

# Our Reasoning is Clearly Fuzzy, So Why Is Crisp Logic So Often Adequate?

Hung T. Nguyen<sup>1,2</sup>, Berlin Wu<sup>3</sup>, and Vladik Kreinovich<sup>4\*</sup>

<sup>1</sup> *Department of Mathematical Sciences, New Mexico State University, Las Cruces, USA*

<sup>2</sup> *Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand*

<sup>3</sup> *Department of Mathematical Sciences, National Chengchi University, Taipei, Taiwan*

<sup>4</sup> *Department of Computer Science, University of Texas at El Paso, TX, USA*

## ABSTRACT

Our reasoning is clearly fuzzy, so why is crisp logic so often adequate? We explain this phenomenon by showing that in the presence of noise, an arbitrary continuous (e.g., fuzzy) system can be well described by its discrete analog. However, as the description gets more accurate, the continuous description becomes necessary.

*Keywords:* Fuzzy logic; Crisp logic; Tsirelson theorem; Quantum physics; Schroedinger paradox; Superposition principle

## 1. Formulation of the Problem

**Fuzzy logic is needed.** Our reasoning is clearly fuzzy. Whenever we use terms like “small,” “young,” etc., there is no crisp boundary: some people are clearly young, some are clearly not young, but there is always an area in between in which a natural answer should be “young to a certain degree.”

Fuzzy logic has been specifically designed to capture this “fuzziness;” see, e.g., [2, 3, 7].

**While fuzzy logic is successful, crisp methods are, surprisingly, successful as well.** Fuzzy logic indeed has many useful applications. However, the puzzling fact is that in many other practical applications, crisp (non-fuzzy) methods work really well.

**What we do in this paper.** In this paper, we provide a possible explanation for this phenomenon.

---

\* Corresponding author: vladik@utep.edu

*Comment.* Some of our results first appeared in [4].

## 2. Basis for Our Explanation: Tsirelson's Theorem

**Tsirelson's observation.** B. S. Tsirelson noticed [6] that in many cases, when we reconstruct the signal from the noisy data, and we assume that the resulting signal belongs to a certain class, the reconstructed signal is often an *extreme* point from this class. For example:

- when we assume that the reconstructed signal is monotonic, the reconstructed function is often (piece-wise) constant;
- if we additionally assume that the signal is smooth (one time differentiable, from the class  $C^1$ ), the result is usually one time differentiable but rarely twice differentiable, etc.

**Geometric explanation for Tsirelson's observation: general idea.** Tsirelson provides an elegant *geometric* explanation to this fact: namely, when we reconstruct a signal from a mixture of a signal and a Gaussian noise, then the *maximum likelihood* estimation (a traditional statistical technique; see, e.g., [5]) means that we look for a signal that belongs to the priori class, and that is the closest (in the  $L^2$ -metric) to the observed "signal + noise."

In particular, if the signal is determined by finitely many (say,  $d$ ) parameters, we must look for a signal  $\vec{s} = (s_1, \dots, s_d)$  from the a priori set  $A \subseteq \mathbb{R}^d$  that is the closest (in the usual Euclidean sense) to the observed values

$$\vec{o} = (o_1, \dots, o_d) = (s_1 + n_1, \dots, s_d + n_d),$$

where  $n_i$  denotes the (unknown) values of the noise.

Since the noise is Gaussian, we can usually apply the *Central Limit Theorem* [5] and conclude that the average value of  $(n_i)^2$  is close to  $\sigma^2$ , where  $\sigma$  is the standard deviation of the noise. In other words, we can conclude that

$$(n_1)^2 + \dots + (n_d)^2 \approx d \cdot \sigma^2.$$

In geometric terms, this means that the distance

$$\sqrt{\sum_{i=1}^d (o_i - s_i)^2} = \sqrt{\sum_{i=1}^d n_i^2}$$

between  $\vec{s}$  and  $\vec{o}$  is  $\approx \sigma \cdot \sqrt{d}$ . Let us denote this distance  $\sigma \cdot \sqrt{d}$  by  $\varepsilon$ .

**Geometric explanation: 2-D case.** Let us (for simplicity) consider the case when  $d = 2$ , and when  $A$  is a convex polygon. Then, we can divide all points  $p$  from the exterior of  $A$  that are  $\varepsilon$ -close to  $A$  into several zones depending on what part of  $A$  is the closest to  $p$ : one of the *sides*, or one of the *edges*.

Geometrically, the set of all points for which the closest point  $a \in A$  belongs to the *side*  $e$  is bounded by the straight lines orthogonal (perpendicular) to  $e$ . The total length of this set is therefore equal to the length of this particular side; hence, the total length of all the points that are the closest to all the sides is equal to the *perimeter* of the polygon. This total length thus does not depend on  $\varepsilon$  at all.

On the other hand, the set of all the points at the distance  $\varepsilon$  from  $A$  grows with the increase in  $\varepsilon$ ; its length grows approximately as the length of a circle, i.e., as  $\text{const} \cdot \varepsilon$ .

When  $\varepsilon$  increases, the (constant) perimeter is a vanishing part of the total length. Hence, for large  $\varepsilon$ :

- the fraction of the points that are the closest to one of the sides tends to 0, while
- the fraction of the points  $p$  for which the *closest* is one of the *edges* tends to 1.

**Geometric explanation: general case.** Similar arguments can be repeated for any dimension. For the same noise level  $\sigma$ , when  $d$  increases, the distance  $\varepsilon = \sigma \cdot \sqrt{d}$  also increases, and therefore, for large  $d$ , for “almost all” observed points  $\vec{o}$ , reconstructed signal is one of the extreme points of the a priori set  $A$ .

Much less probable is that the reconstructed signal belongs to the 1-dimensional face of the set  $A$ , even much less probable that  $s$  belongs to a 2-D face, etc.

**Methodological consequences.** The main *methodological consequence* of this result is that even when the actual state space is *continuous*, when we determine the state from measurements result, we inevitably obtain (most often) one of the *discretely many* states. On the large-scale level, we get one of the few *clusters*. When we add new measurements and thus, get to the next level, each original cluster subdivides into new clusters, etc., so that we get a *hierarchical* structure.

### 3. Our Explanation of the Success of Crisp Techniques

Tsirelson’s result explains why in spite of the clearly fuzzy character of most human reasoning, binary logic describes most of this reasoning pretty well.

Indeed, states with unusual “truth values” (different from 0 and 1) are not an exception, but rather a *general* rule. However, if we do the observations in the presence of some noise (e.g., if we use a not-perfect procedure for describing the values of the membership function), then we will mainly notice the *extreme* points of the set  $[0, 1]$  of the truth values, i.e., the values 0 and 1.

#### 4. In the Future, Fuzzy Techniques Will Be More and More Important

An interesting corollary is that as observations become more accurate, we will observe the actual intermediate fuzzy values as well, and crisp description will become more and more difficult.

*Comment.* Similarly, we can explain Schrödinger's paradox in quantum mechanics (see Appendix).

#### 5. Conclusion

In this paper, we provide a geometric explanation of why in many cases, in spite of a fuzzy nature of human reasoning, crisp models work amazingly well.

#### Acknowledgments

This work is supported, in part, by Chiang Mai University, Thailand, and by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber- ShARE Center of Excellence) and DUE-0926721.

#### References

- [1] R. P. Feynman, R. B. Leighton and M. Sands (2005). *The feynman lectures on physics*, Addison Wesley, Boston, MA.
- [2] G. J. Klir and B. Yuan (1995). *Fuzzy sets and fuzzy logic: Theory and applications*, Prentice Hall, Upper Saddle River, NJ.
- [3] H. T. Nguyen and E. A. Walker (2006). *A first course in fuzzy logic*, CRC Press, Boca Raton, FL.
- [4] H. T. Nguyen, B. Wu and V. Kreinovich (2000). On combining statistical and fuzzy techniques: Detection of business cycles from uncertain data. In H. Mohanty & C. Baral (Eds.), *Trends in information technology, proceedings of the International Conference on Information Technology ICIT* (pp. 69-74), Tata McGraw-Hill, New Delhi, India.
- [5] D. J. Sheskin (2007). *Handbook of parametric and nonparametric statistical procedures*, Chapman & Hall/CRC, Boca Raton, FL.
- [6] B. S. Tsirel'son (1986). A geometrical approach to maximum likelihood estimation for infinite-dimensional Gaussian location. II, *Theory of Probability & Its Applications*, 30, 820-828.
- [7] L. A. Zadeh (1965). Fuzzy sets, *Information and Control*, 8, 338-353.

## Appendix: Explanation of Schrödinger's Paradox

In *classical physics*, it is assumed that for each state of a physical system, every property is either true or false. For example, a particle is either located in a certain interval of space coordinates  $[x - \Delta, x + \Delta]$ , or it is not located inside this interval.

In *quantum mechanics*, in addition to the states in which a particle is located within this interval, and to the states in which the particle is definitely outside it, there are states in which *some* measurements of the coordinate will lead to results *within* the interval, and *some* to the results *outside* this interval.

In such states, we cannot say that a statement “the particle is located in the given interval” is true or that this statement is false; at best, we can determine the *probability* of the “yes” answer. (To describe such unusual “truth value,” quantum logic has been introduced.)

States with unusual “truth values” are not an exception, but rather a *general* rule in quantum mechanics: e.g., for every two states  $\psi$  and  $\psi'$  with certain values  $\lambda \neq \lambda'$  of a measured quantity, there exists a state called their *superposition* in which the value of this quantity is no longer certain. (In the standard formalism of quantum mechanics, where states are described by vectors in a Hilbert space, superposition is simply linear combination.)

Such superposition state is easy to generate.

Schroedinger has shown that this *superposition principle* seemingly contradicts our intuition; see, e.g., [1].

Indeed, suppose that we have a cat in a box, and a light-controlled rifle is aimed at the cat in such a way that a left-polarized photon would trigger the rifle and kill the cat, while the right-polarized photon would keep the cat alive.

If we send a photon with a circular polarization (that is, according to quantum mechanics, a superposition of left- and right-polarized states), we would get (due to the *linear* character of the equations of quantum mechanics), the *superposition* of the states resulting from using left- and right-polarized photons. In other words, we will get a superposition of a *dead* and *alive* cat states. This is, however, something that no one has ever observed: for macroscopic objects (cats included), an object is either dead or alive. Tsirelson's result explains why such non-extremal states are indeed difficult to observe.

Copyright of International Journal of Intelligent Technologies & Applied Statistics is the property of Airiti Press Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.