Likert-Scale Fuzzy Uncertainty from a Traditional Decision Making Viewpoint: It Incorporates Both Subjective Probabilities and Utility Information

Joe Lorkowski and Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
lorkowski@computer.org, vladik@utep.edu

***Abstract*—One of the main methods for eliciting the values of the membership function $\mu(x)$ is to use the Likert scales, i.e., to ask the user to mark his or her degree of certainty by an appropriate mark k on a scale from 0 to n and take $\mu(x) = k/n$. In this paper, we show how to describe this process in terms of the traditional decision making. Our conclusion is that the resulting membership degrees incorporate both probability and utility information. It is therefore not surprising that fuzzy techniques often work better than probabilistic techniques – which only take into account the probability of different outcomes.**

I. Introduction

Fuzzy uncertainty: a usual description. Fuzzy logic (see, e.g., [4], [6], [8]) has been designed to describe imprecise ("fuzzy") natural language properties like "big", "small", etc. In contrast to "crisp" properties like $x \le 10$ which are either true or false, experts are not 100% sure whether a given value x is big or small. To describe such properties P, fuzzy logic proposes to assign, to each possible value x, a degree $\mu_P(x)$ to which the value x satisfies this property:

- the degree $\mu_P(x) = 1$ means that we are absolutely sure that the value x satisfies the property P;
- the degree $\mu_P(x) = 0$ means that we are absolutely sure that the value x does not satisfy the property P; and
- intermediate degrees $0 < \mu_P(x) < 1$ mean that we have *some* confidence that x satisfies the property P but we also have a certain degree of confidence that the value x does not satisfy this property.

How do we elicit the degree $\mu_P(x)$ from the expert? One of the usual ways is to use a *Likert scale*, i.e., to ask the expert to mark his or her degree of confidence that the value x satisfies the property P by one of the marks $0, 1, \ldots, n$ on a scale from 0 to n. If an expert marks m on a scale from 0 to n, then we take the ratio m/n as the desired degree $\mu_P(x)$. For example, if an expert marks her confidence by a value 7 on a scale from 0 to 10, then we take $\mu_P(x) = 7/10$.

For a fixed scale from 0 to n, we only get $n+1$ values this way: 0, $1/n$, $2/n$, …, $(n-1)/n = 1 - 1/n$, and 1. If we want a more detailed description of the expert's uncertainty, we can use a more detailed scale, with a larger value n.

Traditional decision making theory: a brief reminder. Decision making has been analyzed for decades. Efficient models have been developed and tested to describe human decision making, and the resulting tools have been effectively used in business and in other decision areas; see, e.g., [1], [2], [3], [5], [7]. These models are not perfect – this is one the reasons why fuzzy methods are needed – but these tools provide a reasonable first-order approximation description of human decision making.

Need to combine fuzzy techniques and traditional decision making techniques, and the resulting problem that we solve in this paper. Traditional decision making tools are useful but have their limitations. Fuzzy tools are also known to be very useful, in particular, they are known to be useful in control and in decision making (see, e.g., [4], [6]), so a natural idea is to combine these two techniques.

To enhance this combination, it is desirable to be able to describe both techniques in the same terms. In particular, it is desirable to describe fuzzy uncertainty in terms of traditional decision making. To the best of our knowledge, this has not been done before; we hope that our description will lead to useful applications in practical decision making.

Structure of the paper. In our opinion, one of the main reasons why such a description has not been proposed earlier is that there is a lot of confusion and misunderstanding about such basic notions of traditional decision theory as utility, subjective probability, etc. – just like many decision making researchers have misunderstandings about fuzzy techniques. Because of this, we start, in Section 2, with providing a brief overview of the traditional decision theory and its main concepts. In Section 3, we show how the Likert scale selection can be described in these terms. The resulting straightforward description leads to rather complicated optimization problems, and how to solve these optimization problems. As a result, we get an explicit expression that describes the membership values $\mu_P(x)$ in terms of utility and subjective probability. The last section provides conclusion and future work.

II. Traditional Decision Theory and Its Main Concepts: A Brief Overview

Main assumption behind the traditional decision theory. Traditional approach to decision making is based on an assumption that for each two alternatives A' and A'', a user can tell:

- whether the first alternative is better for him/her; we will denote this by $A'' < A'$;
- or the second alternative is better; we will denote this by $A' < A''$;
- or the two given alternatives are of equal value to the user; we will denote this by $A' = A''$.

Towards a numerical description of preferences: the notion of utility. Under the above assumption, we can form a natural numerical scale for describing preferences. Namely, let us select a very bad alternative A_0 and a very good alternative A_1. Then, most other alternatives are better than A_0 but worse than A_1.

For every probability $p \in [0, 1]$, we can form a lottery $L(p)$ in which we get A_1 with probability p and A_0 with probability $1 - p$.

- When $p = 0$, this lottery coincides with the alternative A_0: $L(0) = A_0$.
- When $p = 1$, this lottery coincides with the alternative A_1: $L(1) = A_1$.

For values p between 0 and 1, the lottery is better than A_0 and worse than A_1. The larger the probability p of the positive outcome increases, the better the result:

$$p' < p'' \text{ implies } L(p') < L(p'').$$

Thus, we have a continuous scale of alternatives $L(p)$ that monotonically goes from $L(0) = A_0$ to $L(1) = A_1$. We will use this scale to gauge the attractiveness of each alternative A.

Due to the above monotonicity, when p increases, we first have $L(p) < A$, then we have $L(p) > A$, and there is a threshold separating values p for which $L(p) < A$ from the values p for which $L(p) > A$. This threshold value is called the *utility* of the alternative A:

$$u(A) \stackrel{\text{def}}{=} \sup\{p : L(p) < A\} = \inf\{p : L(p) > A\}.$$

Then, for every $\varepsilon > 0$, we have

$$L(u(A) - \varepsilon) < A < L(u(A) + \varepsilon).$$

We will describe such (almost) equivalence by $\equiv$, i.e., we will write that $A \equiv L(u(A))$.

How to elicit the utility from a user: a fast iterative process. Initially, we know the values $\underline{u} = 0$ and $\overline{u} = 1$ such that $A \equiv L(u(A))$ for some $u(A) \in [\underline{u}, \overline{u}]$.

On each stage of this iterative process, once we know values $\underline{u}$ and $\overline{u}$ for which $u(A) \in [\underline{u}, \overline{u}]$, we compute the midpoint u_{mid} of the interval $[\underline{u}, \overline{u}]$ and ask the user to compare A with the lottery $L(u_{\text{mid}})$ corresponding to this midpoint. There are two possible outcomes of this comparison: $A \le L(u_{\text{mid}})$ and $L(u_{\text{mid}}) \le A$.

- In the first case, the comparison $A \le L(u_{\text{mid}})$ means that $u(A) \le u_{\text{mid}}$, so we can conclude that $u \in [\underline{u}, u_{\text{mid}}]$.
- In the second case, the comparison $L(u_{\text{mid}}) \le A$ means that $u_{\text{mid}} \le u(A)$, so we can conclude that $u \in [u_{\text{mid}}, \overline{u}]$.

In both cases, after an iteration, we decrease the width of the interval $[\underline{u}, \overline{u}]$ by half. So, after k iterations, we get an interval of width 2^{-k} which contains $u(A)$ – i.e., we get $u(A)$ with accuracy 2^{-k}.

How to make a decision based on utility values. Suppose that we have found the utilities $u(A')$, $u(A'')$, …, of the alternatives A', A'', …Which of these alternatives should we choose?

By definition of utility, we have:

- $A \equiv L(u(A))$ for every alternative A, and
- $L(p') < L(p'')$ if and only if $p' < p''$.

We can thus conclude that A' is preferable to A'' if and only if $u(A') > u(A'')$. In other words, we should always select an alternative with the largest possible value of utility. So, to find the best solution, we must solve the corresponding optimization problem.

Before we go further: caution. We are *not* claiming that people estimate probabilities when they make decisions: we know they often don't. Our claim is that when people make *definite* and *consistent* choices, these choices *can* be described by probabilities. (Similarly, a falling rock does not solve equations but follows Newton's equations $ma = m\dfrac{d^2x}{dt^2} = -mg$.) In practice, decisions are often *not* definite (uncertain) and *not* consistent.

How to estimate utility of an action. For each action, we usually know possible outcomes $S_1, \ldots, S_n$. We can often estimate the probabilities $p_1, \ldots, p_n$ of these outcomes.

By definition of utility, each situation S_i is equivalent to a lottery $L(u(S_i))$ in which we get:

- A_1 with probability $u(S_i)$ and
- A_0 with the remaining probability $1 - u(S_i)$.

Thus, the original action is equivalent to a complex lottery in which:

- first, we select one of the situations S_i with probability p_i: $P(S_i) = p_i$;
- then, depending on S_i, we get A_1 with probability $P(A_1 \mid S_i) = u(S_i)$ and A_0 with probability $1 - u(S_i)$.

The probability of getting A_1 in this complex lottery is:

$$P(A_1) = \sum_{i=1}^{n} P(A_1 \mid S_i) \cdot P(S_i) = \sum_{i=1}^{n} u(S_i) \cdot p_i.$$

In this complex lottery, we get:

- A_1 with probability $u = \sum\limits_{i=1}^{n} p_i \cdot u(S_i)$, and
- A_0 with probability $1 - u$.

So, the utility of the complex action is equal to the sum u.

From the mathematical viewpoint, the sum defining u coincides with the expected value of the utility of an outcome.

Thus, selecting the action with the largest utility means that we should select the action with the largest value of expected utility $u = \sum p_i \cdot u(S_i)$.

Subjective probabilities. In practice, we often do not know the probabilities p_i of different outcomes. How can we gauge our subjective impressions about these probabilities?

For each event E, a natural way to estimate its subjective probability is to fix a prize (e.g., \$1) and compare:

- a lottery ℓ_E in which we get the fixed prize if the event E occurs and 0 is it does not occur, with
- a lottery $\ell(p)$ in which we get the same amount with probability p.

Here, similarly to the utility case, we get a value $ps(E)$ for which, for every $\varepsilon > 0$:

$$\ell(ps(E) - \varepsilon) < \ell_E < \ell(ps(E) + \varepsilon).$$

Then, the utility of an action with possible outcomes $S_1, \ldots, S_n$ is equal to $u = \sum_{i=1}^{n} ps(E_i) \cdot u(S_i)$.

Auxiliary issue: almost-uniqueness of utility. The above definition of utility u depends on the selection of two fixed alternatives A_0 and A_1. What if we use different alternatives A_0' and A_1'? How will the new utility u' be related to the original utility u?

By definition of utility, every alternative A is equivalent to a lottery $L(u(A))$ in which we get A_1 with probability $u(A)$ and A_0 with probability $1-u(A)$. For simplicity, let us assume that $A_0' < A_0 < A_1 < A_1'$. Then, for the utility u', we get $A_0 \equiv L'(u'(A_0))$ and $A_1 \equiv L'(u'(A_1))$. So, the alternative A is equivalent to a complex lottery in which:

- we select A_1 with probability $u(A)$ and A_0 with probability $1 - u(A)$;
- depending on which of the two alternatives A_i we get, we get A_1' with probability $u'(A_i)$ and A_0' with probability $1 - u'(A_i)$.

In this complex lottery, we get A_1' with probability $u'(A) = u(A) \cdot (u'(A_1) - u'(A_0)) + u'(A_0)$. Thus, the utility $u'(A)$ is related with the utility $u(A)$ by a linear transformation $u' = a \cdot u + b$, with $a > 0$.

Traditional approach summarized. We assume that

- we know possible actions, and
- we know the exact consequences of each action.

Then, we should select an action with the largest value of expected utility.

III. How to Describe Selection of a Likert Scale in Terms of Traditional Decision Making: Formulation of (and Solution to) the Corresponding Optimization Problem

How do we mark this on a Likert scale? We would like to find out how people decide to mark some values with different labels on a Likert scale. To understand this, let us recall how this marking is done. Suppose that we have Likert scale with $n + 1$ labels 0, 1, 2, …, n, ranging from the smallest to the largest.

Then, if the actual value of the quantity x is very small, we mark label 0. At some point, we change to label 1; let us mark this threshold point by x_1. When we continue increasing x, we first have values marked by label 1, but eventually reach a new threshold after which values will be marked by label 2; let us denote this threshold by x_2, etc. As a result, we divide the range $[\underline{X}, \overline{X}]$ of the original variable into $n + 1$ intervals $[x_0, x_1]$, $[x_1, x_2]$, …, $[x_{n-1}, x_n]$, $[x_n, x_{n+1}]$, where $x_0 = \underline{X}$ and $x_{n+1} = \overline{X}$:

- values from the first interval $[x_0, x_1]$ are marked with label 0;
- values from the second interval $[x_1, x_2]$ are marked with label 1;
- …
- values from the n-th interval $[x_{n-1}, x_n]$ are marked with label $n - 1$;
- values from the $(n+1)$-st interval $[x_n, x_{n+1}]$ are marked with label n.

Then, when we need to make a decision, we base this decision only on the label, i.e., only on the interval to which x belongs. In other words, we make n different decisions depending on whether x belongs to the interval $[x_0, x_1]$, to the interval $[x_1, x_2]$, …, or to the interval $[x_n, x_{n+1}]$.

Decisions based on the Likert discretization are imperfect. Ideally, we should take into account the exact value of the variable x. When we use Likert scale, we only take into account an interval containing x and thus, we do not take into account part of the original information. Since we only use part of the original information about x, the resulting decision may not be as good as the decision based on the ideal complete knowledge.

For example, an ideal office air conditioner should be able to maintain the exact temperature at which a person feels comfortable. People are different, their temperature preferences are different, so an ideal air conditioner should be able to maintain any temperature value x within a certain range $[\underline{X}, \overline{X}]$. In practice, some air conditioners only have a finite number of settings. For example, if we have setting corresponding to 65, 70, 75, and 80 degrees, then a person who prefers 72 degrees will probably select the 70 setting or the 75 setting. In both cases, this person will be somewhat less comfortable than if there was a possibility of an ideal 72 degrees setting.

How do we select a Likert scale: main idea. According to the general ideas of traditional (utility-based) approach to decision making, we should select a Likert scale for which the expected utility is the largest.

To estimate the utility of decisions based on each scale, we will take into account the just-mentioned fact that decisions based on the Likert discretization are imperfect. In utility terms, this means that the utility of the Likert-based decisions is, in general, smaller than the utility of the ideal decision.

Which decision should we choose within each label? In the ideal situation, if we could use the exact value of the quantity

x, then for each value x, we would select an optimal decision $d(x)$, a decision which maximizes the person's utility.

If we only know the label k, i.e., if we only know that the actual value x belongs to the $(k+1)$-st interval $[x_k, x_{k+1}]$, then we have to make a decision based only on this information. In other words, we have to select one of the possible values $\widetilde{x}_k \in [x_k, x_{k+1}]$, and then, for all x from this interval, use the decision $d(\widetilde{x}_k)$ based on this value.

Which value $\widetilde{x}_k$ should we choose: idea. According to the traditional approach to decision making, we should select a value for which the expected utility is the largest.

Which value $\widetilde{x}_k$ should we choose: towards a precise formulation of the problem. To find this expected utility, we need to know two things:

- we need to know the probability of different values of x; these probabilities can be described, e.g., by the probability density function $\rho(x)$;
- we also need to know, for each pair of values x' and x, what is the utility $u(x', x)$ of using a decision $d(x')$ in the situation in which the actual value is x.

In these terms, the expected utility of selecting a value $\widetilde{x}_k$ can be described as

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot u(\widetilde{x}_k, x)\, dx. \tag{1}$$

Thus, for each interval $[x_k, x_{k+1}]$, we need to select a decision $d(\widetilde{x}_k)$ corresponding to the value $\widetilde{x}_k$ for which the expression (1) attains its largest possible value. The resulting expected utility is equal to

$$\max_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot u(\widetilde{x}_k, x)\, dx. \tag{2}$$

How to select the best Likert scale: general formulation of the problem. The actual value x can belong to any of the $n+1$ intervals $[x_k, x_{k+1}]$. Thus, to find the overall expected utility, we need to add the values (2) corresponding to all these intervals. In other words, we need to select the values $x_1, \ldots, x_n$ for which the following expression attains its largest possible value:

$$\sum_{k=0}^{n} \max_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot u(\widetilde{x}_k, x)\, dx. \tag{3}$$

Equivalent reformulation in terms of disutility. In the ideal case, for each value x, we should use a decision $d(x)$ corresponding to this value x, and gain utility $u(x, x)$. In practice, we have to use decisions $d(x')$ corresponding to a slightly different value, and thus, get slightly worse utility values $u(x', x)$. The corresponding decrease in utility $U(x', x) \stackrel{\text{def}}{=} u(x, x) - u(x', x)$ is usually called *disutility*. In terms of disutility, the function $u(x', x)$ has the form

$$u(x', x) = u(x, x) - U(x', x),$$

and thus, the optimized expression (1) takes the form

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot u(x, x)\, dx - \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx.$$

The first integral does not depend on $\widetilde{x}_k$; thus, the expression (1) attains its maximum if and only if the second integral attains its minimum. The resulting maximum (2) thus takes the form

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot u(x, x)\, dx - \min_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx. \tag{4}$$

Thus, the expression (3) takes the form

$$\sum_{k=0}^{n} \int_{x_k}^{x_{k+1}} \rho(x) \cdot u(x, x)\, dx - \sum_{k=0}^{n} \min_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx.$$

The first sum does not depend on selecting the thresholds. Thus, to maximize utility, we should select the values $x_1, \ldots, x_n$ for which the second sum attains its smallest possible value:

$$\sum_{k=0}^{n} \min_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx \to \min. \tag{5}$$

Let is recall that are interested in the membership function. For a general Likert scale, we have a complex optimization problem (5). However, we are not interested in general Likert scales per se, what we are interested in is the use of Likert scales to elicit the values of the membership function $\mu(x)$.

As we have mentioned in Section 1, in an n-valued scale:

- the smallest label 0 corresponds to the value $\mu(x) = 0/n$,
- the next label 1 corresponds to the value $\mu(x) = 1/n$,
- $\ldots$
- the last label n corresponds to the value $\mu(x) = n/n = 1$.

Thus, for each n:

- values from the interval $[x_0, x_1]$ correspond to the value $\mu(x) = 0/n$;
- values from the interval $[x_1, x_2]$ correspond to the value $\mu(x) = 1/n$;
- $\ldots$
- values from the interval $[x_n, x_{n+1}]$ correspond to the value $\mu(x) = n/n = 1$.

The actual value of the membership function $\mu(x)$ corresponds to the limit $n \to \infty$, i.e., in effect, to very large values of n. Thus, in our analysis, we will assume that the number n of labels is huge – and thus, that the width of each of $n+1$ intervals $[x_k, x_{k+1}]$ is very small.

Let us take into account that each interval is narrow. Let us use the fact that each interval is narrow to simplify the expression $U(x', x)$ and thus, the optimized expression (5).

In the expression $U(x', x)$, both values x' and x belong to the same narrow interval and thus, the difference $\Delta x \stackrel{\text{def}}{=}$

$x'-x$ is small. Thus, we can expand the expression $U(x',x) = U(x+\Delta x, x)$ into Taylor series in Δx, and keep only the first non-zero term in this expansion. In general, we have

$$U(x+\Delta, x) = U_0(x) + U_1 \cdot \Delta x + U_2(x) \cdot \Delta x^2 + \ldots,$$

where

$$U_0(x) = U(x,x), \quad U_1(x) = \frac{\partial U(x+\Delta x, x)}{\partial(\Delta x)},$$

$$U_2(x) = \frac{1}{2} \cdot \frac{\partial^2 U(x+\Delta x, x)}{\partial^2(\Delta x)}. \tag{7}$$

Here, by definition of disutility, we get $U_0(x) = U(x,x) = u(x,x) - u(x,x) = 0$. Since the utility is the largest (and thus, disutility is the smallest) when $x' = x$, i.e., when $\Delta x = 0$, the derivative $U_1(x)$ is also equal to 0 – since the derivative of each (differentiable) function is equal to 0 when this function attains its minimum. Thus, the first non-trivial term corresponds to the second derivative:

$$U(x+\Delta x, x) \approx U_2(x) \cdot \Delta x^2,$$

i.e., in other words, that

$$U(\widetilde{x}_k, x) \approx U_2(x) \cdot (\widetilde{x}_k - x)^2.$$

Substituting this expression into the expression

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx$$

that needs to be minimized if we want to find the optimal $\widetilde{x}_k$, we conclude that we need to minimize the integral

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot U_2(x) \cdot (\widetilde{x}_k - x)^2\, dx. \tag{8}$$

This new integral is easy to minimize: if we differentiate this expression with respect to the unknown $\widetilde{x}_k$ and equate the derivative to 0, we conclude that

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot U_2(x) \cdot (\widetilde{x}_k - x)\, dx = 0,$$

i.e., that

$$\widetilde{x}_k \cdot \int_{x_k}^{x_{k+1}} \rho(x) \cdot U_2(x)\, dx = \int_{x_k}^{x_{k+1}} x \cdot \rho(x) \cdot U_2(x)\, dx,$$

and thus, that

$$\widetilde{x}_k = \frac{\int_{x_k}^{x_{k+1}} x \cdot \rho(x) \cdot U_2(x)\, dx}{\int_{x_k}^{x_{k+1}} \rho(x) \cdot U_2(x)\, dx}. \tag{9}$$

This expression can also be simplified if we take into account that the intervals are narrow. Specifically, if we denote the midpoint of the interval $[x_k, x_{k+1}]$ by $\overline{x}_k \stackrel{\text{def}}{=} \dfrac{x_k + x_{k+1}}{2}$, and denote $\Delta x \stackrel{\text{def}}{=} x - \overline{x}_k$, then we have $x = \overline{x}_k + \Delta x$. Expanding the corresponding expressions into Taylor series in terms of a small value Δx and keeping only main terms in this expansion, we get

$$\rho(x) = \rho(\overline{x}_k + \Delta x) = \rho(\overline{x}_k) + \rho'(\overline{x}_k) \cdot \Delta x \approx \rho(\overline{x}_k),$$

where $f'(x)$ denoted the derivative of a function $f(x)$, and

$$U_2(x) = U_2(\overline{x}_k + \Delta x) = U_2(\overline{x}_k) + U_2'(\overline{x}_k) \cdot \Delta x \approx U_2(\overline{x}_k).$$

Substituting these expressions into the formula (9), we conclude that

$$\widetilde{x}_k = \frac{\rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \int_{x_k}^{x_{k+1}} x\, dx}{\rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \int_{x_k}^{x_{k+1}} dx} = \frac{\int_{x_k}^{x_{k+1}} x\, dx}{\int_{x_k}^{x_{k+1}} dx} =$$

$$\frac{\frac{1}{2} \cdot (x_{k+1}^2 - x_k^2)}{x_{k+1} - x_k} = \frac{x_{k+1} + x_k}{2} = \overline{x}_k.$$

Substituting this midpoint value $\widetilde{x}_k = \overline{x}_k$ into the integral (8) and taking into account that on the k-th interval, we have $\rho(x) \approx \rho(\overline{x}_k)$ and $U_2(x) \approx U_2(\overline{x}_k)$, we conclude that the integral (8) takes the form

$$\int_{x_k}^{x_{k+1}} \rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot (\overline{x}_k - x)^2\, dx =$$

$$\rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \int_{x_k}^{x_{k+1}} (\overline{x}_k - x)^2\, dx. \tag{8a}$$

When x goes from x_k to x_{k+1}, the difference $\Delta x = x - \overline{x}_k$ between the value x and the interval's midpoint $\overline{x}_k$ ranges from $-\Delta_k$ to Δ_k, where Δ_k is the interval's half-width:

$$\Delta_k \stackrel{\text{def}}{=} \frac{x_{k+1} - x_k}{2}.$$

In terms of the new variable Δx, the integral in the right-hand side of (8a) has the form

$$\int_{x_k}^{x_{k+1}} (\overline{x}_k - x)^2\, dx = \int_{-\Delta_k}^{\Delta_k} (\Delta x)^2\, d(\Delta x) = \frac{2}{3} \cdot \Delta_k^3.$$

Thus, the integral (8) takes the form

$$\frac{2}{3} \cdot \rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \Delta_k^3.$$

The problem (5) of selecting the Likert scale thus becomes the problem of minimizing the sum (5) of such expressions (8), i.e., of the sum

$$\frac{2}{3} \cdot \sum_{k=0}^{n} \rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \Delta_k^3. \tag{10}$$

Here, $\overline{x}_{k+1} = x_{k+1} + \Delta_{k+1} = (\overline{x}_k + \Delta_k) + \Delta_{k+1} \approx \overline{x}_k + 2\Delta_k$, so $\Delta_k = (1/2) \cdot \Delta\overline{x}_k$, where $\Delta\overline{x}_k \stackrel{\text{def}}{=} \overline{x}_{k+1} - \overline{x}_k$. Thus, (10) takes the form

$$\frac{1}{3} \cdot \sum_{k=0}^{n} \rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \Delta_k^2 \cdot \Delta\overline{x}_k. \tag{11}$$

In terms of the membership function, we have $\mu(\overline{x}_k) = k/n$ and $\mu(\overline{x}_{k+1}) = (k+1)/n$. Since the half-width Δ_k is small, we have

$$\frac{1}{n} = \mu(\overline{x}_{k+1}) - \mu(\overline{x}_k) = \mu(\overline{x}_k + 2\Delta_k) - \mu(\overline{x}_k) \approx \mu'(\overline{x}_k) \cdot 2\Delta_k,$$

thus, $\Delta_k \approx \frac{1}{2n} \cdot \frac{1}{\mu'(\overline{x}_k)}$. Substituting this expression into (11), we get the expression $\frac{1}{3 \cdot (2n)^2} \cdot I$, where

$$I = \sum_{k=0}^{n} \frac{\rho(\overline{x}_k) \cdot U_2(\overline{x}_k)}{(\mu'(\overline{x}_k))^2} \cdot \Delta\overline{x}_k. \tag{12}$$

The expression I is an integral sum, so when $n \to \infty$, this expression tends to the corresponding integral

$$I = \int \frac{\rho(x) \cdot U_2(x)}{(\mu'(x))^2}\, dx. \tag{11}$$

Minimizing (5) is equivalent to minimizing I. With respect to the derivative $d(x) \stackrel{\text{def}}{=} \mu'(x)$, we need to minimize the objective function

$$I = \int \frac{\rho(x) \cdot U_2(x)}{d^2(x)}\, dx \tag{12}$$

under the constraint that

$$\int_{\underline{X}}^{\overline{X}} d(x)\, dx = \mu(\overline{X}) - \mu(\underline{X}) = 1 - 0 = 1. \tag{13}$$

By using the Lagrange multiplier method, we can reduce this constraint optimization problem to the unconstrained problem of minimizing the functional

$$I = \int \frac{\rho(x) \cdot U_2(x)}{d^2(x)}\, dx + \lambda \cdot \int d(x)\, dx, \tag{14}$$

for an appropriate Lagrange multiplier λ. Differentiating (14) with respect to $d(x)$ and equating the derivative to 0, we conclude that $-2 \cdot \frac{\rho(x) \cdot U_2(x)}{d^3(x)} + \lambda = 0$, i.e., that $d(x) = c \cdot (\rho(x) \cdot U_2(x))^{1/3}$ for some constant c. Thus, $\mu(x) = \int_{\underline{X}}^{x} d(t)\, dt = c \cdot \int_{\underline{X}}^{x} (\rho(t) \cdot U_2(t))^{1/3}\, dt$. The constant c must be determined by the condition that $\mu(\overline{X}) = 1$. Thus, we arrive at the following formula (15).

IV. Conclusions and Future Work

Resulting formula. The membership function $\mu(x)$ obtained by using Likert-scale elicitation is equal to

$$\mu(x) = \frac{\int_{\underline{X}}^{x} (\rho(t) \cdot U_2(t))^{1/3}\, dt}{\int_{\underline{X}}^{\overline{X}} (\rho(t) \cdot U_2(t))^{1/3}\, dt}, \tag{15}$$

where $\rho(x)$ is the probability density describing the probabilities of different values of x, $U_2(x) \stackrel{\text{def}}{=} \frac{1}{2} \cdot \frac{\partial^2 U(x + \Delta x, x)}{\partial^2 (\Delta x)}$, $U(x', x) \stackrel{\text{def}}{=} u(x, x) - u(x', x)$, and $u(x', x)$ is the utility of using a decision $d(x')$ corresponding to the value x' in the situation in which the actual value is x.

Comment. The above formula only applies to membership functions like "large" whose values monotonically increase with x. It is easy to write a similar formula for membership functions like "small" which decrease with x. For membership functions like "approximately 0" which first increase and then decrease, we need to separately apply these formula to both increasing and decreasing parts.

Conclusion. The resulting membership degrees incorporate both probability and utility information. This fact *explains why fuzzy techniques often work better than probabilistic techniques* – because the probability techniques only take into account the probability of different outcomes.

Extension to interval-valued case: preliminary results and future work. In this paper, we consider an ideal situation in which

- we have full information about the probabilities $\rho(x)$, and
- the user can always definitely decide between every two alternatives.

In practice, we often only have partial information about probabilities (which can be described by the intervals of possible values of $\rho(x)$) and users are often unsure which of the two alternatives is better (which can be described by interval-valued utilities).

For example, if we have no reasons to believe that some values from the interval $[\underline{X}, \overline{X}]$ are more probable than others and that some values are more sensitive than others, it is natural to assume that $\rho(x) = \text{const}$ and $U_2(x) = \text{const}$, in which case the above formula (15) leads to a linear membership function going from 0 to 1 or from 1 to 0 on the corresponding interval. This may *explain why triangular membership functions* (consisting of two such linear segments) *are successfully used in* many applications of *fuzzy* techniques.

In the future, it is desirable to extend our formulas to the general interval-valued case.

Acknowledgment

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, by Grants 1 T36 GM078000-01 and 1R43TR000173-01 from the National Institutes of Health, and by a grant on F-transforms from the Office of Naval Research. The authors are thankful to the anonymous referees for valuable suggestions.

References

[1] Fishburn P C. *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

[2] Fishburn P C. *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.

[3] Keeney R L and Raiffa H. *Decisions with Multiple Objectives*, John Wiley and Sons, New York, 1976.

[4] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[5] Luce R D and Raiffa R. *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[6] H. T. Nguyen and E. A. Walker, *First Course In Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

[7] Raiffa H. *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.

[8] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.