Why Convex Combinations of Interval Endpoints: Related Explanations for Cases of Data Processing and Decision Making

Christian Servin[1], Olga Kosheleva[2], and Vladik Kreinovich[3]

[1] Computer Science and Information Technology Systems Department, El Paso Community College, 919 Hunter, El Paso, Texas 79915, USA, cservin1@epcc.edu, https://www.christianservin.com

[2] Department of Teacher Education, University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA, olgak@utep.edu, https://www.cs.utep.edu/vladik/olgavita.html

[3] Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA, vladik@utep.edu, https://www.cs.utep.edu/vladik/

Abstract. There are two cases in which it has been empirically shown that a convex combination of the interval's endpoints works better than any other combination: processing interval data and dealing with situations in which we know both approximate probability and possibility and we need to make a decision. In this paper, we provide an explanation of both phenomena.

Keywords: interval uncertainty, convex combination, data processing, possibility, decision making

1 Formulation of the problem

In this paper, we provide a theoretical explanation for the following two different phenomena in which, empirically, a convex combination $\gamma \cdot \overline{x} + (1-\gamma) \cdot \underline{x}$ of the endpoints of an interval $[\underline{x}, \overline{x}]$ leads to the best results.

First phenomenon. The first phenomenon deals with data processing, namely, with one of its simplest cases: linear regression, i.e.., with determining the coefficients a_i of a linear dependence

$$y = a_0 + a_1 \cdot x_1 + \ldots + a_n \cdot x_n \tag{1}$$

based on the approximately known values x_i and y.

Specifically, in several situations j, we know:

- bounds $[\underline{x}_{ij}, \overline{x}_{ij}]$ for the corresponding values x_{ij} of the quantity x_i, and
- bounds $[\underline{y}_j, \overline{y}_j]$ for the corresponding value y_j of the quantity y.

In situations in which we know the exact values of x_{ij} and y_j, a natural way to estimate the coefficients a_i is by using the Least Squares method, i.e., find the values a_j for which the following expression attains the smallest possible value:

$$\sum_j (y_j - (a_0 + a_1 \cdot x_{1j} + \ldots + a_n \cdot x_{nj}))^2. \quad (2)$$

It turns out (see, e.g., [5, 15]) that in the interval-valued case, the best results are obtained if, for an appropriate $\gamma \in [0, 1]$, we apply the Least Squares method to the values

$$x_{ij} = \gamma \cdot \overline{x}_{ij} + (1 - \gamma) \cdot \underline{x}_{ij} \text{ and}$$
$$y_j = \gamma \cdot \overline{y}_j + (1 - \gamma) \cdot \underline{y}_j. \quad (3)$$

Comment. Since processing fuzzy data is, in effect, equivalent to processing α-cut intervals for each γ (see, e.g., [1, 6, 9, 12, 13, 16]), the same technique can be thus naturally extended to the case when we know x_{ij} and y_j with fuzzy uncertainty.

Second phenomenon. The second phenomenon deals with decision making in situations in which we know both:

- approximate probabilities $\widetilde{p}_i$ of different outcomes i, and
- possibility of different outcomes, i.e., in effect, the largest possible probability $\overline{p}_i$ of these outcomes.

It turns out (see, e.g., [2]) that empirically, the best results are obtained if we based our decisions on the probabilities p_i which are equal to a convex combination of approximate probability and possibility:

$$p_i = \gamma \cdot \overline{p}_i + (1 - \gamma) \cdot \widetilde{p}_i. \quad (4)$$

What we do in this paper. In this paper, we provide an explanation for both phenomena.

2 Explanation for the first phenomenon

What we want. We want a technique that, given an interval $[\underline{x}, \overline{x}]$, selects a value from this interval. Let us denote the selected value by $s(\underline{x}, \overline{x})$, where s comes from “select”.

Analysis of the problem. The empirical values x_i and y are, usually, values of physical quantities. A numerical value of a physical quantity depends on the choice of the measuring unit and of the starting point.

- If we replace the measuring unit with another unit which is $a > 0$ times smaller, then all numerical values will multiply by a: $x \mapsto a \cdot x$. For example, if we replace meters with centimeters, then 1.7 m becomes $100 \cdot 1.7 = 170$ cm.
- If we replace the original starting point with a new one which is b units earlier, then this value b will be added to all numerical values: $x \mapsto x + b$. For example, if we replace the 0 point of the Celsius temperature scale with the 0 point of the Klevin scale – which is approximately 273 degree earlier, then 20 C becomes $10 + 273 = 293$ K.
- In general, if we replace both the measuring unit and the starting point, we get a linear transformation $x \mapsto a \cdot x + b$.

Both changes – of the measuring unit and of the starting point – change the numerical value but do not change the physical quantity itself: e.g., a person who is 1.7 m tall is exactly 170 cm tall. Thus, it is reasonable to require that the selection function should not be affected by these changes. For example, the selection corresponding to 1.7 and 1.8 m, when described in centimeters, should be exactly the same as the selection corresponding to 170 and 180 cm. In precise terms, for general a and b and for all $\underline{x} < \overline{x}$, this natural property takes the following form: if $a = s(\underline{x}, \overline{x})$, then $X = s(\underline{X}, \overline{X})$, where $X = a \cdot x + b$, $\underline{X} = a \cdot \underline{x} + b$, and $\overline{X} = a \cdot \overline{x} + b$. In other words, we should always have:

$$s(a \cdot \underline{x} + b, a \cdot \overline{x} + b) = a \cdot s(\underline{x}, \overline{x}) + b. \tag{5}$$

Main result of this section. Let us prove that the selection function that satisfies the property (5) has the desired form – of the convex combination. Let us denote the value $s(0, 1)$ by γ. By definition of the selection function, it must return the value from the input interval. Thus, we have $\gamma \in [0, 1]$, i.e., $0 \le \gamma \le 1$.

Now, for any $\underline{x}_i < \overline{x}_i$, let us take $a = \overline{x}_i - \underline{x}_i$, $b = \underline{x}_i$, $\underline{x} = 0$ and $\overline{x} = 1$. Then,

$$a \cdot \underline{x} + b = (\overline{x}_i - \underline{x}_i) \cdot 0 + \underline{x}_i = \underline{x}_i,$$

$$a \cdot \overline{x} + b = (\overline{x}_i - \underline{x}_i) \cdot 1 + \underline{x}_i = \overline{x}_i,$$

and thus, the formula (5) takes the following form:

$$s(\underline{x}_i, \overline{x}_i) = (\overline{x}_i - \underline{x}_i) \cdot \gamma + \underline{x}_i = \gamma \cdot \overline{x}_i + (1 - \gamma) \cdot \underline{x}_i.$$

3 Explanation for the second phenomenon

Why do we need a different explanation? At first glance, it may seem that we do not need an additional explanation, since we already have a one. But the above explanation only applies to physical quantities whose numerical values depends on the choice of the measuring unit and the starting point. However, in the second phenomenon, we deal with probabilities – and the numerical value of probability is absolute.

So what do we do. Since we cannot directly deal with probabilities, let us take into account where these probabilities are used. As we have mentioned earlier, these probabilities are used to make decisions. Let us therefore briefly recall how decisions are made – or, to be more precise, how decisions *should be* made by rational decision makers. Such recommended decisions are dealt with by *decision theory*: see, e.g., [3, 4, 7, 8, 10, 11, 14].

Decision theory: a brief reminder. To make appropriate decisions, it is important to properly describe people's preferences. For this purpose, decision theory has the notion of *utility* – that enables us to describe preferences in numerical form.

To introduce this notion, we need to select two alternatives:

- a very good alternative A_+ that is better than anything that can actually happen, and
- a very bad alternative A_- that is worse than anything that can actually happen.

Now, to define the utility of each alternative A, we need to compare this alternative, for different values $p \in [0, 1]$, with *lotteries* $L(p)$ in which:

- we get A_+ with probability p and
- we get A_- with the remaining probability $1 - p$.

When $p \approx 0$, the lottery $L(p)$ is close to the very bad alterative A_-. We selected A_- to be worse than anything that we will actually encounter, so we conclude that $L(p)$ is worse than A; we will denote this by $L(p) < A$.

Similarly, when $p \approx 1$, the lottery $L(p)$ is close to the very good alterative A_+. We selected A_+ to be better than anything that we will actually encounter, so we conclude that A is worse than $L(p)$: $A < L(p)$.

Clearly, the larger the probability p of the getting the very good alternative, the better the lottery: if $p < q$, then $L(p) < L(q)$. Thus:

- if $L(q) < A$ and $p < q$, then we have $L(q) < A$, and
- if $A < L(p)$ and $p < q$, then we have $A < L(q)$.

So, the set $\{p : L(p) < A\}$ is closed under adding smaller numbers, and the set $\{p : A < L(p)\}$ is closed under adding larger numbers. And there can be no more than one value p for which A and $L(p)$ are equivalent: when $A \sim L(p)$ and $p < q$, then $L(p) < L(q)$ implies that $A < L(q)$ – thus, $A \not\sim L(q)$. So, there exists a threshold value

$$u(A) \stackrel{\text{def}}{=} \sup\{p : L(p) < A\} = \inf\{p : A < L(p)\}$$

such that:

- if $p < u(A)$, then $L(p) < A$, and
- if $p > u(A)$, the $A < L(p)$.

This threshold value is called the *utility* of the alternative A. Due to the above property, for every positive number $\varepsilon > 0$, no matter how small it is, we have $L(u(A) - \varepsilon) < A < L(u(A) + \varepsilon)$. When ε is sufficiently small, it is not possible to feel the difference between the probabilities $u(A) - \varepsilon$, $u(A)$, and $u(A) + \varepsilon$. In this sense, we can say that the alternative A is *equivalent* to the lottery $L(u(A))$ in which:

– we get A_+ with probability $u(A)$ and
– we get A_- with the remaining probability $1 - u(A)$.

We will denote this equivalence by $A \equiv L(u(A))$.

Since each alternative A is equivalent to the lottery $L(u(A))$ and the best lottery $L(p)$ is the lottery with the largest probability p, we can therefore conclude that the alternative A is better than the alternative B if and only if $u(A) > u(B)$.

The numerical value of utility depends on our selection of A_- and A_+. It can be shown that if we select a different pair (A'_-, A'_+), then instead of the original utilities $u(A)$, we will get $u'(A) = a \cdot u(A) + b$ for some constants $a > 0$ and b that only on the two pairs (A_-, A_+) and (A'_-, A'_+). Thus, similarly to physical quantities, utility is defined modulo a linear transformation.

How can we use this notion to make a decision? Ideally, for each possible decision, we know what are possible outcomes A_i, and what is probability p_i of each outcome. By using the above description, we can determine the utility u_i of each outcome. Thus, the result of this decision is equivalent to a lottery in which we get the outcome A_i with probability p_i. Each outcome is, as we have mentioned, equivalent to a lottery in which we get A_+ with probability u_i and A_- with the remaining probability $1 - u_i$. So, the result of each decision is equivalent to a two-stage lottery, in which:

– first, we select i so that each i has probability p_i, and then
– depending on what i we selected, we select A_+ with probability u_i and A_- with the probability $1 - u_i$.

As a result of this two-stage lottery, we get either A_+ or A_-. The probability u of getting A_+ can be determined by using the law of total probability

$$u = p_1 \cdot u_1 + \ldots + p_n \cdot u_n. \tag{6}$$

In mathematical terms, this formula described the expected value of utility.

Thus, each decision is equivalent to a lottery in which we get A_+ with probability u and A_- with the remaining probability. By definition of utility, this means that the utility of this possible decision is equal to u. Thus, we need to select a decision for which the expected utility u is the largest possible.

Resulting explanation. Since we are interesting in recommendations to decision making, instead of the probabilities $\widetilde{p}$ and $\overline{p}$, let us consider the corresponding utilities

$$\widetilde{u} = \widetilde{p} \cdot u_+ + (1 - \widetilde{p}) \cdot u_- = \widetilde{p} \cdot (u_+ - u_-) + u_- \tag{7}$$

and

$$\overline{u} = \overline{p} \cdot u_+ + (1 - \overline{p}) \cdot u_- = \overline{p} \cdot (u_+ - u_-) + u_-, \tag{8}$$

where u_+ is the utility of the situation when the given event occurs, and u_- is the utility of the situation in which this event does not occur. Since $\overline{p}$ is the upper bound on the possible probabilities, we must have $\widetilde{p} \le \overline{p}$. So:

- if the event is favorable for us, i.e., if $u_- < u_+$, then we have $\widetilde{u} \le \overline{u}$;
- vice versa, if the event is not favorable for us, i.e., if $u_+ < u_-$, then we have $\overline{u} \le \widetilde{u}$.

In these terms, what we want to have is the utility

$$u = p \cdot u_+ + (1 - p) \cdot u_- = p \cdot (u_+ - u_-) + u_- \tag{9}$$

that we shall actually use for decision making.

If $\widetilde{u} = \overline{u}$, then it makes sense to use this value as the desired utility, i.e., to take $u = \widetilde{u} = \overline{u}$. In the case of $\widetilde{u} \ne \overline{u}$, we need to come up with a mapping $u = s(\widetilde{u}, \overline{u})$ that transforms the two utility values into a single utility value. Since, as we have mentioned, utility is defined modulo a linear transformation, it makes sense to require that this mapping should not change if we apply some linear transformation, i.e., if we use a different pair (A_-, A_+). In precise terms, this means that the function $s(\widetilde{u}, \overline{u})$ should satisfy the following requirement:

$$s(a \cdot \widetilde{u} + b, a \cdot \overline{u} + b) = a \cdot s(\widetilde{u}, \overline{u}) + b. \tag{10}$$

This is exactly the same requirement as formula (5), and we have already shown, in the previous section, that when $u_- < u_+$ and thus, $\widetilde{u} < \overline{u}$, this requirement leads to

$$u = s(\widetilde{u}, \overline{u}) = \gamma \cdot \overline{u} + (1 - \gamma) \cdot \overline{u}. \tag{11}$$

When $u_+ < u_-$ and $\overline{u} < \widetilde{u}$, the similar derivation leads to

$$u = s(\widetilde{u}, \overline{u}) = \gamma' \cdot \widetilde{u} + (1 - \gamma') \cdot \overline{u} \tag{12}$$

for some γ', which leads to the formula (11) for $\gamma = 1 - \gamma'$. So, in both case, we get the formula (11).

Substituting the expressions (7)–(9) for u, $\overline{u}$, and $\widetilde{u}$ into the formula (11), we get

$$p \cdot (u_+ - u_-) + u_- =$$
$$\gamma \cdot (\overline{p} \cdot (u_+ - u_-) + u_-) + (1 - \gamma) \cdot (\widetilde{p} \cdot (u_+ - u_-) + u_-), \tag{13}$$

i.e., if we open parentheses:

$$p \cdot (u_+ - u_-) + u_- =$$
$$\gamma \cdot \overline{p} \cdot (u_+ - u_-) + \gamma \cdot u_- + (1 - \gamma) \cdot \widetilde{p} \cdot (u_+ - u_-) + (1 - \gamma) \cdot u_-. \tag{14}$$

One can easily see that terms proportional to u_- cancel each other, so we have

$$p \cdot (u_+ - u_-) = \gamma \cdot \overline{p} \cdot (u_+ - u_-) + (1 - \gamma) \cdot \widetilde{p} \cdot (u_+ - u_-). \tag{15}$$

if we divide both sides of this equality by $u_+ - u_-$, we get:

$$p = \gamma \cdot \overline{p} + (1 - \gamma) \cdot \widetilde{p},$$

which is exactly the desired formula (4). Thus, we get an explanation for the second phenomenon as well.

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), and by the Institute for Risk and Reliability, Leibniz Universitaet Hannover, Germany.

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. D. Dubois, R. Guillaume, C. Marsala, and A. Rico, “Interpreting Fuzzy Decision Trees with Probability-Possibility Mixtures”, In: J.-M. Lesot, S. Vieira, M. Z. Reformat, J. P. Carvalho, F. Batista, B. Bouchon-Meunier, and R. R. Yager (eds.), *Proceedings of the 20th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU 2024, Lisbon, Portugal, July 22–26, 2024*, Vol. 3, Springer Lecture Notes in Networks and Systems, 2025, Vol. 1176, pp. 169–181, https://doi.org/10.1007/978-3-031-73997-2_15
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. Y. Güven, A. Koklu, and T. Kumbasar, “Zadeh’s Type-2 Fuzzy Logic Systems: Precision and High-Quality Prediction Intervals”, *Proceedings of the IEEE World Congress on Computational Intelligence WCCI 2024*, Yokohama, Japan, June 30 – July 5, 2024.
6. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
7. 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.
8. R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
9. J. M. Mendel, *Explainable Uncertain Rule-Based Fuzzy Systems*, Springer, Cham, Switzerland, 2024.

10. 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.
11. H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.
12. H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
13. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
14. H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.
15. J. Ying, S. Kabir, and C. Wagner, "A Restricted Parametrized Model for Interval-Valued Regression", *Proceedings of the 2023 IEEE International Conference on Fuzzy Systems FUZZ-IEEE 2023*, Incheon, Korea, August 13–17, 2023, doi: 10.1109/FUZZ52849.2023.10309686.
16. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.