What Is the Uncertainty of the Result of Data Processing: Fuzzy Analogue of the Central Limit Theorem

Julio C. Urenda[1], Olga Kosheleva[1],
Shahnaz Shahbazova[2] and Vladik Kreinovich[1]

[1]University of Texas at El Paso
El Paso, Texas 79968, USA
jucrenda@utep.edu, olgak@utep.edu, vladik@utep.edu
[2]Azerbaijan Technical University
Baku, Azerbaijan
shahbazova@gmail.com

1. Formulation of the problem

- In the probabilistic approach to uncertainty, the most widely used probability distribution is normal (Gaussian).
- This fact has been empirically confirmed.
- For more than half of the measuring instruments, the probability distribution of the measurement error is close to Gaussian.
- This fact also has a theoretical explanation.
- In most cases, the measurement error is caused by a joint effect of many small factors.
- It is known that the distribution of the sum of a large number of small independent random variables is close to Gaussian.
- This theoretical explanation is known as the *Central Limit Theorem*.

2. Formulation of the problem (cont-d)

- According to this theorem:
  - when the number of summed variables increases,
  - the probability distribution of their sum tends to Gaussian.
- This means exactly that as this number becomes large, the corresponding distribution is close to Gaussian.
- In many practical situations, we do not know the corresponding distributions.
- All we have is expert estimates for the approximation errors.
- These expert estimations are often described by using words from natural language like “small”, “approximately”, etc.
- A natural way to describe these estimates in precise computer-understandable terms is to use fuzzy logic.
- Fuzzy logic was specifically designed for translating natural-language knowledge into such a precise form.

3. Formulation of the problem (cont-d)

- It is reasonable to expect that:
  - if we combine many such estimates,
  - we should also get the resulting overall estimate in a specific form.
- What is this form?
- What is the resulting limit theorem – the analogue of the Central Limit Theorem?
- These are the questions that we study in this talk.

4. Outline of this talk

- First, we analyze the general problem of estimating uncertainty of the result of data processing.
- Then, we review the results related to the probabilistic case.
- After that, we formulate the corresponding fuzzy case as a mathematical problem.
- Finally, we provide a solution to this problem.

5. What is data processing: a brief reminder

- One of the main objectives of science and engineering is:
  - to predict what will happen in the world, and
  - to come up with devices and techniques to make this future most beneficial for us.
- The state of the world is characterized by the values of several quantities.
- For example, the state of the weather is described by temperature, humidity, wind speed, and wind direction.
- So, predicting the future state of the world means predicting the future values of these quantities.
- Similarly, each device, each control strategy can be characterized by some numbers.

6. What is data processing: a brief reminder (cont-d)

- E.g., if we control a car, then, at each moment of time, we need to describe:
  - the value of the acceleration (if any is needed), and – if needed –
  - the angular velocity with which the car is turning.
- So, coming up with the appropriate recommendations means estimating the values of the relevant quantities.
- In both cases, we need to find an estimate $\widetilde{y}$ of each of the desired quantities y:
  - based on all available relevant information,
  - i.e., based on the known estimates $\widetilde{x}_1, \ldots, \widetilde{x}_n$ of the corresponding quantities $x_1, \ldots, x_n$.

7. What is data processing: a brief reminder (cont-d)

- The estimates $\widetilde{x}_i$ may come from measurements or they may come from experts.
- In the following, we will denote the algorithm used for estimating the desired quantity y by $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.
- Running these algorithms is what is usually called *data processing*.

8. How do we select data processing algorithms?

- We select each data processing algorithm so as to best describe the relation between the corresponding quantities y and x_i.
- In other words, we select an algorithm f for which:
  - to the best of our knowledge,
  - the actual values of these quantities satisfy the relation

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

9. Need to take uncertainty into account

- Measurement results are never absolutely accurate.
- Expert estimates are usually even less accurate.
- In both cases, each available estimate $\widetilde{x}_i$ is, in general, different from the actual (unknown) value x_i of the corresponding quantity.
- In other words, there is, in general, a non-zero approximation error

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

- Because of this, the result $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ of data processing is, in general, different from the actual value $y = f(x_1, \ldots, x_n)$.
- There is an uncertainty $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$.
- For practical purposes, it is important to gauge this uncertainty.

10. Need to take uncertainty into account (cont-d)

- For example, suppose that we are prospecting for oil, and we are estimating that a certain area contains 200 million tons.
- Then our actions will depend on how accurate is this estimate:
  - if it is 200 ± 50, then we should start exploiting this area right away;
  - however, if it is 200 ± 300, then maybe there is no oil at all,
  - so it is better to perform further research before investing money in exploitation.

11. Data processing is often hierarchical

- Data processing is often hierarchical, in the following sense.
- Instead of processing all the inputs right away, we divide them into groups – e.g., by time and/or by geographic locations; then:
  - first, we process inputs from each group, resulting in estimates for the combined quantities $z_1, \ldots, z_m$, and
  - then, we use these estimates for z_j to estimate the desired value y.
- This is how votes are counted in nation-wide elections.
- This is how data is often processed.

12. Possibility of linearization

- In most practical situations, the approximation errors Δx_i are relatively small.
- Then, terms quadratic in Δx_i can be safely ignored.
- For example, even if $\Delta x_i \approx 20\%$, the square of this number is 4%, which is much smaller.
- So, we can take into consideration that $x_i = \widetilde{x}_i - \Delta x_i$:
  - expand the expression $\Delta y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(x_1, \ldots, x_n) = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n)$ in Taylor series, and
  - keep only terms linear in Δx_i in this expansion,
  - while ignoring quadratic (and higher order) terms:

$$\Delta y = c_1 \cdot \Delta x_1 + \ldots + c_n \cdot \Delta x_n, \text{ where } c_i \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_i}_{|(\widetilde{x}_1, \ldots, \widetilde{x}_n)}.$$

- This is the main expression that we will use in our analysis of uncertainty of the result of data processing.

13. Linearization in the hierarchical case

- In this case, in the first stage, we get

$$\Delta z_j = c_{j1} \cdot \Delta x_1 + \ldots + c_{jn} \cdot \Delta x_n.$$

- Here many of the coefficients c_{ji} – related to measurements x_i not from the group j – are 0s.

- Then, on the second stage, we get

$$\Delta y = c_1 \cdot \Delta z_1 + \ldots + c_m \cdot \Delta z_m.$$

14. Central Limit Theorem: reminder

- As we have mentioned, measurement errors are usually relatively small.
- Measurement errors corresponding to different measurements are usually independent.
- In practice, the value n is usually large.
- For example, to predict tomorrow's weather, we use thousands of recordings of weather conditions
  - at different locations
  - in different moments of time.
- To analyze an earthquake, we use thousands of values recorded by seismometers:
  - around it,
  - or even, for a serious earthquake, all around the world.

15. Central Limit Theorem: reminder (cont-d)

- Thus, our formula describes the sum of a large number of relatively small independent random variables.
- We have already mentioned earlier that:
  - under reasonable conditions, the resulting distribution is close to Gaussian;
  - this is what the Central Limit Theorem is about.
- Thus, in the probabilistic case, we can conclude, with high confidence, that:
  - in many practical situations,
  - the probability distribution of the uncertainty Δy with which we determine the result y of data processing is close to Gaussian.

16. Beyond the Central Limit Theorem

- As we have commented, the convergence to the Gaussian distribution occurs under some reasonable conditions.
- What happens in the general case – when these conditions are not satisfied?
- To answer this question, let us take into account that data processing is often hierarchical:
- If there is a limit theorem:
  - according to which the probability distributions of the sums are close to distributions of a certain type,
  - then all variables Δz_j have distributions of this type, as well as the variable Δy.
- Thus, these limit distributions must have the following property:
  - a linear combination of thus distributed independent variables
  - should have the distribution of exactly the same type.

17. Beyond the Central Limit Theorem (cont-d)

- In precise terms, when we say that we have a distribution of a certain type, we usually mean that:
  - there is a standard random variable ξ – e.g., normally distributed with mean 0 and standard deviation 1, and
  - all other distributions of this type has the same distribution as $d \cdot \xi$, for some constant d.
- In this case, if d_i is the value of the parameter d corresponding to Δz_j, then we can write Δz_j as $d_j \cdot \xi_j$.
- Then, $\Delta y = c_1 \cdot d_1 \cdot \xi_1 + \ldots + c_n \cdot d_n \cdot \xi_n$,, i.e., $\Delta y = a_1 \cdot \xi_1 + \ldots + a_n \cdot \xi_n$, where $a_j \stackrel{\text{def}}{=} c_j \cdot d_j$.

18. Beyond the Central Limit Theorem (cont-d)

- In these terms, the above requirement states:
  - that each linear combination of identically distributed random variables ξ_j should have the same type of distribution, i.e.,
  - for all possible values a_j, there should be the value a for which the sum $\sum a_j \cdot \xi_j$ has the same probability distribution as $a \cdot \xi$.
- Distributions with this property are known as *infinitely divisible*.
- Gaussian distribution clearly has this property.
- There are other distributions with this property – e.g., Cauchy distribution, with the probability density function

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}.$$

19. What would a limit theorem mean in the fuzzy case: analysis of the problem

- A similar argument can be repeated for the fuzzy case, when:
  - instead of probability distributions,
  - we have membership functions.
- These functions describe:
  - for each possible value x of the corresponding quantity,
  - the degree (scaled to the interval $[0, 1]$) to which this value is possible.
- In this case, similarly to the probabilistic case, the existence of the limit theorem would mean that:
  - all linear combinations
  - are characterized by the same type of membership functions.

20. What would a limit theorem mean in the fuzzy case: analysis of the problem (cont-d)

- This would mean, in particular, that:
  - if the quantities Δz_j are characterized by membership functions of this type,
  - then their linear combination is characterized by a membership function of the same type.
- What does it mean "of the same type"?
- Similarly to the probabilistic case, a natural interpretation is that:
  - we should select one single membership function $\mu_0(x)$, and
  - consider membership functions that describe quantities of the type $d \cdot \xi$, where ξ is described by a membership function $\mu_0(x)$.
- What is the membership function of the quantity $d \cdot \xi$?
- To answer this question, let us recall that we can use different measuring units to describe the same value of the physical quantity.

21. What would a limit theorem mean in the fuzzy case: analysis of the problem (cont-d)

- For example, to describe length, we can use meters, or we can use centimeters:
  - if we replace the original measuring unit with a new one which is d times smaller,
  - then all numerical values are multiplied by d;
  - e.g., 2 meters becomes $2 \cdot 100 = 200$ centimeters.
- In general, the original numerical value x in the new scale is represented as $x' = d \cdot x$.
- Vice versa, the new value x' corresponds, in the original scale, to the value $x = x'/d$.

22. What would a limit theorem mean in the fuzzy case: analysis of the problem (cont-d)

- Thus:
  - if, in the original scale, the degree to which the value x is possible is $\mu_0(x)$,
  - then the degree $\mu(x')$ to which the value x' on the new scale is equal to $\mu_0(x'/d)$.
- So, quantities $d \cdot x$ are described by membership functions $\mu_0(x/d)$.
- In these terms, "membership function of the same type" means that we have a membership function of the type $\mu_0(x/d)$.
- For example, this means that:
  - the membership function of each quantity Δz_j is the same as
  - the membership function of the product $d_j \cdot \xi_j$, where ξ_j has the membership function $\mu_0(x)$.

23. What would a limit theorem mean in the fuzzy case: analysis of the problem (cont-d)

- Thus, if there is a limit theorem, then, similarly to the probabilistic case, we conclude that:
  - if we have several quantities $\xi_1, \ldots, \xi_m$ with the same membership function $\mu_0(x)$,
  - then the membership function for a linear combination (4) should have the same membership function $\mu_0(x/a)$ as the quantity $a \cdot \xi$.
- To describe this requirement in precise terms, let us recall how we can find the membership function corresponding to a linear combination.

24. How to find a membership function corresponding to a linear combination: Zadeh's extension principle

- The value x is a possible value of the linear combination if there are some values ξ_j:
  - which are possible and
  - whose linear combination (4) is equal to x.
- In general, "there exists" means that either this property holds for one combination of values ξ_j or for another combinations of values, etc.:

$$(\xi_1 \text{ is possible } \mathit{and} \ldots \mathit{and}\ \xi_n \text{ is possible } \mathit{and} \sum_{j=1}^{m} a_j \cdot \xi_j = x)\ \mathit{or}$$
$$(\xi'_1 \text{ is possible } \mathit{and} \ldots \mathit{and}\ \xi'_n \text{ is possible } \mathit{and} \sum_{j=1}^{m} a_j \cdot \xi'_j = x)\ \mathit{or}$$
$$\ldots$$

- Here, "or" combines all tuples $(\xi_1, \ldots, \xi_m)$ for which $\sum\limits_{j=1}^{m} a_j \cdot \xi_j = x$.

25. Zadeh's extension principle (cont-d)

- We know that all quantities ξ_j are described by the same membership function $\mu_0(x)$.
- This means that we know, for each value ξ_j, the degree to which this value is possible – this degree is equal to $\mu_0(\xi_j)$.
- According to the general fuzzy methodology:
  - to find the degree of confidence in the above "and"-"or"-combination of such statements,
  - we need to use appropriate "and"- and "or"-operations $f_\&(a, b)$ and $f_\vee(a, b)$ – also known as t-norms and t-conorms.

26. Zadeh's extension principle (cont-d)

- Thus, the desired degree $\mu(x)$ has the form

$$f_\vee\left(f_\&\left(\mu_0(\xi_1),\ldots,\mu_0(\xi_m),d\left(\sum_{j=1}^m a_j\cdot\xi_j=x\right)\right),\right.$$

$$\left.f_\&\left(\mu_0(\xi'_1),\ldots,\mu_0(\xi'_m),d\left(\sum_{j=1}^m a_j\cdot\xi'_j=x\right)\right),\ldots\right).$$

- Which "or"-operation should we choose?
- To make this choice, we need to take into account that:
  - there are infinitely many tuples ξ_j with the desired value x of the linear combination, and thus,
  - infinitely many terms combined by "or".

27. Zadeh's extension principle (cont-d)

- For most "or"-operations (e.g., for $a + b - a \cdot b$):
  - as we combine more and more statements,
  - we will get closer and closer to 1.
- To avoid such a meaningless result, we need to use the only operation that does not increase the value – namely, the operation maximum.
- In this case, we get

$$\mu(x) = \max_{\xi_1,\ldots,\xi_m} f_\& \left(\mu_0(\xi_1), \ldots, \mu_0(\xi_m), d\left(\sum_{j=1}^{m} a_j \cdot \xi_j = x \right) \right).$$

- Here, $d(S)$ is the degree to which the corresponding statement is true.

28. Zadeh's extension principle (cont-d)

- In our case, the statement $\sum\limits_{j=1}^{m} a_j \cdot \xi_j = x$ is either true or false.
  - If this statement is false, its degree is 0, so the whole combination has degree 0.
  - If this statement is true, then its degree is 1, and this does not affect the result of the "and"-operation, since $f_\&(a, 1) = a$.
- Thus, we have

$$\mu(x) = \max_{\xi_j : \sum\limits_{j=1}^{m} a_j \cdot \xi_j = x} f_\& \left(\mu_0(\xi_1), \ldots, \mu_0(\xi_m)\right).$$

- This formula – first derived by Zadeh – is known as *Zadeh's extension principle.*

29. Which "and"-operation should we use?

- We showed which "or"-operation to use.
- A natural next question is: which "and"-operation should we use?
- Some "and"-operations have the form $f_\&(a,b)=f^{-1}(f(a)\cdot f(b))$ for some strictly increasing function $f:[0,1]\to[0,1]$.
- Here $f^{-1}(x)$ denotes the inverse function.
- Such "and"-operations are known as *strictly Archimedean.*
- It is known, that for every "and"-operation $t(a,b)$ and for every $\varepsilon>0$, there exists a strictly Archimedean "and"-operation $f_\&(a,b)$ for which

$$|t(a,b)-f_\&(a,b)|\le\varepsilon \text{ for all } a \text{ and } b.$$

30. Which “and”-operation should we use (cont-d)

- The whole idea of an “and”-operation is that:
  - the value $t(a, b)$ estimates the expert’s degree of certainty in a statement $A \,\&\, B$
  - in a situation when we only know the expert’s degrees of certainty a and b in statements A and B.
- Experts can estimate their degree of certainty only with some accuracy:
  - we can usually distinguish between 7 and 8 on a 0-to-10 scale – which correspond to 0.7 and 0.8;
  - however, it is doubtful that anyone can distinguish between degrees of certainty 0.70 and 0.71;
  - these degrees correspond, for example, to marks 70 and 71 on a 0-to-100 scale.
- For sufficiently small ε, ε-close values are practically indistinguishable.

31. Which "and"-operation should we use (cont-d)

- So, in practice, it would not make any difference if:
  - we use an ε-close strictly Archimedean "and"-operation
  - instead of the original one $t(a,b)$.
- So, from the practical viewpoint, it makes sense to assume that the actual "and"-operation is strictly Archimedean.
- In this case, the above formula takes the following form:

$$\mu(x) = \max_{\xi_j : \sum_{j=1}^{m} a_j \cdot \xi_j = x} f^{-1}(f(\mu_0(\xi_1)) \cdot \ldots \cdot f(\mu_0(\xi_m))).$$

32. What does the limit property mean in this case

- The above limit property means that the function $\mu(x)$ as described by the above formula:
  - also has the same form as the membership function $\mu_0(x)$,
  - i.e., it has the form $\mu(x) = \mu_0(x/d)$ for some value d.
- So, the desired limit property takes the following form:
  - for each tuple $a_1, \ldots, a_m$,
  - there exists a value d for which

$$\mu_0(x/d) = \max_{\xi_j : \sum\limits_{j=1}^{m} a_j \cdot \xi_j = x} f^{-1}(f(\mu_0(\xi_1)) \cdot \ldots \cdot f(\mu_0(\xi_m))).$$

- Let us call membership functions $\mu_0(x)$ satisfying this property *limit membership functions.*
- So, the question is: which membership functions are the limit ones?

33. Let us simplify the problem

- In order to describe all possible limit membership functions, let us first simplify the above limit property as much as possible.
- First, let us avoid the explicit use of the inverse function – since computing the inverse function is, in general, not easy.
- We can achieve this if we apply the function $f(x)$ to both side of the above equality.
- We can take into account that this function is strictly increasing.
- So the largest (max) of its values is attained when x is the largest.
- Then we can conclude that

$$f(\mu_0(x/d)) = \max_{\xi_j : \sum\limits_{j=1}^{m} a_j \cdot \xi_j = x} (f(\mu_0(\xi_1)) \cdot \ldots \cdot f(\mu_0(\xi_m))).$$

- Now, let us make the constraint on ξ_j look simplest.

34. Let us simplify the problem (cont-d)

- For this purpose, let us denote by $v_j \stackrel{\text{def}}{=} a_j \cdot \xi_j$ the terms which are added in this constraint.
- In terms of these new variables v_j, we have $\xi_j = v_j/a_j$.
- So, in terms of v_j, the formula (8) takes the following form:

$$f(\mu_0(x/d)) = \max_{v_j : \sum\limits_{j=1}^{m} v_j = x} (f(\mu_0(v_1/a_1)) \cdot \ldots \cdot f(\mu_0(v_m/a_m))).$$

- A further simplification can be done if we realize that in this formula:
  - we only use the composition of the functions $f(x)$ and $\mu_0(x)$, but
  - we do not use the functions by themselves.
- To simplify the condition, let us therefore denote this composition by

$$\nu(x) \stackrel{\text{def}}{=} f(\mu_0(x)).$$

35. Let us simplify the problem (cont-d)

- In terms of this new function, the above formula takes the following form:
$$\nu(x/d)=\max_{v_j:\sum\limits_{j=1}^m v_j=x}(\nu(v_1/a_1)\cdot\ldots\cdot\nu(v_m/a_m)).$$
- Next, we can replace multiplication – which is more complex than addition – with addition.
- There is a function specifically designed for this purpose – the logarithm function, for which $\ln(a\cdot b)=\ln(a)+\ln(b)$.
- So, instead of using $\mu(x)$, it makes sense to use $\ln(\nu(x))$.
- Since the logarithm is also a strictly increasing function, we conclude that
$$\ln(\nu(x/d))=\max_{v_j:\sum\limits_{j=1}^m v_j=x}(\ln(\nu(v_1/a_1))+\ldots+\ln(\nu(v_m/a_m))).$$

36. Let us simplify the problem (cont-d)

- A further minor simplification comes from the fact that:
  - since the values $\nu(x)$ are smaller than equal to 1,
  - the logarithms of these values are negative (or 0).
- Since it is simpler to deal with positive numbers, let us multiply both sides of the above formula by -1.
- The corresponding operation $x \to -x$ is strictly decreasing, so it changes max to min.
- Thus, for the function $\ell(x) \stackrel{\text{def}}{=} -\ln(\nu(x))$, for which $\nu(x) = \exp(-\ell(x))$, we conclude that

$$\ell(x/d) = \min_{v_j : \sum\limits_{j=1}^{m} v_j = x} (\ell(v_1/a_1) + \ldots + \ell(v_m/a_m)).$$

37. Let us simplify the problem (cont-d)

- In particular, for $m = 2$, when $v_1 + v_2 = x$ and thus, $v_2 = x - v_1$, we conclude that

$$\ell(x/d) = \min_{v_1}(\ell(v_1/a_1) + \ell((x - v_1)/a_2)).$$

- Now, we are ready to analyze this formula.

38. We have reduced our problem to a known problem in convex analysis

- The above formula can be rewritten as

$$\ell_0(x) = \min_{v_1}(\ell_1(v_1) + \ell_2(x - v_1)).$$

- Here, we denoted

$$\ell_0(x) \stackrel{\text{def}}{=} \ell(x/d), \quad \ell_1(x) \stackrel{\text{def}}{=} \ell(x/a_1), \quad \ell_2(x) \stackrel{\text{def}}{=} \ell(x/a_2).$$

- The corresponding combination of the two function is known in *convex analysis*, as the *infimal covolution*, or an *epi-sum*.
- It is usually denoted by $\ell_0 = \ell_1 \,\Box\, \ell_2$.

39. We have reduced our problem to a known problem in convex analysis (cont-d)

- It is known that, under reasonable conditions, this formula can be further simplified if:
  - instead of the original functions $\ell_i(x)$,
  - we use their *Legendre-Fechnel transforms*

$$\ell_i^*(s) = \sup_x (s \cdot x - \ell_i(x)).$$

- Namely, it is known that:
  - the Legendre-Fechnel transform of the infimal convolution of two functions is equal to
  - the sum of their Legendre-Fechnel transforms:

$$\ell_0^*(s) = \ell_1^*(s) + \ell_2^*(s).$$

40. Let us use this reduction

- Let us describe the transform $\ell^*(s)$ of the function $\ell_i(x) = \ell(x/a_i)$ in terms of the Legendre-Fechner transform $F(s)$ of the function $\ell(x)$.
- Indeed, substituting the expression $\ell_i(x) = \ell(x/a_i)$ into the right-hand side of the formula fo f^*, we conclude that

$$\ell_i^*(s) = \sup_x (s \cdot x - \ell(x/a_i)).$$

- So, for the new variable $z \stackrel{\text{def}}{=} x/a_i$, for which $x = a_i \cdot z$, we conclude that

$$\ell_i^*(s) = \sup_z (s \cdot a_i \cdot z) - \ell(z)) = \sup_z ((s \cdot a_i) \cdot z) - \ell(z)), \text{ i.e., } \ell_i^*(s) = F(a_i \cdot s).$$

- Thus, the desired formula takes the following form:

$$F(d \cdot s) = F(a_1 \cdot s) + F(a_2 \cdot s).$$

41. Let us use this reduction (cont-d)

- The requirement is that for every a_1 and a_2, there exists a value $d = d(a_1, a_2)$ for which this property is satisfied.
- Differentiating both sides of this equality by a_2, we conclude that
$$s \cdot F'(a_2 \cdot s) = a \cdot s \cdot F'(d(a_1, a_2) \cdot s).$$
- Here we denoted $a \stackrel{\text{def}}{=} \dfrac{\partial d}{\partial a_2}_{|(a_1, a_2)}$.
- Dividing both sides by s, that $F'(a_2 \cdot s) = a(d, a_2) \cdot F'(c \cdot s)$.
- In particular, for $a_2 = 1$, we conclude that $F'(s) = a(d, 1) \cdot F'(d \cdot s)$, i.e., that $F'(d \cdot s) = A(d) \cdot F'(s)$, where we denoted $A(d) \stackrel{\text{def}}{=} \dfrac{1}{a(d, 1)}$.
- It is known that every continuous solution to this functional equation has the form $F'(s) = b \cdot s^\alpha$.
- Integrating, we conclude that $F(s) = B \cdot s^\beta + C$ for some constants B, β, and C.

42. Let us use this reduction (cont-d)

- Substituting $F(s) = B \cdot s^\beta + C$ into the desired condition, we conclude that $C = 0$ and thus, that $F(s) = B \cdot s^\beta$.
- It is known that:
  - if the Legendre-Fechnel transform of a function is a power law,
  - then the function itself is a power law.
- So $\ell(x) = D \cdot x^\gamma$ for some D and γ, and thus, the function $\nu(x) = \exp(-\ell(x))$ has the form $\nu(x) = \exp(-D \cdot x^\gamma)$.
- Thus, for $\mu(x) = f^{-1}(\nu(x))$, we have $\mu(x) = f^{-1}(\exp(-D \cdot x^\gamma))$.

43. Conclusion: fuzzy analogue of the Central Limit Theorem

- In the probabilistic case:
  - due to the Central Limit Theorem,
  - the uncertainty of the result of data processing is described by a Gaussian distribution
  - or, more generally, by an infinitely divisible distribution.
- Similarly, for the membership function $\mu(\Delta y)$ describing the uncertainty of the result of data processing:
  - when the "and"-operation is the algebraic product, then

$$\mu(\Delta y) = \exp(-D \cdot |\Delta y|^\gamma);$$

  - in general, when the "and"-operation has the form $f_\&(a, b) = f^{-1}(f(a) \cdot f(b))$, then

$$\mu(\Delta y) = f^{-1}(\exp(-D \cdot |\Delta y|^\gamma)).$$

44. 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.