Fuzzy Multivariate Analysis of Variance
by Fuzzy Vector Operation

Abstract

We propose some properties of fuzzy multinomial analysis of variance by fuzzy vector operation with agreement index.

Keywords: fuzzy hypotheses testing, multinomial analysis of variance, agreement index, fuzzy vector, agreement index.

1. Introduction

Our primary propose of statistical test is fuzzy multivariate analysis of variance.

The generalization for simple hypotheses is given by Watanabe and Imaizumi([8]). In Gizegorzewski([2]), the fuzzy hypotheses testing was considered that the data(observation) are vague data and the hypotheses are fuzzy. Also, Kang, Choi and Han([3],[4],[5]) was suggested some estimations of fuzzy variance components for fixed effect model with fuzzy data.

We propose some properties of fuzzy multivariate analysis of variance by fuzzy vector operation with agreement index.

2. Some distance of fuzzy number

We denote by fuzzy number in $\mathcal{E}^n$

$$A=(a_1, a_2, \cdots, a_n)$$

(2.1)

where $a_i(i=1,\cdots,n)$ are projection of $A$ to axis $X_i(i=1,\cdots,n)$, fuzzy number in $R$, respectively.

Definition 2.1. The $\gamma$-level set of fuzzy number in $\mathcal{E}^n$ is define by
\[ [A]^\gamma = \{ (x_1, \ldots, x_n) \in \mathbb{R}^n : (x_1, \ldots, x_n) \in \prod_{i=1}^n [a_i]^\gamma \} \] (2.2)

where \( \prod \) is the Cartesian product of sets.

**Definition 2.2.** Let \( A \) and \( B \) in \( \mathcal{E}_N^\gamma \), for all \( \gamma \in (0, 1] \),

\[ A = B \iff [A]^\gamma = [B]^\gamma \] (2.3)

\[ [A^* B]^\gamma = \prod_{i=1}^n [a_i^* b_i]^\gamma \] (2.4)

where \( ^* \) is operation in \( \mathcal{E}_N^\gamma \) and \( * \) is operation in \( \mathcal{E}_N^\gamma \).

Let \( \prod_{i=0}^n [a_i]^\gamma \), \( 0 < \gamma \leq 1 \), be a given family of nonempty areas.

If \( \prod_{i=0}^n [a_i]^\gamma \subseteq \prod_{i=0}^n [b_i]^\gamma \) for \( 0 < \gamma_1 < \gamma_2 < 1 \) (2.5)

and \( \prod_{i=0}^n \lim_{\gamma \to 0} [a_i]^\gamma = \prod_{i=0}^n [a_i]^\gamma \) (2.6)

then the family \( \prod_{i=0}^n [a_i]^\gamma \), \( 0 < \gamma \leq 1 \), represents the \( \gamma \)-level sets of a fuzzy number \( A \in \mathcal{E}_N^\gamma \), where \( \{ \gamma_n \} \) is a nondecreasing sequence converging to \( \gamma \in (0, 1] \).

Conversely, if \( \prod_{i=0}^n [a_i]^\gamma \), \( 0 < \gamma \leq 1 \), are the \( \gamma \)-level sets of a fuzzy number in \( \mathbb{R}^n \), then conditions (2.5) and (2.6) are true.

We define the metric \( d_{\infty} \) on \( \mathcal{E}_N^\gamma \).

**Definition 2.3.** Let \( A, B \in \mathcal{E}_N^\gamma \),

\[ d_{\infty}(A, B) = \sup \{ d_{\infty}( [A]^\gamma, [B]^\gamma ) : \gamma \in (0, 1) \} \]

\[ = \sup \left\{ d_{\infty} \left( \prod_{i=0}^n [a_i^\gamma, \prod_{i=0}^n [b_i]^\gamma ] : \gamma \in (0, 1) \right) \right\} \]

\[ = \sup \left\{ \sqrt{ \sum_{i=1}^n d_{\infty}( [a_i^\gamma, [b_i]^\gamma ]^2 ) : \gamma \in (0, 1) } \right\} \] (4.7)

where \( d_H \) is Hausdorff distance.

### 3. A fuzzy statistical test

Let \( \bar{x} \) be a random sample from space \( \Omega \) and \( \{ P_\theta, \theta \in \Theta \} \) be a family of fuzzy probability, where \( \theta \) is a parameter and \( \Theta \) is a parameter space. For each \( \phi \in \Theta \), we consider a family of hypotheses \( \{(H_0(\phi), H_1(\phi)) : \phi \in \Theta \} \) and introduce the fuzzy hypothesis as a fuzzy subset.

**Definition 3.1.** The fuzzy hypothesis \( H_f \) is a fuzzy subset of \( \{(H_0(\phi), H_1(\phi)) : \phi \in \Theta \} \) with fuzzy hypothesis membership function \( \chi_H(f(H_0(\phi), H_1(\phi)) \).

We set with simplicity

\[ \chi_H(f) = \chi_H(f(H_0(\phi), H_1(\phi)) \]

and assume normality as convexity.

The fuzzy null hypothesis can be defined as follows.

**Definition 3.2.** The fuzzy null hypothesis \( H_{f, 0} \) is a fuzzy subset of \( \Theta \) with a membership function \( \chi_H(\phi) \). The fuzzy alternative hypothesis \( H_{f, 1} \) is a fuzzy subset of \( \Theta \) defined by the equation

\[ H_{f, 1} = H_{f, 0} \cap \{ \bigcup_{\gamma \in (0, 1)]} \gamma (\{ (a_i, (a_i, [b_i]^\gamma ) : \gamma \geq 0 \} \} \}

where \( \gamma(M) \) stands for the fuzzy set whose membership function of a set \( M \).

The first term on the right hand characteristic function of a set \( M \).

**Definition 3.3.** Let us consider a fuzzy number \( M \subset \mathbb{R} \), which we call the agreement index of \( M \) with regard to \( H_1 \),
the ratio being defined in the following way:
\[ I(M, H) = \frac{\text{area } M \cap H}{\text{area } M} \in [0, 1]. \]  
(3.3)

Using membership function \( R(\alpha, \phi) \) of critical region where \( \alpha \) is significance level, we also define the fuzzy hypothesis membership function \( \chi_{R} \) on \( \{0, 1\} \) as follows.

**Definition 3.4.** We define the real-valued function \( R_a \) on \( \Theta \) as in **Definition 3.1.** The maximum grade membership function of acceptance or rejection is

\[ \chi_{R_a}(0) = \sup_{\phi} \left\{ \text{area } H \phi \cap \text{area } \text{H}(\phi) \right\} \]

\[ \chi_{R_a}(1) = 1 - \chi_{R_a}(0) \]  
(4.4)

Let \( R_a \) denote the fuzzy subset of an entire set \( \{0, 1\} \) defined by \( \chi_{R_a} \), since \( \{0, 1\} \) corresponds to "accept", "reject", the value \( \chi_{R_a}(1) \) and \( \chi_{R_a}(0) \) are equal to the grades of the judgements that the hypothesis is rejected or not rejected.

Now, we show the multivariate statistical properties of our testing method.

**4. Fuzzy MANOVA**

Fuzzy multivariate analysis of variance model for comparing \( a \) population mean vectors is given by

\[ X_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \cdots, a \]

and \( j = 1, 2, \cdots, n \).

where \( X_{ij} = (x_{i\theta1}, x_{i\theta2}, \cdots, x_{i\theta p})' \) is fuzzy vector and \( \epsilon_{ij} \) are independent \( N(0, \Sigma) \) variables.

Here the parameter vector \( \mu \) is an overall fuzzy mean(average) and \( \alpha_i \) represents the \( i \)th treatment fuzzy effect with \( \sum_{i=1}^{a} n_i \alpha_i = 0 \).

The errors for the components of \( X_{ij} \) are correlated, but the fuzzy variance-covariance matrix \( \Sigma \) is the same for the populations.

A vector of observations may be decomposed as suggested by the model.

Thus

\[ X_{\bar{a}} = \bar{X} \otimes (X_{\bar{a}} \otimes \bar{X}), \]

(4.2)

where \( X_{\bar{a}} \) is fuzzy observation vector, \( \bar{X} \) is overall sample fuzzy mean v.s. \( \mu \), \( (X_{\bar{a}} \otimes \bar{X}) \) is estimated treatment fuzzy effect v.s. \( \bar{a} \), \( (X_{\bar{a}} \otimes X_{\bar{a}}) \) is fuzzy residual v.s. \( e_{ij} \).

The decomposition in (4.2) leads to the multivariate analog of the univariate sum of squares. First we note that the cross-product can be written as

\[ (X_{\bar{a}} \otimes \bar{X})(X_{\bar{a}} \otimes \bar{X})' = [(X_{\bar{a}} \otimes \bar{X})(X_{\bar{a}} \otimes \bar{X})][(X_{\bar{a}} \otimes \bar{X})(X_{\bar{a}} \otimes \bar{X})]' \]

(4.3)

The sum of over \( j \) of the middle two expressions is the fuzzy zero matrix become

\[ \sum_{j=1}^{n}(X_{\bar{a}} \otimes \bar{X}) = 0. \]

Next, summing the cross-product over \( i \) and \( j \) yields

\[ \sum_{i=1}^{a} \sum_{j=1}^{n}(X_{ij} \otimes \bar{X})(X_{ij} \otimes \bar{X})' = \sum_{i=1}^{a} n_i (X_{\bar{a}} \otimes \bar{X})(X_{\bar{a}} \otimes \bar{X})' \]

(4.4)

it's means that (Total sum of squares and cross product ) = (treatment(Between) sum of squares and cross product) + (residual(Within) sum of squares and cross product).

The hypothesis of no treatment effects

\[ H_{\bar{a}i}: a_i \approx a_2 \approx \cdots \approx a_a \approx 0 \]

(4.5)

is tested by considering the relative size of treatment and residual fuzzy sums of squares and cross-product.

The degree of freedom correspond involving Wishart densities.
One test of \( H_{J,0}: \alpha_1 \cong \alpha_2 \cong \cdots \cong \alpha_n \cong 0 \) involves generalized variances. We reject \( H_{J,0} \) if the ratio of generalized variance

\[
\Lambda^* = \frac{|W|}{|B + W|}
\]

is too small. The quantity \( \Lambda^* \) proposed originally by Wilks. The exact distribution of \( \Lambda^* \) can be derived for the special cases of \( \rho \geq 1, a = 2 \) with

\[
\left( \frac{\sum n_i - p - 1}{p} \right) \left( \frac{1 - \Lambda^*}{\Lambda^*} \right) \sim F(p, \sum n_i - p - 1).
\]

Bartlett has shown that if \( H_0 \) is true and \( \sum n_i = n \) is large,

\[
\left( n - 1 - \frac{(p + a - 1)\ln \Lambda^*}{2} \right) \sim \chi^2(a; \rho(a - 1))
\]

has approximately a chi-square distribution with \( \rho(a - 1) \) degree of freedom. Consequently, for \( n \) large, we reject \( H_0 \) at significance level \( \alpha \) if

\[
\left( n - 1 - \frac{(p + a - 1)\ln \Lambda^*}{2} \right) \chi^2(a; \rho(a - 1))
\]

5. Example

Suppose an additional variable \( X_0 \) is observed. Arranging the observation pairs \( X_0 \) in rows, the data are

\[
\begin{bmatrix}
8.9, 9.1 & 5.9, 6.1 & 6.9, 7.1 \\
2.9, 3.1 & 1.9, 2.1 & 1.9, 2.1 \\
-0.1, 0.1 & -0.1, 0.1 & -0.1, 0.1 \\
3.9, 4.1 & 0.1, 1.0 & 0.1, 1.0 \\
2.9, 3.1 & 0.1, 0.1 & 0.1, 0.1 \\
7.9, 8.1 & 6.9, 9.1 & 6.9, 7.1
\end{bmatrix}
\]

Thus we have

\[
\begin{bmatrix}
71.17, 84.44 & -17.81, -6.43 \\
-17.81, -6.43 & 43.97, 52.37
\end{bmatrix}
\]

\[
\begin{bmatrix}
7.04, 13.52 & -3.08, 4.2 \\
-3.08, 4.2 & 19.44, 29.12
\end{bmatrix}
\]

and

\[
\Lambda^* = \frac{|W|}{|B + W|} = (0.0149, 0.0009)
\]

Since \( p = 2 \) and \( a = 2 \), an fuzzy test of

\( H_{J,0}: \alpha_1 \cong \alpha_2 \cong 0 \) versus \( H_{J,1} \): at least one

\( \alpha \neq 0 \) is available. To carry out the test, we compare the fuzzy statistic

\[
\left( 1 - \frac{\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) \sum n_i - p - 1 \sim F(a - 1, 4)
\]

with a percentage point of an \( F \)-distribution \( F(0.01; 4, 8) = 7.01 \). Finally, we have the reject degree

\[
\chi^2(a; 1) = 1 - 0.98 \text{ for the hypothesis by agreement index.}
\]

References


