What is the most natural way to propagate subjective interval uncertainty – and why

Ana Tapia-Rosero, Olga Kosheleva, and Vladik Kreinovich

Abstract According to the experimental data, for the sum of two quantities – each of which is described by a subjective interval – the most natural interval is described by a Pythagoras-type formula. In this paper, we show that this experimental result can be explained based on decision theory. Furthermore, we show how this explanation allow us to generalize the empirical formula from addition to general data processing algorithms.

1 Formulation of the problem

Why subjective intervals. In many practical situations, we rely on expert estimates of different quantities – economists estimate current and future characteristics of the economy, medical doctors estimate difficult-to-measure health-related parameters, geoscientists estimate the risk of a strong earthquake, etc.

Expert estimates are usually imprecise. When we ask the experts how accurate their estimates are, they provide a "plus-minus" answer: e.g., 50 ± 10. Such an answer means that, according to this expert, the actual value is most probably located in the interval $[50-10, 50+10]$ – or, more generally, on the interval $[\widetilde{x}-\Delta, \widetilde{x}+\Delta]$, where $\widetilde{x}$ is the original estimate, and Δ describes the expert's estimate of the uncertainty of his/her original estimate. We will call such intervals *subjective*.

Ana Tapia-Rosero
Department of Electrical and Computer Engineering, ESPOL Polytechnic University
Guayaquil, Ecuador, e-mail: atapia@espol.edu.ec

Olga Kosheleva
Department of Teacher Education, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: olgak@utep.edu

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: vladik@utep.edu

Need to propagate subjective interval uncertainty. Expert estimates are used to help solve practical problems: what to do with the economy, how to best treat the patient, what requirements to impose on the new buildings in an earthquake-prone area. To make these decisions, we need to process these estimates. Since expert estimates come with uncertainty, the result of processing these estimates also has uncertainty. It is desirable to estimate this uncertainty, to estimate the accuracy of the result of data processing – i.e., to propagate subjective interval uncertainty via the data processing algorithm.

For processing subjective intervals, it is desirable to come up with operations that best match what humans perceive as the most natural way to process such intervals.

What is known. To decide which operations on subjective intervals are most natural, authors of the papers [19, 21] asked participants to compare several different options. Specifically, in these studies, participants were asked to decide which extensions of arithmetic operations to subjective intervals are the most natural.

To come up with different extensions – and to make the corresponding questions easier to ask – the authors of these papers transformed the input subjective intervals into fuzzy numbers (see, e.g., [1, 7, 13, 17, 18, 24]) – i.e., into nested families of intervals corresponding to different degrees of uncertainty $\alpha \in [0,1]$ – and then asked the participants to compare subjective intervals corresponding to processing fuzzy numbers of different shape. It turned out that for addition $x = x_1 + x_2$, the most natural estimate for the value Δ corresponding to the sum comes from Gaussian-shaped fuzzy numbers – for which $\Delta = \sqrt{\Delta_1^2 + \Delta_2^2}$.

In particular, this value turned out to be more natural than the value $\Delta = \Delta_1 + \Delta_2$ corresponding to the usual interval arithmetic (see, e.g., [6, 10, 12, 14]).

Remaining questions – and what we do in this paper. This empirical result naturally leads to two questions: how to explain it – and how to extend it to other data processing algorithms. In this paper, we provide answers to both questions.

2 How to explain the empirical formula for the sum of two subjective intervals

Our main idea: let us reformulate our problem in statistical terms. In general, a natural measure of deviation from the mean value is the standard deviation; see, e.g., [22]. It is therefore reasonable to interpret the values Δ_i as either equal to – or proportional to – standard deviation: $\Delta_i = k \cdot \sigma_i$ for some constant $k > 0$. In these terms, the question becomes: if we know the standard deviations of two random variables – and we know nothing else – what is then a reasonable estimate for the standard deviation σ of their sum?

Comment. In principle, we can also have bias, i.e., a non-zero mean value. However, a bias, once discovered, it easy to correct – it is sufficient to subtract the mean from all the estimates. So, in this argument, we can assume that the bias has already been eliminated, and the resulting mean is 0.

Let us analyze the reformulated problem. Of course, in the above situation, we cannot uniquely determine the actual value of σ – this value depends on whether the variables are independent or not. What are, in this situation, the possible values of σ?

It is known that, in general, for the sum of two random variables ξ_1 and ξ_2, we have $\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2C_{12}$, where $C_{12} \stackrel{\text{def}}{=} E[(\xi_1 - m_1) \cdot (\xi_2 - m_2)]$ is the covariance of these random variables. It is known that the random variables can be represented as vectors in a multi-D space whose length is σ_i, and the sum of these random variables is simply the sum of these two vectors. In these terms, the problem becomes purely geometric: we know the lengths of two vectors, and we want to know what are the possible values of the length of their sum.

This geometric problem is easy to solve:

- the largest possible length of the sum is when both vectors are parallel and are oriented in the same direction, in which case $\sigma = \sigma_1 + \sigma_2$, and
- the smallest possible length of the sum is when both vectors are parallel and are oriented in different direction, in which case $\sigma = |\sigma_1 - \sigma_2|$.

The same result can be obtained if we take into account a similar formula with correlation, a value between -1 and 1: σ^2 is an increasing function of correlation, so the largest value of σ corresponds to correlation 1 and is equal to the sum, and the smallest value of σ corresponds to correlation -1 and is equal to the absolute value of the difference.

So, we have the whole interval $[|\sigma_1 - \sigma_2|, \sigma_1 + \sigma_2]$, and we need to decide which value from this interval we should select.

Comment. A similar interval can be obtained if we consider the range $[-\Delta, \Delta]$ of the values of the sum $x_1 + x_2$ of two possibly related quantities for which each x_i can take all the values from the interval $[-\Delta_i, \Delta_i]$: the largest possible Δ is $\Delta_1 + \Delta_2$ and the smallest possible is $\Delta = |\Delta_1 - \Delta_2|$; see, e.g., [9].

How can we make this decision? Let us recall what decision theory recommends. To make the desired decision, let us follow recommendations of decision theory. According to decision theory – see, e.g., [3, 4, 8, 11, 15, 16, 20] – decisions of rational people can be described by an appropriate function $u(a)$ known as *utility*: namely, we always select the alternative with the largest value of the utility. The utility is defined in such a way that the utility of a situation in which we have different outcomes with different probabilities is equal to the expected value $E[u(a)]$ of the utility. It is also known that utility is defined modulo an increasing linear transformation: if $u(a)$ is a utility function, then, for each c_0 and $c_1 > 0$, the function $c_{=}0 + c_a \cdot u(a)$ is also utility, with the same properties.

For some alternative a, we do not know the exact value of the utility, we only know the interval $[\underline{u}(a), \overline{u}(a)]$ of its possible values. In this case, according to a discovery made by a Nobelist Leo Hurwicz (see [5, 8, 11]), we should assign, to this alternative, a value $\alpha_H \cdot \overline{u}(a) + (1 - \alpha_H) \cdot \underline{u}(a)$, for some $\alpha_H \in [0,1]$, and there are good arguments to select $\alpha_H = 0.5$.

Let us apply decision techniques to our problem. How does the utility depend on the difference $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ between the estimate $\widetilde{x}$ and the actual value x? To answer this question, in the first approximation, let us use the idea that has been successfully used in physics (see, e.g., [2, 23]): we expand the unknown dependence of the utility on Δx in Taylor series and keep the smallest number of terms necessary to achieve good agreement with qualitative ideas.

In general, any difference $\Delta x \neq 0$ makes decisions worse. For example, while the exact maximum magnitude of a potential earthquake in a given area is uncertain, overestimating its strength may lead to higher construction costs for an event that may not happen. On the other hand, underestimating the potential strength of an earthquake significantly increases the risk of structural failure and human casualties when a major seismic event occurs.

In general, the further Δx from 0, the worse. So, utility attains its maximum when $\Delta x = 0$. Thus, the linear terms in the Taylor expansion is 0, so we need to take a quadratic term into account, i.e., take $u(\Delta x) = a_0 - a_2 \cdot (\Delta x)^2$, for some a_0 and $a_2 > 0$. We can use the fact that the utility function is defined modulo a linear transformation, and apply the transformation $u \mapsto a_2^{-1} \cdot (u - a_0)$, resulting in $u(\Delta x) = -(\Delta x)^2$.

Thus, the utility of the whole random situation is equal to the expected value of this expression, i.e., to $-\sigma^2$. So, in our situation, since the standard deviation σ takes the values from $|\sigma_1 - \sigma_1|$ to $\sigma_1 + \sigma_2$, utility values form an interval

$$[-(\sigma_1 + \sigma)^2, -|\sigma_1 - \sigma_2|^2].$$

Thus, according to the Hurwicz's approach with $\alpha_H = 0.5$, this situation is equivalent to the utility

$$0.5 \cdot (-(\sigma_1 + \sigma)^2) + 0.5 \cdot (-|\sigma_1 - \sigma_2|^2) = -(\sigma_1^2 + \sigma_2^2).$$

We therefore need to select the value σ for which the utility $-\sigma^2$ is equal to this expression: $-\sigma^2 = -(\sigma_1^2 + \sigma_2^2)$. Thus, we get $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Substituting the expressions $\sigma_i = k \cdot \Delta_i$ and $\sigma = k \cdot \Delta$ into this formula and dividing both sides by k, we get exactly the empirical equality $\Delta = \sqrt{\Delta_1^2 + \Delta_2^2}$ that we need to explain.

3 What about other data processing algorithms

What if we consider a general data processing algorithm $f(x_1, \ldots, x_n)$? In this case, we are interested in the difference $\Delta y \stackrel{\text{def}}{=} f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(x_1, \ldots, x_n)$ between the result of data processing and what we would have gotten if we knew the exact values x_i. Here, by definition of Δx_i, we have $x_i = \widetilde{x}_i - \Delta x_i$, so

$$\Delta y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n).$$

Since we keep only the smallest non-trivial terms in the Taylor expansion anyway, let us keep the smaller non-zero terms in the expansion of this expression as well. In this case, the smallest non-zero terms are linear, so

$$\Delta y = \sum_{i=1} c_i \cdot \Delta x_i, \text{ where } c_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}.$$

In this case, imagining that we are talking about the length of the sum of vectors of length $|c_i| \cdot \sigma_i$, we can see that the largest length is when all the vectors are parallel and oriented in the same direction. The smallest possible value is non-zero if one of the terms $|c_j| \cdot \sigma_j$ is larger than the sum of the others, then

$$\underline{\sigma} = |c_j| \cdot \sigma_j - \sum_{i \neq j} |c_i| \cdot \sigma_i.$$

In general, we get an interval $[\underline{\sigma}, \overline{\sigma}]$, thus a utility interval $[-(\overline{\sigma})^2, -(\underline{\sigma})^2]$. If we apply Hurwicz's approach to this formula, then arguments similar to those used in the case of addition, lead us to the following formula for Δ:

- If there exists an index j for which

  $$|c_j| \cdot \Delta_j > \sum_{i \neq j} |c_i| \cdot \Delta_i,$$

  then

  $$\Delta = \sqrt{c_j^2 \cdot \Delta_j^2 + \left(\sum_{i \neq j} |c_i| \cdot \Delta_i \right)^2}.$$

- Otherwise, we have

  $$\Delta = \frac{1}{\sqrt{2}} \cdot \left(\sum_{i=1}^{n} |c_i| \cdot \Delta_i \right).$$

Comment. It should be mentioned that, in general, the value Δ corresponding to, e.g., $x_1 + x_2 + x_3$ is *not* equal to the interval obtained if we first combine x_1 and x_2 and then x_3. For example, for $\Delta_1 = \Delta_2 = \Delta_3 = 1$, the first idea leads to $(1+1+1)/2 = 1.5$, while the second leads to $\sqrt{3} \approx 1.73$.

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 (Center for Collective Impact in Earthquake Science C-CIES), and by the AT&T Fellowship in Information Technology.

It was also supported by a grant from the Hungarian National Research, Development and Innovation Office (NRDI), by the Institute for Risk and Reliability, Leibniz Universitaet Hannover, Germany, and by the European Union under the project ROBOPROX (No. CZ.02.01.01/00/22 008/0004590).

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. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.
3. P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
4. P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.
5. L. Hurwicz, *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.
6. L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2012.
7. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
8. V. Kreinovich, “Decision making under interval uncertainty (and beyond)”, In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.
9. V. Kreinovich, V. M. Nesterov, and N. A. Zheludeva, “Interval methods that are guaranteed to underestimate (and the resulting new justification of Kaucher arithmetic)”, *Reliable Computing*, 1996, Vol. 2, No. 2, pp. 119–124.
10. B. J. Kubica, *Interval Methods for Solving Nonlinear Constraint Satisfaction, Optimization, and Similar Problems: from Inequalities Systems to Game Solutions*, Springer, Cham, Switzerland, 2019.
11. R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
12. G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.
13. J. M. Mendel, *Explainable Uncertain Rule-Based Fuzzy Systems*, Springer, Cham, Switzerland, 2024.
14. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
15. H. T. Nguyen, O. Kosheleva, and V. Kreinovich, “Decision making beyond Arrow’s ‘impossibility theorem’, with the analysis of effects of collusion and mutual attraction”, *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.
16. H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.
17. H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
18. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
19. J. Pisco-Jordán, L. Ramos-Pozo, D. Townsend-Hinostroza, and A. Tapia Rosero, “xf-Calc: towards a fair comparison between approaches for fuzzy numbers arithmetic”, *Proceedings of the 2024 IEEE Latin American Conference on Computational Intelligence LA-CCI*, Bogota, Colombia, November 13–15, 2024. DOI: 10.1109/LA-CCI62337.2024.10814861.

20. H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.
21. L. Ramos-Pozo, J. Pisco-Jordán, and D. Townsend-Hinostroza, "Naturalness and Interpretability of Graphical Representations of Fuzzy Numbers", *Proceedings of the 22nd The Latin American and Caribbean Consortium of Engineering Institutions (LACCEI) International Multi-Conference for Engineering, Education, and Technology: Sustainable Engineering for a Diverse, Equitable, and Inclusive Future at the Service of Education, Research, and Industry for a Society 5.0*, San Jose, Costa Rica, July 17–19, 2024, doi 10.18687/LACCEI2024.1.1.2011 (in Spanish).
22. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.
23. K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2021.
24. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.