Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

On correlation between two real interval sets

To cite this article: P Pandian and K Kavitha 2018 *J. Phys.: Conf. Ser.* **1000** 012055

IOP Conf. Series: Journal of Physics: Conf. Series **1000** (2018) 012055 doi:10.1088/1742-6596/1000/1/012055

On correlation between two real interval sets

P Pandian and K Kavitha

Department of Mathematics, School of Advanced Sciences, VIT University, Vellore-632014, Tamil Nadu, India

E-mail: ppandian@vit.ac.in

Abstract. A new formula for computing the correlation coefficient between two real interval sets is proposed. The proposed formula is based on conventional statistics , interval analysis and optimization theory and also, provides that the coefficient of correlation for interval sets is a sub-interval of the interval [-1,1], but not a number between -1 and 1. Numerical example is presented for understanding the proposed correlation evaluation in interval environment. Further, we extend the proposed correlation computation for two interval sets to the fuzzy environment.

1. Introduction

In conventional statistics [1] , the correlation coefficient defined on crisp sets have been studied. It's value is computed by using a formula and is a crisp real number in the closed interval [-1, 1]. The role of the correlation coefficient between the variables or attributes is more vital in the statistical analysis of crisp data. The correlation analysis is applied for studying the nature of relationship between the variables or attributes. In real life situations, many collected data have the feature of vagueness. The theory of fuzzy sets [8, 22] can be applied to model and to study this kind of data. It is interesting to study how the notation of correlation can be extended to interval sets or fuzzy sets. Now, examining the correlation between fuzzy sets is one of most important since fuzzy correlation and fuzzy regression have been applied in the field of data mining, quality control, marketing, image processing, robot control, medical diagnosis etc.. Many researchers have developed different measures of correlation between membership functions of fuzzy sets, grade values of fuzzy sets or sets of fuzzy numbers.

Murthy et.al. [10] proposed a measure of correlation between two membership functions satisfying some assumptions which lies in the interval $[-1, 1]$. The correlation coefficient between two fuzzy sets was computed by Yu [17]. Dug Hun Hong [8] introduced the correlation of fuzzy numbers. Ding-An Chiang and Nancy P. Lin [7] developed a method to calculate the correlation coefficient for fuzzy data using mathematical statistics which lies in $[-1, 1]$. Wang and Li [18] studied correlation and information energy of interval-valued fuzzy numbers . Chaudhuri and Bhattacharya [5] discussed the correlation between two fuzzy sets defined on the same universal support. Wen-Liang Hung and Jong-Wuu Wu [19] introduced a method to compute the correlation coefficient of fuzzy numbers by means of expected intervals which provides the coefficient correlation within [-1,1]. A method to compute the correlation coefficient of fuzzy numbers using α-cut concept was proposed by Wen-Liang Hung and Jong-Wuu Wu [20]. The fuzzy correlations were discussed by Daniel Ramot et al. [6]. Shiang-Tai Liu and Chiang Kao [15] proposed fuzzy measures for correlation coefficient of fuzzy numbers.

IOP Conf. Series: Journal of Physics: Conf. Series **1000** (2018) 012055 doi:10.1088/1742-6596/1000/1/012055

Fuzzy correlation and regression analysis for fuzzy data were studied by Yongshen Ni [16]. Rahim Saneifard and Rasoul Saneifard [12] presented a method to compute the correlation coefficient for fuzzy data. A new type of correlation coefficient between fuzzy numbers is proposed using the nearest weighted interval approximation of a fuzzy number by Saneifard and Nader Hassasi [14]. Berlin Wu et al. [3] studied correlation coefficient for fuzzy and interval data in Management Science. Wen Long et al. [21] focused on correlation analysis of interval data and proposed a comprehensive weighted correlation coefficient between two interval sets data. Berlin Wu and Chin Fen Hung [4] proposed a method for evaluating Pearson's correlation coefficient for fuzzy interval data.
In the literature, researchers have derived that the coefficient of correlation between interval sets or fuzzy sets is a crisp number which is in the interval [-1,1]. Since the fuzzy sets are not crisp, saying that the coefficient of correlation between them is a crisp number, is not logically true.
This paper is organized as follows: In Section 2, few definitions and basic results related to real intervals and fuzzy numbers are presented. A new formula is derived to compute the correlation coefficient for real interval sets based on interval analysis [11], basic concepts of conventional statistics [1] and optimization technique [13] in Section 3. In Section 4, using the derived correlation coefficient formula for real intervals, the correlation coefficient between two triangular fuzzy numbers is computed and finally, conclusion is presented in Section 5.

2. Preliminaries

The following definitions of the basic arithmetic operators and partial ordering on closed bounded intervals and fuzzy numbers are used in this paper which are found in [8, 10].
Let D denote the set of all closed bounded intervals on the real line R. That is, $D = \{\ [a,b]\ : a \leq b$, a and b are in $R\}$. An interval $[a,b]$ can also be written as follows

$$[a,b] = a + t(b-a); 0 \leq t \leq 1.$$

Definition 2.1. Let$A = [a,b]$and $B = [c,d]$ be in D. Then,
(i)$A \leq B$ if and only if $a \leq b$ and $c \leq d$, that is, $a - c \leq t(d-b); 0 \leq t \leq 1$ and
(ii)$A = B$ if and only if $a = b$ and $c = d$, that is, $a - c = t(d-b); 0 \leq t \leq 1$.

Definition 2.2. Let $A = [a,b] \in$ D . Then,A is said to be positive denoted by $A \geq 0$ if $a \geq 0$.

Definition 2.3. Let A be a classical set and $\mu_A(x)$ be a membership function from A to [0,1]. A fuzzy set $\tilde{A}$ with the membership function $\mu_{\tilde{A}}(x)$ is defined by

$$\tilde{A} = \left\{\ (x, \mu_{\tilde{A}}(x)) : x \in A \text{ and } \mu_{\tilde{A}}(x) \in [0,1]\ \right\}.$$

Definition 2.4. A fuzzy number $\tilde{a}$ is a triangular fuzzy number denoted by (a_1, a_2, a_3) where a_1, a_2 and a_3 are real numbers and its membership function $\mu_{\tilde{a}}(x)$ is given below.

$$\mu_{\tilde{a}}(x) = \begin{cases} 0 & ; \quad x \leq a_1 \\ \frac{(x-a_1)}{(a_2-a_1)} & ; a_1 \leq x < a_2 \\ 1 & ; \quad x = a_2 \\ \frac{(a_3-x)}{(a_3-a_2)} & ; a_2 < x \leq a_3 \\ 0 & ; x \geq a_3 \end{cases}$$

Let $F(R)$ be a set of all triangular fuzzy numbers over R, a set of real numbers.

Definition 2.5: Let $\tilde{a} = (a_1, a_2, a_3)$and $\tilde{b} = (b_1, b_2, b_3)$be in $F(R)$. Then,
(i) $\tilde{a}$ and $\tilde{b}$ are said to be equal if $a_i = b_i$, i =1,2,3, and
(ii) $\tilde{a}$ is said to be less than or equal $\tilde{b}$ if $a_i \leq b_i$, i =1,2,3.

Definition 2.5. Let $\tilde{a} = (a_1, a_2, a_3)$ be in $F(R)$. Then,$\tilde{a}$ is said to be positive if $a_1 \geq 0$.
Definition 2.6. Let $\tilde{a} = (a_1, a_2, a_3)$ be in $F(R)$. Then, $\tilde{a}$ is said to be integer if $a_{\mathrm{i}} \geq 0$, i =1,2,3 are integers.

An interval representation of a fuzzy number is very useful and important. By using the interval (or $\alpha - cut$) representation , it is possible to derive some results in the field of fuzzy numbers based on the field of interval number analysis
A triangular fuzzy number (a, b, c, d) can be represented as an interval number (or $\alpha - cut$) form as follows.

$$(a, b, c) = [a + (b - a)\alpha, c - (c - b)\alpha]; 0 \leq \alpha \leq 1. \tag{1}$$

3. Interval correlation

In this section, a formula for computing correlation between two interval sets is developed based on the conventional statistics, interval analysis and optimization theory.
Let $[X] = \left\{ \mathrm{X_i} = [x_i^1, x_i^2], \mathrm{i} = 1, 2, ..., \mathrm{n} \right\}$ be a set of n real interval x-values and $[Y\} = \left\{ \mathrm{Y_i} = [y_i^1, y_i^2], \mathrm{i} = 1, 2, ..., \mathrm{n} \right\}$ be a set of n real interval y-values.
Now, the point representation of ith intervals,X_i and Y_i are written as follows:
$X_i = tx_i^1 + (1 - t)x_i^2; 0 \leq t \leq 1$ and $Y_i = \theta y_i^1 + (1 - \theta)y_i^2; 0 \leq \theta \leq 1$.
Now, the mean of $X_i's$, $\bar{X} = t\left(\overline{x^1}\right) + (1 - t)\overline{(x^2)}$ and
the mean of $Y_i's$, $\bar{Y} = \theta\left(\overline{y^1}\right) + (1 - \theta)\overline{(y^2)}$
Now, $X_i - \bar{X} = t\left(x_i^1 - \left(\overline{x^1}\right)\right) + (1 - t)\left(x_i^2 - \overline{(x^2)}\right)$ and

$$Y_i - \bar{Y} = \theta\left(y_i^1 - \left(\overline{y^1}\right)\right) + (1 - \theta)\left(y_i^2 - \overline{(y^2)}\right).$$

Now,

$$\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = \sum_{i=1}^{n} \left(\left[t\left(x_i^1 - \left(\overline{x^1}\right)\right) + (1 - t)\left(x_i^2 - \overline{(x^2)}\right)\right]\right.$$
$$\left.\left[\theta\left(y_i^1 - \left(\overline{y^1}\right)\right) + (1 - \theta)\left(y_i^2 - \overline{(y^2)}\right)\right]\right)$$

$$= \sum_{i=1}^{n}\left(x_i^2 - \left(\overline{x^2}\right)\right)\left(y_i^2 - \overline{(y^2)}\right) + \left[\begin{array}{l} \sum_{i=1}^{n}\left(x_i^1 - \left(\overline{x^1}\right)\right)\left(y_i^2 - \overline{(y^2)}\right) \\ -\sum_{i=1}^{n}\left(x_i^2 - \left(\overline{x^2}\right)\right)\left(y_i^2 - \overline{(y^2)}\right) \end{array}\right] t$$
$$+ \left[\sum_{i=1}^{n}\left(x_i^2 - \left(\overline{x^2}\right)\right)\left(y_i^1 - \overline{(y^1)}\right) - \sum_{i=1}^{n}\left(x_i^2 - \left(\overline{x^2}\right)\right)\left(y_i^2 - \overline{(y^2)}\right)\right]\theta$$
$$+ \left[\begin{array}{l} \sum_{i=1}^{n}\left(x_i^1 - \left(\overline{x^1}\right)\right)\left(y_i^1 - \overline{(y^1)}\right) - \sum_{i=1}^{n}\left(x_i^2 - \left(\overline{x^2}\right)\right)\left(y_i^1 - \overline{(y^1)}\right) \\ -\sum_{i=1}^{n}\left(x_i^1 - \left(\overline{x^1}\right)\right)\left(y_i^2 - \overline{(y^2)}\right) + \sum_{i=1}^{n}\left(x_i^2 - \left(\overline{x^2}\right)\right)\left(y_i^2 - \overline{(y^2)}\right) \end{array}\right] t\theta$$
$$= c_{22} + (c_{12} - c_{22})t + (c_{21} - c_{22})\theta + (c_{11} - c_{12} - c_{21} + c_{22})t\theta \tag{2}$$

where

$$c_{11} = \sum_{i=1}^{n}\left(x_i^1 - (\overline{x^1})\right)\left(y_i^1 - \overline{(y^1)}\right), c_{12} = \sum_{i=1}^{n}\left(x_i^1 - (\overline{x^1})\right)\left(y_i^2 - \overline{(y^2)}\right),$$
$$c_{21} = \sum_{i=1}^{n}\left(x_i^2 - (\overline{x^2})\right)\left(y_i^1 - \overline{(y^1)}\right) \text{ and } c_{22} = \sum_{i=1}^{n}\left(x_i^2 - (\overline{x^2})\right)\left(y_i^2 - \overline{(y^2)}\right).$$

Now,

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=\sum_{i=1}^{n}\left(t\left(x_i^1-\left(\overline{x^1}\right)\right)+(1-t)\left(x_i^2-\overline{\left(x^2\right)}\right)\right)^2$$
$$=\left(\sum_{i=1}^{n}\left(x_i^1-\left(\overline{x^1}\right)\right)^2\right)t^2+\left(\sum_{i=1}^{n}\left(x_i^2-\left(\overline{x^2}\right)\right)^2\right)(1-t)^2$$
$$+2\left(\sum_{i=1}^{n}\left(x_i^1-\left(\overline{x^1}\right)\right)\left(x_i^2-\overline{\left(x^2\right)}\right)\right)t(1-t)$$
$$=\left(\sum_{i=1}^{n}\left(x_i^2-\left(\overline{x^2}\right)\right)^2\right)+2\left(\begin{array}{c}\sum_{i=1}^{n}\left(x_i^1-\left(\overline{x^1}\right)\right)\left(x_i^2-\overline{\left(x^2\right)}\right)\\-\left(\sum_{i=1}^{n}\left(x_i^2-\left(\overline{x^2}\right)\right)^2\right)\end{array}\right)t$$
$$+\left(\left(\sum_{i=1}^{n}\left(x_i^1-\left(\overline{x^1}\right)\right)^2\right)+\left(\sum_{i=1}^{n}\left(x_i^2-\left(\overline{x^2}\right)\right)^2\right)\right.$$
$$\left.-2\left(\sum_{i=1}^{n}\left(x_i^1-\left(\overline{x^1}\right)\right)\left(x_i^2-\overline{\left(x^2\right)}\right)\right)\right)t^2$$
$$=a_{22}+2(a_{12}-a_{22})t+(a_{11}+a_{22}-2a_{12})t^2, \tag{3}$$

where

$$a_{11}=\sum_{i=1}^{n}\left(x_i^1-\left(\overline{x^1}\right)\right)^2, a_{22}=\sum_{i=1}^{n}\left(x_i^2-\left(\overline{x^2}\right)\right)^2 \text{ and}$$
$$a_{12}=\sum\left(x_i^1-\left(\overline{x^1}\right)\right)\left(x_i^2-\overline{\left(x^2\right)}\right).$$

Now,

$$\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2=\sum_{i=1}^{n}\left(\theta\left(y_i^1-\left(\overline{y^1}\right)\right)+(1-\theta)\left(y_i^2-\overline{\left(y^2\right)}\right)\right)^2$$
$$=\left(\sum_{i=1}^{n}\left(y_i^1-\left(\overline{y^1}\right)\right)^2\right)\theta^2+\left(\sum_{i=1}^{n}\left(y_i^2-\left(\overline{y^2}\right)\right)^2\right)(1-\theta)^2$$
$$+2\left(\sum_{i=1}^{n}\left(y_i^1-\left(\overline{y^1}\right)\right)\left(y_i^2-\overline{\left(y^2\right)}\right)\right)\theta(1-\theta)$$

$$=\left(\sum_{i=1}^{n}\left(y_i^2-\left(\overline{y^2}\right)\right)^2\right)+2\left(\begin{array}{c}\sum_{i=1}^{n}\left(y_i^1-\left(\overline{y^1}\right)\right)\left(y_i^2-\overline{\left(y^2\right)}\right)\\-\left(\sum_{i=1}^{n}\left(y_i^2-\left(\overline{y^2}\right)\right)^2\right)\end{array}\right)\theta$$
$$+\left(\begin{array}{c}\left(\sum_{i=1}^{n}\left(y_i^1-\left(\overline{y^1}\right)\right)^2\right)+\left(\sum_{i=1}^{n}\left(y_i^2-\left(\overline{y^2}\right)\right)^2\right)\\-2\left(\sum_{i=1}^{n}\left(y_i^1-\left(\overline{y^1}\right)\right)\left(y_i^2-\overline{\left(y^2\right)}\right)\right)\end{array}\right)\theta^2$$
$$=b_{22}+2(b_{12}-b_{22})\theta+(b_{11}+b_{22}-2b_{12})\theta^2 \tag{4}$$

where

$$b_{11} = \sum_{i=1}^{n} \left(y_i^1 - \left(\overline{y^1} \right) \right)^2, b_{22} = \sum_{i=1}^{n} \left(y_i^2 - \left(\overline{y^2} \right) \right)^2 \text{ and}$$

$$b_{12} = \sum_{i=1}^{n} \left(y_i^1 - \left(\overline{y^1} \right) \right) \left(y_i^2 - \overline{(y^2)} \right).$$

Let $r(t,\theta)$ be the correlation coefficient between$X_i's$ values and $Y_i's$ values since each $X_i's$ is a function of t, $0 \le t \le 1$ and each $Y_i's$ is a function θ, $0 \le \theta \le 1$. Note that for each pair of (t,θ), we can find the value of $r(t,\theta)$.
Now, as per basic correlation coefficient formula [1] , $r(t,\theta)$ is given as below:

$$r(t,\theta) = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) \left(Y_i - \bar{Y} \right)}{\sqrt{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2} \times \sqrt{\sum_{i=1}^{n} \left(Y_i - \bar{Y} \right)^2}}$$

Now, from (2), (3) and (4), the above expression $r(t,\theta)$ can be reformed as follows:

$$= \frac{\begin{array}{c} c_{22} + (c_{12} - c_{22})t + (c_{21} - c_{22})\theta \\ +(c_{11} - c_{12} - c_{21} + c_{22})t\theta \end{array}}{\begin{array}{c} \sqrt{a_{22} + 2(a_{12} - a_{22})t + (a_{11} + a_{22} - 2a_{12})t^2} \\ \times \sqrt{b_{22} + 2(b_{12} - b_{22})\theta + (b_{11} + b_{22} - 2b_{12})\theta^2} \end{array}}$$

Since the crisp correlation coefficient always lies between -1 and 1, $r(t,\theta)$ has a real value in $[-1,1]$ for a fixed pair $(t^o,\theta^o)$where $t^o \in [0,1]$ and $\theta^o \in [0.1]$. Therefore, $r(t,\theta)$takes any real value in $[-1,1]$ for each pair of (t,θ), $0 \le t,\theta \le 1$ and $r(t,\theta)$ is not the same value for all pairs of (t,θ),$0 \le t,\theta \le 1$. Thus, the correlation coefficient between two interval sets is an interval and a sub-interval of $[-1,1]$. The interval correlation coefficient can computed from the optimum values of $r(t,\theta)$ subject to $0 \le t \le 1$ and $0 \le \theta \le 1$.
Now, let r_1 be the solution of the non-linear programming problem:
Minimize $r(t,\theta)$ subject to $0 \le t \le 1$ and $0 \le \theta \le 1$ and r_2 be the solution of the non-linear programming problem:
Maximize $r(t,\theta)$ subject to $0 \le t \le 1$ and $0 \le \theta \le 1$.
Now, we may view that r_1 is the minimum level correlation between $\{ \mathrm{X}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ and $\{ \mathrm{Y}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ and r_2 is the maximum level correlation between $\{ \mathrm{X}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ and $\{ \mathrm{Y}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$, the interval $[r] = [r_1, r_2]$ is the interval correlation coefficient between $\{ \mathrm{X}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ and $\{ \mathrm{Y}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$. and also, we can understand that the correlation coefficient between any pair of point sets of the given interval sets is not more than r_2 and not less than r_1. These concepts are very important and useful for taking decision on the relationship between interval sets. So, we define that correlation coefficient between the interval sets between $\{ \mathrm{X}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ and $\{ \mathrm{Y}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$, $[r]$ is $[r] = [r_1, r_2]$. This means that the correlation coefficient between the interval sets varies from r_1 to r_2.

Remark 3.1. The values r_1 to r_2 can be determined using MAT LAB software.

Remark 3.2. (i) If r_1 is positive, then the intervals $\{ \mathrm{X}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ and $\{ \mathrm{Y}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ are positively correlated ;
(ii) If r_2 is negative, then the intervals $\{ \mathrm{X}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ and $\{ \mathrm{Y}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ are negatively correlated;
(iii) If r_1 is negative and r_2 is positive, then the intervals $\{ \mathrm{X}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ and $\{ \mathrm{Y}_i, \mathrm{i} = 1,2,...,\mathrm{n} \}$ are negative-positively correlated and

IOP Conf. Series: Journal of Physics: Conf. Series **1000** (2018) 012055 doi:10.1088/1742-6596/1000/1/012055

(iv) If r_1 and r_2 are zero, then the intervals $\{ \mathrm{X}_i, \mathrm{i} = 1, 2, ..., \mathrm{n} \}$ and $\{ \mathrm{Y}_i, \mathrm{i} = 1, 2, ..., \mathrm{n} \}$ are not linearly correlated.
Now, we explain for computing the coefficient between two real interval sets using the proposed formula with the following numerical example:

Example 3.3. *Compute the interval correlation between two interval sets [X] and [Y] from the following interval data (Table 1):*

[X]	[Y]
[1.5 ; 2.5]	[3.5 ; 4.5]
[3 ; 4]	[5 ; 6]
[4.5 ; 6.5]	[6.5 ; 8.5]
[6.5 ; 7.5]	[6 ; 7]
[8 ; 9]	[8 ; 9]
[9.5 ; 11.5]	[7 ; 9]
[10.5 ; 11.5]	[10 ; 11]
[12 ; 13]	[9 ; 10]

Table 1. Interval data

For understanding, we summarize the values of r_1 and r_2 for different values of t and θ in the following table.

t	θ	r_1	r_2	$[r]$
0	0	0.9153	0.9153	0.9153
0	0.5	0.9134	0.9134	0.9134
0.25	0.25	0.9151	0.9151	0.9151
0.25	0.75	0.9111	0.9111	0.9111
0.5	0.5	0.9147	0.9147	0.9147
0.5	1	0.9086	0.9086	0.9086
0.75	0	0.9061	0.9061	0.9061
0.75	0.75	0.9143	0.9143	0.9143
1	0.25	0.9082	0.9082	0.9082
1	1	0.9135	0.9135	0.9135

Table 2. Values of r_1 and r_2

Note that for each (t, θ), we find the unique correlation coefficient.

Now, using the given interval data, $r(t, \theta)$ is given as follows:

$$r(t,\theta) = \frac{51.8125 + (-1.25)t + (-1.625)\theta + (1.5)t\theta}{\sqrt{98.9688 + (3.25)t + (1.5)t^2} \times \sqrt{32.3750 + (-2.5)\theta + (1.5)\theta^2}}$$

Now, using MATLAB software, we obtain the values r_1 and $r_2$as given below:
$r_1 = 0.9013$ when $t = 1.0000$ and θ= 0.0000 and r_2 =0.9159 when $t = 0.0000$ and θ= 0.1606 . Therefore, the interval correlation coefficient between given two interval data, $[r] = [0.9013, 0.9159]$ and two interval data are positively correlated.

IOP Conf. Series: Journal of Physics: Conf. Series **1000** (2018) 012055 doi:10.1088/1742-6596/1000/1/012055

4. Fuzzy correlation

In this section, the correlation coefficient between two sets of fuzzy numbers is computed by using the formula developed in the Section 3..
Let $\tilde{X} = \left\{ \tilde{\mathrm{X}}_{\mathrm{i}} = (x_i^1, x_i^2, x_i^3), \mathrm{i} = 1, 2, ..., \mathrm{n} \right\}$ be a set of n triangular fuzzy numbers and $\tilde{Y} = \left\{ \tilde{Y}_{\mathrm{i}} = (y_i^1, y_i^2, y_i^3), \mathrm{i} = 1, 2, ..., \mathrm{n} \right\}$ be an another set of n triangular fuzzy numbers.
Now, using (1), the given two fuzzy sets can be put as interval sets as follows:

$$\left[\tilde{X}\right] = \left\{ \begin{array}{l} \left[\tilde{\mathrm{X}}_{\mathrm{i}}\right] = \ [x_i^1 + (x_2^2 - x_i^1)\alpha, x_i^3 - (x_i^3 - x_i^2)\alpha] \\ ,0 \leq \alpha \leq 1; \mathrm{i} = 1, 2, ..., \mathrm{n} \end{array} \right\}$$

and

$$\left[\tilde{Y}\right] = \left\{ \begin{array}{l} \left[\tilde{Y}_{\mathrm{i}}\right] = \ [y_i^1 + (y_i^2 - y_i^1)\alpha, y_i^3 - (y_i^3 - y_i^2)\alpha] \\ ,0 \leq \alpha \leq 1; \mathrm{i} = 1, 2, ..., \mathrm{n} \end{array} \right\}.$$

Now, using the proposed method in the section 3., we compute $[\tilde{r}(t, \theta)]$ for each value of α, $0 \leq \alpha \leq 1$ as follows:
Let r^L be the solution of minimize $[\tilde{r}(t, \theta)]$ subject to $0 \leq t, \theta \leq 1$ for each value of $\alpha, 0 \leq \alpha \leq 1$?and r^U be the solution of maximize $[\tilde{r}(t, \theta)]$ subject to $0 \leq t, \theta \leq 1$for each value of $\alpha, 0 \leq \alpha \leq 1$. Then, the interval correlation coefficient between the interval sets $\left[\tilde{X}\right]$ and $\left[\tilde{Y}\right]$ is $\left[r^L, r^U\right]$,for each value of $\alpha, 0 \leq \alpha \leq 1$. Finally, using the relation (1) and $\left[r^L, r^U\right]$,$0 \leq \alpha \leq 1$, we can obtain the fuzzy correlation coefficient between the fuzzy sets $\tilde{X}$ and $\tilde{Y}$.

Remark 4.1. (i) If r^L, $0 \leq \alpha \leq 1$ are positive, then the fuzzy sets $\tilde{X}$ and $\tilde{Y}$ are positively correlated ;
(ii) If r^U, $0 \leq \alpha \leq 1$ are negative, then the fuzzy sets $\tilde{X}$ and $\tilde{Y}$ are negatively correlated ;
(iii) If r^L is negative and r^U is positive, for some α, $0 \leq \alpha \leq 1$ then the fuzzy sets $\tilde{X}$ and $\tilde{Y}$ are negative-positively correlated and
(iv) If r^L and r^U are zero for $0 \leq \alpha \leq 1$, then the fuzzy sets $\tilde{X}$ and $\tilde{Y}$ are not linearly correlated.
Remark 4.2. : The results in the Remark 4.1. can also be discussed based on the value

$$\alpha,\ 0 \leq \alpha \leq 1.$$

The proposed approach for computing correlation coefficient for sets of fuzzy numbers method is illustrated by using the following numerical example.

Example 4.3. *Compute the correlation coefficient between two sets of triangular fuzzy numbers $\tilde{X}$ and $\tilde{Y}$ from the following fuzzy data as table form (Table 2.):*

Now, the interval form of the given fuzzy data is given below: Now, for each value of $\alpha, 0 \leq \alpha \leq 1$, we obtain the following values of r^L and r^U which are tabulated as given below: Therefore, the fuzzy correlation between $\tilde{X}$ and $\tilde{Y}$ is $(0.5059, 0.5298, 0.5786)$
Now, the graph of the fuzzy correlation coefficient between $\tilde{X}$ and $\tilde{Y}$ is given below:
Remark 4.4. The proposed formula for finding correlation coefficient between triangular fuzzy numbers sets can be used for trapezoidal fuzzy numbers sets since the correlation formula is derived from the real interval sets.

IOP Conf. Series: Journal of Physics: Conf. Series **1000** (2018) 012055 doi:10.1088/1742-6596/1000/1/012055

$\tilde{X}$	$\tilde{Y}$
(40,43,45)	(95,99,102)
(20,21,23)	(60,65,68)
(22,25,27)	(75,79,82)
(40,42,45)	(70,75,80)
(55,57,60)	(85,87,90)
(54,59,62)	(79,81,83)

Table 3. Fuzzy data

$[\tilde{X}]$	$[\tilde{Y}]$
$[40+3\alpha, 45-2\alpha]$	$[95+4\alpha, 102-3\alpha]$
$[20+\alpha, 23-2\alpha]$	$[60+5\alpha, 68-3\alpha]$
$[22+3\alpha, 27-2\alpha]$	$[75+4\alpha, 82-3\alpha]$
$[40+2\alpha, 45-3\alpha]$	$[70+5\alpha, 80-5\alpha]$
$[55+2\alpha, 60-3\alpha]$	$[85+2\alpha, 90-3\alpha]$
$[54+5\alpha, 62-3\alpha]$	$[79+2\alpha, 83-2\alpha]$

Table 4. Interval form of fuzzy data

S.No	$\alpha - value$	r^L	r^U
1	0	0.5059	0.5786
2	0.1	0.5084	0.5746
3	0.2	0.5108	0.5704
4	0.3	0.5133	0.5660
5	0.4	0.5157	0.5614
6	0.5	0.5181	0.5566
7	0.6	0.5205	0.5516
8	0.7	0.5228	0.5465
9	0.8	0.5252	0.5411
10	0.9	0.5275	0.5356
11	1	0.5298	0.5298

Table 5. Values of r^L and r^U

5. Conclusion:

In this paper, the correlation between fuzzy number sets have been studied. A computing formula is derived for finding the correlation coefficient between two real intervals by using mathematical statistics, interval analysis and optimization technique. The correlation coefficient between interval sets by the derived is provided an interval. Using the derived correlation formula, the correlation coefficient between two triangular fuzzy numbers is computed and found that it is a fuzzy number. The interval correlation for interval sets is an interval subset of the interval [-1,1] and the fuzzy correlation for fuzzy number sets is a fuzzy number whose support is a subset of the interval set [-1,1]. The correlation coefficients computed for interval sets and fuzzy sets by our proposed approach tell us the strength of the relationship between the interval sets at each pair of point sets and the fuzzy number sets at each level α, $0 \leq \alpha \leq 1$. The correlation coefficient computed for interval sets and fuzzy number sets by the proposed approach, give us more useful

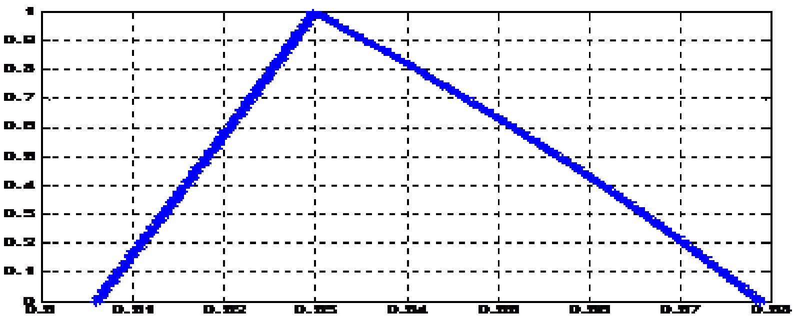

Figure 1. graph of the fuzzy correlation coefficient

information than the correlation coefficients computed by using all existing approaches.

References

[1] S F Arnold 1990 *Mathematical Statistics* Prentice-Hall, Englewood Cliffs, NJ.
[2] Bellman R and L Zadeh 1970 Decision making in fuzzy environment *Manage Sci.* **17** 141-164
[3] Berlin-Wu et al. 2015 Correlation Evaluation with fuzzy data and its application in the management science *Econometrics of Risk* **583** 273-285
[4] Berlin Wu and Chin Fen Hung 2016 Innovative correlation coefficient measurement with fuzzy data *Mathematical Problems in Engineering* **2016** 1-12
[5] Chaudhuri B B and Bhattacharya A 2001 On correlation between two fuzzy sets *Fuzzy Sets and Systems* **118** 447-456
[6] Daniel Ramot et al. 2001 On fuzzy correlations, *IEEE Transactions on Systems Man and Cybernetics-part B: Cybernetics* **31** 381-390
[7] Ding-An Chiang and Nancy P Lin 1999 Correlation of fuzzy sets, *Fuzzy Sets and Systems* **102** 221-226
[8] Hong D H and Hwang S Y 1995 A note on the correlation of fuzzy numbers *Fuzzy Sets and Systems* **75** 77-81
[9] Klir G J and B Yuan 2008 *Fuzzy sets and Fuzzy logic: Theory and Applications* Prentice Hall New Jersey
[10] Murthy C A et al. 1985 Correlation between two fuzzy membership functions *Fuzzy Sets and Systems* **17** 2338
[11] Moore R E 1979 *Methods and Applications of Interval Analysis* SIAM Philadelphia
[12] Rahim Saneifard and RasoulSaneifard 2012 Correlation Coefficient Between Fuzzy Numbers Based On Central Interval *Journal of Fuzzy Set Valued Analysis* **2012** 1- 9
[13] Rao S S 1984 *Optimization : Theory and Applications* 2nd edition Wiley Eastern New Delhi
[14] Saneifard R and Nader Hassasi 2013 Parametric correlation coefficient of fuzzy numbers *African Journal of Business Management* **7** 3410-3415
[15] Shiang-Tai Liu and Chiang Kao 2002 Fuzzy measures for correlation coefficient of fuzzy numbers *Fuzzy Sets and Systems* **128** 267-275
[16] Yongshen Ni 2005 *Fuzzy correlation and regression analysis,* Ph.D. Thesis University of Oklahoma Graduate College Norman
[17] Yu C 1993 Correlation of fuzzy numbers *Fuzzy Sets and Systems* **55** 303-307
[18] Wang G and Li X 1999 Correlation and information energy of interval-valued fuzzy numbers *Fuzzy Sets and Systems* **103** 169-175
[19] Wen-Liang Hung and Jong-Wuu Wu 2001 A note on the correlation of fuzzy numbers by expected interval *International Journal of Uncertainty Fuzziness and Knowledge Based Systems* **9** 517-523
[20] Wen-Liang Hung and Jong-Wuu Wu 2002 Correlation of fuzzy numbers by α-cut method, *International Journal of Uncertainty Fuzziness and Knowledge Based Systems* **10** 725-735
[21] Wen Long et al. 2016 Correlation Analysis of Industry Sectors in Chinas Stock Markets Based on Interval Data *Filomat* **30** 3999-4013
[22] Zadeh L A 1965 Fuzzy sets, *Information and Control* **8** 338-353.