Why 6-Labels Uncertainty Scale in Geosciences: Probability-Based Explanation

Aaron Velasco[1], Julio C. Urenda[2,3], Olga Kosheleva[4], and Vladik Kreinovich[3]
Departments of [1]Earth, Environment, and Resource Sciences, [2]Mathematical Sciences, [3]Computer Science, and [4]Teacher Education
University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA
{aavelasco, jcurenda, olgak, vladik}@utep.edu

Abstract

To describe uncertainty in geosciences, several researchers have recently proposed a 6-labels uncertainty scale, in which one the labels corresponds to full certainty, one label to the absence of any knowledge, and the remaining four labels correspond to the degrees of confidence from the intervals $[0, 0.25]$, $[0.25, 0.5]$, $[0.5, 0.75]$, and $[0.75, 1]$. Tests of this 6-labels scale indicate that it indeed conveys uncertainty information to geoscientists much more effectively than previously proposed uncertainty schemes. In this paper, we use probability-related techniques to explain this effectiveness.

1 Formulation of the Problem

Need to represent uncertainty. In geosciences – as in many other sciences – conclusions are often made with some uncertainty. To get a better understanding of the area's geology, it is this desirable to indicate to what extent we are confident in different features.

Natural ways to represent uncertainty. It is possible to gauge uncertainty by a degree from the interval $[0, 1]$, be it:

- probability coming from the statistical analysis (see, e.g., [9])
- or a subjective ("fuzzy") estimate marked by an expert by a number on the 0-to-1 scale (see, e.g., [1, 3, 4, 6, 7, 11].

Naive idea and why it does not work. In principle, we can add the corresponding degrees to the geological maps. However, this idea has two problems:

- a more minor problem is that the maps are already featuring too much data, and
- a major problem is that most geoscientists – while they get a good intuition about even minor differences in geophysical data, do not have a clear understanding of the difference between, e.g., confidence degrees 60% and 65%.

It is therefore desirable to represent uncertainty in a way that will be easier for geoscientists to grasp.

6-valued representation that works: a brief description. An effective representation – using 6 possible uncertainty labels – was recently described in [10]. In addition to the label "unknown" when we have no information at all, this representation uses 5 labels:

- a label corresponding to full certainty, and
- 4 labels corresponding to degrees d from the intervals $[0, 0.25]$, $[0.25, 0.5]$, $[0.5, 0.75]$, and $[0.75, 1]$.

Tests of using this representation show that it works well for geoscientists.

Challenge. The empirical success of the 6-labels scheme prompts a question: why this representation works well – while other previously proposed scheme did not work so well? For example:

- Why 5 labels and not some other number of labels?
- Why the above four intervals and not some other intervals?

What we do in this paper. In this paper, we use probability ideas to explain this empirical success.

2 Our Explanation

Why no more than 5 uncertainty labels. Why 5 uncertainty labels (not counting the 6th label, when we do not know anything) can be explained by the 7 ± 2 law in psychology (see, e.g., [5, 8]), according to which we naturally divide all the objects into 7 ± 2 groups:

- some people divide into $7 - 2 = 5$ groups,
- some people divide into 7 groups,
- some people divide into $7 + 2 = 9$ groups, etc.

To make a classification easy to use by everyone, it is therefore necessary to use at most 5 labels.

Why exactly 5 uncertainty labels. If we use fewer than 5 labels, we will miss an opportunity to provide more easy-to-grasp information about uncertainty, so we should use exactly 5.

Which intervals? Let us think of how to use the probabilistic approach to answer this question. One of these labels is full uncertainty. What are the intervals corresponding to the remaining four labels? In this paper, we try to use probability techniques to answer this question.

For this purpose, we need to estimate the probability of different degrees. A priori, we have no information about these probabilities. So, we can use Laplace Indeterminacy Principle (related to maximum entropy, see, e.g., [2]) according to which if we have no reason to believe that some value is more or less frequent than the other, then we should assign equal probabilities to both values.

In particular, this means that we should assign equal probability to each degree from the interval $[0, 1]$, i.e., in other words, that we should have a *uniform* distribution of the set of these degrees.

Final explanation. Now that we have select the probability distribution on the set of all degrees, let us use this distribution to we divide these degrees into 4 remaining groups. In general, these groups should correspond to low, medium, etc. certainty. From a common sense viewpoint, if the values $d < d'$ both correspond to, e.g., low certainty, then all the degrees between d and d' should also correspond to low certainty. Thus, for each category, the set of degree corresponding to each category should form a convex subset of the interval $[0, 1]$ – i.e., an interval.

These intervals should cover the whole interval $[0, 1]$. Thus, these intervals should have the form $[0, d_1]$, $[d_1, d_2]$, $[d_2, d_3]$, and $[d_3, 1]$ for some thresholds d_i.

Again, a reasonable idea – corresponding to Laplace Indeterrminacy Principle – is to select these thresholds in such a way that all these four intervals should have the same probability. Since we assumed that the distribution of degrees is uniform, the probability of each interval is equal to the width of this interval. So, these 4 intervals must have the same width:

$$d_1 - 0 = d_2 - d_1 = d_3 - d_2 = 1 - d_3.$$

This leads exactly to the current selection $d_1 = 0.25$, $d_2 = 0.5$, and $d_3 = 0.75$.

Thus, the empirical scale is indeed theoretically explained.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395, and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

References

[1] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.

[2] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.

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

[4] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.

[5] G. A. Miller, "The magical number seven plus or minus two: some limits on our capacity for processing information", *Psychological Review*, 1956, Vol. 63, No. 2, pp. 81–97.

[6] H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.

[7] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.

[8] S. K. Reed, *Cognition: Theories and Application*, SAGE Publications, Thousand Oaks, California, 2022.

[9] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[10] B. Tikoff, T. F. Shipley, E. M. Nelson, R. T. Williams, N. Barshi, and C. Wilson, "Improving the practice of geology through expicit inclusion of scientific uncertainty for data and models", *GSA Today*, 2023, Vol. 33, No. 7, pp. 4–9.

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