The final version was published by Springer: Kovalerchuk, Relationships Between Probability and Possibility Theories, In: Uncertainty Modeling, Studies in Computational Intelligence 683, Kreinovich, V.(Ed.), Springer, 2017, 97-122, DOI 10.1007/978-3-319-51052-1_7

Relationships between Probability and Possibility Theories

Boris Kovalerchuk

Dept. of Computer Science, Central Washington University,
400 E. University Way, Ellensburg, WA 98926, USA
borisk@cwu.edu

Abstract. The goal of a new area of Computing with Words (CWW) is solving computationally tasks formulated in a natural language (NL). The extreme uncertainty of NL is the major challenge to meet this ambitious goal requiring computational approaches to handle NL uncertainties. Attempts to solve various CWW tasks lead to the methodological questions about rigorous and heuristic formulations and solutions of the CWW tasks. These attempts immediately reincarnated the long-time discussion about different methodologies to model uncertainty, namely: Probability Theory, Multi-valued logic, Fuzzy Sets, Fuzzy Logic, and the Possibility theory. The main forum of the recent discussion was an on-line Berkeley Initiative on Soft Computing group in 2014. Zadeh claims that some computing with words tasks are in the province of the fuzzy logic and possibility theory, and probabilistic solutions are not appropriate for these tasks. In this work we propose a useful constructive probabilistic approach for CWW based on sets of specialized K-Measure (Kolmogorov's measure) probability spaces that differs from the random sets. This work clarifies the relationships between probability and possibility theories in the context of CWW.

1 Introduction

Computing with Words (CWW) is a new area [Zadeh, 1999, 2012] that intends to solve computationally the tasks formulated in a natural language (NL). Examples of these tasks are: "What is the sum of (about 5) + (about 10)?" and "What is the possibility or probability that Mary is middle-aged, if she is 43 and is married for about 15 years?"

In these problems we are given *information I* in the form of membership functions (MFs) such as $\mu_{about\text{-}5}$, $\mu_{about\text{-}15\text{-}years}$, μ_{young}, $\mu_{middle\text{-}aged}$, and the respective probability density function (pdfs), e.g., age pdf p_{age}. Similarly for other tasks information *I* can include a membership function μ_{tall}(Height(X)), where X is a person (e.g., John) and a *probability density function* of height p_H(Height(John)) in the height interval $U=[u_1,u_n]$, e.g., u_1 = 160 cm and u_n = 200 cm. The probability space for P_H contains a set of elementary events e_i = "Height of John is u_i": e_1 = "Height of John is u_1", e_2 = "Height of John is u_2",…, e_n = "Height of John is u_n".

Attempts to solve various CWW tasks lead to the methodological questions about rigorous and heuristic formulations and solutions of the CWW tasks which have a tremendous level of uncertainty in both formulations and solutions [Belyakov et al., 2012; Kovalerchuk, 2012-2015].

These attempts immediately reincarnated the long-time discussion about different methodologies to model uncertainty [Zadeh, 1996], namely: probability theory (PrT), multi-valued logic (MVL), Fuzzy Sets (FS), Fuzzy Logic (FL), and the Possibility theory. The main forum of the recent discussion was an on-line Berkeley Initiative on Soft Computing (BISC) group in 2014.

Zadeh claims that some computing with words tasks are in the province of the fuzzy logic, and probabilistic solutions are not appropriate for these tasks. "Representation of grade of membership of probability has no validity" [Zadeh, BISC, 03.07.2014].

He uses this clam for producing a much wider claim of fundamental limitation of the probability theory for computing with words. "Existing logical systems--other than fuzzy logic--cannot reason and compute with information which is described in natural language [Zadeh BISC, 05/27/2014].

In this paper, we show that this and similar claims may be true if we only consider simple probabilistic models and techniques, like the ones used in routine engineering applications. However, modern probability theory contains more sophisticated concepts and ideas that go way beyond these techniques. We show that, by using such more sophisticated concepts and ideas, we can come up with a probabilistic interpretation of fuzzy sets.

To be more precise, such interpretations have been proposed in the past, e.g., the interpretation using random sets [Orlov, 1978; Goodman, 1982; Nguyen, 2005; Nguyen, Kreinovich, 2014]. Unfortunately, the random sets interpretation is quite complicated, requires mathematical sophistication from users, and often a lot of data that is not always realistic in many practical applications to provide an efficient computational tool. In contrast, our interpretation that uses a set of specialized Kolmogorov-type probability measures (K-measures, for short), is intuitively understandable and computationally efficient.

In this discussion Zadeh did not specify his term "no validity". This led to the difficulty for an independent observer to test his opinion. To move from an opinion level to a scientific level we need the independently verifiable *criteria* to test validity of an approach that we propose below for the probabilistic solutions.

Any solution of the task to be a *valid probabilistic solution* must meet criteria of the Axiomatic Probability Theory (APT):

(1) The K-measure spaces (probability spaces) must be created.

(2) All computations must be in accordance with the APT.

In the course of BISC discussion and as result of it we proposed probabilistic formulations and solutions that met criteria (1) and (2) for several CWW test tasks listed below:

1) "Tall John" (Zadeh's task) [Kovalerchuk, 2014],

2) "Hiring Vera, Jane, and Mary" that is based on their age and height (Zadeh's task), [Kovalerchuk, BISC, 03.07.2014],

3) "Degree of circularity of the shape" [Kovalerchuk, BISC, 03/21/2014],

4) "Probably Very is middle-aged" [Kovalerchuk, 2014, BISC 01/17, 2014],

5) "Sum of two fuzzy sets" (Zadeh's task) [Kovalerchuk, 2015].

The proposed K-measure based solutions are at least of the same usefulness as a solution with the fuzzy logic min/max operation. In [Kovalerchuk, 2015] we proposed several alternative explications of task 5 and their valid solutions within APT.

This work focuses on three problems:

P1: Establishing relations between concepts of possibility and probability;

P2: Establishing relations between grades of membership of a fuzzy set and probabilities;

P3: Outlining future rigorous possibility theory.

The paper is organized as follows: problem P1 is considered in Section 2, problem P2 is considered in Section 3, problem P3 is considered in Section 4, and Section 5 provides a conclusion.

Within problem P1 we discuss the following topics:

- Possibilistic Semantics,
- Relationship between possibility and probability in natural language (NL),
- Is NL possibility easier than NL probability for people to use?
- Relation between NL words “possibility” and “probability” and the formal math probability concept,
- NL possibility vs. modal and non-modal probability,
- Can the sum of possibilities be greater than 1?
- Context of probability spaces for modeling possibility.

Within problem P2 we discuss:

- Are fuzzy sets and possibility distributions equivalent?
- Can we model degrees of membership probabilistically by Kolmogorov’s measures spaces (K-spaces)?
- What is a more general concept fuzzy set or probability?
- Is interpreting grade of membership as probability meaningful?
- Relationship among unsharpness, randomness and axiomatic probability?

Within problem P3 we discuss:

- Possibility as upper probability,
- Should evaluation structure be limited by [0,1] or lattice?
- Exact numbers vs. intervals to evaluate uncertainty.

2 Relations between concepts of possibility and probability

2.1 Possibilistic Semantics

The literature on the Possibility Theory is quite extensive [Zadeh, 1978; Dubois, 1084, 2006; Dubois, Nguyen, Prade, 2000; Joslyn, 1995; Yager, 1980, 1993; Giles, 1982; Cooman, Kerre, Vanmassenhov, 1992; Gaines, 1984; Nguyen, 2005; Nguyen, Kreinovich, 2014]. These studies mostly concentrate on conceptual developments.

The important aspects of empirical justification of these developments via psychological and linguistic experiments are still in a very earlier stage. As a result it is difficult to judge whether the current possibility concepts are descriptive or prescriptive. A descriptive theory should reproduce the actual use of the concept of possibility in NL. A prescriptive theory should set up rules for how we should reason about possibilities. Both theories require identification of the semantics of the concept of possibility.

Below we outline semantics of possibility in the natural language based on review in [Joslyn, 1995]:

- An object/event A that did not yet occur (have not yet been seen) can be a possible event, $Pos(A) \geq 0$.
- The object/event A is possible (not prohibited) if $0 < Pos(A) \leq 1$.
- The possibility of the observed object A is a unitary possibility, $Pos(A)=1$.
- If to characterize the object/event A somebody selects the statement "A is possible" over a statement "A is probable" then the chances for A to occur are lower than if the second statement would be selected. This is consistent with the common English phrases "X is possible, but not probable.", i.e., $Pos(X) \geq Pr(X)$.
- Types of possible events:

  Type 1: *Merely possible events* are events that may not occur (earthquake) while occurred in the past, Pos(A) is unknown.

  Type 2: *Eventual events* are events that are guaranteed to occur sooner or later (rain), $Pos(A)=1$

In the section 4.2 we discuss possible deviations from this semantics for some complex situations.

Other aspects of possibilistic semantics are discussed in [Dubois, 2006] with four meanings of the word "*possibility*" identified:

- *Feasibility*, *ease* of achievement, the solution of a problem, satisfying some constraints with linguistic expressions such as "it is possible to solve this problem".
- *Plausibility*, the propensity of events to occur with linguistic expressions such as "it is possible that the train arrives on time".
- *Epistemic*, logical consistency with available information, a proposition is possible, meaning that it does not *contradict* this information. It is an all-or-nothing version of plausibility.
- *Deontic*, possible means *allowed,* permitted by the law.

The expressions below illustrate the semantic difference between NL concepts of possibility and probability where it seems that the expressions with word "probable" are equivalent, but the expressions with word "possible" are not equivalent:

It is *not probable* that "not A" vs. It is probable that A
It is *not possible* that "not A" vs. It is possible that A.

2.2. Is easiness an argument for possibility theory?

Below we analyze arguments for and against the possibility theory based on [Kovalerchuk, 02/03/2014]. Zadeh [BISC, 01/30/2014] argued for the possibility theory by stating: "*Humans find it difficult to estimate probabilities. Humans find it easy to estimate possibilities*." To support this claim he presents "Vera's task" that we reconstruct below from his sketchy description.

Vera is middle-aged and middle-aged can be defined as a probability distribution or a possibility distribution. Age is discrete. It is assumed that neither probability nor possibility distributions are given and a person must use his/her judgment to identify them by answering questions for specific ages, e.g.,

> (Q1)What is your estimate of the *probability* that Vera is 43?
> (Q2) What is your estimate of the *possibility* that Vera is 43?

Zadeh stated that if he would be asked Q1 he "could not come up with an answer. Nor would *anyone* else be able to come up with an answer based on one's perception of middle-aged."

Accordingly answering Q2 must be easier than answering Q1. In fact he equates *possibility* with a *grade of membership*. "Possibility is *numerically equal to grade of membership*. 0.7 is the degree to which 43 fits my perception of middle-aged, that is, it is the grade of membership of 43 in middle-aged." Thus, he is answering question Q3:

> (Q3) What is your estimate of the *grade of membership* that middle-aged Vera is 43?

This question is a result of *interpretation* of possibility as a grade of membership. In other words, Zadeh's statement is not about easiness of possibility vs. probability for humans, but about *ability or inability to interpret grades of membership as probabilities or possibilities*. We had shown the ability to interpret grades of membership as probabilities in [Kovalerchuk, 2014, BISC 01/17, 2014] for Zadeh's examples "Mary is middle-aged" and "John is tall". It is done by reinterpreting probabilistically Zadeh's own solution [Zadeh, 2011] using a set of K-Measure spaces outlined in this paper in section 3.2. For the current example "Vera's age" the reinterpretation is the same. Thus to test easiness of possibility vs. probability for humans other arguments are needed.

2.3. Words possibility and probability vs. formal math probability

In the BISC discussion Zadeh made several statements:

- "Probabilistic information is *not derivable* from possibilistic information and vice versa" [Zadeh, BISC, 04/09/2014].

- “A basic reason is that count and boundary are distinct, *underivable* concepts.” [Zadeh, BISC, 10/27/2014].
- “0.7 is the possibility that Vera is 43 and has *no relation to the probability* of middle-aged” [Zadeh, BISC, 04/09/2014].
- “The possibility that Hans can eat n eggs for breakfast is *independent* of the probability that Hans eats n eggs for breakfast, with the understanding that if the possibility is zero then so is the probability” [Zadeh , BISC, 10/27/2014].

Below we show deficiencies of the first three claims and explain independence in the last claim in a way that is consistent with probabilistic interpretation of the concept of possibility. Zadeh bases these statements on claims that, in large measure, the probability theory is “*count*-oriented”, fuzzy logic is “fuzzy-*boundary*-oriented”, and the possibility theory is “*boundary*-oriented”.

Making these statements, Zadeh references Chang's claim [Chang, 2014] that possibility relates to: “Can it happen?” and probability relates to: “How often?” Piegat [04/10/2014] commented on these statements, noting that Zadeh has accepted for his claims the *frequency interpretation* of probability, while there five or more interpretations of probability [Burdzy, 2009] and in the same way likely multiple interpretations of possibility can be offered depending on some features such as the number of alternatives.

Below we list four well-known different meanings of probability and different people are not equal in using these meanings:

a) probability as it is known in a *natural language* (NL) without any relation to any type of probability theory,

b) probability as it is known in the *frequency-based* probability theory,

c) probability as it is known in the *subjective* probability theory [Wright, Ayton, 1994],

d) probability as it is known in the formal *axiomatic* mathematical probability theory (Kolmogorov’s theory).

We show deficiencies of Zadeh’s statement for meanings (b)-(d) in this section and for meaning (a) later in this paper. It is sufficient to give just a single example of relations for each (b)-(d) that would interpret the possibility that Vera is 43 as a respective probability (b)-(d) of middle-aged.

Consider two events e_1={Vera is middle-aged} and e_2={Vera is not middle-aged}. In the formal mathematical theory of probability, events are objects of any nature. Therefore, we can interpret each of these two events simply as a sequence of words or symbols/labels, e.g., we can assign the number 0.7 to e_1 and the number 0.3 to e_2. In this very formal way we fully satisfy the requirement of (d) for 0.7 and 0.3 to be called probabilities in the Kolmogorov’s axiomatic probability theory [Kolmogorov, 1956]. In addition meanings (b) and (c) allow getting 0.7 and 0.3 meaningfully by asking people questions similar to questions like "What is your subjective belief about event e?" An interpretation (b) is demonstrated in depth by Jozo Dujmovic [2017] in this volume.

To counter Zadeh's claim about the example with n eggs it is sufficient to define two probabilities:

p(Hans *eats* n eggs for breakfast) and p(Hans *can eat* n eggs for breakfast).

The last probability is, in fact, the possibility that Hans eats n eggs for breakfast as the Webster dictionary defines the word "possibility" and we discuss in the next section in details. The idea of that probability (probability of a modal statement with the word "can") was suggested by Cheeseman a long time ago [1986]. Thus, we simply have two different probabilities, because we have different sets of elementary events {eat, not eat} and {can eat, cannot eat}. Nobody claims and expects that two probabilities from two different probability spaces must be equal. In the same way, we have two different probabilities, when we evaluate probabilities of rain with two sets of elementary events {daytime rain, no daytime rain} and {nighttime rain, no nighttime rain}.

2.4. Possibility and probability in natural language

The Webster dictionary provides meanings of the words "possibility" and "probability" in the Natural Language (NL). Table 1 contrasts Webster NL definitions of these words. It shows that the major difference is in "will happen", "is happened" associated with word "probability" and "might happen" associated with word "possibility". In other words, we have different levels of chance to happen.

Table 1. Webster NL definitions of words possibility and probability

Possibility	*Probability*
a *chance* that something *might exist, happen, or be true*	the *chance* that something *will happen*
something that might be done or might happen	something that has a *chance of happening*
abilities or qualities that could make someone or something better in the future.	a measure of *how often* a particular event will happen if something (such as tossing a coin) is done repeatedly

The word "might" expresses a lower chance to happen. Thus even in the pure NL setting there is a strong relation between probability and possibility.

Respectively questions Q1-Q2 can be reformulated into *equivalent* NL forms:

(Q1.1) What is your estimate of the chance that Vera *is* 43?

(Q2.1) What is your estimate of the chance that Vera *might be* 43?

Note that question (Q3): "What is your estimate of the *grade of membership* that middle-aged Vera is 43?" contains term "*grade of membership*" which is a part of the *professional language* not the basic NL, i.e., it requires interpretation. In contrast Q1.1 and Q2.1 do not contain any professional words and phrases. Note that the Webster interpretation of possibility is fully consistent with Cheeseman's modal probability [1986] with the modal word “can” that we discussed above for the modal event “n people can be in the car”, P(n people can be in the car).

2.5. Experiment: Is possibility easier for people than probability in NL?

We conducted an experiment asking students to answer the question Q1.1- Q2.1 in the form shown below, i.e., to answer questions about probability and possibility as it follows from the Webster dictionary:

Please provide your personal opinion for the questions below in the following situation. It is known that Vera is middle-aged.

1. What is the chance that she **is 43**?

   Circle your answer: 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 X

   Circle X if it is difficult for you to assign any number.

2. What the chance that she **might be 43**?

   Circle your answer: 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 X

   Circle X if it is difficult for you to assign any number.

Table 2. Data

Person	*Q1.1*	*Q2.1*
1	0.1	0.3
2	0.1	1
3	0.1	1
1	0.2	0.3
2	0.2	1
3	0.3	0.5
4	0.3	0.6
5	0.3	0.9
6	0.3	1
7	0.3	1
8	0.5	0.8
9	0.5	1
10	0.7	0.8
11	0.7	0.9
12	0.8	1
13	0.8	0.6
14	1	0
15	1	0.4
16	0.4	0.1
17	1	1
18	0.4	0.4
19	0.3	0.3
20	0.1	X
21	X	0.3
22	X	1
23	X	0.3
24	X	1
25	X	X
26	X	X
27	X	X

Table 3. Statistical Analysis

Characteristic	*Q1.1*	*Q2.1*
average answer (all numeric answers)	0.45	0.67
std. dev. (all numeric answers)	0.3	0.34
number of all numeric answers	20	24
average answer for cases with answer Q2.1 > answer Q1.1	0.36	0.81
Std. dev. for cases with answer Q2.1 > answer Q1.1	0.23	0.26
number of answers where answer Q2.1 > answer Q1.1	12	12
average answer for cases with answer Q1.1 > answer Q2.1	0.63	0.49
std. dev. for cases with answer Q1.1 > answer Q2.11	0.37	0.36
number of answers where answer Q1.1 > answer Q2-1	4	4
average answer for cases with answer Q1.1 = answer Q2-1	0.57	0.57
std. dev. for cases with answer Q1.1 = answer Q2.1	0.38	0.38
number of answers with answer Q1.1= answer Q2.1	3	3
number of respondents who refused to answer both questions	3	3
number of respondents who refused to answer Q1.1 only	4	0
number of respondents who refused to answer Q2.1 only	0	1
Total number of respondents who refused to answer.	7	4

In these questions we deliberately avoided both words “probability” and “possibility”, but used the word “chance” in combinations with words “is” and “might be” to provide a meanings of these concepts which is derived from the Webster dictionary. It is important to distinguish wording from semantics of the concepts of “probability” and “possibility”. Questions must preserve semantics of these concepts, but can be worded differently. The word "chance" is more common in NL and is clearer to the most of the people. By asking questions with the word “chance” we preserve the semantics of concepts of “probability” and “possibility” and made the task for the respondents easier.

A total of 27 sophomore Computer Science university students answered these questions. Practically all of them are English native speakers and none of them took fuzzy logic or probability theory classes before. Answering time was not limited, but all answers were produced in a few minutes. Two strong students took an initiative and offered formulas and charts to compute the answers and generated answers based on them. Table 2 shows all answers of this experiment and table 3 shows the statistical analysis of these answers.

This experiment leads to the following conclusion:

1. Most of the students (80%) do not have any problem to answer both questions Q1.1 (probability) and Q2.1 (possibility).
2. Answers produced quickly (in less than 5 min).
3. Answers are consistent with expected higher value for Q2-1 than for Q1.1. The average answer for Q2.1 is 0.67 and the average answer for Q1.1 is 0.45.
4. Out of 27 students three students found difficult to answer both questions (11.1% of all respondents).
5. Out of 27 students four students found difficult to answer only question Q1.1 (14.8% of all respondents).
6. Out of 27 students one student found difficult to answer only question Q2.1 (3.7% of all respondents)

Thus only conclusion 5 can serve as a partial support for Zadeh’s claim that answering Q2.1 is impossible. However, it is applicable only to less than 15% of participants. The other 85% of people have the same easiness/difficulty to answer both Q1-1 and Q2-1.

These results seem to indicate that the probabilistic approach to CWW is at least as good as the possibilistic approach for most of the people who participated in this experiment. The expansion of this experiment to involve more respondents is desirable to check the presented results. We strongly believe that experimental work is necessity in fuzzy logic to guide the production of meaningful formal methods and the scientific justification of existing formal methods, which is largely neglected.

2.6. Is sum of possibilities greater than 1?

Consider a probability space with two events e_{m1} and e_{m2}

e_{m1} = ”A *may* happen”, e_{m2} = ”A *may not* happen”,

Thus, for A = ”rain” we will have events

e_{m1} = ”rain *may* happen”, e_{m2} = ”rain *may not* happen”,

Respectively for A = ”no rain” we will have events

e_{m1} = “no rain *may* happen”, e_{m2} = “no rain *may not* happen”,

Note, that “dry weather” is not equal to “no rain” (negation of rain), because snow is another part of negation of rain in addition to dry weather.

Next consider another probability space with two events without “may” and “may not”,

e_1 = ”A”, e_2 = ”notA”

Denote probabilities in spaces with “may” and “may not” as P_m and without them as P_e, respectively. In both cases in accordance with the definition of probability spaces sums are equal to 1:

$$P_m(e_{m1})+P_m(e_{m2})=1, \quad P_e(e_1)+P_e(e_2)=1,$$

e.g., $P_m(e_{m1}) = 0.4$, $P_e(e_{m2}) = 0.6$, and $P_e(e_1) = 0.2$, $P_e(e_2) = 0.8$.

Note that in example $P_m(e_{m1}) + P_e(e_2)=0.4+0.8 = 1.2 > 1$. In fact, these probabilities are from different probability spaces and are not supposed to be summed up.

The confusion takes place in situations like presented below. Let

e_{m1} = “rain *may* happen” and e_2 = “no rain”

instead of

e_{m2} = “rain *may not* happen”.

Probability of e_2 = “no rain” is probability of a *physical event* (absence of a physical event). In contrast probability of e_{m1} = “rain *may not* happen” is probability of a *mental event*. It depends not only on physical chances that rain will happen, but also on the *mental interpretation* of the word “might” that is quite subjective.

This important difference between events e_2 and e_{m2} can be easily missed. It actually happens with a resulting claim that the possibility theory must be completely different from the probability theory based on such mixing events from different probability spaces and getting the sum of possibilities greater than 1.

2.7. Context of probability spaces

Consider a sum for possibilities of events A and not A: Pos(A)+Pos(not A) and a sum of probabilities Pr(A)+Pr(not A). For the last sum it is assumed in the Probability Theory that we have the *same context* for Pr(A) and Pr(not A) when it is defined that Pr(A)+Pr(not A)=1. Removing the requirement of the same context can make the last sum greater than 1. For instance, let Pr(A)=0.6 for A= "rain", which is computed using data for the last spring month, but Pr(not A) = 0.8 is computed using data for the last summer month. Thus, Pr(A)+Pr(not A) = 0.6+0.8 = 1.4 > 1.

The same *context shift* can happen for computing Pos(A)+Pos(not A). Moreover, for mental events captured by a possibility measure checking that the context is the same and not shifted is *extremely difficult*. Asking "What is the possibility of the rain?", "What is possibility that rain will not happen?", "What is the chance that rain may happen?" and "What is the chance that no rain may happen?" without controlling the context can easily produce the sum that will be greater or less than 1. Sums of possibility values from such "context-free" questions cannot serve as a justification for rejecting the probability theory and for introducing a new possibility theory with the sum greater than 1. The experiments should specifically control that *no context shift* happen.

2.8. Natural Language Possibility vs. modal and non-modal probability

Joslyn [1995] provided an example of the difference between probability and possibility in the ordinary natural language for a six-sided die. Below we reformulate it to show how this difference can be interpreted as a difference between *modal probability* [Cheeseman, 1986] and *non-modal probability*. A six-sided die has six *possible* outcomes (outcomes that *can occur*). It is applied to both balanced and imbalanced dies. In other words, we can say, each face is *completely possible*, possible with *possibility 1*, Pos(s) =1 or *can occur for sure*. In contrast, different faces of the imbalanced die *occur* with different probabilities $P(s_i)$.

Thus, for all sides s_i we have $Pos(s_i)=1$, but probability $P(s_i) < 1$. Respectively $\Sigma_{i=1:6}Pos(s_i)=6$ and $\Sigma_{i=1:6}P(s_i)=1$. At the first glance, it is a strong argument that possibility does not satisfy Kolmogorov's axioms of probability, and respectively possibility is not probability even from the formal mathematical viewpoint, not only as a NL concept.

In fact, $Pos(s_i)$ satisfies Kolmogorov's axioms, but considered not in a single probability space with 6 elementary events (die sides), but in 6 probability spaces with pairs of elementary events: $\{s_1, \text{not } s_1\}$, $\{s_2, \text{not } s_2\}$, …, $\{s_6, \text{not } s_6\}$ each.

Consider $\{s_1, \text{not } s_1\}$ with $Pos(s_1)$ as an answer for the question: "What is the probability that side s_1 *may occur/happen*?" and $Pos(\text{not } s_1)$ as an answer for the question: "What is the probability that side s_1 *may not occur/happen*?". The common sense answers are $Pos(s_1)=1$, and $Pos(\text{not } s_1)=0$ that are fully consistent with the Kolmogorov's axioms. This approach is a core of our approach with the *set of probability spaces* and probability distributions [Kovalerchuk, Klir, 1995; Kovalerchuk, 1996, 2013].

Consider another version of the same situation of mixing of probabilities and possibilities from different spaces as an incorrect way to justify superadditivity for both of them. It is stated in [Raufaste et al, 2003]: "…probabilistic relationship between p(A) and p(not A) is fully determined. By contrast, P(A) and P(not A) are weakly dependent in real life situations like medical diagnosis. For instance, given a particular piece of *evidence*, A can be *fully possible* and not A can be *somewhat possible* at the same time. Therefore, a "superadditivity" inequality stands: Pos(A) + Pos(not A) ≥ 1."

Let us analyze this medical diagnosis example for a pair of events

$$\{A, \text{not } A\} = \{e_1, e_2\} =$$
= {malignant tumor occurred for the patient, malignant tumor did not occur for the patient}

Then according to the probability theory we must have $P_e(e_1)+P_e(e_2)=1$ and it seems reasonable if these probabilities will be based on frequencies of the A and not A under the given evidence. Here in $\{e_1, e_2\}$ the physical entity (malignant tumor) is negated to get e_2 from e_1, and no modality is involved.

Next consider another pair of events (modal events with the words "might" and "might not"):

$\{e_{m1}, e_{m2}\}$= {malignant tumor might occur for the patient, malignant tumor might *not* occur for the patient}

Having $P_m(e_{m1})+P_m(e_{m2})=1$ is also seems reasonable for these events if these probabilities will be based on frequencies of subjective judgments of experts about e_{m1} and e_{m2} under given evidence. The experts will estimate the same physical entity "malignant tumor" for two different (opposite) modalities "can" and "cannot". In other words here we *negate modality* ("can") not a *physical entity* (malignant tumor) as was the case with $\{e_1, e_2\}$. The same property is expected for the events:

$\{e_{m3}, e_{m4}\}$ = {benign tumor might occur for the patient, benign tumor might not occur for the patient}

The pairs $\{e_{m1}, e_{m2}\}$ and $\{e_{m3}, e_{m4}\}$ differ from another pair that mix them:

$\{e_{m1}, e_{m3}\}$ = {malignant tumor might occur for the patient, *no* malignant tumor (benign tumor) *might* occur for the patient}

For this pair, $P_m(e_{m1})+P_m(e_{m3})=1$ seems less reasonable. Here in e_{m3} the *physical entity* (malignant tumor) is *negated*, but *modality* ("might") is the same. In essence, e_{m1} and e_{m3} are from different modal probability spaces, and should not be added, but operated with as it is done in the probability theory with probabilities from different probability spaces.

This example shows that probability spaces for modal entities must be built in a *specific way* that we identified as N1 below. Let A be a modal statement. It has two components: a physical entity E and a modal expression M, denote such A as A=(E,M). Respectively, there are three different not A:

(N1): not A = (E, not M)

(N2): not A = (not E, M)

(N3): not A = (not E, not M)

As we have seen above, it seems more reasonable to expect additivity for N1 than for N2 and N3. The next example illustrates this for N1 and N2. Consider A=(E,M) where E = "sunrise", not E = "sunset", M = "can", and not A=(E, not M) for N1, and not A = (not E, M) for N2.

For N1 the questions are:

"What is the probability that John *can* watch the sunrise tomorrow?" and
"What is the probability that John *cannot* watch the sunrise tomorrow?

For N2 these questions are:

"What is the probability that John can watch the *sunrise* tomorrow?" and
"What is the probability that John can watch the *sunset* tomorrow?

These questions in the possibilistic form can be:

"What is the possibility that John will watch the *sunrise* tomorrow?" and
"What is the possibility that John will watch the *sunset* tomorrow?"

Both probabilities/possibilities P(A) and P(not A) can reach 1 for N2, therefore building spaces with N1 and N3 should be avoided. The situation is the same as in the probability theory itself – not every set of events can be used as a set of meaningful elementary events.

Consider the next example, based on the following joke: "Can misfortune make a man a millionaire? Yes if he is a billionaire. In this example the first two pairs of questions are in N1, and the last pair is in N2:

What is the probability that misfortune *can* make a man a millionaire if he is a billionaire?

What is the probability that misfortune *cannot* make a man a millionaire if he is a billionaire?

What is the possibility that misfortune *will* make a man a millionaire if he is a billionaire?

What is the possibility that misfortune *will not* make a man a millionaire if he is a billionaire?

What is the possibility that misfortune will make a man a millionaire if he is a billionaire?

What is the possibility that negligence will make a man a millionaire if he is a billionaire?

Consider another example on possibility of n people in the car [Cheeseman, 1986]. Let n=10 and A is "10 people are in the car". What is the possibility of "10 people in the car"? Let Pos(A)=0.8 then "What is the possibility of not A?", i.e., possibility of "not 10 people in the car". We need to define a probability space {A, notA}. Note that "not 10 people in the car" includes 9 people. Thus

$$\text{Pos(notA)} \geq \text{Pos}(9),$$

Next it is logical to assume that Pos(9) > Pos(10). Thus, Pos(notA) ≥ 0.8 with

$$\text{Pos(A)} + \text{Pos(notA)} > 1.$$

As we noted above such "superadditivity" often is considered as an argument for the separate possibility theory with superadditivity.

In fact this cannot be such argument because there is a space with normal additivity (P(A)+p(notA)=1). In that space we have a different A. It is not an event that 10 people are in the car, but a modal statement, where A="10 people *can be* in the car", and respectively not A="10 people *cannot be* in the car", e.g., with P(A)=0.8 and P(notA)=0.2.

Thus in general superadditivity situations N2 and N3 can be modeled with a Set (pair) of probability (K-spaces) Spaces (SKS) of N1 type without any superadditivity. This is in line with our approach [Kovalerchuk, 1996] discussed above.

Note that superadditivity can really make sense in the probability theory and be justified, but very differently, and not as an argument for the separate possibility theory. See a chapter by Resconi and Kovalerchuk in this volume.

3. Relations between grades of membership of fuzzy set and probabilities

3.1. What is more general concept fuzzy set, probability or possibility?

Joslyn stated [1995] that neither is fuzzy theory specially related to possibility theory, nor is possibility theory specifically related to fuzzy sets. From his viewpoint "both probability distributions and possibility distributions are *special cases of fuzzy sets*". It is based on the fact an arbitrary fuzzy set can be specialized to be probability or possibility by imposing additional properties. For probability this property is that the sum (integral) or all values must be equal to 1. For possibility this property is that the maximum must be equal to 1.

The specialization of an arbitrary fuzzy set to probability or possibility can be done as follows. Let $m(x)$ be a membership function of a fuzzy set on the interval [a,b]. We can normalize values of $m(x)$ by dividing them by the value S which is the sum (integral) of m(x) values in this interval and get probabilities, $p(x) = m(x)/S$. Alternatively, we can divide $m(x)$ by the value M which is the maximum of m(x) values in the interval [a,b], $\text{pos}(x)=m(x)/M$.

However, these transformations do not establish the actual relation between theories: Fuzzy Set Theory (FST), Fuzzy Logic (FL) and Probability Theory (PrT) where the *main role play operations* with membership functions *m* and probabilities *p*. The concept of the fuzzy set is only a part of the fuzzy set theory we should not base the comparison of the theories based only one concept, but should analyze other critical concepts such as operations.

Probability theory operations of intersection and union ($\cap$ and $\cup$) are *contextual*, $p(x\cap y)=f(p(x), p(y|x))=p(x)p(y|x)$, where conditional probability $p(y|x)$ expresses contextual dependence between x and y. In contrast, operations in FST and FL, i.e., t-norms and t-conforms are "*context-free*", $m(x\&y) = f(m(x),m(y))$, that is no conditional properties, dependencies are captured. In fact FL, FST present rather a *special case* of the Probability Theory because all t-norms are a *subset* of copulas that represent n-D probability distributions [Resconi, Kovalerchuk, 2015 in this volume].

Joslyn [1995] also made an interesting comment that the same researchers, who object to confusion of membership grades with probability values, are not troubled with confusion of membership grades with possibility values. In fact both confusions have been resolved by:

(In1) interpreting a **fuzzy set** as a **set of probability distributions** (**SPD**) not a single probability distribution [Kovalerchuk, 1996, 2013, 2015].

(In2) interpreting **possibility** as **modal probability** as proposed by Cheeseman [1986].

In the next section we illustrate the first interpretation and show how fuzzy sets are interpreted /modeled by SPD. Note that the interpretation of membership functions as a *single probability distribution* is too narrow and often is really confusing. The interpretation In2 already has been discussed in the previous sections.

3.2. What is the relation of a membership function to probability?

Piegat [BISC, 04/10/2014] stated that a membership function of the crisp set is not pdf and not probability distribution, noting that it only informs that a set element belongs to the considered set."

While it is true that in general a membership function (MF) is not a pdf or a probability distribution it is not a necessary to compare them in a typical way, i.e., by one-to-one mapping. This relation can be expressed in different ways.

The relation can be one-to-many similarly to the situation when we compare a linear function and a piecewise linear function. While these functions differ, a set of linear functions accurately represents a piecewise linear function. In the case of MF a set of pdfs represents any MF of a crisp set or a fuzzy set [Kovalerchuk, 1996]. This technique uses a set of probability Spaces that are *Kolmogorov's measure (K-Measure) spaces* (S*KM*) [Kovalerchuk, 2015].

Let $f(x)=1$ be a MF of the interval [a, b]. Consider a pair of events

$\{e_1(x), e_2(x)\}$={x belongs to [a, b], x does not belong to [a, b]}

with pdf $p_x(e(x))=1$, if x belongs to [a,b] else $p_x(e(x)) = 0$. Thus, $p_x(e_1(x)) = f(x) =1$, $p_x(e_2(x))=0$ and $p_x(e_1(x)) + p_x(e_2(x)) = 1$. This is a simplest probability space that satisfies all Kolmogorov's axioms of probability. A set of these pdfs $\mathbf{p} =\{p_x\}$ for all x from [a,b] is equivalent to MF of the crisp set.

Note that here we use the term probability density function (pdf) in a general sense as any function with sum(integral) of its values equal to one on its domain. . It can be a traditional pdf defined on a continuous variable or on a set of any nature continuous or discrete that will be called a set of elementary events. As we will see below typically these sets will be pairs of NL sentences or phrases.

For a fuzzy set with a membership function $f(x)$ we have probabilities

$$p_x(e_1(x)) = f(x),\ p_x(e_2(x)) = 1- f(x)$$

that satisfy all Kolmogorov's axioms of probability, where again a set of these pdfs $\mathbf{p}=\{p_x\}$ for all x from [a,b] is equivalent to MF of the fuzzy set $\langle$ [a,b], f $\rangle$. Figure 1 shows these probability spaces for each x that is human's height h.

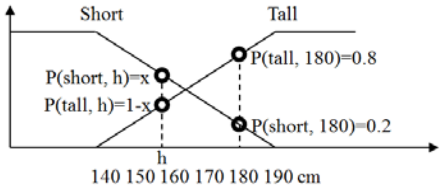

Figure 1. Sets of probability spaces S(h) for elementary events {short(h), tall(h)} for each height h.

In other words MF is a *cross-section* of a set of pdfs **p**. More details are in [Kovalerchuk, 1996]. The advantage of MF over a set of pdfs is in *compactness* of the representation for many MFs such as triangular MFs. Each of these MFs "compresses" an infinite number of pdfs. A probability space S(180) is shown as a pair of circles on a vertical line at point h=180. As we can see from this figure, just two membership functions serve as a compact representation of many simple probability spaces. This is a *fundamental representational advantage* of Zadeh's fuzzy linguistic variables [Zadeh, 1977] vs. multiple small probability spaces.

It is important to note that having two MFs $m_1(x)$ and $m_2(x)$ for two fuzzy sets the operations with them must follow the rules of probabilistic operations AND and OR with sets of respective pdfs $\mathbf{p}_1$ and $\mathbf{p}_2$ on two sets of probability spaces to be *contextual*.

An example of these operations is presented in [Kovalerchuk, 2015] using convolution-based approach. In this context it is very important to analyze the relation of t-norms and t-conorms used in fuzzy logic with sets of probability distributions. One of such analyses is presented in [Resconi, Kovalerchuk, 2015] in this volume.

The common in fuzzy logic "context-free" heuristic min and max operations and other t-norms and t-conorms for AND and OR operations on fuzzy sets are not correct probabilistic operations in general. These context-free operations can only serve as approximations of contextual probabilistic operations for AND and OR operations. For the tasks where the context is critical context-free operations will produce incorrect results.

Thus, membership functions and sets of pdfs are mutually beneficial by combining *fast* development of pdfs (coming from MFs) and rigor of *contextual operations* (coming from probability).

In this sense, fuzzy membership functions and probabilities are *complimentary not contradictory*. It would be incorrect to derive from advantages of MFs outlines above the conclusion that probabilistic interpretation only brings a complication to a simple fuzzy logic process for solving practical problems. The probabilistic interpretation of MFs removes a big chunk of heuristics from the solution and brings a rigorous mechanism to incorporate context into AND and OR operations with fuzzy sets.

3.3. Are fuzzy sets and possibility distributions equivalent?

Zadeh [1978] defined a possibility distribution as a fuzzy set claiming that possibilistic concepts are inherently more appropriate for fuzzy theory than probabilistic concepts.

This equivalence was rejected in earlier studies in [Joslyn, 1995] who reference works with a *wider view* on the concept of possibility that includes:

- its generalization to the *lattice* [Cooman et al., 1992],
- involvement of *qualitative relations* [Dubois, Prade, 1984],
- the semantics of *betting* [Giles, 1982],
- *measurement* theory [Yager, 1980],
- *abstract* algebra [Yager, 1993].

Other alternatives include:

- *modal probabilities* [Cheseman, 1986] and the later work by
- *elaborated lattice* approach [Grabisch, 2009].

Also a general logic of uncertainty based on the lattice can be traced to Gains [Gaines, 1984, Kovalerchuk, 1990]

3.4. Is interpreting grade of membership as probability meaningful?

Zadeh [02/05/2014] objects to equating a grade of membership and a truth value to probability, but equates a grade of membership and a truth value as obvious. He states that "equating grade of membership to probability, is *not meaningful*", and that "consensus-based definitions may be formulated *without* the use of probabilities".

To support this statement he provided an example of a *grade of membership* 0.7 that Vera is middle-aged if she is 43. Below we explicate and analyze his statement for this example that is reproduced as follows.

Grade of Membership that Vera is middle-aged being 43:

$$\text{GradeMembership(Vera is middle-aged}|\ \text{Age} = 43) = 0.7, \tag{1}$$

Truth value for Age=43 to be middle-aged,

$$\text{TruthValue(Age=43 is middle-aged)=0.7.} \tag{2}$$

Possibility that Vera is middle-aged for the Age = 43,

$$\text{Poss(Age} = 43|\ \text{Vera is middle-aged)} = 0.7 \tag{3}$$

Probability that Vera is middle-aged for the Age = 43,

$$\text{Prob(Vera is middle-aged}|\text{ Age} = 43) = 0.7 \tag{4}$$

In other words, Zadeh equates (1), (2) and (3) and objects to equating all of them to (4) as *not meaningful* and *not needed*.

Consider a crisp voting model that equates the grade of membership to probability *meaningfully*. In this model, the statement "Vera is middle-aged" is classified as true or false by each respondent. Respectively the grade of membership is defined as the *average* of votes. Dujmovic [2017] presented this approach in details in his chapter in this volume.

In a *flexible voting model* the voter indicates a grade of membership on the scale from 0 to 1 as a subjective/declarative probability of each voter. In the case of multiple voters the consensus-based grade of membership is defined as the *average* of subjective/declarative probabilities of voters. Thus, a meaningful probabilistic interpretation of a grade of membership exists for long time and is quite simple.

Next we comment on Zadeh's statement that "consensus-based definitions may be formulated *without* the use of probabilities". If the intent of the whole theory is to get results like in (1) then it can be obtained without using the word "probability". In fact both fuzzy logic and probability theory are interested to get more complex answers than (1). This leads to the need to *justify operations* on combination such as t-norms and t-conorms used in fuzzy logic as normative or descriptive. These operations differ from probability operations. Moreover there are multiples of such operations for both conjunction and disjunction in contrast with the probability theory, where the conjunction (intersection) and disjunction (union) are unique. Thus each t-norm and t-conorm must be justified against each other and probabilistic operations. Note that experimental studies [Thole, U., Zimmermann, H-J.,Zysno, 1979; Kovalerchuk, Taliansky, 1992] with humans do not justify min t-norm vs. product t-norm, which is a probabilistic operation for independent events. In general the justification of operations continues to be debated [Belaykov et al., 2012; Kovalerchuk, 2010] while extensive experimental studies with humans still need to be conducted in the fuzzy logic justification research.

3.5. Unsharpness of border and randomness vs. axiomatic probability

Chang [BISC, 2014] differenced fuzzy logic and probability theory stating that (1) *fuzzy sets theory* deals with *unsharpness*, while the *probability theory* deals with *randomness* and (2) the problems with unsharpness *cannot be converted* into "randomness problems". This statement was made to justify that there are problems which cannot be solved by using the probability theory, but which can be solved by using the fuzzy sets and fuzzy logic. In this way, he attempted to counter my arguments that probability theory has a way to solve Zadeh's CWW test tasks by using the formal axioms of probability theory (called by Tschantz [BISC, 03/11/2014] axioms of *K-measure space*).

Tschantz [BISC, 04/07/2014] objected to Chang's statement: "As Boris has said many times, he is not attempting to convert reasoning about vague words into reasoning about randomness. He is merely attempting to *reuse the formal axioms of Kolmogorov*. While these axioms were first found useful in modeling randomness, the fact that they are useful for modeling randomness does not mean that they cannot also be useful for modeling vagueness. To put it another way, someone can use the same model M to model both X or Y even if X and Y are not the same thing. Thus, I do not buy CL Chang's argument that Boris is incorrect**.** If he [Chang] shows that vagueness has some feature that makes it behave in such a way that is so different from randomness that no model can be good at modeling both, then I would buy his argument. However, I would need to know *what that feature is*, what "good" means in this context, and how he proves that no model can be good at both. It seems far easier to point to some contradiction or unintuitive result in Boris's model."

So far no contradiction or unintuitive results have been presented by opponents to reject the approach based on the Space of K-Measures.

4. What is the future rigorous possibility theory?

4.1. Is possibility an upper estimate of probability for complex events?

Below we analyze abilities to interpret possibility Pos as an upper estimate of probability Pr using the current min-max possibility theory where "&" operator is modeled as min and "or" operator is modeled as max.

Section 2.4 had shown how questions with words "possibility" and "probability" can be reworded to be in a comparable form using their Webster definitions. Thus Webster leads to the conclusion that possibility must be an upper estimate of probability in NL for every expression X,

$$Pos(X) \geq Pr(X)$$

Respectively when possibility of A&B is defined as

$$Pos(A\&B) = min(Pos(A),Pos(B))$$

it gives an *upper estimate* of Pr(A&B),

$$Pos(A\&B) = min(Pos(A),Pos(B)) \geq Pr(A\&B)$$

However in contrast when Pos(A or B) is defined as

$$Pos(A \text{ or } B) = max(Pos(A), Pos(B))$$

it gives a *lower estimate* of Pr(A or B)=P(A)+P(B)-P(A&B)

$$\text{Pos(A or B)} = \max(\text{Pos(A)}, \text{Pos(B)}) \leq \text{Pr(AorB)}$$

Thus, the NL property Pos(X) ≥ Pr(X) does not hold when X = (A or B) in the min-max possibility theory advocated by Zadeh and others.

Let us compute the possibility of another composite expression X = A&(B or C) using the min-max possibility theory:

$$\text{Pos(A\&(B or C)} = \min(\text{Pos(A)}, \text{Pos(B or C)})) =$$

$$\min(\text{Pos(A)}, \max(\text{Pos(B)}, \text{Pos(C)})).$$

The value min (Pos(A), Pos(B or C))) is supposed to be an upper estimate for Pr(A&(B or C)). However, min(Pos(A),max(Pos(B), Pos(C))) is not an upper estimate for Pr(A&(B or C)) because max (Pos(B), Pos(C)) is a lower estimate for Pr(B or C). The infusion of this lower estimate to min (Pos(A), Pos(B or C))) can destroy an upper estimate.

Below we show it with a numeric example. Consider sets A, B and C such that A has a significant overlap with (B or C) with probabilities

$$\text{Pr(A)} = 0.4,\ \text{Pr(B)} = 0.3,\ \text{Pr(C)} = 0.32,\ \text{Pr(B or C)} = 0.5,\ \text{Pr(A\&(B or C))} = 0.35$$

From these numbers we see that

$$\text{Pr(A)} = 0.4 < \text{Pr(B or C)} = 0.5$$

$$\text{Pr(B or C)} = 0.5 > \max(\text{Pr(B)}, \text{Pr(C)}) = \max(0.3, 0.32)) = 0.32$$

$$\min(\text{Pr(A)}, \text{Pr(B or C)})) = \min(0.4, 0.5) = 0.4 =$$

$$\text{Pos(A \&(B or C))} > \text{Pr(A\&(B or C))} = 0.35.$$

$$\text{Pos(A \&(B or C))} = \min(\text{Pr(A)}, \max(\text{Pr(B)},\text{Pr(C)})) =$$
$$\min(0.4, 0.32) = 0.32 < \text{Pr(A\&(B or C))} = 0.35$$

Thus, upper estimate 0.4 of probability 0.35 is converted to the lower estimate 0.32 of this probability 0.35.

This confirms that for composite expressions that involve “or” operator, we cannot ensure that possibility of that expression is an upper estimate of its probability in the min-max possibility theory. Thus, a future possibility theory should differ from the current min-max possibility theory to be able to hold the property that Pos(X) ≥ Pr(X) for all composite expressions X.

4.2. Should evaluation structure be limited by [0,1] or lattice?

In Section 3.3 we referenced several studies that generalized the possibility theory to a lattice (partial order of evaluations) from the full order in [0.1]. However, to the best of our knowledge, none of the previous studies went beyond a *lattice structure*. The lattice assumption means that the *absolute positive* and *negative possibility* exists with max value commonly assigned to be equal to 1 and min value to be equal to 0. It does not mean that *x* with Pos(*x*)=1 always happens, but such *x* is one of the real alternatives, e.g., *x* is physically possible. Similarly Pos(*x*)=0 can mean that *x* physically cannot happen.

There are NL concepts where this assumption can be too restrictive. Consider the concept of happiness and a chance that somebody might by absolutely happy. Can somebody be absolutely happy (unhappy) to the level that he/she cannot be happier in the future? The lattice assumption tells that if *x* is absolutely happy at time t, Happy(x)=1, then x cannot be happier later. If *x* tells that he/she was absolutely happy yesterday, and even happier today then under the lattice model this person will be inconsistent or irrational. This person "exceeded" the absolute maximum. Another example is "possibility of absolute understanding". What is a chance that somebody might understand something absolutely today and cannot increase understanding later? In other words, we question the concept of absolute maximum for possibilities such as absolute happiness, Pos(Happy(x))=1 and absolute understanding, Pos(Understand(x))=1. The number of such "unlimited" NL concepts is quite large.

How to build a model to accommodate such behavior? We can simply remove limit 1 (or any constant given in advance) and allow the unlimited values of possibility of happiness, in this example and, in a general measure of possibility.

Other difficulties for constructing an evaluation medium are that often a set of possible alternatives is uncertain, and assigning the possibility values to a known alternative is highly uncertain and difficult. For instance, the NL expression "unlimited possibilities" often means that a set of alternatives is huge, and not fully known, being uncertain to a large extent. Often it is easier to say that *A* is more possible than *B*, and both alternatives are not fully possible, than to give numeric Pos(A) and Pos(B). We simply may have no stable landmark for this. The NL expression "unlimited possibilities" can mean that for a possible *A*, there can exist a *B*, which is more *possible* than *A*. The next difficulty is that adding more possible alternatives to the consideration changes the possibility values. Next the fixed value of max of possibility, e.g., $\max_{x\in X}\mathrm{Pos}(x) = 1$, is not coming from the NL, but is just *imposed* by the mathematical formalisms. We advocate the need for developing the formalism with "unlimited" possibilities.

4.3. From an exact number to an interval

Another common viewpoint is that possibility expresses a *less constraining form of uncertainty* than probability [Dubois and Prade, 1989], and even that the probability theory brings in the excess of constraints [Joslyn, 1995] requiring an *exact number* instead of an interval of belief and plausibility [Bel,Pl] as in the evidence measure in the belief theory.

Again while it is true that probability is a single number, not an interval, the issue is not in the formal expansion to the interval, but in the justification of operations to produce the intervals and operations with intervals. We need well justified *empirical procedures* to get them. Without this generalization, the intervals and other alternative formalisms have little value. Somebody must bring the "life meaning" to it. Respectively the claims of successful applications need to be closely scrutinized to answer a simple question: "Is the application successful because of the method, or in spite of it?" In other words: "Does this method capture the properties of the application better than the alternative methods, in our case the pure probabilistic methods?" It short, deficiency of one method should not be substituted by deficiency of another method.

5. Conclusion

This work clarified the relationships between the probability and the possibility theories as tools for solving Computing with Words tasks that require dealing with extreme uncertainty of the Natural Language. Multiple arguments have been provided to show deficiencies of Zadeh's statement that CWW tasks are in the province of the fuzzy logic and possibility theory, and probabilistic solutions are not appropriate for these tasks. We proposed a useful constructive probabilistic approach for CWW based on sets of specialized K-Measure (Kolmogorov's measure) probability spaces. Next we clarified the relationships between probability and possibility theories on this basis, outlined a future rigorous possibility theory and open problems for its development.

6. References

1. Beliakov, G., Bouchon-Meunier, B., Kacprzyk, J., Kovalerchuk, B., Kreinovich, V., Mendel J.,: Computing With Words: role of fuzzy, probability and measurement concepts, and operations, Mathware & Soft Computing Magazine. Vol.19 n.2, 2012,27-45
2. Burdzy, K., Search for Certainty. On the Clash of Science and Philosophy of Probability. World Scientific, Hackensack, NJ, 2009.
3. Cheeseman, P., Probabilistic vs. Fuzzy Reasoning. In: Kanal, Laveen N. and John F. Lemmer (eds.): Uncertainty in Artificial Intelligence, Amsterdam: Elsevier (North-Holland), 1986, pp. 85–102.
4. Chang, C-L. Berkeley Initiative on Soft Computing (BISC) post, 2014 http://mybisc.blogspot.com.
5. Cooman, Gert de, Kerre, E. and Vanmassenhov, FR., Possibility Theory: An Integral Theoretic Approach, Fuzzy Sets and Systems, v. 46, pp. 287-299, 1992.
6. Dubois, D., Possibility Theory and Statistical Reasoning, 2006
7. Dubois, D., Nguyen H. T., Prade H., 2000. Possibility theory, probability and fuzzy sets: misunderstandings, bridges and gaps. In: D. Dubois H. Prade (Eds), Fundamentals of Fuzzy Sets, The Handbooks of Fuzzy Sets Series, Kluwer, Dordrecht, 343-438.

8. Dubois, D., Prade, H., Measure-Free Conditioning, Probability, and Non-Monotonic Reasoning, in: Proc. 11th Int. Joint Conf. on Artificial Intelligence, pp. 1110-1114, 1989.
9. Dubois, D., Steps to a Theory of Qualitative Possibility", in: Proc. 6th Int. Congress on Cybernetics and Systems, pp. 10-14, 1984.
[1] Dujmovic, J, Berkeley Initiative on Soft Compting (BISC), Post on 4/28/2014. http://mybisc.blogspot.com.
10. Dujmović, J. Relationships Between Fuzziness, Partial Truth, and Probability in the Case of Repetitive Events, In: Uncertainty Modeling, Kreinovich, V.(Ed.), Springer, 2017, 61-71. (This volume).
11. Joslyn, C., Possibilistic processes for complex systems modeling, Ph. D. dissertation, SUNY Binghamton, 1995. http://ftp.gunadarma.ac.id/pub/books/Cultural-HCI/Semiotic-Complex/thesis-possibilistic-process.pdf
12. Gaines, B., Fundamentals of decision: Probabilistic, possibilistic, and other forms of uncertainty in decision analysis. Studies in Management Sciences, 20:47-65, 1984.
13. Giles, R., Foundations for a Theory of Possibility, in: Fuzzy Information and Decision Processes, pp. 183-195, North-Holland, 1982.
14. Goodman, I.R., Fuzzy sets as equivalent classes of random sets, in: R. Yager (Ed.), Fuzzy Sets and Possibility Theory, 1982, pp. 327–343.
15. Grabisch, M., Belief Functions on Lattices, International Journal of Intelligent Systems, Vol. 24, 76–95, 2009.
16. Kolmogorov,A., Foundations of the Theory of Probability, NY, 1956.
17. Kovalerchuk, B., Summation of Linguistic Numbers, Proc. of North American Fuzzy Information Processing Society (NAFIPS) and World Congress on Soft Computing, 08-17-19, 2015, Redmond, WA pp.1-6. DOI: 10.1109/NAFIPS-WConSC.2015.7284161
18. Kovalerchuk, B., Berkeley Initiative on Soft Computing (BISC), Posts on 01/17,2014, 02/03/2014, 03.07.2014.03/08/2014, 03/21/2014http://mybisc.blogspot.com.
19. Kovalerchuk, B., Probabilistic Solution of Zadeh's test problems, in IPMU 2014, Part II, CCIS 443, Springer, pp. 536–545, 2014.
20. Kovalerchuk, B., Quest for Rigorous Combining Probabilistic and Fuzzy Logic Approaches for Computing with Words, in R.Seising, E. Trillas, C. Moraga, S. Termini (eds.): On Fuzziness. A Homage to Lotfi A. Zadeh, SFSC Vol. 216), Springer 2013. Vol. 1. pp. 333-344.
21. Kovalerchuk B.: Interpretable Fuzzy Systems: Analysis of T-norm interpretability, IEEE World Congress on Computational Intelligence, 2010. doi: 0.1109/FUZZY.2010.5584837
22. Kovalerchuk, B., Vityaev, E.: Data Mining in Finance: Advances in Relational and Hybrid Methods (chapter 7 on fuzzy systems), Boston: Kluwer, 2000.
23. Kovalerchuk, B., Context Spaces as Necessary Frames for Correct Approximate Reasoning. International Journal of General Systems. vol. 25 (1) (1996) 61–80
24. Kovalerchuk, B., Klir, G., Linguistic context spaces and modal logic for approximate reasoni.ng and fuzzy-probability comparison. In: Proc. of Third International Symposium on Uncertainty Modeling and Analysis and NAFIPS' 95, IEEE Press, A23-A28 (1995
25. Kovalerchuk, B,, Talianski, V., Comparison of empirical and computed values of fuzzy conjunction. Fuzzy sets and Systems 46:49-53 (1992)
26. Kovalerchuk, B. Analysis of Gaines' logic of uncertainty. In: Proceedings of NAFIPS'90, Eds I.B.Turksen. v.2, Toronto, Canada, 293-295, 1990.
27. Nguyen, H.T., Fuzzy and random sets, Fuzzy Sets and Systems 156 (2005) 349–356.
28. Nguyen, H.T., Kreinovich, V., How to fully represent expert information about imprecise properties in a computer system: random sets, fuzzy sets, and beyond: an overview. Int. J. General Systems 43(6): 586-609 (2014)
29. Orlov A.I., Fuzzy and random sets, Prikladnoi Mnogomiernii Statisticheskii Analyz (Nauka, Moscow), 262-280, 1978 (in Russian).

30. Piegat, A., Berkeley Initiative on Soft Compting (BISC) Post on 04/10/2014. http://mybisc.blogspot.com.
31. Piegat, A., A New Definition of the Fuzzy Set, Int. J. Appl. Math. Comput. Sci., Vol. 15, No. 1, 125–140, 2005
32. Raufaste, E., da Silva Neves, R., Mariné, C., Testing the descriptive validity of Possibility Theory in human judgments of uncertainty, Artificial Intelligence, Volume 148, Issues 1–2, 2003, 197–218.
33. Resconi G., Kovalerchuk, B., Copula as a Bridge between Probability Theory and Fuzzy Logic. (In this volume).
34. Thole, U., Zimmermann, H-J., Zysno, P.: On the Suitability of Minimum and Product Operations for the Intersection of Fuzzy Sets, Fuzzy Sets and Systems 2, 173-186 (1979)
35. Tschantz, M., Berkeley Initiative on Soft Compting (BISC) Posts on 03/11/2014, 04/07/2014, http://mybisc.blogspot.com.
36. Yager, R., On the Completion of Qualitative Possibility Measures, IEEE Trans. on Fuzzy Systems, v. 1:3, 184-194, 1993.
37. Yager, R., Foundation for a Theory of Possibility, J. Cybernetics, v. 10, 177-204, 1980.
38. Wright, G., Ayton, O.,(eds.) Subjective probability. Wiley, Chichester, NY, 1994.
39. Zadeh L., Berkeley Initiative on Soft Compting (BISC), Posts on 01/30/2014, 02/05/2014 , 04/09/2014, 05/27/2014, 10/27/2014 http://mybisc.blogspot.com.
40. Zadeh, L., Computing with Words: Principal Concepts and Ideas, Springer, 2012.
41. Zadeh, L., A Note on Z-numbers, Information Science 181, 2923–2932, 2011.
42. Zadeh, L., From Computing with Numbers to Computing with Words—From Manipulation of Measurements to Manipulation of Perceptions, IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications, Vol. 45, No. 1, 1999, 105-119
43. Zadeh, L.: Discussion: Probability Theory and Fuzzy Logic are Complementary rather than competitive, Technometrics, vol. 37, n 3, 271-276, 1996.
44. Zadeh, L., Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems, 1, 3-28, 1978.
45. Zadeh, L: The concept of linguistic variable and its application to approximate reasoning - 1. Information Sciences 8, 199-249, 1977.