

Relation Between Polling and Likert-Scale Approaches to Eliciting Membership Degrees Clarified by Quantum Computing

Renata Hax Sander Reiser.

Adriano Maron,

Lidiana Visintin

Universidade Federal de Pelotas – PPGC

96010-610, Pelotas, RS, Brazil

{reiser, akmaron, lvisintin}@inf.ufpel.edu.br

Ana Maria Abeijon

Universidade Católica de Pelotas – UCPel

96010-000, Pelotas, RS, Brazil

anabeijon@terra.com.br

Vladik Kreinovich

Department of Computer Science

University of Texas at El Paso

El Paso, TX 79968, USA

vladik@utep.edu

Abstract—In fuzzy logic, there are two main approaches to eliciting membership degrees: an approach based on polling experts, and a Likert-scale approach, in which we ask experts to indicate their degree of confidence on a scale – e.g., on a scale form 0 to 10. Both approaches are reasonable, but they often lead to different membership degrees. In this paper, we analyze the relation between these two approaches, and we show that this relation can be made much clearer if we use models from quantum computing.

I. INTRODUCTION

Need for fuzzy logic. A large part of our knowledge about the world is described in precise terms: we have equations (like Newton’s equation) that describe the dynamics of systems, we have exact value characterizing the results of measuring physical quantities, we know the probabilities of different outcomes in a physical experiment, etc. However, a significant part of our knowledge comes from experts, be it medical doctors of skilled pilots. Expert can often only describe their knowledge by using imprecise (“fuzzy”) words from natural language, such as “small”, “young”, etc. In contrast to well-defined terms, these words are not precise. While a medical doctor can be sure that a skin blemish of size 1 mm is small and a 5-cm size blemish is not small, this doctor may not be 100% certain whether intermediate values are small or not.

How can we describe this uncertainty? When we are absolutely confident in a statement, we declare this statement to be true; in a computer, “true” is usually represented as 1. When we are absolutely confident that a given statement is false, we declare this statement to be false; in a computer, “false” is usually represented as 0. To describe intermediate degrees of confidence, L. A. Zadeh proposed to use numbers between 0 and 1; see, e.g., [3], [5], [9]. These numbers are known as *membership degrees* because, e.g., the expert’s degree of confidence that a 1 cm blemish is small can be viewed as a degree to which 1 cm belongs to the fuzzy set of all small values.

How can we elicit the corresponding degrees? There are many ways to elicit membership degrees from the experts.

One of these methods is *polling*: we ask several experts whether, e.g., a 1 cm blemish is small or not. If 7 out of 10 experts claim that it is small, we assign a degree $7/10 = 0.7$ to the statement that a 1 cm blemish is small. In general, if m out of n experts agree with the statement, we assign it a degree of confidence m/n .

When we only have one expert, we cannot use polling. In this case, we can ask the expert to mark his or her degree of confidence in this statement on a scale, e.g., from 0 to 10; such scales are known as *Likert scales*. If the expert selects 7 on a scale from 0 to 10, we assign, to this statement, a degree $7/10 = 0.7$. In general, if the expert describes his or her confidence in a statement by marking m on a scale from 0 to n , we assign, to this statement, a degree of confidence m/n .

Problem. Both above elicitation methods are reasonable, both lead to reasonable useful results. However, usually, these two methods led to different membership degrees. It is therefore reasonable to find out how these different degrees are connected.

What we do in this paper. In this paper, we analyze the relation between the polling and Likert-scale elicitation techniques, and we show that this relation becomes clearer if we use formulas from quantum computing.

II. RELATION BETWEEN PROBABILISTIC (POLLING) AND LIKERT-SCALE ELICITATIONS: ANALYSIS OF THE PROBLEM

Probabilistic description of polling uncertainty. Formally, the formula m/n for the polling uncertainty is the same as the formula for a frequency (probability) of an event. This formal analogy makes sense. Indeed, our main objective in describing the expert’s knowledge is to use it. For example, we want to know whether a 1 am blemish is small or not because a medical expert describes her recommendations in terms of “small”: one cure is proposed for a small blemish another for a large one. So, a possible way to find out whether a 1 cm

blemish is small or nor is to observe cases when treatment which works for small blemishes was actually used for a 1 cm blemish.

An expert who had such a patient and successfully used a cure intended for small blemishes (or, vice versa, unsuccessfully tried to use a cure intended for big blemishes) will vote that a 1 cm blemish is small. On the other hand, a doctor who, for a 1 cm blemish, unsuccessfully tried a cure intended for small blemishes (or, vice versa, successfully used a cure intended for big blemishes) will vote that a 1 cm blemish is not small.

In such an interpretation, the polling ratio m/n is equal to the frequency with which a randomly selected 1 cm blemish can be cured by a small-blemish cure.

Distinguishable probabilities. When we have a finite number of observations, the frequency is only approximately equal to the probability. For example, if we flip a fair coin (with 1/2 probability of falling heads) 10 times, we do not necessarily get heads exactly $(1/2) \cdot 10 = 5$ times.

The more observations we make, the closer the frequency to the actual probability. However, when the number of observations is limited, we cannot meaningfully distinguish different probabilities. For example, when we flip a coin 10 times, we can only have 11 possible outcomes: 0, 1, ..., 10 – and we have infinitely many possible values of the probability that a coin falls heads. Based on the number of heads, we thus cannot determine the exact probability, we have only finitely many different options.

When we ask someone to mark a point on a Likert scale, it is reasonable to assume that these points correspond to distinguishable probabilities. How can we describe these distinguishable probabilities?

Let us start with the simple case when we have a single observation. In this case, the probability of observing the event is equal to p and the probability of not observing it is equal to $1-p$. In other words, the number X_1 of observed events during a single observation is equal to $X_1 = 1$ with probability p and to $X_1 = 0$ with probability $1-p$. Thus, the expected number $E_1 = E[X_1]$ of observed events during a single observation is equal to $E_1 = p \cdot 1 + (1-p) \cdot 0$. Similarly, the variance $\sigma_1^2 \stackrel{\text{def}}{=} E[(X_1 - E_1)^2]$ is equal to

$$\sigma_1^2 = p \cdot (1-p)^2 + (1-p) \cdot (0-p)^2 = p \cdot (1-p)^2 + (1-p) \cdot p^2 = p \cdot (1-p) \cdot [p + (1-p)] = p \cdot (1-p);$$

(see, e.g., [7]).

The frequency $f = \frac{M}{N}$ corresponding to N observations is equal to the sum of N variables X_1, \dots, X_N corresponding to N observations divided by N :

$$f = \frac{X_1 + \dots + X_N}{N}.$$

It is known that the expected value of the sum is equal to the sum of expected values [7], so

$$E[X_1 + \dots + X_N] = E[X_1] + \dots + E[X_N] = N \cdot p.$$

When a random variable is divided by a constant, its expected value decreases by the same constant [7], so we get

$$E[f] = \frac{E[X_1 + \dots + X_N]}{N} = \frac{N \cdot p}{N} = p.$$

For N independent events, the variance of the sum is equal to the sum of variances [7], so we get

$$\sigma^2[X_1 + \dots + X_N] = \sigma^2[X_1] + \dots + \sigma^2[X_N] = N \cdot p \cdot (1-p).$$

When a random variable is divided by a constant, its variance decreases by the square of this constant [7], so we get

$$\sigma^2[f] = \frac{\sigma^2[X_1 + \dots + X_N]}{N^2} = \frac{N \cdot p \cdot (1-p)}{N^2} = \frac{p \cdot (1-p)}{N}.$$

Due to the Central Limit Theorem [7], for large N , the distribution of the sum $X_1 + \dots + X_n$ is close to Gaussian; thus, the distribution of the frequency $f = \frac{X_1 + \dots + X_N}{N}$ is also close to Gaussian, with mean p and standard deviation

$$\sigma = \sqrt{\frac{p \cdot (1-p)}{N}}.$$

When we have two different events, with probabilities $p < p'$, then the difference $f' - f$ between the corresponding frequencies is equal to the difference between two independent (almost) normal distributions. Thus, this difference is also normally distributed. The mean value of this difference is equal to the difference between the means, i.e., to $p' - p$, and the variance is equal to the sum of the variances, i.e., to

$$\sigma^2 = \frac{p \cdot (1-p)}{N} + \frac{p' \cdot (1-p')}{N}.$$

When can we guarantee – with a certain degree of confidence – that $f' > f$ and that, therefore, $p' > p$? The probability that $f' - f < 0$ is equal to the probability that for a normal distribution with mean $p' - p > 0$ and standard deviation σ , we get a negative value. In statistics, it is known that the probability of being $\geq 2\sigma$ away from the mean E is $\approx 10\%$, the probability to be at least 3σ away from the mean is $\approx 0.1\%$, etc. For each degree of confidence, there is a value k_0 such that values from the interval $(E - k_0 \cdot \sigma, E + k_0 \cdot \sigma)$ are consistent with the observations, while values outside this interval are too improbable to be support the hypothesis that $p' \geq p$. Thus, we can distinguish between probabilities when all consistent values (i.e., all values from the above interval) are positive, i.e., when $E - k_0 \cdot \sigma > 0$ and $E = p' - p > k_0 \cdot \sigma$.

For each value p , what is the next distinguishable value p' ? It is the smallest value $p' = p + \Delta p$ for which

$$p' - p = \Delta p > k_0 \cdot \sqrt{\frac{p \cdot (1-p)}{N} + \frac{p' \cdot (1-p')}{N}} > 0.$$

For large N , we get $p' \approx p$, so

$$\frac{p \cdot (1-p)}{N} + \frac{p' \cdot (1-p')}{N} \approx 2 \cdot \frac{p \cdot (1-p)}{N},$$

and thus, the above condition takes the form

$$\Delta p > k_0 \cdot \sqrt{2} \cdot \sqrt{\frac{p \cdot (1-p)}{N}}.$$

The smallest of such values is

$$\Delta p = k_0 \cdot \sqrt{2} \cdot \sqrt{\frac{p \cdot (1-p)}{N}}.$$

Resulting relation between polling and Likert-scale elicitation techniques. As we have mentioned, our idea is that when we select the values corresponding to the Likert scale, we:

- take $p = 0$ as the first value; this value corresponds to 0 on a scale from 0 to n and, thus to the degree $0/n = 0$,
- take the first distinguishable value as the next value, the value which corresponds to 1 on a scale from 0 to n , and thus, to the degree $1/n$;
- the first valued distinguishable from the $(1/n)$ -value as the value which corresponds to 2 on a scale from 0 to n and thus, to the degree $2/n$;
- etc.

What is the resulting relation between the probability p and the membership degree $\mu(p)$ corresponding to this probability?

For each probability p corresponding to a degree $\mu(p) = \frac{k}{n}$, the next value $p + \Delta p$ corresponds to the degree $\mu(p + \Delta) = \frac{k+1}{n}$. In other words, for every p , we have

$$\mu(p + \Delta p) = \mu(p) + \frac{1}{n}.$$

When we have a reasonable number of observations, then the value Δp (as deduced above) is small. When Δp is small, we have

$$\frac{\mu(p + \Delta p) - \mu(p)}{\Delta p} \approx \lim_{\Delta p \rightarrow 0} \frac{\mu(p + \Delta p) - \mu(p)}{\Delta p} = \mu'(p),$$

where $\mu'(p)$ denotes the derivative. Multiplying both sides of the approximate equality

$$\frac{\mu(p + \Delta p) - \mu(p)}{\Delta p} \approx \mu'(p)$$

by Δp , we conclude that

$$\mu(p + \Delta p) - \mu(p) \approx \mu'(p) \cdot \Delta p.$$

In this approximation, the formula $\mu(p + \Delta p) = \mu(p) + \frac{1}{n}$ takes the form $\mu'(p) \cdot \Delta p = \frac{1}{n}$, or, equivalently,

$$\mu'(p) = \frac{1}{n \cdot \Delta p}.$$

Substituting the above expression for Δp into this formula, we conclude that

$$\mu'(p) = \frac{1}{n \cdot k_0 \cdot \sqrt{2} \cdot \sqrt{\frac{p \cdot (1-p)}{N}}}.$$

We can simplify this expression into

$$\mu'(p) = c \cdot \frac{1}{\sqrt{p \cdot (1-p)}}$$

for some constant c .

For $p = 0$, when no expert claims that the statement is true, we should have $\mu(0) = 0$. Thus, the general expression for $\mu(p)$ can be obtained by integrating the above formula:

$$\mu(p) = c \cdot \int_0^p \frac{dq}{\sqrt{q \cdot (1-q)}}.$$

A textbook way to compute this integral is to use an auxiliary variable t for which $q = \sin^2(t)$. In this case,

$$\sqrt{q} = \sin(t),$$

$$\sqrt{1-q} = \sqrt{1 - \sin^2(t)} = \sqrt{\cos^2(t)} = \cos(t),$$

$$dq = d(\sin^2(t)) = 2 \cdot \sin(t) \cdot \cos(t) \cdot dt,$$

and thus, the integral takes the form

$$\int_0^p \frac{dq}{\sqrt{q \cdot (1-q)}} = \int_0^{t_0} \frac{2 \cdot \sin(t) \cdot \cos(t) \cdot dt}{\sin(t) \cdot \cos(t)} = 2 \cdot \int_0^{t_0} dt = 2t_0,$$

where t_0 is the value corresponding to p , i.e., the value for which $\sin^2(t_0) = p$. So, we conclude that $\mu(p) = 2c \cdot t_0$, where $\sin^2(t_0) = p$. In other words, we have $t_0 = C \cdot \mu$, where we denoted $C \stackrel{\text{def}}{=} \frac{1}{2c}$. Thus, the relation between μ and p takes the form

$$p = \sin^2(C \cdot \mu).$$

The value C can now be determined from the condition that the absolute confidence $\mu = 1$ corresponds to $p = 1$. For $\mu = 1$, we get $p = 1 = \sin^2(C)$, so $C = \frac{\pi}{2}$. Finally, we get the relation between the polling membership value p and the Likert-scale membership value μ :

$$p = \sin^2\left(\frac{\pi}{2} \cdot \mu\right).$$

Discussion. At first glance, this relation looks very mathematical and non-intuitive. We will show, however, that it becomes much clearer if we use the techniques of quantum computing; see, e.g., [6].

III. QUANTUM COMPUTING CLARIFIES THE RELATION BETWEEN POLLING AND LIKERT-TYPE DEGREES

Probabilities in quantum computing: reminder. The main idea of quantum computing is that by using quantum effects, we can often drastically speed up computations. For example, in classical physics, if we want to look for an element in an unsorted array of n elements, then we need at least n computational steps – because if we use fewer steps, we will not look into all n cells and thus, we may miss the desired element. Interestingly, in quantum case, we can perform the search in \sqrt{n} steps (and $\sqrt{n} \ll n$); see, e.g., [1], [2], [6]. This possibility comes from the fact that in quantum physics, in addition to the usual classical states, we can also have *superpositions* of these states.

For a single bit, in addition to the usual states 0 and 1 – which are denoted as $\langle 0|$ and $\langle 1|$, we also have superpositions, i.e., states of the type $a_0 \cdot \langle 0| + a_1 \cdot \langle 1|$, where a_0 and a_1 are, in general, complex numbers (in quantum computing, only real values of a_0 and a_1 are used). Each such state can be described as a vector with coordinates (a_0, a_1) in a 2-D vector space. The corresponding quantum version of a bit is known as a *qubit*.

In a general qubit state $a_0 \cdot \langle 0| + a_1 \cdot \langle 1|$,

- the probability of observing 1 is equal to a_1^2 , and
- the probability of observing 0 is a_0^2 .

Since we always observe either 0 or 1, these probabilities must add up to 1, i.e., we must always have $a_0^2 + a_1^2 = 1$. In geometric terms, this means that the vector (a_0, a_1) must be on the unit circle with a center at 0. Each such vector can be uniquely described by its angle φ with the axis corresponding to $\langle 0|$: in terms of this angle, we have $a_1 = \sin(\varphi)$ and $a_0 = \cos(\varphi)$.

Because of this relation, a state of a qubit is (almost) uniquely determined by the probability $p = a_1^2$ of observing 1. Once we know this probability, we can determine a_1 as $\pm\sqrt{p}$ (this \pm is what we meant by almost uniquely), and we can determine a_0 as $\pm\sqrt{a_0^2} = \pm\sqrt{1-p}$.

Resulting relation between polling and Likert-scale degrees. For each probability p , we can form a qubit state

$$\sqrt{p} \cdot \langle 1| + \sqrt{1-p} \cdot \langle 0|$$

corresponding to this probability; see, e.g., see [4], [8]. For this state:

- on the one hand, due to the geometry of quantum states, we have $p = a_1^2 = \sin^2(\varphi)$;
- on the other hand, due to the above relation between probabilities and Likert-scale values, we have

$$p = \sin^2\left(\frac{\pi}{2} \cdot \mu\right).$$

Thus, we can conclude that the angle φ between the vector corresponding to this state and the vector corresponding to the false state is equal to

$$\varphi = \frac{\pi}{2} \cdot \mu.$$

So, the Likert-scale degree μ can be geometrically interpreted as (proportional to) the angle between the two states:

$$\mu = \frac{2}{\pi} \cdot \varphi.$$

Fuzzy interpretation of a superposition between the two states (see [8]). Superposition is a basic operation in quantum physics. In addition to superposition between the basic states $\langle 0|$ and $\langle 1|$, we can also consider a superposition of states

$$\sqrt{p} \cdot \langle 1| + \sqrt{1-p} \cdot \langle 0|$$

and

$$\sqrt{p'} \cdot \langle 1| + \sqrt{1-p'} \cdot \langle 0|$$

corresponding to uncertainty. To describe a superposition, we can simply add the corresponding vectors $(\sqrt{p}, \sqrt{1-p})$ and $(\sqrt{p'}, \sqrt{1-p'})$, and then “normalize” the resulting sum

$$(\sqrt{p} + \sqrt{p'}, \sqrt{1-p} + \sqrt{1-p'}),$$

i.e., divide it by the length

$$\sqrt{(\sqrt{p} + \sqrt{p'})^2 + (\sqrt{1-p} + \sqrt{1-p'})^2}$$

of this vector sum, to make sure that the resulting vector belongs to the unit circle (and is, thus, a legitimate quantum state). In terms of the probabilities p and p' , the resulting vector has the form

$$(a_1'', a_0'') = \left(\frac{\sqrt{p} + \sqrt{p'}}{\sqrt{(\sqrt{p} + \sqrt{p'})^2 + (\sqrt{1-p} + \sqrt{1-p'})^2}}, \frac{\sqrt{1-p} + \sqrt{1-p'}}{\sqrt{(\sqrt{p} + \sqrt{p'})^2 + (\sqrt{1-p} + \sqrt{1-p'})^2}} \right)$$

with

$$a_1'' = \frac{\sqrt{p} + \sqrt{p'}}{\sqrt{(\sqrt{p} + \sqrt{p'})^2 + (\sqrt{1-p} + \sqrt{1-p'})^2}}.$$

Thus, the probability p'' of observing 1 in this state is equal to

$$p'' = (a_1'')^2 = \frac{(\sqrt{p} + \sqrt{p'})^2}{(\sqrt{p} + \sqrt{p'})^2 + (\sqrt{1-p} + \sqrt{1-p'})^2}.$$

In terms of probabilities, this looks like a very complex expression. However, in terms of angles, it becomes much simpler. Indeed, if we take a sum of two unit vectors at angles φ and φ' from the $\langle 0|$ axis, we get a bisecting vector at an angle

$$\varphi'' = \frac{\varphi + \varphi'}{2}.$$

Since the Likert-scale degree is simply proportional to the angle, we conclude that

$$\mu'' = \frac{\mu + \mu'}{2}.$$

So, superposition corresponds to simple averaging of Likert-scale degrees.

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.

Our sincere thanks to Regivan Santiago, Fernando Gomide, and to all organizers of the Brazilian Congress on Fuzzy Systems CBSF'2012 (Natal, Brazil, November 6–9, 2012) for valuable discussions which motivated this research.

REFERENCES

- [1] L. K. Grover, "A fast quantum mechanical algorithm for database search", *Proc. 28th ACM Symp. on Theory of Computing*, 1996, pp. 212–219.
- [2] L. K. Grover, "Quantum mechanics helps in searching for a needle in a haystack", *Phys. Rev. Lett.*, 1997, Vol. 79, No. 2, pp. 325–328.
- [3] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
- [4] A. Maron, L. Visintin, R. H. S. Reiser, and M. Abeijon, "Interpreting Fuzzy Connectives from Quantum Computing – Case Study in Reichenbach Implication Class", *Proceedings of the Brazilian Congress on Fuzzy Systems CBSF'2012*, Natal, Brazil, November 6–9, 2012.
- [5] H. T. Nguyen and E. A. Walker, *First Course In Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.
- [6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, U.K., 2000.
- [7] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2007.
- [8] L. Visintin, A. Maron, R. H. S. Reiser, M. Abeijon, and V. Kreinovich, "Aggregation Operations from Quantum Computing", submitted to *FUZZ-IEEE'2013*.
- [9] L. A. Zadeh, "Fuzzy sets", *Information and control*, 1965, Vol. 8, pp. 338–353.