한국지능시스템학회 논문지 2011, Vol. 21, No. 1, pp. 1-5
DOI : 10.5391/JKIIS.2011.21.1.1

2변량 정규분포의 독립성에 관한 퍼지 검정

Fuzzy Testing of Independence in Bivariate Normal Distribution

강만기* · 서현아**
Man Ki Kang and Hyun A Seo

* 동의대학교 자연과학대학 데이터정보학과
** 동의대학교 대학원 정보통계학과

요 약

우리는 동의지수법에 의한 이변량정규분포의 변수 간 독립성에 대한 퍼지 상관 검정법을 제안하였다. 이를 위하여 퍼지 데이터의 조건을 제시하고, 퍼지 상관계수와 가설을 기각하고 채택할 정도를 위한 동의지수법을 정의하였다. 또한 예증을 위하여 한 이변량정규분포로부터 퍼지 표본을 추출하여 퍼지 상관계수를 이용한 최강력 불편 퍼지 검정법을 보였다.

키워드 : 애매한 자료, 기각과 채택정도, 퍼지가설검정, 동의지수.

Abstract

We furnished some properties of fuzzy testing of independence for correlation in a bivariate normal distribution by agreement index. First we present some restriction of the fuzzy data, define fuzzy sample correlation coefficient and agreement index for testing hypothesis with acceptance or rejection degree. Also, we show that UMP unbiased fuzzy test and drawing conclusions the fuzzy test.

Key Words : vague data, degree of acceptance and rejection, fuzzy hypotheses testing, agreement index.

1. Introduction

We propose some properties of fuzzy testing of independence in a bivariate normal distribution by agreement index obtained from the fuzzy random samples. The negation of the assertion is taken to be fuzzy null hypothesis H_{f0} and the assertion itself is taken to be the fuzzy alternative hypothesis H_{f1}([1],[6]).

Kang, Lee and Han[2] defined membership function of fuzzy hypotheses, also they found the agreement index by area ratio for fuzzy hypotheses membership function with regard to membership function of fuzzy critical region, thus they obtained the results by the grade for judgement to acceptance or rejection for the fuzzy hypotheses with vague data.

Kang and Seo[4] suggested various type of agreement index of fuzzy hypotheses regard to fuzzy critical region.

First we define the restriction of fuzzy data and fuzzy statistics for repeatedly observed data with alteration error terms. In chapter 3, we introduce some properties of fuzzy sample correlation coefficient with fuzzy data from the bivariate normal distribution.

The various type of fuzzy hypotheses regard to fuzzy critical region for agreement index was shown in chapter 4.

Finally, we have joint fuzzy bivariate normal probability density function and illustrate independence of the bivariate normal distribution by uniformly most powerful unbiased fuzzy test from fuzzy correlation coefficient.

2. Preliminaries

A fuzzy number data A in the real line R is a fuzzy set of characterized by a membership function m_A as $m_A : R \to [0,1]$. A fuzzy number data A is expressed as $A = \int_{x \in R} m_A(x)/x$, with understanding that $m_A(x) \in [0,1]$ represents the grade of membership of x in A.

A fuzzy number A in R is said to be convex if for any real numbers $x, y, z \in R$ with $x \leqq y \leqq z$, $m_A(y) \geqq m_A(x) \wedge m_A(z)$ with $\wedge$ standing for min.

A fuzzy number A is called normal if $\max_x m_A(x) = 1$.

A δ-level set of a fuzzy number data A is denote by $[A]^\delta$ and define the fuzzy number by $[A]^\delta = \{ x \mid m_A(x) \geqq \delta, 0 < \delta < 1 \}$ where δ is precision of data A. Also, δ-level set of a fuzzy number A is a

접수일자 : 2010년 10월 14일
완료일자 : 2011년 11월 20일

convex fuzzy set which is a closed , bounded interval and denote by $[A]^\delta = [A_l^\delta, A_r^\delta]$.

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

$$A \odot B = \int_{x,y \in R} (m_A(x) \wedge m_B(y))/(x \cdot y). \quad (2.1)$$

Let random sample $X_1, X_2, \cdots, X_n$ be convex fuzzy number with membership function $m_{X_1}, m_{X_2}, \cdots, m_{X_n}$, respectively. Then the fuzzy sample mean $\overline{X}$ is defined by (2.1) and denoted by

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 \oplus X_2 \oplus \cdots \oplus X_n). \quad (2.2)$$

Also, $\overline{X}$ is a convex fuzzy number. By the Zadeh's extension principle, we obtain the membership function of fuzzy sample mean as follows :

$$\begin{aligned} m_{\overline{X}} &= m_{z = \frac{1}{n}\Sigma_{i=1}^n X_i}(z) \\ &= m_{z = \frac{1}{n}(X_1 \oplus X_2 \oplus \cdots \oplus X_n)}(z) \\ &= \sup_{z = x_1 + x_2 + \cdots + x_n} min\{m_{X_1}(x_1), \\ &\quad m_{X_2}(x_2), \cdots, m_{X_n}(x_n)\}. \end{aligned} \quad (2.3)$$

Let $[A_i]^\delta = [A_{il}^\delta, A_{ir}^\delta]$. then, for $\delta \in (0,1]$, the δ-level sets of fuzzy sample mean is

$$\begin{aligned} [\overline{X}]^\delta &= \left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]^\delta = \left[\frac{1}{n}\sum_{i=1}^{n} X_{il}^\delta, \frac{1}{n}\sum_{i=1}^{n} X_{ir}^\delta\right] \\ &= \left[\overline{X_l^\delta}, \overline{X_r^\delta}\right]. \end{aligned} \quad (2.4)$$

3. Fuzzy sample correlation coefficient

We study some properties of fuzzy sample correlation coefficient with fuzzy data.

Let X and Y be two fuzzy random variables, $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ be their random sample, X_i and Y_i, for $i = 1, \cdots, n$, be convex fuzzy numbers.

Let $\overline{X}$ and $\overline{Y}$ be fuzzy sample means of X_i and Y_i for $i = 1, \cdots, n$, and $\overline{X}$ and $\overline{Y}$ be convex fuzzy numbers. Then we define fuzzy sample correlation co-efficient by the same methods of the fuzzy sample mean as follows;

$$R_{XY} = \frac{\sum_{i=1}^{n}\{(X_i \ominus \overline{X}) \otimes (Y_i \ominus \overline{Y})\}}{\sqrt{\sum_{i=1}^{n}(X_i \ominus \overline{X})^2 \otimes \sum_{i=1}^{n}(Y_i \ominus \overline{Y})^2}}. \quad (3.1)$$

Also, R_{XY} is a convex fuzzy number.

Let X_i and Y_i, for $i = 1, \cdots, n$, be convex fuzzy numbers, then we define sufficient statistics as

$$\frac{1}{n-1}\sum_{i=1}^{n}\{(X_i \ominus \overline{X}) \otimes (Y_i \ominus \overline{Y})\} = S_{XY},$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i \ominus \overline{X})^2 = S_{XX},$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(Y_i \ominus \overline{Y})^2 = S_{YY}. \quad (3.2)$$

Let $[X_i]^\delta = [X_{il}^\delta, X_{ir}^\delta]$, $[Y_i]^\delta = [Y_{il}^\delta, Y_{ir}^\delta]$, $[\overline{X}]^\delta = [\overline{X_l^\delta}, \overline{X_r^\delta}]$, $[\overline{Y}]^\delta = [\overline{Y_l^\delta}, \overline{Y_r^\delta}]$

then, for $i = 1, \cdots, n$, we obtain the δ-level sets of fuzzy sample correlation coefficient as follows ;

$$\begin{aligned} [R_{XY}]^\delta &= \left[\frac{S_{XY}}{\sqrt{S_{XX} \otimes S_{YY}}}\right]^\delta \\ &= \frac{\sum_{i=1}^{n}\left(\left[X_{il}^\delta - \overline{X_r^\delta}, \widetilde{X_{ir}^\delta} - \overline{X_l^\delta}\right]\right.}{\sqrt{\sum_{i=1}^{n}\left(\left[X_{il}^\delta - \overline{X_r^\delta}, X_{ir}^\delta - \overline{X_l^\delta}\right]\right)^2}} \\ &\cdot \frac{\left.\left[Y_{il}^\delta - \overline{Y_r^\delta}, Y_{ir}^\delta - \overline{Y_l^\delta}\right]\right)}{\sqrt{\sum_{i=1}^{n}\left(\left[Y_{il}^\delta - \overline{Y_r^\delta}, Y_{ir}^\delta - \overline{Y_l^\delta}\right]\right)^2}}. \end{aligned} \quad (3.3)$$

4. Acceptance or rejection degree

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

Choose a membership function $m_X(x)$ whose value is likely to best reflect the plausibility of the fuzzy hy-pothesis being tested.

Let us consider membership function $m_C(x)$ of crit-ical region C, which we will call the agreement index of $m_X(x)$ which regard to $m_C(x)$ ([3],[5]).

Definition 4.1. Let a fuzzy membership function $m_X(x)$, $x \in R$, we consider another membership func-tion $m_C(x)$, $x \in R$, which call the agreement index by the area ratio being defined in the following way;

$$AGI(X, C) = \frac{area(m_X(x) \cap m_C(x))}{area(m_C(x))} \in [0,1]. \quad (4.1)$$

Definition 4.2. We define the grade membership func-tion of rejection or acceptance degree by agreement in-dex for real-valued function R_δ by δ-level on Ω as

$$\Re_\delta(0) = \sup_\theta\left\{\frac{\mathrm{area}\ (\mathrm{m}_{\mathrm{X}_\delta}(\theta) \cap \mathrm{m}_{\mathrm{C}_\delta}(\theta))}{\mathrm{area}\ (\mathrm{m}_{\mathrm{C}_\delta}(\theta))}\right\}, \quad (4.2)$$

$$\Re_{\delta}(1) = 1 - \Re_{\delta}(0) \tag{4.3}$$

for the fuzzy hypothesis testing.

In agreement index, we have the area by $\delta-$level as:

$$area(m_C(x) \cap m_X(x)) = \int_{\delta_0}^{\delta_1} (C_r^{-1}(\delta) - X_l^{-1}(\delta))d\delta$$

$$area\ m_{C_\delta}(\theta) = \int_{\delta_0}^{1} (C_r^{-1}(\delta) - C_l^{-1}(\delta))d\delta \tag{4.4}$$

where C_r, C_l are right and left side line of $m_C(x)$, X_l is left side line of $m_X(x)$ and δ_0 is reliable degree and δ_1 is meeting point of $m_C(x)$ and $m_X(x)$.

Definition 4.3. We have acceptance region and rejection region for the fuzzy critical region m_C as Fig 4.1.

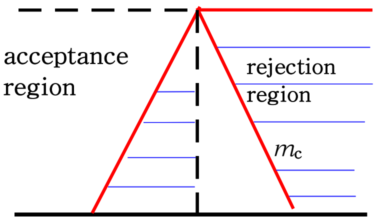
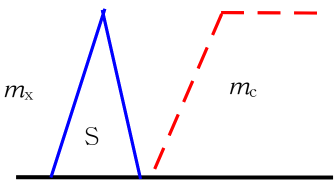
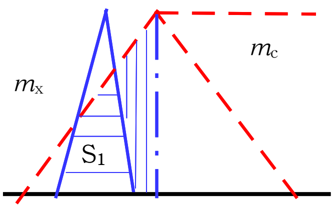


Fig 4.1 Acceptance and rejection region

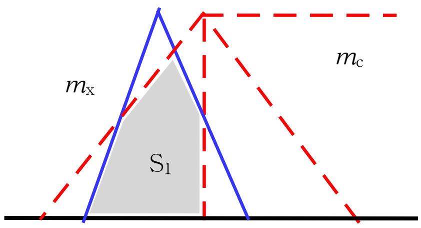
For various kinds of m_X, we can reject the hypotheses by Definition 4.2 as [Fig 4.2].



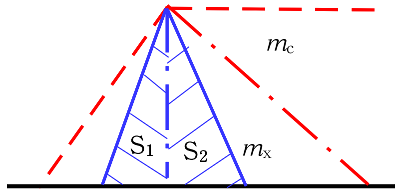
(a)



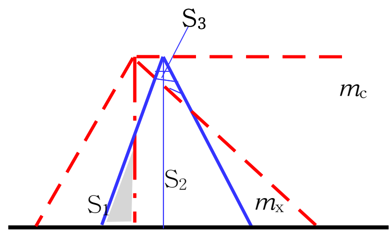
(b)



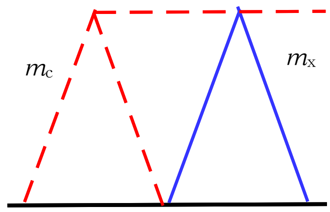
(c)



(d)



(e)



(f)

Fig 4.2 Various type of m_C by m_X

For example, in [Fig 42]-(a) and [Fig 4.2]-(f), we can clearly reject the hypothesis as degree $\Re_{\delta}(0) = 0$ and $\Re_{\delta}(0) = 1$.

For [Fig 4.2]-(b), (c), we have

$$\Re_{\delta}(0) = \frac{area(S_1)}{area(S)} \times \frac{1}{2} \tag{4.5}$$

for left hand side area of center of fuzzy critical region m_C by fuzzy the statistics m_X.

In case of Fig 4.2 (d), we nave

$$\Re_{\delta}(0) = \frac{area(S_1)}{area(S_1)} \times \frac{1}{2} = 0.5 \tag{4.6}$$

it's maintain an uncertain attitude for decision the hypotheses.

Also, we have

$$\Re_{\delta}(0) = \frac{area(S_1)}{area(S_1)} \times \frac{1}{2} + \frac{area(S_3)}{area(S_2)} \times \frac{1}{2} \tag{4.7}$$

in [Fig 4.2] (e).

5. Fuzzy bivariate normal distribution

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ are fuzzy random sample from a bivariate normal probability distribution then we have joint fuzzy bivariate normal probability density function as

$$f(x,y)=\frac{1}{(2\pi\sigma_x\sigma_y\sqrt{1-\rho^2})^n}exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{1}{\sigma_x^2}\sum(x_i-\mu_x)^2-\frac{2\rho}{\sigma_x\sigma_y}\sum(x_i-\mu_x)(y_i-\mu_y)+\frac{1}{\sigma_y^2}\sum(y_i-\mu_2)^2\right)\right] \quad (5.1)$$

Here (μ_x,σ_x^2) and (μ_y,σ_y^2) are the fuzzy mean and fuzzy variance of fuzzy random variable X and Y respectively, and ρ is the fuzzy population correlation coefficient between X and Y of the bivariate normal probability distribution.

We shall consider a fuzzy hypothesis $\rho \simeq 0$ that X and Y are independent, and the corresponding one-sided fuzzy hypothesis $\rho < 0$.

The fuzzy exponential form and the sufficient statistics of the family of densities (5.1) are

$$T_0=\Sigma X_i Y_i,\ T_1=\Sigma X_i^2,\ T_2=\Sigma Y_i^2,\ T_3=\Sigma X_i,\ T_4=\Sigma Y_i \quad (5.2)$$

and

$$\theta_0=\frac{\rho}{\sigma_x\sigma_y(1-\rho^2)},\ \vartheta_1=-\frac{1}{2\sigma_x^2(1-\rho^2)},\ \vartheta_2=-\frac{1}{2\sigma_y^2(1-\rho^2)},\ \vartheta_3=\frac{1}{1-\rho^2}\left(\frac{\mu_x}{\sigma_x^2}-\frac{\mu_y\rho}{\sigma_x\sigma_y}\right),\ \vartheta_4=\frac{1}{1-\rho^2}\left(\frac{\mu_y}{\sigma_y^2}-\frac{\mu_x\rho}{\sigma_x\sigma_y}\right). \quad (5.3)$$

The fuzzy hypothesis $H_{f0}:\rho < 0$ is equivalent to $\theta_0 < 0$. Since the fuzzy sample correlation coefficient of (3.1) is

$$R=\frac{\Sigma(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\Sigma(X_i-\overline{X})^2\Sigma(Y_i-\overline{Y})^2}}.$$

Thus R is unchanged when the X_i and Y_i are replaced by $(X_i-\mu_x)/\sigma_x$ and $(Y_i-\mu_y)/\sigma_y$, the distribution of R does not depend on μ_x, μ_y, σ_x or σ_y, but only on ρ . For $\theta_0 \simeq 0$ it therefore does not depend on $\vartheta_1,\cdots,\vartheta_4$, R is independent of $(T_1, \cdots, T_4)$ when $\theta_0 \simeq 0$. It follow from the UMP(uniformly most powerful) unbiased fuzzy test of H_{f0} rejects when

$$R > C_0 \quad (5.4)$$

or equivalently when

$$\frac{R}{\sqrt{(1-R^2)/(n-2)}} > K_0 \quad (5.5)$$

where C_0 and K_0 are any constants.

The fuzzy statistic R is linear in T_0, and its distribution for $\rho \simeq 0$ of symmetric about 0. The UMP unbiased fuzzy test of the fuzzy hypothesis $\rho \simeq 0$ against the alternatives not ρ is 0, therefore rejects when

$$\frac{|R|}{\sqrt{(1-R^2)/(n-2)}} > K_1 \quad (5.6)$$

for any constant K_1.

6. Illustration

Since $\sqrt{n-2}\ R/\sqrt{1-R^2}$ has the $t-$distribution with $n-2$ degree of freedom when $\rho \simeq 0$, the constants K_0 and K_1 in the above fuzzy tests are given by

$$\int_{K_0}^{\infty} t_{n-2}(x)\,dx=\alpha,\ \int_{K_1}^{\infty} t_{n-2}(x)\,dx=\frac{\alpha}{2}. \quad (6.1)$$

Since the distribution R depends only on the fuzzy correlation coefficient ρ, the same is true of the power of these fuzzy tests.

From a bivariate normal probability distribution, if we have artificial random fuzzy number sample data as;

X	Y
[1.95, 2, 2.05]	[3.95, 4, 4.05]
[2.65, 3, 3.35]	[7.65, 8, 8.35]
[3.85, 4, 4.15]	[5.85, 6, 6.15]
[4.95, 5, 5.05]	[6.95, 7, 7.05]
[5.75, 6, 6.25]	[9.75, 10, 10.25]

then $R=[0.730, 0.778, 0.830]$ by (3.1) and $\frac{R}{\sqrt{(1-R^2)/(n-2)}}=[1.848, 2.144, 2.537]$.

For $H_{f0}:\rho > 0$, if we have fuzzy significance level α =[0.075, 0.100, 0.125] then K_0=[2.113, 2.353, 2.680] from (5.7) by $t-$distribution. So, [1.848, 2.144, 2.537]$> K_0$ is the case of [Fig 4.2]-(c).

Finally, we have rejection degree $\Re_\delta(0)=0.138$ by (4.2).

For $H_{f0}:\rho \simeq 0$, if we have another artificial random fuzzy number sample data as;

X	Y
[1.9, 2, 2.1]	[5.9, 6, 6.1]
[3.8, 4, 4.2]	[8.8, 9, 9.2]
[5.2, 6, 6.8]	[7.2, 8, 8.8]
[7.2, 8, 8.8]	[9.2, 10, 10.8]
[9.1, 10, 10.9]	[10.1, 11, 11.9]

then $\frac{|R|}{\sqrt{(1-R^2)/(n-2)}}=[2.80, 3.667, 4.600]$ and $K_1=[2.680, 3.182, 4.176]$ by (5.7).

Since $[2.80, 3.667, 4.600] > K_1$ is the case of [Fig 4.2]-(e), we have $\Re_\delta(0)=0.877$ by (4.2).

References

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

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

[3] M. K. Kang and Y. R. Park, Fuzzy Binomial Proportion Test by Agreement Index, *Journal of Korean Institute of Intelligent Systems,* Vol 19, Num. 1,pp.19-24, 2009.

[4] M. K. Kang and A. H. Seo, Fuzzy Hypothesis Test by Poisson Test for Most powerful Test, *Journal of Korean Institute of Intelligent Systems,* Vol 19, Num. 6, pp.809-813, 2009.

[5] W. Trutschnig, A strong consistency result for fuzzy relative frequencies interpreted as estimator for the fuzzy-value probability, *Fuzzy Sets and Systems,* 159, pp 259-269, 2008.

[6] Z. Xia, Fuzzy Probability System: fuzzy probability space(1), *Fuzzy Sets and systems,* 120, pp.469-486, 2001.

저 자 소 개

강만기(Man Ki Kang)

한국지능시스템학회 논문지
Vol.15 No.2, Vol.17, No.5 참조

서현아(Hyun A Seo)
현재 : 동의대학교 대학원
정보통계학과 박사과정
관심분야 : 퍼지데이터처리