ISSN (e): 2250 – 3005 || Volume, 08 || Issue, 1|| January – 2018 ||
International Journal of Computational Engineering Research (IJCER)

Use of Statistical Implementation in Measures of Location for Imprecise Data

Dr. Sarjerao Krushnat Powar
Assistant Professor,
Smt. Kasturbai Walchand College of Arts and Science, Sangli,

Abstract
Statistical analysis is usually done on data sets that contain precise information. However, data obtained may not be precise always. Thismay be due to vagueness in the process of measurement itself or due to the values being expressed in linguistic terms. Data that are not precise are referred to as imprecise data or fuzzy data. Developing statistical measures for imprecise data has been pursued by researchers during the past decade. However, the methodologies adopted involve operations on fuzzy numbers through α-cut approach, thereby making the process difficult. Liu (2007) developed a mathematical theory for studying the behavior of fuzzy phenomena known as "Credibility Theory". This theory provides operations on fuzzy numbers that do not depend on the α-cut approach.In this paper, a methodology for obtaining measures of location for imprecise data is developed by using the concepts available in "Credibility Theory". Numerical Illustration for calculating the proposed measures is also provided.
Keywords: *Credibility theory, credibility distribution, fuzzy number, expectation, variance, percentiles*

I. INTRODUCTION

Measures of location play significant role in understanding the behavior of data. Well known measures like median, quartiles etc. are often employed to determine the symmetricity and spread of a given set of data. These measures are often employed in non-parametric inferential procedures like sign test, median test etc. Gibbons and Chakraborti[1] contain details about these testing procedures. Often data obtained from experimental studies are not precise. This may be due to the inherent variability in the process that governs the experimental conditions or may be due to faulty measurements. Such type of data is generally referred to as imprecise or fuzzy data. For example, measuring blood pressure level of an individual is not always precise but rather approximated by the judgment of the investigator. Moreover, the value may be expressed in linguistic terms like mild, low etc. instead of their observed values. Fuzzy data are usually expressed in terms of membership values that denote the degree of belongingness of observed values to some underlying set. Consider the example of reporting the systolic blood pressure level of a patient. Systolic value less than 90 is usually considered "low" in medical diagnostics. Thus two patients with systolic levels less than 90 are grouped into "low" category no matter how much these values are closer to 90. It should however, be emphasized that a patient having blood pressure level 88 is less at risk of belonging to "low" category than compared to a patient having blood pressure level 79. Thus, intuitionally, values closer to 90 should belong to "low" category with lesser degree of belongingness than values that are far less from 90. Thus to get more insight into the data one has to consider not only the observed value but also its belongingness to the underlying set. This can be done by using the fuzzy set theory developed by Zadeh[2]by treating the observed values as fuzzy numbers. Kruse and Meyer, Viertl, Carlsson and Fuller, Bodjanova, Pashaet al.[3–7]discuss methods to determine the mean, median, dispersion of a fuzzy numberusing the fuzzy set theory.However,statistical measures usually deal with a set of observed values rather than a single value and therefore it would be appropriate to develop measures that deal with a set of fuzzy numbers. Hence, in this paper, a new methodology is introduced to calculate the various measures of location like median, quartiles etc. of a set of imprecise data. The proposed method makes use of some results available in credibility theory due to Liu [8]. The advantage of using credibility theory is that it does depend on α-cut approach to perform operations on fuzzy numbers and hence easy to implement. The rest of the paper is organized as follows:

In the next section, preliminary concepts of credibility theory are introduced and some important definitions are given. Section "Percentiles of Fuzzy Numbers" presents the proposed methodology for finding various measures of location of a set of imprecise data. Numerical illustration of the proposed measures is provided in the subsequent section and the conclusion is given in the last section.

II. PRELIMINARIES

To model imprecise values using fuzzy sets, Liu [8] developed a theory called 'Credibility Theory'. This section contains the definitions of certain important concepts based on credibility theory. More details can be found in Liu [8].

Credibility Measure

Let Θ be a nonempty set, and P the power set of Θ (i.e., the largest σ -algebra over Θ). Each element in P is called an event. The set function Cr is called a credibility measure if it satisfies the following axioms:

Axiom 1: $Cr\{\Theta\} = 1$.

Axiom 2: $Cr\{A\} \le Cr\{B\}$ whenever $A \subset B$ Axiom 3: $Cr\{A\} + Cr\{A^C\} \equiv 1$ for any event A Axiom 4:

$Cr\left\{\bigcup_i A_i\right\} = \sup_i Cr\{A_i\}$ for any events $\{A_i\}$ with $\sup_i Cr\{A_i\} < 0.5$.

$Cr\{A\}$ indicates the credibility that the event A will occur. The triplet (Θ, P, Cr) is called a credibility space.

Fuzzy Variable: A fuzzy variable is a (measurable) function from a credibility space (Θ, P, Cr) to the set of real numbers.

Remark:Any function of fuzzy variables defined on the same credibility space is again a fuzzy variable, i.e., the sum or product of two or more fuzzy variables is again a fuzzy variable.

Membership Function: The membership function μ of a fuzzy variable ξ defined on the credibility space (Θ, P, Cr) is given by $\mu(x) = \min\left(2Cr\{\xi = x\}, 1\right), x \in \Re$.

Remark: Membership function represents the degree that the fuzzy variable ξ takes on some prescribed value and it always takes values in the interval [0,1].Depending on the shape of the curve of membership functions, fuzzy variables are defined appropriately. Often used fuzzy variables are triangular and Trapezoidal fuzzy variable.

Triangular Fuzzy Variable: A fuzzy variable fully determined by the triplet (a, b, c) of crisp numbers with $(a < b < c)$ and whose membership function is given by

$$\mu(x) = \begin{cases} \dfrac{x-a}{b-a}, & if \quad a \le x \le b \\ \dfrac{x-c}{b-c}, & if \quad b \le x \le c \\ 0, & otherwise \end{cases}$$

is called a triangular fuzzy variable. If $b - a = c - b$ then the triangular fuzzy variable is said to be symmetric.

Trapezoidal fuzzy variable: A fuzzy variable fully determined by the quadruplet (a, b, c, d) of crisp numbers with $(a < b < c < d)$ and whose membership function is given by

$$\mu(x) = \begin{cases} \dfrac{x-a}{b-a}, & if \quad a \le x \le b \\ 1, & if \quad b \le x \le c \\ \dfrac{x-d}{c-d}, & if \quad c \le x \le d \\ 0, & otherwise \end{cases}$$

is called a trapezoidal fuzzy variable. If $b - a = d - c$ then the trapezoidal fuzzy variable is said to be symmetric. Also if $b = c$ then the trapezoidal fuzzy variable reduces to a triangular fuzzy variable.

Credibility Distribution: The credibility distribution $\Phi : \Re \to [0,1]$ of a fuzzy variable ξ is defined by $\Phi(x) = Cr\{\theta \in \Theta \mid \xi(\theta) \le x\}$.That is, $\Phi(x)$ is the credibility that the fuzzy variable ξ takes a value less than or equal to x .

Expected Value of a Fuzzy Variable:
The expected value of a fuzzy variable ξ denoted by $E(\xi)$ is defined as

$$E(\xi) = \int_0^{+\infty} Cr\{\xi \ge r\}dr - \int_{-\infty}^{0} Cr\{\xi \le r\}dr$$

provided that at least one of the integrals is finite.

Variance of a Fuzzy Variable: The variance of a fuzzy variable ξ with finite expected value e and denoted by $V(\xi)$ is defined as $V(\xi) = E(\xi - e)^2 = \int_0^{+\infty} Cr\{(\xi - e)^2 \ge r\}dr$. Variance of a fuzzy variable provides a measure of spread of the membership function around its expected value.

PERCENTILESOFFUZZY NUMBERS
Consider a set of $'n'$ fuzzynumbers $\xi_1, \xi_2, \ldots, \xi_n$. The $k^{th}, k = 1,2,\ldots,99$ percentile of these fuzzy numbers is defined as the value corresponding to $\left[\frac{(n+1)k}{100}\right]^{th}$ position in the ordered arrangement of $\xi_1, \xi_2, \ldots, \xi_n$.

This definition is similar to that of determining percentiles of a set of crisp numbers. The problem is how to order the set of fuzzy numbers. Li et al.[9] have proposed a method to order fuzzy numbers using their expected value and variance. Let ξ and η be two fuzzy variables defined on the credibility space (Θ, P, Cr). The ordering of ξ and η is defined as follows:

1. $\xi \succ \eta$ if and only if $E[\xi] > E[\eta]$ or $E[\xi] = E[\eta]$ and $V(\xi) < V(\eta)$.
2. $\xi \prec \eta$ if and only if $E[\xi] < E[\eta]$ or $E[\xi] = E[\eta]$ and $V(\xi) > V(\eta)$.
3. $\xi \sim \eta$ if and only if $E[\xi] = E[\eta]$ and $V(\xi) = V(\eta)$.

Thus ξ is greater than or equal to η if and only if $\xi \succ \eta$ or $\xi \sim \eta$ and ξ is lesser than or equal to η if and only if $\xi \prec \eta$ or $\xi \sim \eta$. The above definition can be extended to more than two variables. Using the above definition, one can obtain the k^{th} percentile of fuzzy numbers $\xi_1, \xi_2, \ldots, \xi_n$ as follows:

Step 1: Compute the expected value and variance of the fuzzy numbers $\xi_1, \xi_2, \ldots, \xi_n$.

Step 2: Arrange the $'n'$ fuzzy numbers in increasing order by using the definition of ordering fuzzy numbers given above.

Step 3:Obtain the$k^{t□}$ percentile corresponding to the value of $\left[\frac{(n+1)k}{100}\right]^{t□}$ position in the orderedarrangementof $\xi_1, \xi_2, \dots, \xi_n$.

The median value is obtained at $k = 50$ and the 1st and 3rd quartiles are obtained at $k = 25$ and $k = 75$, respectively.

III. NUMERICAL ILLUSTRATION

The computation of percentiles using the proposed method is illustrated with a numerical example based on an artificial data set consisting of 20 triangular fuzzy numbers as reported in Wen-Liang Hung and Minn-Shen Yang [10]. The fuzzy numbers are given below (Table 1).

Table 1: *Data Set of 20 Triangular Fuzzy Numbers.*

Observation	Triangular fuzzy number	Observation	Triangular fuzzy number
1	(7.29,7.56,8.56)	11	(20.14,20.77,21.24)
2	(6.61,8.56,10.49)	12	(20.8,21.88,22.54)
3	(9.33,9.89,11.09)	13	(20.97,22.45,23.71)
4	(10,10.89,11.77)	14	(22.09,23.88,24.04)
5	(11.66,11.78,12.99)	15	(24.22,24.88,25.52)
6	(11.71,12.9,13.31)	16	(24.73,25.25,26.96)
7	(11.85,13.67,14.57)	17	(23.52,25.47,25.62)
8	(12.97,14.87,16.72)	18	(25.64,26.56,27.19)
9	(13.66,15.45,17.4)	19	(26.24,27.98,29.67)
10	(14.31,15.78,16.2)	20	(27.06,28.77,29.56)

The expected value and variance of a triangular fuzzy variable $\xi = (a, b, c)$ can be obtained using the definitions given in section "Preliminaries" (see Zhigang-Wang and Fanji-Tian)[11].

$$E(\xi) = \frac{a+2b+c}{4} \text{and} V(\xi) = \begin{cases} \frac{33\alpha^3+11\alpha\beta^2+21\alpha^2\beta-\beta^3}{384\alpha}, & \alpha > \beta \\ \frac{\alpha^2}{6}, & \alpha = \beta \\ \frac{33\beta^3+11\alpha^2\beta+21\alpha\beta^2-\alpha^3}{384\alpha}, & \alpha < \beta \end{cases} \quad \text{where } \alpha = b - a \text{ and}$$

$\beta = c - b$.Using the above expressions, the expected value and variance of the triangular fuzzy numbers in Table 1are calculated and is displayedin Table 2 below.

Based on the expected values and variances,the fuzzy numbers given in Table 1 are ordered in increasing order using the method proposed byLiet al. [9] and are given in
Table 3.

Table 2: *Expected Value and Variance of Triangular Fuzzy Numbers.*

Observation	Expected value	Variance	Observation	Expected value	Variance
1	7.7425	0.0277	11	20.7300	0.0223
2	8.5550	2.3944	12	21.7750	0.1761
3	10.0500	0.1136	13	22.3950	0.7276
4	10.8875	0.1038	14	23.4725	0.9348
5	12.0525	0.0194	15	24.8750	0.0310
6	12.7050	0.2167	16	25.5475	0.2734
7	13.4400	1.3130	17	25.0200	1.3058
8	14.8575	2.1365	18	26.4875	0.0974
9	15.4900	2.1005	19	27.9675	1.5004
10	15.5175	0.4849	20	28.5400	1.0009

Table 3: *Ordered Triangular Fuzzy Numbers.*

Ordered Observation	Triangular fuzzy number	Ordered Observation	Triangular fuzzy number
1	(7.29,7.56,8.56)	11	(20.14,20.77,21.24)
2	(6.61,8.56,10.49)	12	(20.8,21.88,22.54)
3	(9.33,9.89,11.09)	13	(20.97,22.45,23.71)
4	(10,10.89,11.77)	14	(22.09,23.88,24.04)
5	(11.66,11.78,12.99)	15	(24.22,24.88,25.52)
6	(11.71,12.9,13.31)	16	(23.52,25.47,25.62)
7	(11.85,13.67,14.57)	17	(24.73,25.25,26.96)
8	(12.97,14.87,16.72)	18	(25.64,26.56,27.19)
9	(13.66,15.45,17.4)	19	(26.24,27.98,29.67)
10	(14.31,15.78,16.2)	20	(27.06,28.77,29.56)

It is to be noted that the positions of observations in Table 1 and Table 3 are same except observations 16 and 17 that are interchanged. The various percentiles of fuzzy numbers in Table 1 can now be obtained using the ordered arrangement in Table 3. For example, the 50th percentile (median) corresponds to the value of 10.5th position in the ordered arrangement i.e. median = value corresponding to 10th position + 0.5 times the difference of values corresponding to 10th and 11th positions in the ordered arrangement. In order to perform operations on triangular fuzzy numbers, the results (1) and (2) mentioned below (Liu)[8] are used.

Results

(1). The sum of two triangular fuzzy variables $\xi = (a_1, a_2, a_3)$ and $\eta = (b_1, b_2, b_3)$ is also a triangular fuzzy variable and is given by $\xi + \eta = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$.

(2). The product of a triangular fuzzy variable $\xi = (a_1, a_2, a_3)$ and a scalar number λ is again a triangular fuzzy variable and is given by $\lambda \cdot \xi = \begin{cases} (\lambda a_1, \lambda a_2, \lambda a_3), & \text{if } \lambda \geq 0 \\ (\lambda a_3, \lambda a_2, \lambda a_1), & \text{if } \lambda < 0 \end{cases}$.

Using the above results, the median value is obtained as median = (16.28, 18.275, 19.665). In a similar manner, the values of 1st and 3rd quartiles, namely Q_1 and Q_3corresponding to $k = 25$ and $k = 75$ are calculated and is given by Q_1 = (11.34, 12.06, 13.4025); Q_3= (22.945, 25.3225, 26.57).

IV. CONCLUSION

In this paper, a method of determining percentiles of a set of fuzzy numbers is developed by making use of the concepts available in credibility theory. The proposed method uses the expected value and variance of fuzzy numbers to get an ordered arrangement. The method is much easier than the usual α-cut approach to perform operations on fuzzy numbers. Further work is being carried out to develop non-parametric test procedures for a set of fuzzy data based on the measures of location.

REFERENCES

[1]. J.D. Gibbons, S.Chakraborti, *Non Parametric Statistical Inference*, 5^{th} Ed. Chapman and Hall: 2010.
[2]. L.A.Zadeh, Fuzzy Sets. *Inform. Control.* 1965; 8(3): 338–353p.
[3]. R. Kruse, K.D.Meyer, *Statistics with Vague Data*, Riedel, Dordrecht, 1987.
[4]. R.Viertl, *Statistical Methods for Fuzzy Data*, John Wiley and Sons Ltd., UK, 2011.
[5]. C.Carlsson, R.Fuller, On Possibilistic Mean Value and Variance of Fuzzy Numbers. *Fuzzy Set Syst.* 2001; 122: 315–326p.
[6]. S.Bodjanova, Median Value and Median Interval of a Fuzzy Number. *Inform. Sciences.* 2005; 172: 73–89p.
[7]. E. Pasha, A. Saiedifar, B.Asady, The Percentiles of Fuzzy Numbers and their Applications. *Iran. J. Fuzzy Syst.* 2009; 6(1): 27–44p.
[8]. B.Liu, *Uncertainty Theory*, Springer: Verlag, Berlin, 2007.
[9]. X.Li, W.Tang, R.Zhao, Ranking Fuzzy Variables by Expected Value and Variance. *6^{th} International Conference on Fuzzy systems and Knowledge Discovery*. 2009; 373–377p.
[10]. Wen-Liang Hung, Miin-Shen Yang, Fuzzy Clustering on LR-Type Fuzzy Numbers with an Application in Taiwanese Tea Evaluation. *Fuzzy Set. Syst.* 2005; 150: 561–577p.
[11]. Zhigang Wang, FanjiTian, A Note on the Expected Value and Variance of Fuzzy Variables. *Int. J. Nonlinear Sci.* 2009; 4: 486–492p.