# Measures of Specificity Used in the Principle of Justifiable Granularity: A Theoretical Explanation of Empirically Optimal Selections

Olga Kosheleva and Vladik Kreinovich

University of Texas at El Paso
El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 1. Granular Computing: a Brief Reminder

- In many practical situations, it is difficult to deal with the whole amount of data.
- It may be that we have too much data.
- Then, it is not feasible to apply the usual data processing algorithms to the data as a whole.
- This is the situation known as *big data*.
- It may be that:
  - while in principle, it is possible to eventually process all the data points,
  - this would take longer time than we have,
  - e.g., when we need to make a decision right away.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 2. Granular Computing (cont-d)

- It may also be that we want to use our intuition to better process the data.
- And to use our intuition, we need to present the data in presentable form.
- There may be other cases when we have too much data.
- To deal with such cases, a natural idea is compress the original data into a smaller set.
- The overall amount of available data can be estimated by multiplying:
  - the overall number of data points
  - by the average amount of bits in each data point.
- In general, each data point does not carry too much information.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 3. Granular Computing (cont-d)

- So the main way to decrease the overall amount of information is to decrease the number of data points.
- Of course, we could simply take a sample from the original data set.
- However, this would deprive us of all the information provided by the un-used data points.
- A much better idea is to each each new "data point" correspond to several original ones.
- This "combined" data point is known as a *granule*.
- The resulting technique is known as *granular computing*.
- The general idea of granular computing can be traced to Lotfi Zadeh.


Granular Computing: ...
How to Combine Data ...
Which Specificity ...
Shift-Invariance: ...
Additional Natural ...
Scale-Invariance: ...
Full Scale-Invariance Is ...
Let Us Combine Shift- ...
Result


Home Page
Title Page
Page 4 of 44
Go Back
Full Screen
Close
Quit

## 4. Granular Computing (cont-d)

- There are many possible types of granules.
- The most widely used type of granular computing is *clustering*, when we:
  - divide all possible objects
  - into several reasonable groups (clusters).
- Another widely used type of granularity is *histograms*, when:
  - we visualize the data
  - by describing the number of data points in different intervals.
- Also, instead of several numerical values,
  - we can consider *intervals* (or, more generally, *sets*)
  - that contain all – or at least most – of the data points.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 5. Granular Computing (cont-d)

- This is done in histograms, this is done in clustering, and this is done in many other practical situations.
- We can consider *fuzzy sets*, that describe:
  - not only which values are possible,
  - but also to what degree different data points are possible.
- We can consider *type-2 fuzzy* or probabilistic granules.
- We can consider *rough sets*, etc.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 6. How to Combine Data Points into a Granule: the Principle of Justifiable Granularity

- Suppose that we have selected a group of data points that we want to compress into a granule.
- Then, the question is which granule to select based on these data points.
- If we include all data points into a granule, the granule often becomes too wide to be useful.
- On the other hand:
  - if the granule is too narrow,
  - it includes only a few of the corresponding points,
  - and is, thus, also rather useless.
- We thus need to achieve a trade-off between coverage and specificity.

# 7. Principle of Justifiable Granularity (cont-d)

- In some cases – e.g., in histogram analysis – there are known methods for selecting the optimal granule.
- However, in the general case, we have to use semi-empirical rules.
- Most of these rules are in good accordance with the decision theory:
  - decisions of a rational decision maker can be described as
  - optimizing the expected value of a *utility function* $u(s)$ – that describes the corresponding preference.


Granular Computing: ...
How to Combine Data ...
Which Specificity ...
Shift-Invariance: ...
Additional Natural ...
Scale-Invariance: ...
Full Scale-Invariance Is ...
Let Us Combine Shift- ...
Result

## 8. Principle of Justifiable Granularity (cont-d)

- In other words, if
  - after making a selection $a$, we get situations $s_1, \ldots, s_n$ with probabilities $p_1(a), \ldots, p_n(a)$,
  - then we should select $a$ for which the expected value $p_1(a) \cdot u(s_1) + \ldots + p_n(a) \cdot u(s_n)$ is the largest.
- One can easily check that:
  - if we replace the utility function $u(a)$ by a re-scaled one $u_1(s) = k \cdot u(s) + \ell$,
  - then we get the same order between selections.
- Vice versa:
  - if two utility functions $u(s)$ and $u_1(s)$ always lead to the same decisions,
  - then $u_1(s) = k \cdot u(s) + \ell$ for some $k > 0$ and $\ell$.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 9. Principle of Justifiable Granularity (cont-d)

- In this sense, utility is similar to physical quantities like time or temperature.
- Their numerical values can change if we select:
  - a different measuring unit and/or
  - a different starting point.
- In our case, when we replace several data points, we lose information.
- So in this case, the utility is negative.
- In our problem, we have two situations.
- For some points, we replace these points with a granule.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 10. Principle of Justifiable Granularity (cont-d)

- The probability $P$ of this replacement can be naturally computed as:
  - the proportion of data points
  - that fit into the corresponding granule.
- This proportion depends on the size $\varepsilon$ of the granule:

$$P = P(\varepsilon).$$

- The larger the size, the higher the proportion.
- The utility of this replacement also depends on the size $\varepsilon$ of the granule: $u = u(\varepsilon)$.
- The larger the size, the smaller the utility.
- Other points do not fit into the granule and are dismissed (or processed in a more complex way).


Granular Computing: ...
How to Combine Data ...
Which Specificity ...
Shift-Invariance: ...
Additional Natural ...
Scale-Invariance: ...
Full Scale-Invariance Is ...
Let Us Combine Shift- ...
Result

## 11. Principle of Justifiable Granularity (cont-d)

- The probability of this dismissal (or alternative processing) is, clearly, the remaining probability $1-P(\varepsilon)$.
- Let us denote the utility of this dismissal (or alternative processing) by $u_0$.
- According to decision making, we thus need to select the size $\varepsilon$ that maximizes the expected utility

$$P(\varepsilon)\cdot u(\varepsilon)+(1-P(\varepsilon))\cdot u_0.$$

- This expression can be equivalently rewritten as $P(\varepsilon)\cdot S(\varepsilon)+u_0$, where we denoted

$$S(\varepsilon)\stackrel{\text{def}}{=}u(\varepsilon)-u_0.$$


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 12. Principle of Justifiable Granularity (cont-d)

- This objective function can be further simplified if we take into account that:
  - subtracting the same value $u_0$ from all the values does not change the order
  - and thus, does not change the optimal selection.
- Thus, we need to select the value $\varepsilon$ for which the product $P(\varepsilon) \cdot S(\varepsilon)$ takes the largest possible value.
- This ideas has indeed been used to select an appropriate granule.
- The probability $P(\varepsilon)$ that is known as the *coverage*.
- The expression $S(\varepsilon)$ – that describes how specific is the granule – is known as *measure of specificity*.
- The idea of maximizing $P(\varepsilon) \cdot S(\varepsilon)$ is known as the *Principle of Justified Granularity* (Pedrycz et al.).

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 13. Which Specificity Functions Work Best?

- The specific selection of the granule size depends on the selection of the measure of specificity.
- Empirical analysis has shown that:
  - out of several measures of specificity that have been tested,
  - the most adequate results are obtained when we use the following two measures of specificity
- The exponential measure of specificity

$$S(\varepsilon) = \text{const} \cdot \exp(-c \cdot \varepsilon).$$

for some constant $c$.

- The power law measure of specificity

$$S(\varepsilon) = \text{const} \cdot (1 - c \cdot \varepsilon)^{\xi}.$$

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 14. What We Do in This Talk

- In this talk, we provide a theoretical explanation for this empirical choice.
- Namely, we show that this choice follows from natural symmetries.
- By definition, $S(\varepsilon)$ differs from the utility function $u(\varepsilon)$ only by an additive constant $u_0$.
- Since, as we have mentioned:
  - the utility function is defined modulo an additive constant $\ell$ anyway,
  - we can as well talk about selecting an appropriate utility function.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 15. Shift-Invariance: Formulation of the First Natural Symmetry

- The data points come from measurements (or from expert estimates).
- Measurements are never absolutely accurate; thus:
  - the measured values are, in general, somewhat different
  - from the actual (unknown) values of the corresponding quantity.
- We usually take the measurement uncertainty into account.
- However, often, there is an additional source of error that we did not think about.
- What if there is indeed such additional source of error, of size $\varepsilon_0$?


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 16. Shift-Invariance (cont-d)

- In this case:
  - when a granule of size $\varepsilon$ includes all appropriate measurement results,
  - for this granule to include the actual values, we must increase the granule size to $\varepsilon + \varepsilon_0$.
- It is reasonable to require that:
  - the relative quality of different granules not change
  - if we take this unknown uncertainty into account.
- In other words, it is reasonable to require that:
  - selections based on the shifted utility

$$u_1(\varepsilon) \stackrel{\text{def}}{=} u(\varepsilon + \varepsilon_0)$$

  - lead to the same choice as selections based on the original utility.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 17. Analysis of the Problem

- We have already mentioned that:
  - when two different utility functions lead to the same selections,
  - we must have $u_1(\varepsilon) = k \cdot u(\varepsilon) + \ell$ for some $k > 0$ and $\ell$.
- The coefficients $k$ and $\ell$ may depend on the shift $\varepsilon_0$.
- Thus, for every $\varepsilon$, there exists the values $k(\varepsilon_0)$ and $\ell(\varepsilon_0)$ for which, for all $\varepsilon > 0$ and $\varepsilon_0$, we have

$$u(\varepsilon + \varepsilon_0) = k(\varepsilon_0) \cdot u(\varepsilon) + \ell(\varepsilon_0).$$


Granular Computing: ...
How to Combine Data ...
Which Specificity ...
Shift-Invariance: ...
Additional Natural ...
Scale-Invariance: ...
Full Scale-Invariance Is ...
Let Us Combine Shift- ...
Result

## 18. Additional Natural Requirement: Smoothness

- It is also reasonable to require that:
  - when we change the granule size a little bit,
  - the utility will also change a little bit.
- In mathematical terms, the utility function $u(\varepsilon)$ should be smooth, i.e., differentiable.
- **Proposition.**
  - *Let $u(\varepsilon)$ be a differentiable function that satisfies the equation*

$$u(\varepsilon+\varepsilon_0)=k(\varepsilon_0)\cdot u(\varepsilon)+\ell(\varepsilon_0);$$

  - *then either $u(\varepsilon)=\text{const}\cdot\exp(-c\cdot\varepsilon)$ for some $c$ or $u(\varepsilon)=\text{const}\cdot(1-g\cdot\varepsilon)$ for some $g$.*
- Thus, we have justified the exponential measure of specificity and the power law for $\xi=1$.


Granular Computing: ...
How to Combine Data ...
Which Specificity ...
Shift-Invariance: ...
Additional Natural ...
Scale-Invariance: ...
Full Scale-Invariance Is ...
Let Us Combine Shift- ...
Result

## 19. Scale-Invariance: Formulation of the Second Natural Symmetry

- The size of the granule is measured in the same units as the values forming this granule.
- For example:
  - if the granule contains values of length,
  - then the size – i.e., the accuracy of representing a value – is also measured by units of length.
- As we have mentioned, the numerical values of a physical quantity depend on the choice of a measuring unit:
  - if we replace the original unit by a new unit which is $\lambda$ times smaller,
  - then all the numerical values are multiplied by $\lambda$.
- *Example:* if we replace meters by centimeters, all numerical values are multiplied by 100: 2 m $\to$ 200 cm.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 20. Scale-Invariance (cont-d)

- When we change the units, the values $\varepsilon$ are replaced by new values $\lambda \cdot \varepsilon$.
- It therefore seems reasonable to require that:
  - the relative quality of different measures of specificity
  - not change if we simply change the measuring unit.
- In other words, the utility function $u_1(\varepsilon) \stackrel{\text{def}}{=} u(\lambda \cdot \varepsilon)$ should be equivalent to $u(\varepsilon)$.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 21. Full Scale-Invariance Is Rarely Possible

- We conclude that for some functions $k(\lambda)$ and $\ell(\lambda)$ depending on $\lambda$, we have:

$$u(\lambda \cdot \varepsilon) = k(\lambda) \cdot u(\varepsilon) + \ell(\lambda).$$

- We already know that, due to shift-invariance, the utility function is either exponential or linear.
- While linear function satisfies the equation, the exponential function does not.
- Thus:
  - if we require both shift- and scale-invariance, we end up with only linear measures of specificity, and
  - we know that empirically, sometimes non-linear measures of specificity work better.
- So, we cannot require both shift- and scale-invariance.
- What can we do?


Granular Computing: ...
How to Combine Data ...
Which Specificity ...
Shift-Invariance: ...
Additional Natural ...
Scale-Invariance: ...
Full Scale-Invariance Is ...
Let Us Combine Shift- ...
Result

## 22. Let Us Combine Shift- and Scale-Invariance

- Combining several different invariances makes perfect sense.
- For example, in the Ohm's Law $V = I \cdot R$ that relates voltage, current, and resistance:
  - if we simply change the unit for current,
  - the law stops working.
- For the formula to remain valid:
  - for each change of the unit for measuring current,
  - we also need to appropriately change the unit for measuring voltage.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 23. Combining Shift- & Scale-Invariance (cont-d)

- In general, such a situation is typical in physics:
  - when a formula is not invariant with respect to one class of transformation,
  - it usually means that for each transformation from this class, there is an appropriate transformation from some related class
  - so that if we apply both transformations at the same time, we get the same formula as before.
- Let us apply this idea to our case.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 24. Resulting Formulation

- We cannot require that the utility function be invariant with respect to arbitrary re-scaling.
- So, let us combine it with shift-invariance the same way as it is done in physics.
- Namely, for every $\lambda$, there exists a value $\varepsilon(\lambda)$ for which:
  - the re-scaled utility function $u(\lambda \cdot \varepsilon)$ is equivalent
  - to the correspondingly shifted one $u(\varepsilon + \varepsilon_0(\lambda))$.
- As we have mentioned, equivalence means that

$$u(\lambda \cdot \varepsilon) = k(\lambda) \cdot u(\varepsilon + \varepsilon_0(\lambda)) + \ell(\lambda).$$


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 25. Result

- **Proposition.**
  - *Let* $u(\varepsilon)$, $\varepsilon_0(\lambda)$, $k(\lambda)$ *and* $\ell(\lambda)$ *be differentiable functions for which, for all* $\lambda$, $\varepsilon$:

$$u(\lambda\cdot\varepsilon)=k(\lambda)\cdot u(\varepsilon+\varepsilon_0(\lambda))+\ell(\lambda).$$

  - *Then, either* $u(\varepsilon)=C\cdot(1-c\cdot\varepsilon)^{\xi}+\mathrm{const}$ *for some* $C$, $c$, *and* $\xi$, *or* $u(\varepsilon)=D\cdot\ln(1-g\cdot\varepsilon)+\mathrm{const}$.
- This result explains the efficiency of the power law measure of specificity.
- One can check that the logarithmic expression is the limit of the power law when $\xi\to 0$.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 26. Conclusions

- In many practical problems, it is beneficial to combine the data into granules.
- For example, when we plot the empirical data,
  - it is often helpful to generate a histogram
  - that shows the frequency with which we encounter values from different intervals.
- This enables us:
  - to see the shape of the corresponding probability distribution
  - which otherwise would be hidden behind the random noise.
- To maximize the effect of such granulation, it is important to select the appropriate size of the granule.

Granular Computing: ...
How to Combine Data ...
Which Specificity ...
Shift-Invariance: ...
Additional Natural ...
Scale-Invariance: ...
Full Scale-Invariance Is ...
Let Us Combine Shift- ...
Result

# 27. Conclusions (cont-d)

- If the granules are too small, the desired dependence will still be hidden behind the noise.
- If the granules are too big, we may lose important details.
- One of the most successful ways to find the proper level of granularity is to use the *Principle of Justified Granularity*.
- According to this principle, we select the granule size for which the product of a measure of coverage and a measure of specificity is the largest possible.
- Theoretically, there are many possible measures of specificity.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 28. Conclusions (cont-d)

- It turns out that empirically, the following two measures lead to the most beneficial granulation:
  - the exponential measure of specificity and
  - the power law measure of specificity.
- In this talk, we show that these empirically successful measures of specificity can be theoretically explained:
  - if we require that the choice of the optimal granularity
  - not depend on the selecting of a measuring unit and
  - not depend on the starting point for measuring the corresponding quantity.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 29. Possible Future Work

- In this paper, we considered the selection of a *single* measure of specificity.
- It seems that we can get even better results if:
  - instead of such a universal measure of specificity,
  - we consider a *family* of specificity measures,
  - so that we will be able, in each practical situation, to select the most appropriate measure.
- It is therefore desirable to extend our result:
  - from selecting the optimal *measure of specificity*
  - to a more complex problem of selecting the optimal *family of measures of specificity*.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 30. Possible Future Work (cont-d)

- This optimal selection of the corresponding family should also be invariant with respect to:
  - changing the measuring unit and
  - changing the starting point for measurement.
- Maybe – just like in our case, these symmetry will be sufficient to select the optimal family?
- Or maybe other ideas are needed to make this selection?
- It is also desirable:
  - to empirically compare different multi-parametric families of measures of specificity,
  - to see which family works the best.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 31. Acknowledgments

This work was supported in part by the US National Science Foundation grant HRD-1242122.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 32. Proof of the First Proposition

- We assumed that the utility function $u(\varepsilon)$ is differentiable.
- Let us prove that in this case, the auxiliary functions $k(\varepsilon_0)$ and $\ell(\varepsilon_0)$ are also differentiable.
- Indeed, if we pick two different values $\varepsilon = \varepsilon_1$ and $\varepsilon = \varepsilon_2 \neq \varepsilon_1$.
- Then, the above formula takes the following form:

$$u(\varepsilon_1 + \varepsilon_0) = k(\varepsilon_0) \cdot u(\varepsilon_1) + \ell(\varepsilon_0);$$

$$u(\varepsilon_2 + \varepsilon_0) = k(\varepsilon_0) \cdot u(\varepsilon_2) + \ell(\varepsilon_0).$$

- Thus, we have a system of two linear equations for the two unknowns $k(\varepsilon_0)$ and $\ell(\varepsilon_0)$.
- By the Cramer's rule, the solution is a rational – hence differentiable – f-n of the coefficient and free terms.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 33. Proof of the First Proposition (cont-d)

- The solution is a differentiable function of the coefficient and free terms.
- Since the function $u(\varepsilon)$ is differentiable, all these coefficients and free terms are also differentiable.
- Thus, we can conclude that the functions $k(\varepsilon_0)$ and $\ell(\varepsilon_0)$ are differentiable.
- We know that all three functions $u(\varepsilon)$, $k(\varepsilon_0)$, and $\ell(\varepsilon_0)$ and differentiable.
- Let us differentiate both sides of the original equation with respect to $\varepsilon_0$.
- As a result, we get the following expression, where, as usual, $f'(x)$ denotes the derivative of the function $f(x)$:

$$u'(\varepsilon+\varepsilon_0)=k'(\varepsilon_0)\cdot u(\varepsilon)+\ell'(\varepsilon_0).$$


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 34. Proof of the First Proposition (cont-d)

- Substituting $\varepsilon_0 = 0$ into this formula, we get

  $u'(\varepsilon) = k_0 \cdot u(\varepsilon) + \ell_0,$ where $k_0 \stackrel{\text{def}}{=} k'(0)$ and $\ell_0 \stackrel{\text{def}}{=} \ell'(0).$

- Since $u' = \dfrac{du}{d\varepsilon}$, we can rewrite the resulting differential equation as $\dfrac{du}{d\varepsilon} = k_0 \cdot u + \ell_0.$
- Let us separate the variables.
- We can do it if we:
  - multiply both sides of this equation by $d\varepsilon$ and
  - divide both sides of this equation by $k_0 \cdot u + \ell_0$.
- As a result, we get the following formula:

$$\frac{du}{k_0 \cdot u + \ell_0} = d\varepsilon.$$


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 35. Proof of the First Proposition (cont-d)

- Here, we have two options:
  - the first option is that $k_0 = 0$;
  - the second option is that $k_0 \neq 0$.
- Let us consider these two options one by one.
- When $k_0 = 0$, integrating both sides, we get:

$$\frac{u}{\ell_0} = \varepsilon + C, \text{ where } C \text{ is an integration constant.}$$

- So, $\dfrac{u}{\ell_0} = C \cdot (1 - g \cdot \varepsilon)$, where we denoted $g \stackrel{\text{def}}{=} -\dfrac{1}{C}$.
- Multiplying both sides of this formula by $\ell_0$, we get

$$u(\varepsilon) = C_1 \cdot (1 - g \cdot \varepsilon), \text{ where } C_1 \stackrel{\text{def}}{=} \ell_0 \cdot C.$$

- Thus, in the case of $k_0 = 0$, we get a linear measure of specificity.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 36. Proof of the First Proposition (cont-d)

- Let us now consider the case when $k_0 \neq 0$.
- In this case, for a new variable $v \stackrel{\text{def}}{=} u + \dfrac{\ell_0}{k_0}$, we have $dv = du$ and $k_0 \cdot u + \ell_0 = k_0 \cdot v$.
- Thus, the above formula takes a simplified form

$$\frac{dv}{k_0 \cdot v} = d\varepsilon.$$

- Integrating both sides of this formula, we get

$$\frac{1}{k_0} \cdot \ln(v) = \varepsilon + C.$$

- Here, $C$ is an integration constant.
- Multiplying both sides of this formula by $k_0$, we get

$$\ln(v) = k_0 \cdot \varepsilon + C_1, \text{ where } C_1 \stackrel{\text{def}}{=} k_0 \cdot C.$$


Granular Computing: ...
How to Combine Data ...
Which Specificity ...
Shift-Invariance: ...
Additional Natural ...
Scale-Invariance: ...
Full Scale-Invariance Is ...
Let Us Combine Shift- ...
Result

## 37. Proof of the First Proposition (cont-d)

- By applying exp to both sides, we get

$$v(\varepsilon) = \exp(\ln(v)) = \exp(k_0 \cdot \varepsilon + C_1) = C_2 \cdot \exp(k_0 \cdot \varepsilon),$$

$$\text{where } C_2 \stackrel{\text{def}}{=} \exp(C_1).$$

- Thus for $u = v - \dfrac{\ell_0}{k_0}$, we get

$$u(\varepsilon) = C_2 \cdot \exp(k_0 \cdot \varepsilon) + \text{const}.$$

- So, in the case when $k_0 \neq 0$, we get the exponential measure of specificity.
- The proposition is proven.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 38. Proof of the Second Proposition

- Similarly to the proof of Proposition 1, let us reduce:
  - the difficult-to-solve functional equation
  - to an easier-to-solve differential equation.
- For this purpose, let us differentiate both side of our equation by $\lambda$.
- As a result, we get the following formula:

$$\varepsilon \cdot u'(\lambda \cdot \varepsilon) =$$

$$k'(\lambda) \cdot u(\varepsilon + \varepsilon_0(\lambda)) + k(\lambda) \cdot u'(\varepsilon + \varepsilon_0(\lambda)) \cdot \varepsilon_0'(\lambda) + \ell'(\lambda).$$

- Let's substitute $\lambda = 1$ into this formula.
- Let's take into account that for $\lambda = 1$, there is no change and thus, $\varepsilon_0(1) = 0$, $k(1) = 1$, and $\ell(1) = 0$.
- So, we get $\varepsilon \cdot u'(\varepsilon) = k_0 \cdot u(\varepsilon) + m_0 \cdot u'(\varepsilon) + \ell_0$, where:

$$k_0 \stackrel{\text{def}}{=} k'(1), \quad m_0 \stackrel{\text{def}}{=} \varepsilon_0'(1), \text{ and } \ell_0 \stackrel{\text{def}}{=} \ell'(1).$$


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 39. Proof of the Second Proposition (cont-d)

- This formula can be rewritten as

$$\varepsilon \cdot \frac{du}{d\varepsilon} = k_0 \cdot u + m_0 \cdot \frac{du}{d\varepsilon} + \ell_0.$$

- Let us now solve this differential equation.
- Moving the terms proportional to $u'$ to the left-hand side, we conclude that

$$(\varepsilon - m_0) \cdot \frac{du}{d\varepsilon} = k_0 \cdot u + \ell_0.$$

- Now, we can separate the variables.
- We can do it if we:
  - multiply both sides by $d\varepsilon$,
  - divide both sides by $\varepsilon - m_0$, and
  - divide both side by $k_0 \cdot u + \ell_0$.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 40. Proof of the Second Proposition (cont-d)

- As a result, we get the following equation:

$$\frac{du}{k_0 \cdot u + \ell} = \frac{d\varepsilon}{\varepsilon - m_0}.$$

- Similarly to the proof of Proposition 1, let us consider two possible cases:
  - case when $k_0 = 0$, and
  - case when $k_0 \neq 0$.
- Let's first consider the case $k_0 = 0$.
- Integrating both sides of the above formula and taking into account that $d(\varepsilon - m_0) = d\varepsilon$, we conclude that

$$\frac{u}{\ell_0} = \ln(m_0 - \varepsilon) + C.$$

- Here, $C$ is an integration constant.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

## 41. Proof of the Second Proposition (cont-d)

- Thus, we have $u(\varepsilon) = \ell_0 \cdot \ln(m_0 - \varepsilon) + C_1$, where
$$C_1 \stackrel{\text{def}}{=} \ell_0 \cdot C.$$
- Here, $m_0 - \varepsilon = m_0 \cdot (1 - g \cdot \varepsilon)$, where $g \stackrel{\text{def}}{=} \dfrac{1}{m_0}$: thus:
$$\ln(m_0 - \varepsilon) = \ln(m_0 \cdot (1 - g \cdot \varepsilon)) = \ln(m_0) + \ln(1 - g \cdot \varepsilon).$$
- Hence, the above formula takes the form
$$u(\varepsilon) = \ell_0 \cdot \ln(1 - g \cdot \varepsilon) + C_2, \text{ where } C_2 \stackrel{\text{def}}{=} C_1 + \ell_0 \cdot \ln(m_0).$$
- So, in the case of $k_0 = 0$, we get the logarithmic measure of specificity.
- Let us now consider the remaining case $k_0 \neq 0$.
- In this case, similarly to the proof of Proposition 1, we can introduce a new variable $v = u + \dfrac{\ell_0}{k_0}$.

Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result

# 42. Proof of the Second Proposition (cont-d)

- Then, the above equation takes the form
$$\frac{dv}{k_0 \cdot v} = \frac{d\varepsilon}{\varepsilon - m_0}.$$
- Integrating both parts of this equation, we get
$$\frac{1}{k_0} \cdot \ln(v) = \ln(m_0 - \varepsilon) + C.$$
- Here $C$ is an integration constant, so:
$$\frac{u}{k_0} = (1 - c \cdot \varepsilon) + C', \text{ where } c \stackrel{\text{def}}{=} \frac{1}{m_0} \text{ and } C' = C + \ln(m_0).$$
- Multiplying both sides by $k_0$, we conclude that
$$\ln(v) = k_0 \cdot \ln(1 - c \cdot \varepsilon) + C_1, \text{ where } C_1 \stackrel{\text{def}}{=} k_0 \cdot C'.$$
- Applying exp to both sides, and taking into account that $\exp(k_0 \cdot \ln(x)) = (\exp(\ln(x)))^{k_0} = x^{k_0}$, we get
$$v = C_2 \cdot (1 - c \cdot \varepsilon)^{k_0}, \text{ where } C_2 \stackrel{\text{def}}{=} \exp(C_1).$$

## 43. Proof of the Second Proposition (cont-d)

- Thus, $u(\varepsilon) = C_2 \cdot (1 - c \cdot \varepsilon)^{k_0} + \text{const}$.
- So, in the case of $k_0 \neq 0$, we get the power law measure of specificity.
- The proposition is proven.


Granular Computing: . . .
How to Combine Data . . .
Which Specificity . . .
Shift-Invariance: . . .
Additional Natural . . .
Scale-Invariance: . . .
Full Scale-Invariance Is . . .
Let Us Combine Shift- . . .
Result