International Journal of Innovative Computing, Information and Control
Volume **8**, Number **10(B)**, October **2012**

 ISSN 1349-4198
pp. 7437–7450

GOODNESS-OF-FIT TEST FOR MEMBERSHIP FUNCTIONS WITH FUZZY DATA

Pei-Chun Lin[1], Berlin Wu[2] and Junzo Watada[1]

[1]Graduate School of Information, Production and System
Waseda University
2-7 Hibikino, Wakamatsu, Kitakyushu 808-0135, Japan
{ peichunpclin; junzo.watada }@gmail.com

[2]Department of Mathematical Sciences
National Chengchi University
No. 64, Sec. 2, Zhi-nan Rd., Taipei 11605, Taiwan
berlin@nccu.edu.tw

Received April 2011; revised September 2011

Abstract. *Conventionally, we use a chi-square test of homogeneity to determine whether the cell probabilities of a multinomial are equal. However, this process of testing hypotheses is based on the assumption of two-valued logic. If we collect questionnaire data using fuzzy logic, i.e., we record the category data with memberships instead of with a 0-1 type, then the conventional test of goodness-of-fit will not work. In this paper, we present a new method, the fuzzy chi-square test, which will enable us to analyze those fuzzy sample data. The new testing process will efficiently solve the problem for which the category data are not integers. Some related properties of the fuzzy multinomial distribution are also described.*

Keywords: Fuzzy set theory, Fuzzy numbers, Membership functions, Sampling survey, Chi-square test for goodness-of-fit

1. **Introduction.** Consider an l-dimensional multinomial vector $n = \{n_1, n_2, \ldots, n_l\}$ with the constraint $\sum_{i=1}^{l} n_i = N$. The Pearson chi-squared statistic test [12] $\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ is a well-known statistic used to test the significance between expected values and observed data. It is clear that large discrepancies between expected and observed values correspond to large chi-square values. However, when values in a sample are expressed as fuzzy numbers, can the data be considered categorical and thus allow us to pursue the conventional chi-square test? For example, when we are asked "How satisfied are you with your life?", we may respond with a fuzzy number (e.g., approximately 70% satisfied and 30% dissatisfied). If we record fuzzy numbers in questionnaires, then the sampling survey is no longer like the conventional one.

Wu and Chang [15] used a fuzzy chi-square test to identify significant differences, but did not provide a theoretical proof. The literature provides various definitions of fuzzy random variables and fuzzy expected values [6, 7, 8, 13]. Most of these definitions are derived from the conventional concept of probability; as a result, it is difficult to derive certain properties specific to fuzzy-numbers under these definitions. More research has focused on fuzzy statistical analysis and its application in social science fields; Casalino et al. [2], Esogbue and Song [4], and Wu and Sun [16] have described the concepts of fuzzy statistics and applied them to social survey. Sun [14] applied fuzzy statistical analysis to lexical semantics computation and Chang et al. [3] proposed a fuzzy inference criterion for assessing process lifetime performance. Moreover, Lin et al. [9] defined a new fuzzy-number weight function that uses a central point and radius to more effectively observe the original fuzzy data. Later, Lin et al. [10] also purposed a method that uses a central point and radius to identify an underlying distribution function, again providing more information about the original fuzzy data.

In this paper, we will show the properties of a certain fuzzy statistic distribution, which is somewhat similar to the conventional statistic distribution. Moreover, we provide an empirical study that uses the fuzzy chi-square test to test hypotheses.

2. **Fuzzy Statistic Analysis.**

2.1. **Definitions.**

Fuzzy set theory, proposed by Zadeh in 1965 [17], deal with vagueness in data. The following definitions will be used in next subsection and Section 3.

Definition 2.1. *Let U be a universal set and $A = \{A_1, A_2, \ldots, A_n\}$ be the set of focal factors in U. For any term or statement X on U, the membership function of $\{A_1, A_2, \ldots, A_n\}$ is $\{\mu_1(X), \mu_2(X), \ldots, \mu_n(X)\}$, where $\mu : U \to [0, 1]$ is a real value function. If the domain of the universal set is discrete, then the fuzzy number x of X can be written as*

$$\mu_U(X) = \sum_{i=1}^{n} \mu_i(X) I_{A_i}(X), \tag{1}$$

where $I_{A_i}(X) = 1$ if $x \in A_i$ and $I_{A_i}(X) = 0$ if $x \notin A_i$.

If the domain of the universal set is continuous, then the fuzzy number x can be written as

$$\mu_U(X) = \int_{A_i \subseteq A} \mu_i(X) I_{A_i}(X). \tag{2}$$

Note that, in the past, a fuzzy number was often written not as it is in Equation (1) but instead as

$$\mu_U(X) = \frac{\mu_1(X)}{A_1} + \frac{\mu_2(X)}{A_2} + \cdots + \frac{\mu_n(X)}{A_n}$$

where "+" stands for "or" and "÷" denotes the membership $\mu_i(X)$ on A_i.

Definition 2.2. *Expected value for fuzzy sample data (data with multiple values).*

Let U be the universal set (a discussion domain), $L = \{L_1, L_2, \ldots, L_k\}$ be a set of k-linguistic variables on U, and $\left\{F_{x_i} = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1, 2, \ldots, n\right\}$ be a sequence of fuzzy random sample on U, m_{ij} ($\sum_{j=1}^{k} m_{ij} = 1$) are the memberships with respect to L_j [11] and have a fuzzy Bernoulli distribution. Then, the expected value for fuzzy sample data is defined as

$$E(F_{x_i}) = \frac{E(m_{i1})}{L_1} + \frac{E(m_{i2})}{L_2} + \cdots + \frac{E(m_{ik})}{L_k}.$$

Definition 2.3. *Variance for fuzzy sample data (data with multiple values).*

As Definition 2.2, the variance for fuzzy sample data is written as follows:

$$Var(F_{x_i}) = \frac{Var(m_{i1})}{L_1} + \frac{Var(m_{i2})}{L_2} + \cdots + \frac{Var(m_{ik})}{L_k}.$$

2.2. **Fuzzy bernoulli and fuzzy binomial distribution.** In this section, we introduce several new distribution functions. A Bernoulli trial is an experiment that has only two possible (and incompatible) outcomes: "success" and "failure" [5]. In general, let $X = 1$ if the outcome of Bernoulli trial is a success, and let $X = 0$ if it is a failure. Then, we say that a fuzzy Bernoulli experiment is a random experiment, the outcome of which can be classified in one of two mutually exclusive and exhaustive ways: success or failure (i.e., we let $X \in [0.5, 1]$ if the outcome of fuzzy Bernoulli trial is a success and $X \in [0, 0.5)$ if it is a failure). Hence, a sequence of fuzzy Bernoulli trials occurs. In such a sequence we let π denote the probability of success for each trial. In addition, we will frequently let $q = 1 - \pi$ denote the probability of failure.

Next, let X be a continuous random variable associated with a fuzzy Bernoulli trial by define as

$$X \in [0.5, 1] \text{ for a success and } X \in [0, 0.5) \text{ for a failure;}$$

that is, the two outcomes, success and failure, are denoted by mutually exclusive parts of the partition set $[0, 1]$.

The probability density function (p.d.f.) of X can be written as

$$f(x) = \begin{cases} 2\pi, & \text{if } x \in [0.5, 1] \\ 2(1-\pi), & \text{if } x \in [0, 0.5) \end{cases}. \tag{3}$$

In this scenario, we say that X has a fuzzy Bernoulli distribution which is denoted by $X \sim FB(1, \pi)$.

We first derive some properties of the fuzzy Bernoulli distribution.

Theorem 2.1. *The fuzzy Bernoulli density function given in (3) is a density function. If $X \sim FB(1, \pi)$, then the expected value of X is $\mu = E(X) = \dfrac{1 + 2\pi}{4}$, and the variance of X is $\sigma^2 = Var(X) = \dfrac{1}{48} + \dfrac{1}{4}\pi(1-\pi)$. Finally, the moment-generating function of X is*

$$M(t) = E(e^{tX}) = \begin{cases} 2\left(\dfrac{e^{\frac{t}{2}} - 1}{t}\right)\left[\pi e^{\frac{t}{2}} + (1-\pi)\right], & \text{if } t \neq 0 \\ 1, & \text{if } t = 0 \end{cases}.$$

Proof: Note that $f(x) \geq 0$. Also note that

$$\int_0^1 f(x)dx = \int_0^{0.5} 2(1-\pi)dx + \int_{0.5}^1 2\pi dx = 2(1-\pi) \cdot x \,|_0^{0.5} + 2\pi \cdot x \,|_{0.5}^1 = 1.$$

Thus, f is a density function.

$$\begin{aligned} E(X) &= \int_{0.5}^1 x \cdot 2\pi dx + \int_0^{0.5} x \cdot 2(1-\pi)dx = \frac{1+2\pi}{4}, \\ Var(X) &= E(X^2) - [E(X)]^2 \\ &= \int_{0.5}^1 x^2 \cdot 2\pi dx + \int_0^{0.5} x^2 \cdot 2(1-\pi)dx - \left(\frac{1+2\pi}{4}\right)^2 = \frac{1}{48} + \frac{1}{4}\pi(1-\pi), \end{aligned}$$

and the moment-generating function of X is

$M(t) = E(e^{tX}) = \displaystyle\int_{0.5}^1 e^{tx} 2\pi dx + \int_0^{0.5} e^{tx} 2(1-\pi)dx = 2\left(\frac{e^{\frac{t}{2}} - 1}{t}\right)\left[\pi e^{\frac{t}{2}} + (1-\pi)\right]$ for $t \neq 0$.

The moment-generating function is not differentiable at zero, but can be calculated by taking $\lim_{t\to 0}$. We present these operations as follows:

$$
\begin{aligned}
M(0) &= \lim_{t\to 0} M(t) = \lim_{t\to 0} 2\left(\frac{e^{\frac{t}{2}}-1}{t}\right)\left[\pi e^{\frac{t}{2}} + (1-\pi)\right] = \lim_{t\to 0} 2\frac{\left(e^{\frac{t}{2}}-1\right)\left[\pi e^{\frac{t}{2}} + (1-\pi)\right]}{t} \\
&= \lim_{t\to 0}\left\{2\left(e^{\frac{t}{2}}\cdot\frac{1}{2}\right)\left[\pi e^{\frac{t}{2}} + (1-\pi)\right] + 2\left(e^{\frac{t}{2}}-1\right)\left(\pi e^{\frac{t}{2}}\cdot\frac{1}{2}\right)\right\} = 1 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(by L'Hospital's Rule).}
\end{aligned}
$$

In a sequence of fuzzy Bernoulli trials, we are often interested in the total number of successes, and we do not consider the order of their occurrence. If we let the random variable M be the number of observed successes in n fuzzy Bernoulli trials, M takes the value of any nonnegative number. To easily denote the fuzzy Binomial distribution, let k denote the number of observed successes, where $2m - n < k \le 2m$ for $k \in N \cup \{0\}$ when $m < n$ and for $k = n$ when $m = n$, then let $n - k$ denote the number of failures occurred. We say that m is the observed value of M and N is defined as the natural number. (The same definitions are used in the following analysis.) The number of ways to select k positions for the k successes in the n trials is $\binom{n}{k}$. Note that, when we know the value of m, the values of k are decided.

The p.d.f. of M can be written as

$$
f(m) = \frac{2}{n}\sum_{k\in\Omega}\binom{n}{k}\pi^k(1-\pi)^{n-k}, \tag{4}
$$

where $\Omega = \{k \in N \cup \{0\} \mid 2m - n < k \le 2m \text{ for } m < n \text{ and } m = n \text{ for } k = n\}$. Another way to represent the p.d.f. is as either

$$
f(m) = \frac{2}{n}\begin{cases}
\binom{n}{0}(1-\pi)^n, & 0 \le m < 0.5, & k = 0 \\
\binom{n}{0}(1-\pi)^n + \binom{n}{1}\pi^1(1-\pi)^{n-1}, & 0.5 \le m < 1, & k = 0,1 \\
\binom{n}{0}(1-\pi)^n + \binom{n}{1}\pi^1(1-\pi)^{n-1} + \binom{n}{2}\pi^2(1-\pi)^{n-2}, & 1 \le m < 1.5, & k = 0,1,2 \\
\vdots, & \vdots, & \vdots \\
\binom{n}{0}(1-\pi)^n + \binom{n}{1}\pi^1(1-\pi)^{n-1} + \cdots + \binom{n}{n-1}\pi^{n-1}(1-\pi)^1, & 0.5(n-1) \le m < 0.5n, & k = 0,1,\ldots,n-1 \\
\binom{n}{1}\pi^1(1-\pi)^{n-1} + \binom{n}{2}\pi^2(1-\pi)^{n-2} + \cdots + \binom{n}{n}\pi^n, & 0.5n \le m < 0.5(n+1), & k = 1,2,\ldots,n \\
\binom{n}{2}\pi^2(1-\pi)^{n-2} + \binom{n}{3}\pi^3(1-\pi)^{n-3} + \cdots + \binom{n}{n}\pi^n, & 0.5(n+1) \le m < 0.5(n+2), & k = 2,3,\ldots,n \\
\vdots, & \vdots, & \vdots \\
\binom{n}{n}\pi^n, & 0.5(2n-1) \le m \le 0.5(n+n), & k = n
\end{cases}
$$

or

$$
f(m) = \frac{2}{n}\begin{cases}
\binom{n}{0}\pi^0(1-\pi)^{n-0}, & 0 \le m < 0.5n, & k = 0 \\
\binom{n}{1}\pi^1(1-\pi)^{n-1}, & 0.5 \le m < 0.5(n+1), & k = 1 \\
\vdots, & \vdots, & \vdots \\
\binom{n}{n}\pi^n(1-\pi)^{n-n}, & 0.5n \le m \le 0.5(n+n), & k = n
\end{cases}.
$$

We say that M has a fuzzy Binomial distribution, which is denoted as $M \sim FB(n,\pi)$. The constants n and π are called the parameters of the fuzzy binomial distribution; they correspond to the number of trials and the probability of success for each trial, respectively.

Theorem 2.2. *The fuzzy Binomial density function given in (4) is a density function. If $M \sim FB(n,\pi)$, then the expected value of M is $\mu = E(M) = n \cdot \frac{1+2\pi}{4}$, and the variance of M is $Var(M) = \frac{n^2}{48} + \frac{n}{4}\pi(1-\pi)$. Finally, the moment-generating function of M is*

$$
M(t) = \begin{cases}
\frac{2}{n}\left(\frac{e^{\frac{n}{2}t}-1}{t}\right)\left[\pi e^{\frac{t}{2}} + (1-\pi)\right]^n, & \text{if } t \ne 0 \\
1, & \text{if } t = 0
\end{cases}.
$$

Proof: Note that $f(m) \geq 0$. Also, using the binomial theorem and integral operation, we have the following:

$$\begin{aligned}\int_0^n f(m)dm &= \int_0^n \frac{2}{n}\sum_{k\in\Omega}\binom{n}{k}\pi^k(1-\pi)^{n-k}dm \\ &= \frac{2}{n}\Big[\int_0^{0.5n}\binom{n}{0}\pi^0(1-\pi)^n dm + \int_{0.5}^{0.5(n+1)}\binom{n}{1}\pi^1(1-\pi)^{n-1}dm + \ldots \\ &\quad + \int_{0.5n}^{0.5(n+n)}\binom{n}{n}\pi^n(1-\pi)^0 dm\Big] \\ &= \frac{2}{n}\sum_{k=0}^{n}\int_{0.5k}^{0.5(n+k)}\binom{n}{k}\pi^k(1-\pi)^{n-k}dm = \frac{2}{n}\sum_{k=0}^{n}\binom{n}{k}\pi^k(1-\pi)^{n-k}\cdot\frac{n}{2} \\ &= \sum_{k=0}^{n}\binom{n}{k}\pi^k(1-\pi)^{n-k} = [\pi+(1-\pi)]^n = 1.\end{aligned}$$

Thus, f is a density function.

Using the binomial theorem and integral operation again, we have the following:

$$\begin{aligned}M(t) = E(e^{tM}) &= \frac{2}{n}\sum_{k=0}^{n}\int_{0.5k}^{0.5(n+k)} e^{tm}\cdot\binom{n}{k}\pi^k(1-\pi)^{n-k}dm \\ &= \frac{2}{n}\cdot\frac{1}{t}\left(e^{\frac{n}{2}t}-1\right)\sum_{k=0}^{n}\binom{n}{k}\left(e^{\frac{t}{2}}\pi\right)^k(1-\pi)^{n-k} \\ &= \frac{2}{n}\cdot\frac{e^{\frac{n}{2}t}-1}{t}\left[\pi e^{\frac{t}{2}}+(1-\pi)\right]^n \text{ for } t \neq 0.\end{aligned}$$

In the case $t = 0$, we can take $\lim_{t\to 0}$ and use L'Hospital's Rule to complete the proof.

Now, we let

$$\psi(t) = \log M(t) = \log\frac{2}{n} + \log\frac{e^{\frac{n}{2}}-1}{t} + n\cdot\log(\pi e^{\frac{t}{2}}+1-\pi),$$

we can get $\psi(0) = \log M(0) = \log 1 = 0$.

Moreover, let $\psi'(t) = \dfrac{t}{e^{\frac{n}{2}t}-1}\cdot\dfrac{e^{\frac{n}{2}t}\frac{n}{2}\cdot t-(e^{\frac{n}{2}t}-1)\cdot 1}{t^2} + n\cdot\dfrac{\pi e^{\frac{t}{2}}\frac{1}{2}}{\pi e^{\frac{t}{2}}+1-\pi}$ and

$$\begin{aligned}\psi''(t) =& \frac{1\cdot\left(e^{\frac{n}{2}t}-1\right)-t\cdot e^{\frac{n}{2}t}\frac{n}{2}}{\left(e^{\frac{n}{2}t}-1\right)^2}\cdot\frac{e^{\frac{n}{2}t}\frac{n}{2}\cdot t-\left(e^{\frac{n}{2}t}-1\right)}{t^2} \\ &+ \frac{t}{e^{\frac{n}{2}t}-1}\cdot\frac{\left[e^{\frac{n}{2}t}\left(\frac{n}{2}\right)^2\cdot t+e^{\frac{n}{2}t}\frac{n}{2}\cdot 1-e^{\frac{n}{2}t}\frac{n}{2}\right]\cdot t^2-\left[e^{\frac{n}{2}t}\frac{n}{2}\cdot t-\left(e^{\frac{n}{2}t}-1\right)\right]\cdot 2t}{t^4} \\ &+ n\cdot\frac{\pi e^{\frac{t}{2}}\left(\frac{1}{2}\right)^2\left(\pi e^{\frac{t}{2}}+1-\pi\right)-\left(\pi e^{\frac{t}{2}}\frac{1}{2}\right)^2}{\left(\pi e^{\frac{t}{2}}+1-\pi\right)^2}.\end{aligned}$$

We can get that

$$\mu = \psi'(0) = \lim_{t\to 0}\frac{\psi(t)-\psi(0)}{t-0} = \lim_{t\to 0}\frac{\psi(t)}{t} = \lim_{t\to 0}\psi'(t) = n\cdot\frac{1+2\pi}{4}$$

(by L'Hospital's Rule)

and $\sigma^2 = \psi''(0) = \lim\limits_{t\to 0} \dfrac{\psi'(t) - \psi'(0)}{t - 0} = \lim\limits_{t\to 0} \dfrac{\psi'(t) - \mu}{t} = \lim\limits_{t\to 0} \psi''(t) = \dfrac{n^2}{48} + \dfrac{n}{4}\pi(1-\pi)$.

In the next section, we will derive the fuzzy multinomial distribution which is expanded from the fuzzy binomial distribution.

3. Fuzzy Multinomial Distribution.

3.1. **Fuzzy trinomial distribution.** Let $M = (M_1, M_2)$ be a bivariate random vector with range $S_n = \{(m_1, m_2) : m_1 \geq 0, m_2 \geq 0 \text{ and } m_1 + m_2 \leq n\}$. That is, m_1 and m_2 are nonnegative real values, such that $m_1 + m_2 \leq n$. Also, let $K_n = \{(k_1, k_2) : 2m_i - n < k_i \leq 2m_i, k_i \in N \cup \{0\} \text{ for } i = 1, 2 \text{ and } k_1 + k_2 \leq n\}$ under the condition S_n. Then, we have a relation between K_n and S_n. When m_1 and m_2 are fixed, k_1 and k_2 are decided. Hence, M has a fuzzy trinomial distribution with parameters n and $\pi = (\pi_1, \pi_2)$, written as $M = (M_1, M_2) \sim FT(n, (\pi_1, \pi_2))$ if M has joint density function

$$f(m_1, m_2) = 2\begin{cases} \left(\dfrac{2}{n}\right)^2 \displaystyle\sum_{k_1}\sum_{k_2} \dfrac{n!}{k_1!k_2!(n-k_1-k_2)!}\pi_1^{k_1}\pi_2^{k_2}(1-\pi_1-\pi_2)^{n-k_1-k_2}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } (k_1, k_2) \in K_n \\ \left(\dfrac{2}{n}\right)^2 \displaystyle\sum_{k_1}\sum_{k_2} \dfrac{n!}{k_1!k_2!(n-k_1-k_2)!}\pi_1^{k_1}\pi_2^{k_2}(1-\pi_1-\pi_2)^{n-k_1-k_2}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } (k_1, k_2) \in K - K_n \end{cases} \tag{5}$$

where $K = \{(k_1, k_2) : k_1 \geq 0, k_2 \geq 0 \text{ and } k_1 + k_2 \leq n\}$ and $(m_1, m_2) \in S_n$.

In the above equations, n is a positive integer and π_1 and π_2 are nonnegative numbers such that $\pi_1 + \pi_2 \leq 1$.

To prove that f is a p.d.f. under S_n, we must extend the set S_n to $\tilde{S_n} \cup A$, where $\tilde{S_n} = \{(m_1, m_2) : 0.5k_i < m_i \leq 0.5(n + k_i), k_i \in N \cup \{0\} \text{ for } i = 1, 2 \text{ and } k_1 + k_2 \leq n\}$ and $A = S - \tilde{S_n}$ where A is measure zero. Note that, S is the set denoted by $S = \{(m_1, m_2) : 0 \leq m_1 \leq n \text{ and } 0 \leq m_2 \leq n\}$.

Theorem 3.1. *The fuzzy trinomial density function given in (5) is a density function. If* $(M_1, M_2) \sim FT(n, (\pi_1, \pi_2))$*, then* $E(M_i) = n \cdot \dfrac{1 + 2\pi_i}{4}$*,* $Var(M_i) = \dfrac{n^2}{48} + \dfrac{n}{4}\pi_i(1 - \pi_i)$*,* $cov(M_1, M_2) = -\dfrac{n}{4}\pi_1\pi_2$*, and the joint moment-generation function is*

$$M(t_1, t_2) = \begin{cases} \left(\dfrac{2}{n}\right)^2 \left(\dfrac{e^{\frac{n}{2}t_1} - 1}{t_1}\right)\left(\dfrac{e^{\frac{n}{2}t_2} - 1}{t_2}\right)\left[\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + (1 - \pi_1 - \pi_2)\right]^n, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } (t_1, t_2) \neq (0, 0) \\ 1, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if } (t_1, t_2) = (0, 0) \end{cases}.$$

Proof: Clearly, $f(m_1, m_2) \geq 0$. Using the trinomial theorem and integral operation, we have the following:

$$\begin{aligned} &\iint_{S_n} f(m_1, m_2)dm_1dm_2 \\ &= \int_0^n \int_0^{n-m_2} f(m_1, m_2)dm_1dm_2 \\ &= \int_0^n \int_0^{n-m_2} 2\left(\frac{2}{n}\right)^2 \sum_{k_1}\sum_{k_2} \frac{n!}{k_1!k_2!(n-k_1-k_2)!}\pi_1^{k_1}\pi_2^{k_2}(1-\pi_1-\pi_2)^{n-k_1-k_2}dm_1dm_2 \end{aligned}$$

$$= \left(\frac{2}{n}\right)^2 \sum_{k_1} \sum_{k_2} \left[\iint_{\tilde{S_n}} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1-\pi_2)^{n-k_1-k_2} dm_1 dm_2 \right.$$
$$\left. + \iint_A 0 \; dm_1 dm_2 \right]$$
$$= \left(\frac{2}{n}\right)^2 \sum_{k_1} \sum_{k_2} \int_{0.5k_2}^{0.5(n+k_2)} \int_{0.5k_1}^{0.5(n+k_1)} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1$$
$$-\pi_2)^{n-k_1-k_2} dm_1 dm_2$$
$$= \sum_{k_1} \sum_{k_2} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1-\pi_2)^{n-k_1-k_2}$$
$$= [\pi_1 + \pi_2 + (1-\pi_1-\pi_2)]^n$$
$$= 1.$$

Thus, f is a density function.

Using the trinomial theorem and integral operation again, for the joint moment-generation function, we have the following:

$$M(t_1, t_2)$$
$$= E(e^{t_1 m_1 + t_2 m_2})$$
$$= \left(\frac{2}{n}\right)^2 \sum_{k_1} \sum_{k_2} \int_{0.5k_2}^{0.5(n+k_2)} \int_{0.5k_1}^{0.5(n+k_1)} e^{t_1 m_1} e^{t_2 m_2} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1$$
$$-\pi_2)^{n-k_1-k_2} dm_1 dm_2$$
$$= \left(\frac{2}{n}\right)^2 \left(\frac{e^{\frac{n}{2}t_1} - 1}{t_1}\right) \left(\frac{e^{\frac{n}{2}t_2} - 1}{t_2}\right) \sum_{k_1} \sum_{k_2} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1-\pi_2)^{n-k_1-k_2}$$
$$= \left(\frac{2}{n}\right)^2 \left(\frac{e^{\frac{n}{2}t_1} - 1}{t_1}\right) \left(\frac{e^{\frac{n}{2}t_2} - 1}{t_2}\right) \left[\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + (1-\pi_1-\pi_2)\right]^n \text{ for } (t_1, t_2) \neq (0,0).$$

The moment-generating function is not differentiable at $(t_1, t_2) = (0, 0)$, but can be calculated by taking $\lim_{(t_1,t_2)\to(0,0)}$. We let $t_1 = r\cos\theta$, $t_2 = r\sin\theta$. If $(t_1, t_2) \to (0, 0)$, then $r \to 0^+$.

Hence, we have

$$M(0,0) = \lim_{(t_1,t_2)\to(0,0)} M(t_1, t_2)$$
$$= \lim_{(t_1,t_2)\to(0,0)} \left(\frac{2}{n}\right)^2 \left(\frac{e^{\frac{n}{2}t_1} - 1}{t_1}\right) \left(\frac{e^{\frac{n}{2}t_2} - 1}{t_2}\right) \left[\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + (1-\pi_1-\pi_2)\right]^n$$
$$= \lim_{r\to 0^+} \left(\frac{2}{n}\right)^2 \left(\frac{e^{\frac{n}{2}r\cos\theta} - 1}{r\cos\theta}\right) \left(\frac{e^{\frac{n}{2}r\sin\theta} - 1}{r\sin\theta}\right) \left[\pi_1 e^{\frac{r\cos\theta}{2}} + \pi_2 e^{\frac{r\sin\theta}{2}} + (1-\pi_1-\pi_2)\right]^n$$
$$= 1.$$

Now, let $\psi(t) = \log M(t)$, where $t = (t_1, t_2)$ is a vector.

We can get that

$$\psi(t) = 2\log\left(\frac{2}{n}\right) + \log\left(\frac{e^{\frac{n}{2}t_1} - 1}{t_1}\right) + \log\left(\frac{e^{\frac{n}{2}t_2} - 1}{t_2}\right)$$
$$+ n\log\left[\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + (1-\pi_1-\pi_2)\right].$$

It implies that $\psi(0) = \log M(0) = \log 1 = 0$.

Moreover, let

$$\psi_i(t) = \frac{\partial\psi(t)}{\partial t_i} = \frac{t_i}{e^{\frac{n}{2}t_i} - 1} \cdot \frac{e^{\frac{n}{2}t_i}\frac{n}{2} \cdot t_i - \left(e^{\frac{n}{2}t_i} - 1\right) \cdot 1}{t_i^2} + n \cdot \frac{\pi_i e^{\frac{t_i}{2}}\frac{1}{2}}{\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2},$$

$$\begin{aligned}\psi_{ii}(t) =& \frac{\partial^2\psi(t)}{\partial t_i^2} = \frac{1 \cdot \left(e^{\frac{n}{2}t_i} - 1\right) - t_i \cdot e^{\frac{n}{2}t_i}\frac{n}{2}}{\left(e^{\frac{n}{2}t_i} - 1\right)^2} \cdot \frac{e^{\frac{n}{2}t_i}\frac{n}{2} \cdot t_i - \left(e^{\frac{n}{2}t_i} - 1\right)}{t_i^2} \\ &+ \frac{t_i}{e^{\frac{n}{2}t_i} - 1} \cdot \frac{\left[e^{\frac{n}{2}t_i}(\frac{n}{2})^2 \cdot t_i + e^{\frac{n}{2}t_i}\frac{n}{2} \cdot 1 - e^{\frac{n}{2}t_i}\frac{n}{2}\right] \cdot t_i^2 - \left[e^{\frac{n}{2}t_i}\frac{n}{2} \cdot t_i - \left(e^{\frac{n}{2}t_i} - 1\right)\right] \cdot 2t_i}{t_i^4} \\ &+ n \cdot \frac{\pi_i e^{\frac{t_i}{2}}\left(\frac{1}{2}\right)^2\left(\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2\right) - \left(\pi_i e^{\frac{t_i}{2}}\frac{1}{2}\right)^2}{\left(\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2\right)^2}\end{aligned}$$

and $\psi_{ij}(t) = \dfrac{\partial^2\psi(t)}{\partial t_j \partial t_i} = \dfrac{\partial}{\partial t_j}\psi_i(t) = n \cdot \dfrac{-\pi_i e^{\frac{t_i}{2}}\frac{1}{2} \cdot \pi_j e^{\frac{t_j}{2}}\frac{1}{2}}{\left(\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2\right)^2}$.

We can get that

$$\mu = EM_i = \psi_i(0) = \lim_{t\to 0}\psi_i(t) = n \cdot \frac{1 + 2\pi_i}{4},$$
$$\sigma^2 = Var(M_i) = \psi_{ii}(0) = \lim_{t\to 0}\psi_{ii}(t) = \frac{n^2}{48} + \frac{n}{4}\pi_i(1 - \pi_i),$$

and $cov(M_1, M_2) = \psi_{12}(0) = \lim_{t\to 0}\psi_{12}(t) = -\dfrac{n}{4}\pi_1\pi_2$.

Theorem 3.2. *Let $(M_1, M_2) \sim FT(n, (\pi_1, \pi_2))$ be a fuzzy trinomial distribution with means π_1 and π_2. Then $M_1 \sim FB(n, \pi_1)$ and $M_2 \sim B(n, \pi_2)$.*

Proof: The marginal moment-generating function of M_1 is calculated as follows:

$$\begin{aligned}M_1(t) &= M(t, 0) \\ &= \lim_{(t_1,t_2)\to(t,0)}\left(\frac{2}{n}\right)^2\left(\frac{e^{\frac{n}{2}t_1} - 1}{t_1}\right)\left(\frac{e^{\frac{n}{2}t_2} - 1}{t_2}\right)\left[\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + (1 - \pi_1 - \pi_2)\right]^n \\ &= \left(\frac{2}{n}\right)^2\left(\frac{e^{\frac{n}{2}t} - 1}{t}\right)\left[\pi_1 e^{\frac{t}{2}} + \pi_2 e^0 + (1 - \pi_1 - \pi_2)\right]^n \lim_{t_2\to 0}\left(\frac{e^{\frac{n}{2}t_2} - 1}{t_2}\right) \\ &= \left(\frac{2}{n}\right)\left(\frac{e^{\frac{n}{2}t} - 1}{t}\right)\left[\pi_1 e^{\frac{t}{2}} + (1 - \pi_1)\right]^n.\end{aligned}$$

This expression is the moment-generating function for $FB(n, \pi_1)$, so $M_1 \sim FB(n, \pi_1)$. The proof for M_2 can be done in a similar way.

Next, we discuss the fuzzy trinomial distribution that will lead to the notation used in the fuzzy multinomial distribution. Let $(M_1, M_2) \sim FT(n, (\pi_1, \pi_2))$, and let $M_3 = n - M_1 - M_2$ and $\pi = 1 - \pi_1 - \pi_2$. Then $M = (M_1, M_2, M_3)$ has the joint density function

$$f(m_1, m_2, m_3) = 2\begin{cases}\left(\dfrac{2}{n}\right)^3 \displaystyle\sum_{k_1}\sum_{k_2}\sum_{k_3}\frac{n!}{k_1!k_2!k_3!}\pi_1^{k_1}\pi_2^{k_2}\pi_3^{k_3}, & \text{if } (k_1, k_2, k_3) \in K_n^* \\ \left(\dfrac{2}{n}\right)^3 \displaystyle\sum_{k_1}\sum_{k_2}\sum_{k_3}\frac{n!}{k_1!k_2!k_3!}\pi_1^{k_1}\pi_2^{k_2}\pi_3^{k_3}, & \text{if } (k_1, k_2, k_3) \in K - K_n^*\end{cases},$$

where $K = \{(k_1, k_2, k_3) : k_1 \geq 0, k_2 \geq 0, k_3 \geq 0$ and $k_1 + k_2 + k_3 \leq n\}$ and where $(m_1, m_2, m_3) \in S_n^*$, S_n^* is denoted as $S_n^* = \{(m_1, m_2, m_3) : m_1 \geq 0, m_2 \geq 0, m_3 \geq 0$ and $m_1 + m_2 + m_3 = n\}$.

Also, let $K_n^* = \{(k_1, k_2, k_3) : 2m_i - n < k_i \leq 2m_i, k_i \in N \cup \{0\}$ for $i = 1, 2, 3$ and $k_1 + k_2 + k_3 = n\}$ under the condition $(m_1, m_2, m_3) \in S_n^*$. Then we have a relation between K_n^* and S_n^*. If m_1, m_2 and m_3 decide, k_1, k_2 and k_3 are decided. It is straightforward to show that (M_1, M_2, M_3) has a joint moment-generating function

$$M(t_1, t_2, t_3) = \begin{cases} \left(\dfrac{2}{n}\right)^3 \left(\dfrac{e^{\frac{n}{2}t_1} - 1}{t_1}\right)\left(\dfrac{e^{\frac{n}{2}t_2} - 1}{t_2}\right)\left(\dfrac{e^{\frac{n}{2}t_3} - 1}{t_3}\right)\left(\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + \pi_3 e^{\frac{t_3}{2}}\right)^n, & \text{if } (t_1, t_2, t_3) \neq (0, 0, 0) \\ 1, & \text{if } (t_1, t_2, t_3) = (0, 0, 0) \end{cases}$$

Note that the joint density function and joint moment-generating function of (M_1, M_2, M_3) are somewhat nicer than those of (M_1, M_2). Also notice that the density functions of (M_1, M_2) and (M_1, M_2, M_3) represent the same model, in which we have n independent replications of an experiment with three possible outcomes.

When $(M_1, M_2) \sim FT(n, (\pi_1, \pi_2))$, the joint distribution of M_1, M_2 and $M_3 = n - M_1 - M_2$ is a special case of the fuzzy multinomial distribution discussed in the following analysis. In this case, we often say that $M = (M_1, M_2, M_3)$ has a three-dimensional fuzzy multinomial distribution and we write $(M_1, M_2, M_3) \sim FM_3(n, (\pi_1, \pi_2, \pi_3))$, where $\pi_3 = 1 - \pi_1 - \pi_2$.

3.2. **Fuzzy multinomial distribution.** We have already considered the situations that involve two and three random variables. Now, we want to extend it to k random variables.

Let $M = (M_1, \ldots, M_k)$ be a k-dimensional random vector with range $S_n = \{(m_1, \ldots, m_k) : m_1 \geq 0, \ldots, m_k \geq 0$ and $m_1 + \cdots + m_k = n\}$. That is, M_i are nonnegative fuzzy-valued random variables whose sum is n. $M = (M_1, \ldots, M_k)$ has a k-dimensional fuzzy multinomial distribution with parameters n and $\pi = (\pi_1, \ldots, \pi_k)$. We write $(M_1, \ldots, M_k) \sim FM_k(n, \pi)$ if M has joint density function

$$f(m_1, \ldots, m_k) = \zeta \begin{cases} \left(\dfrac{2}{n}\right)^k \displaystyle\sum_{k_1} \cdots \sum_{k_k} \frac{n!}{k_1! \ldots k_k!} \pi_1^{k_1} \pi_2^{k_2} \ldots \pi_k^{k_k}, & \text{if } (k_1, k_2, \ldots, k_k) \in K_n \\ \left(\dfrac{2}{n}\right)^k \displaystyle\sum_{k_1} \cdots \sum_{k_k} \frac{n!}{k_1! \ldots k_k!} \pi_1^{k_1} \pi_2^{k_2} \ldots \pi_k^{k_k}, & \text{if } (k_1, k_2, \ldots, k_k) \in K - K_n \end{cases} \tag{6}$$

where $K = \{(k_1, \ldots, k_k) : k_i \geq 0$ for $i = 1, 2, \ldots, k$ and $\sum_{i=1}^k k_i = n\}$, $M \in S_n$, and $\zeta = dim(k-1)$, and where n is a positive integer and π_i are constants such that $\sum_{i=1}^k \pi_k = 1$. Moreover, $K_n = \{(k_1, \ldots, k_k) : 2m_i - n < k_i \leq 2m_i, k_i \in N \cup \{0\}$ for $i = 1, \ldots, k$ and $\sum_{i=1}^k k_i = n\}$. Note that $\sum_{i=1}^k M_i = n$ and hence, $M_k = n - \sum_{i=1}^{k-1} M_i$ and $\pi_k = 1 - \sum_{i=1}^{k-1} \pi_i$. Also note that

$$(M_1, M_2) \sim FM_2(n, (\pi_1, \pi_2)) \Leftrightarrow M_1 \sim FB(n, \pi_1), \quad M_2 = n - M_1,$$

and $(M_1, M_2, M_3) \sim FM_3(n, (\pi_1, \pi_2, \pi_3)) \Leftrightarrow (M_1, M_2) \sim FT(n, (\pi_1, \pi_2), M_3 = n - M_1 - M_2$.

The following theorems illustrate important properties of the fuzzy multinomial distribution.

Theorem 3.3. *The fuzzy multinomial density function in (6) is a joint density function for all positive integers n and for all values $\pi_1, \ldots, \pi_k$ such that $\pi_i \geq 0$ and $\sum_{i=1}^{k} \pi_i = 1$.*

Let $M \sim FM_k(n, \pi)$, where $M = (M_1, \ldots, M_k)$, $\pi = (\pi_1, \ldots, \pi_k)$, $\sum_{i=1}^{k} M_i = n$ and $\sum_{i=1}^{k} \pi_i = 1$. Then, $E(M_i) = n \cdot \frac{1+2\pi_i}{4}$, $Var(M_i) = \frac{n^2}{48} + \frac{n}{4}\pi_i(1-\pi_i)$, $cov(M_i, M_j) = -\frac{n}{4}\pi_i\pi_j$ and the joint moment-generation function can be written as

$$M(t_1, \ldots, t_k) = \begin{cases} \left(\frac{2}{n}\right)^k \left(\frac{e^{\frac{n}{2}t_1} - 1}{t_1}\right) \cdots \left(\frac{e^{\frac{n}{2}t_k} - 1}{t_k}\right) \left(\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + \cdots + \pi_k e^{\frac{t_k}{2}}\right)^n, & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } (t_1, \ldots, t_k) \neq (0, \ldots, 0) & \\ 1, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if } (t_1, \ldots, t_k) = (0, \ldots, 0) & \end{cases}.$$

If $M \sim FM_k(n, \pi)$, then $M_i \sim FB(n, \pi_i)$ and $(M_i, M_j) \sim FT(n, (\pi_i, \pi_j))$.

Proof: The proof is given in the same way as those for Theorems 3.1 and 3.2.

The next theorem gives a normal approximation, which has a widely range of uses.

Theorem 3.4. *Let $X_i \sim FB(1, \pi)$ and $\bar{X}_n = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{M}{n}$, where $M \sim FB(n, \pi)$ and $M = \sum_{i=1}^{n} x_i$. Suppose that $\mu = E(X_i)$ is finite and that $\sigma^2 = Var(X_i) < \infty$. Then $\frac{M - n \cdot \frac{1+2\pi}{4}}{\left[\frac{n^2}{48} + \frac{n}{4}\pi(1-\pi)\right]^{\frac{1}{2}}} \xrightarrow{d} N(0,1)$ as $n \to \infty$.*

Proof: Because $X_i \sim FB(1, \pi)$, we have the following:

$$\mu = E(X_i) = \frac{1+2\pi}{4} \text{ and } \sigma^2 = Var(X_i) = \frac{1}{48} + \frac{1}{4}\pi(1-\pi).$$

Moreover, $\bar{X}_n = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{M}{n}$, where $M \sim FB(n, \pi)$, $\mu = E(M) = n \cdot \frac{1+2\pi}{4}$ and $Var(M) = \frac{n^2}{48} + \frac{n}{4}\pi(1-\pi)$.

Hence,

$$\mu = E(\bar{X}_n) = E\left(\frac{M}{n}\right) = \frac{1}{n}E(M) = \frac{1}{n} \cdot \left(n \cdot \frac{1+2\pi}{4}\right) = \frac{1+2\pi}{4},$$

and $Var(\bar{X}_n) = Var\left(\frac{M}{n}\right) = \frac{1}{n^2}Var(M) = \frac{1}{n^2}\left[\frac{n^2}{48} + \frac{n}{4}\pi(1-\pi)\right] = \frac{1}{48} + \frac{1}{4n}\pi(1-\pi)$.

Using the central limit theorem, we have the following

$$\frac{\bar{X}_n - \mu}{\left[\frac{1}{48} + \frac{1}{4n}\pi(1-\pi)\right]^{\frac{1}{2}}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$

Hence, $\frac{M - n \cdot \frac{1+2\pi}{4}}{\left[\frac{n^2}{48} + \frac{n}{4}\pi(1-\pi)\right]^{\frac{1}{2}}} \xrightarrow{d} N(0,1)$ as $n \to \infty$.

We have introduced some new distributions based on fuzzy theory. Now, we can use these distributions to derive a fuzzy chi-square statistic for goodness-of-fit.

3.3. Goodness-of-fit for membership functions with fuzzy data. The l-sample fuzzy multinomial model

Let M_i be independent l-dimensional random vectors ($M_i \sim FM_k(n_i, \pi_i)$) where n_i are integers and π_i are unknown parameter vectors. This model is called the l-sample fuzzy multinomial model. The primary purpose of this model is to test the equality of all π_i. Let M_{ij} be the jth component of M_i, let π_{ij} be the jth component of π_i and let L_j denote linguistic variables for $j = 1, 2, \ldots, k$. Table 1 shows these notations.

TABLE 1. The table of membership M_{ij} in L_j

	L_1	L_2	$\ldots$	L_k	Total
M_1	M_{11}	M_{12}	$\ldots$	M_{1k}	$M_{1\cdot} = n_1$
M_2	M_{21}	M_{22}	$\ldots$	M_{2k}	$M_{2\cdot} = n_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
M_l	M_{l1}	M_{l2}	$\ldots$	M_{lk}	$M_{l\cdot} = n_l$
Total	$M_{\cdot 1}$	$M_{\cdot 2}$	$\ldots$	$M_{\cdot k}$	$N = \sum_{i=1}^{l} n_i$

Theorem 3.5. *A_{ij} is an unbiased estimator of π_{ij} for the l-sample fuzzy multinomial model M_{ij}, where $A_{ij} = \dfrac{2M_{ij}}{n_i} - \dfrac{1}{2}$.*

Proof: We know that $M_i \sim FM_k(n_i, \pi_i)$ implies that $M_{ij} \sim FB(n_i, \pi_{ij})$. Thus, we have $E(M_{ij}) = n_i \cdot \dfrac{1 + 2\pi_{ij}}{4}$ and $Var(M_{ij}) = \dfrac{n_i^2}{48} + \dfrac{n_i}{4}\pi_{ij}(1 - \pi_{ij})$. We can then derive $E\left(\dfrac{1}{2}\left(\dfrac{4M_{ij}}{n_i} - 1\right)\right) = E\left(\dfrac{2M_{ij}}{n_i} - \dfrac{1}{2}\right) = E(A_{ij}) = \pi_{ij}$. Hence, A_{ij} is an unbiased estimation of π_{ij}.

Now, we test $H_0 : \pi_1 = \pi_2 = \cdots = \pi_l$ against H_1: H_0 not being true. Under the null hypothesis H_0, the π_i values are all equal, so we let

$$\pi_1 = \pi_2 = \cdots = \pi_l = \pi_0, \text{ where } \pi_0 = (\pi_{01}, \pi_{02}, \ldots, \pi_{0k})'.$$

Therefore, a sensible estimator for the expected frequency for the jth cell in the ith sample is

$$\hat{E}_{ij} = n_i \cdot \frac{1 + 2\hat{\pi}_{0j}}{4} = n_i \cdot \frac{1}{4}\left[1 + 2\left(\frac{2M_{\cdot j}}{N} - \frac{1}{2}\right)\right] = n_i \cdot \frac{M_{\cdot j}}{N},$$

where $M_{\cdot j} = \sum_i M_{ij}$ and $N = \sum_i n_i$.

Let

$$\hat{U}_k = \sum_{i=1}^{l}\sum_{j=1}^{k-1}\left\{\frac{(M_{ij} - \hat{E}_{ij})^2}{\hat{B}_{ij}} + \frac{4\left[\sum_{j=1}^{k-1}(M_{ij} - \hat{E}_{ij})\left(1 - \frac{n_i^2}{48\hat{B}_{ij}}\right)\right]^2}{n_i - 4\sum_{j=1}^{k-1}\hat{B}_{ij}\left(1 - \frac{n_i^2}{48\hat{B}_{ij}}\right)^2}\right\},$$

where $\hat{B}_{ij} = \dfrac{1}{2}\hat{E}_{ij} - \dfrac{n_i}{8} + \dfrac{n_i^2}{48}$.

Here $\hat{U}_k$ is a fuzzy chi-square distribution and has $(l-1)(k-1)$ degrees of freedom.

Because the distribution of $\hat{U}_k$ is approximately $\chi^2(l-1)(k-1)$, we will reject H_0 if $\hat{U}_k \geq \chi^2_\alpha(l-1)(k-1)$, where α is the desired significance level of the test.

Theorem 3.6. $\hat{U}_k \xrightarrow{d} \chi^2(l-1)(k-1)$

Proof: It is enough for us to show that $U_{k_n} \xrightarrow{d} \chi^2(k-1)$.

Let $M_n = (M_{n1}, M_{n2}, \ldots, M_{nk})'$, $\pi = (\pi_1, \pi_2, \ldots, \pi_k)'$, $E_n = (E_{n1}, E_{n2}, \ldots, E_{nk})'$, and V be the $K \times K$ matrix, whose ith diagonal element is $V_{ii} = \frac{1}{4}\pi_i(1-\pi)$ and whose (i,j)th off-diagonal element is $V_{ij} = -\frac{1}{4}\pi_i\pi_j$. So, $Var(X_i) = \frac{1}{48} + \frac{1}{4}\pi_i(1-\pi_i)$, and $cov(X_i, X_j) = -\frac{1}{4}\pi_i\pi_j$, for $i \neq j$.

First, let us show that $n^{-1}(M_n - E_n)$ is approximately $N_k\left(0, \frac{V}{n} + \frac{1}{48}I_k\right)$. Because $M_n \sim FM_k(n, \pi)$, where $\pi = (\pi_1, \pi_2, \ldots, \pi_k)'$ and $M_n = n\bar{X}_n$. We can get that $E(\bar{X}_n) = \frac{1+2\pi}{4}$ and $Var(\bar{X}_n) = \frac{V}{n} + \frac{1}{48}I_k$. Therefore, $E_n = E(M_n) = n \cdot \frac{1+2\pi}{4}$ and $Var(M_n) = n \cdot V + \frac{n^2}{48}I_k$. Using the multinomial central limit theorem, $\bar{X}_n - \frac{1+2\pi}{4}$ is approximately $N_k\left(0, \frac{V}{n} + \frac{1}{48}I_k\right)$. We can get that $\frac{n\bar{X}_n - n\frac{1+2\pi}{4}}{n}$ is approximately $N_k\left(0, \frac{V}{n} + \frac{1}{48}I_k\right)$. Hence, $n^{-1}(M_n - E_n)$ is approximately $N_k\left(0, \frac{V}{n} + \frac{1}{48}I_k\right)$.

Because V is not invertible, let M_n^* and E_n^* be the $(k-1)$-dimensional vectors and let V^* be the $(k-1) \times (k-1)$-dimensional matrix. Then, $n^{-1}(M_n^* - E_n^*)$ is approximately $N_{k-1}\left(0, \frac{V^*}{n} + \frac{1}{48}I_{k-1}\right)$.

We know that [1]

$$T_n = [n^{-1}(M_n^* - E_n^*)]' \left(\frac{V^*}{n} + \frac{1}{48}I_{k-1}\right)^{-1} [n^{-1}(M_n^* - E_n^*)] \sim \chi^2(k-1),$$

which implies that

$$T_n = (M_n^* - E_n^*)' \left(nV^* + \frac{n^2}{48}I_{k-1}\right)^{-1} (M_n^* - E_n^*) \sim \chi^2(k-1),$$

where $nV^* + \dfrac{n^2}{48}I_{k-1} = \dfrac{1}{4}D_{k-1} + \dfrac{n^2}{48}I_{k-1} - \dfrac{1}{4n}F_n^*F_n^{*\prime}$.

Let

$$D_{k-1} = \begin{bmatrix} n\pi_1 & 0 & \ldots & 0 \\ 0 & n\pi_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & n\pi_{k-1} \end{bmatrix}$$

and $F_n = (n\pi_1, n\pi_2, \ldots, n\pi_k)'$.

We also know the following [1]:

If A is a $q \times q$ invertible symmetric matrix, b and c are q-dimensional vectors, and $d \neq 0$ is a number, then $c'(A - d^{-1}bb')^{-1}c = c'A^{-1}c + \dfrac{(c'A^{-1}b)^2}{d - b'A^{-1}b}$. Then, we let $C_n = M_n^* - E_n^*$, $A = \dfrac{1}{4}D_{k-1} + \dfrac{n^2}{48}I_{k-1}$, $b = F_n^*$ and $d = 4n$. Hence, we can get that

$$\begin{aligned} T_n =& (M_n^* - E_n^*)' \left(nV^* + \frac{n^2}{48}I_{k-1}\right)^{-1} (M_n^* - E_n^*) \\ =& (M_n^* - E_n^*)' \left(\frac{1}{4}D_{k-1} + \frac{n^2}{48}I_{k-1}\right)^{-1} (M_n^* - E_n^*) + \frac{\left[(M_n^* - E_n^*)' \left(\frac{1}{4}D_{k-1} + \frac{n^2}{48}I_{k-1}\right)^{-1} F_n^*\right]^2}{4n - F_n^{*\prime} \left(\frac{1}{4}D_{k-1} + \frac{n^2}{48}I_{k-1}\right)^{-1} F_n^*} \\ =& \sum_{i=1}^{k-1} \left[\frac{(M_{ni} - E_{ni})^2}{B_{ni}} + \frac{4\left[\sum_{i=1}^{k-1}(M_{ni} - E_{ni})\left(1 - \frac{n^2}{48B_{ni}}\right)\right]^2}{n - 4\sum_{i=1}^{k-1} B_{ni}\left(1 - \frac{n^2}{48B_{ni}}\right)^2} \right], \end{aligned}$$

where $B_{ni} = \dfrac{1}{2}E_{ni} - \dfrac{n}{8} + \dfrac{n^2}{48}$.

Hence, $U_{k_n} = T_n \xrightarrow{d} \chi^2(k-1)$.

To compute the degrees of freedom in Theorem 3.6, note that there are $(l-1)$ degrees of freedom for each of the k populations. Thus, there are $k(l-1)$ degrees of freedom for the whole model. Under the null hypothesis, we estimate $(l-1)$ independent parameters, the components of π_0. (Note that $\sum \pi_{0j} = 1$.) Therefore, we would expect the degrees of freedom for this hypothesis to be $k(l-1)-(l-1) = (k-1)(l-1)$.

4. **Empirical Studies.**

Example 4.1. *A manager wants to have a dinner party with his staffs at the end of this year. He wants to know whether the different sexes will make difference choices about dinner style. He conducts a sampling survey and randomly chooses 100 samples (50 males and 50 females) from the company. During the answering process, people are asked to reply via two methods: a conventional reply and a fuzzy response. In the conventional reply, people can only choose one answer. In the fuzzy response, people can answer the question with percentages (the total percentage is 100%). For instance, a staff could respond that he has a preference 70% for Chinese style, 20% for Japanese style and 10% for Korean style. The manager then sums up the percentages and the results are as follows:*

Table 2. Respose in conventional way

Style Voter	Chinese Style	Japanese Style	Korean Style	Thailande Style	others	Chi-square test of homogeneity
Male	27	12	2	3	6	$\chi^2 = 3.46 >$
Female	26	14	2	0	8	$0.71 = \chi^2_{0.95}(4)$

Table 3. Respose in fuzzy way

Style Voter	Chinese Style	Japanese Style	Korean Style	Thailande Style	others	Chi-square test of homogeneity
Male	19.86	16.09	4.45	3.6	6	$\chi^2 = 0.18 >$
Female	18.7	13.5	6.82	4.29	6.69	$0.71 = \chi^2_{0.95}(4)$

The null hypothesis H_0 is that there is no difference in dinner style choice against the alternative hypothesis H_1 is that H_0 is not being true. With a significance level $\alpha = 0.95$, we can see the two results from the different responses are different. For the conventional response, we reject the null hypothesis, while for the fuzzy reply, we accept the null hypothesis. In other words, for the conventional reply, sex affects the response about dinner style, while for the fuzzy reply, sex does not affect the response about dinner style.

Although we did not illustrate the feasibility of the test statistic in this paper, we provided a method to evaluate fuzzy questionnaire which is more approaching to human though. The statistic test we proposed in this paper is more flexible with fuzzy numbers than a corresponding statistic test with real numbers.

5. **Conclusions.** We seldom use the fuzzy survey in the social sciences. One of the reasons is that it is difficult to find an appropriate fuzzy testing process. In this paper, we provided a formula, called fuzzy chi-square test to deal with fuzzy data. We used the fuzzy binomial distribution to find expected value and variance. Hence, we could find the estimator for π_{ij} in l-sample fuzzy multinomial model. Moreover, we used the central limit theorem to obtain an approximately normal distribution. We used a proof similar to that

of the conventional Pearson's chi-square test to describe the fuzzy chi-square teat. We also presented an example in Section 4 where we used two methods, the conventional chi-square test and the fuzzy chi-square test, to test the hypothesis. Outstanding questions that will lead to future improvements are as follows:

1. Does the size of sample affect the sensitivity of the result?
2. How can we prove that A_{ij} is the best estimator for π_{ij}?
3. Although we proved the fuzzy chi-square test in this paper, the result is a little complex and hard to calculate.

REFERENCES

[1] S. F. Arnold, *Mathematical Statistics*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[2] G. Casalino, F. Memola and C. Minutolo, A model for evaluation of laser welding efficiency and quality using an artificial neural network and fuzzy logic, *Proc. of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol.218, no.6, pp.641-646, 2004.

[3] J.-F. Chang, B.-M. Hsu, M.-H. Shu and C.-S. Yang, Fuzzy inference for assessing process lifetime performance, *International Journal of Innovative Computing, Information and Control*, vol.3, no.6(B), pp.1729-1742, 2007.

[4] A. O. Esogbue and Q. Song, On optimal defuzzification and learning algorithms: Theory and applications, *Fuzzy Optimization and Decision Making*, vol.2, no.4, pp.283-296, 2003.

[5] R. A. Johnson and B. Gourik, *Statistics: Principles and Methods*, 2nd Edition, Wiley, New York, 1992.

[6] H. Kwakernaak, (a) Fuzzy random variables: Definitions and theorems, *Information Sciences*, vol.15, no.1, pp.1-29, 1978.

[7] H. Kwakernaak, (b) Fuzzy random variables: Algorithms and examples for the discrete case, *Information Sciences*, vol.17, no.3, pp.253-278, 1979.

[8] Y. Liu, Z. Qiao and G. Wang, Fuzzy random reliability of structures based on fuzzy random variables, *Fuzzy Sets and Systems*, vol.86, no.3, pp.345-355, 1997.

[9] P.-C. Lin, B. Wu and J. Watada, Kolmogorov-Smirnov two sample test with continuous fuzzy data, *Integrated Uncertainty Management and Applications, AISC*, vol.68, pp.175-186, 2010.

[10] P.-C. Lin, J. Watada and B. Wu, Database of fuzzy probability distribution function and its application, *The International Symposium on Innovative Management, Information and Production*, Hangzhou, China, 2010.

[11] H. Nguyen and B. Wu, *Fundamentals of Statistics with Fuzzy Data*, Springer, Netherlands, 2006.

[12] K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophy Magazine Series*, vol.5, no.50, pp.157-172, 1990.

[13] M. L. Puri and D. Ralescu, Fuzzy random variables, *Journal of Mathematical Analysis and Application*, vol.114, no.2, pp.409-422, 1986.

[14] C. Sun, Soft computing and statistical analysis for lexical semantics, *International Journal of Innovative Computing, Information and Control*, vol.4, no.10, pp.2731-2747, 2008.

[15] B. Wu and S. K. Chang, On testing hypothesis of fuzzy sample mean, *Japan Journal of Industrial and Applied Mathematics*, vol.24, no.2, pp.197-209, 2007.

[16] B. Wu and C. Sun, Interval-valued statistics, fuzzy logic, and their use in computational semantics, *Journal of Intelligent and Fuzzy Systems*, vol.11, no.1-2, pp.1-7, 2001.

[17] L. A. Zadeh, Fuzzy sets, *Information and Control*, vol.8, no.3, pp.338-353, 1965.