Fuzzy Bootstrap Test for the Mean and Variance with Dp,q-Distance

Bahram Sadeghpour Gildeh and Sedigheh Rahimpour

Abstract—**Testing statistical hypothesis is a main topic in statistical inference. In this paper, we consider the problem of testing a simple hypothesis about the mean and variance of a fuzzy random variable with the help of Dp,q-distance. Concerning the hypothesis testing, the bootstrap techniques have empirically shown to be efficient and powerful. By means of simulation and some examples we show that the bootstrap method is a powerful tool in the statistical hypothesis testing about the parameters of fuzzy random variables.**

Index Terms—**Bootstrap, Dp,q-distance, fuzzy random variable, testing of hypothesis.**

I. Introduction

The concept of fuzzy sets was introduced by Zadeh [9] to describe non-statistical uncertainty (inexactness, vagueness). Fuzzy random variables defined by Puri and Ralescu [5] deal with both kinds of uncertainty: the randomness and the vagueness.

One of the primary purposes of statistical inference is to test hypothesis. Testing hypothesis with fuzzy data was considered by Tanaka, Okuda and Asaie [7]. Some methods of statistical inference with fuzzy data are reviewed by Viertl [8]. Taheri and Arefi [6] exhibit an approach to test fuzzy hypotheses based on fuzzy test statistic. By means of central limit theorem, Korner [11] proposed an asymptotic test for the one sample problem about the mean value of fuzzy random variable.

The bootstrap using fuzzy data is developed in different approaches. Montenegro, Colubi, Casals and Gil [4], Jimenez-Gamero, Pino-Mejias and Rojas-Medar [3] and Colubi [2] have considered the problem of hypothesis testing about the mean of a fuzzy random variable. Akbari and Rezaei [1] present a fuzzy bootstrap test for variance.

The paper is organized as follows. Section 2 contains some preliminary. In Section 3 we provide the problem of hypothesis testing about the mean value of fuzzy random variable and an example are considered in order to illustrate them. In Section 4, the hypothesis testing procedure for variance of fuzzy random variable is described. Finally, in Section 5 we conclude the paper.

II. Preliminaries

Definition 2.1: Let X be a universal set, then a fuzzy set $\tilde{A}$ of X is defined by its membership function $\tilde{A}: X \longrightarrow [0,1]$, where $\tilde{A}(x)$ is the membership grade of x to $\tilde{A}$.

Manuscript received October 9, 2012; revised December 07, 2012.

The authors are with Department of Statistics, University of Mazandaran, Babolsar, Iran (e-mail: sadeghpour@umz.ac.ir; rahimpour.s@gmail.com).

Definition 2.2: for each $0 \leq \alpha \leq 1$, the α-level set of $\tilde{A}$ is defined by

$$A_\alpha = \{x \in X \mid \tilde{A}(x) \geq \alpha\}. \quad (1)$$

Definition 2.3: A fuzzy number is a fuzzy set of $\Re$ such that the following conditions are satisfied:

1) $\tilde{A}$ is normal, i.e. $\{x \in \Re \mid \tilde{A}(x) = 1\}$ is non empty,
2) $\tilde{A}$ is convex, i.e. $\forall x_1, x_2 \in \Re$ and $\lambda \in [0,1]$:
$$\tilde{A}(\lambda x_1 + (1-\lambda)x_2) \geq min\,(\tilde{A}(x_1), \tilde{A}(x_2)),$$
3) $\tilde{A}$ is upper semicontinuous with compact support, i.e. $\forall \epsilon > 0,\ \exists \delta > 0;\ |x-y| < \delta \Rightarrow \tilde{A}(x) < \tilde{A}(y) + \epsilon$

According to the above definition, α-level set of a fuzzy number is a closed interval, denoted by $A_\alpha = [A_\alpha^-, A_\alpha^+]$, where

$$A_\alpha^- = inf\{x \in \Re\,;\ \tilde{A}(x) \geq \alpha\},$$
$$A_\alpha^+ = sup\{x \in \Re\,;\ \tilde{A}(x) \geq \alpha\}. \quad (2)$$

$F(\Re)$ denote the set of all fuzzy numbers. By Zadeh's extension principle for each $\tilde{A}, \tilde{B} \in F(\Re)$ and $\lambda \in \Re$ we have

$$(\tilde{A} \oplus \tilde{B})_\alpha = A_\alpha + B_\alpha = \{a + b \mid a \in A_\alpha, b \in B_\alpha\},$$
$$(\lambda \odot \tilde{A})_\alpha = \lambda\,.\,A_\alpha = \{\lambda\, a + b \mid a \in A_\alpha\}. \quad (3)$$

Definition 2.4: [10] The $D_{p,q}$-distance between two fuzzy numbers $\tilde{A}, \tilde{B} \in F(\Re)$ indexed by parameters $1 \leq p \leq \infty$ and $0 \leq q \leq 1$, is a nonnegative function on $F(\Re) \times F(\Re)$ give as follows:

$$D_{p,q}(\tilde{A}, \tilde{B}) = \begin{cases} [(1-q)\int_0^1 |A_\alpha^- - B_\alpha^-|^p\, d\alpha + q\int_0^1 |A_\alpha^+ - B_\alpha^+|^p d\alpha]^{1/p} & ;\ p < \infty \\ (1-q)\ sup_{0<\alpha\leq 1}(|A_\alpha^- - B_\alpha^-|) + q\ inf_{0<\alpha\leq 1}(|A_\alpha^+ - B_\alpha^+|) & ;\ p = \infty. \end{cases}$$

In this paper, we suppose that $p = 2$ and $q = \frac{1}{2}$.

Definition 2.5: The membership function of an LR-fuzzy number $\tilde{A} = (m, l, r)_{LR}$ is

$$\tilde{A}(x) = \begin{cases} L\left(\frac{m-x}{l}\right) & ;\ x < m \\ 1 & ;\ x = m \\ R\left(\frac{x-m}{r}\right) & ;\ x > m, \end{cases} \quad (4)$$

where L, R: $\Re^+ \longrightarrow [0,1]$ are fixed left-continuous and non-increasing functions with $L(0)=R(0)=1$. The functions L and R are called left and right shape functions, m is the modal point and l, r≥ 0 are the left and right spreads, respectively, of the LR-fuzzy number.

The $D_{2,\frac{1}{2}}$ -distance of two LR-fuzzy numbers $\tilde{A} = (m_A, l_A, r_A)_{LR}$ and $\tilde{B} = (m_B, l_B, r_B)_{LR}$ is

$$D^2_{2,\frac{1}{2}}(\tilde{A}, \tilde{B}) = (m_A - m_B)^2 + L_2(l_A - l_B)^2 + R_2(r_A - r_B)^2$$
$$+2(m_A - m_B)\big(R_1(r_A - r_B) - L_1(l_A - l_B)\big),$$

DOI: 10.7763/IJMLC.2013.V3.265

where

$$L^{(-1)}(\alpha) = sup\{x \in \Re \ ; \ L(x) \geq \alpha\} \ , \ \ R^{(-1)}(\alpha) = sup\{x \in \Re \ ; \ R(x) \geq \alpha\},$$

$$L_1 = \frac{1}{2}\int_0^1 L^{(-1)}(\alpha) \ d\alpha \quad , \quad L_2 = \frac{1}{2}\int_0^1 (L^{(-1)}(\alpha))^2 \ d\alpha,$$

and R_1, R_2 are similarly defined.

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition 2.6: [5] A mapping $\tilde{X} : \Omega \rightarrow F(\Re)$ is said to be a fuzzy random variable if and only if

$$\{(\omega, x) : x \in X_\alpha(\omega)\} \in \mathcal{A} \times \mathcal{B},$$

where $\mathcal{B}$ denote the σ-field of Borel set in $\Re$.

Definition 2.7: [2] If $\tilde{X}$ be a fuzzy random variable, the expected value of $\tilde{X}$ is the unique fuzzy subset of $\Re$ (if it exist), denoted by $\tilde{E}(\tilde{X})$, such that for all $\alpha \in [0,1]$ we have:

$$(E(\tilde{X}))_\alpha = E(X_\alpha) = [E(X_\alpha^-), E(X_\alpha^+)]. \quad (5)$$

Definition 2.8: If $E(sup_{x \in \chi_0} |x|^2) < \infty$, the variance of $\tilde{X}$ on the basis of $D_{p,q}$-distance is defined by

$$\sigma^2 = DVar(\tilde{X}) = E(\left[D_{2,q}\left(\tilde{X}, \tilde{E}(\tilde{X})\right)\right]^2), \quad (6)$$

where χ_0 is the closure of the α-level set.

III. Bootstrap Hypothesis Test for the Mean of Fuzzy Random Variable

In this Section we introduce a way to get bootstrap test for the mean based on fuzzy data.

Let $\tilde{X}$ be a fuzzy random variable such that $E(sup_{x \in \chi_0} |x|) < \infty$ and let $\tilde{X}_1, \dots, \tilde{X}_n$ be a random sample obtained from $\tilde{X}$, the sample fuzzy mean value is given by

$$\tilde{\bar{X}}_n = \frac{\tilde{X}_1 \oplus \dots \oplus \tilde{X}_n}{n}. \quad (7)$$

Theorem 3.1 The sample fuzzy mean value is an unbiased estimator of the fuzzy parameter $\tilde{E}(\tilde{X})$.

Proof: $\forall \ \alpha \in [0,1]$, we have

$$\left(E(\tilde{\bar{X}}_n)\right)_\alpha = [E(\bar{X}_{n\alpha}^-), E(\bar{X}_{n\alpha}^+)] = [\mu_\alpha^-, \mu_\alpha^+] = \mu_\alpha$$

Let $\tilde{X}_1, \dots, \tilde{X}_n$ be i.i.d fuzzy random variable. Given $\tilde{X}_1, \dots, \tilde{X}_n$, let $\underline{\tilde{X}}^* = (\tilde{X}_1^*, \dots, \tilde{X}_n^*)$ be a bootstrap sample, that is, $\tilde{X}_1^*, \dots, \tilde{X}_n^*$ are i.i.d fuzzy random variable such that

$$P_*(\tilde{X}_i^* = \tilde{X}_j) = \frac{1}{n} \quad , \ j = 1,2,..,n, i = 1,2,..,n.$$

where P_* denotes the bootstrap probability law, that is, the conditional probability given the original sample $\tilde{X}_1, \dots, \tilde{X}_n$.

We want to test the null hypothesis $H_0 : \tilde{E}(\tilde{X}_1) = \tilde{\mu}_0$ against the alternative $H_1 : \ \tilde{E}(\tilde{X}_1) \neq \tilde{\mu}_0$. We can expressed these hypotheses in terms of $D_{2,\frac{1}{2}}$:

$$\begin{cases} H_0 : \ D^2_{2,\frac{1}{2}}(\tilde{E}(\tilde{X}_1), \tilde{\mu}_0) = 0 \\ H_1 : \ D^2_{2,\frac{1}{2}}(\tilde{E}(\tilde{X}_1), \tilde{\mu}_0) > 0. \end{cases}$$

Let $\underline{\tilde{X}}^* = (\tilde{X}_1^*, \dots, \tilde{X}_n^*)$ be a bootstrap sample obtained from $\tilde{X}_1, \dots, \tilde{X}_n$. Then for testing H_0, we consider the following test: reject H_0 whenever

$$T_n = \frac{D^2_{2,\frac{1}{2}}(\tilde{\bar{X}}_n, \tilde{\mu}_0)}{S_n^2} > t_{1-\alpha}, \quad (8)$$

where $t_{1-\alpha}$ is the $(1-\alpha)$-quantile of the distribution of the bootstrap statistic $T_n^* = \frac{D^2_{2,\frac{1}{2}}(\tilde{\bar{X}}_n^*, \tilde{\bar{X}}_n)}{S_n^{*2}}$ and

$$\tilde{\bar{X}}_n^* = \frac{\tilde{X}_1^* \oplus \dots \oplus \tilde{X}_n^*}{n}, \qquad S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n} D^2_{2,\frac{1}{2}}(\tilde{X}_i^*, \tilde{\bar{X}}_n^*)$$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} D^2_{2,\frac{1}{2}}(\tilde{X}_i \ , \tilde{\bar{X}}_n \ ).$$

Or equivalently, the test reject H_0 if

$$p_{boot} = P_*(T_n^* \geq t_{obs}) \leq \alpha, \quad (9)$$

where t_{obs} is the observed value of the test statistic T_n.

The testing procedure can be applied in practice as the follows:

1) Compute the value of the statistic T_n.
2) Fix the bootstrap population to be $\tilde{X}_1, \dots, \tilde{X}_n$.
3) Obtain a sample of i.i.d. fuzzy random variables $(\tilde{X}_1^*, \dots, \tilde{X}_n^*)$ from the bootstrap population.
4) Compute the value of the bootstrap statistic T_n^*.
5) Steps 3 and 4 should be repeated B times to get a set of B estimators $(\{T_n^{*1}, .., T_n^{*B}\})$.
6) Compute the bootstrap $t_{1-\alpha}$ and p-value.

Example 3.1 Suppose that we have taken a fuzzy random sample of size *n*=10 from a population and that we have observed the LR- fuzzy data of Table I.

TABLE I: Fuzzy Random Sample of Size *N*=10

n	Observation	n	Observation
1	(2.9683,0.0727,0.1665)LR	6	(3.3631,0.7665,0.9047)LR
2	(2.4351,0.6316,0.4865)LR	7	(3.2805,0.4777,0.5045)LR
3	(2.4589,0.8847,0.8977)LR	8	(3.5347,0.2378,0.5163)LR
4	(3.6093,0.2727,0.9092)LR	9	(3.2832,0.2749,0.3190)LR
5	(2.9602,0.4364,0.0606)LR	10	(2.0515,0.3593,0.9866)LR

Suppose we are interested in a bootstrap test for the following hypotheses:

$$H_0 : \ \tilde{E}(\tilde{X}_1) = (3,0.5,0.5)_{LR} \ ,$$
$$H_1 : \ \tilde{E}(\tilde{X}_1) \neq (3,0.5,0.5)_{LR}.$$

If *B*=5000, the percentiles and bootstrap distribution of T_n^{*b} are shown in Table II and Fig. 1.

TABLE II: Percentiles of The Bootstrap Distribution of T_n^{*b}

α	0.01	0.05	0.95	0.99
$t_{1-\alpha}$	0.002	0.005	0.557	1.518

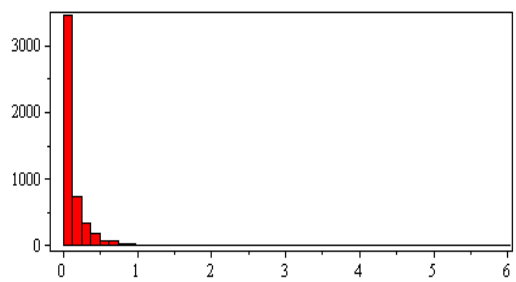

Fig. 1. Bootstrap distribution of T_n^{*b}.

Now, for $\alpha = 0.05$ or $\alpha = 0.01$ H_0 will not be rejected, because T_n=0.004 is less than the critical value $t_{1-\alpha}$, or the bootstrap *p-value* =0.967 is greater than α.

IV. A Fuzzy Bootstrap Test for the Variance of Fuzzy Random Variable

The aim of this section is testing the null hypothesis

$$H_0 : \sigma^2 \leq \sigma_0^2 \quad \text{versus} \quad H_1 : \sigma^2 > \sigma_0^2.$$

Theorem 4.1: The sample fuzzy variance value (S_n^2) is an unbiased estimator of the parameter $\sigma^2 = DVar(\tilde{X})$.

Proof: We have

$$E(S_n^2) = E\{\frac{1}{2(n-1)}\sum_{i=1}^{n}(\int_0^1 (X_{i\alpha}^- - \overline{X}_{n\alpha}^-)^2\, d\alpha + \int_0^1 (X_{i\alpha}^+ - \overline{X}_{n\alpha}^+)^2\, d\alpha)\}$$

$$= \frac{1}{2}\int_0^1 \frac{1}{n-1} E\{\sum_{i=1}^{n} (X_{i\alpha}^- - \overline{X}_{n\alpha}^-)^2\}\, d\alpha + \frac{1}{2}\int_0^1 \frac{1}{n-1} E\{\sum_{i=1}^{n} (X_{i\alpha}^+ - \overline{X}_{n\alpha}^+)^2\}\, d\alpha$$

$$= \frac{1}{2}\int_0^1 E(X_\alpha^- - \mu_\alpha^-)^2\, d\alpha + \frac{1}{2}\int_0^1 E(X_\alpha^+ - \mu_\alpha^+)^2\, d\alpha = \sigma^2.$$

Suppose that we have a fuzzy random sample $\tilde{X}_1, \ldots, \tilde{X}_n$ from $\tilde{X}$. Let $\underline{\tilde{X}}^* = (\tilde{X}_1^*, \ldots, \tilde{X}_n^*)$ be a bootstrap sample obtained from $\tilde{X}_1, \ldots, \tilde{X}_n$. Then, we have that:

In testing the null hypothesis $H_0 : \sigma^2 \leq \sigma_0^2$ at the nominal significance level α, H_0 should be rejected whenever

$$G_n = \frac{(n-1)S_n^2}{\sigma_0^2} > t_{1-\alpha}, \qquad (10)$$

where $t_{1-\alpha}$ is the $(1-\alpha)$-quantile of the null bootstrap distribution of $G_n^* = \frac{(n-1)S_n^{*2}}{S_n^2}$ and S_n^2 , S_n^{*2} are as before.

Or equivalently, the test reject H_0 if

$$p_{boot} \leq \alpha \qquad (11)$$

Example 4.1 Suppose that we have taken a fuzzy random sample of size n=12 from a population and that we have observed the LR- fuzzy data of Table III.

Suppose we are interested in a bootstrap test for the following hypotheses:

$$H_0 : \sigma^2 \leq 729 \,, H_1 : \sigma^2 > 729.$$

If B=5000, the percentiles and bootstrap distribution of G_n^{*b} are shown in Table IV and Fig. 2.

TABLE III: Fuzzy Random Sample of Size *N*=12

n	observation	n	observation
1	(35,33,36)LR	7	(73,70,76)LR
2	(82,80,84)LR	8	(70,65,73)LR
3	(87,85,87)LR	9	(103,100,105)LR
4	(90,90,90)LR	10	(56,54,58)LR
5	(63,60,66)LR	11	(40,40,42)LR
6	(70,70,72)LR	12	(96,94,99)LR

TABLE IV: Percentiles of the Bootstrap Distribution of G_N^{*B}

α	0.01	0.05	0.95	0.99
χ_7^2	1.239	2.167	14.067	18.475
χ_{11}^2	3.053	4. 575	19.675	24.725
$t_{1-\alpha}$	2.84	4.48	15.77	18.25

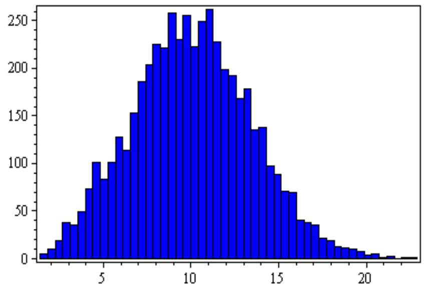

Fig. 2. Bootstrap distribution of G_n^{*b}.

Now, for $\alpha = 0.05$ or $\alpha = 0.01$ H0 will not be rejected, because G_n=9.052 is less than the critical value $t_{1-\alpha}$, Or the bootstrap p-value =0.608 is greater than α.

V. Conclusions

In this paper, we have described a bootstrap method for the mean and variance of fuzzy random variables with $D_{p,q}$distance.

In the classical context, the bootstrap has become a very powerful tool for estimating the sampling distribution of a statistic and its characteristics. In this paper, we have shown that the development of an adequate bootstrap theory in the fuzzy context would be very profitable because in this context the asymptotic approximations are, in most cases, difficult to handle and hence they are useless to make inferences.

References

[1] M. G. Akbari and A. Rezaei, “Bootstrap statistical inference for the variance based on fuzzy data,” *Austrian Journal of Statistics*, vol. 38, pp. 121-130, 2009.

[2] A. Colubi, “Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy-and real-valued data,” *Fuzzy Sets and Systems*, 2007.

[3] M. D. Jimenez-Gamero, R. Pino-Mejias, and M. A. Rojas-Medar, “A bootstrap test for the expectation of fuzzy random variable,” *Comput. Statist. Data Anal*, 2004.

[4] M. G. Montenegro, A. Colubi, M. R. Casals, and M. A. Gil, “Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable,” Metrika, vol. 59, pp. 31-49, 2004.

[5] M. L. Puri and D. A. Ralescu, “Fuzzy random variables,” *J. math. Anal. Appl*. vol. 114, pp. 409-422, 1986.

[6] S. M. Taheri and M. Arefi, “Testing fuzzy hypotheses based on fuzzy statistics,” *Soft Computing,* vol. 13, pp. 617-625, 2008.

[7] H. Tanaka, T. Okuda, and K. Asai, “Fuzzy information and decision in statistical model,” in: Gupta, M. M., *et al.* (Eds.), *Advances in Fuzzy Sets Theory and Applications*, North-Holland, Amsterdam, pp. 303-320, 1979.

[8] R. Viertl, “Univariate statistical analysis with fuzzy data,” *Computational Statistics and Data analysis*, vol. 51, pp. 133-147, 2006.

[9] L. A. Zadeh, “Fuzzy sets,” *Inform. Control*, vol. 8, pp. 338-353, 1965.

[10] B. Sadeghpour Gildeh, and D. Gien, “La distance-$D_{p,q}$ et le coefficient de correlation enter deux variables aleatories floues,” *Rencontres francophonessur la logique floue et ses applications LFA 01*, pp. 97-102, 2001.

[11] R. Korner, “An asymptotic α-tests for the expectation of random fuzzy variables,” *J. Statist. Plann. Inference*, vol. 83, pp. 331-346, 2000.