Sankhyā : The Indian Journal of Statistics
2009, Volume 71-B, Part 2, pp. 206-221

Statistical Inference about the Variance of Fuzzy Random Variables

M. GH. Akbari
University of Birjand, Southern Khorasan, Iran
A. H. Rezaei
University of Mashhad, Mashhad, Iran
Y. Waghei
University of Birjand, Southern Khorasan, Iran

Abstract

The variance of a fuzzy random variable plays an important role as a measure of central tendency. Some of the main contributions in this topic are consolidated and discussed in this paper. In case of the hypothesis testing problem, bootstrap techniques (Efron and Tibshirani, 1993) have empirically been shown to be efficient and powerful. Algorithms to apply these techniques in practice and some illustrative examples are provided. We also describe a bootstrap method for estimating the variance that is designed for the testing of hypotheses problem for fuzzy data based on the L_2 metric.

AMS (2000) *subject classification.* Primary 60A86.
Keywords and phrases. Fuzzy canonical number; L_2 metric; fuzzy random sample; testing of hypotheses; bootstrap method.

1 Introduction

Statistical analysis, in its traditional form, is based on the crispness of data, the associated random variables, the point estimation techniques used, the hypotheses and parameters of interest and so on. There can be many different situations in which the above mentioned concepts are imprecise. On the other hand, the theory of fuzzy sets is a well established tool for formulation and analysis of imprecise and subjective concepts. Testing of hypotheses with fuzzy data is therefore an important practical problem. The theory of statistical inference (confidence interval and testing of hypotheses) in fuzzy environments has been discussed and developed through different approaches in the literature.

Filzmoser and Viertl (2004) present an approach for statistical testing on the basis of fuzzy values by introducing the fuzzy p value. Torabi et al. (2006) try to develop a new approach for testing fuzzy hypotheses when the available data are also fuzzy. They state and prove a generalized Neyman-Pearson Lemma for such a problem. Some methods of statistical inference with fuzzy data are reviewed by Viertl (2006). Buckley (2005, 2006) studies the problem of statistical inference in a fuzzy environment. Thompson and Geyer (2007) have proposed the fuzzy p value in latent variable problems. Taheri and Arefi (2009) exhibit an approach for testing fuzzy hypotheses based on fuzzy test statistics. Akbari and Rezaei (2009) describe a bootstrap method for variance that is designed directly for hypothesis testing in case of fuzzy data based on the Yao-Wu signed distance. Parchami et al. (2009) consider the problem of testing of hypotheses when the hypotheses are fuzzy and the data are crisp. They first introduce the notion of a fuzzy p value by applying the extension principle, and then present an approach for testing fuzzy hypotheses by comparing a fuzzy p value and a fuzzy significance level, based on a comparison of two fuzzy sets.

The bootstrap technique for fuzzy data has also been developed through different approaches. Montenegro et al. (2004) have presented an asymptotic one-sample procedure. Gonzalez et al. (2006) have shown that the one-sample method of testing the mean of a fuzzy random variable can be extended to general ones (more precisely, to those whose range is not necessarily finite and whose values are fuzzy subsets of finite-dimensional Euclidean space). In this paper we construct a new method for bootstrap testing of hypotheses in fuzzy environments in an approach which is completely different from those mentioned above. For this purpose we organize the material in the following manner; in Section 2 we describe some basic concepts of canonical fuzzy numbers, the L_2 metric, fuzzy random variables, variance and fuzzy random samples. In Section 3 we summarize the testing of hypotheses results for the one-sample case. Section 4 provides a bootstrap hypothesis testing technique for the two sample problem. Finally, a brief conclusion is provided in Section 5.

2 Preliminaries

In this section we study canonical fuzzy numbers, L_2 metric, fuzzy random variable, variance and fuzzy random samples.

2.1. Canonical numbers. Let X be the universal space, then a fuzzy subset $\widetilde{x}$ of X is defined by its membership function $\mu_{\widetilde{x}} : X \rightarrow [0,1]$. We denote by $\widetilde{x}_\alpha = \{x : \mu_{\widetilde{x}}(x) \geq \alpha\}$ the $\alpha-$cut set of $\widetilde{x}$ and $\widetilde{x}_0$ is the closure of the set $\{x : \mu_{\widetilde{x}}(x) > 0\}$, and
(1) $\widetilde{x}$ is called normal fuzzy set if there exist $x \in X$ such that $\mu_{\widetilde{x}}(x) = 1$;
(2) $\widetilde{x}$ is called convex fuzzy set if $\mu_{\widetilde{x}}(\lambda x + (1-\lambda)y) \geq \min(\mu_{\widetilde{x}}(x)\ ,\ \mu_{\widetilde{x}}(y))$ for all $\lambda \in [0,1]$;
(3) the fuzzy set $\widetilde{x}$ is called a fuzzy number if $\widetilde{x}$ is normal convex fuzzy set and its $\alpha-$cut sets, is bounded $\forall \alpha \neq 0$;
(4) $\widetilde{x}$ is called a closed fuzzy number if $\widetilde{x}$ is fuzzy number and its membership function $\mu_{\widetilde{x}}$ is upper semicontinues;
(5) $\widetilde{x}$ is called a bounded fuzzy number if $\widetilde{x}$ is a fuzzy number and its membership function $\mu_{\widetilde{x}}$ has compact support.

If $\widetilde{x}$ is a closed and bounded fuzzy number with $x^L_\alpha = \inf\{x : x \in \widetilde{x}_\alpha\}$ and $x^U_\alpha = \sup\{x : x \in \widetilde{x}_\alpha\}$ and its membership function be strictly increasing on the interval $[x^L_\alpha, x^L_1]$ and strictly decreasing on the interval $[x^U_1, x^U_\alpha]$, then $\widetilde{x}$ is called canonical fuzzy number.

Let " $\odot$ " be a binary operation $\oplus$ or $\ominus$ between two canonical fuzzy numbers $\widetilde{a}$ and $\widetilde{b}$. The membership function of $\widetilde{a} \odot \widetilde{b}$ is defined by

$$\mu_{\widetilde{a}\odot\widetilde{b}}(z) = \sup_{x \circ y = z} \min\{\mu_{\widetilde{a}}(x), \mu_{\widetilde{b}}(y)\}$$

for $\odot = \oplus$ or $\ominus$ and $\circ = +$ or $-$.

In the following, let $\odot_{int}$ be a binary operation $\oplus_{int}$ or $\ominus_{int}$ between two closed intervals $\widetilde{a}_\alpha = [a^L_\alpha, a^U_\alpha]$ and $\widetilde{b}_\alpha = [b^L_\alpha, b^U_\alpha]$. Then $\widetilde{a}_\alpha \odot_{int} \widetilde{b}_\alpha$ is defined by

$$\widetilde{a}_\alpha \odot_{int} \widetilde{b}_\alpha = \{z \in \mathcal{R} : z = x \circ y,\ x \in \widetilde{a}_\alpha,\ y \in \widetilde{b}_\alpha\}.$$

If $\widetilde{a}$ and $\widetilde{b}$ be two closed fuzzy numbers. Then $\widetilde{a} \oplus \widetilde{b}$ and $\widetilde{a} \ominus \widetilde{b}$ are also closed fuzzy numbers. Furthermore, we have

$$\begin{aligned}(\widetilde{a} \oplus \widetilde{b})_\alpha &= \widetilde{a}_\alpha \oplus_{int} \widetilde{b}_\alpha = [a^L_\alpha + b^L_\alpha, a^U_\alpha + b^U_\alpha]\\(\widetilde{a} \ominus \widetilde{b})_\alpha &= \widetilde{a}_\alpha \ominus_{int} \widetilde{b}_\alpha = [a^L_\alpha - b^U_\alpha, a^U_\alpha - b^L_\alpha].\end{aligned}$$

2.2. L_2 metric. Now we define a distance between fuzzy numbers, which will be used later.

Several ranking methods have been proposed so far, by Cheng (1998), Yao and Wu (2000) Modarres and Sadi-Nezhad (2001), Nojavan and Ghazanfari (2006), Puri and Ralescu (1983), and Akbari and Rezaei (2009).

In this paper we use another metric for canonical fuzzy numbers that is called the L_2 metric.
Given a real number $x \in \mathcal{R}$, we can induce a fuzzy number $\widetilde{x}$ with membership function $\mu_{\widetilde{x}}(r)$ such that $\mu_{\widetilde{x}}(x) = 1$ and $\mu_{\widetilde{x}}(r) < 1$ for $r \neq x$. We call $\widetilde{x}$ as a fuzzy real number induced by the real number x.

Let $\mathcal{F}(\mathcal{R})$ be the set of all fuzzy real numbers induced by the real numbers $\mathcal{R}$. We define the relation $\sim$ on $\mathcal{F}(\mathcal{R})$ as $\widetilde{x}_1 \sim \widetilde{x}_2$ iff $\widetilde{x}_1$ and $\widetilde{x}_2$ are induced by the same real number x. Then $\sim$ is an equivalence relation, which induces the equivalence classes $[\widetilde{x}] = \{\widetilde{a} : \widetilde{a} \sim \widetilde{x}\}$. The quotient set $\mathcal{F}(\mathcal{R})/\sim$ is the set of all equivalence classes. We call $\mathcal{F}(\mathcal{R})/\sim$ as the fuzzy real number system. In practice, we take only one element $\widetilde{x}$ from each equivalence class $[\widetilde{x}]$ to from the fuzzy real number system $(\mathcal{F}(\mathcal{R})/\sim)$ that is,

$$(\mathcal{F}(\mathcal{R})/\sim) = \{\widetilde{x} : \widetilde{x} \in [\widetilde{x}], \widetilde{x} \;\; is\ the\ only\ element\ from\ [\widetilde{x}]\}.$$

If the fuzzy real number system $(\mathcal{F}(\mathcal{R})/\sim)$ consists all of canonical fuzzy real numbers then we call $(\mathcal{F}(\mathcal{R})/\sim)$ as the canonical fuzzy real number system.
For each $\alpha-$cuts of $\widetilde{a} \in \mathcal{F}(\mathcal{R}^n)$ the support function $S_{\widetilde{a}_\alpha}$ is defined as $S_{\widetilde{a}_\alpha}(t) = \sup_{x \in \widetilde{a}_\alpha} \ll x, t \gg$, $t \in S^{n-1}$, S^{n-1} the $(n-1)-$dimensional unit sphere in $\mathcal{R}^n$. Using support function we define L_2 metric

$$\delta_2(\widetilde{a}, \widetilde{b}) = \left(n \int_0^1 (\rho_2(\widetilde{a}_\alpha, \widetilde{b}_\alpha))^2 d\alpha \right)^{\frac{1}{2}} \qquad \widetilde{a}, \widetilde{b} \in \mathcal{F}(\mathcal{R}^n),$$

where

$$\rho_2(\widetilde{a}_\alpha, \widetilde{b}_\alpha) = \left(\int_{S^{n-1}} |S_{\widetilde{a}_\alpha}(t) - S_{\widetilde{b}_\alpha}(t)|^2 \mu(dt) \right)^{\frac{1}{2}}.$$

Note that μ is the normalized Lebesgue measure on S^{n-1}.

This metric is very realistic because

- it implies very good statistical properties in connection with variance;
- it involves distances between extreme points;
- it is distance with convenient statistical features.

EXAMPLE 2.1. As an example of a canonical fuzzy set on $\mathcal{R}$, consider so-called LR-fuzzy numbers $\widetilde{a} = (\mu, l, r)_{LR}$ with central value $\mu \in \mathcal{R}$, left and right spread $l \in \mathcal{R}^{\geq 0}$, $r \in \mathcal{R}^{\geq 0}$, decreasing left and right shape functions

$L : \mathcal{R}^{\geq 0} \to [0,1]$, $R : \mathcal{R}^{\geq 0} \to [0,1]$ with $L(0) = R(0) = 1$, i.e. a fuzzy set $\widetilde{a}$ with

$$\mu_{\widetilde{a}}(x) = \begin{cases} L(\frac{\mu - x}{l}) & x \leq \mu \\ \\ R(\frac{x-\mu}{r}) & x \geq \mu \end{cases}$$

An LR-fuzzy number $\widetilde{a} = (\mu, l, r)_{LR}$ with $L = R$ and $l = r = \varepsilon$ is called symmetric LR-fuzzy number and abbreviated by $\widetilde{a} = (\mu - \varepsilon, \mu, \mu + \varepsilon)$.

Let $\widetilde{a}_i = (\mu_i, l_i, r_i)_{LR}$; $i = 1, 2$. We have

$$\widetilde{a}_{i\alpha} = [\mu_i - L^{-1}(\alpha) l_i, \mu_i + R^{-1}(\alpha) r_i] \qquad i = 1, 2,$$

furthermore

$$S_{\widetilde{a}_{i\alpha}}(t) = \begin{cases} -\mu_i + L^{-1}(\alpha) l_i & t = -1 \\ \\ \mu_i + R^{-1}(\alpha) r_i & t = 1. \end{cases}$$

Thus

$$\delta_2^2(\widetilde{a}_1, \widetilde{a}_2) = (\mu_1 - \mu_2)^2 + \frac{1}{2}\int_0^1 (L^{-1}(\alpha))^2 d\alpha \, (l_1 - l_2)^2 + \frac{1}{2}\int_0^1 (R^{-1}(\alpha))^2 d\alpha \, (r_1 - r_2)^2$$

$$- \int_0^1 (L^{-1}(\alpha)) d\alpha \, (\mu_1 - \mu_2)(l_1 - l_2) + \int_0^1 (R^{-1}(\alpha)) d\alpha \, (\mu_1 - \mu_2)(r_1 - r_2)$$

For symmetric fuzzy numbers $\widetilde{a}_i = (\mu_i - \varepsilon_i, \mu_i, \mu_i + \varepsilon_i)$; $i = 1, 2$. We have

$$\delta_2^2(\widetilde{a}_1, \widetilde{a}_2) = (\mu_1 - \mu_2)^2 + \int_0^1 (L^{-1}(\alpha))^2 d\alpha \, (\varepsilon_1 - \varepsilon_2)^2.$$

2.3. Fuzzy random variable, variance and fuzzy random sample. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A compact convex random set (Cr.s.) $\mathbf{X}$ is a Borel measurable function from $(\Omega, \mathcal{F}, P)$ to $(\mathbf{X}, \mathcal{B}, P_{\mathbf{X}})$, where $P_{\mathbf{X}}$ is the probability measure induced by $\mathbf{X}$ and is called the distribution of the Cr.s. X , i.e.,

$$P_{\mathbf{X}}(A) = P(\mathbf{X} \in A) = \int_{\mathbf{X} \in A} dP \qquad \forall A \in \mathcal{B}.$$

DEFINITION 2.1. *A fuzzy random variable (F r.v.) is a Borel measurable function* $\widetilde{\mathbf{X}} : \Omega \to \mathcal{F}(\mathcal{R}^n)$ *where*

$$\{(\omega, x) : \omega \in \Omega \ , \ x \in \widetilde{\mathbf{X}}_\alpha(\omega)\} \in \ \mathcal{F} \times \mathcal{B} \qquad \forall \alpha \in [0,1].$$

Then all $\alpha-$cuts of $\widetilde{\mathbf{X}}$ are Cr.s. and further more The above definition used here is equivalent to the often used definition by Puri and Ralescu (1986), and for $n = 1$ to the definition by Kwakernaak (1978).

LEMMA 2.1. *Let $\mathcal{F}(\mathcal{R})$ be a canonical fuzzy real number system. Then $\widetilde{X}$ is a F r.v. iff X_α^L and X_α^U are random variables for all $\alpha \in [0,1]$.*

The expected value $\widetilde{E}(\widetilde{\mathbf{X}})$ of the F r.v. $\widetilde{\mathbf{X}}$ is defined by

$$\widetilde{E}_\alpha(\widetilde{\mathbf{X}}) = \{E(\mathbf{X}) | \mathbf{X} : \Omega \to \mathcal{R}^n,\ \mathbf{X}(\omega) = \widetilde{\mathbf{X}}_\alpha(\omega)\}.$$

DEFINITION 2.2. *The variance of a F r.v. $\widetilde{\mathbf{X}}$ is defined as $\nu(\widetilde{\mathbf{X}}) = E[\delta_2^2(\widetilde{\mathbf{X}},\ E(\widetilde{\mathbf{X}}))]$. Using $E_\alpha(\widetilde{\mathbf{X}}) = E(\widetilde{\mathbf{X}}_\alpha)$ and $S_{E(\widetilde{\mathbf{X}}_\alpha)}(t) = E(S_{\widetilde{\mathbf{X}}_\alpha}(t))$ this can be written as*

$$\nu(\widetilde{\mathbf{X}}) = n \int_0^1 \int_{S^{n-1}} Var(S_{\widetilde{\mathbf{X}}_\alpha}(t))\mu(dt)d\alpha.$$

Näther (2006) defined an scalar multiplication between $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ given by

$$< \widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}} >= n \int_0^1 \int_{S^{n-1}} S_{\widetilde{\mathbf{X}}_\alpha}(t) S_{\widetilde{\mathbf{Y}}_\alpha}(t)\mu(dt)d\alpha,$$

thus

$$\nu(\widetilde{\mathbf{X}}) = E < \widetilde{\mathbf{X}}, \widetilde{\mathbf{X}} > - < E(\widetilde{\mathbf{X}}), E(\widetilde{\mathbf{X}}) >$$

and similarly

$$\begin{aligned} Cov(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) &= n \int_0^1 \int_{S^{n-1}} Cov(S_{\widetilde{\mathbf{X}}_\alpha}(t), S_{\widetilde{\mathbf{Y}}_\alpha}(t))\mu(dt)d\alpha \\ &= E < \widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}} > - < E(\widetilde{\mathbf{X}}), E(\widetilde{\mathbf{Y}}) > . \end{aligned}$$

DEFINITION 2.3. *Let $\widetilde{X}$ and $\widetilde{Y}$ be two F r.v.'s. We say that $\widetilde{X}$ and $\widetilde{Y}$ are independent iff each random variable in the set $\{X_\alpha^L, X_\alpha^U : 0 \le \alpha \le 1\}$ is independent with any random variable in the set $\{Y_\alpha^L, Y_\alpha^U : 0 \le \alpha \le 1\}$.*

DEFINITION 2.4. *We say $\widetilde{X}$ and $\widetilde{Y}$ are identically distributed iff X_α^L, Y_α^L are identically distributed, and X_α^U, Y_α^U are identically distributed for all $\alpha \in [0,1]$.*

DEFINITION 2.5. *We say* $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \widetilde{X}_2, ..., \widetilde{X}_n)$ *is a fuzzy random sample iff* $\widetilde{X}_i$*'s are independent and identically distributed.*

LEMMA 2.2. *Let* $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \widetilde{X}_2, ..., \widetilde{X}_n)$ *be a fuzzy random sample. The sample fuzzy variance value* $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n \delta_2^2\left(\ \widetilde{X}_i, \overline{\widetilde{\mathbf{X}}}\ \right)$ *is an unbiased estimator of the parameter* $\nu(\widetilde{X})$*; where* $\overline{\widetilde{\mathbf{X}}}$ *is the sample fuzzy mean value* $\frac{1}{n}\oplus_{i=1}^n \widetilde{X}_i$.

PROOF. We have

$$
\begin{aligned}
&E[S_n^2] \\
&= \frac{1}{n-1}\sum_{i=1}^n \int_0^1\int_{-1}^1 E[S_{\widetilde{X}_{i\alpha}}(t) - S_{\overline{\widetilde{\mathbf{X}}}_\alpha}(t)]^2 dt d\alpha \\
&= \frac{1}{n-1}\sum_{i=1}^n \int_0^1\int_{-1}^1 E[S_{\widetilde{X}_{i\alpha}}(t) - E(S_{\widetilde{X}_{i\alpha}}(t)) + E(S_{\widetilde{X}_{i\alpha}}(t)) - S_{\overline{\widetilde{\mathbf{X}}}_\alpha}(t)]^2 dt d\alpha \\
&= \frac{1}{n-1}\sum_{i=1}^n \left[\nu(\widetilde{X}_i) + \int_0^1\int_{-1}^1 \left\{Var[S_{\overline{\widetilde{X}}_\alpha}(t)] - 2Cov(S_{\widetilde{X}_{i\alpha}}(t), S_{\overline{\widetilde{\mathbf{X}}}_\alpha}(t))\right\} dt d\alpha\right] \\
&= \frac{1}{n-1}\sum_{i=1}^n \left[\ \nu(\widetilde{X}_i) + \tfrac{\nu(\widetilde{X}_i)}{n} - 2\tfrac{\nu(\widetilde{X}_i)}{n}\ \right] \\
&= \nu(\widetilde{X}).
\end{aligned}
$$

LEMMA 2.3. *Consider Lemma 2.2. For given crisp value* S_n^2

$$\lim_{n\to\infty} S_n^2 = \nu(\widetilde{X}).$$

PROOF. It is a special condition of Strong Law of Large Numbers.

LEMMA 2.4. *If* $Var(X)$ *be a variance of the crisp random variable* X*, then we have*

$$\nu(X) = Var(X).$$

PROOF. From Example 2.1, it is obvious.

3 Bootstrap Testing of Hypotheses for the Single-sample Case

In this section we describe a method for bootstrap hypothesis testing in the single-sample case using fuzzy data.

We note that, in classical testing of hypotheses, there is a relationship between interval estimation and testing of hypothesis. We describe a bootstrap method that is designed based on relationship between interval estimation and testing of hypotheses, for hypothesis testing for fuzzy data based on L_2 metric.

Suppose that we have fuzzy random samples $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \widetilde{X}_2, ..., \widetilde{X}_n)$, and we want to test the fuzzy null hypotheses

H_0: the variance of observations is σ_0^2

versus

H_1: the variance of observations is not σ_0^2.

We need a distribution that estimates the population of treatment times under H_0. Note first that the empirical distribution (i.e., putting probability $\frac{1}{n}$ on each member of $\widetilde{\mathbf{X}}$) is not an approximate estimate for distribution because is does not obey H_0. Somehow we need to obtain an estimate of the population that has variance σ_0^2. A simple way is to translate the empirical distribution so that is has the desired variance. In other word, we use as our estimated null distribution the empirical distribution on the values

$$\widetilde{X}_{ci} = \frac{\sigma_0 \widetilde{X}_i}{S_n} \qquad i = 1, 2, ..., n,$$

since, under the hypothesis H_0,

$$\begin{aligned}
S_{cn}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\int_0^1\int_{-1}^{1}[S_{\widetilde{X}_{ci\alpha}}(t) - S_{\overline{\widetilde{\mathbf{X}}}_{c\alpha}}(t)]^2 dt d\alpha, \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\int_0^1\int_{-1}^{1}[\frac{\sigma_0}{S_n} S_{\widetilde{X}_{i\alpha}}(t) - \frac{\sigma_0}{S_n} S_{\overline{\widetilde{\mathbf{X}}}_{\alpha}}(t)]^2 dt d\alpha \\
&= \frac{\sigma_0^2}{S_n^2}\frac{1}{n-1}\sum_{i=1}^{n}\int_0^1\int_{-1}^{1}[S_{\widetilde{X}_{i\alpha}}(t) - S_{\overline{\widetilde{\mathbf{X}}}_{\alpha}}(t)]^2 dt d\alpha \\
&= \sigma_0^2,
\end{aligned}$$

as $\frac{\sigma_0}{S_n} > 0$ and, for any $r > 0$, $S_{r\widetilde{X}_\alpha} = r S_{\widetilde{X}_\alpha}$.

We generate B bootstrap fuzzy random sample $\widetilde{\mathbf{X}}^{*^1}, \widetilde{\mathbf{X}}^{*^2}, ..., \widetilde{\mathbf{X}}^{*^B}$ (i.e., each $\widetilde{\mathbf{X}}^{*^b}$ is a fuzzy sample of size n drawn randomly and replacement from $\widetilde{\mathbf{X}}$). We have the following:

- ***In testing the null hypothesis*** $H_0: \ \sigma^2 = \sigma_0^2$ ***at the nominal significance level*** $\gamma \in [0,1]$***,*** H_0 ***should be rejected whenever***

$$\chi^2 = \frac{(n-1)S_n^2}{\sigma_0^2} > z_{1-\frac{\gamma}{2}} \quad or \quad \chi^2 = \frac{(n-1)S_n^2}{\sigma_0^2} < z_{\frac{\gamma}{2}}$$

where z_γ ***is the*** $100(1-\gamma)$ ***fractile of the distribution of the bootstrap*** $\chi^{2*^b} = \frac{(n-1)S_{cn}^{2*^b}}{\sigma_0^2}$ ***and with***

$$S_{cn}^{2*^b} = \frac{1}{n-1}\sum_{i=1}^{n}\int_0^1\int_{-1}^1 [S_{\widetilde{X}^{*b}_{ci\alpha}}(t) - S_{\overline{\widetilde{\mathbf{X}}}^{*b}_{c\alpha}}(t)]^2 dt d\alpha \qquad b = 1, 2, ..., B.$$

If $B \times \gamma$ is not an integer, the following procedure can be used. Assuming $\gamma \leq 0.5$, let $k = [(B+1)\gamma]$, the largest integer $\leq (B+1)\gamma$. Then we define the empirical γ and $1-\gamma$ quantizes by the kth largest and $(B+1-k)$th largest values of χ^{2*^b}, respectively.

Now we brief the steps of bootstrap testing of hypotheses for one-sample as the follows:

Computation of the bootstrap test statistics for testing $H_0 : \sigma^2 = \sigma_0^2$
1. Draw B samples of size n with replacement from $(\widetilde{x}_i; i = 1, 2, ..., n)$.
2. compute the value of the bootstrap statistics $\chi^{2*^b} \quad b = 1, 2, ..., B$.
3. compute the bootstrap $z_{\frac{\gamma}{2}}$ and $z_{1-\frac{\gamma}{2}}$ values.

Algorithm 1. One-sample bootstrap test algorithm

Example 3.1. Consider a fuzzy random sample of size $n = 15$ from a population, given in Table 1.

TABLE 1. FUZZY RANDOM SAMPLE OF SIZE $n = 15$ FROM A POPULATION

N	Observation	N	Observation	N	Observation
1	$(33, 35, 36)$	6	$(60, 63, 66)$	11	$(41, 41, 41)$
2	$(81, 82, 84)$	7	$(70, 70, 72)$	12	$(54, 56, 58)$
3	$(85, 87, 87)$	8	$(70, 73, 76)$	13	$(40, 40, 42)$
4	$(90, 90, 90)$	9	$(69, 70, , 73)$	14	$(94, 96, 99)$
5	$(53, 53, 55)$	10	$(38, 40, 40)$	15	$(39, 41, 44)$

Now suppose that we want to test the following hypotheses

$$\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 \neq \sigma_0^2. \end{cases}$$

For $\gamma = 0.05$, The percentiles and histogram (for $\sigma_0^2 = 400$) of χ^{2*^b} based on 10000 bootstrap samples are shown in Table 2 and figure 1 respectively.

TABLE 2. σ_0^2, χ^2, $z_{\frac{\gamma}{2}}$, $z_{1-\frac{\gamma}{2}}$ AND *result*

σ_0^2	χ^2	$z_{\frac{\gamma}{2}}$	$z_{1-\frac{\gamma}{2}}$	*result*
100	39.57	6.54	18.95	*reject* H_0
225	17.58	7.35	19.22	*accept* H_0
400	9.89	6.74	19.11	*accept* H_0
625	6.33	6.2	18.9	*accept* H_0
900	4.4	6.91	19.41	*reject* H_0
1225	3.23	6.51	19.58	*reject* H_0

EXAMPLE 3.2. The water levels of a river in a year can not be measured in an exact way because of the fluctuation. Under this consideration, the more appropriate way to describe the water levels are to say that the water levels are around 35,40,38,42,41,34,44,... meters. For example the phrases "approximately 35" should be regarded as a fuzzy number $\widetilde{35}$ that will be realize thought fuzzy sets theory. We may come across more realistic expressions about parameter would be considered as: "less than 35", "more than 35", "essentially less than 35 ", and so on.

For some fuzzy observations of $\widetilde{x}_1 = \widetilde{9}$, $\widetilde{x}_2 = \widetilde{7}$, $\widetilde{x}_3 = \widetilde{15}$, $\widetilde{x}_4 = \widetilde{17}$, $\widetilde{x}_5 = \widetilde{11}$, $\widetilde{x}_6 = \widetilde{9}$, $\widetilde{x}_7 = \widetilde{20}$, $\widetilde{x}_8 = \widetilde{19}$, $\widetilde{x}_9 = \widetilde{21}$ with triangular membership function,

$$\mu_{\widetilde{x}_i}(y) = \begin{cases} \frac{y - x_i + a}{a} & x_i - a \leq y \leq x_i \\ \\ \frac{x_i + b - y}{b} & x_i \leq y \leq x_i + b, \end{cases}$$

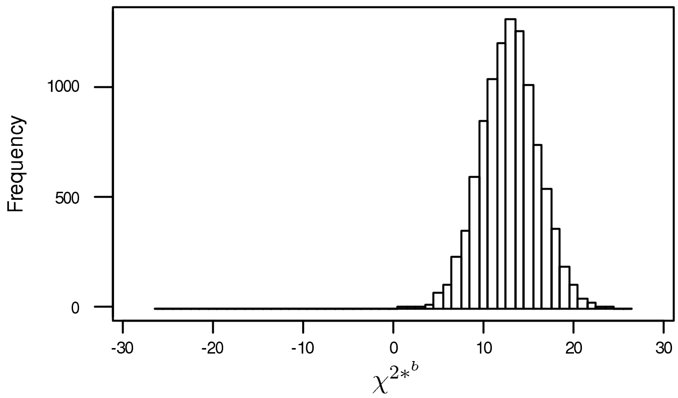

Figure 1: Bootstrap distribution of χ^{2*^b} for $\sigma_0^2 = 400$

for each $1 \leq i \leq 9$ and $a, b \geq 0$. Now suppose that we want to test the following hypotheses

$$\begin{cases} H_0: & \sigma^2 = 9 \\ H_1: & \sigma^2 \neq 9. \end{cases}$$

For $\gamma = 0.05$, we have $\chi^2 = 3.11$, $z_{\frac{\gamma}{2}} = 2.325$, and $z_{1-\frac{\gamma}{2}} = 15.88$. Thus we accept H_0.

4 Bootstrap Testing of Hypotheses for the Two-sample Case

In this section we describe one method to bootstrap testing of hypotheses for the two-sample case based on fuzzy data.

Let $\widetilde{\mathbf{z}} = (\widetilde{z}_1, \widetilde{z}_2, ..., \widetilde{z}_n)$ and $\widetilde{\mathbf{y}} = (\widetilde{y}_1, \widetilde{y}_2, ..., \widetilde{y}_m)$ be fuzzy random samples from possibly different probability distributions, and we wish to test the null hypothesis

H_0 : the variance of first population (σ_1^2) is equal to variance of second population (σ_2^2).

H_1 : the variance of first population (σ_1^2) is not equal to variance of second population (σ_2^2).

Denote the combined fuzzy sample by $\widetilde{\mathbf{x}} = (\widetilde{z}_1, \widetilde{z}_2, ..., \widetilde{z}_n, \widetilde{y}_1, \widetilde{y}_2, ..., \widetilde{y}_m)$ and let a probability $\frac{1}{n+m}$ be assigned to each member of $\widetilde{\mathbf{x}}$. We have:

- ***In testing the null hypothesis*** $H_0 : \sigma_1^2 = \sigma_2^2$ ***at the nominal significance level*** $\gamma \in [0,1]$***,*** H_0 ***should be rejected whenever***

$$\xi = \frac{(n-1)S_{1n}^2}{(m-1)S_{2n}^2} > z_{1-\frac{\gamma}{2}} \quad or \quad \xi = \frac{(n-1)S_{1n}^2}{(m-1)S_{2n}^2} < z_{\frac{\gamma}{2}}$$

where z_γ ***is the*** $100(1-\gamma)$ ***fractile of the distribution of the bootstrap*** $\xi^{*^b} = \frac{(n-1)S_{1cn}^2}{(m-1)S_{2cn}^2}$ ***and with***

$$S_{1cn}^{2*^b} = \frac{1}{n-1}\sum_{i=1}^{n}\int_0^1\int_{-1}^{1}[S_{\widetilde{Z}_{ci\alpha}^{*b}}(t) - S_{\overline{\widetilde{\mathbf{Z}}}_{c\alpha}^{*b}}(t)]^2 dt d\alpha \qquad b = 1, 2, ..., B,$$

$$S_{2cn}^{2*^b} = \frac{1}{n-1}\sum_{i=1}^{n}\int_0^1\int_{-1}^{1}[S_{\widetilde{Y}_{ci\alpha}^{*b}}(t) - S_{\overline{\widetilde{\mathbf{Y}}}_{c\alpha}^{*b}}(t)]^2 dt d\alpha \qquad b = 1, 2, ..., B.$$

Now we introduce the steps of bootstrap for the two-sample case as follows:

Computation of the bootstrap test statistics for testing H_0
1. Draw B samples of size $n+m$ with replacement from $\widetilde{\mathbf{x}}$. Call the first n observations $\widetilde{\mathbf{z}}^{*^b}$ and the remaining m observations $\widetilde{\mathbf{y}}^{*^b}$, for $b = 1, 2, ..., B$.
2. Compute the value of the bootstrap statistics ξ^{*^b} $b = 1, 2, ..., B$.

Algorithm 2. Two-sample bootstrap test algorithm

EXAMPLE 4.1. Consider fuzzy random samples of sizes $n = 9$ and $m = 8$ respectively from two population. The corresponding triangular fuzzy data is given in Table 3.

Table 3. Fuzzy random samples of sizes $n = 9$ and $m = 8$ from two populations

The first population	The second population
$(19, 20, 22)$	$(92, 94, 95)$
$(47, 50, 53)$	$(68, 68, 70)$
$(100, 100, 100)$	$(43, 44, 45)$
$(58, 60, 61)$	$(57, 59, 61)$
$(49, 50, 51)$	$(99, 99, 99)$
$(90, 92, 93)$	$(82, 83, 85)$
$(9, 10, 10)$	$(23, 23, 23)$
$(25, 27, 28)$	$(40, 42, 43)$
$(46, 46, 46)$	

The percentiles and histogram, using 10000 bootstrap samples, of ξ^{*^b} are given in Table 4 and Figure 2 respectively.

Table 4. Percentiles of the bootstrap distribution of ξ^{*^b}

Percentile	0.005	0.025	0.05	0.1	0.9	0.95	0.975	0.995
Bootstrap	0.04	0.225	0.37	0.552	3.316	4.33	5.77	10.21

For $\gamma = 0.05$, we have $\xi = 0.999$, so we accept H_0.

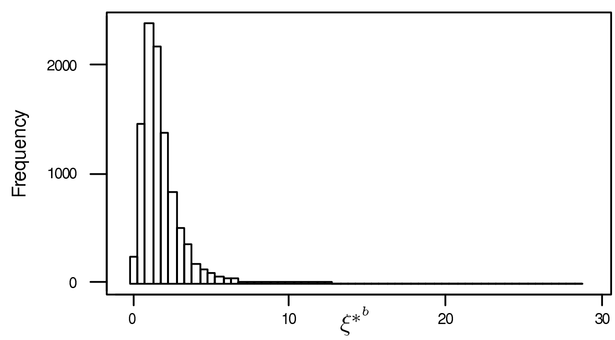

Figure 2: Bootstrap distribution of ξ^{*^b}

EXAMPLE 4.2. A food company wished to test two different package designs for a new product. Seventeen stores, with approximately equal sales volumes, are selected as the experimental units. Package designs 1 is assigned to nine stores and package designs 2 is assigned to eight stores.

The data are displayed in Table 5, while the percentiles and histogram, using 10000 bootstrap samples, of ${\xi^*}^b$ are given in Table 6 and Figure 3 respectively.

For $\gamma = 0.05$, we have $\xi = 5.64$, so H_0 is rejected.

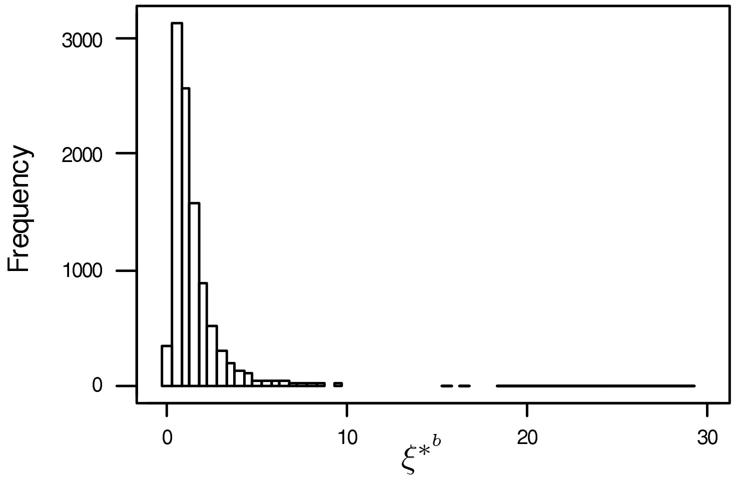

Figure 3: Bootstrap distribution of ${\xi^*}^b$

TABLE 5. FUZZY DATA FOR SALES VOLUMES

Store	Package design 1	Package design 2
1	$(14, 15, 15)$	$(13, 15, 15.5)$
2	$(22, 22, 22)$	$(14, 14, 14)$
3	$(9, 9, 10)$	$(17.5, 18, 19)$
4	$(11, 12, 13)$	$(13, 14, 14.5)$
5	$(9.5, 10, 11)$	$(11.5, 12, 12.6)$
6	$(18.5, 19, 19.2)$	$(17, 18, 19)$
7	$(19, 19, 19.5)$	$(14, 15, 16)$
8	$(25, 27, 28)$	$(12.5, 13, 13)$
9	$(16, 17, 19)$	

Table 6. Percentiles of the bootstrap distribution of ξ^{*^b}

Percentile	0.005	0.025	0.05	0.1	0.9	0.95	0.975	0.995
Bootstrap	0.115	0.211	0.301	0.403	2.735	3.7	4.874	9.003

5 Conclusions

In this paper, we have proposed a technique for bootstrap testing of hypotheses for one-sample and two-sample problem for fuzzy data based on the L_2 metric. Application of the proposed method to test the variance, correlation and other parameters of linear models, such as regression models and design of experiment, will be natural extensions of this approach and consitute potential future work. Furthermore we can construct bootstrap confidence intervals for crisp and fuzzy parameters.

Acknowledgements

The authors are grateful to the referees and the editor for their suggestions.

References

Akbari, M. GH. and Rezaei, A. (2009). Bootstrap statistical inference for the variance based on fuzzy data. *Austrian Journal of Statistics*, **38**, 121-130.

Akbari, M. GH. and Rezaei, A. (2009). Order statistics using fuzzy random variables. *Statistics and Probability Letters* **79**, 1031-1037.

Buckley, J. J. (2005). *Fuzzy probabilities : New Approach and Applications.* Springer-Verlag, Berlin, Heidelberg.

Buckley, J. J. (2006). *Fuzzy Probability and Statistics.* Springer-Verlag, Berlin, Heidelberg.

Cheng, C. (1998). A new approach for ranking fuzzy numbers by distance method. *Fuzzy Sets and Systems* **95**, 307-317.

Gonzalez-Rodriguez, G., Montenegro, M., Colubi, A. and Gil, M. A. (2006). Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data.*Fuzzy Sets and Systems* **157**, 2608-2613.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman-Hall, New York.

Filzmoser, P. and Viertl, R. (2004). Testing of hypotheses with fuzzy data: the fuzzy p-value. *Metrika* **59**, 21-29.

Kwakernaak H. (1978). Fuzzy random variables I. *Information Sciences* **15**, 1-29.

MODARRES, M. AND SADI-NEZHAD, S. (2001). Ranking fuzzy numbers by preference ratio. *Fuzzy Sets and Systems* **118**, 429-436.

MONTENEGRO, M., COLUBI, A., CASALS, M. R. AND GIL, M. A. (2004). Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable, *Metrika* **59**, 31-49.

NÄTHER, W. (2006). Regression with fuzzy data. *Computational Statistics and Data Analysis* **51**, 235-252.

NOJAVAN, M. AND GHAZANFARI, M. (2006). A fuzzy ranking method by desirability index. *Journal of Intelligent and Fuzzy Systems* **17**, 27-34.

PARCHAMI, A, TAHERI, S. M. AND MASHINCHI, M. (2009). Fuzzy p-value in testing fuzzy hypotheses with crisp data. *Statistical Papers*, Online.

PURI, M. L. AND RALESCU, D. A. (1983). Differential of fuzzy functions. *Journal of Mathematical Analysis and Applications* **114**, 552-558.

PURI, M. L. AND RALESCU, D. A. (1986). Fuzzy random variables. *Journal of Mathematical Analysis and Applications* **114**, 409-422.

TAHERI, S. M. AND AREFI, M. (2009). Testing fuzzy hypotheses based on fuzzy statistics. *Soft Computing* **13**, 617-625.

THOMPSON, E. A. AND GEYER C. J. (2007). Fuzzy p-values in latent variable problems. *Biometrika* **94**, 49-60.

TORABI, H., BEHBOODIAN, J. AND TAHERI, S. M. (2006). Neyman-Pearson Lemma for fuzzy hypotheses testing withy vague data. *Metrika* **64**, 289-304.

VIERTL, R. (2006). Univariate statistical analysis with fuzzy data. *Computational Statistics and Data Analysis* **51**, 133-147.

YAO, J. S. AND WU, K. (2000). Ranking fuzzy numbers based on decomposition principle and signed distance. *Fuzzy Sets and Systems* **11**, 275-288.

M. GH. AKBARI
DEPARTMENT OF STATISTICS
FACULTY OF SCIENCES
UNIVERSITY OF BIRJAND
SOUTHERN KHORASAN
IRAN
E-mail: akbari@isical.in

A. H. REZAEI
DEPARTMENT OF STATISTICS
SCHOOL OF MATHEMATICAL SCIENCES
FERDOWSI
UNIVERSITY OF MASHHAD
MASHHAD
IRAN

Y. WAGHEI
DEPARTMENT OF STATISTICS
FACULTY OF SCIENCES
UNIVERSITY OF BIRJAND
SOUTHERN KHORASAN
IRAN

Paper received November 2007; revised November 2008.