International Journal of
Innovative Management, Information & Production ISME International©2011 ISSN 2185-5439
Volume **2**, Number **2**, December **2011** PP. 1-7

A DATABASE FOR A NEW FUZZY PROBABILITY DISTRIBUTION FUNCTION AND ITS APPLICATION

PEICHUN LIN[1], JUNZO WATADA[2] AND BERLIN WU[3]
[1,2]Graduate School of Information
Production and System, Waseda University
2-7 Hibikino, Wakamatsu, Kitakyushu, 808-0135 Japan.
[1]peichunpclin@gmail.com
[2]junzow@osb.att.ne.jp

[3]Department of Mathematical Sciences
National Chengchi University, Taipei, Taiwan
berlin@nccu.edu.tw

ABSTRACT. *This paper provides a definition of a new fuzzy distribution function, which is further classified by some of its properties. We build a database to provide the information about the fuzzy distribution functions. Some examples illustrate the method for testing whether two fuzzy samples have been drawn from the same distribution function. The results show that it is useful to classify the fuzzy numbers into different groups based on properties in the database.*

Keywords: Fuzzy Data**;** Fuzzy Probability Distributions; Fuzzy Statistics and Data Analysis; Test Hypothesis

1. Introduction. Conventional statistical analysis is pursued using numerical values, so called crisp values. The values are not supposed to be any fuzzy numbers. The traditional statistical analysis is considered with associated random variables, point estimation techniques, hypotheses and parameters of interest and so on Akbari et al. (2009). Two-sample test statistic is one of the most useful nonparametric methods to compare two samples, as it is sensitive to differences in both the location and the shape of the empirical cumulative distribution functions of the two samples (Conover, 1971). But when data are given vaguely, that is, the given data are fuzzy numbers, knowing the distribution function of fuzzy data plays a pivotal rule in dealing with problems in the real world. The distribution function must be estimated under the specified condition or in the situation given in advance. In real life, vague information is widely used when describing data in natural language (Nguyen and Wu, 2006). On the other hand, in dealing with fuzzy data, the underlying probability distribution of the fuzzy data is unknown. The situation makes it difficult to put such information into statistical terms. Therefore, the objective of the paper is to establish a technique to handle such information and knowledge.

In order to find the probability distribution function with fuzzy data, we have to take the concept of fuzzy statistic into consideration. Following Zadeh (1965, 1973), fuzzy data are treated here. Fundamental statistics, such as mean, median and mode are useful measurements in illustrating some characteristics of the sample distribution. Recently, the fuzzy statistical analysis and applications are studied and focused on the social science such as Casalino et al. (2004), Dubois and Prade (1980), Wu (1995), and Wu and Sun (2001), which demonstrate the central role of fuzzy statistic and apply it to social survey. All research studies used to deal with problems using the central points of fuzzy numbers. Pei-Chun Lin, et al. defined a new weight function in terms of central point and radius. The function will show more effective to observe original fuzzy data.

Specifically, the objective of this paper is to define fuzzy distribution function, and clarify its property. The rest of the paper consists of the following. Section 2 gives the brief review of related studies. In Section 3, the main procedure is described to construct the database with fuzzy distribution function. In Section 4, we describe some empirical studies. Finally, the concluding remarks and the topics of further studies are addressed in Section 5.

2. Problem Statement and Preliminaries. Let us introduce some definitions using LR-type fuzzy numbers (Esogbue and Song, 2003), which are used in the next section.

Definition 1. A fuzzy number A=[a, b, c, d] is called trapezoidal with tolerance interval [*b*, *c*], left width L=b-a and right width U=d-c if its membership function has the following form:

$$\mu_A(x) = \begin{cases} 1 - \dfrac{b-t}{L} & if \ \ b - L \le t \\ 1 & if \ \ b \le t \le c, \\ 1 - \dfrac{t-c}{U} & if \ \ c \le t \le c \cdot \\ 0 & otherwise \end{cases}$$

and it is also denoted by $A \equiv (b, c, L, U)$.

Definition 2. Especially, if $b = c$, the fuzzy number is called a triangular fuzzy number $B = [a, b, d]$, the left width $L = b - a$ and right width $U = d - b$. Its membership function is written in the following form:

$$\mu_B(x) = \begin{cases} 1 - \dfrac{b-t}{L} & if \ \ b - L \le t : \\ 1 & if \ \ t = b, \\ 1 - \dfrac{t-d}{U} & if \ \ d \le t \le d \cdot \\ 0 & otherwise \end{cases}$$

and it is also denoted by $B \equiv (b, L, U)$.

Definition 3. An interval fuzzy number is denoted as $C = [a, d]$ with central point $o = \frac{d+a}{2}$ and radius $l = \frac{d-a}{2}$. We also use the notation $C \equiv (o, l)$.

Moreover, two operations on LR-type fuzzy numbers, namely fuzzy addition and scalar multiplication, are given as follows.

Lemma 1 Let $A_1 = (a, b, c, d)$ and $A_2 = (\bar{a}, \bar{b}, \bar{c}, \bar{d})$ be any two fuzzy numbers and λ be any real number. Then

(a) $A_1 + A_2 = (a + \bar{a}, b + \bar{b}, c + \bar{c}, d + \bar{d})$.

(b) $\lambda A_1 = (\lambda a, \lambda b, \lambda c, \lambda d)$, $\forall\ \lambda \geq 0$.

From the above definitions and Lemma 1, we also need to know how to calculate the fuzzy expected value and fuzzy variance. We give a brief statement in the following subsection.

Definition 4. Fuzzy Expected Value

Let A, B and C be continuous fuzzy random variables on the probability space $(\Omega, \mathcal{F}, P)$. Then the fuzzy expected value of A, B and C are defined as follows.

If $A_i \equiv (b_i, c_i, L_i, U_i)$ is trapezoidal fuzzy numbers, $\forall\ i = 1,2, \dots, n$, then

$$E(A_i) = (E(b_i), E(c_i), E(L_i), E(U_i)).$$

Especially, if $B_i \equiv (b_i, L_i, U_i)$ is triangular fuzzy numbers, $\forall\ i = 1,2, \dots, n$, then

$$E(B_i) = (E(b_i), E(L_i), E(U_i)).$$

Moreover, if $C_i = (o_i, l_i)$ is interval fuzzy numbers, $\forall\ i = 1,2, \dots, n$, then

$$E(C_i) = (E(o_i), E(l_i)).$$

Definition 5. Fuzzy Variance

Let A, B and C be continuous fuzzy random variables on the probability space $(\Omega, \mathcal{F}, P)$. Then the fuzzy expected value of A, B and C are defined as follows.

If $A_i \equiv (b_i, c_i, L_i, U_i)$ is trapezoidal fuzzy numbers, $\forall\ i = 1,2, \dots, n$, then

$$var(A_i) = (var(b_i), var(c_i), var(L_i), var(U_i)).$$

Especially, if $B_i \equiv (b_i, L_i, U_i)$ is triangular fuzzy numbers, $\forall\ i = 1,2, \dots, n$, then

$$var(B_i) = (var(b_i), var(L_i), var(U_i)).$$

Moreover, if $C_i = (o_i, l_i)$ is interval fuzzy numbers, $\forall\ i = 1,2, \dots, n$, then

$$var(C_i) = (var(o_i), var(l_i)).$$

When we know the probability distribution function with each elements of A, B and C, we can easily find out the fuzzy expected value and fuzzy variance.

In the next section, we want to give a database of fuzzy distribution function.

2. Database of Fuzzy Probability Distribution Functions.

2.1. Fuzzy Probability Distribution Function (FPDF). In order to demonstrate the database of distribution function with fuzzy data, we need to define a new function which is called Fuzzy Probability Distribution Function (FPDF).

Definition 6. f is a Fuzzy Probability Distribution Function (FPDF) defined on $\mathbb{R}^3$ with location and scale parameter, (b, L, U), where b is the second element of triangular fuzzy numbers $[a, b, c]$, $L = b - a$ and $U = c - b$. Moreover, this function is denoted by $f_F(b, L, U)$.

Notation If b_i, L_i *and* $U_i$re independent, then fuzzy probability distribution function can be written in the following:

$$f_{F_i}(b,L,U) \equiv f_i(b) * f_i(L) * f_i(U), \quad \forall\, i = 1,2,\dots n$$

Here, we give the procedure of building fuzzy probability distribution functions with triangular fuzzy data.

Procedure of building fuzzy probability distribution functions with triangular fuzzy data:

Step 1: Get the triangular fuzzy numbers.

Step 2: Computing b_i, L_i and U_i, $\forall\, i = 1,2,\dots,n$.

Step 3: Identify the underlying distribution by simulating b_i, L_i and U_i,

$$\forall\, i = 1,2,\dots,n.$$

Step 4: Get the distribution functions of $f_i(b)$, $f_i(L)$ and $f_i(U)$, $\forall\, i = 1,2,\dots n$.

Step 5: Compute the fuzzy distribution function $f_{F_i}(b,L,U)$, $\forall\, i = 1,2,\dots n$.

When we know this procedure, we can construct the database of fuzzy probability distribution functions. We give this concept in the following subsection.

2.2. Database of Fuzzy Probability Distribution Functions. When we know how to find out the fuzzy distribution functions, we can build the database of each fuzzy distribution function as shown in TABLE 1.

TABLE 1. Database of fuzzy probability distribution functions

$f_{F_i}(b_i,L_i,U_i)$	b_i	L_i	U_i	Property
$f_{F_1}(b_1,L_1,U_1)$	D_{b_1}	D_{L_1}	D_{U_1}	I
$f_{F_2}(b_2,L_2,U_2)$	D_{b_2}	D_{L_2}	D_{U_2}	II
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$f_{F_n}(b_n,L_n,U_n)$	D_{b_n}	D_{L_n}	D_{U_n}	N

Property 1. If we have $b_1 \sim D_{b_1} = \Gamma(k_1,\theta_1)$, $L_1 \sim D_{L_1} = \Gamma(k_2,\theta_2)$ and $U_1 \sim D_{U_1} = \Gamma(k_3,\theta_3)$, where $\Gamma(k_i,\theta_i), i = 1,2,3$, is gamma distribution function, then we can find the expected value and variance easily. We use moment method estimator (MME) to estimate our parameter in each distribution function. Moreover, b, L and U have the same property as gamma distribution function. We will show the detail statement in the next section.

Property 2. If we have $b_1 \sim D_{b_1} = W(\lambda,\kappa)$,$L_1 \sim D_{L_1} = \Gamma(k_4,\theta_4)$ and $U_1 \sim D_{U_1} = \Gamma(k_5,\theta_5)$, where $W(\lambda,\kappa)$ is Weibull distribution function and $\Gamma(k_i,\theta_i), i = 4,5$, is gamma distribution function, then we can find the expected value and variance easily. We use moment method estimator (MME) to estimate our parameter in each distribution function. Moreover, b has the same property as Weibull distribution function, L and U have the same property as gamma distribution function. We will also show the detail statement in the next section.

We get the surface plot by one kind of Property of b, L and U in FIGURE 1.

Definition 7. Let $X_1 \equiv (b_1,L_1,U_1)$ and $X_2 \equiv (b_2,L_2,U_2)$ are two triangular fuzzy numbers. We say that X_1 and X_2 have the same property. It means that b_1 and b_2 have the same distribution, L_1 and L_2 have the same distribution and U_1 and U_2 have the same distribution. Otherwise, we say that they don't have the same properties.

In the next section, we give some empirical studies to demonstrate how to find out the fuzzy distribution function and its application.

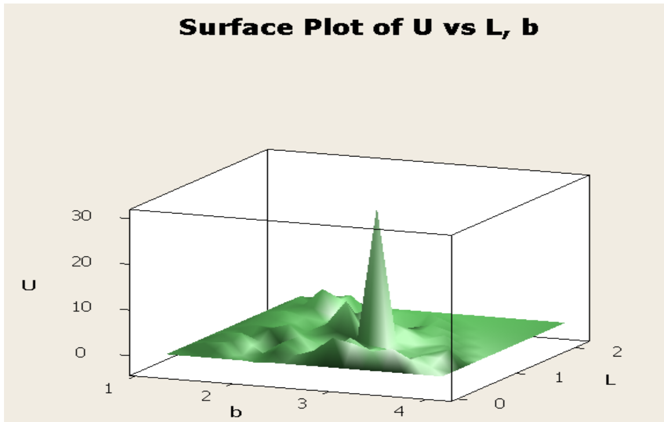

FIGURE 1. Surface plot of U vs L and b

4. Empirical Studies. Example 4.1, A Japanese dining hall manager planned to introduce a new boxed lunch service near universities in Taipei city and decided to take a survey to investigate which place will have good returns to sell a boxed lunch for students. He gives the questionnaires to two schools' (Shin Chien University and Chungchi University) students. Samples were randomly selected in each school. The investigator asked them the following questions: 1. more than which times (real numbers) can you accept to buy a boxed lunch a week? 2. How many times (real values) should you buy a boxed lunch a week? 3. More than which times (real values) can't you accept to buy a boxed lunch? We collected those answers of the questionnaires and get triangular fuzzy numbers. The answers are shown in Table 2.

Moreover, we found the fuzzy distribution function by simulating each parameter. We gave the results as Table 3.

TABLE 2. Triangular fuzzy numbers of two schools

Shin Chien University	(2,4,9)	(3,3,5)	(1,2,3)	(3,3,7)	(2,3,4)	(2,3,3)
Chungchi University	(6,7,8)	(4,4,5)	(1,1,2)	(2,3,5)	(4,5,8)	(3,5,6)

TABLE 3. The probability distribution function for parameter of b_i, L_i and U_i

	b_i	L_i	U_i
Shin Chien University (X)	$\Gamma(11.07, 0.27)$	$\Gamma(3.65, 0.39)$	$\Gamma(1.80, 1.21)$
Chungchi University (Y)	$W(2.44, 3.52)$	$\Gamma(7.25, 0.17)$	$\Gamma(2.92, 0.57)$

We could see that b_1 and b_2 have different distribution function, L_1 and L_2 have the same distribution function and U_1 and U_2 have the same distribution function. By Definition 7, we concluded that the two schools have different property of the acceptable times of buying a boxed lunch a week.

Example 4.2, In the same problem as Example 4.1, we used the method in Pei-Chun Lin, *et al.* to test whether two samples have the same distribution function. First, we found the 200 weight values of triangular fuzzy numbers. Then we could get the cumulative distribution function of $S_{100}(X)$, $S_{100}(Y)$ and $|S_{100}(X)-S_{100}(Y)|$. We give the result in Table 4.

TABLE 4. The cumulative distribution function of $S_{100}(X)$, $S_{100}(Y)$ and $|S_{100}(X)-S_{100}(Y)|$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$S_{100}(X)$	0.00	0.00	0.01	0.01	0.02	0.04	0.05	0.09	0.10	0.10	0.10	0.11	0.12	0.15	0.15	0.16	0.23	0.23
$S_{100}(Y)$	0.01	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.06	0.08	0.08	0.08	0.08	0.09	0.09	0.09	0.10
$\|S_{100}(X)-S_{100}(Y)\|$	0.01	0.02	0.01	0.02	0.01	0.01	0.02	0.06	0.07	0.04	0.02	0.03	0.04	0.07	0.06	0.07	0.14	0.13

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
$S_{100}(X)$	0.25	0.25	0.25	0.25	0.29	0.31	0.31	0.31	0.33	0.33	0.35	0.41	0.43	0.43	0.43	0.43	0.43	0.45
$S_{100}(Y)$	0.10	0.14	0.19	0.21	0.21	0.21	0.24	0.25	0.25	0.30	0.30	0.30	0.30	0.34	0.37	0.42	0.47	0.47
$\|S_{100}(X)-S_{100}(Y)\|$	0.15	0.11	0.06	0.04	0.08	0.10	0.07	0.06	0.08	0.03	0.05	0.11	0.13	0.09	0.06	0.01	0.04	0.02

	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
$S_{100}(X)$	0.49	0.49	0.49	0.52	0.62	0.63	0.63	0.64	0.64	0.64	0.64	0.73	0.73	0.73	0.80	0.80	0.90	0.91
$S_{100}(Y)$	0.47	0.50	0.52	0.52	0.52	0.52	0.54	0.54	0.60	0.61	0.63	0.63	0.67	0.71	0.71	0.75	0.75	0.75
$\|S_{100}(X)-S_{100}(Y)\|$	0.02	0.01	0.03	0.00	0.10	0.11	0.09	0.10	0.04	0.03	0.01	0.10	0.06	0.02	0.09	0.05	0.15	0.16

	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
$S_{100}(X)$	0.91	0.91	0.91	0.94	0.95	0.95	0.98	0.98	0.98	0.99	0.99	0.99	1	1	1	1	1	1
$S_{100}(Y)$	0.76	0.78	0.86	0.86	0.86	0.87	0.87	0.88	0.90	0.90	0.91	0.93	0.93	0.94	0.95	0.96	0.99	1
$\|S_{100}(X)-S_{100}(Y)\|$	0.15	0.13	0.05	0.08	0.09	0.08	0.01	0.10	0.08	0.09	0.08	0.06	0.07	0.06	0.05	0.04	0.01	0

From Table 4, the test statistic was obtained:

$$D = max|S_{100}(X)-S_{100}(Y)| = 0.16.$$

at a significance level $\alpha = 0.05$. For large numbers, D must be at least 0.192 in order to reject H_0, for $1.36\sqrt{\frac{100+100}{100\cdot 100}} = 0.192$. Since the observed value did not exceed the critical value, we did not reject H_0. We concluded that the two schools have the same distribution of the acceptable times of buying a boxed lunch a week.

5. Conclusion. In this paper, we defined a new fuzzy distribution function. Moreover, we constructed a database of fuzzy distribution functions. We executed many simulations of real fuzzy data to build up a complete database. But we have also some studies to work on in the future as follows:

(1) How is the range between two parameters of each underlying distribution function defined so as that we can get a good simulation in the properties? Moreover, how can we say that two samples have the same property?

(2) Can we test whether the original fuzzy data distribute to some properties in the database or not?

REFERENCES

[1] A. O. Esogbue and Q. Song (2003), On optimal defuzzification and learning algorithms: theory and applications, *Fuzzy Optimization and Decision Making*, vol.2, no.4, pp.283.

[2] B. Wu (1995), Fuzzy statistical analysis: a new research method on the sampling survey, *Research Letters of National Chengchi University*, vol.2, pp.65-80.

[3] B. Wu and C. Sun (2001), Interval-valued statistics, fuzzy logic, and their use in computational semantics, *Journal of Intelligent and Fuzzy Systems*, vol.11, no.1-2, pp.1-7.

[4] D. Dubois and H. Prade (1980), Systems of linear fuzzy constrains, *Fuzzy Sets and Systems*, vol.3, no.1, pp.37-48.

[5] G. Casalino, F. Memola and C. Minutolo (2004), A model for evaluation of laser welding efficiency and quality using an artificial neural network and fuzzy logic, *Proceedings of the Institution of Mechanical Engineers*, vol.218, no.6, pp.641.

[6] H. Nguyen and B. Wu (2006), *Fundamentals of Statistics with Fuzzy Data*, Springer, Netherlands.

[7] L. A. Zadeh (1965), Fuzzy Sets, *Information and Control*, vol.8, no.3, pp.338-353.

[8] L. A. Zadeh (1973), Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Trans. Syst. Man Cybern*, SMC-3, no.1, pp.28-44.

[9] M. G. Akbari., A. H. Rezaei and Y. Waghei (2009), Statistical Inference about the Variance of Fuzzy Random Variables, *The Indian Journal of Statistics*, vol. 71-b, no. 2, pp.206-221.

[10] P. C. Lin, B. Wu and J. Watada (2010), Kolmogorov-Smirnov two sample test with continuous fuzzy data, *Integrated Uncertainty Management and Applications*, AISC 68, pp. 175-186.

[11] W. J. Conover (1971), *Practical Nonparametric Statistics*, New York.