Probabilistic Interpretation of non-probabilistic approaches to uncertainty: interval and fuzzy

Vladik Kreinovich
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

1. Outline

- Traditionally, uncertainty was handled exclusively by probabilistic techniques.
- However, since the 1960s, many non-probabilistic uncertainty-related approaches have appeared.
- Example: interval and fuzzy approaches.
- On the positive side, these new approaches have led to many successful practical applications.
- On the other hand, there has been a lot of controversy related to these new approaches.
- This controversy is related to the presumed inconsistency between:
  - these new approaches and
  - the traditional probabilistic viewpoint.

2. Outline (cont-d)

- In this talk, we show that a significant part of this controversy is caused by misunderstandings, both:
  - on behalf of promoters of new approaches and
  - on behalf of the researchers pursuing probabilistic approach.
- To clarify this situation, we recall that: one of the main objectives of science and engineering is to make decisions.
- Decision theory research has shown that:
  - a consistent approach to uncertainty is equivalent to
  - assuming a – possibly subjective – probability distribution.
- It turns out that:
  - in line with this general result,
  - both interval and fuzzy approaches can be naturally interpreted in probabilistic terms.

3. Outline (cont-d)

- With such an interpretation in mind:
  - these new approaches are *not* – as some researchers still think – alternatives to probabilistic approach to uncertainty;
  - what these approaches do is they provide new *algorithms* for solving several important classes of practical problems,
  - algorithms which are, for these problems, more efficient than the use of traditional generic probabilistic techniques.

4. First Example – Interval Uncertainty: Why?

- The main objectives of science and engineering are:
  - to describe the current state of the world,
  - to predict the future state of the world, and
  - to come up with gadgets and algorithms that would make the future state better.
- To describe the state of the world, we need to know the values of the corresponding quantities.
- The values of these quantities mostly come from measurements.
- Measurements are never absolutely accurate.
- The measurement result $\widetilde{x}$ is usually somewhat different from the actual (unknown) value x.
- In other words, the measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ is non-zero.
- In many cases, we can determine the probability distribution of Δx.

5. Interval Uncertainty: Why (cont-d)

- For that, we use a much more accurate measuring instrument (MI) called a *standard*.
- To find the distribution of Δx, we repeatedly measure the same quantities by the calibrated MI and the standard MI.
- Since the standard MI is much more accurate, we can ignore its measurement error.
- So, we can take the difference between measurement results as the measurement error.
- After many measurements, we get a histrogram, from which we can find a distribution for Δx.
- However, there are two cases when this is not done.
- First case is state-of-the-art measurements.
- For them, we do not have any more accurate MIs – Hubble telescope, super-colliders.

6. Interval Uncertainty: Why (cont-d)

- In this case, we cannot determine the probability distribution.
- At best, we can have an upper bound Δ on the absolute value of the measurement error: $|\Delta x| \le \Delta$.
- If we do not even know an upper bound, then after a measurement, all real values are possible.
- This is not a measurement, this is a wild guess.
- Once we know Δ, then, based on the measurement result, we can conclude that the actual value x is in the interval $\mathbf{x}_i = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.
- This is known as *interval uncertainty*.

7. Interval Uncertainty: Why (cont-d)

- Second case is practical measurements, on the factory floor.
- Nowadays, sensors are cheap, but their calibration is very expensive.
- So, in manufacturing, sensors are rarely thoroughly calibrated.
- In many cases, we just use the upper bounds on the measurement errors provided by the sensor's manufacturers.
- In this case, we also have interval uncertainty.

8. Interval Computations: Why and How

- To predict the future value y of a quantity, we can use the known relation $y = f(x_1, \ldots, x_n)$ between:
  - the desired future value of a quantity y and
  - the current values x_i of related quantities.
- In the interval case, we only know that $x_i \in \mathbf{x}_i$.
- Thus, the only thing we can conclude about y is that

$$y \in \mathbf{y} = [\underline{y}, \overline{y}] = f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=} \{f(x_1, \ldots, x_n) : x_i \in \mathbf{x}_i\}.$$

- Computing this range $\mathbf{y}$ is known as *interval computations*.
- It is known that in general, this problem is NP-hard already for quadratic functions.
- This means that, unless P = NP, no feasible algorithm can solve all instances of this problem.
- This computation is easy if $f(x_1, \ldots, x_n)$ is monotonic.

9. Interval Computations: Why and How (cont-d)

- For example, for addition $f(x_1, x_2) = x_1 + x_2$, we have

$$[\underline{x}_1, \overline{x}_1] + [\underline{x}_2, \overline{x}_2] = [\underline{x}_1 + \underline{x}_2, \overline{x}_1 + \overline{x}_2].$$

- For subtraction, we have:

$$[\underline{x}_1, \overline{x}_1] - [\underline{x}_2, \overline{x}_2] = [\underline{x}_1 - \overline{x}_2, \overline{x}_1 - \underline{x}_2].$$

- Multiplication $x_1 \cdot x_2$ is either increasing or decreasing, depending on the signs of x_i, so:

$$[\underline{x}_1, \overline{x}_1] \cdot [\underline{x}_2, \overline{x}_2] = [\min(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2),$$

$$\max(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2)].$$

- For inverse, we get $\dfrac{1}{[\underline{x}_1, \overline{x}_1]} = \left[\dfrac{1}{\overline{x}_1}, \dfrac{1}{\underline{x}_1}\right]$ if $0 \notin [\underline{x}_1, \overline{x}_1]$.

- Division is computers is implemented as $x_1/x_2 = x_1 \cdot (1/x_2)$, so we can combine these two operations.

10. Interval Computations: Why and How (cont-d)

- What to do in the general case?
- In a computer, only arithmetic operations are hardware supported.
- All other computations are performed as a sequence of arithmetic operations.
- For example, $\exp(x)$ is computed by computing the first several terms in the corresponding Taylor series.
- So, a natural first idea is to replace each arithmetic operation with an operation on intervals.
- It can be proven, by induction over number of operations, that this way, we get an *enclosure*, i.e., $\mathbf{Y} \supseteq \mathbf{y}$.
- The problem is that in many cases, we get too wide an interval.
- How can we get narrower intervals?
- One idea is to use monotonicity.

11. Interval Computations: Why and How (cont-d)

- In general, monotonicity may not be evident.
- To check monotonicity w.r.t. x_i, we need to check whether the partial derivative is always non-negative or non-positive.
- To check this, we can compute an enclosure $\mathbf{d}_i = [\underline{d}_i, \overline{d}_i]$ for

$$\frac{\partial f}{\partial x_i}(\mathbf{x}_1, \ldots, \mathbf{x}_n).$$

- The enclosure $\mathbf{d}_i$ can be computed, e.g., by the natural first idea.
- If $\underline{d}_i \geq 0$, the function is increasing, if $\overline{d}_i \leq 0$, it is decreasing.
- What to do in the non-monotonic case?
- One idea is to use centered form:

$$\mathbf{y} \subseteq f(\widetilde{x}_1, \ldots, \widetilde{x}_n) + \sum_{i=1}^{n} \mathbf{d}_i \cdot [-\Delta_i, \Delta_i].$$

12. Interval Computations: Why and How (cont-d)

- Another natural idea is bisection:
  - we divide one of the intervals into two halves,
  - this divides the range $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$ into two halves;
  - we estimate the range of f over each half and take the union of the two estimates.
- Summarizing – to estimate the range, we:
  - first check monotonicity w.r.t. each variable;
  - if a function is monotonic w.r.t. x_i, then we need to only consider values $\underline{x}_i$ and $\overline{x}_i$, i.e., to consider functions of $n-1$ variables;
  - if f is not monotonic, we use centered form;
  - if we want a more accurate estimate, we bisect and repeat the procedure again.
- There have been many practical applications of these interval techniques.

13. Why Not Assume Uniform Distribution?

- *Situation:* in many practical applications, it is very difficult to come up with the probabilities.
- *Traditional engineering approach:* use probabilistic techniques.
- *Problem:* many different probability distributions are consistent with the same observations.
- *Solution:* select one of these distributions – e.g., the one with the largest entropy.
- *Example – single variable:* if all we know is that $x \in [\underline{x}, \overline{x}]$, then MaxEnt leads to a uniform distribution.
- *Example – multiple variables:* different variables are independently distributed.

14. Limitations of Uniform Distributions

- *Example:* simplest algorithm $y = x_1 + \ldots + x_n$.
- *Measurement errors:* $\Delta x_i \in [-\Delta, \Delta]$.
- *Analysis:* $\Delta y = \Delta x_1 + \ldots + \Delta x_n$.
- *Worst case situation:* $\Delta y = n \cdot \Delta$.
- *Maximum Entropy approach:* due to Central Limit Theorem, Δy is $\approx$ normal, with $\sigma = \Delta \cdot \dfrac{\sqrt{n}}{\sqrt{3}}$.
- *Why this may be inadequate:* we get $\Delta \sim \sqrt{n}$, but due to correlation, it is possible that $\Delta = n \cdot \Delta \sim n \gg \sqrt{n}$.
- *Conclusion:* using a single distribution can be very misleading, especially if we want guaranteed results.
- Examples: high-risk application areas such as space exploration or nuclear engineering.

15. How Can This Be Interpreted in Probabilistic Terms?

- What does interval uncertainty $[\underline{x}, \overline{x}]$ means in probabilistic terms?
- It means that the only information that we have about the probability distribution is that it is located on the interval $\mathbf{x}_i$.
- This interpretation explains the main formula of interval computations: $\mathbf{y} = f(\mathbf{x}_1, \ldots, \mathbf{x}_n)$.
- Namely:
  - if all we know about a joint distribution is that each marginal belongs to $\mathbf{x}_i$,
  - then the class of possible distributions of y is the class of all possible distributions of $\mathbf{y}$.
- Clearly, if each marginal is in $\mathbf{x}_i$, the distribution of $y = f(x_1, \ldots, x_n)$ is located on $\mathbf{y}$.
- So, to prove this result, it is sufficient to prove that every distribution on $\mathbf{y}$ can be thus represented.

16. Probabilistic Interpretation (cont-d)

- Indeed, every distribution can be viewed as a convex combination of point distributions corresponding to $y \in \mathbf{y}$.
- So, it is sufficient to represent such point distributions.
- By definition of the range, each value $y \in \mathbf{y} = f(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ can be represented as $y = f(x_1, \ldots, x_n)$ for some $x_i \in \mathbf{x}_i$.
- Thus, if we have point distributions for each i, we get the desired point distribution for y.
- The statement is thus proven.
- Since interval is part of probabilistic framework, why use it?
- We could instead try different joint distributions, find the range for y for each of them, and take the union of all these ranges.
- The advantage of interval computations techniques is that it solves this problem much faster than this testing.

17. Probabilistic Interpretation (cont-d)

- From this viewpoint,
  - interval approach is *not* – as some researchers may think – alternatives to probabilistic approach to uncertainty;
  - what this approach does is it provides new *algorithms* for solving several important classes of practical problems,
  - algorithms which are, for these problems, more efficient than the use of traditional generic probabilistic techniques.

18. Why Fuzzy Logic: A Practical Problem

- Lotfi A. Zadeh was a specialist in control and systems.
- His textbook *Linear System Theory: The State Space Approach* (with Charles A. Desoer) was a classic.
- It provided optimal solutions to many important control problems – optimal within the existing models.
- But, surprisingly, in many practical situations, "optimal" control was worse than control by human experts.
- Clearly, something was missing from the corresponding models.
- So Zadeh asked experts what is missing.

19. What Experts Said

- Many experts explained what was wrong with the "optimal" control.
- However, these explanations were given in imprecise natural-language terms.
- For example, a expert driver can say:
  - if a car in front is close, and
  - if this car slows down a little bit,
  - then a driver should hit the breaks slightly.
- Until Zadeh, engineers would try to extract precise strategy from the expert; they would ask an expert:
  - a car is 5 m close, and
  - it slows down from 60 to 55 km/h,
  - for how long and with what force should we hit the brakes?

20. Problem with the Traditional AI Approach

- Most people cannot answer this question.
- Those who answer give a somewhat random number – and different number every time.
- If we implement exactly this force, we get a weird control – much worse than when a human drives.
- If we instead apply optimization:
  - the resulting control is optimal for the exact weight of the car;
  - but if a new passenger enters the car – the problem changes;
  - the previous optimal control is not longer optimal;
  - this control can be really bad.
- If we simply ignore expert rules, we also get a suboptimal control.

21. What Zadeh Proposed to Solve This Problem: Main Idea

- Also, what we want is imprecise:
  - e.g., for an elevator, we want a smooth ride,
  - but it is difficult to describe this in precise terms.
- Zadeh had an idea:
  - in situations when we can only extract imprecise (fuzzy) rules from the experts,
  - instead of ignoring these rules,
  - let us develop techniques that transform these fuzzy rules into a precise control strategy.
- Zadeh invented the corresponding technique – it is the technique he called *fuzzy logic*.

22. Fuzzy Technique Illustrated on a Simple Example

- Zadeh's technique can be illustrated on a simple example of a thermostat:
  - if we turn the knob to the right, the temperature T increases;
  - if we turn it to the left, the temperature decreases;
  - our goal is to maintain a comfortable temp. T_0.
- Experts can formulate rules on how the angle u to which we rotate the knob depends on T:
  - if the temperature is practically comfortable, no control is needed;
  - if the temperature is slightly higher than desired, cool the room a little bit;
  - if the temperature is slightly lower than desired, heat up the room a little bit; etc.

23. Simple Example (cont-d)

- In terms of the difference $x \stackrel{\text{def}}{=} T - T_0$:
  - if x is negligible, u should be negligible;
  - if x is small positive, then u should be small negative;
  - if x is small negative, then u should be small positive, etc.
- By using abbreviations N for "negligible", SP for "small positive"', and SN for "small negative", we get:

$$N(x) \Rightarrow N(u); \quad SP(x) \Rightarrow SN(u); \quad SN(x) \Rightarrow SP(u); \ldots$$

- A control u is reasonable for given x ($R(x,u)$) if one of these rules is applicable: $R(x,u) \Leftrightarrow$

$$(N(x)\,\&\,N(u)) \vee (SP(x)\,\&\,SN(u))\,\&\,(SN(x)\,\&\,SP(u)) \vee \ldots$$

24. Three Stages of Fuzzy Control Technique

- To translate this into precise formula, we need:
  - to translate $N(x)$, $N(u)$, ... into precise terms,
  - to interpret "and" and "or", and then
  - to translate the resulting property $R(x, u)$ into a single control value $\overline{u}$.
- Since we deal with "and" and "or", this technique is related to logic.
- Since we deal with imprecise ("fuzzy") statements, Zadeh called it *fuzzy logic*.
- Let us explain all three stages of fuzzy logic technique.

25. First Stage

- First stage: how can we interpret "x is negligible"?
- For traditional (precise) properties like "$x > 5^\circ$", the property is either true or false.
- Here, to some folks, 5 degrees is negligible, some feel a difference of 2 degrees.
- And no one can select an exact value – so that, say 1.9 is negligible but 2.0 is not.
- Same thing: there is no exact threshold separating"close" from "not close".
- At best, expert can mark the *degree* to which x is negligible on a scale from, say, 0 to 10.
- If an expert marks 7 on a scale from 0 to 10, we say that his degree of confidence that x is negligible is $7/10$.

26. Second Stage

- This way, we can find the degrees of $N(x)$, $N(u)$, $SP(x)$, etc.
- Based on these degrees, we need to estimate degrees of propositional combinations $N(x)\,\&\,N(u)$, etc.
- Ideally, we can ask the expert for degrees of all such combinations.
- However, for n basic statements, there are 2^n such combinations.
- For $n = 30$, we have $2^{30} \approx 10^9$ combinations.
- It is not possible to ask 10^9 questions.
- So, we need to be able to estimate the degree $d(A\,\&\,B)$ based on degrees $a = d(A)$ and $b = d(B)$.
- The algorithm $d(A\,\&\,B) \approx f_\&(a, b)$ for such an estimation is known as an *"and"-operation.*

27. Second Stage (cont-d)

- For historical reasons, "and"-operations are also knows as *t-norms*.
- What are natural properties of "and"-operations?
- Since $A \,\&\, B$ means the same as $B \,\&\, A$, this operation must be commutative: $f_\&(a,b) = f_\&(b,a)$.
- Since $A \,\&\, (B \,\&\, C)$ means the same as $(A \,\&\, B) \,\&\, C$, the "and"-operation must be associative.
- There are also natural requirements of monotonicity, continuity,

$$f_\&(1,1) = 1, f_\&(0,0) = f_\&(0,1) = f_\&(1,0) = 0, \ldots$$

- All such operations are known.

28. Examples of ““And”- and “Or”-Operations

- W may want to also require that $A \,\&\, A$ means the same as A: $f_\&(a,a)=a$.
- In this case, we get $f_\&(a,b)=\min(a,b)$.
- This is one of the most widely used “and”-operations.
- Others include $f_\&(a,b)=a\cdot b$, etc.
- Similar properties hold for “or”-operations $f_\vee(a,b)$ (a.k.a. t-conorms).
- For example, if we require that $A\vee A$ means the same as A, we get $f_\vee(a,b)=\max(a,b)$.
- Others include $f_\vee(a,b)=a+b-a\cdot b$, etc.

29. Third (Final) Stage and Resulting Success Stories

- By applying "and"- and "'or"-operations, we get, for each u, the degree $R(x,u)$ to which u is reasonable.
- Now, we need to select a single control value $\overline{u}$.
- It is reasonable to use Least Squares, with $R(x,u)$ as weights: $\int R(x,u)\cdot(u-\overline{u})^2\,du \to \min$.
- The resulting formula is known as *centroid defuzzification*:

$$\overline{u} = \frac{\int R(x,u)\cdot u\,du}{\int R(x,u)\,du}.$$

- This technique has led to many successes:
  - fuzzy-controlled trains and elevators provide smooth ride;
  - fuzzy rice cookers produce tasty rice; etc.

30. Misunderstandings

- Fuzzy ideas were (and are) often misunderstood.
- Some folks falsely believe that in fuzzy logic, $d(A \,\&\, B)$ is uniquely determined by $d(A)$ and $d(B)$.
- They think that a simple counterexamples to this Straw-man belief can prove that fuzzy logic is wrong.
- Some falsely believed that Zadeh recommended min and max only.
- In reality, in his very first fuzzy paper he introduced other operations as well.
- Some believed that Zadeh wanted to replace probabilities with fuzzy logic.
- In reality, he always emphasized the need to have 100 flowers bloom.

31. Why This Should Be Interpretable in Probabilistic Terms

- One of the main objectives of science and engineering is to make decisions.
- Decision theory research has shown that:
  - a consistent approach to uncertainty is equivalent to
  - assuming a – possibly subjective – probability distribution.
- Suppose that an expert is uncertain about some future event E.
- How can we describe this uncertainty in probabilistic terms?
- To do this, we ask the expert to compare two "lotteries":
  - a lottery $L(E)$ in which the expert will get \$100 if E happens and 0 otherwise; and
  - a lottery $L(p)$ in which the experts gets \$100 with probability p, and 0 otherwise.

32. Probabilistic Interpretation of Fuzzy (cont-d)

- For $p = 0$, clearly $L(E)$ is better: $L(0) < L(E)$.
- For $p = 1$, clearly, $L(1)$ is better: $L(E) < L(1)$.
- One can show that there exists a threshold

$$p_0 = \sup\{p : L(p) < L(E)\} = \inf\{p : L(E) < L(p)\}.$$

- For this threshold, for every $\varepsilon > 0$, we have

$$L(p_0 - \varepsilon) < L(E) < L(p_0 + \varepsilon).$$

- In this sense, the expert's belief in E is equivalent to having a subjective probability p_0 that E will happen.

33. How Is Fuzzy Techniques Helpful in Statistical Analysis?

- In many practical situations:
  - we know the probabilities $p_1, \ldots, p_n$ of individual events E_1, ..., E_n, and
  - we would like to know the probabilities of different propositional combinations, such as $E_1 \,\&\, E_2$.
- To describe all such probabilities, it is sufficient to find the probabilities of all "and"-combinations

$$E_{i_1} \,\&\, \ldots \,\&\, E_{i_m}.$$

- If the events are independent, the answer is easy:

$$p(E_{i_1} \,\&\, \ldots \,\&\, E_{i_m}) = p(E_{i_1}) \cdot \ldots p(E_{i_m}).$$

34. Statistical Analysis (cont-d)

- However, often:
  - we know that the events are not independent,
  - but we do not have enough data to find out the exact dependence.
- Traditional statistical approach was to assume some prior joint distribution.
- The problem is that different prior distributions lead to different answers.
- In statistical analysis, we usually select the easiest-to-process distribution.
- However, real life is often complex – so why should we select the simplest method?

35. Statistical Analysis (cont-d)

- One of the main fuzzy ideas is to select an appropriate "and"-operation for:
  - converting probabilities $a = p(A)$ and $b = p(B)$
  - into an estimate $f_\&(a, b)$ for $p(A \,\&\, B)$.
- A natural requirements that estimates for $A \,\&\, B$ and $B \,\&\, A$ should be the same lead to commutativity

$$f_\&(a, b) = f_\&(b, a).$$

- The requirement that estimates for $A\& \,(B \,\&\, C)$ and $(A \,\&\, B) \,\&\, C$ coincide lead to associativity.
- The corresponding "and"-operation should be experimentally determined.

36. Relation to MYCIN

- This idea, in effect, formalizes the procedure successfully used for Stanford's MYCIN.
- This was the world's first successful expert system – designed for diagnosing rare blood diseases.
- Interestingly, MYCIN's authors first thought that their "and"-operation describes general human reasoning.
- However, when they tried to apply it to geophysics, they realized that we need a different $f_{\&}(a, b)$.
- This makes sense – in geophysics:
  - we start digging for oil if there is a good chance of success,
  - even if further tests could clarify the situations.
- In contrast, in medicine, we do not recommend a surgery unless we have made all possible tests.

37. Comment: This Was A Simplified Description of Fuzzy

- The above description of fuzzy only contains the main ideas, real-life applications are more complex.
- First, just like experts cannot say with what force their press the brakes, they cannot tell whet exactly is their degree of confidence.
- An expert can say 7 or 8 on a scale of 0 to 10, but cannot distinguish between 70/100 and 71/100.
- Thus, a more adequate description of expert's confidence is not a number but an interval of possible values.
- An expert may also say how confident she is about each degree – so we have a type-2 fuzzy degree.
- This leads to control which is closer to expert's – and thus, better: smoother, more stable, etc.

38. This Was A Simplified Description (cont-d)

- Second, centroid defuzzification does not always work.
- For example, if we want to avoid an obstacle in front, we can steer to the left or to the right.
- The situation is completely symmetric, thus the defuzzified value is symmetric.
- So it leads us straight into the obstacle.
- Thus, we need to only select control values for which degree of confidence exceeds some threshold.
- Third, we also often have additional constraints – which could also be fuzzy.
- Finally, we often want not just to follow expert, but to optimize – thus further improving their advice.
- Optimization under fuzzy uncertainty can also be handled by fuzzy logic techniques.

39. General Conclusion

- Interval and fuzzy approaches are *not* – as some researchers still think – alternatives to probabilistic approach to uncertainty.
- What these approaches do is they provide new *algorithms* for solving several important classes of practical problems.
- Algorithms which are, for these problems, more efficient than the use of traditional generic probabilistic techniques.

40. 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), and
  - HRD-1834620 and HRD-2034030 (CAHSI Includes).
- It was also supported 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.