퍼지 확률에 의한 이항분포

The Binomial Distribution with Fuzzy Valued Probability

강만기[1], 서현아[2], 박영래[2], 최규탁[3]

1 부산시 진구 동의대학교 데이터정보학과
E-mail: mkkang@deu.ac.kr
2 부산시 진구 동의대학교 데이터정보학과 대학원
3 부산시 사상구 경남정보대학 경영학부

Abstract

We introduce some properties for fuzzy binomial distributions with fuzzy valued probability. First we define fuzzy type I error and type II error for fuzzy relative frequency and agreement index. And we show that an fuzzy power function and fuzzy binomial frequency function for binomial proportion test.

Key Words : power function, fuzzy probability, fuzzy binomial frequency, agreement index.

1. Introduction

A discussion of the formulation of a fuzzy relative frequency problem and the steps for solving it requires the introduction of a number of definitions and concepts. Before proceeding with this topic in its full generality, we develop its basic ideas in terms of a specific problem in which the chance behavior is governed.

When an investigation is aimed at establishing an assertion with substantive support obtained from the sample, 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 H_1.

In testing a null fuzzy hypothesis H_0 against an fuzzy alternative H_1, our attitude is to uphold H_0 as true 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. Falsely rejecting H_0 is viewed as more serious error than failing to reject H_0 when H_1 is true.

Thus, we introduction some properties for fuzzy power function of performance of a test. First we define fuzzy type I error and type II error for the probability of the two types of error. And we show that an fuzzy statistical test by large number and graph of fuzzy probability mass function for binomial distribution.

2. Fuzzy valued probability

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

From the probabilistic point of view the unavoidable fuzziness of measurements has various far-reaching consequences.

Suppose that (Ω, A, P) is a probability space and that $X : \Omega \to R$ is a random variable that cannot be observed precisely, but that is possible to observe the interval $\widetilde{X}(s) = [\underline{X}(s), \overline{X}(s)]$, which contains the true value $X(s)$ for every $s \in \Omega$ will be call the sample point.

Definition 2.1. Assuming that $\widetilde{X}$ is a random interval and so-called lower and upper probability, $\underline{\pi}$ and $\overline{\pi}$ respectively, defined by

$$\underline{\pi}(B) := P(\{s\in\Omega : \widetilde{X}(s)\subseteq B\}),$$
$$\overline{\pi}(B) := P(\{s\in\Omega : \widetilde{X}(s)\cap B\neq\emptyset\}) \quad (2.1)$$

for every Borel set $B\in\mathrm{B}(R)$. Obviously $\underline{\pi}(B)$ and $\overline{\pi}(B)$ can be interpreted as lower and upper bound for $P(\{s\in\Omega : X(s)\in B\})$.

Given a sequence $\widetilde{X}_1, \widetilde{X}_2, \ldots$ of pairwise independent random intervals with the same distribution as $\widetilde{X}$ define the upper and lower relative frequency, $\underline{f}_n(B, s)$ and $\overline{f}_n(B, s)$, respectively, for every Borel set $B\in\mathrm{B}(R)$, every $s\in\Omega$ and every $n\in N$ analogous to (2.1) by

$$\underline{f}_n(B, s) := \frac{1}{n} m\{i\in\{1,\cdots,n\} : \widetilde{X}_i(s)\subseteq B\}$$
$$, \overline{f}_n(B, s) := \frac{1}{n} m\{i\in\{1,\cdots,n\} : \widetilde{X}_i\cap B\neq\emptyset\}$$
$$. \quad (2.2)$$

Definition 2.2. Fix a Borel set $A\in\mathrm{B}(R)$ and for every $\widetilde{X}_i (i\in N)$ define two $\{1,0\}$-valued random variables $\underline{Z}_i$, $\overline{Z}_i$ by

$$\underline{Z}_i(s) := \begin{cases}1 & \text{if } \widetilde{X}_i(s)\subseteq A,\\ 0 & otherwise.\end{cases}$$
$$\overline{Z}_i(s) := \begin{cases}1 & \text{if } \widetilde{X}_i(s)\cap A\neq\emptyset,\\ 0 & otherwise.\end{cases} \quad (2.3)$$

It follows immediately that both equalities

$$\lim_{n\to\infty}\underline{h}_n(A, s) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\underline{Z}_i(s) = (\underline{Z}_1) = \underline{\pi}(A)$$

and

$$\lim_{n\to\infty}\overline{h}_n(A, s) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\overline{Z}_i(\omega) = E(\overline{Z}_1) = \overline{\pi}(A)$$
$$(2.4)$$

hold for p-almost every $s\in\Omega$, which using the Hausdorff metric δ_H can be summarized by the fact that

$$\lim_{n\to\infty}\delta_H[\underline{h}_n(A,s), \overline{h}_n(A,s)], [\underline{\pi}(A), \overline{\pi}(A)]) = 0 \quad (2.5)$$

holds for p-almost every $s\in\Omega$.

Remark. In the one-dimensional case it can furthermore be shown that the fuzzy random variable X^* is equivalent to the measurability of the boundaries $\underline{X}_\gamma, \overline{X}_\gamma$, of the γ-cuts, defined by

$$[\underline{X}_\gamma(s), \overline{X}_\gamma(s)] := \widetilde{X}_\gamma(s) = [\widetilde{X}(s)]_\gamma \quad (2.6)$$

for every $\gamma\in(0, 1]$ and every $s\in\Omega$.

Futhermore it follows directly from the definition that $\underline{h}_{n,\gamma}(B)$ is monotonically increasing and that $\overline{h}_{n,\gamma}(B)$ is monotonically decreasing as function α. Consequently

$$[\underline{h}_{n,\gamma_1}(B), \overline{h}_{n,\gamma_1}(B)] \supseteq [\underline{h}_{n,\gamma_2}(B), \overline{h}_{n,\gamma_2}(B)] \quad (2.7)$$

holds for $\gamma_1\le\gamma_2$ and $\gamma_1, \gamma_2\in(0, 1]$.

3. Agreement index

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.

Definition 3.1. Let a fuzzy membership function $\chi_H(x), x\in R$, we consider another membership function $\chi_A(x), x\in R$, which we call the agreement index of A with regard to H the ratio being defined in the following way;

$$R(A, H) = (area(\chi_A(x)\cap\chi_H(x)) / area(\chi_A(x)) \in [0, 1] \quad (3.1)$$

as shown in Figure 3.1.

Definition 3.2. We define agreement index by real-valued function R_γ on $\psi\in\Theta$ as the maximum grade membership function of acceptance or rejection is

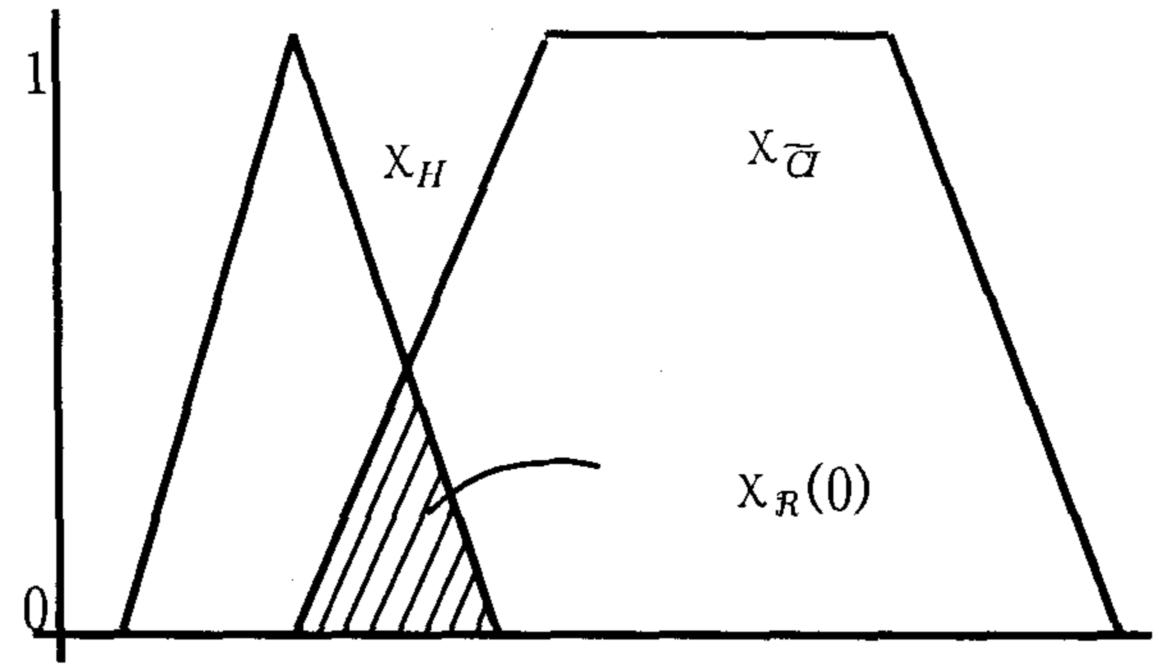
$$\chi_{R_\gamma}(0) = \sup_\psi\{area\ (\chi_{H_\gamma}(\psi)\cap\chi_{\overline{T}_\gamma}(\psi)) / area\ \chi_{H_\gamma}(\psi)\} \quad (3.2)$$
$$\chi_{R_\gamma}(1) = 1 - \chi_{R_\gamma}(0) \quad (3.3)$$

for the fuzzy hypothesis testing as Figure 1.

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

$$area(\chi_A(x)\cap\chi_H(x)) = \int_{\gamma_0}^{\gamma_1}(A_r^{-1}(\gamma) - H_l^{-1}(\gamma))d\gamma \quad (3.4)$$

where A_r, A_l are right and left side line of $\chi_A(x)$, H_l is left side line of $\chi_H(x)$ and γ_0 is reliable degree and γ_1 is meeting point of $\chi_A(x)$ and $\chi_H(x)$.



[Figure 3.1]

4. Fuzzy power function

A test of the fuzzy null hypothesis H_0 is to be rejected is called the fuzzy rejection region of the test. A test of completely specified by a fuzzy test statistic and the fuzzy rejection region.

	Unknown true state of nature	
test concludes	H_0 true $p<p_0$	H_0 falselse $(p>p_0)$
Do not reject H_0	Correct	Wrong (type II error)
Reject H_0	Wrong (type I error)	Correct

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

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

$\tilde{\beta} = \tilde{P}$[type II error] = $\tilde{P}$[not rejecting 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 β depends on the value over the range covered by H_1.

$\gamma(p) = P$ [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 line 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 line. In this range, $1-\gamma(p) = \tilde{P}$[retain H_0]= $\tilde{P}$[type II error]= $\beta(p)$. 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 state of nature.

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

$$X > n_i \tag{4.1}$$

then rejection probabilities for fuzzy test is

$$r(p) = P(X > n_i) \tag{4.2}$$

such as fuzzy number

$$[r^l{}_{n_i}(p), r^c{}_{n_i}(p), r^r{}_{n_i}(p)] \tag{4.3}$$

in Figure 4.1.

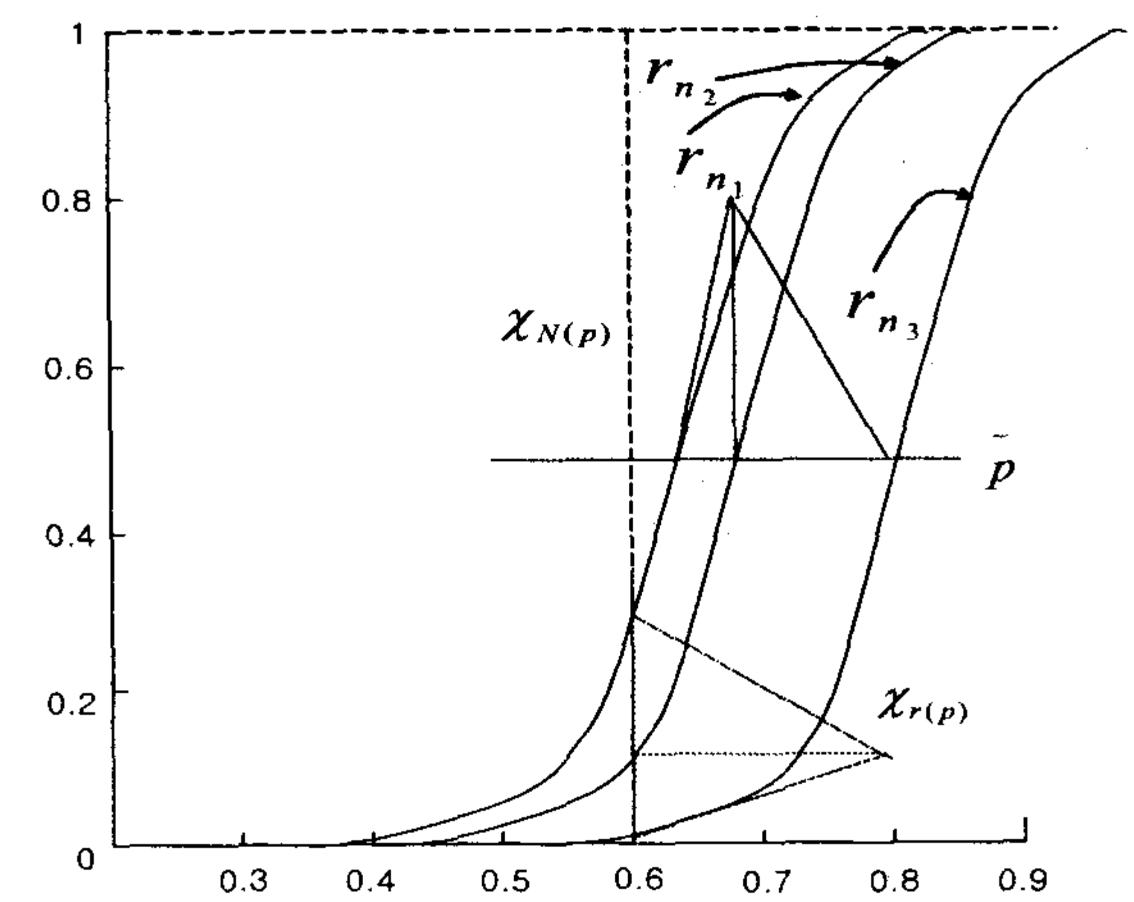


Figure 4.1 Power curves for the tests

5. Fuzzy binomial distribution

By (2.6), $\widetilde{X_\gamma}(s)$ have fyzzy probability $\tilde{p}$ and tests for $\tilde{p}$ will be based on the sufficient statistic $X(s) \sim BIN(n, \tilde{p})$.

Theorem 5.1 $BIN(n, \tilde{p})$ is fuzzy distribution with probability mass function

$$f(x) = nCx\,\tilde{p}^{\,x}(1-\tilde{p})^{n-x},\ x = 0, 1, \cdots, n \tag{5.1}$$

Proof If we take one place point $\forall q \in \tilde{p}$ then 1) $f(x) = nCx\,q^{\,x}(1-q)^{n-x} \geq 0$

2) $\sum_{x=0}^{20} nCx\,q^{\,x}(1-q)^{n-x} = 1$.

By the resolution identity,

$nCx\tilde{p}^{x}(1-\tilde{p})^{n-x}$ has binomial p.m.f..

Theorem 5.2

For large n an approximate size α test of $H_0: \tilde{p}<\tilde{p}_0$ against $H_a: \tilde{p}>\tilde{p}_0$ is to reject H_0 if

$$\tilde{Z}_0 = \frac{\tilde{X}_Y(s)-\tilde{p}_o}{\sqrt{\tilde{p}_0(1-\tilde{p}_0)/n}} > Z_{1-\alpha} \quad (5.2)$$

An approximate size α test of $H_0: \tilde{p}>\tilde{p}_0$ against $H_a: \tilde{p}<\tilde{p}_0$ is to reject H_0 if

$$\tilde{Z}_0 < -Z_{1-\alpha}$$

An approximate size α test of $H_0: \tilde{p}\approx\tilde{p}_0$ against $H_a: \tilde{p}\neq\tilde{p}_0$ is to reject H_0 of

$$\tilde{Z}_0 < -Z_{1-\alpha/2} \text{ or } \tilde{Z}_0 > Z_{1-\alpha/2}$$

These results follow from the fact that when $\tilde{p}=\tilde{p}_0$, then

$$\frac{\tilde{X}_Y(s)-\tilde{p}_o}{\sqrt{\tilde{p}_0(1-\tilde{p}_0)/n}} \xrightarrow{d} Z\sim N(0,1) \quad (5.3)$$

As in the previous one-sided examples, the probability of rejecting a true H_0 will be less than α for other values of $\tilde{p}$ in the null hypotheses.

If we have a specification of the null hypothesis as

$$H_0: \tilde{p}< \widetilde{0.6} \quad (5.4)$$

where

$$\chi_A(p) = \begin{cases} 20p-11, & 0.55\le p\le 0.6 \\ -20p+13, & 0.6\le p\le 0.65 \end{cases} \quad (5.5)$$

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

$$r_c(\tilde{p}_0) = \{P(X>c|\tilde{p}\} = 1-P\{X<c-1|\tilde{p}\}$$
$$=1-\sum_{x=0}^{c-1}\binom{20}{x}\tilde{P}_0^x(1-\tilde{p}_0)^{20-x} \quad (5.6)$$

The fuzzy binomial frequency function is shown as Figure 5.1.

If we have $\tilde{X}_Y(s)=[0.680,0.880]$ and $\tilde{Z}_0=[0.55,0.65]$ then rejection degree is $\chi_{R_\alpha}(1)=0.1875$ by (3.2) and (3.3).

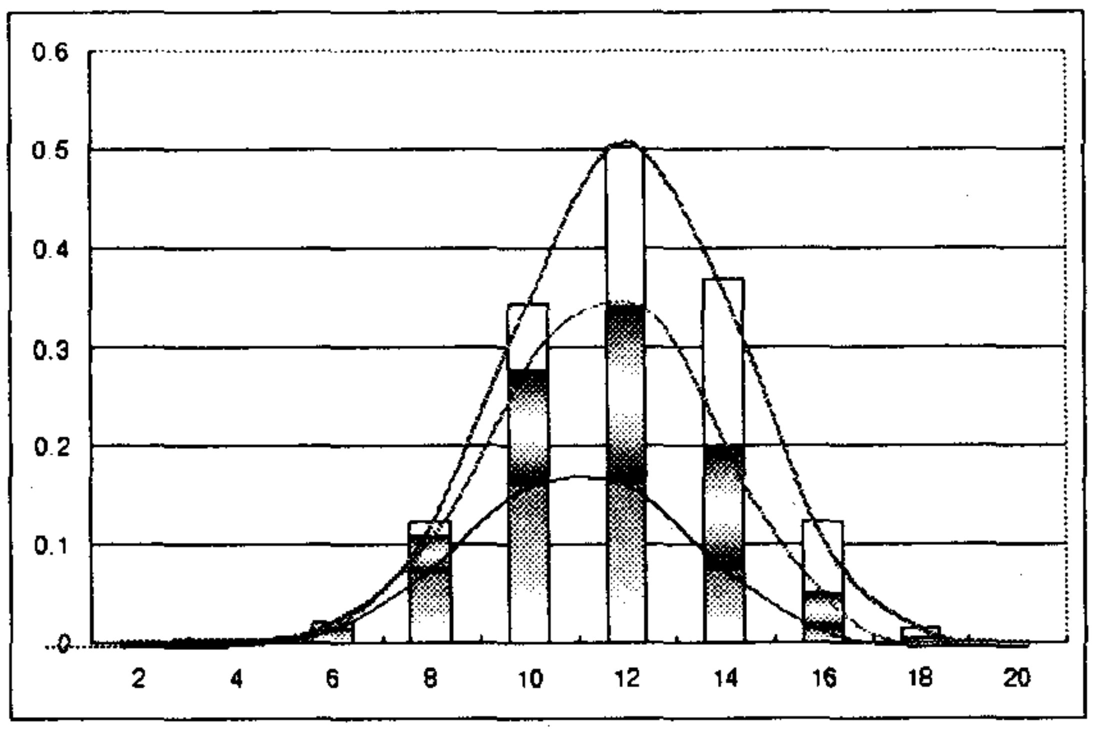


Figure 5.1 fuzzy binomial frequency function

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 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 with Application to Des Crispy Statistics, *Fuzzy Sets and Systems,* 50, pp 143-165, 1992.

[6] 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.

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

[8] Z. Xia, Fuzzy Probability System: fuzzy probability system, *Fuzzy Sets and Systems 120* , pp.469-486, 2001.