Fuzzy hypotheses testing by $\alpha-$level

Man-Ki Kang[1], Ji-Ypung Jung[1], Woo-Song Park[1],Chang-Eun Lee[1], Gue-Tak Choi[2]
[1]Dept. Data Information Science, Dongeui University
(mkkang@dongeui.ac.kr)
[2] Kyungnam College of Information & Technology

Abstract

We propose some properties of fuzzy p-value and fuzzy significance level to the test statistics for the fuzzy hypotheses testing. Appling the principle of agreement index, we suggest two method for fuzzy hypothesis testing by fuzzy rejection region and fuzzy p-value with fuzzy hypothesis to separately $\alpha-$level.

Key words: critical region, degree of acceptance and rejection, p-value, hypotheses testing, significance level, agreement index.

1. Preliminaries

Our primary purposes of statistical inference is to test hypotheses.

In the traditional approach to hypotheses testing all the concepts are precise and well defined. However if we consider vagueness into observations, we would be faced that hypotheses are quitely new and interesting problems.

Watanabe and Imaizumi(1993) have ideas for fuzzy hypotheses testing methods by based on fuzzy hypothesis membership function with possibility.

Grzegorzewski(2000) proposed fuzzy test for testing statistical hypotheses with vague data and discussed the robustness of the test with possibility.

Kang and Lee(2000, 2002, 2003) defined fuzzy hypotheses membership function and obtained the fuzzy confidence interval from fuzzy test statistic, also they found the agreement index by area for fuzzy hypotheses membership function and membership function of fuzzy confidence interval, thus they obtained the results by the grade for judgement to acceptance or rejection for the fuzzy hypotheses

Now, we propose some properties of fuzzy hypotheses testing by critical region and p-value with agreement index for fuzzy number.

We considered the fuzzy hypothesis

$$H_{f,0}: \theta \simeq \psi, \quad \theta \in \Theta$$

constructed by a set

$$\{(H_0(\psi), H_1(\psi)) | \psi \in \Theta\}$$

with membership function $m_{H_f}(\psi)$ where Θ is parameter space.

A fuzzy number A in $\mathcal{R}$ is said to be convex if for any real numbers $x, y, z \in \mathcal{R}$ with $x \le y \le z$,

$$m_A(y) \ge m_A(x) \wedge m_A(z) \qquad (1.1)$$

with $\wedge$ standing for minimum.

A fuzzy number A is called normal if the following holds

$$\bigvee_x m_A(x) = 1. \qquad (1.2)$$

2. Fuzzy test statistics

In this section, we introduce the computation of fuzzy numbers fuzzy statistics from fuzzy sample.

Let A and B be fuzzy numbers in $\mathcal{R}$ and let $\odot$ be a binary operation defined in $\mathcal{R}$. Then the operation $\odot$ can be extended to the fuzzy numbers A and B by defining the relation(the extension principle).

Let $A, B \subset \mathcal{R}$, $\forall x, y, z \in \mathcal{R}$:

$$m_{A_r \odot B_r}(z) = \bigvee_{z = x \cdot y} (m_{A_r}(x) \wedge m_{B_r}(y)). \qquad (2.1)$$

A modelling the fuzziness of data were described the fuzziness of a sample $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n)$. As a precise sample x of n precise realizations $x_i \in \mathcal{R}$ may be regarded as vector in $\mathcal{R}^n$, a fuzzy sample $\tilde{x}$ of n fuzzy realizations $\tilde{x}_i \in \mathcal{F}(\mathcal{R})$ may be regarded as a fuzzy vector.

A piecewise continuous function $m_{\tilde{x}}: \mathcal{R} \to [0,1]$, fuzzy number $\tilde{x}$ in $\mathcal{F}(\mathcal{R})$ if the family

$$[\tilde{x}]^{\gamma} = \{x \in \mathcal{R},\ m_{\tilde{x}}(x) \ge \gamma\}, \quad \forall \gamma \in (0,1]$$

has the following properties:

$[\tilde{x}]^{\gamma}$ is not empty for $\gamma = 1$, or closed, finite interval $[[\tilde{x}]^{\gamma}_l, [\tilde{x}]^{\gamma}_r]$ in $\mathcal{R}$ and simple connected, compact in $\mathcal{R}^n$.

Assuming that a precise sample x is given, the sample mean $\bar{x}$ is the image of the sample x under the function $f(x)$ given by $f(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Definition 2.1. Let $\tilde{x} \in \mathcal{F}(\mathcal{R}^n)$ be a fuzzy sample with $m_{\tilde{x}}(\cdot)$ and γ-cut representation $\{[\tilde{x}]^{\gamma}: \gamma \in (0,1]\}$, and let $f: \mathcal{R}^n \to Y$, $Y \subseteq \mathcal{R}$, be a real valued continuous mapping. The fuzzy image $\tilde{u} = f(\tilde{x})$ of the fuzzy number data $\tilde{x}$ under the mapping $f(\cdot)$ is defined by the following characterizing function :

$$m_{\tilde{u}}(u) = \sup_{x \in X_u} m_{\tilde{x}}(x),$$

$$X_u = \{x \in \mathcal{R}^n : f(x) = u\},\ \forall u \in \mathcal{R}. \quad (2.2)$$

with $\sup_{x \in X} m_{\tilde{x}}(x) := 0$, if X_u is empty.

Definition 2.2. From the fuzzy sample, the sample mean is a fuzzy number $\tilde{\bar{x}} \in \mathcal{F}(\mathcal{R})$ with characterizing function $m_{\tilde{\bar{x}}}(\cdot)$ given by

$$m_{\tilde{\bar{x}}}(y') = \sup_{\boldsymbol{x} \in X_y} m_{\tilde{\bar{x}}}(\boldsymbol{x}),$$

$$X_y = \left\{ \boldsymbol{x} \in \mathcal{R}^n : \frac{1}{n}\sum_{i=1}^{n} x_i = y \right\},\ \forall y \in \mathcal{R} \qquad (2.3)$$

The γ-cut representation is given by

$$[\tilde{\bar{x}}]^{\gamma} = \left[\min_{\boldsymbol{x} \in [\tilde{x}]^{\gamma}} f(\boldsymbol{x}),\ \max_{\boldsymbol{x} \in [\tilde{x}]^{\gamma}} f(\boldsymbol{x}) \right], \quad \forall \gamma \in (0,1] \qquad (2.4)$$

where $f(\boldsymbol{x}) = g(x_1, x_2, \cdots, x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i)$.

Definition 2.3. If $\tilde{\boldsymbol{x}}$ in Definition 2.1 is a minimum rule of fuzzy sample, the γ-cut representation of fuzzy sample mean is given by

$$[\tilde{\bar{x}}]^{\gamma} = \left[\frac{1}{n}\sum_{i=1}^{n}[\tilde{x}_i]^{\gamma}_l,\ \frac{1}{n}\sum_{i=1}^{n}[\tilde{x}_i]^{\gamma}_r \right],\ \forall \gamma \in (0,1] \qquad (2.5)$$

$[\tilde{x}_i]^{\gamma} = [[\tilde{x}_i]^{\gamma}_l, [\tilde{x}_i]^{\gamma}_r]$ is the γ-cut of the fuzzy data point $\tilde{x}_i$.

Defiinition 2.4. The sample variance and the sample standard deviation are fuzzy numbers $\mathfrak{S}^2 \in \mathcal{F}(\mathcal{R})$ and $\mathfrak{S} \in \mathcal{F}(\mathcal{R})$ with $\Gamma(\mathfrak{S}^2) \subseteq \mathcal{R}^+$ and $\Gamma(\mathfrak{S}) \subseteq \mathcal{R}^+$.

An approximate solution is based on the introduction of a *standardized fuzzy sample* $\tilde{z} = (\tilde{z}_1, \tilde{z}_2, \cdots, \tilde{z}_n)$ define by $\tilde{z}_i = \tilde{x}_i - \bar{\tilde{x}}$.

The γ-cut of each $\tilde{z}_i$ is given by

$$[\tilde{z}_i]^\gamma = [[\tilde{z}_i]_l^\gamma, [\tilde{z}_i]_r^\gamma] = [[\tilde{x}_i]_l^\gamma - [\bar{\tilde{x}}]_r^\gamma, [\tilde{x}_i]_r^\gamma - [\bar{\tilde{x}}]_l^\gamma] \quad (2.6)$$

Now renumber that the sample variance s^2 of sample x is related to the second empirical moment of the standardized sample by $z = (x_1 - \bar{x}, x_2 - \bar{x}, \cdots, x_n - \bar{x})$ by

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} z_i^2 \ . \quad (2.7)$$

A minimization rule fuzzy sample this approximation results in the following γ-cut representation of fuzzy sample variance and the fuzzy sample standard deviation, respectively:

$$[\tilde{s}^2]^\gamma = \left[\frac{1}{n-1} \sum_{i=1}^{n} [\tilde{z}_i^2]_l^\gamma, \frac{1}{n-1} \sum_{i=1}^{n} [\tilde{z}_i^2]_r^\gamma \right], \quad \forall \gamma \in (0,1] \quad (2..8)$$

$$[\tilde{s}]^\gamma = \left[\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} [\tilde{z}_i^2]_l^\gamma}, \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} [\tilde{z}_i^2]_r^\gamma} \right], \quad \forall \gamma \in (0,1] \quad (2.9)$$

$$[z_i^2]_l^\gamma = \begin{cases} ([\tilde{z}_i]_l^\gamma)^2, & [\tilde{z}_i]_l^\gamma \geq 0 \\ ([\tilde{z}_i]_r^\gamma)^2, & [\tilde{z}_i]_r^\gamma \leq 0 \\ 0, & oterwise \end{cases} \quad (2.10)$$

$$[z_i^2]_r^\gamma = \begin{cases} ([\tilde{z}_i]_r^\gamma)^2, & [\tilde{z}_i]_l^\gamma + [\tilde{z}_i]_r^\gamma \geq 0 \\ ([\tilde{z}_i]_l^\gamma)^2, & oterwise\ . \end{cases} \quad (2.11)$$

Thus, $[\tilde{z}_i]^\gamma = [[\tilde{z}_i]_l^\gamma, [\tilde{z}_i]_r^\gamma]$ are the γ-cuts of the components of the standardized fuzzy sample $\tilde{z}$.

3. Fuzzy hypothesis test

Let x be a random sample from sample space Ω. Let $\{P_\theta, \theta \in \Theta\}$ be a family of fuzzy probability distribution, where θ is a parameter vector and Θ is a parameter space.

The fuzzy hypothesis H_f can be interpreted as a hypothesis '$\theta \simeq \psi$', which is obtained by adding fuzziness to an ordinary null hypothesis H_0: $\theta = \psi_0$, we note that $H_{f.0}$: $\theta \simeq \psi_0$ for such a fuzzy hypothesis.

Definition 3.1. We define agreement index by real-valued function R_γ on Θ as the maximum grade membership function of acceptance or rejection is

$$m_{R_\gamma}(0) = \sup_\psi \{ area\ (m_{H_\gamma}(\psi) \cap m_{T_\gamma}(\psi)) / area\ m_{H_\gamma}(\psi) \} \quad (3.1)$$

$$m_{R_\gamma}(1) = 1 - m_{R_\gamma}(0) \quad (3.2)$$

for the fuzzy hypothesis testing.

4. Example

For example, a city health dependent wishes to determine if the mean bacteria count per unit volume of water at a lake beach exceeds the safety level of about 200. Researchers have collected 10 water samples of unit volume and have found the bacteria count to be fuzzy sample as:

[174,176],[188,192],[213, 217],[196,200], [183,185],[205,209],[207, 213],[191,195], [194,198], [179,181]

Do the data indicate a cause for concern?

Considering the failure to detect a violation of the safety level can result in serious consequence, formulation of the null hypothesis should be

$$H_{f.0}: \mu \simeq 202$$

Since the counts are spread over a wide range, an approximation by a continuous distribution is not unrealistic for inference about the mean. Assuming that the measurements consistence a fuzzy sample from a normal population and that the level of significance $\alpha = 0.1$. We employ the t-test

$$\tilde{t} = \frac{(\tilde{X} - 202)}{\tilde{s}/\sqrt{10}} \quad (4.1)$$

with rejection region $t \leq -t_{0.1/2}$ or $t \geq t_{0.1/2}$. From the t-table we determine that $t_{0.05} = 1.833$ with $d.f. = 9$.

Computations from the sample data yield

$\tilde{X} = [193.0 + 1.8\gamma, 196.6 - 1.9\gamma]$

$\tilde{s}^2 = [106.311 + 51.511\gamma + 14.8\gamma^2,$

$$268.533 - 110.711\gamma + 14.8\gamma^2]$$

If we have $\gamma = 1$,

$$\tilde{t} = \frac{|\bar{X} - 202|}{\tilde{s}/\sqrt{10}}$$

$$= [1.733, 1.733] > t(0.05; 9) = 1.833$$

thus we accept the alternative hypothesis $H_{f.0}$.

If we have $\gamma = 0.0$,

$$\tilde{t} = \frac{|\bar{X} - [201 + \gamma, 203 - \gamma]|}{\tilde{s}/\sqrt{10}} = [0.849, 3.067]$$

and $t(0.05; 9) = 1.833$,

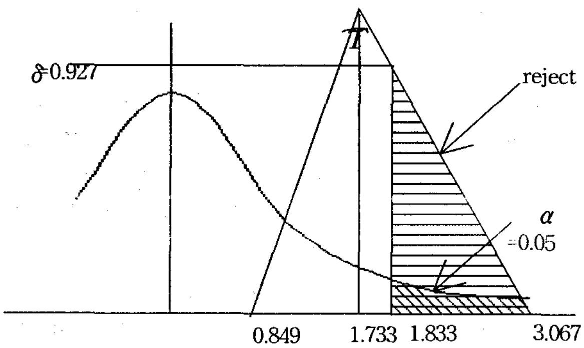
By α-level, if we have $\alpha_0 = 0$

$$\int_0^{0.927} (-1.333\alpha + 3.067 - 1.833) d\alpha /$$

$$\int_0^1 ((-1.333\alpha + 3.067) - (0.884\alpha + 0.849)) d\alpha$$

$$= 0.516. \qquad (4.2)$$

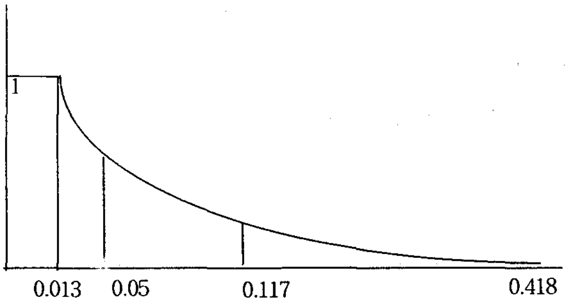
Thus, we accept the alternative hypothesis $H_{f.0}$ degree of $m_{\tilde{R}_\gamma}(0) = 0.516$ by [Definition 3.1] from [Figure 4.1],



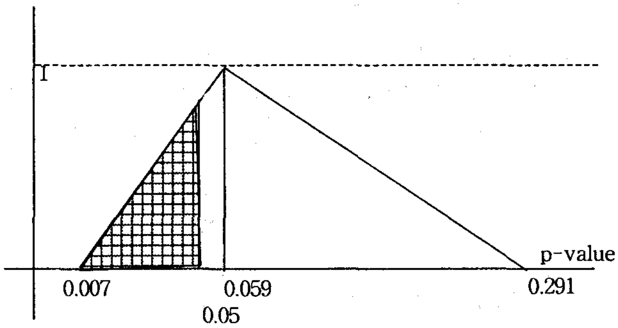
[Figure 4.1]

and we can show that the membership function of rejection degree by p-value to [Figure 4.2].

Another hand, by degree of p-value, we also can accept the alternative hypothesis $H_{f.0}$ degree of $m_{\tilde{R}_\gamma}(0) = 0.163$ by [Definition 3.1] from [Figure 4. 3] as the (4.2). Thus , if we can compare the properties of the fuzzy test statistics with distribution function, the test by p-value is excellent testing statistical hypothesis rather than t-test statistics.



[Figure 4.2]



[Figure 4.3]

References

[1] P. X. Gizegorzewski, Testing Hypotheses with vague data, *Fuzzy Sets and Systems.* 112 , pp.501-510, 2000.

[2] M. K. Kang, G. T. Choi and C. E. Lee, On Statistical for Fuzzy Hypotheses with Fuzzy Data, *Proceeding of Korea Fuzzy Logic and Intelligent System Society Fall Conference*, Vol. 10, Num. 2, 2000.

[3] M. K. Kang, G. T. Choi and S. I .Han, A Bayesian Fuzzy Hypotheses testing with Loss Function, *Proceeding of Korea Fuzzy Logic and Intelligent Systems Society Fall Conference,* Vol 13, Num. 2, 2002.

[4] M. K. Kang, Lee, C. E. and Han, S. I. . Fuzzy Hypotheses Testing for Hybrid Numbers by Agreement Index, *Far East Journal of Theoretical Statistics.* ,10(1), 2003.

[5] S. F. Schnatter, S. On Statistical inference for Fuzzy Data whit Application to Des Crispy Statistics, *Fuzzy Sets and Systems,* 50, pp 143-165, 1992.

[6] N. Watanabe, T. Imaizumi, A Fuzzy Statistical Test of Fuzzy Hypotheses, *Fuzzy Sets and systems* 53 , pp.167-178, 1993.