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

Aaron Velasco[1], Julio Urenda[2,3], Vladik Kreinovich[3], and Olga Kosheleva[4]

[1]Department of Earth, Environmental, and Resource Sciences
[2,3]Departments of [2]Mathematics, [3]Computer Science and [4]Teacher Education
University of Texas at El Paso, El Paso, Texas 79968, USA
aavelasco@utep.edu, jcurenda@utep.edu, vladik@utep.edu, olgak@utep.edu

1. 6-valued scale: a brief description

- In geoscience – 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.
- It is possible to gauge this uncertainty by a degree from the interval $[0, 1]$.
- It can be probability coming from the statistical analysis.
- It can be a subjective ("fuzzy") estimate marked by an expert by a number on the 0-to-1 scale.
- In principle, we can add these degrees to the geological maps.
- However, this idea has two problems.
- The minor one is that the maps are already featuring too much data.

2. 6-valued scale: a brief description (cont-d)

- The major one 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 confidence degrees 60% and 65%.
- It is therefore desirable to represent uncertainty in a way that will be easier for geoscientists to grasp.
- Such a representation – using 6 possible uncertainty labels – was recently described.
- 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

$$[0, 0.25],\ [0.25, 0.5],\ [0.5, 0.75], \text{ and } [0.75, 1].$$

3. Challenge

- Tests of using this representation show that it works well for geoscientists.
- This empirical success prompts a question:
  - why this representation works well,
  - while other previously proposed scheme did not work so well?
- Why 5 values? Why the above four intervals and not some others?
- In this talk, we use probability ideas to explain this empirical success.

4. Our explanation

- Why 5 uncertainty labels – not counting the 6th label, when we do not know anything?
- This can be explained by the 7 ± 2 law in psychology, according to which we naturally divide all the objects into 7 ± 2 groups.
- Some people divide into 5 groups, some into 7, some into 9.
- To make a classification easy to use by everyone, it is therefore necessary to use at most 5 labels.
- If we use less than 5, we will miss an opportunity to provide more easy-to-grasp information about uncertainty.
- So we should use exactly 5.
- One of these labels is full uncertainty.
- What are the intervals corresponding to the remaining four labels?
- To use probability techniques to answer this question, we need to estimate the probability of different degrees.

5. Our explanation (cont-d)

- A priori, we have no information about these probabilities.
- So, we can use Laplace Indeterminacy Principle (related to maximum entropy) 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 $[0, 1]$.
- In other words, we should have a uniform distribution of the set of these degrees.
- Now, we need to divide these degrees into 4 remaining groups.
- In general, these groups should have the form $[0, d_1]$, $[d_1, d_2]$, $[d_2, d_3]$, and $[d_3, 1]$ for some thresholds d_i.

6. Our explanation (cont-d)

- Again, a reasonable idea is to select these thresholds in such a way that all these four intervals should have the same probability.
- We assumed that the distribution of degrees is uniform.
- So, the probability of each interval is equal to the width of this interval.
- Thus, we must have $d_1 - 0 = d_2 - d_1 = d_3 - d_2 = 1 - d_3$.
- 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.

7. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).