

The Spearman and Kendall rank correlation coefficients between intuitionistic fuzzy sets

Eulalia Szmidt and Janusz Kacprzyk

Systems Research Institute Polish Academy of Sciences, ul.Newelska 6, 01-447 Warsaw, Poland
 WIT – Warsaw School of Information Technology ul. Newelska 6, 01-447 Warsaw, Poland
 {szmidt, kacprzyk }@ibspan.waw.pl

Abstract

This paper is a continuation of our previous work on *Pearson's coefficient r* and we discuss here the concept of the *Spearman correlation coefficient* and the *Kendall correlation coefficient* between Atanassov's intuitionistic fuzzy sets (A-IFSs, for short) to measure the degree of association between the A-IFSs when the assumption that the data distributions are normal is not valid or when data are in the form of ranks.

Keywords: Intuitionistic fuzzy sets, correlation, the Spearman correlation coefficient, the Kendall correlation coefficient

1. Introduction

It is often important to know the relationships between relevant variables while building models describing a process or a system. The Pearson correlation coefficient r measuring a linear relationship between the variables is one of the most frequently used tools in statistics (cf. Rodgers and Nicewander [14]). Generally, correlation indicates how well two normally distributed variables move together in a linear way (Rodgers and Nicewander [14], Aczel [1]). When the assumption about the normal distributions of the variables considered is not valid or the data are in the form of ranks, we use other measures of the degree of association between two variables, namely the *Spearman rank correlation coefficient* r_s (e.g., Aczel [1]) or the *Kendall τ correlation coefficient* (Kendall [9]).

As Zadeh has observed [37], [38], most of information relevant to probabilistic analyzes is imprecise, and there is imprecision and fuzziness not only in probabilities, but also in events, relations and properties. In this context, the probabilistic concepts should also be extended to fuzzy models and their generalizations. The analysis of relationships between the A-IFSs, which are a generalization of fuzzy sets, seems to be of a vital importance, too. We have already discussed in detail the *Pearson correlation coefficient r* (Szmidt and Kacprzyk [32]). In this paper we jointly discuss the *Spearman rank correlation coefficient* r_s (Szmidt and Kacprzyk [33]) and the *Kendall τ correlation coefficient* to find and indicate some common properties and differences.

Both the *Spearman rank correlation coefficient* and the *Kendall τ correlation coefficient* for the A-IFSs are

a generalization of their crisp counterparts (it boils down to them, and preserve the same properties each).

Moreover, both correlation coefficients take into account all three terms describing the A-IFS, i.e., the membership values, non-membership values, and hesitation margins. We show that each term plays an important role in data analysis and decision making.

2. A Brief Introduction to Intuitionistic Fuzzy Sets

One of the possible generalizations of a fuzzy set in X (Zadeh [36]), given by

$$A' = \{< x, \mu_{A'}(x) > | x \in X\} \quad (1)$$

where $\mu_{A'}(x) \in [0, 1]$ is the membership function of the fuzzy set A' , is Atanassov's intuitionistic fuzzy set A (Atanassov [2], [3]), namely:

$$A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\} \quad (2)$$

where: $\mu_A : X \rightarrow [0, 1]$ and $\nu_A : X \rightarrow [0, 1]$ such that $0 \leq \mu_A(x) + \nu_A(x) \leq 1$, and $\mu_A(x), \nu_A(x) \in [0, 1]$ denote the degree of membership and a degree of non-membership of $x \in A$, respectively, and the *hesitation margin* of $x \in A$ is:

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \quad (3)$$

The $\pi_A(x)$ expresses a lack of knowledge of whether x belongs to A or not (Atanassov [3]); obviously, $0 \leq \pi_A(x) \leq 1$, for each $x \in X$;

The hesitation margin turns out to be important while considering the distances (Szmidt and Kacprzyk [16], [18], [25], entropy (Szmidt and Kacprzyk [20], [27]), similarity (Szmidt and Kacprzyk [28]) for the A-IFSs, etc. i.e., the measures that play a crucial role in virtually all information processing tasks. The hesitation margin is shown to be indispensable also in the ranking of intuitionistic fuzzy alternatives as it indicates how reliable (sure) information represented by an alternative is (cf. Szmidt and Kacprzyk [29], [30]).

The use of A-IFSs instead of fuzzy sets implies the introduction of additional degrees of freedom (non-memberships and hesitation margins) into the set description. Such a generalization of fuzzy sets gives us an additional possibility to represent imperfect knowledge which may lead to describing many real problems in a more adequate way. This is confirmed by successful

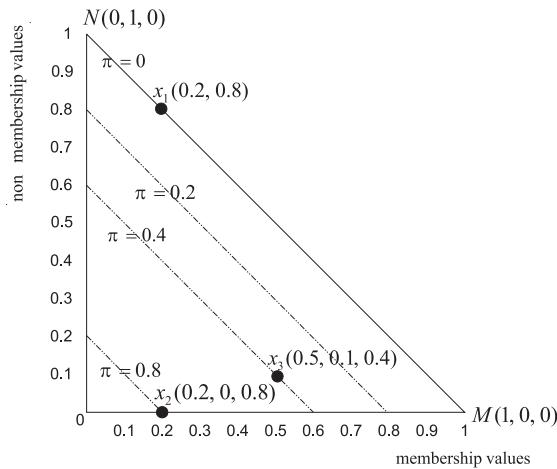


Figure 1: Geometrical representation

applications of A-IFSs to group decision making, negotiations, voting and other situations are presented in Szmidt and Kacprzyk [15], [17], [19], [21], [22], [23], [24], [26], [31], Szmidt and Kukier [34], [35].

2.1. A geometrical representation

One of possible geometrical representations of an intuitionistic fuzzy sets is given in Figure 1 (cf. Atanassov [3]). It is worth noticing that although we use a two-dimensional figure (which is more convenient to draw in our further considerations), we still adopt our approach (e.g., Szmidt and Kacprzyk [18], [25], [20], [27]), [28]) taking into account all three terms (membership, non-membership and hesitation margin values) describing an intuitionistic fuzzy set. Any element belonging to an intuitionistic fuzzy set may be represented inside an MNO triangle. In other words, the MNO triangle represents the surface where the coordinates of any element belonging to an A-IFS can be represented. Each point belonging to the MNO triangle is described by the three coordinates: (μ, ν, π) . Points M and N represent the crisp elements. Point $M(1, 0, 0)$ represents elements fully belonging to an A-IFS as $\mu = 1$, and may be seen as the representation of the ideal positive element. Point $N(0, 1, 0)$ represents elements fully not belonging to an A-IFS as $\nu = 1$, i.e. can be viewed as the ideal negative element. Point $O(0, 0, 1)$ represents elements about which we are not able to say if they belong or not belong to an A-IFS (the intuitionistic fuzzy index $\pi = 1$). Such an interpretation is intuitively appealing and provides means for the representation of many aspects of imperfect information. Segment MN (where $\pi = 0$) represents elements belonging to the classic fuzzy sets ($\mu + \nu = 1$). For example, point $x_1(0.2, 0.8, 0)$ (Figure 1), like any element from segment MN represents an element of a fuzzy set. A line parallel to MN describes the elements with the same values of the hesitation margin. In Figure 1 we can see point $x_3(0.5, 0.1, 0.4)$ representing an element with the hesitation margin equal 0.4, and point $x_2(0.2, 0, 0.8)$ representing an element with the hesitation margin equal 0.8. The closer a line is

parallel to MN is to O , the higher the hesitation margin.

3. Correlation, the Spearman and Kendall Rank Correlation Coefficients between crisp sets

The correlation coefficient (Pearson's r) between two variables is a measure of the linear relationship between them.

The correlation coefficient is 1 in the case of a positive (increasing) linear relationship, -1 in the case of a negative (decreasing) linear relationship, and some value between -1 and 1 in all other cases. The closer the coefficient is to either -1 or 1, the stronger the correlation between the variables.

Definition 1 Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a random sample of size n from a joint probability density function $f_{X,Y}(x, y)$, let \bar{X} and \bar{Y} be the sample means of variables X and Y , respectively, then the sample correlation coefficient $r(X, Y)$ is given as (e.g., [14]):

$$r(A, B) = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{(\sum_{i=1}^n (x_i - \bar{X})^2 \sum_{i=1}^n (y_i - \bar{Y})^2)^{0.5}} \quad (4)$$

where: $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i$.

When the assumption that the data distributions are normal is not valid or when data are in the form of ranks, we may use the *Spearman rank correlation coefficient* or the *Kendall rank correlation coefficient*.

Definition 2 Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a random sample of size n . To compute the *Spearman correlation coefficient* we rank all the observations of the first variable. Next, we independently rank the values of the second variable. The *Spearman rank correlation coefficient* is the *Pearson correlation coefficient* applied to the ranks R , namely (e.g., Aczel [1]):

$$r(A, B) = \frac{\sum_{i=1}^n (R(x_i) - \frac{1}{2}(n+1))(R(y_i) - \frac{1}{2}(n+1))}{(\sum_{i=1}^n (R(x_i) - \frac{1}{2}(n+1))^2 \sum_{i=1}^n (R(y_i) - \frac{1}{2}(n+1))^2)^{0.5}} \quad (5)$$

and in (5) the average values for X and Y are the same and equal to $0.5(n+1)$.

When there are not two values of X or two values of Y with the same rank (so called ties), there is an easier way to compute the Spearman correlation coefficient, namely (e.g., Aczel [1], Myers and Well [11]):

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (6)$$

where d_i , $i = 1, \dots, n$ are the differences in the ranks of x_i and y_i : $d_i = R(x_i) - R(y_i)$. If there are ties (two

values of X or two values of Y with the same rank), but the number of ties is small compared with n , (6) is still useful (cf. Aczel [1]).

The Spearman correlation coefficient fulfills the requirements of the correlation measures. As (6) was obtained from the Pearson coefficient for ranks, it fulfills the same properties as the Pearson coefficient (see e.g., Rodgers and Nicewander [14]), i.e.:

1. $r_s(A, B) = r_s(B, A)$,
2. If $A = B$ then $r_s(A, B) = 1$,
3. $|r_s(A, B)| \leq 1$.

When the variables X and Y are perfectly positively related, i.e., when X increases whenever Y increases, r_s is equal to 1. When X and Y are perfectly negatively related, i.e., when X increases whenever Y decreases, r_s is equal to -1. r_s is equal to zero when there is no relation between X and Y . Values between -1 and 1 give a relative indication of the degree of association between X and Y . In other words, $-1 \leq r_s \leq 1$.

Definition 3 (e.g., Nelsen [12]) Let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be a set of joint observations from two random variables X and Y respectively, such that all the values of (x_i) and (y_i) are unique. Any pair of observations (x_i, y_i) and (x_j, y_j) are said to be concordant if the ranks for both elements agree: that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. They are said to be discordant, if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant, nor discordant.

For n observations, (i.e., $(i, j) \in \{1, \dots, n\}^2$) the number of concordant C , discordant D , tied pairs T in X , and tied pairs U in Y is denoted:

$$C = |\{(i, j) | x_i < x_j \text{ and } y_i < y_j\}| \quad (7)$$

$$D = |\{(i, j) | x_i < x_j \text{ and } y_i > y_j\}| \quad (8)$$

$$T = |\{(i, j) | x_i = x_j\}| \quad (9)$$

$$U = |\{(i, j) | y_i = y_j\}| \quad (10)$$

The Kendall τ coefficient is defined as [9]:

$$\tau = \frac{C - D}{\frac{1}{2}n(n - 1)} \quad (11)$$

where: C - number of concordant pairs; D - number of discordant pairs, and

$$|\tau| \leq 1$$

For the perfect agreement between the two rankings (i.e., all pairs are concordant) the coefficient has value 1. For the perfect disagreement between the two rankings (i.e., all pairs are discordant) the coefficient has value -1. All other arrangements produce the value of τ between -1 and 1 (increasing values imply increasing agreement between the rankings). For completely independent rankings, the coefficient has value 0.

If two values of X or two values of Y with the same rank (so called ties) occur, the following formula is used

(Kendall [9]):

$$\tau_b = \frac{c - D}{\sqrt{\frac{1}{2}(n(n - 1) - T)} \sqrt{\frac{1}{2}(n(n - 1) - U)}} \quad (12)$$

where: T - number of ties in X (number of pairs for which $x_i = x_j$); U - number of ties in Y (number of pairs for which $y_i = y_j$).

4. Correlation, and the Spearman and Kendall Rank Correlation Coefficients between the A-IFSSs

In Szmidt and Kacprzyk [32] we proposed a correlation coefficient for two A-IFSSs, A and B , so that we could express not only a relative strength but also a positive or negative relationship between A and B . We took into account all three terms describing an A-IFSSs (the membership, non-membership values and the hesitation margins) because each of them influences the results (cf. [32]).

Suppose that we have a random sample $x_1, x_2, \dots, x_n \in X$ with a sequence of paired data $[(\mu_A(x_1), \nu_A(x_1), \pi_A(x_1)), (\mu_B(x_1), \nu_B(x_1), \pi_B(x_1))], [(\mu_A(x_2), \nu_A(x_2), \pi_A(x_2)), (\mu_B(x_2), \nu_B(x_2), \pi_B(x_2))], \dots, [(\mu_A(x_n), \nu_A(x_n), \pi_A(x_n)), (\mu_B(x_n), \nu_B(x_n), \pi_B(x_n))]$ which correspond to the membership values, non-memberships values and hesitation margins of the A-IFSSs A and B defined on X , then the correlation coefficient $r_{A-IFS}(A, B)$ is given by Definition 4.

Definition 4 The correlation coefficient $r_{A-IFS}(A, B)$ between two A-IFSSs, A and B in X , is (Szmidt and Kacprzyk [32]):

$$r_{A-IFS}(A, B) = \frac{1}{3}(r_1(A, B) + r_2(A, B) + r_3(A, B)) \quad (13)$$

where

$$r_1(A, B) = \sum_{i=1}^n (\mu_A(x_i) - \bar{\mu}_A)(\mu_B(x_i) - \bar{\mu}_B) / \sum_{i=1}^n (\mu_A(x_i) - \bar{\mu}_A)^2)^{0.5} (\sum_{i=1}^n (\mu_B(x_i) - \bar{\mu}_B)^2)^{0.5} \quad (14)$$

$$r_2(A, B) = \sum_{i=1}^n (\nu_A(x_i) - \bar{\nu}_A)(\nu_B(x_i) - \bar{\nu}_B) / \sum_{i=1}^n (\nu_A(x_i) - \bar{\nu}_A)^2)^{0.5} (\sum_{i=1}^n (\nu_B(x_i) - \bar{\nu}_B)^2)^{0.5} \quad (15)$$

$$r_3(A, B) = \sum_{i=1}^n (\pi_A(x_i) - \bar{\pi}_A)(\pi_B(x_i) - \bar{\pi}_B) / \sum_{i=1}^n (\pi_A(x_i) - \bar{\pi}_A)^2)^{0.5} (\sum_{i=1}^n (\pi_B(x_i) - \bar{\pi}_B)^2)^{0.5} \quad (16)$$

where: $\overline{\mu_A} = \frac{1}{n} \sum_{i=1}^n \mu_A(x_i)$, $\overline{\mu_B} = \frac{1}{n} \sum_{i=1}^n \mu_B(x_i)$,
 $\overline{\nu_A} = \frac{1}{n} \sum_{i=1}^n \nu_A(x_i)$, $\overline{\nu_B} = \frac{1}{n} \sum_{i=1}^n \nu_B(x_i)$, $\overline{\pi_A} = \frac{1}{n} \sum_{i=1}^n \pi_A(x_i)$, $\overline{\pi_B} = \frac{1}{n} \sum_{i=1}^n \pi_B(x_i)$,

The proposed correlation coefficient (13) depends on two factors: the amount of information expressed by the membership and non-membership degrees (14)–(15), and the reliability of information expressed by the hesitation margins (16).

Remark: Analogously as for the crisp and fuzzy data, $r_{A-IFS}(A, B)$ makes sense for the A-IFS type variables whose values vary. If, for instance, the temperature is constant and the amount of ice cream sold is the same, then it is impossible to conclude about their relationship (as, from the mathematical point of view, we avoid zero in the denominator).

The correlation coefficient $r_{A-IFS}(A, B)$ (13) fulfills the following properties:

1. $r_{A-IFS}(A, B) = r_{A-IFS}(B, A)$,
2. If $A = B$ then $r_{A-IFS}(A, B) = 1$,
3. $|r_{A-IFS}(A, B)| \leq 1$.

The above properties are not only fulfilled by the correlation coefficient $r_{A-IFS}(A, B)$ (13) but also by all of its three components (14)–(16).

Remark: It is should be emphasized that $r_{A-IFS}(A, B) = 1$ occurs not only for $A = B$ but also in the cases of a perfect linear correlation of the data (the same concerns each component (14)–(16)).

From Definition 4, we immediately obtain the *Spearman rank correlation coefficient* because it is the usual (Pearson) correlation coefficient applied to the ranks.

Definition 5 The Spearman rank correlation coefficient is defined as:

$$r_{s_{A-IFS}} = \frac{1}{3}(r_{s_1} + r_{s_1} + r_{s_1}) \quad (17)$$

where r_{s_i} , $i = 1, \dots, 3$ are the Spearman rank correlation coefficients between A and B with respect to their membership function, non-membership function, and hesitation margin, given as:

$$r_{s_1} = 1 - \frac{6 \sum_{i=1}^n d_{1i}^2}{n(n^2 - 1)} \quad (18)$$

where d_{1i} , $i = 1, \dots, n$ are the differences in the ranks with respect to the membership functions: $d_{1i} = R(\mu_A(x_i)) - R(\mu_B(x_i))$,

$$r_{s_2} = 1 - \frac{6 \sum_{i=1}^n d_{2i}^2}{n(n^2 - 1)} \quad (19)$$

where d_{2i} , $i = 1, \dots, n$ are the differences in the ranks with respect to the non-membership functions: $d_{2i} = R(\nu_A(x_i)) - R(\nu_B(x_i))$, and

$$r_{s_3} = 1 - \frac{6 \sum_{i=1}^n d_{3i}^2}{n(n^2 - 1)} \quad (20)$$

Table 1: Example 1 - calculations of (18)

Patient No.	μ_A	$R(\mu_A)$	μ_B	$R(\mu_B)$	d_1	d_1^2
1	0.1	1	0.2	1	0	0
2	0.3	3	0.42	2	1	1
3	0.39	4	0.5	3	1	1
4	0.42	5	0.53	4	1	1
5	0.2	2	0.65	5	-3	9

Table 2: Example 1 - calculations of (19)

Patient No.	ν_A	$R(\nu_A)$	ν_B	$R(\nu_B)$	d_2	d_2^2
1	0.05	1	0.12	1	0	0
2	0.15	2	0.28	3	-1	1
3	0.3	3	0.31	5	-2	4
4	0.48	4	0.29	4	0	0
5	0.72	5	0.18	2	3	9

where d_{3i} , $i = 1, \dots, n$ are the differences in the ranks with respect to the hesitation margins: $d_{3i} = R(\pi_A(x_i)) - R(\pi_B(x_i))$.

Obviously, for the *Spearman rank correlation* (17) the same properties as for the *Pearson correlations coefficient* are valid, i.e.:

1. $r_{s_{A-IFS}}(A, B) = r_{s_{A-IFS}}(B, A)$
2. If $A = B$ then $r_{s_{A-IFS}}(A, B) = 1$
3. $|r_{s_{A-IFS}}(A, B)| \leq 1$

The separate components of the *Spearman rank correlation* (17) i.e., (18)–(20) fulfill the above properties, too.

Obviously, in the case of crisp sets, $r_{s_{A-IFS}}$ (17) reduces to r_s (6).

Table 3: Example 1 - calculations of (20)

Patient No.	π_A	$R(\pi_A)$	π_B	$R(\pi_B)$	d_3	d_3^2
1	0.85	5	0.68	5	0	0
2	0.55	4	0.3	4	0	0
3	0.31	3	0.19	2	1	1
4	0.1	2	0.27	3	-1	1
5	0.08	1	0.17	1	0	0

Example 1 For instance, while administering any medication, an expert clinician should make decision based on the context of the individual patient and his/her own past experience of the expected effect (e.g. Helgason and Jobe [5], [6]). The effects may be positive (expressed by a membership value), negative (expressed by a non-membership value), and difficult to foresee (expressed by a hesitation margin), for a specific patient. Suppose that two new medicines A and B are tested, and their effects on 5 patients are the following (Figure 2):

$$A = \{(x_1, 0.1, 0.05, 0.85), (x_2, 0.3, 0.15, 0.55), (x_3, 0.39, 0.3, 0.31), (x_4, 0.42, 0.48, 0.1), (x_5, 0.2, 0.72, 0.08)\} \quad (21)$$

$$B = \{(x_1, 0.2, 0.12, 0.68), (x_2, 0.42, 0.28, 0.3), (x_3, 0.5, 0.31, 0.19), (x_4, 0.53, 0.29, 0.27), (x_5, 0.65, 0.18, 0.17)\} \quad (22)$$

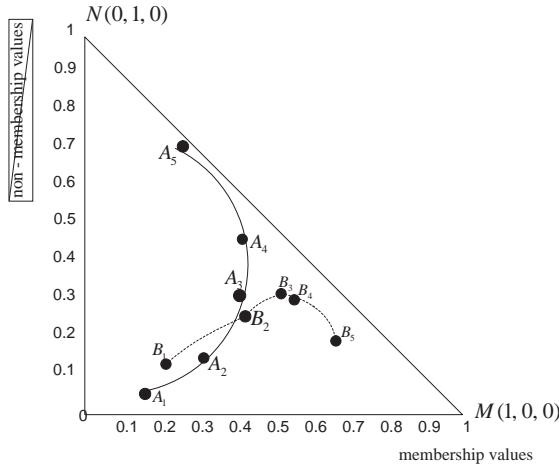


Figure 2: Data from Example 1

where, for example, the positive effects of medicine A on the first patient (x_1) are expressed by the membership value equal to 0.1, the negative effects are expressed by the non-membership value equal to 0.05, and effects difficult to foresee are expressed by the hesitation margin equal to 0.85; etc.

As the assumption of a normal distribution may be inappropriate, we use the Spearman rank correlation coefficient to conclude about the association between the two medicines A and B . Due to (17), we examine first the positive effects expressed by μ_A , and μ_B (Table 1) so as to obtain the value of (18). As $\sum_{i=1}^5 d_{1i}^2 = 12$, from (18) we obtain $r_{s_1} = 1 - ((6 * 12)/(5 * 24)) = 0.4$. Repeating the same calculations for the negative effects expressed by ν_A , and ν_B (Table 2), we obtain the value of r_{s_2} (19) equal to 0.3. Finally, we examine effects that are difficult to foresee (Table 3), and obtain the value of r_{s_3} (20) equal to $1 - ((6 * 2)/(5 * 24)) = 0.9$. As a result, from (17), we obtain $r_{s_{A-IFS}} = 0.53$. But from the point of view of an expert clinician all three components may seem interesting. Let us notice that in this example the data are such that r_{s_3} influences substantially the final result. If we consider a relationship between A and B only in the categories of the positive effects and negative effects, we obtain an average relationship equal to $0.5(0.4 + 0.3) = 0.35$ which suggests that the relationship between A and B is not strong. On the other hand, r_{s_3} (20) equal to 0.9 suggests that both medicines are related as far as unforeseen effects are concerned. It may be an important information from a medical point of view.

By showing in detail the above small example we wanted to illustrate that all three terms describing the A-IFSs, namely the membership values, non-membership values, and hesitation margins are important from the point of view of assessing the correlation between the A-IFSs. It seems obvious as each of the term “speaks” for another aspect of a system described via an A-IFS. The membership values and non-membership values represent our knowledge whereas the hesitation margins represent a lack of knowledge. It seems that from a point

of view of decision making we can not exclude from the models a lack of knowledge which in real world is the subject of equal or even more interest than “sure knowledge” (represented in the A-IFS models by the membership and non-membership values).

Now we will propose the definition of the *Kendall rank correlation coefficient* for the A-IFSs, and will use it to solve the same Example 1.

Definition 6 The Kendall rank correlation coefficient τ_{A-IFS} between two A-IFSs, A and B in X is defined as:

$$\tau_{A-IFS} = \frac{1}{3}(\tau_1 + \tau_2 + \tau_3) \quad (23)$$

where: τ_i , $i = 1, \dots, 3$ are the Kendall rank correlation coefficients between A and B with respect to their membership values, non-membership values, and hesitation margin values, given as:

$$\tau_1 = \frac{C_\mu - D_\mu}{\frac{1}{2}n(n-1)} \quad (24)$$

where: C_μ – the number of concordant pairs with respect to the membership values; D_μ – the number of discordant pairs with respect to the membership values, i.e.:

$$C_\mu = |\{(i, j) | \mu_A(x_i) < \mu_A(x_j) \text{ and } \mu_B(x_i) < \mu_B(x_j)\}| \quad (25)$$

$$D_\mu = |\{(i, j) | \mu_A(x_i) < \mu_A(x_j) \text{ and } \mu_B(x_i) > \mu_B(x_j)\}| \quad (26)$$

$$\tau_2 = \frac{C_\nu - D_\nu}{\frac{1}{2}n(n-1)} \quad (27)$$

where: C_ν – the number of concordant pairs with respect to the non-membership values; D_ν – the number of discordant pairs with respect to the non-membership values, i.e.:

$$C_\nu = |\{(i, j) | \nu_A(x_i) < \nu_A(x_j) \text{ and } \nu_B(x_i) < \nu_B(x_j)\}| \quad (28)$$

$$D_\nu = |\{(i, j) | \nu_A(x_i) < \nu_A(x_j) \text{ and } \nu_B(x_i) > \nu_B(x_j)\}| \quad (29)$$

and

$$\tau_3 = \frac{C_\pi - D_\pi}{\frac{1}{2}n(n-1)} \quad (30)$$

where: C_π – the number of concordant pairs with respect to the hesitation margins; D_π – the number of discordant pairs with respect to the hesitation margins, i.e.:

$$C_\pi = |\{(i, j) | \pi_A(x_i) < \pi_A(x_j) \text{ and } \pi_B(x_i) < \pi_B(x_j)\}| \quad (31)$$

$$D_\pi = |\{(i, j) | \pi_A(x_i) < \pi_A(x_j) \text{ and } \pi_B(x_i) > \pi_B(x_j)\}| \quad (32)$$

For the *Kendall rank correlation coefficient* τ (23) between the A-IFSs the same properties as for its crisp set counterpart are valid, i.e.:

Table 4: Example 1 - calculations of (24)

pair	(0.2, 0.65)	(0.2, 0.42)	(0.2, 0.5)	(0.2, 0.53)	((0.65, 0.42)	(0.65, 0.5)	(0.65, 0.53)	(0.42, 0.5)	(0.42, 0.53)	(0.5, 0.53)
score	+1	+1	+1	+1	-1	-1	-1	+1	+1	+1

Table 5: Example 1 - calculations of (27)

pair	(0.12, 0.28)	(0.12, 0.31)	(0.12, 0.29)	(0.12, 0.18)	((0.28, 0.31)	(0.28, 0.29)	(0.28, 0.18)	(0.31, 0.29)	(0.31, 0.18)	(0.29, 0.18)
score	+1	+1	+1	+1	+1	+1	-1	-1	-1	-1

Table 6: Example 1 - calculations of (30)

pair	(0.17, 0.27)	(0.17, 0.19)	(0.17, 0.3)	(0.17, 0.68)	((0.27, 0.19)	(0.27, 0.3)	(0.27, 0.68)	(0.19, 0.3)	(0.19, 0.68)	(0.3, 0.68)
score	+1	+1	+1	+1	-1	+1	+1	+1	+1	+1

$$1. |\tau_{A-IFS}| \leq 1$$

$$2. \tau_{A-IFS}(A, B) = \tau_{A-IFS}(B, A)$$

The interpretation of the obtained values is the same as in a case of the crisp sets. The perfect agreement of the rankings (i.e., the two rankings are the same) produces the value of the coefficient τ_{A-IFS} equal to 1. The perfect disagreement of the rankings (i.e., one ranking is the reverse of the other) produces the value of the coefficient τ_{A-IFS} equal to -1. All other arrangements produce the value of the coefficient τ_{A-IFS} which lies between -1 and 1, and increasing values imply increasing agreement between the rankings. For the completely independent rankings, the coefficient τ_{A-IFS} is equal to 0. Each of the three components of the *Kendall rank correlation coefficient* (23) i.e., (24), (27), and (30) fulfill the above properties, too.

It is worth emphasizing that the *Kendall rank correlation coefficient* is an intuitively simple measure of the strength of relationship between the A-IFSSs. Just the same as in a case of crisp sets, for the *Kendall correlation coefficient* τ_{A-IFS} , the ratio of the concordant to discordant pairs of observations equals $(1 + \tau_{A-IFS})/(1 - \tau_{A-IFS})$.

Remark: Definition 6 was introduced for the cases without ties. If ties occur, the counterparts of (12) replace the formulas (24), (27), and (30).

Now we will calculate τ_{A-IFS} (23) for the data (21)–(22) of Example 1.

We calculate τ_{A-IFS} in three steps using the three terms describing an A-IFS (membership values, non-membership values, and the hesitation margins) which are responsible for the three components (τ_1 , τ_2 , τ_3) of τ_{A-IFS} (23).

First, we make use of the membership values of A and B . The membership values of the elements in A are: 0.1, 0.3, 0.39, 0.42, 0.2, so first we must order them obtaining: 0.1, 0.2, 0.3, 0.39, 0.42, and the corresponding membership values of the elements in B are: 0.2, 0.65, 0.42, 0.5, 0.53. The details of the calculations of τ_1 due to (25)–(26) are given in Table 4. It is worth mentioning that we calculate τ_1 using the actual observations without first converting them to ranks: we con-

sider all possible pairs of the membership values of the elements in B – for each element we consider only the pairs build from elements occurring after the element considered, e.g., for 0.42 we consider two pairs: (0.42, 0.5) and (0.42, 0.53), whereas for 0.5 we consider one pair only: (0.5, 0.53). If the second element in a pair is bigger than the first element, the score is equal 1, if the second element is smaller than the first element, the score is equal to -1 (see (Table 4)) As we have $C_5^2 = 10$ pairs, and the obtained score (Table 4) is equal $7 - 3 = 4$, in result τ_1 (24) is equal to 0.4.

Next, we make use of the non-membership values. The non-membership values of the elements in A : 0.05, 0.15, 0.3, 0.48, 0.72 are already ordered (increase), and the non-membership values of the elements in B are: 0.12, 0.28, 0.31, 0.29, 0.18. In result, performing the same calculations as previously (details in Table 5), we obtain τ_2 (27) equal to 0.2.

Finally, we calculate τ_3 (30). The hesitation margins of the elements in A , i.e.: 0.85, 0.55, 0.31, 0.1, 0.08 decrease, so we arrange them in increasing order obtaining: 0.08, 0.1, 0.31, 0.55, 0.85, with the respective order of the elements of B : 0.17, 0.27, 0.19, 0.3, 0.68. Further calculations are like previously – details in Table 6. In effect τ_3 (30) is equal to 0.8.

Therefore, finally, from (23) we obtain $\tau_{A-IFS}(A, B) = \frac{1}{3}(0.4 + 0.2 + 0.8) = 0.47$.

If we exclude from considerations the hesitation margins, and take into account two components (24) and (27) only, we obtain $\tau_{A-IFS}(A, B) = \frac{1}{2}(0.4 + 0.2) = 0.3$. So the situation is the same as it was for the Spearman rank correlation, i.e., τ_3 influences substantially the final result (τ_3 (30) equal to 0.8 suggests that both medicines are strongly associated as far as unforeseen effects are concerned which may be an important information from a medical point of view).

Certainly, we may find such cases for which τ_3 (30) does not influence the Kendall coefficient τ_{A-IFS} (23) in a sense of the final result (an obtained number). But such situations are exceptional, not a rule.

It is worth mentioning again that each of the three components of both r_{SA-IFS} (17) and τ_{A-IFS} (23) should be examined carefully, and assessed separately

when decision making. More, the aggregated forms of $r_{s_{A-IFS}}$ (17) and τ_{A-IFS} (23) should not be used in decision making while their negative and the positive components have equal importance for the final decision.

We can also notice that the relationship between the *Spearman correlation coefficient* (17) and the *Kendall correlation coefficient* (23) for the A-IFSs follows the rule explored widely for crisp sets (cf. e.g., Fredricks and Nelsen [4]), namely the values of the *Spearman correlation coefficient* are usually bigger than the values of the *Kendall correlation coefficient*.

5. Conclusions

We presented the concepts of the *Spearman and Kendall correlation coefficients* for the A-IFSs. The coefficients are a generalization of their counterparts for the crisp sets, i.e., they fulfill the same properties, and reduce to their well known forms for the crisp sets.

Next, all three terms describing the A-IFS are taken into account in our analysis of both correlation coefficients (the membership values, non-membership values and hesitation margins). Each term plays an important role in data analysis and decision making, so that each of them should be reflected while assessing the relationship between the A-IFSs no matter which correlation coefficient is used.

Acknowledgment

Partially supported by the Ministry of Science and Higher Education Grant Nr N N519 384936.

References

- [1] Aczel A. D. (1998), Complete business statistics. Richard D. Irvin, Inc.
- [2] Atanassov K. (1983), Intuitionistic Fuzzy Sets. VII ITKR Session. Sofia (Centr. Sci.-Techn. Libr. of Bulg. Acad. of Sci., 1697/84) (in Bulgarian).
- [3] Atanassov K. (1999), Intuitionistic Fuzzy Sets: Theory and Applications. Springer-Verlag.
- [4] Fredricks G. A., Nelsen R. B. (2007) On the relationship between Spearman's ρ and Kendall's τ for pairs of continuous random variables. *Journal of Statistical Planning and Inference*, 137, 2143–2150.
- [5] Helgason C. M. and Jobe T. H. (2003), Perception based reasoning and fuzzy cardinality provide direct measures of causality sensitive to initial conditions in the individual patient. *International Journal of Computational Cognition*, 1, 79–104.
- [6] Helgason C. M., Watkins F. A. and Jobe T. H. (2002), Measurable differences between sequential and parallel diagnostic decision processes for determining stroke subtype: A representation of interacting pathologies. *Thromb Haemost*, 88, 210–212.
- [7] Henrysson S. (1971), Gathering, analyzing, and using data on test items. In: *Educational Measurement*, ed. Thorndike R. L., Washington D.C. American Council on Education, 130–159.
- [8] Kendall M. G. (1938) A new measure of rank correlation. *Biometrika*, 30 (1-2) 81–93.
- [9] Kendall M. G. (1970) Rank correlation methods, fourth ed., Charles Griffin & Co., London.
- [10] Kenneth S. Kendler, Josef Parnas (2008) Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology. Johns Hopkins University Press.
- [11] Myers, J. L. and Well A. W. (2003) Research Design and Statistical Analysis (second edition ed.). Lawrence Erlbaum.
- [12] Nelsen R.B. (2001), Kendall tau metric. In: Hazewinkel, Michiel, Encyclopaedia of Mathematics, Springer, ISBN 978-1556080104,
- [13] Noether G.E. (1981) Why Kendall Tau? *Teaching Statistics*, 3(2), 41–43.
- [14] Rodgers J. L. and W. Alan Nicewander W. A. (1988) Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42(1), 59–66.
- [15] Szmidt E. and Kacprzyk J. (1996c) Remarks on some applications of intuitionistic fuzzy sets in decision making, *Notes on IFS*, 2(3), 22–31.
- [16] Szmidt E. and Kacprzyk J. (1997) On measuring distances between intuitionistic fuzzy sets, *Notes on IFS*, 3(4), 1–13.
- [17] Szmidt E. and Kacprzyk J. (1998) Group Decision Making under Intuitionistic Fuzzy Preference Relations. *IPMU'98*, 172–178.
- [18] Szmidt E. and Kacprzyk J. (2000) Distances between intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, 114(3), 505–518.
- [19] Szmidt E. and Kacprzyk J. (2000) On Measures on Consensus Under Intuitionistic Fuzzy Relations. *IPMU 2000*, 1454–1461.
- [20] Szmidt E., Kacprzyk J. (2001) Entropy for intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 118 (3), 467–477.
- [21] Szmidt E. and Kacprzyk J. (2001) Analysis of Consensus under Intuitionistic Fuzzy Preferences. *Proc. Int. Conf. in Fuzzy Logic and Technology*. De Montfort Univ. Leicester, UK, 79–82.
- [22] Szmidt E. and Kacprzyk J. (2002a) Analysis of Agreement in a Group of Experts via Distances Between Intuitionistic Fuzzy Preferences. *Proc. 9th Int. Conf. IPMU 2002*, 1859–1865.
- [23] Szmidt E. and Kacprzyk J. (2002b) An Intuitionistic Fuzzy Set Based Approach to Intelligent Data Analysis (an application to medical diagnosis). In A. Abraham, L.Jain, J. Kacprzyk (Eds.): *Recent Advances in Intelligent Paradigms and Applications*. Springer-Verlag, 57–70.
- [24] Szmidt E. and J. Kacprzyk J. (2002c) An Intuitionistic Fuzzy Set Based Approach to Intelligent Data Analysis (an application to medical diagnosis). In A. Abraham, L. Jain, J. Kacprzyk (Eds.): *Recent Advances in Intelligent Paradigms and Applications*. Springer-Verlag, 57–70.

- [25] Szmidt E. and Kacprzyk J. (2006) Distances Between Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. IEEE IS'06, 716–721.
- [26] Szmidt E. and Kacprzyk J. (2006) An Application of Intuitionistic Fuzzy Set Similarity Measures to a Multi-criteria Decision Making Problem. ICAISC 2006, LNAI 4029, Springer-Verlag, 314–323.
- [27] Szmidt E. and Kacprzyk J. (2007). Some problems with entropy measures for the Atanassov intuitionistic fuzzy sets. Applications of Fuzzy Sets Theory. LNAI 4578, 291–297. Springer-Verlag.
- [28] Szmidt E. and Kacprzyk J. (2007a). A New Similarity Measure for Intuitionistic Fuzzy Sets: Straightforward Approaches may not work. 2007 IEEE Conf. on Fuzzy Systems, 481–486.
- [29] Szmidt E. and Kacprzyk J. (2008) A new approach to ranking alternatives expressed via intuitionistic fuzzy sets. In: D. Ruan et al. (Eds.) Computational Intelligence in Decision and Control. World Scientific, 265–270.
- [30] Szmidt E. and Kacprzyk J. (2009). Amount of information and its reliability in the ranking of Atanassov's intuitionistic fuzzy alternatives. In: Recent Advances in decision Making, SCI 222. E. Rakus-Andersson, R. Yager, N. Ichalkaranje, L.C. Jain (Eds.), Springer-Verlag, 7–19.
- [31] Szmidt E., Kacprzyk J.: Ranking of Intuitionistic Fuzzy Alternatives in a Multi-criteria Decision Making Problem. In: Proceedings of the conference: NAFIPS 2009, Cincinnati, USA, June 14–17, 2009, IEEE, ISBN: 978-1-4244-4577-6.
- [32] Szmidt E. and Kacprzyk J. (2010) Correlation between intuitionistic fuzzy sets. LNAI 6178 (Computational Intelligence for Knowledge-Based Systems Design, Eds. E.Hullermeier, R. Kruse, F. Hoffmann), 169–177.
- [33] Szmidt E. and Kacprzyk J. (2010) The Spearman rank correlation coefficient between intuitionistic fuzzy sets. In: Proc. 2010 IEEE Int. Conf. on Intelligent Systems IEEE'IS 2010, London, 276–280.
- [34] Szmidt E. and Kukier M. (2006). Classification of Imbalanced and Overlapping Classes using Intuitionistic Fuzzy Sets. IEEE IS'06, London, 722–727.
- [35] Szmidt E. and Kukier M. (2008) A New Approach to Classification of Imbalanced Classes via Atanassov's Intuitionistic Fuzzy Sets. In: Hsiao-Fan Wang (Ed.): Intelligent Data Analysis : Developing New Methodologies Through Pattern Discovery and Recovery. Idea Group, 85–101.
- [36] Zadeh L.A. (1965) Fuzzy sets. Information and Control, 8, 338–353.
- [37] Zadeh L.A. (2002) Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. Journal of Statistical Planning and Inference 105, 233–264.
- [38] Zadeh L.A. (2006) Generalized theory of uncertainty (GTU)-principal concepts and ideas. Computational Statistics and Data Analysis, 51, 15–46.