FUZZY DATA IN STATISTICS

Milan Mareš

The development of effective methods of data processing belongs to important challenges of modern applied mathematics and theoretical information science. If the natural uncertainty of the data means their vagueness, then the theory of fuzzy quantities offers relatively strong tools for their treatment. These tools differ from the statistical methods and this difference is not only justifiable but also admissible. This relatively brief paper aims to summarize the main fuzzy approaches to vague data processing, to discuss their main advantages and also their essential limitations, and to specify their place in the wide scale of information and knowledge processing methods effective for vague data.

Keywords: fuzzy quantity, extension principle, fuzzy data

AMS Subject Classification: 03E72, 08A72, 62–07

1. INTRODUCTION

The, mostly quantitative, data which are to be processed frequently look exactly determined but they rarely are such. The indeterminism more or less hidden in their structure may be of different types: imprecision or approximation, randomness, enormous complexity complicating full understanding, and also vagueness. The last type of indeterminism is very frequently connected with the properties of natural language used during the process of data acquisition. In the communication among people, certain degree of vagueness is not only acceptable but even useful – it means its flexibility and adaptability to new situations and phenomena. On the other hand, the exact mathematical methods used for vague input data may cause serious difficulties or at least doubts when the obtained output data are interpreted and, moreover, applied to practical situations. The world appears rather more uncertain than the results aim to state. Consequently, the development of adequate mathematical tools for processing vagueness became quite urgent in the second half of the twentieth century (see [12]).

For long time, probability theory and probability-based statistics appeared to be sufficient for mathematical modelling of any type of indeterminism. Anyhow, the first attempts to artificial intelligence, intelligent systems, and to related problems pointed at a different type of uncertainty. In many situations, it could be processed by probabilistic methods, as well. Only the derived results were sometimes rather different from the experience-based expectation. The probabilistic models did not reflect the type of uncertainty in question.

The theory of fuzzy sets formulated in [12] attempts to handle the uncertainty of vague expressions in the natural language. The first reaction on the new model of uncertainty was not indifferent at all. On one side, hundreds of papers (yes, many of them were mathematically very primitive) suggested useful applications of fuzziness, on the other, numerous authors tried to show that all fuzzy set theoretical models can be (better!) reformulated in the probabilistic language and that the whole fuzzy sets theory is nothing more than some "pidgeon probability". Let us admit that some elementary concepts of fuzzy sets, the so called membership functions, optically remember essentially simplified probability densities. Since that time, the development of fuzzy set theory brought new interesting theoretical and applied results (and also the most active fundamentalists on both sides became older and more conciliatory), and the parallel existence of both theories, as well as their justification are generally accepted. It is known that the difference between their models results from their ability to describe essentially different phenomena, which also means that the methods of their processing are different. There exist probabilistic questions and probabilistic answers, as well as fuzzy theoretical questions and answers. In some, relatively rare, cases the choice between them is rather subjective but in the great majority of situations one of both models is evidently more adequate.

One of the opportunities in which both types of uncertainty meet, is the statistical processing of random phenomena which are described by (usually quantitative) expressions of vague natural language, like "*about 5*", "*almost 8*", "*something between 10 and 15*", but also, in the extreme cases, "*many*", "*several*", "*sufficiently low*" and many others. In such case, the classical statistician usually "looks in another direction", substitutes the vague quantities by some deterministic representatives or, in the better case, treats the vagueness as some special case of randomness, however it does not correspond with the essence of the modelled phenomena. Let us accept a paradigm that there exist another possibility. We may admit the fact that the observed results of random experiment are described vaguely, that the fuzzy sets and, especially, fuzzy quantities properly respect this vagueness, and that there exist effective tools for processing the vague (fuzzy) data similarly, even if not identically, like the usual deterministic data. This paper is to offer a brief overview of such tools and to discuss their specificity.

2. FUZZINESS AND RANDOMNESS

Both concepts mentioned in the heading deal with some forms of uncertainty. The randomness is well-known, here we discuss the position of fuzziness in the "territory" not covered by probability.

For those who were not interested in the fuzzy sets, yet, and also to fix the elementary notations, we briefly introduce the basic concepts and symbols of fuzzy set theory, and discuss the relation between fuzzy and random phenomena. The acceptance of that relation means the main problem for probabilists meeting the fuzziness.

Let us consider a basic set $\mathcal{U}$ which is called the *universum*. Each *fuzzy set* on $\mathcal{U}$, denoted by A, is defined by its *membership function* $\mu_A : \mathcal{U} \to [0,1]$. The membership function extends the concept of the deterministic characteristic function with values in $\{0,1\}$, where the interpretation of $\mu_A(x) = 0$ and $\mu_A(x) = 1$ is the same like in the deterministic case, and $\mu_A(x) \in (0,1)$ means that $x \in A$ only with some possibility, quantitatively determined by the membership value $\mu_A(x)$.

We denote by $\mathcal{P}(\mathcal{U})$ the class of all deterministic subsets of $\mathcal{U}$, and by $\mathcal{F}(\mathcal{U})$ the class of all fuzzy sets on $\mathcal{U}$. IF $M \in \mathcal{P}(\mathcal{U})$ then we define by $\langle M \rangle \in \mathcal{F}(\mathcal{U})$ the fuzzy set for which

$$\mu_{\langle M \rangle} = 1 \quad \text{if } x \in M, \qquad \mu_{\langle M \rangle} = 0 \quad \text{if } x \notin M. \tag{1}$$

The classical set-theoretical operations are easily extended on $\mathcal{F}(\mathcal{U})$ as follows. If A, $B \in \mathcal{F}(\mathcal{U})$ then

$$A \subset B \quad \text{iff } \mu_A(x) \le \mu_B(x) \text{ for all } x \in \mathcal{U}, \qquad \text{(subset)}, \tag{2}$$

$$\overline{A} \quad \text{iff } \mu_{\overline{A}}(x) = 1 - \mu_A(x),\ \forall\, x \in \mathcal{U}, \qquad \text{(complement)}, \tag{3}$$

$$A \cup B \quad \text{iff } \mu_{A \cup B}(x) = \max(\mu_A(x),\, \mu_B(x)),\ \forall\, x \in \mathcal{U} \qquad \text{(union)}, \tag{4}$$

$$A \cap B \quad \text{iff } \mu_{A \cap B}(x) = \min(\mu_A(x),\, \mu_B(x)),\ \forall\, x \in \mathcal{U} \ \text{(intersection)}. \tag{5}$$

Even if the above operations extend their deterministic analogies, some of their properties may be rather surprising. It is, e. g., not difficult to construct a fuzzy set which is a subset of its own complement.

Having introduced the elementary formalism of fuzzy sets, we may comment the eventual discussion about the similarity between fuzziness and randomness.

An example may be better than long argumentation. The question: "*How old is this man?*" leads to a probabilistic answer: "*He is probably about 30*" and it is formally represented by a probability distribution over the set of natural numbers with modus near 30. If we add "*But he can be also almost 40*", the formal representation of both sentences is still a probability distribution over naturals – we have only to increase the probabilities of number near 40 and, consequently, to decrease the probabilities of values about 30. This decreasing follows naturally from both – the definition of probability distribution (with total probability 1), and the intuitive conclusion that increasing the change of 40 we automatically decrease the chances of some other values.

Let us, now, consider another question: "*What is the acceptable age of new employee?*" We may answer "*Acceptable age is about 30,*" and to represent this answer by a fuzzy subset of natural numbers with a membership function. If we, similarly to the previous case, add "*But almost 40 would not be bad, too,*" then we may simply increase the membership values of numbers nearing 40. In this case, we need not proportionally decrease other values of the membership function as the sum of all membership is not limited. Note, please, that this is no imperfectness of fuzzy sets but a simple reflection of the fact that the verbal category "acceptable age" regards each age separately – the acceptability of one age does not influence this property of another. To be "acceptable" is a rather uncertain quality, but this uncertainty is not the randomness, it is vagueness connected with the used verbal expression. The question in this second case is not probabilistic but fuzzy set theoretical.

The fact that the probabilities and memberships are further processed by different tools (the addition and multiplication on one side and maximum or minimum on the other) follows from the essential properties of both uncertainties.

Very formally, there really exists a strong connection between fuzzy sets and random events. Namely, the event that a point $x \in \mathcal{U}$ belongs to the fuzzy set A can be interpreted as random event, binary distributed, with probabilities $(\mu_A(x), 1 - \mu_A(x))$. Then a fuzzy set A is a parametrized class of binary random events $\{A_x : x \in \mathcal{U}\}$, where each A_x is an event of the above type. In this sense, fuzzy sets respect the probabilistic rules. Of course, the model of vague phenomenon as a parametrized class of binary distributions is not very handy, and also the processing of fuzziness by means of maxima and minima is not typical for the probabilistic models.

Comparing the probabilistic and fuzzy set theoretical model of uncertainty, it is useful to stress one great advantages of probabilities – the large numbers law. It connects the abstract concept of probability with very concrete phenomenon of the material world, the relative frequency. It supports the numerical values of probabilistic measures with the everyday experience and, moreover, it offers an effective method of quantification of actual values of probabilities. Nothing like this exists in the fuzzy set theory – no asymptotic properties relate the membership functions to vagueness, no matter how it could be quantified.

3. FUZZY QUANTITIES

The quantitative data represent a typical object of statistical processing. If those data are vague then they may be represented by a specific type of fuzzy set over the real line called fuzzy quantity. More formally, *fuzzy quantity* a is a fuzzy subset of R with membership function $\mu_a : R \to [0, 1]$ such that

$$\text{there exists } x_a \in R \text{ such that } \mu_a(x_a) = 1, \tag{6}$$

$$\text{there exist } x_a^1 < x_a^2 \in R \text{ such that } \mu_a(x) = 0 \text{ for } x \notin [x_a^1, x_a^2]. \tag{7}$$

The number x_a fulfilling (6) is called the *modal value* of a. The set of all fuzzy quantities is denoted $Q(R)$, and preserving the notations used above, $Q(R) \subset \mathcal{F}(R)$.

The fuzzy quantities, as defined above, include a wide scale of formal structures – membership functions. Nevertheless, in many practical applications, much narrower subset of $Q(R)$ including very simple fuzzy quantities is sufficient for modelling practical problems. Let us consider four numbers $x_a^1 \le x_a \le x_a' \le x_a^2$ and a fuzzy quantity a with membership function μ_a such that

$$\begin{aligned} \mu_a(x) &= (x - x_a^1)/(x_a - x_a^1) && \text{for } x \in (x_a^1, x_a) \\ &= 1 && \text{for } x \in [x_a, x_a'], \\ &= (x_a^2 - x)/(x_a^2 - x_a') && \text{for } x \in (x_a', x_a^2), \\ &= 0 && \text{else,} \end{aligned} \tag{8}$$

where we suppose that $x_a^1 \ne x_a$ and $x_a^2 \ne x_a'$. If some of those inequalities turns into equality then (8) can be easily modified by omitting the relevant interval (which becomes empty). Such fuzzy quantity is called *trapezoidal* or also a *fuzzy interval*. If $x_a = x_a'$ then it turns into *triangular* fuzzy quantity, called also *fuzzy number*.

The fuzzy quantities extend the traditional concept of number by admitting also alternative possible values of the considered quantity. It means that also a crisp number $r \in R$ can be considered for fuzzy quantity with extremally reduced membership function. We denote this quantity by $\langle r \rangle \in Q(R)$, and put

$$\mu_{\langle r \rangle}(r) = 1, \qquad \mu_{\langle r \rangle}(x) = 0 \quad \text{for } x \neq r. \tag{9}$$

As the fuzzy quantities extend the real numbers, it is desirable to extend also the classical algebraic operations over R to the set $Q(R)$. The usual method of doing it is called *extension principle*. Its general formulation is as follows.

Let $a_1, a_2, \ldots, a_n \in Q(R)$ and let $f(x_1, x_2, \ldots, x_n) : R \times R \times \cdots \times R \to R$ be a real-valued function of $n \in N$ real variables. Then its *extension* $f : Q(R) \times Q(R) \times \cdots \times Q(R) \to Q(R)$ is a mapping such that the membership function of $f(a_1, a_2, \ldots, a_n) \in Q(R)$ is

$$\mu_{f(a_1,a_2,\ldots,a_n)}(x) = \sup_{\substack{x_1 \in R,\, x_2 \in R, \ldots, x_n \in R \\ x = f(x_1, x_2, \ldots, x_n)}} \left[\min\left(\mu_{a_1}(x_1), \mu_{a_2}(x_2), \ldots, \mu_{a_n}(x_a)\right)\right]. \tag{10}$$

General formula (10) gains simpler forms for particular algebraic operations.

The *sum of fuzzy quantities* a and b is a fuzzy quantity $a \oplus b$ with membership function

$$\mu_{a \oplus b}(x) = \sup_{y \in R} \left[\min(\mu_a(y), \mu_b(x - y))\right], \quad x \in R. \tag{11}$$

The *crisp product* of real number $r \in R$ and fuzzy quantity $a \in Q(R)$ is a fuzzy quantity $r \cdot a$ with membership function

$$\begin{aligned} \mu_{r \cdot a}(x) &= \mu_a(x/r) \quad \text{if } x \in R,\ r \neq 0, \\ &= \mu_{\langle 0 \rangle}(x) \quad \text{if } r = 0 \end{aligned} \tag{12}$$

which implies

$$\mu_{-a}(x) = \mu_{-1 \cdot a}(x) = \mu_a(-x) \quad \text{for } x \in R. \tag{13}$$

More general *fuzzy product* of a, $b \in Q(R)$ is a fuzzy quantity $a \odot b$, where

$$\begin{aligned} \mu_{a \odot b}(x) &= \sup_{\substack{y \in R \\ y \neq 0}} \left[\min(\mu_a(y),\, \mu_b(x/y))\right], \quad x \in R,\ x \neq 0 \\ &= \max\left(\mu_a(0),\, \mu_b(0)\right) \qquad \text{for } x = 0. \end{aligned} \tag{14}$$

Observation 1. If $r \in R$, $a \in Q(R)$, then $r \cdot a = \langle r \rangle \odot a$.

Similar formulas can be derived for other binary operations. For our purpose, the above formulas (in most cases the sum and crisp product) are usually sufficient. In some situations, the second power can be useful, too.

If $a \in Q(R)$, then also $a^2 \in Q(R)$ and, in accordance with (10),

$$\begin{aligned} \mu_{a^2(x)} &= \mu_a\left(\sqrt{x}\right) \quad \text{for } x \in R,\ x \geq 0 \\ &= 0 \qquad \text{for } x < 0 \end{aligned} \tag{15}$$

(as $y^2 \geq 0$ for all $y \in R$).

The fuzzy quantities can be ordered, similarly to deterministic numbers. In fact, there exist many different approaches to the binary ordering relation over $Q(R)$, the rational ones are summarized and analyzed in [5] partly also in [8]. The most natural appears the ordering paradigm due to which the ordering relation between vague elements is to be vague, i. e., fuzzy, as well. Respecting this principle, we define the fuzzy ordering $\succsim$ of fuzzy quantities as a fuzzy subset of the cartesian product $Q(R \times Q(R))$ with membership function $\nu_{\succsim}$ such that for any ordered pair a, b of fuzzy quantities

$$\nu_{\succsim}(a,b) = \sup_{x,y \in R,\, x \geq y} [\min(\mu_a(x),\, \mu_b(y))] \tag{16}$$

means the possibility that $a \succsim b$.

The above concepts represent the tools applicable whenever the classical statistical methods are to be transmitted into the environment of fuzzy data. Their practical applicability depends on their computational properties.

4. ALGEBRAIC PROPERTIES OF FUZZY QUANTITIES

The operations and relations over fuzzy quantities, described in the previous section rather differ from those over deterministic real numbers. More about them is summarized, e. g., in [1, 2, 7, 8]. The most essential difference consists in the validity of some intuitively expected algebraic properties.

In this and all following sections, the equation $a = b$ for a, $b \in Q(R)$ means the total identity of their membership functions, $\mu_a(x) = \mu_b(x)$ for all $x \in R$.

Let us start with the group properties (see [7, 8]), and consider the operation of sum $\oplus$. Then it is easy to derive the *commutativity* and *associativity* for a, b, $c \in Q(R)$

$$a \oplus b = b \oplus a, \qquad (a \oplus b) \oplus c = a \oplus (b \oplus c).$$

There exists exactly one zero *fuzzy quantity* $o \in Q(R)$ such that for any $a \in Q(R)$, $a \oplus o = a$, namely, $o = \langle 0 \rangle$. Unfortunately, except very special (and rather degenerated) cases, for no general fuzzy quantity $a \in Q(R)$ there exists $(-a) \in Q(R)$ such that $a \oplus (-a) = o = \langle 0 \rangle$. This group property (*opposite element*) is not fulfilled.

The group properties of the fuzzy product $\odot$ are quite analogous. It is commutative and associative, there exists exactly one *unit-element* $\langle 1 \rangle$ such that $a \odot \langle 1 \rangle = a$ for any $a \in Q(R)$. Also the existence of *reverse elements* is not fulfilled.

The crisp product $r \cdot a$, where $a \in Q(R)$, $r, r_1, r_2 \in R$ is a simple operation fulfilling

$$r \cdot a = a \cdot r, \qquad r_1 \cdot (r_2 \cdot a) = (r_1 \cdot r_2) \cdot a.$$

Using both operations, sum and product, parallelly, we can meet the problem of their *distributivity*. Generally, the equation between $a \odot (a \oplus c)$ and $(a \odot b) \oplus (a \odot c)$, where a, b, $c \in Q(R)$, is fulfilled for special cases, summarized in [1, 2] and recollected in [7, 8]. Luckily, one of the special cases covers the case

$$r \cdot (a \oplus b) = (r \cdot a) \oplus (r \cdot b), \quad r \in R,\; a,\, b \in Q(R),$$

meanwhile the symmetric case, the equality between $r_1 \cdot a \oplus r_2 \cdot a$ and $(r_1 + r_2) \cdot a$ is not generally true (hence, $a \oplus a$ is not generally equal to $2 \cdot a$).

Observation 2. Let $a, b \in Q(R)$ be trapezoidal, characterized by quadruples $(x_a^1, x_a, x'_a, x_a^2)$ and $(x_b^1, x_b, x'_b, x_b^2)$, respectively, and let $r \in R$. Then $(-a)$, $r \cdot a$, $a \oplus b$ are trapezoidal, characterized by

$$\left(-x_a^2,\, -x'_a,\, -x_a,\, -x_b^1\right),$$
$$\left(r \cdot x_a^1,\, r \cdot x_a,\, r \cdot x'_a,\, r \cdot x_a^2\right) \text{ if } r > 0, \quad \left(r \cdot x_a^2,\, r \cdot x'_a,\, r \cdot x_a,\, r \cdot x'_a\right) \text{ if } r < 0,$$
$$\left(x_a^1 + x_b^1,\, x_a + x_b,\, x'_a + x'_b,\, x_a^2 + x_b^2\right),$$

respectively.

Let us note that for trapezoidal a, b, their fuzzy product $a \odot b$ is not generally trapezoidal. Nevertheless, its approximation by trapezoidal fuzzy quantity is possible and in some applications quite sufficient.

5. WHERE ARE THE ROOTS OF PROBLEMS?

The rather exotic behaviour of fuzzy quantities can be quite uncomfortable in applications. We are used to expect some algebraic properties fulfilled by quantitative data, and feel rather surprised, when, e. g., $a \oplus a \neq 2 \cdot a$. To understand the structure of fuzziness in the numerical environment, it is useful to analyze the essence of these inconsistencies. It was done in several papers and the results are summarized in [7] and [8].

Those results show that the roots of the disproportions between deterministic and fuzzy quantities follow from the combined effect of the used concepts of fuzzy zero and equality between fuzzy quantities.

Let us define *modal fuzzy zero* as any fuzzy quantity $s \in Q(R)$ for which

$$\mu_s(x) = \mu_s(-x) \quad \text{for all } x \in R, \text{ and } \mu_s(0) = 1. \tag{17}$$

It is easy to prove that for any $a \in Q(R)$, $a \oplus (-a)$ is modal fuzzy zero. On the other hand, if $s \in Q(R)$ is modal fuzzy zero then for $a \in Q(R)$ generally $a \oplus s$ is not equal to a (except the special case of $s = \langle 0 \rangle$). To achieve the validity of the zero-element-axiom, at least in a weakened form, we define the *additive equivalence relation* $\sim_\oplus$. If $a, b \in Q(R)$, $a \sim_\oplus b$, iff there exist modal fuzzy zeros s, s' such that

$$a \oplus s = b \oplus s'. \tag{18}$$

Then it is easy to show that for any $a, b, c \in Q(R)$ and modal fuzzy zero s,

$$a \oplus b \sim_\oplus b \oplus a, \qquad a \oplus (b \oplus c) \sim_\oplus (a \oplus b) \oplus c,$$
$$a \oplus s \sim_\oplus a, \quad \text{and} \quad a \oplus (-a) \sim_\oplus s.$$

In this weakened sense, the group properties are fulfilled. Moreover, for $a, b \in Q(R)$ and $r \in R$

$$(r \cdot a) \oplus (r \cdot b) \sim_\oplus r \cdot (a \oplus b).$$

The opposite distributivity is not generally fulfilled but it is true for an important class of fuzzy quantities. Namely, we say that $a \in Q(R)$ is *almost trapezoidal* iff there exists trapezoidal b such that $a \sim_{\oplus} b$. Then for any almost trapezoidal $a \in Q(R)$ and any $r_1, r_2 \in R$

$$(r_1 \cdot a) \oplus (r_2 \cdot a) \sim_{\oplus} (r_1 + r_2) \cdot a.$$

Observation 3. Fuzzy quantity $a \in Q(R)$ is almost trapezoidal iff there exist trapezoidal $b \in Q(R)$ and modal fuzzy zero $s \in Q(R)$ such that $a = s \oplus b$.

The above concept of the modal fuzzy zero is evidently natural for representing the zero among fuzzy quantities.

Observation 4. For any $a \in Q(R)$ and any modal fuzzy zero s, the fuzzy product $a \odot s$ is also modal fuzzy zero.

Observation 5. As shown in [7, 8], any symmetric fuzzy quantity $t \in Q(R)$ such that $\mu_t(x) = \mu_t(-x)$ for all $x \in R$ displays all the fuzzy-zero properties mentioned above (except the general zero-modality $\mu_t(0) = 1$). It means that such fuzzy quantities can be considered for fuzzy zeros or, vice-versa, for simple fuzzy quantities with interesting property of symmetry – due to the specificity of the used model.

6. PROBLEM OF CUMMULATING VAGUENESS

The computations based on the extension principle combine the uncertainties of particular fuzzy quantities, and cummulate them into the uncertainty of the result.

In most frequented cases, it means the growth of the uncertainty of results.

Let us denote for $a \in Q(R)$ the real numbers

$$\inf(a) = \inf\left(x \in R : \mu_a(x) > 0\right), \quad \sup(a) = \sup\left(x \in R : \mu_a(x) > 0\right),$$
$$\mathrm{var}(a) = \sup(a) - \inf(a).$$

Observation 6. Let $a, b \in Q(R)$ and $r \in R$. Then it is easy to see that

$$\begin{aligned}
&\sup(a \oplus b) = \sup(a) + \sup(b),\\
&\inf(a \oplus b) = \inf(a) + \inf(b),\\
&\mathrm{var}(a) + \mathrm{var}(\mathrm{b}) \geq \mathrm{var}(a \oplus b) \geq \max(\mathrm{var}(a), \mathrm{var}(b)),\\
&\sup(r \cdot a) = r \cdot \sup a, \quad \inf(r \cdot a) = r \cdot \inf(a),\\
&\mathrm{var}(r \cdot a) = r \cdot \mathrm{var}(a)\\
&\sup(-a) = -\inf(a), \quad \inf(-a) = -\sup(a), \quad \mathrm{var}(-a) = \mathrm{var}(a),\\
&\sup(a \odot b) = \max\Big[\sup(a) \cdot \sup(b), \sup(a) \cdot \inf(b),\\
&\qquad\qquad \inf(a) \cdot \inf(b), \inf(a) \cdot \sup(b)\Big],\\
&\inf(a \odot b) = \min\Big[\sup(a) \cdot \sup(b), \sup(a) \cdot \inf(b),
\end{aligned}$$

$$\inf(a)\cdot\inf(b),\ \inf(a)\cdot\sup(b)\Big],$$
$$\inf(a^2) = \min\left[(\inf(a))^2, (\sup(a))^2\right],$$
$$\sup(a^2) = \max\left[(\inf(a))^2, (\sup(a))^2\right].$$

The characteristics var(a) represents the extent of possible values of a, in other words the extent of uncertainty connected with a. Especially, the uncertainty of a sum of many fuzzy quantities may be so large that it practically eliminates the informational value of the possible values – the extent of possible values may be much larger than the significant (let us say, modal) value of the resulting fuzzy quantity. This discrepancy cannot be eliminated by the choice of the procedure applying the extension principle, e. g., by the integration of several algebraical operations (summations, e. g.) in one function.

Observation 7. Let $a_1, a_2, \ldots, a_n \in Q(R)$, let $f : R^n \to R$ be such that $f(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \cdots + x_n$ and let

$$a = f(a_1, a_2, \ldots, a_n),$$

in accordance with the extension principle (10). Let $\overline{a} \in Q(R)$ be such that

$$\overline{a} = a_1 \oplus a_2 \oplus \cdots \oplus a_n,$$

where (11) was used. Then $a = \overline{a}$ and, consequently, $\mathrm{var}(a) = \mathrm{var}(\overline{a})$.

The consequences of the relations summarized in Observation 6 means that the enormous growth of uncertainty regards rather economic (and econometric) cummulative factors (like the durations and flows in the Critical Path Methods, aggregated quantities, etc.). Many significant statistical characteristics, like the mean value or other moments, are influenced much less.

7. DECOMPOSITION PRINCIPLE

The extension principle and methods based on it are natural and prevailing in the theory of fuzzy quantities. Nevertheless, an alternative approach was suggested and analyzed in several papers, e. g. in [9], where it is summarized.

It follows from the idea that a fuzzy quantity is, in fact, a hybrid concept composed from three components. It combines a numerical quantitative value, the uncertainty included in the verbal formulation of the quantity, and the subjective uncertainty following from the reliability of the source of fuzzy data. Each of these components can (and would) be processes separately by means of its specific tools, and the results can be recombined in the construction in the outcome fuzzy quantity.

The previous heuristic description can be formalized as follows. Let us consider a fuzzy quantity a with membership function μ_a. Then we decompose the function μ_a into

— real number $x_a \in R$ which is called the *deterministic core* of a,

— real, continuous and increasing function $f : a : R \to R$ such that $f_a(0) = 0$, called the *scale* of a,

— real function $\varphi_a : R \to [0,1]$, such that $\varphi_a(0) = 1$, called *shape* of a,

such that

$$\mu_a(x) = \varphi_a\left(f_a(x) - f_a(x_a)\right), \quad x \in R. \tag{19}$$

We say that a is *characterized* by the triple (x_a, f_a, φ_a), and obviously φ_a is in some sense a "normalized" structure of the membership, meanwhile f_a (or its gradient) represents the extent of uncertainty of the source of data a and x_a is its modal value.

If a, $b \in Q(R)$ are characterized by the triples (x_a, f_a, φ_a), (x_b, f_b, φ_b), respectively, then, computing an algebraic operation with them (e. g., their sum $a + b$ or fuzzy product) we may proceed separately the crisp cores x_a, x_b (usually by classical algebraic operations), shapes φ_a, φ_b may be adequately processed rather by "fuzzy logical" operations (like maximum, minimum) or by suitable functional operations, and the scales $f_a \cdot f_b$ may be processed by functional operations, described, e. g., in [10]. In all cases, the choice of actual operations may be harmonized with the type and demands of the actual application.

Doing so, we compute the core x_{a+b}, shape φ_{a+b} and scale f_{a+b}, and by (19) we construct the membership function μ_{a+b} of the resulting fuzzy quantity $a + b$.

This approach stresses the essential subjectivity of fuzzy set theoretical models. On the other hand it may avoid some disadvantages of more rigid procedures of the extensional principle. However, it extends the choice of methods being for disposal whenever fuzzy quantities are to be processed. The applicability of this approach for the extension of classical statistical methods to the environment of fuzzy quantities is still rather discutable – the variability of the possible analytic results may usually contradict with the sophisticated structure of statistical description and analysis of reality. Nevertheless, deeper knowledge of the properties of these concepts, as well as the standardization of some of the above choices, can open the way to more reliable applications even of the above model.

8. CONCLUSIVE REMARKS

The extension of the classical statistical methods to the environment of fuzzy input data is not a mechanical action consisting of a few formal changes in definitions. The previous chapters aimed to summarize, namely, the tools for the fuzzy quantities processing, and to point at their properties (or relevant sources in the literature). It is useful to note that some significant steps were already done (see, e. g., [4]). Nevertheless, some new ways to the adequate statistical processing became open in the last years.

The alternative model of fuzzy quantities based on their decomposition and separate processing was mentioned in the previous Chapter 7.

An interesting approach to fuzziness which is close to probabilistic model was done in [11], and it may be applicable even for statistical problems.

Very wide new way is open by strong generalization of the concept of fuzzy quantities, based on the notions of triangular norms, aggregation operators, binary copulas and related models (see, e. g., [5, 6] and also [2]). The degree of generalization used in these views on data processing offers a unitary approach to some probabilistic, fuzzistic, and some other methods, which could be used for the study of the mutual transfer of their particular procedures. Deeper analysis of these topics farly extends the acceptable form of a journal paper.

ACKNOWLEDGEMENT

The research whose results are summarized here was partly summarized by Grant Agency of Czech Republic, grant No. 402/04/1026, by Grant Agency of the Academy of Sciences of CR, grant No. A 1075301, and by the Centre of Applied Research DAR (project of the Ministry of Education of CR, No. 1M0572).

REFERENCES

[1] D. Dubois and H. Prade: Fuzzy numbers: An overview. In: Analysis of Fuzzy Information (J. C. Bezdek, ed.), Vol. 2, CRC-Press, Boca Raton 1988, pp. 3–39.

[2] D. Dubois, E. E. Kerre, R. Mesiar, and H. Prade: Fuzzy interval analysis. In: Fundamental of Fuzzy Sets, Vol. 1, Kluwer Acad. Publ, Dodrecht 2000, pp. 483–581.

[3] J. Kacprzyk and M. Fedrizi (eds.): Fuzzy Regression Analysis. Omnitech Press, Physica–Verlag, Warsaw 1992.

[4] E. E. Kerre and X. Wang: Reasonable properties for the ordering of fuzzy quantities. Part I, Part II. Fuzzy Sets and Systems *118* (2001), 375–385, 387–405.

[5] E. P. Klement, R. Mesiar, and E. Pap: Triangular Norms. Kluwer Acad. Publ., Dordrecht 2000.

[6] T. Calvo, G. Mayor, and R. Mesiar (eds.): Aggregation Operators. Physica–Verlag, Heidelberg 2002.

[7] M. Mareš: Computation Over Fuzzy quantities CRC–Press, Boca Raton 1994.

[8] M. Mareš: Weak arithmetics of fuzzy numbers. Fuzzy Sets and Systems *91* (1997), 2, 143–154.

[9] M. Mareš and R. Mesiar: Verbally generated fuzzy quantities and their aggregation. In: Analysis of Fuzzy Information (J. C. Bezdek, ed.), Vol. 2, CRC-Press, Boca Raton 1988, pp. 291–3352.

[10] M. Mareš and R. Mesiar: Dual meaning of verbal quantities. Kybernetika *38* (2002), 6, 709–716.

[11] Tran Quoc Chien: Medium distances of probability fuzzy-points and an application to linear programming. Kybernetika *25* (1989), 6, 494–504.

[12] L. A. Zadeh: Fuzzy sets. Inform. and Control *8* (1965), 3, 338–353.

Milan Mareš, Institute of Information Theory and Automation — Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 09 Praha 8. Czech Republic.
e-mail: mares@utia.cas.cz