한국지능시스템학회 논문지 2009, Vol. 19, No. 1, pp. 19-24

동의지수에 의한 퍼지 이항비률 검정

Fuzzy Binomial Proportion Test by Agreement Index

강만기* · 박영례**
Man-Ki Kang* and Young-Rye Park**

* 동의대학교 자연과학대학 데이터정보학과
** 동의대학교 대학원 정보통계학과

요 약

반복적으로 관측된 데이터가 애매한 경우에 관측자료를 퍼지화하여 이에 대한 확률을 정의하고, 퍼지 확률공간에 의한 제I 일종의 오류와 제II종의 오류를 보이며, 동의지수 방법으로 퍼지 이항비률 검정법을 제안하고 예증한다.

Abstract

We propose some properties for fuzzy binomial proportion test by agreement index. First we define fuzzy probability space and fuzzy type I error and type II error for the fuzzy probability of the two type errors. Also, we show that a fuzzy power function of performance for a fuzzy hypothesis test and drawing conclusions from the test.

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

1. Preliminaries

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

We obtained from the fuzzy samples, the negation of the assertion is taken to be the fuzzy null hypothesis H_0 and the assertion itself is taken to be the fuzzy alternative hypothesis H_1.

In testing a null fuzzy hypothesis H_0 against an alternative fuzzy hypothesis H_1, our attitude is to uphold H_0 as true degree unless the data speak strongly against it, in which case, H_0 should be rejected in favor of H_1 by degree of acceptance and rejection([1],[5]).

Kang, Choi and Lee[2] and defined fuzzy hypotheses membership function also they found the agreement index by area for fuzzy hypotheses membership function and membership function of fuzzy critical region. Also Kang, Choi and Han[3] obtained the results by the grade for judgement to acceptance or rejection for the fuzzy hypotheses.

Thus, we introduction some properties of fuzzy binomial proportion test by agreement index. First we define fuzzy probability space and fuzzy type I error and type II error for the fuzzy probability of the two types of errors. Also, we show that a fuzzy power function of performance of a fuzzy test and drawing conclusions from a test.

We considered the fuzzy hypothesis

$$H_0: \theta \simeq \psi \ \text{ or } \ H_0: \theta < \psi,\ \theta \in \Theta \tag{1.1}$$

constructed by a set

$$\{(H_0(\psi), H_1(\psi)) | \psi \in \Theta\} \tag{1.2}$$

with membership function $m_H(\psi)$ where Θ is parameter space.

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

$$m_A(y) \geqq m_A(x) \wedge m_A(z) \tag{1.3}$$

with $\wedge$ standing for minimum.

A fuzzy number A is called normal if the following holds

$$\bigvee_x m_A(x) = 1 \tag{1.4}$$

An δ-level set of a fuzzy number A is a set denoted by $[A]^\delta$ and is defined by

$$[A]^\delta = \{x | m_A(x) \geqq \delta,\ 0 < \delta \leqq 1\} \tag{1.5}$$

An δ-level set of fuzzy number A is a convex fuzzy set which is a closed bounded interval denoted by $[A]^\delta = [A_l^\delta, A_r^\delta]$.

접수일자 : 2008년 9월 5일
완료일자 : 2009년 2월 10일

2. Fuzzy probability

The concept of probability is relevant to experiments that have same what uncertain outcoms. Thus uncertain outcom is fuzzy concepts.

We denote by $x=x(s)$ the possible outcome of an fuzzy random experiment of subject s and will be call the fuzzy sample point.

Definition 2.1. The sample of an fuzzy random experiment of subject s is a pair $(\Omega(s),\Phi)$([6]), where
(1) $\Omega(s)$ is the set of all possible outcomes of the fuzzy random experiment of subject s
(2) Φ is a σ-field of subjects of $\Omega(s)$.

In general, call $\Omega(s)$ the valuation set of subject s, which includes all possible observation result or includes the observation range of subject s. Let $S(s)$ be the collection of all subjects of $\Omega(s)$; then it is the σ-field of subjects of $\Omega(s)$. So, $(\Omega(s),S(s))$ is the sample space of subject s.

The first method is that any set A of $S(s)$ is called an event; the second method is that is a statement about the results of an fuzzy random experiment determines an event.

Let Φ be the σ-field that is generated by such an event family, any set $A\in\Phi$ is called an event.

We must note that now we do not have any probability measure on sample space (Ω,S), next let us discuss how one can construct a probability on (Ω,S) by the sample of fuzzy random experiment, satisfying the Kolmogorov axiom.

The fuzzy random experiment can be repeated under identical fuzzy condition by one and the same subject s, for $i=1,2,\cdots$; we denote by $E_k=E_k(s)$ the k-th trial of subject s, and by $x_k=x_k(s)$ the sample point which is a result of E_k; we say that $E(s)=\{E_k(s)\}$.

For each $A\in S$, let $m_A(x)$ be given by

$$0\leqq m_A(x)\leqq 1 \qquad (2.1)$$

That is, it is the membership function of fuzzy events A, and for each $x_k\in\Omega_n$, $m_A(x_k)$ have provided the time information of the occurrence of events A, which has some uncertainty from the randomness of samples, Ω_n. To provide the time information of occurrence of events A under identical conditions, take the weighted mean([4]) of all these $m_A(x_k)$, $1\le k\le n$, which have the same weight, that is, for each $A\in S$, let

$$P(A,n,s)=\frac{1}{n}\sum_{k=1}^{n}m_A(x_k)=\frac{1}{n}\sum_{k=1}^{n}m_A(x_k(s)) \qquad (2.2)$$

be the fuzzy probability of occurrence of fuzzy events A in trials n.

Thus we have following two propositions.

Proposition 2.1 For any fuzzy random experiment, let (Ω,S) be its common sample space. Then in a sequence of fuzzy random experiment of subject $s(\in S)$, the probability $P(A,n,s)$, $n=1,2,\cdots$, has all the properties as follows. For each $A\in S$, and for each $n\ge 1$;
(1) $0\le P(A,n,s)\le 1$ (2.3)
(2) $P(\Omega,n,s)=1$ (2.4)
(3) If $A_i\in S$ and $A_i\cap A_j=\phi$, $i\ne j$; $i,j=1,2,\cdots$, let

$$A=\bigcup_{i=1}^{\infty}A_i,\ \text{then}\ P(A,n,s)=\sum_{i=1}^{\infty}P(A_i,n,s). \qquad (2.5)$$

Proof. For each $1\le k\le n$, if we have fuzzy samples $\omega_k(s)$ then $m_A(x_k(s))$ is fixed a set functions satisfying $0\leq m_A(x_k(s))\leqq 1$ for each $A\in S$ and $m_\Omega(x_k(s))=1$ is clear.

For any sequence $\{A_i\}$ of disjoint sets of S, let $A=\cup_{k=1}^{\infty}A_i$, if $x_k(s)\in A$, $1\le k\le n$ then there exist i such that $x_k(s)\in A_i$. If $x_k(s)$ was checked n_i times in A_i then we have

$$\frac{1}{n}\sum_{k=1}^{n_i}m_{A_i}(x_k(s))=P(A_i,n,s)\ ,\ \text{where}\ \sum_{i=1}^{\infty}n_i=n$$

Thus we have $P(A,n,s)=\sum_{i=1}^{\infty}P(A_i,n,s)$.

It follows $0\le P(A,n,s)\le 1$ for each $A\in S$,

$P(\Omega,n,s)=1$ and $P(A,n,s)=\sum_{i=1}^{\infty}P(A_i,n,s)$.

□

Proposition 2.2. We have that
(1) $P(A^c,n,s)=1-P(A,n,s)$. (2.6)
(2) $P(\phi,n,s)=0$. (2.7)
(3) If $A_1,A_2\in\Phi$ then

$$P(A_1\cup A_2,n,s)$$
$$=P(A_1,n,s)+P(A_2,n,s)-P(A_1\cap A_2,n,s). \qquad (2.8)$$

Proof. Since

$$P(A^c,n,s)=\frac{1}{n}\sum_{k=1}^{n}m_{A^c}(x_k)=\frac{1}{n}\sum_{k=1}^{n}(1-m_A(x_k(s)))$$
$$=\frac{1}{n}\sum_{k=1}^{n}1-\frac{1}{n}\sum_{k=1}^{n}m_A(x_k(s))=1-\frac{1}{n}\sum_{k=1}^{n}m_A(x_k(s)).$$

So we have

$$P(A^c,n,s)=1-P(A,n,s).$$

From $P(A_1\cup A_2,n,s)=\frac{1}{n}\sum_{k=1}^{n}m_{A_1\cup A_2}(x_k)$

$$=\frac{1}{n}\sum_{k=1}^{n}Max\{m_{A_1}(x_k),\ m_{A_2}(x_k)\}.$$

Put

$$S_1=\{x_k\in S|\,m_{A_1}(x_k)\ge m_{A_2}(x_k)\},$$

by n_1 times for $A_1(x_k)$ and

$$S_2=\{x_k \in S \mid m_{A_1}(x_k) < m_{A_2}(x_k)\},$$

by n_2 times for $A_2(x_k)$ and $S=S_1 \cup S_2$, $n_1+n_2=n$. Then

$$P(A_1 \cup A_2, n, s)=\frac{1}{n}\sum_{k=1}^{n_1} m_{A_1}(x_k|S_1)+\frac{1}{n}\sum_{k=1}^{n_2} m_{A_2}(x_k|S_2).$$

Since

$$P(A_1 \cap A_2, n, s)=\frac{1}{n}\sum_{k=1}^{n} m_{A_1 \cap A_2}(x_k)$$
$$=\frac{1}{n}\sum_{k=1}^{n} Min\{m_{A_1}(x_k),\ m_{A_2}(x_k)\}$$
$$=\frac{1}{n}\sum_{k=1}^{n_2} m_{A_1}(x_k|S_2)+\frac{1}{n}\sum_{k=1}^{n_1} m_{A_2}(x_k|S_1)$$

□

Thus we have

$$P(A_1 \cup A_2, n, s)=P(A_1, n, s)+P(A_2, n, s)-P(A_1 \cap A_2, n, s)$$

So we have

1. For set $\Omega(s)$ of subject $s(\in S)$.
2. Any σ-field Φ in $\Omega(s)$.
3. Set function $P(A,n,s)$ is a normal measure on ϕ.

So every such triple $(\Omega(s), \Phi, P(A,n,s))$ will be call a probability space according to the viewpoint of modern probability theory.

3. Acceptance or rejection degree

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.

Choose a fuzzy hypothesis H whose value is likely to best reflect the plausibility of the fuzzy hypothesis being tested.

Let us consider fuzzy membership function $m_A(x)$, which we will call the agreement index of $m_A(x)$ which regard to $m_H(x)$.

Definition 3.1. Let a fuzzy membership function $m_H(x)$, $x \in R$, we consider another membership function $m_A(x)$, $x \in R$, which call the agreement index, the ratio being defined in the following way;

$$R(A,H)=\frac{area(m_A(x) \cap m_H(x))}{area(m_A(x))} \in [0,1] \qquad (3.1)$$

as shown in Figure 3.1.

Formulate the structure of the rejection degree region by saying that fuzzy hypotheses H_0 should be rejected if the observed value of fuzzy test statistics T is too large, too small or intermediate, as the case may be in the following way.

Definition 3.2. We define the maximum grade membership function of acceptance or rejection degree by agreement index for real-valued function R_δ by δ-level on Θ as

$$m_{\Re_\delta}(0)=\sup_\psi\left\{\frac{\text{area }(m_{H_\delta}(\psi) \cap m_{T_\delta}(\psi))}{\text{area } m_{H_\delta}(\psi)}\right\} \qquad (3.2)$$

$$m_{\Re_\delta}(1)=1-m_{\Re_\delta}(0) \qquad (3.3)$$

for the fuzzy hypothesis testing as Figure 3.1.

Definition 3.3. In agreement index, we have the area by δ-level as:

$$area(m_A(x) \cap m_H(x))=\int_{\delta_0}^{\delta_1}(A_r^{-1}(\delta)-H_l^{-1}(\delta))d\delta$$

$$area\ m_{H_\delta}(\theta)=\int_{\delta_0}^{1}(A_r^{-1}(\delta)-A_l^{-1}(\delta))d\delta \qquad (3.4)$$

where A_r, A_l are right and left side line of $m_A(x)$, H_l is left side line of $m_H(x)$ and δ_0 is reliable degree and δ_1 is meeting point of $m_A(x)$ and $m_H(x)$.

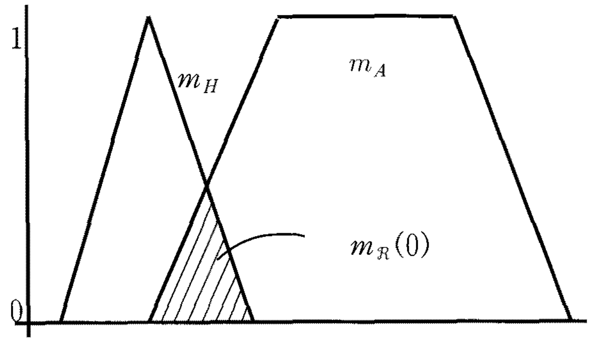


Figure 3.1. Agreement index of $m_A(x)$ regard to $m_H(x)$.

4. The two types of fuzzy error

A fuzzy test of the fuzzy null hypothesis is a course of separation of a fuzzy set if values of a fuzzy random variable X for which H_0 is would be rejected. The fuzzy random variable whose value serves to determinable the action is called the fuzzy test statistic, and the fuzzy set of its values for which H_0 is to be rejected is called the fuzzy rejection region of the test. A fuzzy test of specifiable by a fuzzy test statistic and the fuzzy rejection region is denoted by

$$H_0: p \prec p_0 \text{ or } H_0: p \simeq p_0$$

where $\prec$ is less than or similarity.

Considering the unknown state of nature and the possible results from applying a fuzzy test, one of the following situations will arise:

	UNKNOWN TRUE STATE OF NATURE	
TEST concludes	H_0 true $p<p_0$	H_0 false $(p>p_0)$
Do not reject H_0	Correct degree	Wrong (type Ⅱ error)
Reject H_0	Wrong (type Ⅰ error)	Correct degree

Fuzzy Type Ⅰ error : rejection degree of H_0 when H_0 is true.

Fuzzy Type Ⅱ error : failure degree to reject H_0 when H_1 is true.

The probabilities of the two types of error

$\tilde{\alpha}=P$[type Ⅰ error] = P[rejection degree of H_0 when H_0 is true]

$\tilde{\beta}=P$[type Ⅱ error] = P[not rejecting degree H_0 when H_1 is true]

The probability $\tilde{\alpha}$ depends on the particular value of the parameter in the range covered by H_0, whereas $\tilde{\beta}$ depends on the value over the range covered by H_1 and

$\gamma(p)=P$ [the test rejects H_0 when the true value of the parameter is p].

Under H_0, p is restricted to the range $p<p_0$, which is to the left of the middle vertical membership function in Figure 4.1. In this part of the graph, the rejection probability $\gamma(p)$ is, by definition, the same as the type Ⅰ error probability $\tilde{\alpha}(p)$. Under H_1, the range of p is $p>p_0$, which is to the right of the middle vertical membership function. In this range, $1-\gamma(p)=P$[retain H_0]=P[type Ⅱ error]=$\tilde{\beta}(p)$. Thus the graph of the rejection probability curve $\gamma(p)$ of a test provides a complete picture of the performance of the test for all possible contingencies with regard to the true degree state of nature.

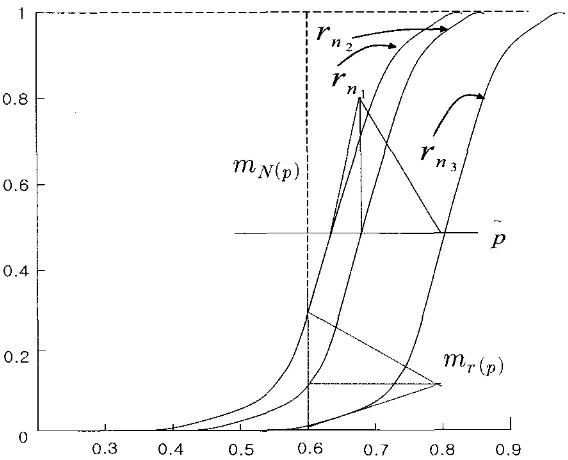


Figure 4.1 Power curves for the tests

5. Drawing conclusions from a test

In binomial proportion test, if we have a fuzzy test with rejection region

$$X>n_i\ ,\ i=1,2,\cdots,n \quad (5.1)$$

then rejection probabilities for fuzzy test is

$$r(p)=P(X>n_i) \quad (5.2)$$

such as power function fuzzy number

$$[r^l_{n_i}(p),r^c_{n_i}(p),r^r_{n_i}(p)] \quad (5.3)$$

in Figure 4.1.

If we have a specification probability of Chapter 2 for the null hypothesis as

$$H_0: P(A,n,s) < \widetilde{0.6} \quad (5.4)$$

where $m_A(p)=\begin{cases}10p-5, & 0.5\le p\le 0.6\\ -10p+7, & 0.6\le p\le 0.7\end{cases}$.

and we denote observation value of $P(A,n,s)$ by $\tilde{p}$ with trial 20 then X is a random variable with possible values of $0,1,2,\cdots,20$ of $B(X,20,\widetilde{0.6})$.

Thus we have power function

$$\gamma_c(\tilde{p_0})=P\{X>c|\tilde{p}\}=1-P\{X<c-1|\tilde{p}\}$$
$$=1-\sum_{x=0}^{c-1}\binom{20}{x}\tilde{P_0}^x(1-\tilde{p_0})^{20-x}.$$

If we have the level of significance 0.05,

for $\delta=1.0$, we have $\tilde{p}$= [0.6, 0.6, 0.6] as

$\tilde{p}$	0.5	0.6	0.7
15	0.021	0.126	0.416
16	0.006	0.051	0.238
17	0.001	0.016	0.107
18	0	0.004	0.035

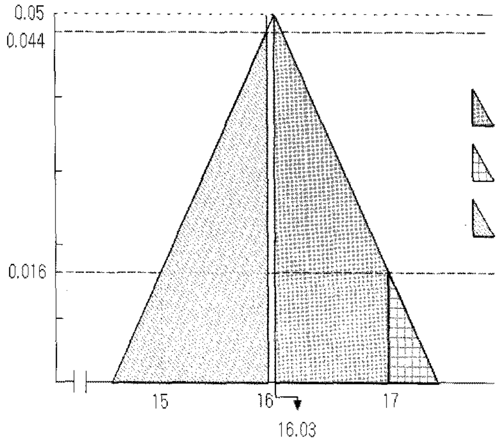


Figure 5.1. Test statistics for $\delta = 1.0$

thus we have $S_1 = 0.037$, $S_2 = 0.004$, $S_3 = 0.032$ by Figure 5.1, the acceptance degree for $X \geq 17$ is $m_{\Re_r}(1) = \frac{S_2}{S_1} = 0.108$ and the acceptance degree for $X \geq 16$ is $m_{\Re_r}(1) = \frac{S_3}{S_1} = 0.867$.

For $\delta = 0.5$, Then we have $\tilde{p} = [0.55, 0.60, 0.65]$ as

	0.55	0.6	0.65
16	0.019	0.051	0.118
17	0.005	0.016	0.044
18	0.001	0.004	0.012
19	0	0.001	0.002

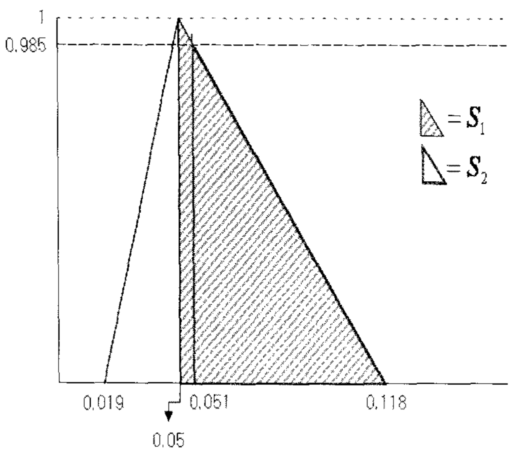


Figure 5.2. Test statistics for $\delta = 0.5$

thus we have $S_1 = 0.034$, $S_2 = 0.033$ by Figure 5.2, the acceptance degree for $X = 16$ is $m_{\Re_r}(0) = \frac{S_2}{S_1} = 0.971$.

For $\delta = 0.0$, we have $\tilde{p} = [0.5, 0.6, 0.7]$ as

$\tilde{p}$	0.5	0.6	0.7
15	0.021	0.126	0.416
16	0.006	0.051	0.238
17	0.001	0.016	0.107
18	0	0.004	0.035

thus we have $S_1 = 0.094$, $S_2 = 0.093$ by Figure 5.3, the acceptance degree for $X = 15$ is $m_{\Re_r}(0) = \frac{S_2}{S_1} = 0.99$.

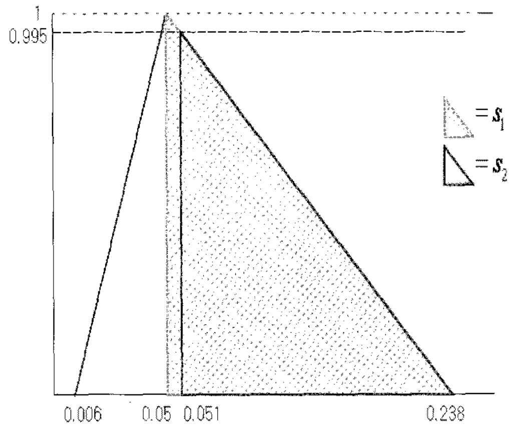


Figure 5.3. Test statistics for $\delta = 0$

From the fuzzy binomial proportion test, we have a part of the fuzzy binomial probability mass function as Figure 5.4.

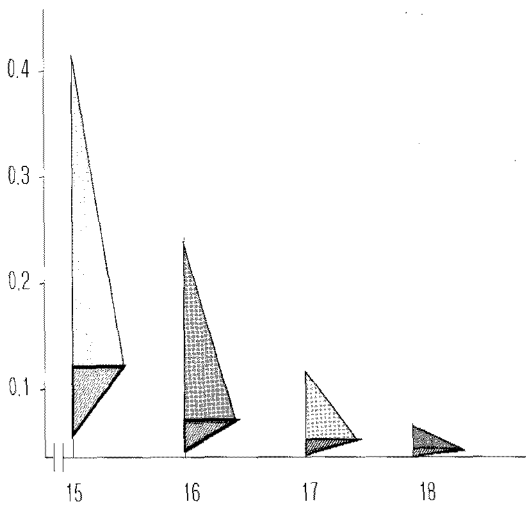


Figure 5.4. Some parts of fuzzy binomial probability mass function

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 Test 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, 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.

[4] W. Trutschnig, A Strong Consistency Rresult for Fuzzy Relative Frequencies Interpreted as Estimator for the Fuzzy-value Probability, *Fuzzy Sets and Systems,* 159, pp 259-269, 2008.

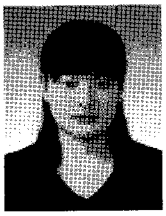
[5] N. Watanabe and T. Imaizumi, A Fuzzy Statistical Test of Fuzzy Hypotheses, *Fuzzy Sets and systems,* 53 , pp.167-178, 1993.

[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)

한국퍼지 및 지능시스템학회 논문지 Vol.15 No.5(2005), Vol.17 No.5(2007) 참조



박영례(Young-Rye Park)
현재 : 동의대학교 대학원 정보통계학과 석사과정

관심분야 : 퍼지데이터처리