International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems
Vol. 0, No. 0 (1993) 000–000


# FUZZY/PROBABILITY ~ FRACTAL/SMOOTH

HUNG T. NGUYEN
*Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88003, USA, email hunguyen@nmsu.edu*

VLADIK KREINOVICH
*Department of Computer Science, University of Texas at El Paso, El Paso, TX 79968, USA, email vladik@cs.utep.edu*

BERLIN WU
*Department of Mathematical Sciences, National Chengchi University, Taipei, Taiwan, email berlin@math.nccu.edu.tw*

Received 15 January 1999
Revised 21 June 1999

Many applications of probability theory are based on the assumption that, as the number of cases increase, the relative frequency of cases with a certain property tends to a number – *probability* that this property is true. L. Zadeh has shown that in many real-life situations, the frequency oscillates and does not converge at all. It is very difficult to describe such situations by using methods from traditional probability theory. Fuzzy logic is not based on any convergence assumptions and therefore, provides a natural description of such situations. However, a natural next question arises: how can we describe this oscillating behavior? Since we cannot describe it by using a *single* parameter (such as probability), we need to use a *multi-D* formalism. In this paper, we describe an optimal formalism for describing such oscillations, and show that it complements traditional probability techniques in the same way as fractals complement smooth curves and surfaces.

*Keywords*: Multi-D degrees of belief, fractal, fuzzy, probability.

## 1. Introduction

One of the most natural ways to describe a degree of belief $d(A)$ in a statement $A$ is by asking $N$ experts and by taking the ratio $d_N(A) = Y(N,A)/N$ of those who believe in $A$ as the desired degree $d(A)$[2,1,4,8]. Ideally, the more experts we ask (i.e., the larger $N$), the better estimates we get; in mathematical terms, as $N$ increases, the estimates $d_N(A)$ converge to the *actual* (*limit*) value $d(A)$. However, in real life, there are problems with this definition of degree of confidence.

The *first problem* is that in some situations, this definition assigns *the same* degree of confidence $d(A) = d(A')$ to two different statements while it is intuitively clear that our confidence in the first statement is much larger than our confidence in the second one. Let us give an example. As $A$, we can take a statement which is clearly false; then, $d(A) = 0$. As $A'$, we take a statement on the "cutting edge" of science, a statement the truth of which has just been discovered, and which is still only known to the top experts in the field. For this statement:

- while $N$ is smaller than the number of these top researchers, the value $Y(N, A')$ grows, but
- as soon as $N$ exceeds the number of these top researchers, the value $Y(N, A')$ stays constant, does not increase with $N$ and, therefore, the ratio $d_N(A') = Y(N, A')/N$ tends to 0 as $N \to \infty$.

If we simply use the limit as $d(A')$, then we would have $d(A') = 0 = d(A)$, while intuitively, our degree of confidence in $d(A')$ is much higher than in $A$.

The *second problem* is that in some real-life situations, the sequence $d_N(A)$ does not seem to tend to any limit at all. For example, we may have a statement $A$ which seems intuitively true (e.g., that "optimism helps to fight a disease"), contradicts to the accepted science, but has been recently re-discovered and experimentally confirmed (so recently that it is not yet common knowledge among experts). Then, as we increase the number $N$ of experts, the ratio $d_N(A)$ exhibits the following oscillating behavior:

- at first, when $N$ is reasonably small, so that we only consider top experts in the field, we have $Y(N, A) \approx N$, and $d_N(A) \approx 1$;
- then, as we start including more and more experts who are not yet at the top research level, the number $Y(N, A)$ stays fixed, while $N$ increases, so we get $d_N(A) \approx 0$;
- finally, when we increase $N$ to such an extent that our list of experts starts including people with commonsense reasoning, the value $Y(N, A)$ again starts increasing as $Y(N, A) \approx N$, and the ratio $d_N(A)$ shoots back to 1.

In more sophisticated examples, we may have even more oscillations. For example, L. Zadeh gives an example of such oscillating behavior in estimating the probability that he (or any other person) will have a tax audit:

- first, we can consider all people in the US, and get a certain probability;
- as we go from the US as a whole to California, the probability of an audit increases;
- as we further narrow down the list to only people from Berkeley (thus, excluding Silicon Valley and Los Angeles), the probability goes down again;
- if we only consider middle-class people from Berkeley, the probability goes up again;
- as we further restrict ourselves to university professors, this probability goes down, etc.

The point that Zadeh makes is that it is very difficult to describe such an oscillating process by using methods from probability theory, which presumes a convergence. A natural next question is: how can we describe this oscillating behavior? Since we cannot describe it by using a *single* parameter (such as probability), we need to use a *multi-D* formalism.

We want to describe possible asymptotics of $Y(N,A)$ (and of the ratio $d_N(A)$) as $N$ increases. In the traditional probability approach, we have a one-parametric family of asymptotics $Y(N,A) \sim p \cdot N$, with a parameter $p$ (which leads to $d_N(A) \sim p$). In a more general multi-D case, it is natural to consider families with *several* parameters, i.e., families of the type $\{C_1 \cdot f_1(N) + \ldots + C_n \cdot f_n(N)\}$, where $f_1(N), \ldots, f_n(N)$ are given functions, and $C_i$ are arbitrary constants. We would like to describe the families which are *the best* in describing expert estimates. Since we do not have a precise formalization of what "the best" means, the problem of choosing the best family is the problem of optimization under an uncertain criterion. In[14], we have described a general formalism for solving such problems, and we have shown that this formalism is in good accordance with the empirical optimality of different fuzzy, neural, genetic, and other techniques. So, we will use this approach to describe the best families.

## 2. Optimal in what sense?

**The main idea.** We are looking for the *best* (*optimal*) choice of a potential function.

Normally, the word "best" is understood in the sense of some *numerical* optimality criterion. However, in our case of *fuzzy* choice, it is often difficult to formulate the exact *numerical* criterion. Instead, we assume that there is an *ordinal* criterion, i.e., that we can compare arbitrary two choices, but that we cannot assign numerical values to these choices.

It turns out that in many cases, there are reasonable *symmetries*, and it is natural to assume that the (ordinal) optimality criterion is invariant with respect to these symmetries. Then, we are able to describe all choices that are optimal with respect to some invariant ordinal optimality criteria.

This general approach was described and used in[3,9,10,14,15,16], in particular, for fuzzy control. In this section, we will show that this approach is applicable to fuzzy elicitation as well.

**Let us borrow from the experience of modern physics and use symmetries.** In modern physics, symmetry groups are a tool that enables to compress complicated differential equations into compact form (see, e.g.,[7]). Moreover, the very differential equations themselves can be uniquely deduced from the corresponding symmetry requirements (see, e.g.,[5,6]).

**It is possible to use symmetry.** As we have mentioned, in our previous papers, we have shown that the symmetry group approach can be used to find optimal membership functions, optimal t-norms and t-conorms, and optimal defuzzification procedures.

It is therefore reasonable to expect that the same approach can also be used to choose the best potential function for fuzzy elicitation.

**What is a criterion for choosing a family of functions?** What does it mean to choose a *best* family of functions? It means that we have some *criterion* that enables us to choose between the two families.

Traditionally, optimality criteria are *numerical*, i.e., to every family $F$, we assign some value $J(F)$ expressing its quality, and choose a family for which this value is maximal (i.e., when $J(F) \geq J(G)$ for every other alternative $G$). However, it is not necessary to restrict ourselves to such numeric criteria only.

For example, if we have several different families $F$ that have the same adequacy $P(F)$, we can choose between them the one that has the minimal computational complexity $C(F)$. In this case, the actual criterion that we use to compare two families is not numeric, but more complicated:

> *A family $F_1$ is better than the family $F_2$ if and only if*
>
> – *either $P(F_1) > P(F_2)$,*
> – *or $P(F_1) = P(F_2)$ and $C(F_1) < C(F_2)$.*

A criterion can be even more complicated.

The only thing that a criterion *must* do is to allow us, for every pair of families $(F_1, F_2)$, to make one of the following conclusions:

- the first family is better with respect to this criterion (we'll denote it by $F_1 \succ F_2$, or $F_2 \prec F_1$);
- with respect to the given criterion, the second family is better ($F_2 \succ F_1$);
- with respect to this criterion, the two families have the same quality (we'll denote it by $F_1 \sim F_2$);
- this criterion does not allow us to compare the two families.

Of course, it is necessary to demand that these choices be consistent.

> For example, if $F_1 \succ F_2$ and $F_2 \succ F_3$ then $F_1 \succ F_3$.

**The criterion must be final, i.e., it must pick the unique family as the best one.** A natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family).

The reason for this demand is very simple: If a criterion *does not choose* any family at all, then it is of no use. If *several* different families are the best according to this criterion, then we still have the problem of choosing the best among them. Therefore we need some additional criterion for that choice, like in the above example:

If several families $F_1, F_2, \ldots$ turn out to have the same adequacy ($P(F_1) = P(F_2) = \ldots$), we can choose among them a family with minimal computational complexity ($C(F_i) \to \min$).

So what we actually do in this case is abandon that criterion for which there were several "best" families, and consider a new "composite" criterion instead: $F_1$ is better than $F_2$ according to this new criterion if either it was better according to the old criterion, or they had the same quality according to the old criterion and $F_1$ is better than $F_2$ according to the additional criterion.

In other words, if a criterion does not allow us to choose a unique best family, it means that this criterion is not final, we'll have to modify it until we come to a final criterion that will have that property.

**The criterion must not change whether we count experts or schools of experts.** When we talk about counting experts, we can literally count them. However, this may not always be the best approach, because the whole idea of increasing $N$ is to increase the *diversity* of the experts, and so, if, e.g., two experts belong to a single school of researchers (and therefore, have similar views), it may not make big sense to interview both. Instead, we should interview people from different schools, and count these *schools*, not individual researchers.

When we count *researchers*, the value $Y(N, A)$ is always an integer. When we count *schools*, and the school is divided on this particular issue (the larger schools we take, the more frequent such a situation will be), then it is natural to add 1/2 (or whatever ratio is in this school) to the total number of schools that believe in $A$. In this case, the value of $Y(N, A)$ is not necessarily an integer: it can be an arbitrary *rational* number. In this case, it is natural to assume that the function $f(N)$ that approximates $Y(N, A)$ can take arbitrary real values.

The notion of the "school of researchers" may mean different things; for example:

- we can divide the researchers into large groups whose views are more or less similar, but can be different in details;
- we can also divide the researchers into very small groups with practically identical views.

The exact mathematical form of an approximating function $f(N)$ depends on how we count these "schools of scientists". If we re-define a school, and in our new definition, a school is $\lambda$ times smaller, then $N$ old schools correspond to $N' = \lambda \cdot N$ new schools, and, correspondingly, the original number $b = f(N) = Y(N, A)$ of schools who believed in $A$ changes to $b' = \lambda \cdot Y(N, A)$. In terms of the new number of school $N' = \lambda \cdot N$, this new dependence takes the form $b' = f_\lambda(N')$, where $f_\lambda(N) = \lambda \cdot f(N/\lambda)$.

It is reasonable to assume that the relative quality of different families should not change if we simply change the units, i.e., if the family $F$ is better than a family $G$, then the transformed family $F_\lambda$ should also be better than the family $G_\lambda$. Now, we are ready for the formal definition.

## 3. Definitions and the Main Result

By a *family* $F$, we mean a family of all functions of the type $f(x) = C_1 \cdot f_1(x) + \ldots + C_n \cdot f_n(x)$, where $f_i(x)$ are differentiable functions from $R$ to $R$.

A pair of relations $(\prec, \sim)$ on a set $\Phi$ is called *consistent* if it satisfies the following conditions, for every $F, G, H \in \Phi$: (1) if $F \prec G$ and $G \prec H$ then $F \prec H$; (2) $F \sim F$; (3) if $F \sim G$ then $G \sim F$; (4) if $F \sim G$ and $G \sim H$ then $F \sim H$; (5) if $F \prec G$ and $G \sim H$ then $F \prec H$; (6) if $F \sim G$ and $G \prec H$ then $F \prec H$; (7) if $F \prec G$ then it is not true that $G \prec F$, and it is not true that $F \sim G$.

Assume a set $\Phi$ is given. Its elements will be called *alternatives*. By an *optimality criterion*, we mean a consistent pair $(\prec, \sim)$ of relations on the set $\Phi$ of all alternatives. If $F \succ G$ we say that $F$ is *better* than $G$; if $F \sim G$ we say that the alternatives $F$ and $G$ are *equivalent* with respect to this criterion. We say that an alternative $F$ is *optimal* (or *best*) with respect to a criterion $(\prec, \sim)$ if for every other alternative $G$ either $F \succ G$ or $F \sim G$. We say that a criterion is *final* if there exists an optimal alternative, and this optimal alternative is unique.

In this paper, we will consider optimality criteria on the set $\Phi$ of all families.

Let $\lambda > 0$ be a positive real number. By a *$\lambda$-rescaling* of a function $f(x)$, we mean a function $f_\lambda(x) = \lambda \cdot f(x/\lambda)$. By a *$\lambda$-rescaling* $F_\lambda$ of a family of functions $F$ we mean the family consisting of $\lambda$-rescalings of all functions from $F$.

We say that an optimality criterion on $\Phi$ is *unit-invariant* if for every two families $F$ and $G$ and for every number $\lambda > 0$, the following two conditions are true: i) if $F$ is better than $G$ in the sense of this criterion (i.e., $F \succ G$), then $F_\lambda \succ G_\lambda$; ii) if $F$ is equivalent to $G$ in the sense of this criterion (i.e., $F \sim G$), then $F_\lambda \sim G_\lambda$.

**Theorem.**[14] *If a family $F$ is optimal in the sense of some optimality criterion that is final and unit-invariant, then every function $f(x)$ from this family $F$ is equal to a linear combination of the functions of the type*

$$f(x) = x^\alpha \cdot (\ln(x))^p \cdot \sin(\beta \cdot \ln(x) + \varphi),$$

*where $p$ is a non-negative integer, and $\alpha$, $\beta$ and $\varphi$ are real numbers.*

This theorem was proven in[14]. This result (to be more precise, the appearance of sin) explains the above-described "oscillating" behavior of $Y(N, A)$.

As shown in[14], for $n = 1$, the only possible families are $\{C_1 \cdot N^\alpha\}$ considered in[11,13]: for $\alpha = 1$, we get functions corresponding to a probability approach; in[11,13], we also give an interpretation for families corresponding to $\alpha \neq 1$.

For $n = 2$, we already have a possibility of an oscillating function

$$f(N) = C_1 \cdot N + C_2 \cdot N \cdot \sin(\beta \cdot \ln(N)).$$

For this function, as sin oscillates between $-1$ and 1, the ratio $d_N(A) = f(N)/N$ oscillates between $C_1 - C_2$ and $C_1 + C_2$. Thus, it is natural to say that the corresponding degree of certainty is an *interval* $[C_1 - C_2, C_1 + C_2]$. However, the exact form gives us more information than the interval: namely, it also describes the "oscillation rate" $\beta$.

## 4. Analogy between fuzzy and fractal

The difference between the situations which are easily describable by methods of *probability* theory, and more complicated situations which require *non-probabilistic* (*fuzzy*) description can be described in terms of the asymptotics of $Y(N, A)$:

- in the probabilistic case, $Y(N, A) \sim p \cdot N$, while
- in the fuzzy case, we have a more general asymptotics, e.g., $Y(N, A) \sim p \cdot N^\alpha$.

This difference is similar to the difference between *smooth* curves (or surfaces) and *fractal* curves (surfaces)[12]. Indeed, according to the definition of a fractal, a *fractal* is a set of a fractal (non-integer) dimension, and dimension of a set is defined in terms of finite approximations. For a given real number $\varepsilon$, we say that a finite set $S \subseteq A$ is an *$\varepsilon$-approximation* to a set $A$ if every point $a \in A$ is $\varepsilon$-close to one of the points from $S$. The smaller $\varepsilon$, the more points we need to approximate a given set $A$. Thus, as a natural measure of complexity of a set $A$, we can take, for every $\varepsilon$, the smallest number $N_\varepsilon(A)$ of points which are necessary to approximate the set $A$ with an accuracy $\varepsilon$.

In the simplest case when the set $A$ is the interval $[0, 1]$, we can explicitly compute $N_\varepsilon(A)$. Indeed, in this set $S$, we need a point which is $\varepsilon$-close to 0, i.e., which is $\le \varepsilon$. This point $s_1$ covers everything from 0 to $s_1 + \varepsilon$. The further this point from 0, the less area is left for other points to cover, so the smallest possible number of points occurs when $s_1$ is at the farthest possible location, i.e., when $s_1 = \varepsilon$. The above-chosen point $s_1$ covers all points from 0 to $s_1 + \varepsilon = 2\varepsilon$, so to cover points right after $2\varepsilon$, we need a next point at a location $s_2 \le 2\varepsilon + \varepsilon = 3\varepsilon$; a similar argument shows that the smallest number of points is when we take $s_2 = 3\varepsilon$. Similarly, $s_k = (2k - 1) \cdot \varepsilon$. So, we need $N_\varepsilon([0, 1]) \sim 1/2\varepsilon$ points to cover the entire interval $[0, 1]$. Similarly, for any other smooth curve $\gamma$, $N_\varepsilon(\gamma) \sim \text{const}/\varepsilon$.

Similarly, for a 2-D smooth *surface* $A$, we have $N_\varepsilon(A) \sim \text{const}/\varepsilon^2$.

For a 3-D domain with a smooth boundary, we have $N_\varepsilon(A) \sim \text{const}/\varepsilon^3$.

In all these smooth cases, $N_\varepsilon(A) \sim \text{const}/\varepsilon^D$, where $D$ is the dimension of the set $A$. In view of this fact, we can define *dimension* for *non-smooth* sets $A$ as a parameter $\alpha$ for which $N_\varepsilon(A) \sim \text{const}/\varepsilon^\alpha$. Then, a set is a *fractal* if either $\alpha$ is not an integer, or if the asymptotics of $N_\varepsilon(A)$ is more complicated than $\text{const}/\varepsilon^\alpha$.

Similarly, in our case:

- when $Y(N, A) \sim p \cdot N$, we get a *probabilistic* situation, and
- when $Y(N, A) \sim \text{const} \cdot N^\alpha$ for some $\alpha \ne 1$, or if the asymptotics of $Y(N, A)$ is more complicated than $N^\alpha$, we have a non-probabilistic (*fuzzy*) situations.

Thus, we can say that fuzzy is a generalization of probability in the same sense in which fractals are a generalization of smooth surfaces:

$$\frac{\text{Fuzzy}}{\text{Probability}} = \frac{\text{Fractal}}{\text{Smooth}}.$$

## Acknowledgments

This work was supported in part by Taiwan's NSC grant No. NSC85-2415-H004-013b, by NASA under cooperative agreement NCC5-209, by NSF grant No. DUE-9750858 and CDA-9522207, by the United Space Alliance grant No. NAS 9-20000 (PWO C0C67713A6), by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, grant No. F49620-95-1-0518, and by NSA grant No. MDA904-98-1-0564.

## References

1. J. M. Blin, "Fuzzy relations in group decision theory", *J. of Cybernetics* **4** (1974) 17–22.
2. J. M. Blin and A. B. Whinston, "Fuzzy sets and social choice", *J. of Cybernetics* **3** (1973) 28–36.
3. B. Bouchon-Meunier *et al.*, "On the formulation of optimization under elastic constraints (with control in mind)", *Fuzzy Sets and Systems* **81** (1996) 5–29.
4. D. Dubois and H. Prade, *Fuzzy sets and systems: theory and applications* (Academic Press, N.Y., London, 1980).
5. A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry: towards mathematical foundations", *International Journal of Theoretical Physics* **36** (1997) No. 4, 1009–1020.
6. A. M. Finkelstein, V. Kreinovich, and R. R. Zapatrin, "Fundamental physical equations uniquely determined by their symmetry groups," *Lecture Notes in Mathematics*, (Springer-Verlag, Berlin-Heidelberg-N.Y., Vol. 1214, 1986) 159–170.
7. A. Frank and K. B. Wolf, *Symmetries in physics* (Springer-Verlag, Berlin, N.Y., 1992).
8. G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications* (Prentice Hall, Upper Saddle River, NJ, 1995).
9. V. Kreinovich, C. Quintana, and R. Lea, "What procedure to choose while designing a fuzzy control? Towards mathematical foundations of fuzzy control", *Working Notes of the 1st International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems* (College Station, TX, 1991) 123–130.
10. V. Kreinovich *et al.*, "What non-linearity to choose? Mathematical foundations of fuzzy control", *Proceedings of the 1992 International Conference on Fuzzy Systems and Intelligent Control* (Louisville, KY, 1992) 349–412.
11. C. Langrand, V. Kreinovich and H. T. Nguyen, "Two-dimensional fuzzy logic for expert systems", *Sixth International Fuzzy Systems Association World Congress* (San Paulo, Brazil, July 22–28, 1995) **1** 221–224.
12. B. B. Mandelbrot, *The fractal geometry of Nature* (Freeman, San Francisco, 1982).
13. H. T. Nguyen and V. Kreinovich, "Towards theoretical foundations of soft computing applications", *Int'l J. on Uncertainty, Fuzziness, and Knowledge-Based Systems* **3** (1995) No. 3, 341–373.
14. H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science* (Kluwer, Dordrecht, 1997).
15. E. R. Scerri, V. Kreinovich, P. Wojciechowski, and R. R. Yager, "Ordinal explanation of the periodic system of chemical elements", *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems (IJUFKS)* **6** (1998) No. 4, 387–399.
16. M. H. Smith and V. Kreinovich, "Optimal strategy of switching reasoning methods in fuzzy control", In H. T. Nguyen *et al.* (eds.), *Theoretical aspects of fuzzy control* (J. Wiley, N.Y., 1995) 117–146.