Fuzzy-Related Ideas Help on All Stages of Knowledge and Data Processing: Applied Case Studies

Christian Servin

Information Technology Department
El Paso Community College
919 Hunter, El Paso TX 79915, USA
cservin@gmail.com

1. Need for Knowledge and Data Processing: Reminder

- To make appropriate decisions, we need to know the state of the world.
- In other words, we need to know the values of all relevant quantities.
- Some quantities we can measure directly: e.g., temperature outside.
- Other quantities are difficult (or even impossible) to measure directly.
- To estimate such quantity y, we:
  - measure or estimate auxiliary quantities $x_1, \ldots, x_n$ whose relation to y is known, and then
  - use this relation to estimate y.

2. Stages of Knowledge and Data Processing

- Measurements are never absolutely accurate.
- It is therefore important to gauge their accuracy.
- For measurements of spatial quantities, it is also important to gauge accuracy of measuring spatial locations.
- This accuracy is known as *(spatial) resolution.*
- Often, we have some hypothesis about the real world.
- In such situations, one of the main objectives of data processing is to test the corresponding hypothesis.
- We can thus conclude that there are three main stages of knowledge and data processing:
  - gauging measurement accuracy and resolution,
  - testing hypotheses
  - estimating the values of the desired quantities.

3. What We Do In This Talk

- In this talk, we show, on application examples, that fuzzy-related ideas are very helpful on all these stages.
- Specifically, we show that fuzzy-related ideas help:
  - in gauging measurement accuracy and spatial resolution,
  - in testing hypotheses, and
  - in estimating the values of the desired quantities.
- If time allows, we also show:
  - how to enhance the use of fuzzy techniques, and
  - how to go beyond simple $[0, 1]$-based fuzzy techniques.

Part I

Fuzzy-Related Ideas Help in Gauging Measurement Accuracy: Case of Normal Distributions

4. Need to Determine Accuracies of Measurement Instruments

- Most information comes from measurements.
- Measurement results are never absolutely accurate.
- The measurement result $\widetilde{x}$ is, in general, different from the actual (unknown) value x of the corr. quantity.
- To properly process data, it is therefore important to know how accurate are our measurements.
- Ideally, we would like to know:
  - what are the possible values of measurement errors $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$, and
  - how frequent are different possible values of Δx.
- In other words, we would like to know the probability distribution on the set of all possible values of Δx.

5. How Accuracies Are Usually Determined: by Using a Second, Much More Accurate Measuring Instrument

- A usual way to find the desired probability distribution is:
  - to have a second measuring instrument
  - which is much more accurate than the one that we want to estimate.
- In this case, the measurement error $\Delta x_2 = \widetilde{x}_2 - x$ of this second instrument is much smaller than $\Delta x = \widetilde{x} - x$.
- Thus, the difference $\widetilde{x} - \widetilde{x}_2 = (\widetilde{x} - x) - (\widetilde{x}_2 - x) \approx \Delta x$ can serve as a good approximation to Δx.
- From the sample of such differences, we can therefore find the desired probability distribution for Δx.

6. What If We Do Not Have a More Accurate Measuring Instrument?

- But what if the measuring instrument whose accuracy we want to estimate is among the best?
- In this case, we do not have a much more accurate measuring instrument.
- What can we do in this case?
- Usually, there are *several* measuring instrument of the type that we want to analyze.
- Due to measurement errors, for same quantity, these instruments produce different measurement results.
- Let's try to extract the measurement accuracy from the differences between these measurement results.

7. Two Possible Situations

- In some cases, we have a stable manufacturing process.
- This process produces several practical identical measuring instruments.
- For these instruments, the probability distributions of measurement error are the same.
- In such cases, all we need to find is this common probability distribution.
- In other cases, we cannot ignore the differences between different instruments.
- In such cases, for each individual measuring instrument, we need to find its own probability distribution.

8. What is Known: Case of Normal Distribution

- Often, the measurement error is caused by the joint effect of numerous independent small factors.
- In such situations, the Central Limit Theorem implies that this distribution is close to Gaussian.
- A Gaussian distribution is uniquely determined by its mean (bias) and standard deviation σ.
- When we only know the differences, we cannot determine the bias:
  - it could be that all the measuring instruments have the same bias, and
  - we will never determine that since we only see the differences.
- Thus, we should limit ourselves to the *random component* $\Delta x - E[\Delta x]$ of the measurement error.

9. Case of Normal Distribution (cont-d)

- For this "re-normalized" measurement error Δx, the mean is 0, so, all we need to determine is σ.
- For two identical independent instruments, $\widetilde{x}_2 - \widetilde{x}_1$ is also normal, with variance $V = \sigma^2 + \sigma^2 = 2\sigma^2$.
- Thus, once we experimentally determine the variance V of $\widetilde{x}_2 - \widetilde{x}_1$, we *can* compute $\sigma^2 = \dfrac{V}{2}$.
- When instruments are different, the variance of $\widetilde{x}_i - \widetilde{x}_j$ is equal to $V_{ij} = \sigma_i^2 + \sigma_j^2$.
- Thus, for empirical variances V_{12}, V_{23}, and V_{13}, we have:

$$V_{12} = \sigma_1^2 + \sigma_2^2, \quad V_{23} = \sigma_2^2 + \sigma_3^2, \quad V_{13} = \sigma_1^2 + \sigma_3^2.$$

- So, we *can* compute $\sigma_1^2 = \dfrac{V_{12} + V_{13} - V_{23}}{2}$,

$$\sigma_2^2 = \frac{V_{12} + V_{23} - V_{13}}{2}, \text{ and } \sigma_3^2 = \frac{V_{13} + V_{23} - V_{12}}{2}.$$

10. What If Distributions Are Not Gaussian?

- Empirical analysis shows that only slightly more than a half of instruments have Gaussian measurement errors.
- What happens in the non-Gaussian case?
- It is known that, sometimes, we simply cannot uniquely reconstruct the corresponding distributions.
- In this talk, we explain when such a reconstruction is possible and when it is not possible.

11. Idea: Let Us Use Moments

- A Gaussian distribution with zero mean is uniquely determined by its second moment $M_2 = \sigma^2$.
- This means that all higher moments $M_k \stackrel{\text{def}}{=} E[(\Delta x)^k]$ are uniquely determined by the value M_2.
- In general, we may have values of M_k which are different from the corresponding Gaussian values.
- Thus, to describe a general distribution, in addition to M_2, we also need to describe M_3, M_4, ...
- If we know all the moments, then we can uniquely determine the corresponding probability distribution.
- Indeed, the usual way to represent a random variable Δx is by its pdf $\rho(\Delta x)$.

12. Let Us Use Moments (cont-d)

- In many situations, it is convenient to use its *characteristic function*

$$\chi(\omega) \stackrel{\text{def}}{=} E[\exp(\mathrm{i}\cdot\omega\cdot\Delta x)] = \int \rho(\Delta x)\cdot\exp(\mathrm{i}\cdot\omega\cdot\Delta x)\, d(\Delta x).$$

- From the mathematical viewpoint, the characteristic function is the Fourier transform of the pdf.
- It is known that we can uniquely reconstruct a function from its Fourier transform: by inverse transform.
- On the other hand,

$$\exp(z) = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \ldots + \frac{z^k}{k!} + \ldots, \text{ so}$$

$$\chi(\omega) = 1 - \frac{1}{2}\cdot\omega^2\cdot M_2 + \ldots + \frac{\mathrm{i}^k}{k!}\cdot\omega^k\cdot M_k + \ldots$$

- If we know M_k, then we know $\chi(\omega)$, hence $\rho(x)$.

13. Important Fact: For a Symmetric Distribution, Odd Moments Are Zeros

- For a symmetric distribution, for which $\rho(-\Delta x) = \rho(\Delta x)$, all odd moments M_{2s+1} are equal to 0:

$$M_{2s+1} = \int \rho(\Delta x) \cdot (\Delta x)^{2s+1}\, d(\Delta x).$$

- Indeed, if we replace Δx to $\Delta x' \stackrel{\text{def}}{=} -\Delta x$, then $d(\Delta x) = -d(\Delta x')$, $(\Delta x)^{2s+1} = -(\Delta x')^{2s+1}$ and thus, the above integral takes the form

$$M_{2s+1} = -\int \rho(-\Delta x') \cdot (\Delta x')^{2s+1}\, d(\Delta x') =$$

$$-\int \rho(\Delta x') \cdot (\Delta x')^{2s+1}\, d(\Delta x').$$

- So $M_{2s+1} = -M_{2s+1}$ and hence, $M_{2s+1} = 0$.

14. Case When Have Several Identical Measuring Instruments

- In this case, all measuring instruments have the same moments $M_k = E[(\Delta x)^k]$.
- The only available information consists of the differences $\Delta x_1 - \Delta x_2 = \widetilde{x}_1 - \widetilde{x}_2$.
- We can thus determine the moments $M'_k = E[(\Delta x_1 - \Delta x_2)^k]$.
- We would like to find M_k.
- For $k = 2$, we have $M'_2 = 2M_2$, so $M_2 = 0.5M'_2$.
- For $k = 3$, we get $M'_3 = 0$.
- So, the only case when we can reconstruct M_3 is when we know it already.
- One such case is when we know that the distribution is symmetric.

15. When the Probability Distribution of the Measurement Error Is Symmetric

- For a symmetric distribution, all odd moments are equal to 0.
- Thus, to uniquely determine a symmetric distribution, it is sufficient to determine all its even moments M_{2s}.
- We show, by induction, that we can reconstruct all even moments.
- We already know that we can reconstruct M_2.
- Let us assume that we already know how to reconstruct the moments $M_2, \ldots, M_{2s}$.
- To reconstruct $M_{2s+2} = E[(\Delta x)^{2s+2}]$, let us use

$$M'_{2s+2} = E[(\Delta x_1 - \Delta x_2)^{2s+2}].$$

16. Case of Symmetric Distribution (cont-d)

- Here,

$$(\Delta x_1 - \Delta x_2)^{2s+2} = (\Delta x_1)^{2s+2} - (2s+2)\cdot(\Delta x_1)^{2s+1}\cdot \Delta x_2 + \ldots$$

- So,

$$M'_{2s+2} = M_{2s+2} + \frac{(2s+2)\cdot(2s+1)}{1\cdot 2}\cdot M_{2s}\cdot M_2 - \ldots + M_{s+2}.$$

- Thus,

$$M_{2s+2} = \frac{M'_{2s+2}}{2} - \frac{1}{2}\cdot\frac{(2s+2)\cdot(2s+1)}{1\cdot 2}\cdot M_{2s}\cdot M_2 - \ldots$$

- We know the value M'_{2s+2}, and we assumed that we already know $M_2, \ldots, M_{2s}$.
- Thus, we can uniquely determine M_{2s+2}.
- Induction proves that we can indeed determine all the even moments.

17. Case of Different Measuring Instruments

- We observe the moments $M'_{k,i,j} = E[(\Delta x_i - \Delta x_j)^k]$.
- We want to find the moments $M_{k,i} = E[(\Delta x_i)^k]$.
- We already know that in general, the reconstruction is not possible.
- Let us show, by induction, that reconstruction is possible if one of the distributions $i = 1$ is symmetric.
- For $k = 2$, we have $M'_{2,i,j} = M_{2,i} + M_{2,j}$, so we can find $M_{2,i}$; e.g., $M_{2,1} = \dfrac{M'_{2,1,2} + M'_{2,1,3} - M'_{2,2,3}}{2}$.
- Similar formulas help reconstruct even moments.
- $M'_{2s+1,i,1} = M_{2s+1,i} + \frac{(2s+1)\cdot 2s}{1\cdot 2} \cdot M_{2s-1,i} \cdot M_{2,1} + \ldots,$ so

$$M_{2s+1,i} = M'_{2s+1,i,1} - \frac{(2s+1)\cdot 2s}{1\cdot 2} \cdot M_{2s,i} \cdot M_{2,1} - \ldots.$$

Part II

Fuzzy-Related Ideas Help in Gauging Measurement Accuracy: General Case

18. Need to Gauge Accuracy

- The results $\widetilde{x}$ of measuring of estimating a quantity x are never absolutely accurate: $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x \neq 0$.
- *It is important* to know the accuracy Δx.
- *Sometimes*, we do not know this accuracy; then we must estimate it based on the data itself.
- *Usually*, we compare $\widetilde{x}$ with the result $\widetilde{x}_s$ of meas. x w/a more accurate ("standard") measuring instrument:

$$\text{since } \widetilde{x}_s \approx x, \text{ we have } \Delta x \approx \widetilde{x} - \widetilde{x}_s.$$

- *Problem:* how to calibrate state-of-the-art measurements? examples:
  - in the environmental sciences, we measure Carbon and heat fluxes via Eddy covariance towers;
  - in geosciences, we use seismic, gravity, etc., to find density at different depths and locations.

19. Normally Distributed Measurement Errors with Mean 0 and Unknown Variance V

- If we have two similar MIs, then $\widetilde{x}^{(1)} - \widetilde{x}^{(2)} = \Delta x^{(1)} - \Delta x^{(2)}$ is normally distributed w/variance $V' = 2V$.
- From the sample of differences, we estimate V' and estimate V as $V'/2$.
- *Example:* two nearby Eddy Covariance towers.
- In geosciences, we usually have only one seismic map, only one gravity map, etc.
- In general, we have several measurement results $\widetilde{x}^{(i)}$ with variances V_i.
- Here, the variance e_{ij} of the difference $\widetilde{x}^{(i)} - \widetilde{x}^{(j)}$ is equal to $e_{ij} = V_i + V_j$.
- We have 3 equations $e_{12} = \ldots$, $e_{23} = \ldots$, $e_{13} = \ldots$ for 3 unknown variances, so, e.g., $V_1 = \dfrac{e_{12} + e_{13} - e_{23}}{2}$.

20. Need to Go Beyond Normal Distributions, and Resulting Problem

- The distribution of measurement errors is sometimes *not normal* (e.g., in measuring fluxes).
- *In such cases,* in addition to variance V, we need to know skewness and other characteristics.
- *In general,* reconstruction of an asymmetric distribution is not unique.
- *Proof:* Δx and $\Delta y = -\Delta x$ lead to the same distribution for differences $\widetilde{x}^{(1)} - \widetilde{x}^{(2)} = \Delta x^{(1)} - \Delta x^{(2)}$.
- *Natural questions:*
  - which characteristics of the distribution Δx *can* we reconstruct?
  - what are efficient *algorithms* for this reconstruction?

21. Let's Use Fourier Analysis

- We want to find is the probability density $\rho(z)$ describing the distribution of the measurement error $z \stackrel{\text{def}}{=} \Delta x$.
- In order to find the unknown probability density, we will first find its Fourier transform
$$F(\omega) = \int \rho(z) \cdot e^{\mathrm{i}\cdot\omega\cdot z}\, dz = E\left[e^{\mathrm{i}\cdot\omega\cdot z}\right].$$
- Such a mathematical expectation is also known as a *characteristic function* of the random variable z.
- Based on the observed values of the difference $z^{(1)} - z^{(2)}$, we can estimate its characteristic function
$$D(\omega) = E\left[e^{\mathrm{i}\cdot\omega\cdot(z^{(1)}-z^{(2)})}\right].$$
- It is known that $D(\omega) = F(\omega) \cdot F^*(\omega) = |F(\omega)|^2$.
- How can we reconstruct the complex-valued function $F(\omega)$ if we only know its absolute value?

22. Is It Possible to Estimate Accuracy?

- Theoretically, we can consider all possible values of the difference $z^{(1)} - z^{(2)}$.
- In practice, we can only get values proportional to the smallest measuring unit h (e.g., $h = 1$ cm).
- In the 1-D case, the Fourier transform takes the form $F(\omega) = \sum\limits_{k=0}^{N} p_k \cdot s^k$, where $s \stackrel{\text{def}}{=} e^{\mathrm{i}\cdot\omega\cdot h}$.
- In the multi-D case, we have $z = (k_1 \cdot h_1, k_2 \cdot h_2, \ldots)$, and $F(\omega) = \sum\limits_{k_1=0}^{N_1} \sum\limits_{k_2=0}^{N_2} \ldots p_{k_1,k_2,\ldots} \cdot s_1^{k_1} \cdot s_2^{k_2} \cdot \ldots$
- In terms of polynomials, the question takes the following form:
  - we know the values $D(s) = |P(s)|^2 = P(s) \cdot P^*(s)$ for some polynomial $P(s)$,
  - we need to reconstruct this polynomial $P(s)$.

23. Is It Possible to Estimate Accuracy (cont-d)

- In 1-D case, each complex-valued polynomial of degree N has, in general, N complex roots $s^{(1)}$, $s^{(2)}$, etc.
- Thus, $P(s) = \text{const} \cdot (s - s^{(1)}) \cdot (s - s^{(2)}) \cdot \ldots$ and $|P(s)|^2 = \text{const} \cdot (s - s^{(1)}) \cdot (s - s^{(1)})^* \cdot \ldots$
- There are many factors, so there are many ways to represent it as a product – reconstruction is not unique.
- In the multi-D case, a generic polynomial *cannot* be represented as a product of polynomials.
- E.g., to describe a polynomial $\sum\limits_{k=0}^{n} \sum\limits_{l=1}^{n} c_{kl} \cdot s_1^k \cdot s_2^l$ of degree n, we need $(n+1)^2$ coefficients.
- When polyn. multiply, degrees add: $s^m \cdot s^{m'} = s^{m+m'}$.
- Thus, if $P(s)$ is a product of two polynomials, one has a degree $m < n$, and the other degree $n - m$.

24. Proof (cont-d)

- If $P(s)$ is a product of two polynomials, one has a degree $m < n$, and the other degree $n - m$.
- In general:
  - we need $(m+1)^2$ coefficients to describe a polynomial of degree m and
  - we need $(n-m+1)^2$ coefficients to describe a polynomial of degree $n-m$,
  - so to describe arbitrary products of such polynomials, we need $(m+1)^2 + (n-m+1)^2$ coefficients.
- In general, the total number of coefficients is smaller than $(n+1)^2$.
- So, a general polynomial cannot be represented as a product of two polynomials.

25. Conclusion

- We have shown that a general polynomial cannot be represented as a product of two polynomials.
- Thus, $D(s) = P(s) \cdot P^*(s) = Q(s) \cdot Q^*(s)$ implies that $Q(s) = P(s)$ or $Q(s) = P^*(s)$.
- In the first case, we get $\rho(x)$.
- In the second case, we get $\rho(-x)$.
- So, in general, only $\rho(x)$ and $\rho(-x)$ are consistent with the observed differences $\Delta x^{(1)} - \Delta x^{(2)}$.
- Thus, we *can* reconstruct the distribution $\rho(x)$ of measurement errors – modulo $x \to -x$.

26. Practical Question: How to Gauge the Accuracy

- *Problem:* find a function $\rho(z)$ which satisfies the following two conditions:
  - $\rho(z) \geq 0$ for all z, and
  - $|F(\omega)|^2 = D(\omega)$ for given $D(\omega)$.
- *Method:* of successive projections.
- We start with an arbitrary function $\rho^{(0)}(z)$.
- On the k-th iteration, starting with the result $\rho^{(k-1)}(z)$ of the previous iteration, we:
  - find the closest function $\rho'(x)$ to $\rho^{(k-1)}(z)$ which satisfies the 1st condition;
  - then, find the closest function $\rho^{(k)}(x)$ to $\rho'(z)$ which satisfies the 2nd condition.
- We continue this process until it converges.

27. Resulting Algorithm

- We start with an arbitrary function $\rho^{(0)}(z)$.
- On the k-th iteration, we start with the function $\rho^{(k-1)}(z)$ obtained on the previous iteration, and we:
  - first, we compute $\rho'(z) = \max(0, \rho^{(k-1)}(z))$;
  - then, we apply Fourier transform to $\rho'(z)$ and get $F'(\omega)$;
  - after that, we compute $F^{(k)}(\omega) = \dfrac{\sqrt{|D(\omega)|}}{|F'(\omega)|} \cdot F'(\omega)$;
  - as the next approx. $\rho^{(k)}(z)$, we take the result of applying the inverse Fourier transform to $F^{(k)}(\omega)$.
- We continue this process until it converges; this enables us to recover many $\rho(x)$.

Part III
Fuzzy-Related Ideas Help in Gauging Spatial Resolution

28. Outline

- How to estimate relative spatial resolution of different maps or images of the same area under uncertainty?
- We consider probabilistic and fuzzy approaches.
- We show that both approaches lead to the same estimate.
- This makes us somewhat more confident that this joint result is reasonable.

29. Formulation of the Problem

- Different measurements results in maps of different spatial resolution.
- Example: geosciences.
  - gravity data are more accurate, but spatial resolution is low;
  - seismic data are less accurate, but have higher spatial resolution.
- Different techniques provide different pieces of information about the area.
- We would like to have a map that contains all this information.
- For that, we need to fuse the corresponding maps.
- To properly fuse the maps, we need to know their relative spatial resolution.

30. Need to Perform Estimations under Uncertainty

- We need to fuse images $\widetilde{I}_i(x)$, $i = 1, 2, \ldots$ corresponding to different spatial resolution.
- Spatial resolution refers to the relation between $\widetilde{I}_i(x)$ and the actual (ideal) image $I(x)$.
- In many cases, we have uncertainty: we do not know the relation between $\widetilde{I}_i(x)$ and $I(x)$.
- Traditionally, uncertainty in science and engineering is handled by the probabilistic approach.
- This approach originated in situations when we can experimentally determine the frequencies (probabilities).
- It is also actively used when we only have partial (or even no) information about these probabilities.

31. Probabilistic Approach: Examples

- Example 1:
  - if we have two alternatives and
  - we have no reason to assume that one of the them is more frequent than the other one,
  - then it makes sense to assume that these two alternatives have equal probabilities $1/2$.
- Example 2:
  - if we only know that a quantity x is located on an interval $[0, 1]$, and
  - we do not know which values from this interval are more probable or less probable,
  - it makes sense to assume that all these possible values are equally probable, i.e.,
  - that we have a uniform probability distribution on the interval $[0, 1]$.

32. Fuzzy Approach: Idea

- The probabilistic approach "makes up" the unknown probabilities based on common sense.
- Another idea is to explicitly *formalize* the common-sense ideas.
- *Problem:* commonsense ideas are often described by imprecise ("fuzzy") words from natural language.
- *Solution:* use *fuzzy techniques*, specifically developed to formalize natural-language knowledge.

33. Relation Between $\widetilde{I}_i(x)$ and $I(x)$: Probabilistic Approach

- Spatial uncertainty means that the value located at a point x is observed as corr. to $\widetilde{x} = x + \Delta x \approx x$.
- As a result, each value $I(x)$ gets distributed to values $I(x + \Delta x)$, for the corresponding random variable Δx.
- In general, there are many independent sources of spatial uncertainty.
- Δx can therefore be represented as a sum of many small independent random variables.
- Under reasonable assumptions, the distribution of such sums is close to Gaussian (*Central Limit Theorem*).
- So, we conclude that Δx is normally distributed.

34. Probabilistic Approach (cont-d)

- We conclude that Δx is normally distributed.
- In the isotropic case, the probability density is:

$$\rho(\Delta x) = \frac{1}{2\pi \cdot \sigma} \cdot \exp\left(-\frac{\|\Delta x\|^2}{2\sigma^2}\right).$$

- Each original value $I(x)$ is thus distributed, with this density, among the neighboring values.
- The observed $\widetilde{I}(y)$ can be obtained by adding the values $I(x) \cdot \rho(\Delta x)\, d\Delta x$ corresponding to $x + \Delta x = y$:

$$\widetilde{I}(y) = \int I(x) \cdot f(y - x)\, dx =$$

$$\text{const} \cdot \int I(x) \cdot \exp\left(-\frac{\|y - x\|^2}{2\sigma^2}\right)\, dx.$$

35. Relation Between $\widetilde{I}(x)$ and $I(x)$: Fuzzy Approach

- In fuzzy approach, we explicitly formalize the corresponding commonsense knowledge.
- The corresponding rules for each observed value $\widetilde{I}(y)$ are straightforward:

  If x is close to y, then $\widetilde{I}(y)$ is close to $I(x)$.

- Under reasonable assumptions, the way to describe closeness is by using a Gaussian membership function.
- In the isotropic case, we have

$$\mu(y-x) = \exp\left(-\frac{\|y-x\|^2}{2\sigma^2}\right).$$

- For each y, the value $\widetilde{I}(y)$ is equal to $I(x)$ with degree of membership $\mu(y-x)$.

36. Fuzzy Approach (cont-d)

- For each y, $\widetilde{I}(y)$ is $I(x)$ with degree $\mu(y-x)$, where
$$\mu(y-x)=\exp\left(-\frac{\|y-x\|^2}{2\sigma^2}\right).$$
- To transform this fuzzy information into a single (crisp) value, we can use, e.g., centroid defuzzification
$$\widetilde{I}(y)=\frac{\int I(x)\cdot\mu(y-x)\,dx}{\int \mu(y-x)\,dx}.$$
- The denominator is a constant not depending on y.
- Substituting the expression for the Gaussian membership function into this formula, we conclude that
$$\widetilde{I}(y)=\text{const}\cdot\int I(x)\cdot\exp\left(-\frac{\|y-x\|^2}{2\sigma^2}\right)\,dx.$$
- This is exactly the same formula as in the probability approach.

37. How to Estimate Relative Spatial Resolution: First Approximation

- Specifically, we have two images $\widetilde{I}_1(x)$ and $\widetilde{I}_2(x)$.
- According to our formulas, we have

$$\widetilde{I}_1(y) = C_1 \cdot \int I(x) \cdot \exp\left(-\frac{\|y-x\|^2}{2\sigma_1^2}\right)\,dx$$

$$\widetilde{I}_2(y) = C_2 \cdot \int I(x) \cdot \exp\left(-\frac{\|y-x\|^2}{2\sigma_2^2}\right)\,dx.$$

- We want to find the values σ_1 and σ_2.
- Formulas involving convolution are greatly simplified if we use Fourier transform:
  - if $h(y) = \int f(y) \cdot g(x-y)\,dx$,
  - then $H(\omega) = F(\omega) \cdot G(\omega)$.

38. First Approximation (cont-d)

- The Fourier transform $G(\omega)$ of a Gaussian $g(x) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)$ is $G(\omega) = \exp\left(-\frac{1}{2}\cdot\|\omega\|^2\cdot\sigma^2\right)$. Thus:

$$F_1(\omega) = C_1 \cdot F(\omega) \cdot \exp\left(-\frac{1}{2}\cdot\|\omega\|^2\cdot\sigma_1^2\right);$$

$$F_2(\omega) = C_2 \cdot F(\omega) \cdot \exp\left(-\frac{1}{2}\cdot\|\omega\|^2\cdot\sigma_2^2\right).$$

- From the above equations, one can conclude that

$$F_2(\omega) = C \cdot F_1(\omega) \cdot \exp\left(-\frac{1}{2}\cdot\|\omega\|^2\cdot(\sigma_2^2-\sigma_1^2)\right).$$

- Thus, the only information about σ_i that we can extract from the maps is the difference $\Delta \stackrel{\text{def}}{=} \sigma_2^2 - \sigma_1^2$.
- It is reasonable to call this difference *relative spatial resolution* of the two images (maps).

39. A More Realistic Description

- In addition to spatial blurring, there is also inevitable additive noise (measurement error), so

$$\widetilde{I}_i(y) = C_1 \cdot \int I(x) \cdot \exp\left(-\frac{\|y-x\|^2}{2\sigma_i^2}\right)\, dx + n_i(y).$$

- As a result, for Fourier transforms, we get

$$F_i(\omega) = C_1 \cdot F(\omega) \cdot \exp\left(-\frac{1}{2} \cdot \|\omega\|^2 \cdot \sigma_i^2\right) + N_i(\omega).$$

- From these equations, we can conclude that

$$F_2(\omega) = C \cdot F_1(\omega) \cdot \exp\left(-\frac{1}{2} \cdot \|\omega\|^2 \cdot (\sigma_2^2 - \sigma_1^2)\right) + N(\omega),$$

where $C \stackrel{\text{def}}{=} \dfrac{C_1}{C_2}$ & $N(\omega) \stackrel{\text{def}}{=} N_2(\omega) - N_1(\omega) \cdot \exp\left(-\dfrac{1}{2} \cdot \|\omega\|^2 \cdot \Delta\right)$.

- This is a model that we will use to reconstruct the relative spatial resolution Δ.

40. How to Find C

- We have concluded that

$$F_2(\omega) = C \cdot F_1(\omega) \cdot \exp\left(-\frac{1}{2} \cdot \|\omega\|^2 \cdot (\sigma_2^2 - \sigma_1^2)\right) + N(\omega).$$

- The coefficient C can be found, e.g., by comparing the overall energy, i.e., by comparing the values for $\omega = 0$.
- For this value, we get $F_2(0) = C \cdot F_1(0) + N(0)$, so

$$C \approx \frac{F_2(0)}{F_1(0)}.$$

- Once C is estimated, we can divide $\widetilde{I}_2(x)$ (and thus, $F_2(\omega)$) by C, and get a simpler relation:

$$F_2(\omega) = F_1(\omega) \cdot \exp\left(-\frac{1}{2} \cdot \|\omega\|^2 \cdot \Delta\right) + N(\omega).$$

41. Taking Noise into Account

- In many practical cases, we do not know the exact characteristics of the additive noise.
- We only know noise order of magnitude n, such that $N(\omega) \approx n$.
- For frequencies ω for which $N_2(\omega) \approx n$, the whole observed value may be caused by noise.
- The corresponding values $N_i(\omega)$ do not carry any information about the actual image.
- Thus, they do not carry any information about Δ.
- So, we must only consider "above-noise" frequencies, for which $|F_2(\omega)| \geq c \cdot n$ for some constant $c \gg 1$.
- For these frequencies, we have

$$F_2(\omega) \approx F_1(\omega) \cdot \exp\left(-\frac{1}{2} \cdot \|\omega\|^2 \cdot \Delta\right) \text{ with accuracy } \approx n.$$

42. Taking Noise into Account (cont-d)

- The values of the Fourier transform are, in general, complex numbers.
- A complex number z can be characterized by its absolute value (modulus) $|z|$ and its phase.
- As one can see from the formulas, Δ does not affect the phases, so it is sufficient to consider absolute values:

$$|F_2(\omega)| \approx |F_1(\omega)| \cdot \exp\left(-\frac{1}{2} \cdot \|\omega\|^2 \cdot \Delta\right) \text{ with accuracy } \approx n.$$

- The above formula non-linearly depends on Δ.
- We can reduce this dependence to linear by using logarithms $\ell_i(\omega) \stackrel{\text{def}}{=} \ln(|F_i(\omega)|)$:

$$\ell_2(\omega) \approx \ell_1(\omega) - \frac{1}{2} \cdot \|\omega\|^2 \cdot \Delta \text{ with accuracy } \approx \frac{n}{|F_2(\omega)|}.$$

43. Resulting Formulas

- *Reminder:* for $\ell_i(\omega) = \ln(|F_i(\omega)|)$, we have:

$$\ell_2(\omega) \approx \ell_1(\omega) - \frac{1}{2} \cdot \|\omega\|^2 \cdot \Delta \text{ with accuracy} \approx \frac{n}{|F_2(\omega)|}.$$

- So, $\Delta \approx \dfrac{2(\ell_1(\omega) - \ell_2(\omega))}{\|\omega\|^2}$ with accuracy $\approx \dfrac{2n}{|F_2(\omega)| \cdot \|\omega\|^2}$.

- The Least Square Method for this problem leads to

$$\Delta = 2 \cdot \frac{\int (\ell_1(\omega) - \ell_2(\omega)) \cdot |F_2(\omega)|^2 \cdot \|\omega\|^2 \, d\omega}{\int |F_2(\omega)|^2 \cdot \|\omega\|^4 \, d\omega}.$$

- Here, integration is over frequencies ω for which $|F_2(\omega)| \geq c \cdot n$ for some pre-selected $c \gg 1$.

- *Preliminary results* show that this method correctly reconstructs the relative spatial resolution.

Part IV
Fuzzy-Related Ideas Help in Hypothesis Testing

44. Need for t-Tests

- Biomedical researchers continuously look for possible relations between relevant quantities.
- Such relations may help in preventing and curing diseases.
- Once a hypothesis is made about such a relation, it is necessary to test whether it is confirmed by the data.
- For such hypothesis testing, t-tests are most widely used.
- A t-test can check, whether two samples come from distributions with the same mean.
- Example: checking whether the average blood pressure decreases after a proposed treatment.

45. Need to Preserve Privacy

- In traditional statistics, we assume that we know the exact values of the corresponding quantities.
- In biomedical research, however, it is important to preserve patients' privacy and confidentiality.
- Knowing the exact values of age, height, weight, etc., one can uniquely identify the patient.
- One of the most efficient ways to preserve privacy is thus to replace the exact values with intervals containing such values.
- Example: instead of the exact age, we only store an interval containing this age:
  - between 20 and 30, or
  - between 30 and 40, etc.

46. Resulting Computational Challenge

- We want to estimate the value of a statistic s.
- We know how the statistic depends on the sample values $x_1, \ldots, x_n$.
- For example, for the t-test, we estimate a statistic t.
- The hypothesis is confirmed, with given confidence α, if this value is below a certain threshold t_α: $t \in [0, t_\alpha]$
- Example: the mean is $s = \dfrac{1}{n} \cdot \sum_{i=1}^{n} x_i$.
- For privacy-protected data, instead of the exact values x_i, we only know the intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$.
- Different values $x_i \in \mathbf{x}_i$ lead, in general, to different values of the corresponding statistic s.
- In particular, for different $x_i \in \mathbf{x}_i$ and $y_i \in \mathbf{y}_i$, we have different values $t(x_1, \ldots, x_n, y_1, \ldots, y_n)$.

47. Computational Challenge (cont-d)

- To confirm the hypothesis, we need to check that $t(x_1, \ldots, y_1, \ldots) \le t_\alpha$ for all $x_i \in \mathbf{x}_i$ and $y_i \in \mathbf{y}_i$.
- This is equivalent to $\overline{t} \le t_\alpha$, where

$$\overline{t} \stackrel{\text{def}}{=} \max\{t(x_1, \ldots, y_1, \ldots) : x_i \in \mathbf{x}_i, y_i \in \mathbf{y}_i\}.$$

- To reject the hypothesis, we need to check that $t(x_1, \ldots, y_1, \ldots) > t_\alpha$ for all $x_i \in \mathbf{x}_i$ and $y_i \in \mathbf{y}_i$.
- This is equivalent to $\underline{t} > t_\alpha$, where

$$\underline{t} \stackrel{\text{def}}{=} \min\{t(x_1, \ldots, y_1, \ldots) : x_i \in \mathbf{x}_i, y_i \in \mathbf{y}_i\}.$$

- Thus, we need to compute the range

$$[\underline{t}, \overline{t}] = \{t(x_1, \ldots, y_1, \ldots) : x_i \in \mathbf{x}_i, y_i \in \mathbf{y}_i\}.$$

48. Interval Computations

- Computation under interval uncertainty about inputs is known as *interval computations*.
- In general, computing the range is NP-hard.
- This means, crudely speaking, that no feasible algorithm can solve all instances of this problem.
- In some cases, feasible algorithms are possible.
- For example, it is easy to compute the range of the mean $s = \frac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$.
- Since this function in monotonic in all x_i, the range is

$$[\underline{s}, \overline{s}] = \left[\frac{1}{n} \cdot \sum_{i=1}^{n} \underline{x}_i, \frac{1}{n} \cdot \sum_{i=1}^{n} \overline{x}_i\right].$$

- We provide efficient algorithms for computing t-tests under privacy-motivated interval uncertainty.

49. Versions of t-Test: Reminder

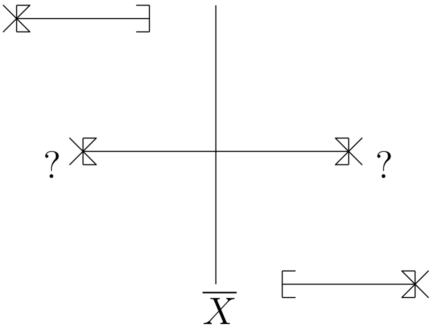
- Mean $\overline{X}=\dfrac{1}{n}\cdot\sum\limits_{i=1}^{n}x_i$, variance $s^2=\dfrac{1}{n-1}\cdot\sum\limits_{i=1}^{n}(x_i-\overline{x})^2$.
- For testing that the actual mean μ is μ_0: $t=\dfrac{\overline{X}-\mu_0}{s/\sqrt{n}}$.
- For testing that the means are equal ($\mu_1=\mu_2$), case of equal sample sizes $n_1=n_2$ and equal variance:
  $t=\dfrac{\overline{X}_1-\overline{X}_2}{s_{X_1X_2}\cdot\sqrt{2/n}}$, where $s_{X_1X_2}=\sqrt{\dfrac{1}{2}\cdot(s^2_{X_1}+s^2_{X_2})}$.
- Case of unequal sample sizes $n_1\ne n_2$, equal variance:

$$t=\frac{\overline{X}_1-\overline{X}_2}{s_{X_1X_2}\cdot\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}},\ s_{X_1X_2}\stackrel{\text{def}}{=}\sqrt{\frac{(n_1-1)s^2_{X_1}+(n_2-1)s^2_{X_2}}{n_1+n_2-2}}.$$

- General case: $t=\dfrac{\overline{X}_1-\overline{X}_2}{s_{\overline{X}_1-\overline{X}_2}}$, $s_{\overline{X}_1-\overline{X}_2}\stackrel{\text{def}}{=}\sqrt{\dfrac{s^2_{X_1}}{n_1}+\dfrac{s^2_{X_2}}{n_2}}$.

50. Intuitive Idea

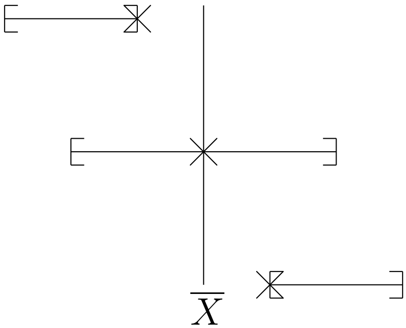
- All expressions for t have the form $\dfrac{\overline{X}}{s}$.
- Thus, the smallest value $\underline{t}$ of t is attained when s is the largest.
- This means that for each i, we select x_i which are as far from the mean as possible.



- For intervals $[\underline{x}_i, \overline{x}_i]$ containing $\overline{X}$, we have two options: $x_i = \underline{x}_i$ and $x_i = \overline{x}_i$.
- For all other intervals $[\underline{x}_i, \overline{x}_i]$, we have only one option.

51. Idea (cont-d)

- Similarly, the largest value $\overline{t}$ of t is attained when s is the smallest.
- This means that for each i, we select x_i which are as close from the mean as possible.



- For each interval $[\underline{x}_i, \overline{x}_i]$, we have only one option:
  - if $\overline{x}_i < \overline{X}$, then $x_i = \overline{x}_i$;
  - if $\overline{X} < \underline{x}_i$, then $x_i = \underline{x}_i$;
  - if $\underline{x}_i \leq \overline{X} \leq \overline{x}_i$, then $x_i = \overline{X}$.

52. Towards Algorithm for $\bar{t}$

- A function $f(x)$ attains its maximum on $[\underline{x}, \overline{x}]$:
  - either inside the interval, then $\dfrac{df}{dx} = 0$;
  - or for $x_i^M = \underline{x}$, then $\dfrac{df}{dx} \leq 0$;
  - or for $x^M = \overline{x}$, then $\dfrac{df}{dx} \geq 0$.
- So, for every i, when the maximum $t = \bar{t}$ is attained:
  - either when $\underline{x}_i < x_i^M < \overline{x}_i$ and $\dfrac{\partial t}{\partial x_i} = 0$;
  - or when $x_i^M = \underline{x}_i$ and $\dfrac{\partial t}{\partial x_i} \leq 0$;
  - or when $x_i^M = \overline{x}_i$ and $\dfrac{\partial t}{\partial x_i} \geq 0$.
- Here, $\dfrac{\partial t}{\partial x_i} \sim x_i - c$ for some quadratic c indep. on i.

53. Towards Algorithm for $\bar{t}$ (cont-d)

- When $\underline{x}_i \le c \le \overline{x}_i$, we cannot have $x_i^M = \underline{x}_i$ and $x_i^M = \overline{x}_i$, so ${}_i^M$ is in between, so $\dfrac{\partial t}{\partial x_i} = 0$ and $x_i^M = c$.
- Similarly, when $\overline{x}_i \le c$, we have $x_i^M = \overline{x}_i$.
- When $c \le \underline{x}_i$, we have $x_i^M = \underline{x}_i$.
- In all three cases, x_i^M is the point closest to c.
- Let's sort all endpoints of the intervals:

$$x_{(1)} \le x_{(2)} \le \ldots$$

- The value c is in one of the zones $[x_{(k)}, x_{(k+1)}]$.
- For each zone k, for each i, we either know x_i^M, or we know that $x_i^M = c$.
- Substituting $x_i = x_i^M$ into the quadratic expression $c(x_1, \ldots, x_n)$, we get a quadratic equations for c.

54. Algorithm for $\bar{t}$ (cont-d)

- After solving the quadratic equation, we find c.
- Thus, we know all the values x_i.
- Based on these values, we compute the value t corresponding to the k-th zone.
- We repeat this for each pair of X_1- and X_2-zone.
- The largest of the computed values t is the desired maximum $\bar{t}$.
- For a sample of size n, we have $2n$ bounds, so we have $2n + 1 = O(n)$ zones.
- Thus, we have $O(n) \cdot O(n) = O(n^2)$ pairs of zones.
- For each pair of zone, we need $O(n)$.
- Thus, overall, we need $O(n^2) \cdot O(n) = O(n^3)$ steps.
- So, our algorithm is indeed feasible.

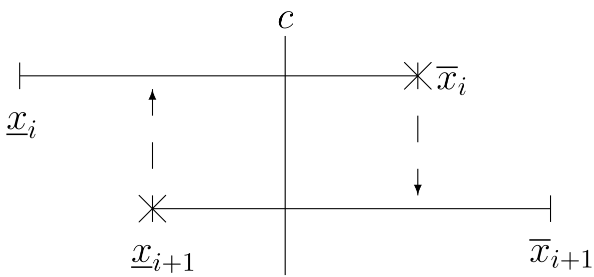
55. Towards Algorithm for $\underline{t}$

- A function $f(x)$ attains its minimum on $[\underline{x}, \overline{x}]$:
  - either inside the interval, then $\dfrac{df}{dx} = 0$;
  - or for $x^m = \underline{x}$, then $\dfrac{df}{dx} \geq 0$;
  - or for $x^m = \overline{x}$, then $\dfrac{df}{dx} \leq 0$.
- So, for every i, when the minimum $t = \overline{t}$ is attained:
  - either when $\underline{x}_i < x_i^m < \overline{x}_i$ and $\dfrac{\partial t}{\partial x_i} = 0$;
  - or when $x_i^m = \underline{x}_i$ and $\dfrac{\partial t}{\partial x_i} \geq 0$;
  - or when $x_i^m = \overline{x}_i$ and $\dfrac{\partial t}{\partial x_i} \leq 0$.
- Here, $\dfrac{\partial t}{\partial x_i} \sim x_i - c$ for some quadratic c indep. on i.

56. Towards Algorithm for $\underline{t}$ (cont-d)

- When $c < \underline{x}_i$, we cannot have $x_i^m = \underline{x}_i$ and $\underline{x}_i < x_i^m < \overline{x}_i$, so $x_i^m = \overline{x}_i$.
- Similarly, when $\overline{x}_i < c$, we have $x_i^m = \underline{x}_i$.
- When $\underline{x}_i \le c \le \overline{x}_i$, we can have both $x_i^m = \underline{x}_i$ and $x_i^m = \overline{x}_i$.
- For privacy data, intervals $[\underline{x}_i, \overline{x}_i]$ can be sorted so that $\underline{x}_i \le \underline{x}_{i+1}$ and $\overline{x}_i \le \overline{x}_{i+1}$.
- Let us show that min is attained when $x_i^m \le x_i^{m+1}$.
- Indeed, the only possibility for $x_i^m \le x_i^{m+1}$ is when both intervals contain c, $x_i^m = \overline{x}_i$, and $x_{i+1}^m = \underline{x}_{i+1}$.
- In this case, since t is symmetric w.r.t. all x_i we can swap these values and take $x_i^m = \underline{x}_{i+1}$, and $x_{i+1}^m = \overline{x}_i$.

57. Swap and Resulting Algorithm



- We see that the resulting tuple is not minimizing.
- Thus, there exists k for which the minimizing sequence x_i^m has the form $(\underline{x}_1, \ldots, \underline{x}_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$.
- We have such thresholds k_1 and k_2 for both samples.
- There are n^2 pairs of such thresholds.
- For each pair, we know the values x_i and thus, we can compute t by using time $O(n)$.
- The smallest of these values t is the desired value $\underline{t}$.

Home Page
Title Page
◀◀ ▶▶
◀ ▶

Go Back
Full Screen
Close
Quit

58. This Algorithm Is Feasible and Can Be Further Improved

- The algorithm takes time $O(n^2) \cdot O(n) = O(n^3)$ and is, thus, feasible.
- How can we make computations faster?
- When we change from k to $k+1$, only one value changes x^m_{k+1}, from $\underline{x}_{k+1}$ to $\overline{x}_{k+1}$.
- Thus, we can change $\overline{X}_i$ and S_{X_i} is $O(1)$ steps.
- With this improvement, we can compute $\underline{t}$ is time $O(n^2)$.

Part V

Fuzzy-Related Ideas Help in Estimating the Values of the Desired Quantities: A Practical Example

59. How to Gauge Risk

- Engineers estimate the largest strength s_0 of historic floods and other natural disasters.
- Then they design the buildings so that they can withstand such disasters.
- However, there is always a possibility that the disaster strength S exceeds s_0.
- *Examples:* hurricane Katrina, Fukushima, etc.
- We cannot guarantee that $S \le s_0$.
- So, we should at least require $p = \text{Prob}(S > s_0) \le p_0$ for some small p_0.
- E.g., for manned space flights, NASA used $p_0 = 10^{-3}$.
- For reliability of a cell in a computer memory, we need $p_0 \ll 10^{-9}$: else one of the cells will be always faulty.

60. How to Gauge Risk (cont-d)

- It is also desirable to know how much damage will come, on average, if the threshold x_0 is exceeded.
- For each possible value S of the corresponding disaster strength, we estimate the corresponding damage X.
- Let x_p denote the damage corresponding to s_0, then

$$S \geq x_0 \text{ if and only if } X \geq x_p.$$

- Thus, we need to know the *expected shortfall*

$$\text{ES}_p \stackrel{\text{def}}{=} E[X \mid X \geq x_p].$$

- The values x_p and ES_p is how we gauge the risk.
- Similar two measures are used in finance to describe the risk that an investment would result in a big loss.

61. How to Estimate ES_p: Ideal Case

- In the ideal case, we know the probability distribution that describes possible values of the damage X.
- A distribution is usually described by its *cumulative distribution function* (cdf) $F(x) \stackrel{\text{def}}{=} \text{Prob}(X \le x)$.
- The probability p_0 of exceeding the threshold x_p is equal to $1 - F(x_p)$, so $F(x_p) = 1 - p_0 = p$.
- For each p, the value x_p for which $F(x_p) = p$ is known as the p-th *quantile*:
  - for $p = 0.5$, we get the median;
  - for $p = 0.25$ and $p = 0.75$, we get *quartiles*, etc.
- The conditional expectation can then be computed as the ratio $\text{ES}_p = \dfrac{\int_{x_p}^{\infty} x\, dF(x)}{1 - p}$.

62. In Practice, We Only Have Partial Information About the Probabilities

- In practice, we rarely know the exact values of all the probabilities:
  - instead of the exact values $F(x)$ corresponding to different values x,
  - we only know an *interval* $[\underline{F}(x), \overline{F}(x)]$ that contains the actual (unknown) value $F(x)$.
- Such interval-valued cdf is known as a *probability box* (*p-box*, for short).
- More generally:
  - we may have several intervals $[\underline{F}(x), \overline{F}(x)]$;
  - these intervals correspond to different degrees of certainty $\alpha \in [0, 1]$.
- So, $F(x)$ is a sequence of embedded intervals, i.e., in effect, a *fuzzy number*.

63. How to Gauge Risk Under Interval (p-Box) and Fuzzy Uncertainty?

- For different cdfs $F(x) \in [\underline{F}(x), \overline{F}(x)]$ within a p-box, we get different quantiles x_p:
  - the smallest value x_p corresponds to the largest values $\overline{F}(x)$ of the cdf; while
  - the largest value x_p corresponds to the smallest values $\underline{F}(x)$ of the cdf.
- Thus, possible values of the quantile x_p form an interval $[\underline{x}_p, \overline{x}_p]$ in which $\overline{F}(\underline{x}_p) = \underline{F}(\overline{x}_p) = p$.
- To handle the fuzzy case, we take into account that:
  - for all $y = f(x_1, \ldots, x_n)$ with fuzzy x_i,
  - the alpha-cut ${}^\alpha\mathbf{y} \stackrel{\text{def}}{=} \{y : \mu(y) \geq \alpha\}$ of the result is equal to the range

$$f({}^\alpha\mathbf{x}_1, \ldots, {}^\alpha\mathbf{x}_n) = \{f(x_1, \ldots, x_n) : x_1 \in {}^\alpha\mathbf{x}_1, \ldots, x_n \in {}^\alpha\mathbf{x}_n(\alpha)\}.$$

64. Need to Gauge Risk Under Interval (p-Box) and Fuzzy Uncertainty (cont-d)

- So, to find the α-cut of the quantile x_p, we can:
  - compute the interval $[\underline{x}_p, \overline{x}_p]$
  - when each $F(x)$ belongs to the corresponding α-cut of the fuzzy number $\mathbf{F}(x)$.
- This straightforward computation is possible since the dependence of x_p on $F(x)$ is monotonic.
- So, the largest values of x_p is attained for smallest $F(x)$, and vice versa.
- For ES_p, there is no such clear monotonicity.
- We thus need a new algorithm for estimating ES_p under interval and fuzzy uncertainty.

65. What We Do

- We provide efficient algorithms for computing ES_p under interval (p-box) and fuzzy uncertainty.
- From the algorithmic viewpoint:
  - the problem of computing the expected shortfall under fuzzy uncertainty
  - can be reduced to the case of interval (p-box) uncertainty.
- Thus, we only need an algorithm for the interval (p-box) uncertainty.

66. Algorithm: Case of Interval Uncertainty

- We are given a p-box $[\underline{F}(x), \overline{F}(x)]$ and a probability p.
- We want to find the range $[\underline{\text{ES}}_p, \overline{\text{ES}}_p]$ of possible values of ES_p when cdf $F(x)$ is in this p-box.
- First, we compute $\overline{\text{ES}}_p$ as ES_p corresponding to $F(x) = \underline{F}(x)$, i.e., as the ratio:

$$\frac{1}{1-p} \cdot \int_{\overline{x}_p}^{\infty} x\, d\underline{F}(x), \text{ where } \overline{x}_p \stackrel{\text{def}}{=} (\underline{F})^{-1}(p).$$

- Then, we compute $\underline{\text{ES}}_p$ as ES_p corresponding to $F(x) = \overline{F}(x)$, i.e., as the ratio:

$$\frac{1}{1-p} \cdot \int_{\underline{x}_p}^{\infty} x\, d\overline{F}(x), \text{ where } \underline{x}_p \stackrel{\text{def}}{=} (\overline{F})^{-1}(p).$$

67. Algorithm: Case of Fuzzy Uncertainty

- We have a fuzzy-valued cdf $\mathbf{F}(x)$, i.e., we have the α-cuts ${}^\alpha\mathbf{F}(x) = [{}^\alpha\underline{F}(x), {}^\alpha\overline{F}(x)]$.
- We are also given a probability p.
- We want to to compute the α-cuts ${}^\alpha\mathbf{ES}_p = [{}^\alpha\underline{\mathrm{ES}}_p, {}^\alpha\overline{\mathrm{ES}}_p]$ of the expected shortfall $\mathbf{ES}_p$.
- First, we compute ${}^\alpha\underline{\mathrm{ES}}_p$ as ES_p corresponding to ${}^\alpha\underline{F}(x)$, i.e., as the ratio

$$\frac{1}{1-p} \cdot \int_{{}^\alpha\overline{x}_p}^\infty x\, d\,{}^\alpha\underline{F}(x), \text{ where } {}^\alpha\overline{x}_p \stackrel{\text{def}}{=} ({}^\alpha\underline{F})^{-1}(p).$$

- Then, we compute ${}^\alpha\overline{\mathrm{ES}}_p$ as ES_p corresponding to ${}^\alpha\overline{F}(x)$, i.e., as the ratio

$$\frac{1}{1-p} \cdot \int_{{}^\alpha\underline{x}_p}^\infty x\, d\,{}^\alpha\overline{F}(x), \text{ where } {}^\alpha\underline{x}_p \stackrel{\text{def}}{=} ({}^\alpha\overline{F})^{-1}(p).$$

68. Appendix: Analysis of the Problem

- We have $\text{ES}_p = \dfrac{1}{1-p}\cdot I$, where $I \stackrel{\text{def}}{=} \int_{x_p}^\infty x\,dF(x)$; so:
  - ES_p attains its smallest possible value $\underline{\text{ES}}_p$ when I attains its smallest possible value $\underline{I}$;
  - ES_p attains its largest possible value $\overline{\text{ES}}_p$ when I attains its largest possible value $\overline{I}$.
- The integral I has an infinite upper bound.
- This integral can be thus represented as a limit of integrals I_T with a finite upper bound T when $T\to\infty$:

$$I = \lim_{T\to\infty} I_T, \text{ where } I_T \stackrel{\text{def}}{=} \int_{x_p}^T x\,dF(x).$$

- Thus, for very large T, we have $I \approx I_T$.

69. Analysis of the Problem (cont-d)

- $I_T = \int_{x_p}^T x\,dF(x)$ can be integrated by part:

$$I_T = x \cdot F(x)|_{x_p}^T - \int_{x_p}^T F(x)\,dx =$$

$$T \cdot F(T) - x_p \cdot F(x_p) - \int_{x_p}^T F(x)\,dx.$$

- For large T, we have $F(T)$ practically equal to $\lim\limits_{T\to\infty} F(T) = 1$, so $T \cdot F(T) = T$.
- By definition of a quantile x_p, we have $F(x_p) = p$, so

$$I_T = T - x_p \cdot p - \int_{x_p}^T F(x)\,dx.$$

70. When Does the Expression $I_T = T - x_p \cdot p - \int_{x_p}^T F(x)\,dx$ Attain Its Largest Value?

- Let $F^{\max}(x)$ be a cdf for which I_T is the largest, and let $x_p^{\max}$ be the corresponding value of x_p.
- For fixed x_p, the integral I_T is a decreasing function of the values $F(x)$.
- Thus, I_T is the largest when all $F(x)$ are the smallest.
- We have two limitations on the values $F(x)$ for $x \ge x_p$:
  - $\underline{F}(x) \le F(x) \le \overline{F}(x)$ from a given p-box;
  - $F(x) \ge p$ from $F(x_p) = p$ and monotonicity of $F(x)$.
- These constraints $\underline{F}(x) \le F(x) \le \overline{F}(x)$ and $F(x) \ge p$ can be equivalently described by a single constraint

$$\max(\underline{F}(x), p) \le F(x) \le \overline{F}(x).$$

71. When I_T Attains Its Largest Value?

- Thus, the smallest possible values of $F(x)$ correspond to $F(x) = \max(\underline{F}(x), p)$.
- When $\underline{F}(x) \geq p$, we have $\max(\underline{F}(x), p) = \underline{F}(x)$ and hence $F(x) = \underline{F}(x)$.
- The equality $\underline{F}(x) = p$ is equivalent to $x = \overline{x}_p$, thus the condition $\underline{F}(x) \geq p$ is equivalent to $x \geq \overline{x}_p$.
- When $\underline{F}(x) < p$, i.e., when $x < \overline{x}_p$, then $F(x) = p$; so:

$$\int_{x_p^{\max}}^{T} F(x)\,dx = \int_{x_p^{\max}}^{\overline{x}_p} p\,dx + \int_{\overline{x}_p}^{T} \underline{F}(x)\,dx =$$

$$(\overline{x}_p - x_p^{\max}) \cdot p + \int_{\overline{x}_p}^{T} \underline{F}(x)\,dx, \text{ and}$$

$$I_T = T - x_p^{\max} \cdot p - (\overline{x}_p - x_p^{\max}) \cdot p - \int_{\overline{x}_p}^{T} \underline{F}(x)\,dx.$$

72. When I_T Attains Its Largest Value: Result

- We have shown that

$$I_T = T - x_p^{\max} \cdot p - (\overline{x}_p - x_p^{\max}) \cdot p - \int_{\overline{x}_p}^{T} \underline{F}(x)\,dx.$$

- The two terms $x_p^{\max} \cdot p$ and $(\overline{x}_p - x_p^{\max}) \cdot p$ can be easily combined into a single term $\overline{x}_p \cdot p$, so

$$I_T = T - \overline{x}_p \cdot p - \int_{\overline{x}_p}^{T} \underline{F}(x)\,dx.$$

- Here, $\overline{x}_p$ is the quantile corresponding to the lower endpoint $\underline{F}(x)$ of the p-box.
- So, we can conclude that the above expression is the value of the I_T corresponding to $F(x) = \underline{F}(x)$.
- Thus, the largest value of the integral I_T – and hence, of ES_p – is attained when $F(x) = \underline{F}(x)$.

73. When I_T Attains Its Smallest Value?

- Let $x_p^{\min}$ be the value corresponding to the cdf $F^{\min}(x)$ for which this integral is the largest possible.
- This means, in particular, that:
  - among all cdfs $F(x)$ with the same value of the p-th quantile $x_p^{\min}$ (i.e., for which $F(x_p^{\min}) = p$),
  - this particular cdf $F^{\min}(x)$ leads to the smallest possible value of the integral I_T.
- I_T is a decreasing function of the values $F(x)$.
- Thus, this integral is the smallest when all the values $F(x)$ are the largest.
- Under the limitations $\max(\underline{F}(x), p) \le F(x) \le \overline{F}(x)$, the largest possible values are $F(x) = \overline{F}(x)$.
- Thus, the smallest value of the integral I_T – and hence, of ES_p – is attained when $F(x) = \overline{F}(x)$.

74. General Conclusion

Fuzzy-related ideas are helpful on every stage of data and knowledge processing.

Part VI
How to Enhance the Use of Fuzzy Techniques

75. Relations Are Ubiquitous

- Many real-life quantities $x_1, \ldots, x_n$ are *related:*
  - once we know the value of one or more of the quantities,
  - this knowledge restricts possible values of other quantities.
- In some cases, we have a *functional* relation – the values of the quantities $x_1, \ldots, x_{n-1}$ uniquely determine x_n.
- *Example:* Ohm's law $V = I \cdot R$.
- In many other cases, however, we have relations which are not functional.
- In mathematical terms, a relation between real-valued quantities x_i is defined as a mapping $R : \mathbb{R}^n \to \{0, 1\}$:

$$R(x_1, \ldots, x_n) = 1 \Leftrightarrow (x_1, \ldots, x_n) \text{ is possible.}$$

76. Real-Life Relations Are Often Fuzzy

- In practice, about some combinations $(x_1, \ldots, x_n)$, we are not 100% sure whether they are possible.
- We must describe, for each combination $x = (x_1, \ldots, x_n)$, our degree of certainty that x is possible.
- In the computer, "true" is usually represented as 1, and "false" as 0.
- It is therefore natural to represent intermediate degrees of certainty as numbers from the interval $[0, 1]$:
  - the larger the number,
  - the larger our degree of confidence.
- The resulting mapping $R : \mathbb{R}^n \to [0, 1]$ is known as a *fuzzy relation*.

77. Need For a Concise Representation of a Fuzzy Relation

- Theoretically, each of the quantities x_i can have infinitely many different values.
- Due to measurement uncertainty, for each variable x_i, we only have finitely many distinguishable values

$$x_{i1}, \ldots, x_{ij}, \ldots, x_{iN_i}.$$

- In principle, we can store the degrees of certainty corresponding to all $N_1 \cdot \ldots \cdot N_n$ combinations.
- However, when n is large, the resulting number of values becomes astronomically high.
- We therefore need to come up with a more concise representation of fuzzy relations.

78. Fuzzy Rules As a Natural Concise Representation of Fuzzy Relations

- Many fuzzy relations come from *fuzzy rules*, i.e., from a combination of rules of the type

$$\text{“if } A_{r,1}(x_1) \text{ and } \ldots \text{ and } A_{r,n-1}(x_{n-1}) \text{ then } A_{r,n}(x_n)\text{”}.$$

- The ubiquity of such rules comes from the fact that this is how experts often describe their decisions.
- E.g.: “if a car in front is close, and it starts breaking seriously, one needs to hit the brakes hard right away”.
- Such rules use imprecise (fuzzy) words like “close”, “seriously”, “hard”, thus we need fuzzy techniques.
- One of the most common ways to formalize the fuzzy rules it the *Mamdani approach*.

79. Mamdani Approach: Reminder

- A tuple $(x_1, \ldots, x_n)$ is reasonable if for one of the rules, conditions and conclusions are satisfied:

$$(A_{1,1}(x_1) \,\&\, \ldots \,\&\, A_{1,n-1}(x_{n-1}) \,\&\, A_{1,n}(x_n)) \vee \ldots \vee$$

$$(A_{n_r,1}(x_1) \,\&\, \ldots \,\&\, A_{n_r,n-1}(x_{n-1}) \,\&\, A_{n_r,n}(x_n)).$$

- We use an "and"-operation (t-norm) $f_\&(a,b)$ and an "or"-operation (t-conorm) $f_\vee(a,b)$ to represent $\&$ and $\vee$:

$$d(x_1, \ldots, x_n) = f_\vee(d_1(x_1, \ldots, x_n), \ldots, d_{n_r}(x_1, \ldots, x_n)), \text{ where}$$

$$d_r(x_1, \ldots, x_n) = f_\&(A_{r,1}(x_1), \ldots, A_{r,n-1}(x_{n-1}), A_{r,n}(x_n)).$$

- Fuzzy rules are a natural concise way of representing a relation.

- Thus, it is reasonable to try to approximate a given fuzzy relation by an appropriate family of rules.

80. Why t-Norms and t-Conorms: Reminder

- Often, we know the expert's degrees of confidence $a = d(A)$ and $b = d(B)$ in two statements A and B.
- We want to estimate the expert's degree of confidence in $A \,\&\, B$ or $A \vee B$ based on a and b.
- The resulting estimates $f_{\&}(a,b)$ and $f_{\vee}(a,b)$ are known as "and"- and "or"-operations.
- The composite statements $A \,\&\, B$ and $B \,\&\, A$ are equivalent to each other.
- Thus, we require that the estimates $f_{\&}(a,b)$ and $f_{\&}(b,a)$ be equal, i.e., that $f_{\&}(a,b)$ be *commutative.*
- Also, $(A \,\&\, B) \,\&\, C$ and $A \,\&\, (B \,\&\, C)$ are equivalent, hence $f_{\&}$ should be associative: $f_{\&}(f_{\&}(a,b),c) = f_{\&}(a, f_{\&}(b,c))$.
- Similarly, $f_{\vee}$ should be commutative and associative.

81. What About Distributivity?

- Statements $A \,\&\, (B \vee C)$ and $(A \,\&\, B) \vee (A \,\&\, C)$ are also equivalent to each other.
- So why not require distributivity

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

- Distributivity does hold when $f_\vee(a, b) = \max(a, b)$.
- *Problem:* distributivity only holds for $f_\vee = \max$, while other t-conorms are also useful.
- *Proof:* for $b = c = 1$, we get $f_\&(a, 1) = f_\vee(f_\&(a, 1), f_\&(a, 1))$, hence $a = f_\vee(a, a)$, which implies max.
- *Solution:* require $f_\&(a, f_\vee(b, c)) = f_\vee(f_\&(a, b), f_\&(a, c))$ only when $f_\vee(b, c) < 1$.
- *Example:* $f_\&(a, b) = a \cdot b$, $f_\vee(a, b) = \min(a + b, 1)$.

82. Need to Describe Distributive "And"- and "Or"-Operations

- The main objective of this talk is to approximate a general fuzzy relation by fuzzy rules.
- We explained why it is reasonable to require that the "and"- and "or"-operations are distributive.
- Our goal is thus to approximate a general fuzzy relation by fuzzy rules that use distributive operations.
- We would like to produce an algorithm which is applicable for each distributive pairs of operations.
- So, to approach this approximate problem, let us see how we can describe such a generic pair.
- To come up with such a description, let us first recall how we can describe a generic "or"-operation.

83. Different Types of "Or"-Operations: Reminder

- Some "or"-operations are *Archimedean*: for all $a, b \in (0, 1)$ there is an n s.t. $f_\vee(a, a, \ldots, a)$ (n times) $> b$.
- Example: "algebraic sum" $f_\vee(a, b) = a + b - a \cdot b$.
- All such operations are isomorphic to addition: $f_\vee(a, b) = \psi^{-1}(\psi(a) + \psi(b))$ for some $\psi(a)$.
- We also have $\max(a, b)$ and operations isomorphic to $\min(a + b, 1)$: $f_\vee(a, b) = \psi^{-1}(\min(\psi(a) + \psi(b), 1))$.
- Every "or"-operation is isomorphic to a lexicographic combination of:
  - Archimedean operations,
  - max, and
  - operations isomorphic to $f_\vee(a, b) = \min(a + b, 1)$.

84. We Can Consider Approximate Descriptions

- The main purpose of "or"-operation $f_\vee(a,b)$ is to *estimate* the expert's degree of belief $d(A \vee B)$.
- We can thus replace an "or"-operation with a close one without changing its estimation accuracy.
- *Known:* for every "or"-operation $f_\vee(a,b)$ and for every $\varepsilon > 0$, there is an ε-close Archemedean $f'_\vee(a,b)$:

$$|f'_\vee(a,b) - f_\vee(a,b)| \le \varepsilon.$$

- Alas, for Archimedean operations, $b, c < 1$ imply $f_\vee(b,c) < 1$, so distributivity implies $f_\vee = \max$.
- We thus need to approximate by a non-Archimedean operation.
- *New Result:* for every $\varepsilon > 0$, each "or"-operation can be ε-approximated by an operation isomorphic to

$$\min(a+b, 1).$$

85. A New Universal Approximation Result: Idea of the Proof

- *Known:* each Archimedean "or"-operation has the form $f'_\vee(a,b) = \psi^{-1}(\psi(a)+\psi(b))$ for some function ψ.
- For every $\delta > 0$, consider a new function $\psi'(a)$ s.t.:
  - $\psi'(a) = \psi(a)$ for all $a \le 1-\delta$ and
  - $\psi'(a) = \psi(1-\delta) + (a-(1-\delta))$ for all $a \in (1-\delta, 1]$.
- For this function $\psi'(a)$, we form an "or"-operation

$$f''_\vee(a,b) \stackrel{\text{def}}{=} (\psi')^{-1}(\min(\psi'(a)+\psi'(b), \psi'(1)).$$

- When $\delta \to 0$, we have $f''_\vee(a,b) \to f'_\vee(a,b)$.
- Thus, for sufficiently small $\delta > 0$, $f''_\vee(a,b)$ is close to $f'_\vee(a,b)$ and thus, to $f_\vee(a,b)$.

- In terms of $\psi''(a) \stackrel{\text{def}}{=} \dfrac{\psi'(a)}{\psi'(1)}$, we get

$$f''_\vee(a,b) = (\psi'')^{-1}(\min(\psi''(a) + \psi''(b), 1)).$$

86. Let Us Use the New Approximation Result to Describe All Distributive Pairs

- Each "or"-operation can be, with arbitrary accuracy, approximated by an operation isom. to $\min(a+b,1)$.
- Thus, for all practical purposes, we can assume that the actual "or"-operation is isomorphic to $\min(a+b,1)$.
- In the corresponding scale, $c = f_\vee(a,b)$ takes the form $c' = \min(a'+b',1)$, and distributivity means

$$b' + c' < 1 \Rightarrow f_\&(a', b' + c') = f_\&(a', b') + f_\&(a', c').$$

- In other words, for each a', the function $b' \to f_\&(a', b')$ is a monotonic additive function of b'.
- It is known that all monotonic additive functions have the form $f(x) = k \cdot x$.
- Thus, we have $f_\&(a', b') = k(a') \cdot b'$ for some $k(a')$.

87. Let's Describe All Distributive Pairs (cont-d)

- For $f_\&(a', b') = k(a') \cdot b'$, commutativity means
$$k(a') \cdot b' = k(b') \cdot a'.$$
- Dividing both sides of this equality by $a' \cdot b'$, we conclude that $\dfrac{k(a')}{a'} = \dfrac{k(b')}{b'}$.
- In other words, we conclude that the ratio $\dfrac{k(a')}{a'}$ has the same value for all possible values $a' \in [0, 1]$.
- In other words, we conclude that this ratio is a constant; let us denote this constant by r.
- Then, from $\dfrac{k(a')}{a'} = r$, we conclude that $k(a') = r \cdot a'$.
- Therefore, $f_\&(a', b') = k(a') \cdot b' = r \cdot a' \cdot b'$.
- From the requirement that $f_\&(1, 1) = 1$, we conclude that $r = 1$ and thus, $f_\&(a', b') = a' \cdot b'$.

88. General Description of Distributive Pairs: Result

- Each "or"-operation can be, with arbitrary accuracy, approximated by an operation isom. to $\min(a+b,1)$.
- Thus, for all practical purposes, we can assume that the actual "or"-operation is isomorphic to $\min(a+b,1)$.
- Under this assumption, each distributive pair is isomorphic to the pair consisting of:
  - an "and"-operation $f_\vee(a,b)=\min(a+b,1)$ and
  - the algebraic-product "and"-operation

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

89. Approximating a Fuzzy Relation by Fuzzy Rules: What We Propose

- *Reminder:*
  - we have a fuzzy relation $R(x_1, \ldots, x_n)$,
  - we have a distributive pair of "and"- and "or"-operations, and
  - we want to represent R as

$$R(x_1, \ldots, x_n) = f_\vee(d_1(x_1, \ldots, x_n), \ldots, d_{n_r}(x_1, \ldots, x_n)),$$

$$\text{where } d_r(x_1, \ldots, x_n) = f_\&(A_{r,1}(x_1), \ldots, A_{r,n}(x_n)).$$

- Since operations are distributive, after rescaling $\psi''(a)$, we get $\min(a + b, 1)$ and product.
- Thus, the desired representation takes the form

$$R'(x_1, \ldots, x_n) = \min\left(\sum_{r=1}^{n_r} \prod_{i=1}^{n} A'_{ri}(x_i), 1\right).$$

90. What We Propose (cont-d)

- Approximating a function $R'(x_1, \ldots, x_n)$ by expressions $\sum\limits_{r=1}^{n_r} \prod\limits_{i=1}^{n} A'_{ri}(x_i)$ is known as *tensor decomposition.*
- Many efficient algorithms have been developed for solving this problem.
- We therefore propose to solve the original problem of approximating a relation $R(x_1, \ldots, x_n)$ as follows:
  - first, we form $R'(x_1, \ldots, x_n) = \psi''(R(x_1, \ldots, x_n))$;
  - then, we use a tensor decomposition algorithm to find the $A'_{ri}(x_i)$ approximating $R'(x_1, \ldots, x_n)$;
  - finally, we "re-scale" the resulting functions $A'_{ri}(x_i)$ back to the original scale, i.e., form functions

$$A_{ri}(x_i) \stackrel{\text{def}}{=} (\psi'')^{-1}(A'_{ri}(x_i)).$$

- Thus, we approximate the relation R by fuzzy rules.

91. Discussion

- We are interested in representations with non-negative values $A_{ri}(x_i)$.
- Most tensor decomposition algorithms allow representations with functions of arbitrary sign.
- So we may end up with negative values of $A_{ri}(x_i)$.
- This is OK if all we are interested in is approximation.
- However, we may want an approximation s.t. $A_{ri}(x_i)$ are membership functions, i.e., $A_{ri}(x_i) \geq 0$.
- To achieve this, we replace each negative value by the closest non-negative one, i.e., by 0.
- It should be mentioned, however, that this replacement may somewhat decrease the approximation accuracy.

Part VII

How to Go Beyond Simple $[0,1]$-Based Fuzzy Techniques

92. Fuzzy Logic: Reminder

- In the traditional two-valued logic, every statement is either true or false.
- In the computer these values are represented as, correspondingly 1 and 0.
- These two values cannot capture a situation when an expert is not 100% sure about his/her statement.
- To capture such expert uncertainty, L. Zadeh came up with an idea of fuzzy logic, where for each statement:
  - instead of two possible truth values 0 and 1,
  - we can have degrees of certainty that can take any values from 0 to 1.
- We now need to extend propositional operations from $\{0, 1\}$ to $[0, 1]$.

93. Fuzzy Logic (cont-d)

- We need to extend propositional operations from $\{0, 1\}$ to $[0, 1]$.
- From the purely mathematical viewpoint, there are many such extensions.
- It is desirable to preserve as many properties of the 2-valued logic as possible.
- Usually, "and"- and "or"-operations are selected to be commutative and associative.
- This still leaves us with plenty of different choices.
- It is therefore desirable, among all such operations, to select those that satisfy additional properties.

94. Distributivity

- One of such additional natural properties is distributivity, that $A \,\&\, (B \vee C)$ is equivalent to $(A \,\&\, B) \vee (A \,\&\, C)$:

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

- If we require this for all a, b, and c, then

$$f_\vee(a, b) = \max(a, b).$$

- It is known that sometimes, the expert's use of "or" is better described by other "or"-operations.
- It is reasonable to restrict the above equality to cases when $f_\vee(b, c) < 1$.
- Then, "and"- and "or"-operations are equivalent to $f_\&(a, b) = a \cdot b$ and $f_\vee(a, b) = \min(a + b, 1)$.

95. Need to Go Beyond $[0,1]$

- The $[0,1]$-based fuzzy logic captures many features of expert uncertainty.
- However, in some situations, it is not fully adequate to distinguish between different situations; e.g.:
  - if we have no information about a given statement,
  - then it makes sense to describe this uncertainty by the midpoint 0.5.
- On the other hand:
  - if have exactly as many arguments supporting S as supporting $\neg S$,
  - then it also makes sense to describe this uncertainty by the value 0.5.
- In both situations, the truth value is the same, but the uncertainty is different.

96. Need for 2-D Extensions

- If we add an argument in support of S, then:
  - in the first case, we now have an argument supporting S and no arguments supporting $\neg S$,
  - so the truth value of S should drastically increase;
  - in the second case, the numbers of statement supporting S and $\neg S$ remains almost equal;
  - so, the truth value should not change much.
- To distinguish between such situations, it is desirable:
  - to supplement the $[0, 1]$-valued degree of belief
  - with an additional number (or numbers).
- The simplest case: use *one* additional number.
- Thus, we use *two* numbers to describe our degree of certainty in a given statement.

97. 2-D Extensions Should Be Commutative, Associative, and Distributive

- From the commonsense viewpoint, logical operations are commutative, associative, and distributive.
- It is thus reasonable to require that the 2-D extensions of satisfy these three properties.
- The most widely used 2-D extension is *interval-valued* fuzzy logic.
- There, our degree of certainty in a statement is described by an interval $[\underline{d}, \overline{d}] \subseteq [0, 1]$.
- This enables us to clearly distinguish between the above two situations:
  - the case of complete uncertainty is naturally described by the interval $[0, 1]$, while
  - the case when equally many arguments for S and for $\neg S$ is described by $[0, 5, 0.5] = \{0.5\}$.

98. Distributive 2-D Extensions

- In principle, we can extend different t-norms and t-conorms to the interval-valued case:

$$f([\underline{a},\overline{a}],[\underline{b},\overline{b}])\stackrel{\text{def}}{=}\{f(a,b): a\in[\underline{a},\overline{a}] \text{ and } b\in[\underline{b},\overline{b}]\}.$$

- In particular, for $a\cdot b$ and $a+b$, we get

$$[\underline{a},\overline{a}]\cdot[\underline{b},\overline{b}]=[\underline{a}\cdot\underline{b},\overline{a}\cdot\overline{b}];\quad [\underline{a},\overline{a}]+[\underline{b},\overline{b}]=[\underline{a}+\underline{b},\overline{a}+\overline{b}].$$

- The resulting interval-valued logic is distributive.
- Another useful 2-D distributive extension of the usual fuzzy logic is the *complex-valued* fuzzy logic.
- In this logic, degrees of belief can take any complex values $a+b\cdot \mathrm{i}$, with $\mathrm{i}\stackrel{\text{def}}{=}\sqrt{-1}$.
- The complex-valued logic lacks a clear justification and clear interpretation.
- Thus, it is not as widely used an interval-valued one.

99. Are There Other Extension?

- At first glance, it looks like:
  - the above two extensions have been rather arbitrarily chosen, and
  - in principle, there are many other extensions.
- We show that interval-valued and complex-valued are the *only* possible 2-D distributive extensions.
- This result elevates complex-valued fuzzy logic:
  - from the status of one of the mathematically possible extensions
  - to a much higher status of one of the two possible extensions.
- This will, hopefully lead to a more frequent use of complex-valued fuzzy logic.

100. 2-D Logic: Set of Possible Values

- Let $\odot$ and $\oplus$ be 2-D extensions of $\cdot$ and $+$.
- Let x be a 2-D element different from real numbers.
- On this extended set, we want to allow multiplication.
- Thus, we need to consider elements of the type $b \odot x$ for arbitrary real numbers b.
- We also want to allow addition between real numbers a and the products $b \odot x$: $a \oplus (b \odot x)$.
- The set of all such elements depends on two parameters a and b and is, thus, 2-dimensional.
- We are interested in 2-D extensions.
- Thus, the desired extension cannot contain any other elements.
- So, each extension is the set of all the elements of the type $a \oplus (b \odot x)$.

101. Addition (“Or”-Operation) on the Set of Possible Values

- Due to commutativity and associativity of $\oplus$, we get

$$(a \oplus (b \odot x)) \oplus (a' \oplus (b' \odot x)) = (a \oplus a') \oplus ((b \odot x) \oplus (b' \odot x)).$$

- Here, a and a' are both real numbers, so

$$(a \oplus (b \odot x)) \oplus (a' \oplus (b' \odot x)) = (a + a') \oplus ((b \odot x) \oplus (b' \odot x)).$$

- Distributivity implies $(b \odot x) \oplus (b' \odot x) = (b \oplus b') \odot x$, so $(b \odot x) \oplus (b' \odot x) = (b + b') \odot x$.
- Substituting this expression into the above formula for $(a \oplus (b \odot x)) \oplus (a' \oplus (b' \odot x))$, we get

$$(a \oplus (b \odot x)) \oplus (a' \oplus (b' \odot x)) = (a + a') \oplus ((b + b') \odot x).$$

- In other words, we have a component-wise addition.

102. Multiplication ("And"-Operation) on the Set of Possible Values

- Due to distributivity, we have

$$(a \oplus (b \odot x)) \odot (a' \oplus (b' \odot x)) =$$

$$(a \odot a') \oplus ((a \odot b' + a' \odot b) \odot x) \oplus ((b \odot b') \odot (x \odot x)).$$

- Since for real numbers, the new operations $\odot$ and $\oplus$ are simply multiplication and addition, we get:

$$(a \oplus (b \odot x)) \odot (a' \oplus (b' \odot x)) =$$

$$(a \cdot a') \oplus ((a \cdot b' + a' \cdot b) \odot x) \oplus ((b \cdot b') \odot (x \odot x)).$$

- Thus, to describe the product of the new objects, it is sufficient to know the value of $x \odot x$.

- Since all the new elements have the form $a \oplus (b \odot x)$, we thus have $x \odot x = p \oplus (q \odot x)$ for some p and q.

103. Multiplication Simplified

- Let us show that we can simplify the formula for $x \odot x$ by re-selecting the element x.
- First, instead of x, we can select $x' = x \oplus \left(-\frac{q}{2}\right)$; then, $x' \odot x' = p'$, where $p' \stackrel{\text{def}}{=} p + \frac{q^2}{4}$.
- Thus, without losing generality, we can assume that $x \odot x = p$ for some real number p.
- We will consider 3 cases: $p > 0$, $p < 0$, and $p = 0$.
- When $p > 0$, we can simplify the above formula even more, by considering $x'' = \frac{1}{\sqrt{p}} \odot x$; then, $x'' \odot x'' = 1$.
- If we interpret $a \oplus (b \odot x'')$ as $[a - b, a + b]$, then $\oplus$ and $\odot$ become interval addition and multiplication.

104. Cases When $p < 0$ and When $p = 0$

- When $p < 0$, we can take $x'' = \dfrac{1}{\sqrt{|p|}} \odot x$, then $x'' \odot x'' = -1$, and we get *complex-valued* fuzzy logic.
- When $p = 0$, we get

  $$(a \oplus (b \odot x)) \odot (a' \oplus (b' \odot x)) = (a \cdot a') \oplus ((a \cdot b' + a' \cdot b) \odot x).$$

- Let us show that this formula corresponds to *linearized* approach to uncertainty.
- We are interested in a quantity y which depend on the directly measured quantities $x_1, \ldots, x_n$ as

  $$y = f(x_1, \ldots, x_n).$$

- We use the results $\widetilde{x}_i$ of measuring x_i to estimate y as

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

105. Case When $p = 0$ (cont-d)

- We assume that measurement results $\widetilde{x}_i$ are reasonably accurate.
- So, we can safely ignore the terms that are quadratic (or higher order) in terms of the measurement errors

$$\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i.$$

- Thus, $y = \widetilde{y} + \sum\limits_{i=1}^{n} y_i \cdot \Delta x_i$, where $y_i \stackrel{\text{def}}{=} -\dfrac{\partial f}{\partial x_i}$.
- If we have a second quantity $y' = \widetilde{y}' + \sum\limits_{i=1}^{n} y'_i \cdot \Delta x_i$, then their sum and product have the form

$$y + y' = (\widetilde{y} + \widetilde{y}') + \sum_i (y_i + y'_i) \cdot \Delta x_i;$$

$$y \cdot y' = (\widetilde{y} \cdot \widetilde{y}') + \sum_i (\widetilde{y} \cdot y'_i + \widetilde{y}' \cdot y_i) \cdot \Delta x_i.$$

106. Case When $p = 0$: Conclusion

- In particular, for $n = 1$, we have exactly the above formulas corresponding to $p = 0$:

$$y + y' = (\widetilde{y} + \widetilde{y}') + (y_1 + y_1') \cdot \Delta x_1;$$

$$y \cdot y' = (\widetilde{y} \cdot \widetilde{y}') + (\widetilde{y} \cdot y_1' + \widetilde{y}' \cdot y_1) \cdot \Delta x_1.$$

- Thus, the case $p = 0$ corresponds to a special case of interval-valued fuzzy logic:
  - when intervals are narrow
  - so that we can ignore terms which are quadratic in terms of their width.

107. Conclusion

We have thus shown that there are only two types of distributive 2-D fuzzy logic:

- interval-valued or
- complex-valued.