Membership Functions I: Comparing Methods of Measurement

Jean-Lou Chameau and Juan Carlos Santamarina

School of Civil Engineering Purdue University West Lafayette, Indiana

ABSTRACT

A set of questionnaires was prepared to compare four elicitation methods: point estimation, interval estimation, membership function exemplification, and pairwise comparison. Technical and nontechnical topics were involved, assuming subjects performed either as measurement instruments or as repositories of knowledge. The advantages, disadvantages, and limitations of each method are discussed on the basis of the evaluation of the answers given by 22 professors and graduate students. The analysis of the results presented herein is extended in Part II, where trends in fuzziness are discussed. Both Parts I and II are oriented toward practical applications.

KEYWORDS: *elicitation, fuzzy sets, hedges, measurement, membership function*

INTRODUCTION

Uncertainty results from the limited comprehension of a phenomenon. At first glance, it may be surprising that humans have not evolved to overcome this limitation. However, a critical look at human cognitive behavior indicates that there is a trend to maximize performance, and not to maximize precision or to minimize effort. Hence, although cognitive components could improve, there may not be a need to fully comprehend a complex phenomenon. Instead, a top-level model could be used to represent it, but it must include the uncertainty that such abstraction originates. Several models, fuzzy sets among them, can be used to represent this uncertainty; however, context-specific characteristics will favor the use of one over the others (or a combination).

The goal of this study is not the measurement of membership values per se, but the elicitation of knowledge (fuzzy) about a phenomenon, for practical applications. The concern is not with minor variations in membership, but with the appropriateness of membership functions that would be used in a fuzzy analysis. The results of the study are presented in two parts. Part I compares four methods to develop membership functions and their results. Part II centers on an evaluation of responses to the questionnaires used in this study, seeking for general trends in fuzziness.

QUESTIONNAIRE AND ELICITATION METHODS

An introductory course on fuzzy sets taught by the senior author provided the opportunity to gather a group of 22 engineering professors and graduate students with interest in fuzzy sets. They answered a set of seven written questionnaires, which included the following topics: frequency of failure of engineering systems, causes of failures of engineering systems, stability of buildings, darkness of colors for photogrammetric tasks, relative sizes of figures, and age. These topics have different technical relevance and, more important, different underlying assumptions with respect to the role of assessors: In the case of colors and relative size of figures, the subjects worked as measurement instruments; in the other four topics they performed as repositories of information learned through experience or cultural influences, and thus the task dealt with the elicitation of this information. Note that while the latter case is relevant in acquiring knowledge to develop knowledge systems, the former is essential for the use of such system.

Four elicitation methods were used: point estimation, interval estimation, exemplification, and pair-by-pair comparison. These methods differ in assumed scale for membership, elicitation format, and processing of measurements.

Point Estimation

This method requires individuals to select one element within a list (number, object, quality, etc.) or a point on a "reference" axis (numerical or verbal scale) that best answers the question. A typical task is: "Classify color A according to its degree of darkness." Another example is: "Circle the number of years that best separates the group of old people from that of not-old people." A list of numbers is provided in this case, with special attention given to the selection of its range, so that it will not affect the assessors' answers.
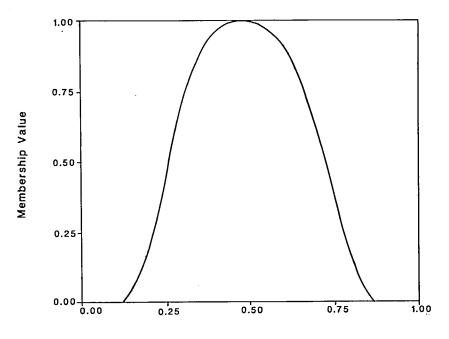
The technique used is a "binary direct rating method" [1], in which membership is determined as a proportional function of the number of answers favoring that particular element or level of stimulus. Accumulation is done for strictly increasing or decreasing membership functions. The membership values

are normalized with respect to the largest one for both unimodal (bell-shaped) and monotonic membership functions (a smoothing function may be used before normalization). A characteristic membership function obtained using this method is presented in Figure 1 for the darkness of the color "emerald green."

Interval Estimation

This method is similar to point estimation, but in this case subjects are allowed to select a reasonable range of possible values (or interval on a reference axis) that best answers the question. Typical tasks are: "Give an interval in which you estimate the color lies" and "Give a range for the age that best separates the group of old and very old people." In the first case a segment is provided with ends representing the extremes of the scale and a few intermediate levels in between.

This is also a binary direct rating method, and therefore membership is obtained as in point estimation but considering all levels in the response intervals. Note that each assessor provides a set of values; thus this is essentially a set-valued statistical approach, equivalent to the "falling shadow" proposed by Wang et al. [2]. Different assumptions can be made for the distribution of

Reference Axis= Darkness

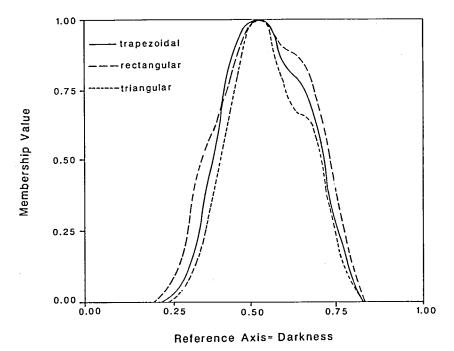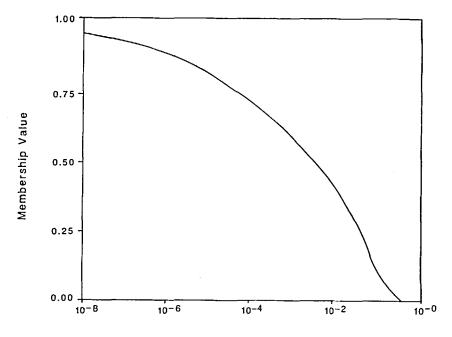**Figure 1.** Point Estimation. Color: Emerald Green

Figure 2. Interval Estimation. Color: Emerald Green. Different Weight Distributions

support (weight) within the interval: rectangular (uniform), triangular, trapezoi-
dal, or continuous curves like pi, s, or z. Figure 2 shows characteristic
unimodal functions obtained with rectangular, trapezoidal, and triangular weight
distributions.

Membership Function Exemplification

In this method assessors are required to provide the object membership to
several discrete points on the reference axis. Typical tasks are: "Give the degree
of belongingness of color A to the set of very dark colors" and "Give the degree
of membership of each of the listed ages to the set of old people." If there are
more than one assessor, the answers for each discrete point can be averaged and
the results normalized. (Note the difference between averaging measures of
membership used here, and aggregating responses to determine membership
used in the previous methods.) The method was applied to unimodal and
monotonic functions. A strictly decreasing membership function obtained with
this method is shown in Figure 3 for the representation of acceptable probability
of failure for civil engineering structures according to the answers to this
questionnaire.

This method is referred as "continuous direct rating" and is the one preferred

Reference Axis= Probability of Failure

Figure 3. Membership Function Exemplification. Acceptable Probability of Failure for Civil Engineering Structures

by Turksen and Norwich in their experiments [1, 3]. However, they performed averages on the responses of one assessor, while in this study the average is on the answers given by different assessors. The effect of this difference is discussed in Part II of the paper.
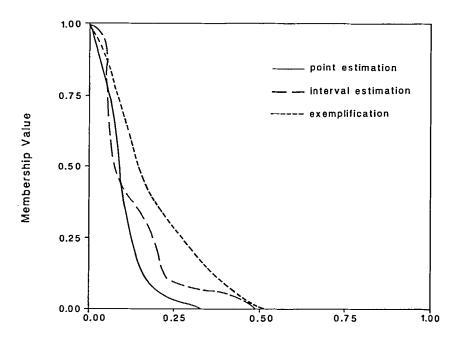
Pairwise Comparison

This method was introduced by Saaty [4, 5]. It consists of comparing the strength by which two objects possess the quality being analyzed. Typical questions in this method are: "Which color, A or B, has the property of darkness more strongly, and how much more? and "Which of the two ages, *n* or *m*, best represents the age of old people? Indicate how strong its representation is compared to the other." A numerical scale is provided to express the relative strength of the property. Comparisons are repeated for all pairs of objects. A nonsymmetric full matrix of relative weights is formed. The degrees of belongingness are the components of the eigenvector corresponding to the maximum eigenvalue. The method also gives a measure of the consistency, but not of the quality, of the answers of each assessor. Examples of membership

functions obtained with this method and of the assumptions made (especially with regard to the existence of membership on a ratio scale) are presented and discussed in the following sections.

TYPICAL RESULTS

Unimodal membership functions (Figures 4 and 5) were constructed for each of nine colors ranging from white to black, using point and interval estimations and exemplification. Good agreement was found in both position and width of the membership functions developed by the different procedures. The maximum discrepancies occurred for colors of intermediate darkness, that is, for membership functions located in the central part of the reference scale. Fuzzy sets obtained by the exemplification method are in most cases wide, fuzzier, than those obtained with the other two methods.

Monotonic functions were obtained by all methods for the "Old People" questionnaire. Strictly increasing membership functions establishing the boundaries between old and not old, and between very old and not very old, were formed (Figures 6 and 7). Good agreement in position and steepness of the

Reference Axis= Darkness

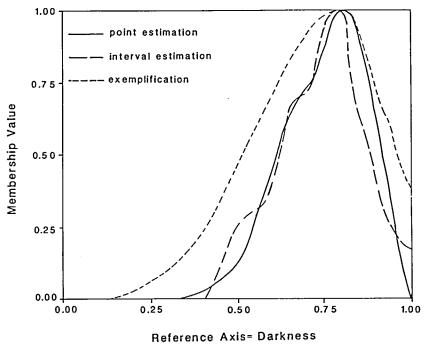Figure 4. Comparison of Methods. Color: White

Figure 5. Comparison of Methods. Color: Dark Green

functions was found in both tasks. The main observation is the significant shift of the functions in Figure 7 with respect to those in Figure 6. Although minor steepening is also observed, these results support more the concept of shift [6] than the one of steepening [7] as the effect of the linguistic hedge "very." This issue is further addressed in Part II of this paper. The age of assessors affected their response, with a tendency to shift answers to the right for older individuals, and in particular in the "old–not old" task. Partial agreement was found between these results and those obtained by other researchers [7, 8]; differences may be due to the age of the assessors, the wording of the question, and the response mode.

ADVANTAGES, DISADVANTAGES, AND LIMITATIONS

Implications of this study that have direct impact on the use of the four elicitation methods under evaluation are presented next. It is noted that the results from these four methods were based on either aggregation or averaging of the responses from several assessors. This approach may seem in contradiction with the inherently individual-dependent nature of uncertainty. However,
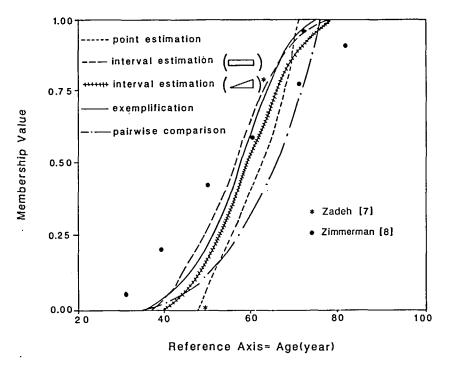
Figure 6.   Comparison of Methods. Age: Old—Not Old

such an approach is justified if the focus of attention is not on an individual's evaluation (comprehension) of a phenomenon but on the phenomenon itself. This is the situation faced in engineering and certainly other fields when one wants to develop knowledge systems for general use.

The point estimation method provides a fast response mode and a simple processing of the answers. The use of existing data and histograms satisfying the possibility-probability consistency principle is an extension of this approach. Its main disadvantage is the contradiction between the fuzziness of the perception and the crispness of the response mode. The difficulty involved in this "defuzzification" process imposed on the assessors causes dispersion in the answers that increases the fuzziness of the resulting membership functions.

The main advantages of the interval estimation method are its simplicity, the fact that it can be quickly answered, and a response mode that allows for fuzziness. Such flexible response permits subjects to transfer the vagueness in their conceptualization of the problem to the answer. As a result, membership functions of lower fuzziness than those developed by the other methods are obtained. A similar observation was also made by Kochen and Badre [9], who noted the difficulty individuals found in reducing the fuzziness of the question
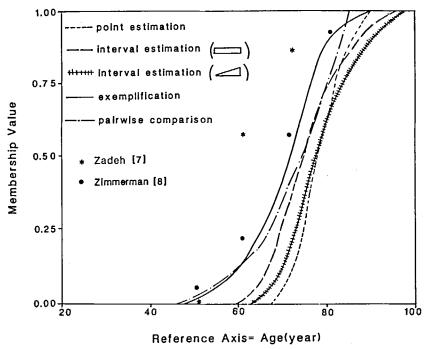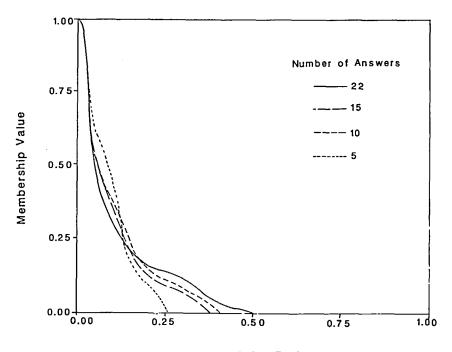
Figure 7. Comparison of Methods. Age: Very Old—Not Very Old

during their answering process. Differences in the fuzziness of results should be taken as a measure of the influence exerted by the method of measurement.

The interval estimation method has the inherent disadvantage of requiring a minimum number of assessors. A parametric study was performed to evaluate the importance of this limitation. Plots of membership functions obtained with 5, 10, 15, and 22 answers showed that as few as five assessors were sufficient to obtain representative functions for the tasks considered. Figures 8 and 9 show extreme and middle-of-scale examples, respectively, constructed using a trapezoidal weight distribution.

Finally, when a triangular, trapezoidal, or other nonuniform function for the distribution of support within the interval is assumed, the interval estimation method can be classified in between a binary and a continuous direct rating approach. Such implementation leads to narrower (i.e., less fuzzy) membership functions than those obtained with uniform weight in the interval (Figure 2).

The exemplification method gives membership functions without further processing. However, the implementation may become cumbersome if an extensive discretization of the reference axis is needed and repeated measurements are made. Implementations of the method using computer graphics and a set of predefined "membership shapes" may reduce this inconvenience. It is

Reference Axis= Darkness

Figure 8. Interval Estimation. Number of Assessors

possible to verify the measurements by reverse rating, as Turksen did [1]; such verification is also possible for the other methods used in this study. Finally, this technique assumes that individuals can express membership values without biases; limitation of this capacity and the crisp nature of the response resulted in fuzzier sets than those obtained by other methods (Figures 4 and 5).

The construction of membership functions with the pairwise comparison method is not as straightforward as the theory would indicate. It is assumed that the components of the eigenvector corresponding to the maximum eigenvalue are the degrees of belongingness of the objects. Saaty [4] recommends normalizing each component of the eigenvector with respect to the sum of all components. Thus, results are affected by the number of elements being compared, and no component could ever be 0 or 1.0.

The procedure followed in this study was to include the extremes or a pair of elements of known membership in the set of alternatives. After measurement and processing, bounds were fixed and intermediate values corrected. There is an underlying fundamental issue with this method: the assumption that membership is on a ratio scale. As expressed by Norwich and Turksen [3], "just because a subject is able to give ratio judgements ... , we cannot assume that the
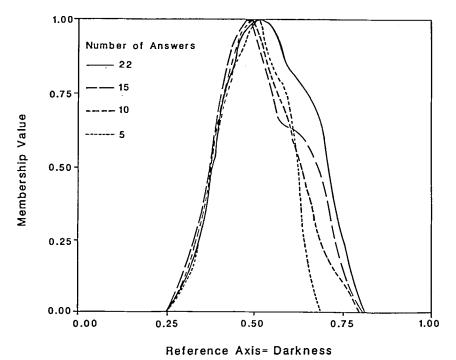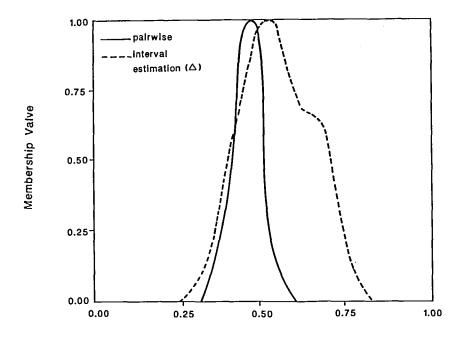
Figure 9. Interval Estimation. Number of Assessors (Emerald Green)

subject is using a ratio scale for membership." These authors concluded that membership exists at best on an interval scale when values are obtained from subjects' answers. A discussion of experimental results and axioms of measurement supporting this conclusion are found in several papers [1, 3, 10].

An alternative use of pairwise comparison is "comparison-based point estimation": The positions of a set of stimuli on the reference axis are determined from each assessor by pairwise comparison, and the membership function for each stimulus is calculated from the aggregation of the answers from different assessors (as was done in point estimation). Figure 10 compares the result of this alternative approach with the membership function based on interval estimation for the color "emerald green"; a membership function of quite low fuzziness is obtained.

The use of pairwise comparison to improve point estimation is attractive; however, some limitations still apply:

- The variable must be on a ratio scale.
- The positions of a pair of stimuli on the reference axis are required.
- Individuals' ability to compare is assumed.
- The choice of a scale for the comparisons may affect the results.

Reference Axix = Darkness

Figure 10. Alternative Use of Pairwise Comparison: Comparison-Based Point Estimation

- There is a rapid increase in the number of questions with the number of levels of stimuli (if n is the number of objects under consideration, then $(n^2 - n)/2$ is the number of answers needed).

The extent of the ability to compare was tested with a questionnaire on the relative size of rectangles, with ratios ranging from 1 to 890. Plots of the mean estimated ratio, of the dispersion between answers, and of the individuals' appreciation of the uncertainty of their answers versus the correct ratio indicate that individuals perform best for ratios up to 10 (see also Part II of this paper).

The effect of the comparison scale was tested with a questionnaire on the overturning stability of buildings. Three scales were used: 1 to 5, 1, to 9 (suggested by Saaty [5]), and "open scale" (i.e., real ratio as perceived by the subject). It was found that the wider the scale, the more inconsistent the answers of each assessor, the higher the variation among assessors, the more the shift to the "stable" side, and the wider the membership functions. Interviews with a few subjects after the questionnaire was answered revealed that individuals found great difficulty in estimating the "real ratio."

In addition to the above limitations, the pairwise comparison method should

Table 1. Comparison of Methods

Item	Method			
	Point	Interval	Exemplification	Pairwise
Number of answers from each assessor[a]	1	1	m	$(n^2 - n)/2$
Minimum number of assessors[b]	15–20	5–10	1–3	1–3
Ease of response[c]	5.86	8.00	5.14	5.18
Expected consistency[c]	5.53	7.68	5.43	5.23
Expected quality[c]	5.59	6.68	5.91	6.41

[a] n = number of objects; m = number of discretization points.

[b] Based on the questionnaire "Old People" (aggregation from different assessors is assumed).

[c] Average value of individual responses.

not be used for qualities or issues that are controversial. A questionnaire intended to rate the most prevalent causes of civil engineering failures markedly exhibited this limitation: in 5 out of 9 pairs, the selection of either option in the pair was supported by a similar number of assessors.

The last page of the set of questionnaires was designed to capture the subjects' opinion on the four methods that were tested. Individuals were required to grade the methods on a scale of 1 to 10 (lowest grade to highest grade) according to expected consistency, ease of response, and expected quality of information. Table 1 summarizes these results as well as other comparisons based on one of the questionnaires that used the four techniques. The method of interval estimation is favored in most categories, in particular with regard to ease of response.

CONCLUSIONS

Four methods for obtaining membership functions were selected as possible candidates for practical applications: point estimation, interval estimation, membership function exemplification, and pairwise comparison. A set of questionnaires was designed to determine their advantages, disadvantages, and limitations.

Good agreement was observed between the results obtained with point estimation, exemplification, and interval estimation, the last one leading to the least fuzzy membership functions. The method based on the estimation of intervals was found to offer a number of advantages that make it very suitable for practical applications. Its main attribute is to allow for the transfer of the fuzziness in the perception of the object directly into the answer, minimizing the difficulty involved in the response process.

The pairwise comparison method (as suggested by Saaty) faces several difficulties with regard to its assumptions as well as its implementation. Of particular concern is its validity to measure membership. An alternative use of the method to improve the point estimation technique appears possible.

ACKNOWLEDGMENTS

References

1. Turksen, I. B., Measurement of membership functions, in *Applications of Fuzzy Set Theory in Human Factors,* (W. Karwowski and A. Mital, Eds.), Elsevier, New York, 55–67, 1986.

2. Wang, P., Liu, X., and Sanchez, E., Set-valued statistics and its applications to earthquake engineering, *Fuzzy Sets Syst.* **18**, 347–356, 1986.

3. Norwich, A. M., and Turksen, I. B., A model for the measurement of membership and the consequences of its empirical implementation, *Fuzzy Sets Syst.* **12**, 1–25, 1984.

4. Saaty, T. L., Measuring the fuzziness of sets, *J. Cybernetics* **4**, 53–61, 1974.

5. Saaty, T. L., A scaling method for priorities in hierarchical structures, *J. Math. Psych.* **15**, 234–281, 1977.

6. Hersh, H. M. and Caramazza, A., A fuzzy set approach to modifiers and vagueness in natural language, *J. Exp. Psych.: General,* **105**, 254–276, 1976.

7. Zadeh, L. A., A fuzzy-set-theoretic interpretation of linguistic hedges, *J. Cybernetics,* **2**, 4–34, 1972.

8. Zimmermann, H. J., *Fuzzy Set Theory and Its Applications,* Kluwer-Nijhoff, Hingham 1985.

9. Kochen, M., and Badre, A. N., On the precision of adjectives which denote fuzzy sets, *J. Cybernetics,* **4,** 49–59, 1974.

10. Turksen, I. B., Measurement of fuzziness: An interpretation of the axioms of measurement, *Proc. of the Conference on Fuzzy Information Knowledge Representation and Decision Analysis,* July 19–22, Marseille, France, 97–102, 1983.