Why Kappa Regression?

Julio C. Urenda[1], Orsolya Csiszár[2,3], Gábor Csiszár[4],
József Dombi[5], György Eigner[3],
Olga Kosheleva[1], and Vladik Kreinovich[1],

[1]University of Texas at El Paso, USA
[2]University of Applied Sciences Esslingen, Germany
[3]Óbuda University, Budapest, Hungary
[4]University of Stuttgart, Germany
[5]University of Szeged, Hungary

E-mails: vladik@utep.edu, orsolya.csiszar@nik.uni-obuda.hu,
gabor.csiszar@mp.imw.uni-stuttgart.de, dombi@inf.u-szeged.hu,
eigner.gyorgy@nik.uni-obuda.hu, olgak@utep.edu, vladik@utep.edu

1. Empirical facts

- In many practical situations, probability distributions with the following cdf work well:

$$F(x) = \text{Prob}(X \le x) = \frac{1}{1 + C \cdot \left(\dfrac{b - x}{x - a}\right)^{\lambda}}.$$

- Such distributions are known as *kappa-regression distributions*.
- Fuzzy processing with similar membership functions also works well: $\mu(x) = \dfrac{1}{1 + C \cdot \left(\dfrac{b - x}{x - a}\right)^{\lambda}}$.
- Often, other families of probability distributions – e.g., Gaussian – work better.
- Still, kappas work very well. How can we explain this?

2. A known limit case

- In the limit, kappa regressions become *logistic* distribution

$$F(x) = \frac{1}{1 + C \cdot \exp(-k \cdot x)}.$$

- So, let us first try to understand why this limit case has been very successful.

3. Idea of invariance

- Let us recall how real-life phenomena are described and explained in the first place.
- Modern science – especially physics – has been very successful.
- We can predict many events.
- But what is the general basis for all these predictions?
  - we observe that the Sun goes up day after day, and
  - we conclude that in the similar situations, the Sun will go up again.
- We observe, at different locations, that if you drop a pen, it will fall with the acceleration of 9.81 m/sec^2.
- So we conclude that in similar situations, it will fall down with the same acceleration.

4. Idea of invariance (cont-d)

- We observe, in many cases, that mechanical bodies follow Newton's laws.
- So we conclude that in the similar situations, the same laws will be observed.
- In all these cases, we conclude that:
  - when we change a situation to a similar one,
  - e.g., by moving to a different location on Earth or to a different day, etc.,
  - the processes will remain similar.
- The idea that physical properties don't change if we perform some transformations is called *invariance*.

5. Idea of invariance (cont-d)

- Invariances – also called *symmetries* in physics – are indeed one of the fundamental ideas of modern physics.
- Many new theories – starting with the theory of quarks – are proposed:
  - not by writing down differential equations,
  - but by describing the corresponding invariances.

6. What are the simplest invariances?

- Some invariances – e.g., the ones used in quark theory – are rather complicated.
- Let us start with the simplest possible invariances.
- These invariances are related to the fact that:
  - when we write equations,
  - we operate with numerical values of the physical quantities.
- To describe physical quantities by numbers, we need to select a measuring unit and a starting point.
- For example, we can measure time starting:
  - with Year 0 – as in the commonly used calendar –
  - or with any other moment of time;
- After the French revolution, the new calendar started with the year of the revolution as the first year.

7. What are the simplest invariances (cont-d)

- We can also change a measuring unit – e.g., count days or months instead of years.
- In general:
  - if you replace the original measuring unit with a new unit which is c times smaller,
  - then all numerical values are multiplied by c:

$$x \to c \cdot x.$$

- E.g., if we replace meters with centimeters, all numerical values will be multiplied by 100.
- 2 m social distance will become $2 \cdot 100 = 200$ cm.
- By a *scaling*, we mean a transformation (function) $f(x) = c \cdot x$ for some $c > 0$.

8. What are the simplest invariances (cont-d)

- Similarly:
  - if we replace the original starting point with the one which is x_0 units earlier,
  - then all numerical values increase by x_0:

$$x \to x + x_0.$$

- By a *shift*, we mean a transformation $f(x) = x + x_0$ for some x_0.
- In many physical situations, there is no preferred starting point.
- So, we expect that the processes remain similar:
  - if we change the starting point,
  - i.e., if we replace all numerical values x with shifted values $x + x_0$.

9. What are the simplest invariances (cont-d)

- Similarly, in many physical situations, there is no preferred measuring unit.
- So, we expect that the processes remain similar:
  - if we replace the measuring unit,
  - i.e., if we replace all numerical values x with re-scaled values $c \cdot x$.

10. How can we apply these ideas to probability distributions?

- Of course:
  - if we change the units of one of the quantities,
  - then, to preserve the same equations, we need to accordingly change the units of related quantities.
- For example, let us start with the formula $d = v \cdot t$ that the distance is velocity times time.
- Let us change the unit for time from hours to seconds.
- Then, to preserve the formula, we need to corresponding change the units for velocity: e.g., from km/h to km/sec.
- In probability theory, there is a natural way to change probabilities: the Bayes formula.

11. Bayes formula

- If we have a new observation E, then the previous probability $P_0(H)$ of a hypothesis H changes to:

$$P(H\,|\,E)=\frac{P(E\,|\,H)\cdot P_0(H)}{P(E\,|\,H)\cdot P_0(H)+P(E\,|\,\neg H)\cdot P_0(\neg H)}=$$

$$\frac{P(E\,|\,H)\cdot P_0(H)}{P(E\,|\,H)\cdot P_0(H)+P(E\,|\,\neg H)\cdot (1-P_0(H))}=$$

$$\frac{P_0(H)}{P_0(H)+r\cdot(1-P_0(H))}.$$

- Here, we denoted $r\stackrel{\text{def}}{=}\dfrac{P(E\,|\,\neg H)}{P(E\,|\,H)}$.

12. Bayes formula (cont-d)

- So, a natural idea is to require that:
  - if we apply a reasonable transformation to x,
  - e.g., change the starting point or change the measuring unit,
  - then the probability distribution will change according to the Bayes formula.
- We say that cdfs $F(x)$ and $G(x)$ are *equivalent* if for some real number r, we have:

$$G(x) = \frac{F(x)}{F(x) + r \cdot (1 - F(x))}.$$

- This equivalence divides all possible cumulative distribution functions into equivalence classes.

13. Bayes formula (cont-d)

- It is reasonable to call an equivalence class f-*invariant* if this class does not change under a transformation f.
- This definition can be equivalently described in terms of the cdfs from the f-invariant equivalence class.
- We say that $F(x)$ is f-*invariant*, if $F(f(x))$ and $F(x)$ are equivalent, i.e., if for some $r > 0$, we have

$$F(f(x)) = \frac{F(x)}{F(x) + r \cdot (1 - F(x))}.$$

14. What probability distributions satisfy this invariance requirement?

- **Result.** For each cumulative distribution function $F(x)$, the following two conditions are equivalent:
  - $F(x)$ is invariant with respect to all shifts;
  - $F(x)$ is a logistic distribution.
- The Bayes formula becomes simpler if we consider the *odds* $O \stackrel{\text{def}}{=} \dfrac{P}{1-P}$:

$$O' = \frac{P'}{1-P'} = \frac{1}{r} \cdot \frac{P}{1-P} = s \cdot O, \text{ where we denoted } s \stackrel{\text{def}}{=} \frac{1}{r}.$$

- In these terms, shift-invariance means

$$O(x+x_0) = s(x_0) \cdot O(x) \text{ for some } s.$$

15. What probability distributions satisfy this invariance requirement (cont-d)

- Each cumulative distribution function $F(x)$ is monotonic and thus, measurable.
- Thus, the odds function is also measurable.
- It is known that all measurable solutions of the above functional equation have the form $O(x) = c \cdot \exp(k \cdot x)$.
- So, for $P = \dfrac{1}{1 + \dfrac{1}{O}}$, we get the logistic distribution.

16. What About the Fuzzy Case?

- The Bayes formula is not applicable to membership functions.
- So, we need a different explanation.
- Let us recall that one of the possible ways to get membership degrees is to poll experts.
- If m out of n experts think that the given statement is true, we assign to it the degree of confidence m/n.
- For example:
  - we can say that a person of a certain age is young to a degree 0.7
  - if 70% of the experts consider this person young.

17. Resulting transformations

- For statements that require true expertise – we ask top experts, of whose opinion we are most confident.
- Suppose that out of n top experts, m thought that the given statement is true.
- Then we assign, to this statement, the degree of confidence $\mu = m/n$.
- The problem is that in many practical situations, there are very few top experts: the number n is small.
- In this case, we have a very limited number of possible degrees.
- For example, when $n = 5$, we only have 6 possible values: 0, 1/5, 2/5, 3/5, 4/5, and 1.
- The only way to make a more meaningful distinction is to use a larger value of n, i.e., to ask more experts.

18. Resulting transformations (cont-d)

- However, in the presence of the top experts, other not-so-top experts may be:
  - either silent,
  - or simply follow the opinion of their peers.
- If we ask n' more experts and the new experts are silent, then the new degree of confidence is

$$\mu' = m/(n + n').$$

- In terms of the original degree of confidence $\mu = m/n$, we have $\mu' = c \cdot \mu$, where $c \stackrel{\text{def}}{=} n/(n + n')$.
- What if the new experts follow the majority of top experts – and if this majority confirms our statement.
- Then the new degree of confidence is

$$\mu' = (m + n')/(n + n').$$

19. Resulting transformations (cont-d)

- In terms of the original degree of confidence μ, we have $\mu' = c \cdot \mu + a$, where $a \stackrel{\text{def}}{=} n'/(n + n')$.
- In both cases, we have a linear transformation $\mu \to \mu'$.
- A similar linear transformation occurs if:
  - some of the new experts remain silent, and
  - some follow the majority of top experts.
- So, linear transformations make sense for fuzzy degrees as well.

20. Beyond linear transformations

- In principle, not all functions are linear.
- For example, the Bayes formula describes a non-linear transformation.
- Let us look for a general class of transformations w.r.t. which physical properties can be invariant.
- Clearly:
  - if the properties do not change when we apply a transformation $x' = f(x)$,
  - and do not change if we then apply the transformation $x'' = g(x')$,
  - then going from x to $x'' = g(x') = g(f(x))$ also does not change the properties.
- Thus, the class of possible transformations must be closed under composition.

21. Beyond linear transformations (cont-d)

- Similarly:
  - if the physical properties do not change when we go from x to $y = f(x)$,
  - then the transition back, from y to $x = f^{-1}(y)$, also preserves all physical properties.
- So, the class of possible transformation must contain the inverse transformation.
- In mathematical terms, this means that the class of all possible transformations much be a *group*.
- Also, we want this to be constructive, we want to be able to simulate such transformations on a computer.

22. Beyond linear transformations (cont-d)

- At any given moment of time, a computer can only store and use finitely many parameters; thus:
  - elements of the desired transformation group
  - must be uniquely determined by the values of finitely many parameters.
- In mathematical terms, this means that the corresponding group must be *finite-dimensional*.
- It is known that under reasonable conditions:
  - any finite-dimensional transformation group that contains all linear transformation
  - contains only fractional-linear transformations

$$f(x) = \frac{A + B \cdot x}{C + D \cdot x}.$$

- So, we will call them *r-transformations* (r for "reasonable").

23. Which reasonable transformations preserve the interval $[0,1]$?

- Possible degrees of confidence form the interval $[0,1]$.
- It is therefore reasonable to look for transformations that preserve this interval, i.e., map $[0,1]\to[0,1]$.
- Such transformations have the form

$$f(x)=\frac{x}{x+r\cdot(1-x)} \text{ for some real number } r.$$

- So, we say that the membership functions $\mu(x)$ and $\nu(x)$ are *equivalent* if for some real number r, we have:

$$\nu(x)=\frac{\mu(x)}{\mu(x)+r\cdot(1-\mu(x))}.$$

24. Which reasonable transformations preserve the interval $[0,1]$ (cont-d)

- We say that a membership function $\mu(x)$ is f*-invariant* if $\mu(f(x))$ and $\mu(x)$ are equivalent.
- For each membership function $\mu(x)$, the following two conditions are equivalent to each other:
  - $\mu(x)$ is invariant with respect to all shifts;
  - $\mu(x)$ is described by the formula

$$\mu(x) = \frac{1}{1 + C \cdot \exp(-k \cdot x)}.$$

25. Another Special Case

- So far, we considered invariance w.r.t. shifts.
- What if we require that the cdf be invariant with respect to changing the measuring unit $x \to c \cdot x$.
- For each cumulative distribution function $F(x)$, the following two conditions are equivalent to each other:
  - $F(x)$ is invariant with respect to all scalings;
  - $F(x)$ is described by the formula

$$F(x) = \frac{1}{1 + C \cdot x^{-k}}.$$

26. General Case

- The general kappa-regression distribution is concentrated, with probability 1, on the interval (a, b).
- This means that in this case, we cannot apply shift-invariance – since there is a natural starting value a.
- We cannot apply scale-invariance – since there is a natural measuring unit, e.g., the difference $b - a$.
- If we want to use invariances, we need to use some more general transformations.
- We have shown that reasonable requirements lead to fractional-linear transformations.
- So, we get the following result.

27. General Case

- **Theorem.** Let $a < b$. For each cdf $F(x)$, the following two conditions are equivalent to each other:
  - $F(x)$ is invariant with respect to all r-transformations that preserve the interval $[a, b]$;
  - $F(x)$ is a kappa-regression distribution.
- A similar result holds for membership functions.
- So, we have explained the efficiency of kappa-regression distributions and membership functions.
- They are the only ones which satisfy the reasonable invariance conditions.

28. Acknowledgment

This work was supported in part by:

- the grant TUDFO/47138-1/2019-ITM from the Ministry of Technology and Innovation, Hungary
- 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), and
- by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.