

Fuzziness and Roughness of Non-precise Quantities

Slavka Bodjanova

Texas A&M University-Kingsville, Kingsville, U.S.A.

Abstract: Connections between fuzzy sets and rough sets are studied in description of non-precise quantities. Characterization of empirical distribution of non-precise observations in non-precise categories is proposed.

Zusammenfassung: Die Beziehung zwischen unscharfen Mengen und rauen Mengen wird für die Beschreibung von unscharfen Größen untersucht. Ferner werden Verteilungsfunktionen für unscharfe Daten angegeben.

Keywords: Fuzzy Sets, Rough Sets, Non-Precise Data, Empirical Distribution, Frequency Function.

1 Introduction

Many problems arising in scientific investigation generate non-precise data incorporating non-statistical uncertainty. A non-precise observation of a quantitative variable can be described by a special type of membership function defined on the set of all real numbers called a fuzzy number or a fuzzy interval. This membership function is a fuzzy set. Fuzzy set theory originated by Zadeh (1965) relies on ordering relations that express intensity (degree) of membership of an object in a set. Several researchers studied applications of statistical methods to fuzzy data. Among them, for example, Kruse and Meyer (1987), Kacprzyk and Federizzi (1988), Frühwirth-Schnatter (1992) and Römer and Kandel (1995). Viertl (1996) provided comprehensive guidelines for exploratory analysis and statistical inference of non-precise data.

Pawlak (1982) has proposed rough set methodology as a new approach in handling analysis of non-precise concepts. In this methodology any non-precise concept is characterized by a pair of precise concepts called the lower and the upper approximations. Rough set theory is based on equivalence relations describing partitions made of classes of indiscernible objects. This new approach proved to be useful in many applications (see Slowinski, 1992; Ziarko, 1994; Polkowski and Skowron, 1998).

The purpose of data analysis is to gain information from data. The concept of information is connected with the concept of uncertainty. Information may be incomplete, not fully reliable, vague or deficient in other ways. These various information deficiencies result in different types of uncertainty (see Klir and Yuan, 1995). Rough sets and fuzzy sets are two distinct models of imperfect knowledge. Indiscernibility, studied by the theory of rough sets, refers to the granularity (coarseness) of knowledge. Vagueness, studied by the theory of fuzzy sets, is due to the fact that categories of natural language are usually sets with smooth (not crisp) boundaries. Many measures of uncertainty have been developed in order to quantify the amount of information in non-precise data. It has been proven in numerous ways that the sensible measure of uncertainty in probability theory is the Shannon entropy (see Shannon, 1948). Nonprobabilistic entropies and indetermination measures in the setting of fuzzy sets theory were proposed by Empotz (1981). A

review of measures of fuzziness can be found in Klir (1995). Pawlak (1991) introduced a measure of roughness of a non-precise concept in an approximation space. Recently, research articles combining fuzziness and roughness in decision making applications have appeared in Pal and Skowron (1999).

This paper shows applications of techniques from the theory of fuzzy sets and the theory of rough sets to basic analysis of non-precise data. Each non-precise observation described by a fuzzy set can be viewed as a non-precise quantity in the approximation space (\mathbb{R}, I) where \mathbb{R} is the set of all real numbers and I is the identity relation on $\mathbb{R} \times \mathbb{R}$. Then the following questions arise: How fuzzy and how rough is the non-precise quantity? Is there a meaningful way of reducing its fuzziness and roughness?

Suppose that a finite number n of non-precise observations of a quantitative random variable X is given. These observations create a non-precise sample \mathcal{S} . One of the basic tasks in analysis of empirical data is to group observations from \mathcal{S} into a few conveniently designated categories (intervals) defined on the domain of X . For example, categories $C_1 = \text{small values}$, $C_2 = \text{medium values}$ and $C_3 = \text{large values}$. Each category represents a vague concept and therefore should be characterized by a fuzzy interval. This leads to the following questions: How to describe empirical distribution of non-precise observations in non-precise categories? How to derive a non-precise frequency function?

A notion of fuzzy frequency function in the case of precise categories and non-precise observations was introduced by Viertl (1996). His work can be generalized to non-precise categories. A non-precise frequency function is defined as a fuzzy set on the set $\mathbb{N}_n = \{1, \dots, n\}$. From the point of view of rough sets, this function is a non-precise concept in the approximation space (\mathbb{N}_n, I) where I is the identity relation on $\mathbb{N}_n \times \mathbb{N}_n$. One can ask: How fuzzy and how rough is the non-precise frequency function?

This paper answers the questions posted above using some basic techniques from the theory of rough sets and the theory of fuzzy sets. The necessary background for both theories is given in Section 2. Section 3 deals with roughness and fuzziness of non-precise quantities. For $\alpha \in (0, 1]$ measure of α -fuzziness and measure of α -roughness are introduced. It is shown that imprecision of a fuzzy quantity can be reduced by its approximation by an α -sharper version called a generalized α -cut. Section 4 presents a method for calculating non-precise counts of non-precise observations from a sample \mathcal{S} in non-precise categories from a family \mathcal{C} . The conditions under which \mathcal{C} becomes a fuzzy cover, fuzzy partition, or a weak fuzzy partition of the range of \mathcal{S} are explained. Then a quantitative evaluation of fuzziness and roughness of empirical distribution of non-precise data is given. The proposed techniques are illustrated on a small sample of non-precise observations in Section 5.

The goal of this study is twofold. Firstly, it aims to show connections between fuzzy sets and rough sets in description of non-precise quantities. Secondly, it provides some basic tools for exploratory analysis of non-precise data.

2 Theoretical Background

2.1 Fuzzy Sets

Let X be a set of objects. A fuzzy set A on X is defined by membership function

$$A : X \rightarrow [0, 1]. \quad (1)$$

A crisp set is a special case of fuzzy set in which the membership function is restricted to $\{0, 1\}$. The largest membership coefficient in a given fuzzy set A is called its height and is denoted by h_A . When $h_A = 1$, A is called a normal fuzzy set, otherwise, it is called subnormal. Given a number $\alpha \in (0, 1]$, an α -cut of fuzzy set A is defined for all $x \in X$ by

$$A_\alpha(x) = \begin{cases} 1 & \text{if } A(x) \geq \alpha, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

A fuzzy set A can be reconstructed from its α -cuts as follows: for all $x \in X$:

$$A(x) = \sup_{\alpha \in (0, 1]} \alpha \cdot A_\alpha(x), \quad (3)$$

where sup denotes supremum.

The support and the core of a fuzzy set A , denoted by S_A and C_A respectively, are crisp subsets of X such that for all $x \in X$:

$$S_A(x) = \begin{cases} 1 & \text{if } A(x) > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

and

$$C_A(x) = \begin{cases} 1 & \text{if } A(x) = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

The standard complement, \overline{A} , of a fuzzy set A with respect to the set X is defined for all $x \in X$ by the equation

$$\overline{A}(x) = 1 - A(x). \quad (6)$$

Given two fuzzy sets, A and B , their standard intersection $A \cap B$, and standard union $A \cup B$, are defined for all $x \in X$ by

$$(A \cap B)(x) = \min\{A(x), B(x)\}, \quad (7)$$

and

$$(A \cup B)(x) = \max\{A(x), B(x)\}. \quad (8)$$

We say that fuzzy set A is a subset of fuzzy set B , denoted $A \subset B$, if $A(x) \leq B(x)$ for all $x \in X$.

For any fuzzy set A defined on a finite universal set X , its scalar cardinality $|A|$ is given by the formula

$$|A| = \sum_{x \in X} A(x). \quad (9)$$

When X is an infinite set and P is a measure on X , then its cardinality is defined by

$$|A| = \int_X A(x) dP(x). \quad (10)$$

Fuzzy sets that are defined on the set of real numbers \mathbb{R} are called fuzzy quantities. A fuzzy interval is any normal fuzzy quantity with bounded support whose α -cuts for all $\alpha \in (0, 1]$ are closed crisp interval of real numbers. A fuzzy number A is a special fuzzy interval for which $A(x) = 1$ for exactly one $x \in \mathbb{R}$.

Let a, b, c, d are real numbers such that $a \leq b \leq c \leq d$. Then a trapezoidal fuzzy interval A denoted by quadruple $[a, b, c, d]$ has the membership function

$$A(x) = \begin{cases} \frac{x-a}{b-a} & \text{if } a < x < b, \\ 1 & \text{if } b \leq x \leq c, \\ \frac{d-x}{d-c} & \text{if } c < x < d, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

The α -cut of A for $\alpha \in (0, 1)$ is

$$A_\alpha(x) = \begin{cases} 1 & \text{if } f^{-1}(\alpha) \leq x \leq g^{-1}(\alpha), \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

and for $\alpha = 1$

$$A_\alpha(x) = \begin{cases} 1 & \text{if } b \leq x \leq c, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

Figure 1 depicts trapezoidal fuzzy interval $[151, 161, 169, 179]$ (solid line) together with its α -cut at $\alpha = 0.6$ (dashed line).

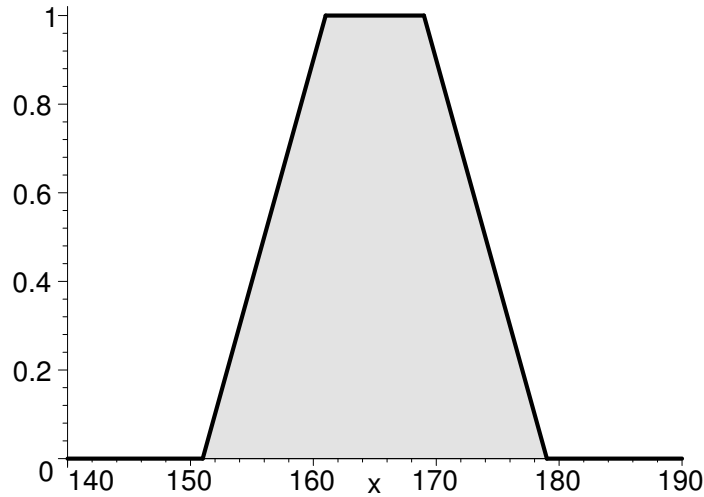


Figure 1: Trapezoidal fuzzy interval and its 0.6-cut

Some vague concepts can be represented by *half-open* trapezoidal fuzzy intervals. For example, the concept *very small values* is described by trapezoidal fuzzy interval A where $A(x) = 1$ for all $x \leq c$, and denoted by $(., c, d]$. Analogously, vague concept *very large*

values is represented by trapezoidal fuzzy interval A where $A(x) = 1$ for all $x \geq b$, and denoted by $[a, b, .)$. If $b = c$ in (12) we get a triangular fuzzy number characterized by the triple $[a, b, d]$.

For quantitative evaluation of the amount of uncertainty in a fuzzy set, a measure of fuzziness of a fuzzy set is needed. In general, a measure of fuzziness is a nonnegative real function ϕ defined on the set of all fuzzy subsets of X which satisfies the following properties:

1. $\phi(A) = 0$ if and only if A is a crisp set,
2. $\phi(A)$ attains its maximum if and only if $A(x) = 0.5$ for all $x \in X$,
3. $\phi(A) \leq \phi(B)$ when set A is sharper than set B , which means that $A(x) \leq B(x)$ when $B(x) \leq 0.5$ and $A(x) \geq B(x)$ when $B(x) \geq 0.5$ for all $x \in X$.

More information about fuzzy sets can be found in Dubois and Prade (1980) or Klir and Yuan (1995).

2.2 Rough Sets

Let U denote a nonempty universal set, and let $R \subset U \times U$ be an equivalence relation. The pair (U, R) is called a knowledge base or an approximation space. Given an arbitrary set $A \subset U$, it may not be possible to describe A precisely in the approximation space (U, R) . Instead, one may characterize A by a pair of lower and upper approximations defined as follows:

$$\underline{R}(A) = \{x \in U : [x]_R \subset A\}, \quad (14)$$

$$\overline{R}(A) = \{x \in U : [x]_R \cap A \neq \emptyset\}, \quad (15)$$

where $[x]_R$ is the equivalence class containing x . The pair $(\underline{R}(A), \overline{R}(A))$ is called the rough set with a reference set A .

A set A is called R -definable if and only if $\underline{R}(A) = \overline{R}(A)$. Otherwise A is rough with respect to R . The set $BN_R(A) = \overline{R}(A) - \underline{R}(A)$ is called the borderline region of A . The greater is the borderline region of a set the lower is the accuracy of approximation of the set in approximation space (U, R) . Pawlak (1991) introduced an accuracy measure as follows:

$$\gamma_R(A) = \frac{|\underline{R}(A)|}{|\overline{R}(A)|}. \quad (16)$$

Obviously, $0 \leq \gamma_R(A) \leq 1$. If $\gamma_R(A) = 1$, the borderline region of A is empty and the set A is R -definable. In order to express the degree of inexactness (roughness) of a set Pawlak (1991) suggested the measure

$$\rho_R(A) = 1 - \gamma_R(A) \quad (17)$$

and referred to it as a *roughness* of A .

A rough set A can be characterized by a single membership function

$$A(x) = \frac{|A \cap [x]_R|}{|[x]_R|}, \quad (18)$$

for all $x \in U$. It is obvious that $A(x) = 1$ if and only if $x \in \underline{R}(A)$ and $A(x) = 0$ if and only if $x \notin \overline{R}(A)$. An element $x \in U$ belongs to the borderline region $BN_R(A)$ if and only if $0 < A(x) < 1$.

Research on the theory of rough sets increases steadily and several extensions of Pawlak's classical rough sets theory have been developed (see Polkowski and Skowron, 1998).

3 Fuzziness and Roughness

Klir and Yuan (1995) suggested that fuzziness of a fuzzy set can be measured by distance between its membership function and the membership function (characteristic function) of the nearest crisp set. This idea can be generalized as follows:

Definition 1. Let A be a fuzzy set and let $\alpha \in (0, 1]$. Then α -fuzziness of A is given by

$$\phi_\alpha(A) = d(A, A_\alpha), \quad (19)$$

where A_α is α -cut of A and $d : [0, 1] \times [0, 1] \rightarrow [0, \infty)$ is a metric distance.

In this paper the Hamming distance, defined by the sum of absolute values of differences will be used.

Theorem 1. Let A be a trapezoidal fuzzy interval characterized by quadruple $[a, b, c, d]$. Then

$$\phi_\alpha(A) = \int_{\mathbb{R}} |A(x) - A_\alpha(x)| dx = \frac{(b-a) + (d-c)}{\alpha} (\alpha^2 + (1-\alpha)^2). \quad (20)$$

Proof. According to (13) A_α is interval of real numbers

$[f^{-1}(\alpha), g^{-1}(\alpha)] = [a + (b-a)\alpha, d - (d-c)\alpha] = [\gamma_1, \gamma_2]$. Then

$$\begin{aligned} \phi_\alpha(A) &= \int_{\mathbb{R}} |A(x) - A_\alpha(x)| dx = \int_a^{\gamma_1} \frac{x-a}{b-a} dx + \int_{\gamma_1}^b (1 - \frac{x-a}{b-a}) dx + \int_c^{\gamma_2} (1 - \frac{d-x}{d-c}) dx + \int_{\gamma_2}^d \frac{d-x}{d-c} dx = \\ &= \frac{(\gamma_1-a)\alpha}{2} + \frac{(b-\gamma_1)(1-\alpha)}{2} + \frac{(\gamma_2-c)(1-\alpha)}{2} + \frac{(d-\gamma_2)\alpha}{2} = \frac{(b-a)\alpha^2}{2} + \frac{(b-a)(1-\alpha)^2}{2} + \frac{(d-c)(1-\alpha)^2}{2} + \frac{(d-c)\alpha^2}{2} = \\ &= \frac{(b-a)+(d-c)}{2} (\alpha^2 + (1-\alpha)^2). \quad \square \end{aligned}$$

It is easy to verify the following properties:

1. $\phi_\alpha(A) = \phi_{1-\alpha}(A)$ for $\alpha \in [0.5, 1]$,
2. $\phi_{0.5} = \min_{\alpha \in (0,1]} \phi_\alpha(A)$.

Because in this case the nearest crisp set to a trapezoidal fuzzy interval A is its α -cut $A_{0.5}$ fuzziness of A is evaluated by

$$\phi(A) = \phi_{0.5}(A) = (b-a) + (d-c). \quad (21)$$

Dubois and Prade (1980) called the quantities $(b-a)$ and $(d-c)$ the left and the right spread of A , respectively. Then fuzziness of A calculated by (22) is actually the total spread of A . Fuzziness of a triangular fuzzy number $A = [a, b, d]$ is $\phi(A) = d-a = |S_A|$.

Assume approximation space (\mathbb{R}, I) where \mathbb{R} is the set of all real numbers and I is the identity relation on \mathbb{R} . Then, a fuzzy set A defined on \mathbb{R} can be characterized by a pair of lower and upper approximations

$$\underline{I}(A) = \{x \in \mathbb{R} : [x]_I \subset A\}, \quad (22)$$

$$\overline{I}(A) = \{x \in \mathbb{R} : [x]_I \cap A \neq \emptyset\}. \quad (23)$$

Because I is the identity relation, the membership functions (characteristic functions) of sets in (22) and (23) are

$$\underline{I}(A)(x) = \begin{cases} 1 & \text{if } A(x) = 1, \\ 0 & \text{otherwise,} \end{cases} \quad (24)$$

and

$$\overline{I}(A)(x) = \begin{cases} 1 & \text{if } A(x) > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (25)$$

respectively. Then $\underline{I}(A)$ is the core of A and $\overline{I}(A)$ is the support of A . According to (18) roughness of A in approximation space (\mathbb{R}, I) is given by

$$\rho_I(A) = 1 - \frac{|\underline{I}(A)|}{|\overline{I}(A)|} = 1 - \frac{|C_A|}{|S_A|}. \quad (26)$$

Measure of roughness of a fuzzy quantity can be generalized as follows.

Definition 2. Let A be a fuzzy quantity and let $\alpha \in (0, 1]$. Then α -roughness of A in approximation space (\mathbb{R}, I) is given by

$$\rho_{I_\alpha}(A) = 1 - \frac{|C_{A_\alpha}|}{|S_A|}, \quad (27)$$

where A_α is α -cut of A .

Because this paper will consider only the approximation space (\mathbb{R}, I) , the notation $\rho(A)$ instead of $\rho_I(A)$ and the notation $\rho_\alpha(A)$ instead of ρ_{I_α} will be used.

Theorem 2. Let A be a trapezoidal fuzzy interval characterized by quadruple $[a, b, c, d]$. Then

$$\rho_\alpha(A) = \alpha \left(1 - \frac{c - b}{d - a} \right). \quad (28)$$

Proof. A_α is a crisp set and therefore $C_{A_\alpha} = A_\alpha$. Because A is a trapezoidal fuzzy interval its α -cut is crisp interval of real numbers $A_\alpha = [a + (b - a)\alpha, d - (d - c)\alpha]$. Therefore $|C_{A_\alpha}| = d - (d - c)\alpha - (a + (b - a)\alpha)$.

Support of A is crisp interval of real numbers $S_A = (a, d)$ and $|S_A| = d - a$. Then $\rho_\alpha(A) = \frac{d - a - [(d - a) - (d - c)\alpha - (b - a)\alpha]}{d - a} = \frac{(b - a)\alpha + (d - c)\alpha}{d - a} = \frac{(d - a)\alpha + (b - c)\alpha}{d - a} = \alpha \left(1 - \frac{b - c}{d - a} \right)$. \square

It is easy to verify the following properties:

1. $\rho_{1-\alpha}(A) = \frac{1-\alpha}{\alpha} \rho_\alpha(A)$ for $\alpha \in [0.5, 1)$,
2. $\rho_{\alpha_1}(A) \leq \rho_{\alpha_2}(A)$ if $\alpha_1 \leq \alpha_2$.

Because $C_{A_\alpha} = A_\alpha$ for $\alpha = 1$, roughness of a trapezoidal fuzzy interval A is evaluated by

$$\rho(A) = \rho_1(A) = 1 - \frac{c - b}{d - a}. \quad (29)$$

Roughness of a triangular fuzzy number $A = [a, b, d]$ is $\rho(A) = 1$.

Measures of roughness and fuzziness of a trapezoidal fuzzy interval $A = [a, b, c, d]$ are related. It is easy to verify that

$$\rho(A) = \frac{\phi(A)}{d - a} = \frac{\phi(A)}{|S_A|}. \quad (30)$$

Let \mathcal{A} be a family of n fuzzy quantities A_1, \dots, A_n . Then fuzziness and roughness of \mathcal{A} are evaluated by

$$\phi(\mathcal{A}) = \frac{1}{n} \sum_{i=1}^n \phi(A_i) \quad (31)$$

and

$$\rho(\mathcal{A}) = \frac{1}{n} \sum_{i=1}^n \rho(A_i). \quad (32)$$

Values of $\phi_\alpha(A)$ and $\rho_\alpha(A)$ provide information about the error occurred when a fuzzy quantity A is approximated by its α -cut A_α . This approximation separates coefficients $A(x)$ as follows: *Large coefficients* if $A(x) \geq \alpha$, and *small coefficients* if $A(x) < \alpha$. Because A_α is a crisp set, $\rho_\alpha(A_\alpha) = \phi_\alpha(A_\alpha) = 0$. This approximation might be too restrictive, especially when the membership function of a fuzzy set is a continuous function. For example, if $\alpha = 0.6$, then the coefficient $A(x_1) = 0.61$ is considered to be large and the coefficient $A(x_2) = 0.59$ small. For a smooth transition between large and small coefficients in approximation of a fuzzy quantity A by a quantity with lower roughness and fuzziness, a generalized α -cut of A is proposed.

Definition 3. Let A be a fuzzy quantity and let $\alpha \in (0, 1]$. A generalized α -cut of A is a trapezoidal fuzzy interval $\mathcal{G}_\alpha(A)$ which satisfies the following properties:

If $\alpha = 0.5$ then $\mathcal{G}_\alpha(A) = A_\alpha$,

if $\alpha > 0.5$ then

$$\mathcal{G}_\alpha(A)(x) \geq A(x) \quad \text{when } A(x) \geq \alpha, \quad (33)$$

$$\mathcal{G}_\alpha(A)(x) \leq A(x) \quad \text{when } 1 - \alpha \leq A(x) < \alpha, \quad (34)$$

$$\mathcal{G}_\alpha(A)(x) = 0 \quad \text{when } A(x) < 1 - \alpha, \quad (35)$$

if $\alpha < 0.5$ then

$$\mathcal{G}_\alpha(A)(x) \leq A(x) \quad \text{when } A(x) < \alpha, \quad (36)$$

$$\mathcal{G}_\alpha(A)(x) \geq A(x) \quad \text{when } \alpha \leq A(x) < 1 - \alpha, \quad (37)$$

$$\mathcal{G}_\alpha(A)(x) = 1 \quad \text{when } A(x) \geq 1 - \alpha. \quad (38)$$

Theorem 3. Let A be a fuzzy quantity and let A_α , $\mathcal{G}_\alpha(A)$ be α -cut and generalized α -cut of A , respectively. Then $\phi_\alpha(\mathcal{G}_\alpha(A)) \leq \phi_\alpha(A)$ and $\rho_\alpha(\mathcal{G}_\alpha(A)) \leq \rho_\alpha(A)$.

Proof. From Definition 3 follows that $(\mathcal{G}_\alpha(A))_\alpha = \{x : \mathcal{G}_\alpha(A)(x) \geq \alpha\} = \{x : A(x) \geq \alpha\} = A_\alpha$. Therefore $\phi_\alpha(\mathcal{G}_\alpha(A)) = d(\mathcal{G}_\alpha(A), A_\alpha)$.

For $A(x) \geq \alpha$ we have: $A(x) \leq \mathcal{G}_\alpha(A)(x) \leq A_\alpha(x) = 1$. For $A(x) < \alpha$ we have: $0 = A_\alpha(x) \leq \mathcal{G}_\alpha(A)(x) \leq A(x)$. Therefore $d(\mathcal{G}_\alpha(A), A_\alpha) \leq d(A, A_\alpha)$, which implies that $\phi_\alpha(\mathcal{G}_\alpha(A)) \leq \phi_\alpha(A)$.

From Definition 3 follows that if $\mathcal{G}_\alpha(A)(x) > 0$ then $A(x) > 0$. Therefore support of $\mathcal{G}_\alpha(A)$ is a subset of support of A . Then $\rho_\alpha(\mathcal{G}_\alpha(A)) = 1 - \frac{|(\mathcal{G}_\alpha(A))_\alpha|}{|S_{\mathcal{G}_\alpha(A)}|} \leq 1 - \frac{|A_\alpha|}{|S_A|} = \rho_\alpha(A)$. \square

Theorem 4. Let A be a trapezoidal fuzzy interval characterized by quadruple $[a, b, c, d]$. Then generalized α -cut of A is a trapezoidal fuzzy interval $\mathcal{G}_\alpha(A)$ characterized by quadruple $[a_1, b_1, c_1, d_1]$, where for $\alpha \in [0.5, 1]$:

$$a_1 = a + (b - a)(1 - \alpha), \quad (39)$$

$$b_1 = a + \frac{(b - a)(\alpha - (1 - \alpha)^2)}{\alpha}, \quad (40)$$

$$c_1 = d - \frac{(d - c)(\alpha - (1 - \alpha)^2)}{\alpha}, \quad (41)$$

$$d_1 = d - (d - c)(1 - \alpha), \quad (42)$$

and for $\alpha \in (0, 0.5]$:

$$a_1 = a + \frac{(b - a)\alpha^2}{1 - \alpha}, \quad (43)$$

$$b_1 = a + (b - a)(1 - \alpha), \quad (44)$$

$$c_1 = d - (d - c)(1 - \alpha), \quad (45)$$

$$d_1 = d - \frac{(d - c)\alpha^2}{1 - \alpha}. \quad (46)$$

Proof. Let $\alpha \in [0.5, 1]$. Then, according to (33) – (35), support of $\mathcal{G}_\alpha(A)$ is equal to $(1 - \alpha)$ -cut of A . Therefore

$S_{\mathcal{G}_\alpha(A)} = (a_1, d_1) = (a + (b - a)(1 - \alpha), d - (d - c)(1 - \alpha)) = A_{1-\alpha}$, which means that $a_1 = a + (b - a)(1 - \alpha)$ and $d_1 = d - (d - c)(1 - \alpha)$.

Because α -cut of $\mathcal{G}_\alpha(A)$ is the same as α -cut of A , we have that $\mathcal{G}_\alpha(A)(a + (b - a)\alpha) = \alpha$ and $\mathcal{G}_\alpha(A)(d - (d - c)\alpha) = \alpha$.

Point $[b_1, \mathcal{G}_\alpha(A)(b_1)] = [b_1, 1]$ is on the line determined by points $[a_1, 0]$ and $[a + (b - a)\alpha, \alpha]$. Point $[c_1, \mathcal{G}_\alpha(A)(c_1)] = [c_1, 1]$ is on the line determined by points $[d_1, 0]$ and $[d - (d - c)\alpha, \alpha]$. Then

$$1 - 0 = \frac{\alpha - 0}{a + (b - a)\alpha - a_1}(b_1 - a_1)$$

and

$$b_1 = \frac{a + (b - a)\alpha - a_1}{\alpha} + a_1.$$

After some algebraic manipulations we get (40).

Analogously,

$$1 - 0 = \frac{\alpha - 0}{d - (d - c)\alpha - d_1}(c_1 - d_1)$$

and

$$c_1 = \frac{d - (d - c)\alpha - d_1}{\alpha} + d_1.$$

After some algebraic manipulations we get (41).

Let $\alpha \in (0, 0.5]$. Then according to (36) – (38) core of $\mathcal{G}_\alpha(A)$ is equal to $(1 - \alpha)$ -cut of A . Therefore

$C_{\mathcal{G}_\alpha(A)} = [b_1, c_1] = [a + (b - a)(1 - \alpha), d - (d - c)(1 - \alpha)] = A_{1-\alpha}$, which means that $b_1 = a + (b - a)(1 - \alpha)$ and $c_1 = d - (d - c)(1 - \alpha)$.

Because α -cut of $\mathcal{G}_\alpha(A)$ is the same as α -cut of A , we have that $\mathcal{G}_\alpha(A)(a + (b - a)\alpha) = \alpha$ and $\mathcal{G}_\alpha(A)(d - (d - c)\alpha) = \alpha$.

Point $[a_1, \mathcal{G}_\alpha(A)(a_1)] = [a_1, 0]$ is on the line determined by points $[b_1, 1]$ and $[a + (b - a)\alpha, \alpha]$. Point $[d_1, \mathcal{G}_\alpha(A)(d_1)] = [d_1, 0]$ is on the line determined by points $[c_1, 1]$ and $[d - (d - c)\alpha, \alpha]$. Then

$$0 - 1 = \frac{1 - \alpha}{b_1 - (a + (b - a)\alpha)}(a_1 - b_1)$$

and

$$a_1 = \frac{(a + (b - a)\alpha - b_1)}{1 - \alpha} + b_1.$$

After some algebraic manipulations we get (43).

Analogously,

$$0 - 1 = \frac{1 - \alpha}{c_1 - (d - (d - c)\alpha)}(d_1 - c_1)$$

and

$$d_1 = \frac{(d - (d - c)\alpha - c_1)}{1 - \alpha} + c_1.$$

After some algebraic manipulations we get (46).

If we use $\alpha = 0.5$ in (39) – (46) we get that $a_1 = b_1 = a + (b - a)\alpha$ and $c_1 = d_1 = d - d(d - c)\alpha$. This verifies that $\mathcal{G}_\alpha(A) = A_\alpha$ for $\alpha = 0.5$.

□

Generalized α -cut of a fuzzy quantity A is a fuzzification of its crisp α -cut A_α . Fuzzy quantity $\mathcal{G}_\alpha(A)$ is α -sharper than A , because $\mathcal{G}_\alpha(A)(x) \leq A(x)$ for $A(x) < \alpha$ and $\mathcal{G}_\alpha(A)(x) \geq A(x)$ for $A(x) \geq \alpha$. Generalized α -cut and ordinary α -cut are the same if and only if $\alpha = 0.5$. Because $\mathcal{G}_\alpha(A)$ is a fuzzy set for $\alpha \neq 0.5$, one can obtain its generalized α -cut $\mathcal{G}_\alpha(\mathcal{G}_\alpha(A))$. This procedure can be repeated. Each successive approximation leads to an α -sharper (less α -fuzzy and less α -rough) non-precise quantity. The limit case is the crisp α -cut A_α .

Theorem 5. Let $\mathcal{G}_\alpha(\mathcal{G}_\alpha(A)) = (\mathcal{G}_\alpha(A))^2$, $\mathcal{G}_\alpha(\mathcal{G}_\alpha(\mathcal{G}_\alpha(A))) = (\mathcal{G}_\alpha(A))^3$, etc.

Then $\lim_{n \rightarrow \infty} (\mathcal{G}_\alpha(A))^n = A_\alpha$.

Proof is in Bodjanova (1999b).

Figure 2 depicts fuzzy interval $[151, 161, 169, 179]$ (solid line) together with its generalized α -cut at $\alpha = 0.6$ (dashed line).

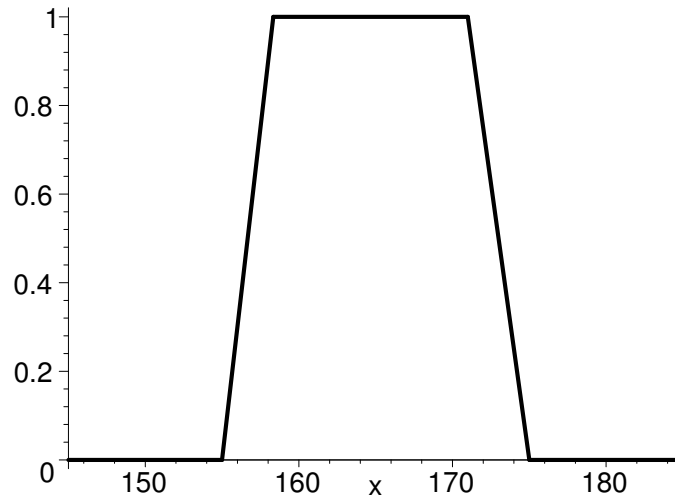


Figure 2: Trapezoidal fuzzy interval and its generalized 0.6-cut

4 Distribution of Non-Precise Quantities

Assume a sample \mathcal{S} of n non-precise observations of a continuous random variable X . Suppose that all non-precise observations are described by fuzzy numbers X_1, \dots, X_n . For each $i \in \mathbb{N}_n = \{1, \dots, n\}$ support of X_i is interval of real numbers (a_i, b_i) . Let $a_S = \min_i \{a_i\}$ and $b_S = \max_i \{b_i\}$. The interval $R_S = [a_S, b_S]$ will be called the range of sample \mathcal{S} . Range R_S can be divided in k non-precise categories ($2 \leq k < n$) described by vague linguistic expressions, and characterized by fuzzy intervals. For example, $C_1 = \text{very small values}$, $C_2 = \text{small values}$, $C_3 = \text{medium values}$, $C_4 = \text{large values}$ and $C_5 = \text{very large values}$.

Let $\mathcal{C} = \{C_j\}_{j=1}^k$ be a family of non-precise categories defined on R_S . Then

1. If for each $x \in R_S$ there is a $C_j \in \mathcal{C}$ such that $C_j(x) > 0$, family \mathcal{C} is called a *fuzzy cover* of R_S .
2. If for each $x \in R_S$: $0 < \sum_{j=1}^k C_j(x) \leq 1$, family \mathcal{C} is called a *weak fuzzy partition* of R_S .
3. If for each $x \in R_S$: $\sum_{j=1}^k C_j(x) = 1$, family \mathcal{C} is called a *fuzzy partition* of R_S .

Assignment of non-precise observations (fuzzy numbers) to non-precise categories (fuzzy intervals) is based on degree of inclusion of fuzzy sets. There are numerous inclusion grades in literature (see Dubois and Prade, 1980). This paper uses the degree of

inclusion given by

$$\epsilon(X_i, C_j) = \frac{|X_i \cap C_j|}{|X_i|}. \quad (47)$$

Because X_i is a fuzzy number and C_j is a fuzzy interval, we get

$$\epsilon(X_i, C_j) = \frac{\int_{\mathbb{R}} \min\{X_i(x), C_j(x)\} dx}{\int_{\mathbb{R}} X_i(x) dx}. \quad (48)$$

Comparison of (47) with (18) leads to the conclusion that $\epsilon(X_i, C_j)$ is rough membership coefficient $C_j(X_i)$ in approximation space (\mathcal{S}, I) , where I is the identity relation on $\mathcal{S} \times \mathcal{S}$. Therefore, each non-precise category C_j can be considered a rough set in (\mathcal{S}, I) defined by the following lower and upper approximations:

$$\underline{I}(C_j) = \{X_i \in \mathcal{S} : \epsilon(X_i, C_j) = 1\}, \quad (49)$$

$$\overline{I}(C_j) = \{X_i \in \mathcal{S} : \epsilon(X_i, C_j) > 0\}. \quad (50)$$

$\underline{I}(C_j)$ includes the elements from sample \mathcal{S} which for sure belong to C_j and $\overline{I}(C_j)$ includes the elements which possibly belong to C_j . Roughness of approximation of C_j in (\mathcal{S}, I) can be interpreted as roughness of distribution of elements from \mathcal{S} in category C_j and measured by

$$\rho(\mathcal{S}, C_j) = 1 - \frac{|\underline{I}(C_j)|}{|\overline{I}(C_j)|}. \quad (51)$$

Then the roughness of distribution of elements from sample \mathcal{S} in family of categories \mathcal{C} is

$$\rho(\mathcal{S}, \mathcal{C}) = \frac{1}{k} \sum_{j=1}^k \rho(\mathcal{S}, C_j). \quad (52)$$

The distribution is called *exact* if and only if $\rho(\mathcal{S}, \mathcal{C}) = 0$. Otherwise the distribution is *rough* or *non-exact*.

Let E_j be a fuzzy set defined on \mathcal{S} such that $E_j(X_i) = \epsilon(X_i, C_j)$. Then, $\underline{I}(C_j)$ given by (49) is the core of E_j , and $\overline{I}(C_j)$ given by (50) is the support of E_j . For $\alpha \in (0, 1]$ the α -cut of E_j is

$$(E_j)_\alpha = \{X_i \in \mathcal{S} : \epsilon(X_i, C_j) \geq \alpha\}. \quad (53)$$

Therefore α -fuzziness of distribution of elements from sample \mathcal{S} in category C_j can be evaluated by α -fuzziness of E_j . Hence

$$\phi_\alpha(\mathcal{S}, C_j) = d(E_j, (E_j)_\alpha), \quad (54)$$

and α -fuzziness of distribution of elements from sample \mathcal{S} in family of categories \mathcal{C} is

$$\phi_\alpha(\mathcal{S}, \mathcal{C}) = \frac{1}{k} \sum_{j=1}^k \phi_\alpha(\mathcal{S}, C_j). \quad (55)$$

Fuzziness of distribution of elements from \mathcal{S} in family \mathcal{C} is

$$\phi(\mathcal{S}, \mathcal{C}) = \phi_{0.5}(\mathcal{S}, \mathcal{C}). \quad (56)$$

Analogously, α -roughness of elements from sample \mathcal{S} in category C_j can be evaluated by α -roughness of E_j . Hence

$$\rho_\alpha(\mathcal{S}, C_j) = 1 - \frac{|(E_j)_\alpha|}{|S_{E_j}|}, \quad (57)$$

and α -roughness of distribution of elements from sample \mathcal{S} in family of categories \mathcal{C} is

$$\rho_\alpha(\mathcal{S}, \mathcal{C}) = \frac{1}{k} \sum_{j=1}^k \rho_\alpha(\mathcal{S}, C_j). \quad (58)$$

Another way of evaluation of roughness and fuzziness of empirical distribution of non-precise data is based on evaluation of a fuzzy frequency function. A notion of fuzzy frequency function in the case of precise categories and non-precise observations was introduced by Viertl (1996) and generalized by Bodjanova (1999a) as follows:

Definition 4. Let \mathcal{S} be a sample of n non-precise observations (fuzzy quantities). Let C_j be a class (fuzzy interval) defined on the range of \mathcal{S} . Then the fuzzy frequency function $w_j : \mathbb{N}_n \rightarrow [0, 1]$ describing distribution of elements from \mathcal{S} in the class C_j is given for all $t \in \mathbb{N}_n$ by

$$w_j(t) = \max_{\Omega_t} \{ \min \{ \epsilon(X_{c(1)}, C_j), \epsilon(X_{c(2)}, C_j), \dots, \epsilon(X_{c(t)}, C_j) \} \}, \quad (59)$$

where Ω_t is the set of all possible combinations $[c(1), c(2), \dots, c(t)]$ of t elements from $\{1, 2, \dots, n\}$.

The value of $w_j(t)$ is interpreted as the degree of possibility that at least t elements from \mathcal{S} are included in C_j . It is obvious that w_j must be a nonincreasing function on \mathbb{N}_n . Then α -fuzziness of distribution of elements from sample \mathcal{S} in category C_j is given by

$$\phi_\alpha(\mathcal{S}, C_j) = d(w_j, (w_j)_\alpha), \quad (60)$$

and α -fuzziness of distribution of elements from sample \mathcal{S} in family of categories \mathcal{C} is

$$\phi_\alpha(\mathcal{S}, \mathcal{C}) = \frac{1}{k} \sum_{j=1}^k \phi_\alpha(\mathcal{S}, C_j). \quad (61)$$

Analogously, α -roughness of elements from sample \mathcal{S} in category C_j can be evaluated by α -roughness of w_j in approximation space (\mathbb{N}_n, I) . Hence

$$\rho_\alpha(\mathcal{S}, C_j) = 1 - \frac{|(w_j)_\alpha|}{|S_{w_j}|}, \quad (62)$$

and α -roughness of distribution of elements from sample \mathcal{S} in family of categories \mathcal{C} is

$$\rho_\alpha(\mathcal{S}, \mathcal{C}) = \frac{1}{k} \sum_{j=1}^k \rho_\alpha(\mathcal{S}, C_j). \quad (63)$$

5 Application

The construction of a membership function of a non-precise observation depends on the field of application (see Viertl, 1996). It is assumed in this section that each non-precise observation X_i is a triangular fuzzy number $[a_i, b_i, d_i]$. Constants a_i, b_i and d_i can be estimated by a human observer or calculated from a set of repeated measurements (crisp real numbers) associated with i -th object. Then, for example, a_i = first quartile, b_i = median, and d_i = third quartile of measurements.

Let the quantitative variable X of interest be a water level of a river measured (observed) in centimeters. A random sample \mathcal{S} of 12 non-precise measurements described by triangular fuzzy numbers X_i is in Table 1. Graphical representation of this sample is on Figure 3. Fuzziness of \mathcal{S} calculated according to (32) is $\phi(\mathcal{S}) = 14.17$. Because all non-precise observations are characterized by triangular fuzzy numbers, roughness of \mathcal{S} is $\rho(\mathcal{S}) = 1$ (see formula (32)). This amount of fuzziness and roughness of \mathcal{S} might be too high for further analysis. In order to reduce them, one may approximate each fuzzy observation X_i by its generalized α -cut $\mathcal{G}_\alpha(X_i)$. Let us assume, for example, $\alpha = 0.7$. Then $\mathcal{G}_{0.7}(X_i)$ is determined by quadruple $[a_{1i}, b_{1i}, c_{1i}, d_{1i}]$ given by formulas (39)-(42). Generalized 0.7-cuts with their fuzziness and roughness are in Table 2. Then

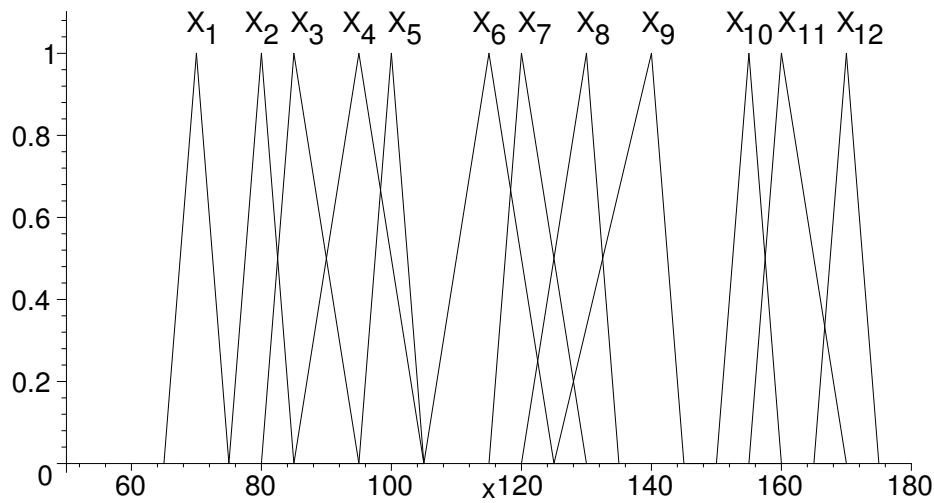
$$\phi(\mathcal{G}_{0.7}(\mathcal{S})) = \frac{1}{12} \sum_{i=1}^{12} \phi(\mathcal{G}_{0.7}(X_i)) = 8.14, \quad (64)$$

$$\rho(\mathcal{G}_{0.7}(\mathcal{S})) = \frac{1}{12} \sum_{i=1}^{12} \rho(\mathcal{G}_{0.7}(X_i)) = 0.823. \quad (65)$$

Table 1: Non-precise observations of water level

X_i	a_i	b_i	d_i
X_1	65	70	75
X_2	75	80	85
X_3	80	85	95
X_4	85	95	105
X_5	95	100	105
X_6	105	115	125
X_7	115	120	130
X_8	120	130	135
X_9	125	140	145
X_{10}	150	155	160
X_{11}	155	160	170
X_{12}	165	170	175

Fuzziness of \mathcal{S} has been reduced by 44.6 % and its roughness by 17.7%. In further exploratory analysis of \mathcal{S} one may use instead of original fuzzy numbers X_i their approximations by trapezoidal fuzzy intervals $\mathcal{G}_{0.7}(X_i)$.

Figure 3: Non-precise sample \mathcal{S}

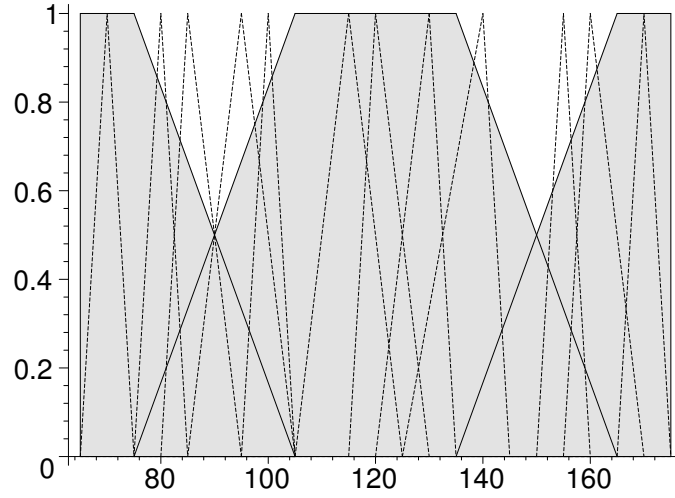
The range of sample \mathcal{S} is interval of real numbers $R_{\mathcal{S}} = [65, 175]$. Assume that $R_{\mathcal{S}}$ is divided into categories *low water level*, *medium water level*, *high water level* described by fuzzy intervals $C_1 = [65, 65, 75, 105]$, $C_2 = [75, 105, 135, 165]$ and $C_3 = [135, 165, 175, 175]$, respectively. Family $\mathcal{C} = \{C_j\}_{j=1}^3$ is a fuzzy partition of $R_{\mathcal{S}}$ such that $\phi(\mathcal{C}) = 40$ and $\rho(\mathcal{C}) = 0.389$. Fuzzy quantities $X_i \in \mathcal{S}$ together with categories $C_j \in \mathcal{C}$ are depicted on Figure 4.

Table 2: Generalized 0.7-cut of X_i

$\mathcal{G}_{0.7}(X_i)$	a_{1i}	b_{1i}	c_{1i}	d_{1i}	$\phi(\mathcal{G}_{0.7}(X_i))$	$\rho(\mathcal{G}_{0.7}(X_i))$
X_1	66.5	69.4	70.6	73.5	5.8	0.829
X_2	76.5	79.4	80.6	83.5	5.8	0.829
X_3	81.5	84.4	86.3	92.0	8.6	0.819
X_4	88.0	93.7	96.3	102.0	11.4	0.814
X_5	96.5	99.4	100.6	103.5	5.8	0.829
X_6	108.0	113.7	116.3	122.0	11.4	0.814
X_7	116.5	119.4	121.3	127.0	8.6	0.819
X_8	123.0	128.7	130.6	133.5	8.6	0.819
X_9	129.5	138.1	140.6	143.5	11.5	0.821
X_{10}	151.5	154.4	155.6	158.5	5.8	0.829
X_{11}	156.5	159.4	161.3	167.0	8.6	0.819
X_{12}	166.5	169.4	170.6	173.5	5.0	0.829

Calculation of grades of inclusion $\epsilon(X_i, C_j)$ is based on formula (48). Because $X_i = [a_i, b_i, d_i]$, cardinality $|X_i| = \frac{1}{2}(d_i - a_i)$. Each category C_j is a trapezoidal fuzzy interval $[a_j, b_j, c_j, d_j]$. If $X_i \subset C_j$ then $|X_i \cap C_j| = |X_i|$ and $\epsilon(X_i, C_j) = 1$. If $X_i \cap C_j = \emptyset$ then

$|X_i \cap C_j| = 0$ and $\epsilon(X_i, C_j) = 0$. If $X_i \neq X_i \cap C_j \neq \emptyset$, then one of the following

Figure 4: Non-precise categories on the range of \mathcal{S}

cases occurs:

1. There is only one real number x_1 such that $X_i(x_1) = C_j(x_1) = h_1 \in (0, 1)$ and $a_i \leq d_j$. For example, X_{11} and C_2 in our sample. Then

$$|X_i \cap C_j| = \int_{a_i}^{x_1} X_i(x) dx + \int_{x_1}^{d_j} C_j(x) dx = \frac{(d_j - a_i)h_1}{2}. \quad (66)$$

2. There is only one real number x_2 such that $X_i(x_2) = C_j(x_2) = h_2 \in (0, 1)$ and $a_j \leq d_i$. For example, X_9 and C_3 in our sample. Then

$$|X_i \cap C_j| = \int_{a_j}^{x_2} C_j(x) dx + \int_{x_2}^{d_i} X_i(x) dx = \frac{(d_i - a_j)h_2}{2}. \quad (67)$$

3. There are two real numbers x_1, x_2 such that $x_1 < x_2$, $X_i(x_1) = C_j(x_1) = h_1$, $X_i(x_2) = C_j(x_2) = h_2$ and $h_1, h_2 \in (0, 1)$. For example, X_4 and C_2 in our sample. Then

$$\begin{aligned} |X_i \cap C_j| &= \int_{a_i}^{x_1} X_i(x) dx + \int_{x_1}^{x_2} C_j(x) dx + \int_{x_2}^{d_i} X_i(x) dx \\ &= \frac{(x_2 - a_i)h_1}{2} + \frac{(d_i - x_1)h_2}{2}. \end{aligned} \quad (68)$$

Values of x_1, x_2 and values of h_1, h_2 for each combination of X_i and C_j are in Table 3 and Table 4, respectively.

Table 3: Points x_p where $C_j(x_p) = X_i(x_p) \in (0, 1)$

X_i	C_1		C_2		C_3	
	x_1	x_2	x_1	x_2	x_1	x_2
X_1	--	--	--	--	--	--
X_2	79.3	81.0	--	83.6	--	--
X_3	83.6	90.0	81.0	90.0	--	--
X_4	90.0	--	90.0	97.5	--	--
X_5	96.4	--	99.0	100.7	--	--
X_6	--	--	--	--	--	--
X_7	--	--	--	--	--	--
X_8	--	--	--	--	--	--
X_9	--	--	138.3	141.0	--	143.6
X_{10}	--	--	152.1	159.0	153.0	156.4
X_{11}	--	--	156.4	--	159.0	161.3
X_{12}	--	--	--	--	--	--

Table 4: Values h_p of $C_j(x_p) = X_i(x_p) \in (0, 1)$

X_i	C_1		C_2		C_3	
	h_1	h_2	h_1	h_2	h_1	h_2
X_1	--	--	--	--	--	--
X_2	0.857	0.800	--	0.287	--	--
X_3	0.713	0.500	0.200	0.5000	--	--
X_4	0.500	--	0.500	0.750	--	--
X_5	0.287	--	0.800	0.857	--	--
X_6	--	--	--	--	--	--
X_7	--	--	--	--	--	--
X_8	--	--	--	--	--	--
X_9	--	--	0.890	0.800	--	0.287
X_{10}	--	--	0.430	0.200	0.600	0.713
X_{11}	--	--	0.287	--	0.800	0.877
X_{12}	--	--	--	--	--	--

Table 5 provides information about $|C_j \cap X_i|$ and $\epsilon(C_j, X_i)$. From this table, using (59), fuzzy frequency functions w_j associated with categories C_j are determined. A fuzzy frequency function w_j represents the fuzzy count of observations from sample \mathcal{S} in category C_j .

Table 5: Inclusion of X_i in C_j

X_i	$ C_1 \cap X_i $	$ C_2 \cap X_i $	$ C_3 \cap X_i $	$\epsilon(X_i, C_1)$	$\epsilon(X_i, C_2)$	$\epsilon(X_i, C_3)$
X_1	5.00	0.00	0.00	1.000	0.000	0.000
X_2	4.85	1.44	0.00	0.970	0.288	0.000
X_3	6.42	4.50	0.00	0.856	0.600	0.000
X_4	5.00	8.75	0.00	0.500	0.875	0.000
X_5	1.43	4.85	0.00	0.286	0.970	0.000
X_6	0.00	10.00	0.00	0.000	1.000	0.000
X_7	0.00	7.50	0.00	0.000	1.000	0.000
X_8	0.00	7.50	0.00	0.000	1.000	0.000
X_9	0.00	9.80	1.44	0.000	0.980	0.144
X_{10}	0.00	2.73	4.42	0.000	0.546	0.880
X_{11}	0.00	1.43	7.34	0.000	0.191	0.979
X_{12}	0.00	0.00	5.00	0.000	0.000	1.000

Fuzzy frequency function w_1 associated with C_1 is

$$w_1(t) = \begin{cases} 1 & \text{if } t = 1, \\ 0.970 & \text{if } t = 2, \\ 0.856 & \text{if } t = 3, \\ 0.500 & \text{if } t = 4, \\ 0.286 & \text{if } t = 5, \\ 0 & \text{otherwise.} \end{cases} \quad (69)$$

Fuzzy frequency function w_2 associated with C_2 is

$$w_2(t) = \begin{cases} 1 & \text{if } 1 \leq t \leq 3, \\ 0.980 & \text{if } t = 4, \\ 0.970 & \text{if } t = 5, \\ 0.875 & \text{if } t = 6, \\ 0.600 & \text{if } t = 7, \\ 0.546 & \text{if } t = 8, \\ 0.288 & \text{if } t = 9, \\ 0.191 & \text{if } t = 10, \\ 0 & \text{otherwise.} \end{cases} \quad (70)$$

Fuzzy frequency function w_3 associated with C_3 is

$$w_3(t) = \begin{cases} 1 & \text{if } t = 1, \\ 0.979 & \text{if } t = 2, \\ 0.884 & \text{if } t = 3, \\ 0.144 & \text{if } t = 4, \\ 0 & \text{otherwise.} \end{cases} \quad (71)$$

Then

$$\phi(\mathcal{S}, C_1) = \phi_{0.5}(w_1) = |1 - 1| + |0.97 - 1| + |0.856 - 1| + |0.5 - 1| + |0.286 - 0| = .960,$$

$$\phi(\mathcal{S}, C_2) = \phi_{0.5}(w_2) = |0.288 - 0| + |0.6 - 1| + |0.875 - 1| + |0.97 - 1| + |0.98 - 1| + |0.546 - 1| + |0.19 - 0| = 1.507,$$

$$\phi(\mathcal{S}, C_3) = \phi_{0.5}(w_3) = |0.144 - 0| + |0.884 - 1| + |0.979 - 1| = 0.281.$$

Fuzziness of distribution of elements from \mathcal{S} in categories from \mathcal{C} is

$$\phi(\mathcal{S}, \mathcal{C}) = \frac{1}{3}[0.960 + 1.507 + 0.281] = 0.916. \quad (72)$$

Analogously,

$$\rho(\mathcal{S}, C_1) = \rho_1(w_1) = 1 - \frac{1}{5} = \frac{4}{5} = 0.80,$$

$$\rho(\mathcal{S}, C_2) = \rho_1(w_2) = 1 - \frac{3}{10} = \frac{7}{10} = 0.70$$

$$\rho(\mathcal{S}, C_3) = \rho_1(w_3) = 1 - \frac{1}{4} = \frac{3}{4} = 0.75.$$

Roughness of distribution of elements from \mathcal{S} in categories from \mathcal{C} is

$$\rho(\mathcal{S}, \mathcal{C}) = \frac{1}{3}[0.8 + 0.7 + 0.75] = 0.75. \quad (73)$$

The fuzzy frequency functions w_1 , w_2 and w_3 are displayed on generalized histogram on Figure 5. A generalized histogram (see Bodjanova, 1999a) consists of combination of

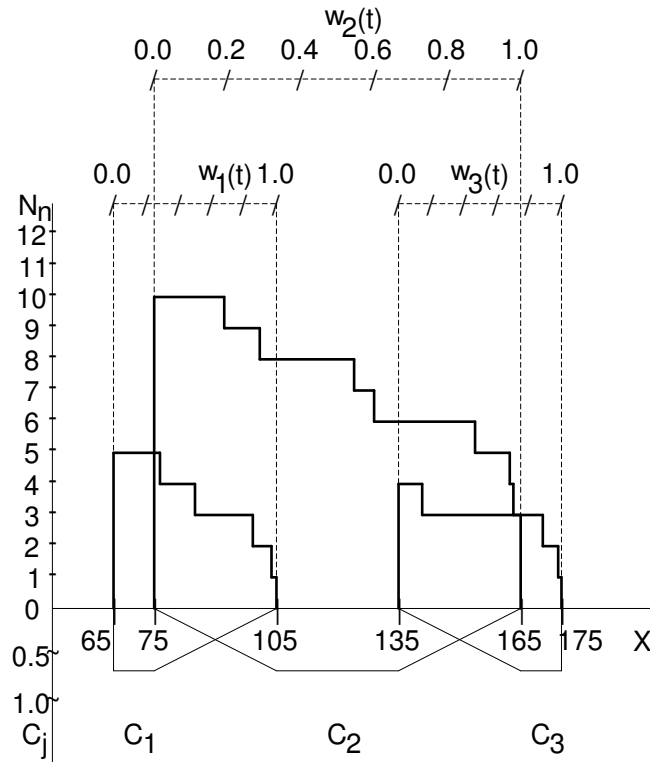


Figure 5: Generalized histogram of \mathcal{S}

fuzzy intervals C_j and fuzzy frequency functions w_j . The horizontal x -axis on Figure 5 is scaled according to the values of the range $R_{\mathcal{S}}$. The vertical axis has two different parts. The lower part (below the x -axis) represents the scale for coefficients of fuzzy functions C_j increasing from 0 to 1 in the direction to the bottom of the figure. Thus the lower part of a generalized histogram depicts fuzzy intervals C_j . The upper part of the vertical axis

(above the x -axis) includes elements from the set N_n . Thus the upper part of a generalized histogram can depict fuzzy frequency functions w_j . A scale for each w_j is needed. Because $w_j(t) \in [0, 1]$ and w_j is associated with C_j , a line segment parallel to the support of $C_j = (a_j, d_j)$ is drawn at least as far above the x -axis as the horizontal line determined by points $[a_j, n], [d_j, n]$. This line segment is partitioned into 10 equidistant intervals with endpoints marked by $0.0, 0.1, 0.2, \dots, 1.0$ and serves as the scale for coefficients $w_j(t)$. Graph of w_j consists of line segments $\{[t-1, w_j(t)], [t, w_j(t)]\}_{t=1}^n$. The area below the function $w_j(t)$ and above the support of C_j can be called a *fuzzy bar* of the generalized histogram and it is a graphical representation of frequency of elements from S in category C_j . For example, fuzzy frequency function w_2 given by (70) is determined by sequence of line segments

$\{[0, 1], [1, 1]\}; \{[1, 1], [2, 1]\}; \{[2, 1], [3, 1]\}; \{[3, 0.98], [4, 0.98]\}; \{[4, 0.97], [5, 0.97]\};$
 $\{[5, 0.875], [6, 0.875]\}; \{[6, 0.6], [7, 0.6]\}; \{[7, 0.546], [8, 0.546]\}; \{[8, 0.288], [9, 0.288]\};$
 $\{[9, 0.191], [10, 0.191]\}; \{[10, 0], [11, 0]\}; \{[11, 0], [12, 0]\}.$

Figure 6 shows the fuzzy bar associated with category C_2 in details.

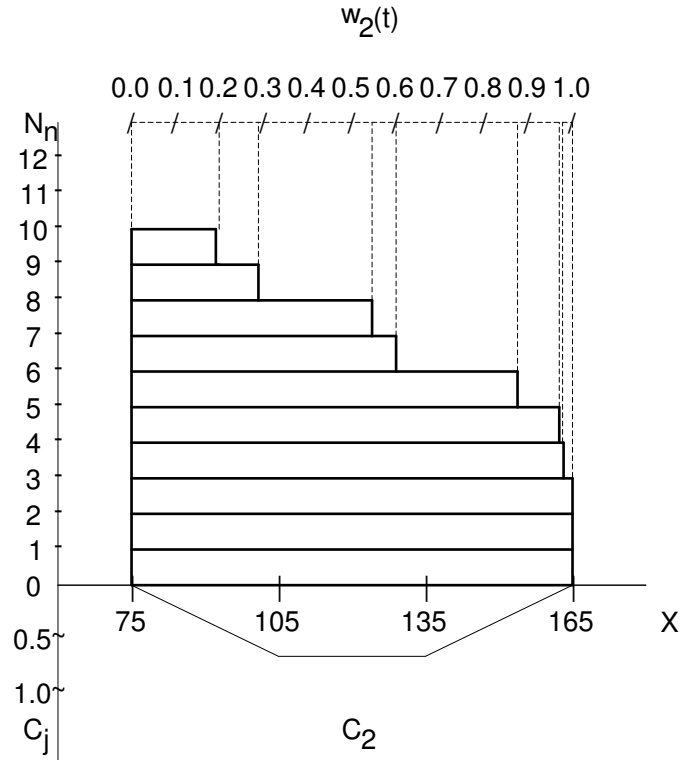


Figure 6: Fuzzy frequency of category C_2

6 Conclusion

Several connections between the theory of fuzzy sets and the theory of rough sets have been shown in dealing with characterization of non-precise quantities and characterization of empirical distributions of non-precise data in non-precise categories.

From a theoretical point of view, the notions of α -sharpness, α -fuzziness and generalized α -cut have been introduced. Each of these notions deserves careful study and will be the object of further investigation.

From a practical point of view, this paper presented some simple techniques for evaluation of uncertainty in a sample of non-precise data. Computational steps illustrated on a small example provide guidelines for application of suggested methods to analysis of real data.

References

- S. Bodjanova. A generalized histogram. To appear in *Fuzzy Sets and Systems*, 1999a.
- S. Bodjanova. Roughness and Fuzziness. Technical report, Texas A&M University-Kingsville, 1999b.
- D. Dubois and H. Prade. *Fuzzy Sets and Systems. Theory and Applications*. Academic Press, New York, 1980.
- H. Empotz. Nonprobabilistic entropies and indetermination measures in the setting of fuzzy sets theory. *Fuzzy Sets and Systems*, 5:307–317, 1981.
- S. Frühwirth-Schnatter. On statistical inference for fuzzy data with applications to descriptive statistics. *Fuzzy Sets and Systems*, 50:143–165, 1992.
- J. Kacprzyk and M. Federizzi. *Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision Making*. Springer-Verlag, Berlin, 1988.
- G.J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic. Theory and Applications*. Prentice Hall, Upper Saddle River, 1995.
- R. Kruse and K.D. Meyer. *Statistic with Vague Data*. Reidel, Dordrecht, 1987.
- S.K. Pal and A. Skowron, (eds.). *Rough Fuzzy Hybridization. A New Trend in Decision Making*. Springer-Verlag, Singapore, 1999.
- Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11: 341–356, 1982.
- Z. Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer, Dordrecht, 1991.
- L. Polkowski and A. Skowron, (eds.). *Rough Sets in Knowledge Discovery. Methodology and Applications*. Physica-Verlag, Heidelberg, 1998.
- Ch. Römer and A. Kandel. Statistical tests for fuzzy data. *Fuzzy Sets and Systems*, 72: 1–26, 1995.
- C.E. Shannon. The mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.

- R. Slowinski, (ed.). *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*. Kluwer, Dordrecht, 1992.
- R. Viertl. *Statistical Methods for Non-Precise Data*. CRC Press, Boca Raton, 1996.
- L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:383–353, 1965.
- W.P. Ziarko, (ed.). *Rough Sets, Fuzzy Sets and Knowledge Discovery*. Springer-Verlag, Berlin, 1994.

Author's address:

Dr. Slavka Bodjanova
Texas A&M University-Kingsville
Department of Mathematics
Kingsville, TX 78363
U.S.A.

Tel.: (361)-593-2238

Fax: (361)-593-3518

Email: kfsb000@tamuk.edu