafer approach. In: Yager, RR.; Kacprzyk, J.; Pedrizzi, M., editors. Advances in the Dempster-Shafer Theory of Evidence. New York: Wiley; 1994. p. 555-576.

Chopra S. Affine arithmetic-type techniques for handling uncertainty in expert systems. International Journal of Intelligent Technologies and Applied Statistics. 2008; 1(1):59–110.

Cormen, ThH; Leiserson, CE.; Rivest, RL.; Stein, S. Introduction to Algorithms. Cambridge, MA: MIT Press; 2009.

Dempster AP. Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics. 1967; 38:325–339.

Goodman, IR. Fuzzy sets as equivalence classes of random sets. In: Yager, R., et al., editors. Fuzzy Sets and Possibility Theory. Oxford, UK: Pergamon Press; 1982. p. 327-432.

Jaynes, ET. Probability Theory: The Logic of Science. Cambridge, Massachusetts: Cambridge University Press; 2003.

Klir, G.; Yuan, B. Fuzzy Sets and Fuzzy Logic: Theory and Applications. Upper Saddle River, NJ: Prentice Hall; 1995.

Mendel, JM. Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions. Upper Saddle River, New Jersey: Prentice-Hall; 2001.

Mendel, JM.; Wu, D. Perceptual Computing: Aiding People in Making Subjective Judgments. New York: IEEE Press and Wiley; 2010.

Nguyen, HT. An Introduction to Random Sets. Boca Raton, Florida: Chapman and Hall/CRC; 2006.

Nguyen, HT.; Kreinovich, V.; Xiang, G. Random fuzzy sets. In: Wang, H-F., editor. Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery. Hershey, Pennsylvania: IGI Global; 2008. p. 18-44.

Nguyen, HT.; Walker, EA. First Course in Fuzzy Logic. Boca Raton, FL: CRC Press; 2006.

Nilsson NJ. Probabilistic logic. Artifical Intelligence. 1986; 28(1):71–87.

Shafer, G. A Mathematical Theory of Evidence. Princeton, New Jersey: Princeton University Press; 1976.

Yager, RR.; Kacprzyk, J.; Pedrizzi, M., editors. Advances in the Dempster-Shafer Theory of Evidence. New York: Wiley; 1994.

Yager, R.; Liu, L., editors. Classical Works of the Demspter-Shafer Theory of Belief Functions. Berlin, Heidelberg, New York: Springer-Verlag; 2008.