From Interval (Set) and Probabilistic Granules to Set-and-Probabilistic Granules of Higher Order

Vladik Kreinovich

Abstract In this chapter, we provide a natural motivation for granules of higher order, and we show that these granules provide a unified description of different uncertainty formalisms such as random sets, Dempster-Shafer approach, fuzzy sets, imprecise probabilities, and Bayesian statistics. We also prove that for fuzzy uncertainty, granules of second order are sufficient.
Keywords: Granules of Higher Order, Random Sets, Dempster-Shafer Approach, Fuzzy Sets, Imprecise Probabilities, Bayesian Statistics.

1 Introduction

Techniques for representing uncertainty: current situation. Many techniques have been proposed (and successfully used) to describe and process uncertainty:

- sets, in particular, intervals (Rabinovich 2005);
- probability distributions (Jaynes 2003), (Rabinovich 2005);
- imprecise probabilities (Walley 1989), (Ferson 2002), (Ferson et al. 2003);
- Dempster-Shafer approach;
- fuzzy sets;
- interval-valued (and, more generally, type-2) fuzzy sets (Mendel 2001), (Mendel Wu 2010);
- Bayesian statistical techniques (Jaynes 2003), (Gelman et al. 2004);
- and many other different approaches.

Natural questions. From this variety, come natural questions:

- Which of the techniques should we apply in different practical situations?
- Are there yet-to-be-discovered better techniques for processing uncertainty?

Vladik Kreinovich
University of Texas at El Paso, El Paso, TX 79968, USA, e-mail: vladik@utep.edu

- How can we combine uncertainty described by different formalisms?

How to answer these questions: a need for classification. To answer all these questions, it is necessary to come up with a reasonable classification of different uncertainty techniques. Such classification would enable us:

- to meaningfully compare different techniques, for the purpose of deciding which is better in different situations,
- to check whether all possible techniques in this classification have been invented, or new (possible better) techniques are still possible, and
- to meaningfully combine these techniques – since they appear as particular cases of a general approach to uncertainty.

What we do in this paper. In this paper, we start with the question of where our data comes from, what are the corresponding uncertainties. Based on the corresponding from-scratch analysis, we explain how different uncertainty techniques appear, and thus, come up with a natural classification.

In the process of this classification, we provide a natural motivation for granules of higher order. We also show that these granules provide a unified description of different uncertainty formalisms.

2 From an Ideal Exact Description to Interval (Set) and Probabilistic Granules

Idealized objects. To describe the physical world, we identify *objects*: elementary particles, atoms, molecules, solid bodied, stars, etc.

An ideal object should be well-defined. For example, in a geographic description, when we define forests, lakes, rivers, etc., we should be able to determine the exact boundary between a lake and a river that flows into this lake, the exact boundary between a forest and a nearby grassy area, etc.

An ideal object should also be reasonably stable with time. From this viewpoint,

- a river is a reasonable geographic object, because its path does not change much for a long time, while
- a puddle – which is often easily visible too – is not a reasonable geographic object, because it can disappear in the course of hours.

Group objects. The above description applied to *individual* objects: we can talk about the height of an individual man, the speed with which the individual river flows, etc.

In physics, we are also interested in "group" objects: the mass of an electron, the magnetic moment of an ion of a specific type. Similarly, in manufacturing, we are interested in the speed and/or fuel efficiency of a certain type of a car, in the

frequency of a certain type of a laser, etc. In all these cases, instead of dealing with an *individual* object, we have a *collection* of similar objects.

In the ideal case, we assume that all these objects are *identical*, so whatever we can observe based on one of these objects can be applied to others as well. For example, if we apply safety tests to several cars of the same make and model, and these tests are successful, we conclude that all the cars of this make and model are safe.

Idealized description of idealized objects. Ideally, we should have a full description of each (idealized) object. In other words, for each possible quantity (such as location, mass, height, etc.), we should know the exact value of this quantity for the given object.

For many physical quantities, there are several different ways to measure the value of this quantity; ideally, all these ways should lead to the same numerical result. For example, a GPS location of the car can be measured based on different parts of this car; all these measurements should lead to the same result. Similarly, a width of a wooden plank should be the same no matter where we measure it.

In mathematical terms, an (idealized) exact description of an (idealized) object means that we know, for this object, the exact values $s_1, \ldots, s_n$ of all possible quantities of interest.

All such tuples $s = (s_1, \ldots, s_n)$ of real numbers form an n-dimensional space $\mathbb{R}^n$.

Usually, not all tuples are physically possible. For example, mass is always non-negative, velocity is always limited by the speed of light, etc. The actual state must therefore belong to the set $S \subseteq \mathbb{R}^n$ of all physically possible tuples.

Comment. In this paper, for simplicity, we assume that this set S is known. To get a more realistic description, we must take into account that this set is not exactly known – e.g., the speed of light is only approximately known.

Objects are not ideal: aleatoric uncertainty. As we have mentioned, ideally, objects should be well-defined and stable (not changing with time).

In practice, objects are often not well defined. For example, when a river flows into the lake, it is often not clear where is the boundary between the river and the lake.

Objects are also not perfectly stable: they change, slightly but change. For example, the weight of a person slightly changes – when she breathes in and breathes out, when she sweats, etc.

As a result, for the same object, even an ideally accurate measuring instrument can measure different values:

- the measurement results differ with time because the object changes,
- these results differ because we may select different boundaries for the object, etc.

This "objective" difference is known as *aleatoric* uncertainty.

Objects are not perfectly identical: another case of aleatoric uncertainty. In the ideal case, we assumed that all the "instance" of a group object are identical. In reality, different objects are slightly different. For example, different cars of the same make and model may have slightly different fuel economy characteristics.

Different measuring procedures can lead to different values: yet another case of objective (aleatoric) uncertainty. In the ideal case, all possible procedures for measuring a given quantity should lead to the exact same result. In practice, different procedure may lead to slightly different results.

For example, for a non-ideal wooden plank, its width may differ slightly from one location to another.

Measurements are not 100% accurate: epistemic uncertainty. In the ideal case, we assumed that all the measurements lead to the exact values of the measured quantity. In practice, measurements are never 100% accurate; the result $\widetilde{x}$ of each measurement is, in general, slightly different form the actual (unknown) value x of the desired quantity: $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x \neq 0$; see, e.g., (Rabinovich 2005).

Thus, even in the ideal case, when the object is well-defined and stable – i.e., where there is no aleatoric ("objective") uncertainty, measurement results can be different. This uncertainty is "subjective" in the sense that it characterizes not the object itself, but rather our *knowledge* about this object. Such uncertainty is called *epistemic*.

How to describe aleatoric uncertainty: set (interval) granules. In the ideal case, each object is characterized by a single state $s \in S$.

Due to aleatoric uncertainty, for the same object, we may get *different* states $s \in S$, depending on:

- at what time we measure the corresponding quantities;
- how we define the boundaries of the object;
- which object from the group of (almost) identical objects we take; and
- which of the possible measuring procedures we use to measure the corresponding quantities.

Thus, to fully characterize the situation, we must know which states s are possible for this object and which are not. In other words, in view of the aleatoric uncertainty, to describe an object, we need to know the *set* $\mathbf{s} \subseteq S$ of all possible states characterizing this object. Different states $s \in \mathbf{s}$ are "equally" possible, so this set forms a single set *granule* characterizing the object.

This set $\mathbf{s}$ is usually *connected* – in the sense that there is a continuous transition between every two possible states $s, s' \in \mathbf{s}$, In the 1-D case, this connectivity implies that with every two possible states s and s', all intermediate states are also possible – i.e., that the set $\mathbf{s}$ of all possible states is an *interval*.

How to describe aleatoric uncertainty: probabilistic granules. In addition to knowing which states are possible and which are not, it is also desirable to know how frequent are different possible states.

For example, within a population of cars of the same make and model, it is desirable not only to know that the fuel efficiency of an individual car may be lower than on average, it is also desirable to know how frequent are such low-efficiency situation:

- if they are rare, it is acceptable, but

- if such situations are frequent, this is a strong argument against buying this particular model.

Thus, it is desirable to know, for each possible value $s \in \mathbf{s}$, a frequency (= probability) that this value occurs when measuring this object. If we know this information, then, instead of a set granule, we have a *probabilistic* granule.

Description of epistemic uncertainty: set (interval) and probabilistic granules. Similarly, we can describe *epistemic* uncertainty.

Indeed, once we know the results $\widetilde{s} = (\widetilde{s}_1, \ldots, \widetilde{s}_n)$ of measuring the desired quantities, we would like to know which states s are compatible with these measurement results. In other words, we need to know the set $\mathbf{s}$ of all such states – i.e., a set (interval) granule.

In addition to knowing which states $s \in S$ are possible and which are not possible, it is also desirable to know which values $s \in \mathbf{s}$ are more frequent and which are less frequent – i.e., it is desirable to know the *frequency* (*probability*) of different value s. Thus, we also arrive at the need to consider a probability distribution on the set $\mathbf{s}$ of possible states – i.e., a probabilistic granule.

Yet another reason for granulation: need to speed up decision making. Even when we know the actual element s from S with a good accuracy, it is still often reasonable, when making a decision, to ignore this difficult-to-process accurate information and to base our decision on the *type* of an object – i.e., on the granule to which this object belongs.

For example, when an animal attacks a person, it makes sense to ignore the animal's eye color and other details of the animal and concentrate of the type of the animal: e.g., is it a small (mostly harmless) dog or a dangerous tiger.

3 Need for Granules of Granules – i.e., for Granules of Higher Order

Need for granules of higher order. Ideally, to characterize the uncertainty, we describe

- either the set of possible values
- or the probability distribution on the set of possible values.

This description assumes that we know *exactly* which values are possible and which values are not possible – and we know the *exact* values of the corresponding probabilities.

In practice, we are not always sure which values are possible and which are not, and do not know the exact values of the corresponding probabilities. In other words, our knowledge of the corresponding uncertainty is also uncertain.

Thus, instead of a single set-valued granule, we have a *granule* of possible set-valued granules – a construction which can be naturally described as a granule of

higher order. Similarly, instead of a single probabilistic granule, we have a *granule* of possible probabilistic granules – a construction which can also be naturally described as a granule of higher order.

Motivation for restricting ourselves to granules of second order. For example, instead of a single set of possible states, we can have a class of possible sets – a second order construction. This class may also not be exactly known – so we should consider class of possible classes, a third order construction.

However, from the computational viewpoint, a set is already difficult to process, a class of sets is even more complex, and a class of classes is practically impossible to analyze.

Comment. For fuzzy uncertainty, a deeper argument in favor of second-order granules is given in the Appendix; this argument was first outlined in (Nguyen Kreinovich 1998) and (Kreinovich Nguyen 2001).

4 Second-Order Granules: Natural Classification

Classification: general idea. An ideal description of uncertainty is to describe as a granule – i.e.,

- either a set of possible values
- or a probability distribution on the set of possible values.

In reality, we do not have the exact knowledge of the corresponding granule (i.e., of the set or of the probability distribution).

When we did not know the exact state, we considered either the set of all possible states or a probability distribution on the set of possible states. Similarly, when we do not know the exact granule, we have to consider:

- either a set of possible granules,
- or a probability distribution on the set of possible granules.

In each of these two cases, we have two subcases, depending on whether we consider set-valued (interval) or probabilistic granules. Thus, we arrive at the four possible situations:

1. a set of possible sets;
2. a set of possible probability distributions;
3. a probability distribution on the class of possible sets; and
4. a probability distribution on the class of possible probability distributions.

What we plan to show. At first glance, we have a very mathematical classification. However, as we will show, these four types of second order granules correspond to well-known and well-used types of uncertainty – such as random sets, Dempster-Shafer approach, fuzzy sets, imprecise probabilities, and Bayesian statistics.

Thus, the idea of second order granules provides a natural unified description of different formalisms for describing uncertainty.

First case: set of possible sets. In the first case, instead of selecting a single set of possible values, we select a *class* of possible sets. This idea is actively used in representation of uncertainty. For example, in the *rough set* approach, each set S is represented by a set $\underline{S}$ that is contained in S and a set $\overline{S}$ that contains S. In this case, the only information that we have about the actual (unknown) set S is that $\underline{S} \subseteq S \subseteq \overline{S}$, i.e., that S belongs to the *set interval* $[\underline{S}, \overline{S}] \stackrel{\text{def}}{=} \{S : \underline{S} \subseteq S \subseteq \overline{S}\}$. General set intervals – not necessarily generated by rough sets – are also actively used; see, e.g., (Yao Li 1996), (Yao et al. 2008).

Second case: a set of probability distributions. In this case, instead of selecting a single probability distribution, we select a set of possible probability distributions. This description of uncertainty is known as *imprecise probability*; see, e.g., (Walley 1991).

An important particular case of imprecise probability is the case of p-boxes (probability boxes), i.e., interval bounds on the values of the cumulative distribution function $F(x) \stackrel{\text{def}}{=} \text{Prob}(\xi \leq x)$; see, e.g., (Ferson 2002), (Ferson et al. 2003).

Third case: a probability distribution on the class of possible sets. In this case, instead of selecting a single set, we select a probability distribution on the class of all possible sets. In other words, we assign, to each set, a probability that this is indeed the right set.

When we assign a probability to each number, we get a *random number*. When we assign a probability to each vector, we get a *random vector*. Similarly, when we assign a probability to each set, we get a *random set.* Random sets have indeed been actively applied in description and processing of uncertainty; see, e.g., (Nguyen 2008) and references therein.

In the discrete case, a random set means that we assign, to different sets A, values $m(A) \geq 0$ such that $\sum\limits_{A} m(A) = 1$. This is exactly the widely used *Dempster-Shafer* approach to describing uncertainty.

Relation to fuzzy. One of the reason why random sets are important in describing uncertainty is that they provide a reasonable alternative description of *fuzzy sets*. Specifically, one way to assign a membership degree to a statement like "25 is young" is to take N experts, ask these experts whether a person who is 25 years old is young, and take the portion $N(25)/N$ of experts who agree with this statement as the desired degree $\mu_{\text{young}}(25)$. This procedure assumes

- that the experts are equally important – because we give equal weight to opinions of different experts, and
- that each expert is capable of giving a precise (crisp) answer to each such question.

For each of N experts i ($1 \leq i \leq N$), there us denote, by S_i, the set of all ages which for which, for this expert, the person is young. In general, different experts have

different sets S_i, but it is possible that two or more different experts have the same set.

As we have mentioned, we assume that the experts are equally important, i.e., that each expert gets assigned the same probability $1/N$. So, if a set S_i occurs as an opinion of only one expert, we assign it the probability $1/N$; if it occurs as the common opinion of several (k) experts, we assign it the probability k/N. Thus, we have probabilities assigned to different sets, i.e., we have a random set. In terms of this random set, the degree to which, say, 25 is young, is simply equal to the sum of the probabilities of all the sets S_i that include 25, i.e., to the probability that 25 belongs to the random set.

In general, for the corresponding random set S, for every value x, the membership degree $\mu(x)$ is equal to the probability that $x \in S$: $\mu(x) = \text{Prob}(x \in S)$.

Fourth case: a probability distribution on the class of probability distributions. In this case, instead of selecting a single probability distribution, we select a probability distribution on the set of these distributions.

Theoretically, we can have an infinite-dimensional class of probability distributions, i.e., a class in which we need to know the values of infinitely many parameters to uniquely determine the distribution. In practice, in a computer, we can only store finitely many values of the parameters. Thus, from the practical viewpoint, each class of probability distributions is characterized by the values of finitely many parameters. There may be only two parameters – like in 1-D Gaussian distributions, there can be many more parameters as in more sophisticated classes, but there are always finitely many parameters.

In this case, selecting a probability distribution from the class means selecting the values of these parameters. Thus, the above situation can be described as follows:

- instead of selecting a unique set of parameters characterizing a probability distribution,
- we select a probability distribution on the set of these parameters.

This idea describes the *Bayesian* statistical approach, whose main idea is indeed to select a (prior) distribution on the set of all possible values of different parameters; see, e.g., (Jaynes 2003), (Gelman et al. 2004).

Summary. Second order granules approach covers many known uncertainty formalisms as particular cases:

	set of ...	probability distributions on ...
... sets	set intervals; rough sets	random sets; Dempster-Shafer approach; fuzzy approach
... probability distributions	imprecise probability; p-box	Bayesian statistics

What if we also consider fuzzy. In the above text, we only consider set-valued (interval) and probabilistic granules. These granules correspond to "objective"

(aleatoric) and measurement uncertainty. Expert estimates lead to fuzzy uncertainty, where we have fuzzy granules – i.e., fuzzy sets. If we add the possibility of fuzzy granules, then we get new possibilities of second-order granules:

- a set of fuzzy granules, i.e., a set of fuzzy sets – e.g., an interval-valued fuzzy set;
- a probability distribution on the set of fuzzy granules, i.e., a *random fuzzy set*; see, e.g., (Li et al. 2002);
- a fuzzy set of fuzzy sets – i.e., a general type-2 fuzzy set;
- a fuzzy class of sets or probability distributions – something which was tried in imprecise probability research.

Third order granules? In contrast to the second order granules which have many practical applications, third order ones are rarely used. A few practical examples include interval-valued fuzzy sets; see, e.g., (Nguyen Kreinovich 1995), (Nguyen et al. 1997), (Mendel 2001), (Mendel et al. 2010):

- since a fuzzy set can be interpreted as a random set,
- an interval-valued fuzzy set – i.e., a set of possible fuzzy sets – can be interpreted as a set of possible random sets, i.e., as a third order granule.

5 Conclusion

In this paper, we provided a natural motivation for granules of higher order, and showed that these granules provide a unified description of different uncertainty formalisms such as random sets, Dempster-Shafer approach, fuzzy sets, imprecise probabilities, and Bayesian statistics.

We also prove that within this general description, most reasonable uncertainty formalisms have already been discovered: for example, we prove that for fuzzy uncertainty, granules of second order are sufficient.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, by Grant MSM 6198898701 from MŠMT of Czech Republic, and by Grant 5015 "Application of fuzzy logic with operators in the knowledge based systems" from the Science and Technology Centre in Ukraine (STCU), funded by European Union.

The author is greatly thankful to the volume editors and to the anonymous referees for valuable suggestions.

References

Ferson S (2002) RAMAS Risk Calc 4.0: Risk assessment with uncertain numbers. CRC Press: Boca Raton, Florida

Ferson S, Kreinovich V, Ginzburg L, Myers DS, Sentz K (2003) Constructing Probability Boxes and Dempster-Shafer Structures. Sandia National Laboratories, Report SAND2002-4015, January 2003.

Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian Data Analysis. Chapman & Hall/CRC, Boca Raton, Florida

Jaynes ET (2003) Probability Theory: The Logic of Science. Cambridge University Press: Cambridge, Massachusetts

Kreinovich V, Nguyen HT (2001) 1st Order, 2nd Order, What Next? Do We Really Need Third-Order Descriptions: A View From A Realistic (Granular) Viewpoint. In: Proceedings of the Joint 9th World Congress of the International Fuzzy Systems Association and 20th International Conference of the North American Fuzzy Information Processing Society IFSA/NAFIPS 2001, Vancouver, Canada, July 25-28, 2001, pp. 1908–1913.

Li S, Ogura Y, Kreinovich V (2002) Limit Theorems and Applications of Set Valued and Fuzzy Valued Random Variables. Kluwer Academic Publishers, Dordrecht

Mendel J (2001) Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions. Prentice-Hall, Upper Saddle River, New Jersey

Mendel JM, Wu D (2010) Perceptual Computing: Aiding People in Making Subjective Judgments. IEEE Press and Wiley

Nguyen (2008) An Introduction to Random Sets, Chapman & Hall/CRC, Boca Raton, Florida

Nguyen HT, Kreinovich V (1995) Towards theoretical foundations of soft computing applications. International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems (IJUFKS) 3(3):341–373

Nguyen HT, Kreinovich V (1998) Possible new directions in mathematical foundations of fuzzy technology: a contribution to the mathematics of fuzzy theory, In: Phuong NP, Ohsato A (eds.), Proceedings of the Vietnam-Japan Bilateral Symposium on Fuzzy Systems and Applications VJFUZZY'98, HaLong Bay, Vietnam, 30th September–2nd October, 1998, pp. 9–32

Nguyen HT, Kreinovich V, Zuo Q (1997) Interval-valued degrees of belief: applications of interval computations to expert systems and intelligent control. International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems (IJUFKS) 5(3):317–358

Rabinovich S (2005) Measurement Errors and Uncertainties: Theory and Practice. Springer Verlag, New York

Walley P (1991) Statistical Reasoning with Imprecise Probabilities. Chapman & Hall: New York

Yao JT, Yao YY, Kreinovich V, Pinheiro da Silva P, Starks SA, Xiang G, Nguyen HT (2008) Towards more adequate representation of uncertainty: from intervals to

set intervals, with the possible addition of probabilities and certainty degrees. Proceedings of the IEEE World Congress on Computational Intelligence WCCI'2008, Hong Kong, China, June 1–6, 2008, pp. 983–990.

Yao YY, Li X (1996) Comparison of rough-set and interval-set models for uncertain reasoning. Fundamenta Informaticae 27:289–298

Appendix: 2nd Order is Sufficient for Fuzzy Uncertainty

Second order descriptions: the main idea. Experts are often not 100% certain in the statements they make; therefore, in the design of knowledge-based systems, it is desirable to take this uncertainty into consideration. Usually, this uncertainty is described by a number from the interval $[0,1]$; this number is called *subjective probability*, *degree of certainty*, etc.

One of the main problems with this approach is that we must use *exact* numbers from the interval $[0,1]$ to represent experts' degrees of certainty; an expert may be able to tell whether his degree of certainty is closer to 0.9 or to 0.5, but it is hardly possible that an expert would be able to meaningfully distinguish between degrees of certainty, say, 0.7 and 0.701. If you ask the expert whether his degree of certainty about a certain statement A can be described by a certain number d (e.g., $d = 0.701$), the expert will, sometimes, not be able to give a definite answer, she will be uncertain about it. This uncertainty can be, in its turn, described by a number from the interval $[0,1]$. It is, therefore, natural to represent our degree of certainty in a statement A not by a *single* (crisp) number $d(A) \in [0,1]$ (as in the $[0,1]$-based description), but rather by a *function* $\mu_{\mathbf{d}(A)}$ which assigns, to each possible real number $d \in [0,1]$, a *degree* $\mu_{\mathbf{d}(A)}(d)$ with which this number d can be the (desired) degree of certainty of A. This is called a *second-order* description of uncertainty.

Third and higher order descriptions. In second-order description, to describe a degree with which a given number $d \in [0,1]$ can be a degree of certainty of a statement A, we use a *real number* $\mu_{\mathbf{d}(A)}(d)$. As we have already mentioned, it is difficult to describe our degree of certainty by a single number. Therefore, to make this description even more realistic, we can represent each degree of certainty $d(P(x))$ not by a (more traditional) $[0,1]$-based description, but by a *second order* description. As a result, we get the *third order* description.

Similarly, to make our description even more realistic, we can use the third order descriptions to describe degrees of certainty; then, we get *fourth order* uncertainty, etc.

Third order descriptions are not used: why? Theoretically, we can define third, fourth order, etc., descriptions, but in practical applications, only second order descriptions were used so far (see, e.g., (Nguyen Kreinovich 1995), (Nguyen et al. 1997), (Mendel 2001), (Mendel et al. 2010)). Based on this empirical fact, it is natural to conclude that third and higher order descriptions are not really necessary. We will show that this conclusion can be theoretically justified.

First step in describing uncertainty: set of uncertainty-describing words. Let us first describe the problem formally. An expert uses words from a natural language to describe his degrees of certainty. In every language, there are only finitely many words, so we have a finite set of words that needs to be interpreted. We will denote this set of words by W.

Second step: a fuzzy property described by a word-valued "membership function". If we have any property P on a universe of discourse U, an expert can describe, for each element $x \in U$, his degree of certainty $d(x) \in W$ that the element x has the property P.

Traditional fuzzy logic as a first approximation: numbers assigned to words describing uncertainty. Our ultimate goal is to provide a computer representation for each word $w \in W$. In the traditional $[0,1]$-based description, this computer representation assigns, to every word, a *real number* from the interval $[0,1]$; in general, we may have some other computer representations (examples will be given later). Let us denote the set of all possible computer representations by S.

In the first approximation, i.e., in the first order description, we represent each word $w \in W$, which describes a degree of uncertainty, by an element $s \in S$ (e.g., by a real number from the interval $[0,1]$). In this section, we will denote this first-approximation computer representation of a word w by $s = \|w\|$.

If the set S is too small, then it may not contain enough elements to distinguish between different expert's degree of belief: this was exactly the problem with classical $\{0,1\}$-based description, in which we only have two possible computer representations – "true" and "false" – that are not enough to adequately describe the different degrees of certainty. We will therefore assume that the set S is rich enough to represent different degrees of certainty.

In particular, the set $[0,1]$ contains infinitely many points, so it should be sufficient; even if we only consider computer-representable real numbers, there are still much more of them (millions and billions) than words in a language (which is usually in hundreds of thousands at most), so we can safely make this "richness" assumption. In mathematical terms, it means that two different degrees of belief are represented by different computer terms, i.e., that if $w_1 \neq w_2$, then $\|w_1\| \neq \|w_2\|$.

First approximation is not absolutely adequate. The problem with the first-order representation is that the relation between words $w \in W$ and computer representation $s \in S$ is, in reality, also imprecise. Typically, when we have a word $w \in W$, we cannot pick a single corresponding representative $s \in S$; instead, we may have *several* possible representatives, with different degrees of adequacy.

Actual description of expert uncertainty: word-valued degree to which a word describes uncertainty. In other words, instead of a *single* value $s = \|w\|$ assigned to a word w, we have *several* values $s \in S$, each with its own degree of adequacy; this degree of adequacy can also be described by an expert, who uses an appropriate word $w \in W$ from the natural language.

In other words, for every word $w \in W$ and for ever representation $s \in S$, we have a degree $w' \in W$ describing to what extent s is adequate in representing w. Let us

represent this degree of adequacy by $a(w,s)$; the symbol a represents a function $a : W \times S \to W$, i.e., a function that maps every pair (w,s) into a new word $a(w,s)$.

Second-order description of uncertainty as a second approximation to actual uncertainty. So, the meaning of a word $w \in W$ is represented by a *function* a which assigns, to every element $s \in S$, a degree of adequacy $a(w,s) \in W$. We want to represent this degree of adequacy in a computer; therefore, instead of using the word $a(w,s)$ itself, we will use the computer representation $\|a(w,s)\|$ of this word. Hence, we get a *second-order* representation, in which a degree of certainty corresponding to a word $w \in W$ is represented not by a *single* element $\|w\| \in S$, but by a *function* $\mu_w : S \to S$, a function which is defined as $\mu_w(s) = \|a(w,s)\|$.

Second-order description is not 100% adequate either; third-, fourth-order descriptions, etc. The second-order representation is also not absolutely adequate, because, to represent the degree $a(w,s)$, we used a single number $\|a(w,s)\|$. To get a more adequate representation, instead of this single value, we can use, for each element $s' \in S$, a degree of adequacy with which the element s' represents the word $a(w,s)$. This degree of adequacy is also a word $a(a(w,s),s')$, so we can represent it by an appropriate element $\|a(a(w,s),s')\|$. Thus, we get a *third-order* representation, in which to every element s, we assign a second-order representation. To get an even more adequate representation, we can use fourth- and higher order representations.

Let us express this scheme formally.

Definition 1.

- *Let W be a finite set; element of this set will be called* words.
- *Let U be set called a* universe of discourse. *By a* fuzzy property *P, we mean a mapping which maps each element $x \in U$ into a word $P(x) \in W$; we say that this word* described the degree of certainty *that x satisfies the property P.*
- *By a* first-approximation uncertainty representation, *we mean a pair $\langle S, \|.\| \rangle$, where:*
  - *S is a set; elements of this set will be called computer representations; and*
  - *$\|.\|$ is a function from W to S; we say that an element $\|w\| \in S$* represents *the word w.*
- *We say that an uncertainty representation is* sufficiently rich *if for every two words $w_1, w_2 \in W$, $w_1 \neq w_2$ implies $\|w_1\| \neq \|w_2\|$.*

Definition 2. *Let W be a set of words, and let S be a set of computer representations. By an* adequacy *function, we mean a function $a : W \times S \to W$; for each word $w \in W$, and for each representation $s \in S$, we say that $a(w,s)$ describes the degree to which the element s* adequately describes *the word w.*

Definition 3. *Let U be a universe of discourse, and let S be a set of computer representations. For each $n = 1, 2, \ldots$, we define the notions of n-*th order degree of certainty *and of a n-th order fuzzy set, by the following induction over n:*

- *By a* first-order degree of certainty, *we mean an element $s \in S$ (i.e., the set S_1 of all first-order degrees of certainty is exactly S).*

- *For every n, by a n*-th order fuzzy set*, we mean a function $\mu : U \to S_n$ from the universe of discourse U to the set S_n of all n-th order degrees of certainty.*
- *For every $n > 1$, by a n*-th order degree of certainty*, we mean a function s_n which maps every value $s \in S$ into an $(n-1)$-th order degree of certainty (i.e., a function $s_n : S \to S_{n-1}$).*

Definition 4. *Let W be a set of words, let $\langle S, \|.\| \rangle$ be an uncertainty representation, and let a be an adequacy function. For every $n > 1$, and for every word $w \in W$, we define the n*-th order degree of uncertainty $\|w\|_{a,n} \in S_n$ corresponding to the word w as follows:*

- *As a first order degree of uncertainty $\|w\|_{a,1}$ corresponding to the word w, we simply take $\|w\|_{a,1} = \|w\|$.*
- *If we have already defined degrees of orders $1, \ldots, n-1$, then, as an n-th order degree of uncertainty $\|w\|_{a,n} \in S_n$ corresponding to the word w, we take a function s_n which maps every value $s \in S$ into a $(n-1)$-th order degree $\|a(w,s)\|_{a,n-1}$.*

Definition 5. *Let W be a set of words, let $\langle S, \|.\| \rangle$ be an uncertainty representation, let a be an adequacy function, and let P be a fuzzy property on a universe of discourse P. Then, by a n-th order fuzzy set (or a n*-th order membership function*) $\mu_{P,a}^{(n)}(x)$ corresponding to P, we mean a function which maps every value $x \in U$ into an n-th order degree of certainty $\|P(x)\|_{a,n}$ which corresponds to the word $P(x) \in W$.*

We will prove that for properties which are *non-degenerate* in some reasonable sense, it is sufficient to know the *first* and *second* order membership functions, and then the others can be uniquely reconstructed. Moreover, if we know the membership functions of first two orders for a non-degenerate *class* of fuzzy properties, then we will be able to reconstruct the higher order membership functions for *all* fuzzy properties from this class.

Definition 6.

- *We say that a fuzzy property P on a universe of discourse U is* non-degenerate *if for every $w \in W$, there exists an element $x \in U$ for which $P(x) = w$.*
- *We say that a class $\mathscr{P}$ of fuzzy properties P on a universe of discourse U is* non-degenerate *if for every $w \in W$, there exists a property $P \in \mathscr{P}$ and an element $x \in U$ for which $P(x) = w$.*

Comment. For example, if $W \neq \{0,1\}$, then every crisp property, i.e., every property for which $P(x) \in \{0,1\}$ for all x, is *not* non-degenerate (i.e., degenerate).

Proposition 1. *Let W be a set of words, let $\langle S, \|.\| \rangle$ be a sufficiently rich uncertainty representation, let U be a universe of discourse. Let P and P' be fuzzy properties, so that P is non-degenerate, and let a and a' be adequacy functions. Then, from $\mu_{P,a}^{(1)} = \mu_{P',a'}^{(1)}$ and $\mu_{P,a}^{(2)} = \mu_{P',a'}^{(2)}$, we can conclude that $\mu_{P,a}^{(n)} = \mu_{P',a'}^{(n)}$ for all n.*

Comments.

- In other words, under reasonable assumptions, for each property, the information contained in the first and second order fuzzy sets is sufficient to reconstruct all higher order fuzzy sets as well; therefore, in a computer representation, it is sufficient to keep only first and second order fuzzy sets.
- This result is somewhat similar to the well-known result that a Gaussian distribution can be uniquely determined by its moments of first and second orders, and all higher order moments can be uniquely reconstructed from the moments of the first two orders.
- It is possible to show that the non-degeneracy condition is needed, because if a property P is not non-degenerate, then there exist adequacy functions $a \neq a'$ for which $\mu^{(1)}_{P,a} = \mu^{(1)}_{P,a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{P,a'}$, but $\mu^{(3)}_{P,a} \neq \mu^{(3)}_{P,a'}$ already for $n = 3$.

Proposition 2. *Let W be a set of words, let $\langle S, \|.\| \rangle$ be a sufficiently rich uncertainty representation, let U be a universe of discourse. Let $\mathscr{P}$ and $\mathscr{P}'$ be classes of fuzzy properties, so that the class $\mathscr{P}$ is non-degenerate, and let $\varphi : \mathscr{P} \to \mathscr{P}'$ be a 1-1-transformation, and let a and a' be adequacy functions. Then, if for every $P \in \mathscr{P}$, we have $\mu^{(1)}_{P,a} = \mu^{(1)}_{\varphi(P),a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{\varphi(P),a'}$, we can conclude that $\mu^{(n)}_{P,a} = \mu^{(n)}_{\varphi(P),a'}$ for all n.*

Comment. So, even if we do not know the adequacy function (and we do not know the corresponding fuzzy properties $P \in \mathscr{P}$), we can still uniquely reconstruct fuzzy sets of all orders which correspond to all fuzzy properties P.

Proof of Propositions 1 and 2. Proposition 1 can be viewed as a particular case of Proposition 2, when $\mathscr{P} = \{P\}$, $\mathscr{P}' = \{P'\}$, and φ maps P onto P'. Therefore, to prove both Propositions 1 and 2, it is sufficient to prove Proposition 2.

We will show that under the conditions of Proposition 2, from $\mu^{(1)}_{P,a} = \mu^{(1)}_{\varphi(P),a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{\varphi(P),a'}$, we will be able to conclude that $\varphi(P) = P$ for all $P \in \mathscr{P}$, and that $a = a'$; therefore, we will easily conclude that $\mu^{(n)}_{P,a} = \mu^{(n)}_{\varphi(P),a'}$ for all n.

Indeed, by definition of the first membership function, for every $x \in U$, we have $\mu^{(1)}_{P,a}(x) = \|P(x)\|$. Thus, from the equality $\mu^{(1)}_{P,a} = \mu^{(1)}_{\varphi(P),a'}$, we conclude that for every $P \in \mathscr{P}$, we have $\|P(x)\| = \|\varphi(P)(x)\|$ for all $x \in U$. Since the uncertainty representation is assumed to be sufficiently rich, we can conclude that $\varphi(P)(x) = P(x)$ for all $x \in U$, i.e., that $\varphi(P) = P$ for every $P \in \mathscr{P}$.

Let us now show that $a = a'$, i.e., that for every $w \in W$ and for every $s \in S$, we have $a(w,s) = a'(w,s)$. Indeed, since $\mathscr{P}$ is a non-degenerate class, there exists a value $x \in U$ and a property $P \in \mathscr{P}$ for which $P(x) = w$. Let us consider the equality of the second order membership functions for this very P. Since $\varphi(P) = P$, the given equality $\mu^{(2)}_{P,a} = \mu^{(2)}_{\varphi(P),a'}$ can be simplified into the following form: $\mu^{(2)}_{P,a} = \mu^{(2)}_{P,a'}$. Let us consider this equality for the above-chosen value x (for which $P(x) = w$). For this x, by definition of the second-order membership function, $\mu^{(2)}_{P,a}(x) = \|P(x)\|_{a,2} = \|w\|_{a,2}$; and similarly, $\mu^{(2)}_{P,a'}(x) = \|P(x)\|_{a,2} = \|w\|_{a',2}$; thus, $\|w\|_{a,2} = \|w\|_{a',2}$.

By definition, $\|w\|_{a,2}$ is a function which maps every value $s \in S$ into a 1-st order degree $\|a(w,s)\|_{a,1} = \|a(w,s)\|$. Thus, from the equality of the functions $\|w\|_{a,2}$

and $\|w\|_{a',2}$, we can conclude that their values at a given s are also equal, i.e., that $\|a(w,s)\| = \|a'(w,s)\|$. Since the uncertainty structure is sufficiently rich, we conclude that $a(w,s) = a'(w,s)$. The proposition is proven.

Proof of a comment after Proposition 1. Since P is *not* non-degenerate, there exists a value $w_0 \in W$ which cannot be represented as $P(x)$ for any $x \in U$. Let us pick arbitrary elements $x_0 \in U$ and $s_0 \in S$, and define $a(w,s)$ and $a'(w,s)$ as follows:

- first, we define $a(w,s) = a'(w,s)$ for all words w of the type $w = P(x)$: namely, we take $a(P(x_0),s_0) = a'(P(x_0),s_0) = w_0$ and take arbitrary other values for different pairs (w,s) with $w = P(x)$;
- then, we define $a(w,s)$ and $a'(w,s)$ for the remaining pairs (w,s): namely, we take $a(w_0,s_0) = w_0$, $a'(w_0,s_0) = P(x_0) \neq w_0$, and we define a and a' arbitrarily for all other pairs (w,s).

Let us show that for thus chosen adequacy functions, the membership functions of first and second order coincide, but the membership functions of the third order differ. Indeed:

- For the *first* order, we have, for every x, $\mu^{(1)}_{P,a}(x) = \|P(x)\|$ and similarly, $\mu^{(1)}_{P,a'}(x) = \|P(x)\|$; therefore, $\mu^{(1)}_{P,a}(x) = \mu^{(1)}_{P,a'}(x)$ for all x. Hence, $\mu^{(1)}_{P,a} = \mu^{(1)}_{P,a'}$.
- For the *second* order, for every x, $\mu^{(2)}_{P,a}(x)$ is a function which maps $s \in S$ into a value $\|a(P(x),s)\|_{a,1} = \|a(P(x),s)\|$. Similarly, $\mu^{(2)}_{P,a'}(x)$ is a function which maps $s \in S$ into a value $\|a'(P(x),s)\|_{a',1} = \|a'(P(x),s)\|$. For words w of the type $P(x)$, we have defined a and a' in such a way that $a(w,s) = a'(w,s)$; therefore, $\|a(P(x),s)\| = \|a'(P(x),s)\|$ for all x and s. Thus, $\mu^{(2)}_{P,a} = \mu^{(2)}_{P,a'}$.
- Finally, let us show that the *third* order membership functions differ. We will show that the values of the functions $\mu^{(3)}_{P,a}$ and $\mu^{(3)}_{P,a'}$ differ for $x = x_0$. Indeed, by definition of the third order membership function,
  - $\mu^{(3)}_{P,a}(x_0)$ is a function which maps every s into the value $\|a(P(x_0),s)\|_{a,2}$, and
  - $\mu^{(3)}_{P,a'}(x_0)$ is a function which maps every s into the value $\|a'(P(x_0),s)\|_{a',2}$.

  To prove that these function are different, it is sufficient to show that their values differ for *some* values s; we will show that they differ for $s = s_0$, i.e., that $\|a(P(x_0),s_0)\|_{a,2} \neq \|a'(P(x_0),s_0)\|_{a',2}$. By our construction of a, we have $a(P(x_0),s_0) = a'(P(x_0),s_0) = w_0$, so the inequality that we need to prove takes the form $\|w_0\|_{a,2} \neq \|w_0\|_{a',2}$.
  By definition, $\|w_0\|_{a,2}$ is a function which maps every value $s \in S$ into $\|a(w_0,s)\|_{a,1} = \|a(w_0,s)\|$. Similarly, $\|w_0\|_{a',2}$ is a function which maps every value $s \in S$ into $\|a'(w_0,s)\|_{a,1} = \|a'(w_0,s)\|$. For s_0, according to our construction of a and a', we have $a(w_0,s_0) = w_0 \neq P(x_0) = a'(w_0,s_0)$. Thus, since the uncertainty representation is sufficiently rich, we conclude that $\|a(w_0,s_0)\| \neq \|a'(w_0,s_0)\|$, and therefore, that $\|w_0\|_{a,2} \neq \|w_0\|_{a',2}$ and $\mu^{(3)}_{P,a} = \mu^{(3)}_{P,a'}$.

The statement is proven.