3. Fuzzy Statistic Distribution

Before introducing new statistic distribution, we first define that how to find the expected value and variance for fuzzy sample data.

3.1 Expected Value and Variance for Fuzzy Sample Data

Definition 3.1 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*

$$\{Fx_i = \frac{m_{i1}}{L_1} + \frac{m_{i2}}{L_2} + \cdots + \frac{m_{ik}}{L_k}, i = 1, 2, \ldots, n\}$$

be a sequence of random fuzzy sample on U *,*

$$m_{ij} \ \ (\sum_{j=1}^{k} m_{ij} = 1) \text{ is the memberships with respect to } L_j$$

(*Nguyen and Wu 2006* [*8*]) *and has the fuzzy Bernoulli distribution. Then, the expected value for fuzzy sample data is defined as*

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

Definition 3.2 Variance for fuzzy sample data

As definition above, we have the variance for fuzzy sample data as following:

$$\text{var}(Fx_i) = \frac{\text{var}(m_{i1})}{L_1} + \frac{\text{var}(m_{i2})}{L_2} + \cdots \frac{\text{var}(m_{ik})}{L_k}$$

3.2 Fuzzy Bernoulli and Fuzzy Binomial Distribution

In this section, we want to introduce some new distribution functions. We have known that (e.g. [5]) a Bernoulli trial is an experiment which has only two possible (incompatible) outcomes, which we shall label "success" and "failure". In general, let $X = 1$ if the outcome of Bernoulli trial is a success and $X = 0$ if it is a failure. Now, we say that a Fuzzy Bernoulli experiment is a random experiment, the outcome of which can be classified in but one of two mutually exclusive and exhaustive ways, say, 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 on each trial. In addition, we will frequently let $q = 1 - \pi$ denote the probability of failure.

Now, let X be a continuous random variable associated with a Fuzzy Bernoulli trial by defining it as follows:

$$X\,(\text{success}) \in [0.5, 1] \quad \text{and} \quad X\,(\text{failure}) \in [0, 0.5)$$

That is, the two outcomes, success and failure, are denoted by mutually part of a partition set $[0, 1]$.

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

$$f(x) = 2\begin{cases} \pi & \text{if } x \in [0.5, 1] \\ 1-\pi & \text{if } x \in [0, 0.5) \end{cases} \tag{3.1}$$

We say that X has a Fuzzy Bernoulli distribution, and denoted by $X \sim FB(1,\pi)$.

We fist derive some properties of the fuzzy Bernoulli distribution.

Theorem 3.3

a. *The Fuzzy Bernoulli density function given in* (*3.1*) *is a density function.*

b. *If* $X \sim FB(1,\pi)$, *then the expected value of* X *is*

$$\mu = \quad E(X) = \frac{1+2\pi}{4},$$

and the variance of X *is*

$$\sigma^2 = \quad Var(X) = \frac{1}{48} + \frac{1}{4}\pi(1-\pi).$$

Finally, the moment-generating function of X *is*

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

Proof:

a. Note that $f(x) \geq 0$.

Also, $\int_0^1 f(x)dx$

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

So that f is a density function.

b. $E(X)=\int_{0.5}^{1} x\cdot 2\pi\,dx+\int_{0}^{0.5} x\cdot 2(1-\pi)\,dx=\dfrac{1+2\pi}{4}$,

$$Var(X)=E(X^2)-[E(X)]^2=\int_{0.5}^{1} x^2\cdot 2\pi\,dx+\int_{0}^{0.5} x^2\cdot 2(1-\pi)\,dx-(\frac{1+2\pi}{4})^2$$

$=\dfrac{1}{48}+\dfrac{1}{4}\pi(1-\pi)$, and the moment-generating function of X is

$$M(t)=E(e^{tX})=\int_{0.5}^{1} e^{tx}\cdot 2\pi\,dx+\int_{0}^{0.5} e^{tx}\cdot 2(1-\pi)\,dx$$

$$=2\pi\cdot\frac{1}{t}e^{tx}\Big|_{0.5}^{1}+2(1-\pi)\frac{1}{t}e^{tx}\Big|_{0}^{0.5}$$

$$=2\pi\cdot\frac{1}{t}(e^{t}-e^{\frac{t}{2}})+2(1-\pi)\cdot\frac{1}{t}(e^{\frac{t}{2}}-1)$$

$$=2\pi\frac{e^{t}-e^{\frac{t}{2}}}{t}+2(1-\pi)\frac{e^{\frac{t}{2}}-1}{t}$$

$$=2\cdot\frac{1}{t}[\pi e^{t}+(1-2\pi)e^{\frac{t}{2}}-(1-\pi)]$$

$$=2\cdot\frac{1}{t}(\pi e^{\frac{t}{2}}+(1-\pi))(e^{\frac{t}{2}}-1)$$

$$=2(\frac{e^{\frac{t}{2}}-1}{t})[\pi e^{\frac{t}{2}}+(1-\pi)] \quad \text{for} \quad t\neq 0.$$

The moment-generating function is not differentiable at zero, but the moments can be calculated by differentiating and then taking $\lim_{t\to 0}$. We present it as following:

$$M(0)=\lim_{t\to 0}M(t)=\lim_{t\to 0}2(\frac{e^{\frac{t}{2}}-1}{t})[\pi e^{\frac{t}{2}}+(1-\pi)]=\lim_{t\to 0}2\frac{(e^{\frac{t}{2}}-1)[\pi e^{\frac{t}{2}}+(1-\pi)]}{t}$$

$$=\lim_{t\to 0}\{2(e^{\frac{t}{2}}\cdot\frac{1}{2})[\pi e^{\frac{t}{2}}+(1-\pi)]+2(e^{\frac{t}{2}}-1)[\pi e^{\frac{t}{2}}\cdot\frac{1}{2}]\} \quad \text{(by L'Hospital's Rule)}$$

$=1$.

In a sequence of Fuzzy Bernoulli trials, we are often interested in the total number of successes and not in the order of their occurrence. If we let the random variable M equal the number of observed successes in n Fuzzy Bernoulli trials, the possible values of M are any nonnegative numbers. In order to easily denote the Fuzzy Binomial distribution, let k successes occur, where $2m-n<k\leq 2m$ for $k\in N\cup\{0\}$ as $m<n$ and $k=n$ as $m=n$, then $n-k$ failures occur. On the above, we say that m is the observed numbers of M and N is defined by natural numbers. (The same definitions are in the following.) The number of ways of

selecting 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 is decided (see *Figure 3.1*).

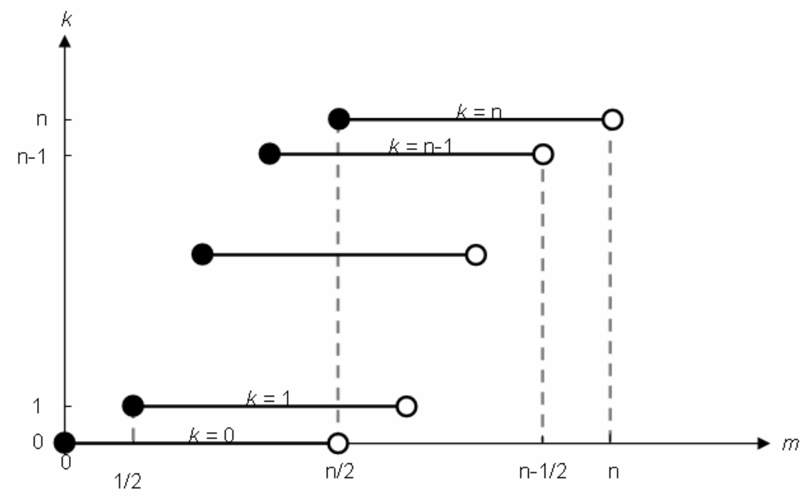

Figure 3.1. The relation of m and k.

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

$$f(m) = \frac{2}{n}\sum_{k\in\Omega}\binom{n}{k}\pi^k(1-\pi)^{n-k}, \tag{3.2}$$

where $\Omega = \{k \in N \cup \{0\} \,|\, 2m - n < k \le 2m \ for\ m < n \ and\ m = n \ for\ k = n\}$.

Another way to present the p.d.f. is like that

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

or

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

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

Theorem 3.4

a. The Fuzzy Bernoulli density function given in (3.2) is a density function.

b. 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

$$\sigma^2 = Var(M) = \frac{n^2}{48}+\frac{n}{4}\pi(1-\pi).$$

Finally, the moment-generating function of M is

$$M(t)=\begin{cases}\frac{2}{n}\cdot\frac{e^{\frac{n}{2}t}-1}{t}[\pi e^{\frac{t}{2}}+(1-\pi)]^n & if\ t\neq 0\\ 1 & if\ t=0\end{cases}.$$

Proof:

a. Note that, $f(m)\ge 0$.

Also, by the binomial theorem and integral operation,

$$\int_0^n f(m)dm=\int_0^n \frac{2}{n}\sum_{k\in\Omega}\binom{n}{k}\pi^k(1-\pi)^{n-k}dm$$

$$=\frac{2}{n}\{\int_0^{0.5n}\binom{n}{0}\pi^0(1-\pi)^{n-0}dm+\int_{0.5}^{0.5(n+1)}\binom{n}{1}\pi^1(1-\pi)^{n-1}dm+\cdots+\int_{0.5n}^{0.5(n+n)}\binom{n}{n}\pi^n(1-\pi)^{n-n}dm\}$$

$$=\frac{2}{n}\sum_{k=0}^{n}\int_{0.5k}^{0.5(n+k)}\binom{n}{k}\pi^k(1-\pi)^{n-k}dm=\frac{2}{n}\sum_{k=0}^{n}\binom{n}{k}\pi^k(1-\pi)^{n-k}\cdot m\Big|_{0.5k}^{0.5(n+k)}$$

$$=\frac{2}{n}\sum_{k=0}^{n}\binom{n}{k}\pi^k(1-\pi)^{n-k}\cdot\frac{n}{2}=\sum_{k=0}^{n}\binom{n}{k}\pi^k(1-\pi)^{n-k}=[\pi+(1-\pi)]^n=1.$$

So that f is a density function.

b. By the binomial theorem and integral operation again,

$$M(t) = E(e^{tM}) = \frac{2}{n}\sum_{k=0}^{n}\int_{0.5k}^{0.5(n+k)} e^{tm}\cdot\binom{n}{k}\pi^k(1-\pi)^{n-k}\,dm$$

$$= \frac{2}{n}\sum_{k=0}^{n}\binom{n}{k}\pi^k(1-\pi)^{n-k}\cdot\frac{1}{t}e^{tm}\Big|_{0.5k}^{0.5(n+k)}$$

$$= \frac{2}{n}\cdot\frac{1}{t}\sum_{k=0}^{n}\binom{n}{k}\pi^k(1-\pi)^{n-k}[e^{\frac{t}{2}(n+k)} - e^{\frac{t}{2}k}]$$

$$= \frac{2}{n}\cdot\frac{1}{t}(e^{\frac{n}{2}t}-1)\sum_{k=0}^{n}\binom{n}{k}(e^{\frac{t}{2}}\pi)^k(1-\pi)^{n-k}$$

$$= \frac{2}{n}\cdot\frac{e^{\frac{n}{2}t}-1}{t}[\pi e^{\frac{t}{2}}+(1-\pi)]^n \quad \text{for } t\neq 0.$$

As $t=0$, the same proof as theorem 3.3 b.

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

And $\psi(0) = \log M(0) = \log 1 = 0$.

$$\psi'(t) = \frac{t}{e^{\frac{n}{2}t}-1}\cdot\frac{e^{\frac{n}{2}t}\frac{n}{2}\cdot t-(e^{\frac{n}{2}t}-1)\cdot 1}{t^2} + n\cdot\frac{\pi e^{\frac{t}{2}}\frac{1}{2}}{\pi e^{\frac{t}{2}}+1-\pi}.$$

So that $\mu = \psi'(0) = \lim_{t\to 0}\frac{\psi(t)-\psi(0)}{t-0} = \lim_{t\to 0}\frac{\psi(t)}{t} = \lim_{t\to 0}\psi'(t)$ (by L'Hospital's Rule)

$$= \lim_{t\to 0}[\frac{t}{e^{\frac{n}{2}t}-1}\cdot\frac{e^{\frac{n}{2}t}\frac{n}{2}\cdot t-(e^{\frac{n}{2}t}-1)}{t^2} + n\cdot\frac{\pi e^{\frac{t}{2}}\frac{1}{2}}{\pi e^{\frac{t}{2}}+1-\pi}] = n\cdot\frac{1+2\pi}{4}.$$

Moreover,

$$\psi''(t) = \frac{1\cdot(e^{\frac{n}{2}t}-1)-t\cdot e^{\frac{n}{2}t}\frac{n}{2}}{(e^{\frac{n}{2}t}-1)^2}\cdot\frac{e^{\frac{n}{2}t}\frac{n}{2}\cdot t-(e^{\frac{n}{2}t}-1)}{t^2}$$

$$+\frac{t}{e^{\frac{n}{2}t}-1}\cdot\frac{[e^{\frac{n}{2}t}(\frac{n}{2})^2\cdot t+e^{\frac{n}{2}t}\frac{n}{2}\cdot 1-e^{\frac{n}{2}t}\frac{n}{2}]\cdot t^2-[e^{\frac{n}{2}t}\frac{n}{2}\cdot t-(e^{\frac{n}{2}t}-1)]\cdot 2t}{t^4}$$

$$+ n \cdot \frac{\pi e^{\frac{t}{2}}(\frac{1}{2})^2(\pi e^{\frac{t}{2}}+1-\pi)-(\pi e^{\frac{t}{2}}\frac{1}{2})^2}{(\pi e^{\frac{t}{2}}+1-\pi)^2} .$$

Hence $\sigma^2 = \psi''(0) = \lim_{t\to 0}\frac{\psi'(t)-\psi'(0)}{t-0} = \lim_{t\to 0}\frac{\psi'(t)-\mu}{t} = \lim_{t\to 0}\psi''(t)$

(by L'Hospital's Rule)

$$= \frac{n^2}{48}+\frac{n}{4}\pi(1-\pi) .$$

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

3.3 Fuzzy Multinomial Distribution

First, we want to introduce Fuzzy trinomial distribution, and then extension it to the multinomial distribution.

Fuzzy trinomial distribution

The Fuzzy binomial distribution counts the fuzzy number of "successes" in n independent replications of an experiment with two possible outcomes.

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

$$f(m_1, m_2) = 2\begin{cases} (\frac{2}{n})^2\sum_{k_1}\sum_{k_2} \frac{n!}{k_1!k_2!(n-k_1-k_2)!}\pi_1^{k_1}\pi_2^{k_2}(1-\pi_1-\pi_2)^{n-k_1-k_2} & if \;\; (k_1, k_2) \in K_n \\ (\frac{2}{n})^2\sum_{k_1}\sum_{k_2} \frac{n!}{k_1!k_2!(n-k_1-k_2)!}\pi_1^{k_1}\pi_2^{k_2}(1-\pi_1-\pi_2)^{n-k_1-k_2} & if \;\; (k_1, k_2) \in K - K_n \end{cases}$$

where $K = \{(k_1, k_2) : k_1 \geq 0, k_2 \geq 0, \; and \; k_1 + k_2 \leq n\}$ & $(m_1, m_2) \in S_n$. (3.3)

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

In order 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

$$\tilde{S}_n = \{(m_1, m_2) : 0.5k_i \le m_i < 0.5(n + k_i), k_i \in N \cup \{0\} \; for\, i = 1,2 \; and\; k_1 + k_2 \le n\} \quad \&$$

$$A = S - \tilde{S}_n$$

with A is measure zero. Note that, S is the set denoted by

$$S = \{(m_1, m_2) : 0 \le m_1 \le n \,\&\, 0 \le m_2 \le n\}.$$

Theorem 3.5

a. *The Fuzzy trinomial density function given in (3.3) is a density function.*

b. *If* $(M_1, M_2) \sim FT(n, (\pi_1, \pi_2))$, *then* $EM_i = n \cdot \frac{1 + 2\pi_i}{4}$,

$Var(M_i) = \frac{n^2}{48} + \frac{n}{4}\pi_i(1 - \pi_i)$, $\operatorname{cov}(M_1, M_2) = -\frac{n}{4}\pi_1\pi_2$, *and the joint moment-generation function is*

$$M(t_1, t_2) = \begin{cases} (\frac{2}{n})^2 \cdot (\frac{e^{\frac{n}{2}t_1} - 1}{t_1})(\frac{e^{\frac{n}{2}t_2} - 1}{t_2})[\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + (1 - \pi_1 - \pi_2)]^n & if\ (t_1, t_2) = (0,0) \\ 1 & if\ (t_1, t_2) \ne (0,0) \end{cases}$$

Proof:

a. Clearly $f(m_1, m_2) \ge 0$.

By the trinomial theorem and integral operation,

$$\iint_{S_n} f(m_1, m_2) dm_1 dm_2 = \int_0^n \int_0^{n - m_2} f(m_1, m_2) dm_1 dm_2$$

$$= \int_0^n \int_0^{n - m_2} 2(\frac{2}{n})^2 \sum\sum_{(k_1, k_2) \in K} \frac{n!}{k_1! k_2! (n - k_1 - k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1 - \pi_1 - \pi_2)^{n - k_1 - k_2} dm_1\, dm_2$$

$$= 2(\frac{2}{n})^2 \sum\sum_{(k_1, k_2) \in K} \frac{1}{2} \iint_{\tilde{S}_n \cup A} \frac{n!}{k_1! k_2! (n - k_1 - k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1 - \pi_1 - \pi_2)^{n - k_1 - k_2} dm_1 dm_2$$

$$= (\frac{2}{n})^2 \sum\sum_{(k_1, k_2) \in K} [\iint_{\tilde{S}_n} \frac{n!}{k_1! k_2! (n - k_1 - k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1 - \pi_1 - \pi_2)^{n - k_1 - k_2} dm_1 dm_2 +$$

$$\iint_A 0\, dm_1 dm_2]$$

$$= (\frac{2}{n})^2 \sum\sum_{(k_1, k_2) \in K} \int_{0.5k_2}^{0.5(n + k_2)} \int_{0.5k_1}^{0.5(n + k_1)} \frac{n!}{k_1! k_2! (n - k_1 - k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1 - \pi_1 - \pi_2)^{n - k_1 - k_2} dm_1\, dm_2$$

$$= (\frac{2}{n})^2 \sum\sum_{(k_1,k_2)\in K} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1-\pi_2)^{n-k_1-k_2} \cdot [m_1 \Big|_{0.5k_1}^{0.5(n+k_1)}] \cdot [m_2 \Big|_{0.5k_2}^{0.5(n+k_2)}]$$

$$= (\frac{2}{n})^2 \sum\sum_{(k_1,k_2)\in K} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1-\pi_2)^{n-k_1-k_2} \cdot (\frac{n}{2})(\frac{n}{2})$$

$$= \sum\sum_{(k_1,k_2)\in K} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1-\pi_2)^{n-k_1-k_2}$$

$$= [\pi_1 + \pi_2 + (1-\pi_1-\pi_2)]^n = 1 .$$

So that f is a density function.

b. By the trinomial theorem and integral operation again, we have, for the joint moment-generation function,

$$M(t_1,t_2) = E(e^{t_1 m_1 + t_2 m_2})$$

$$= (\frac{2}{n})^2 \sum\sum_{(k_1,k_2)\in K} \int_{0.5k_2}^{0.5(n+k_2)} \int_{0.5k_1}^{0.5(n+k_1)} e^{t_1 m_1} \cdot e^{t_2 m_2} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1-\pi_2)^{n-k_1-k_2} dm_1\, dm_2$$

$$= (\frac{2}{n})^2 \sum\sum_{(k_1,k_2)\in K} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1-\pi_2)^{n-k_1-k_2} \cdot [\frac{e^{t_1 m_1}}{t_1} \Big|_{0.5k_1}^{0.5(n+k_1)}] \cdot [\frac{e^{t_2 m_2}}{t_2} \Big|_{0.5k_2}^{0.5(n+k_2)}]$$

$$= (\frac{2}{n})^2 (\frac{e^{\frac{n}{2}t_1} - 1}{t_1})(\frac{e^{\frac{n}{2}t_2} - 1}{t_2}) \sum\sum_{(k_1,k_2)\in K} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} \pi_1^{k_1} \pi_2^{k_2} (1-\pi_1-\pi_2)^{n-k_1-k_2}$$

$$= (\frac{2}{n})^2 (\frac{e^{\frac{n}{2}t_1} - 1}{t_1})(\frac{e^{\frac{n}{2}t_2} - 1}{t_2}) [\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + (1-\pi_1-\pi_2)]^n \quad \text{for} \quad (t_1,t_2) \neq (0,0) .$$

The moment-generating function is not differentiable at $(t_1,t_2) = (0,0)$, but the moments can be calculated by differentiating and then taking $\lim_{(t_1,t_2)\to(0,0)}$.

Then $M(0) = M(0,0) = \lim_{(t_1,t_2)\to(0,0)} M(t_1,t_2)$

$$= \lim_{(t_1,t_2)\to(0,0)} (\frac{2}{n})^2 (\frac{e^{\frac{n}{2}t_1} - 1}{t_1})(\frac{e^{\frac{n}{2}t_2} - 1}{t_2}) [\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + (1-\pi_1-\pi_2)]^n$$

(Let $t_1 = r\cos\theta$, $t_2 = r\sin\theta$. If $(t_1,t_2) \to (0,0)$, then $r \to 0^+$.)

$$= \lim_{r\to 0^+} (\frac{2}{n})^2 (\frac{e^{\frac{n}{2}r\cos\theta} - 1}{r\cos\theta})(\frac{e^{\frac{n}{2}r\sin\theta} - 1}{r\sin\theta}) [\pi_1 e^{\frac{r\cos\theta}{2}} + \pi_2 e^{\frac{r\sin\theta}{2}} + (1-\pi_1-\pi_2)]^n$$

$$= 1 .$$

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

Therefore,

$$\psi(t) = 2\cdot\log(\frac{2}{n}) + \log(\frac{e^{\frac{n}{2}t_1}-1}{t_1}) + \log(\frac{e^{\frac{n}{2}t_2}-1}{t_2}) + n\cdot\log(\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2),$$

and $\psi(0) = \log M(0) = \log 1 = 0$.

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

$$\mu = EM_i = \psi_i(0) = \lim_{t\to 0}\psi_i(t) = \lim_{t\to 0}[\frac{t_i}{e^{\frac{n}{2}t_i}-1}\cdot\frac{e^{\frac{n}{2}t_i}\frac{n}{2}\cdot t_i - (e^{\frac{n}{2}t_i}-1)}{t_i^2} + n\cdot\frac{\pi_i e^{\frac{t_i}{2}}\frac{1}{2}}{\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2}]$$

$$= n\cdot\frac{1+2\pi_i}{4},$$

$$\psi_{ii}(t) = \frac{\partial^2\psi(t)}{\partial t_i^2} = \frac{1\cdot(e^{\frac{n}{2}t_i}-1) - t_i\cdot e^{\frac{n}{2}t_i}\frac{n}{2}}{(e^{\frac{n}{2}t_i}-1)^2}\cdot\frac{e^{\frac{n}{2}t_i}\frac{n}{2}\cdot t_i - (e^{\frac{n}{2}t_i}-1)}{t_i^2}$$

$$+\frac{t_i}{e^{\frac{n}{2}t_i}-1}\cdot\frac{[e^{\frac{n}{2}t_i}(\frac{n}{2})^2\cdot t_i + e^{\frac{n}{2}t_i}\frac{n}{2}\cdot 1 - e^{\frac{n}{2}t_i}\frac{n}{2}]\cdot t_i^2 - [e^{\frac{n}{2}t_i}\frac{n}{2}\cdot t_i - (e^{\frac{n}{2}t_i}-1)]\cdot 2t_i}{t_i^4}$$

$$+n\cdot\frac{\pi_i e^{\frac{t_i}{2}}(\frac{1}{2})^2(\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2) - (\pi_i e^{\frac{t_i}{2}}\frac{1}{2})^2}{(\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2)^2}.$$

$$\sigma^2 = \text{var}(M_i) = \psi_{ii}(0) = \lim_{t\to 0}\psi_{ii}(t) = \frac{n^2}{48} + \frac{n}{4}\pi_i(1-\pi_i),$$

$$\psi_{ij}(t) = \frac{\partial^2\psi(t)}{\partial t_j\partial t_i} = \frac{\partial}{\partial t_j}(\psi_i(t))$$

$$= \frac{\partial}{\partial t_j}[\frac{t_i}{e^{\frac{n}{2}t_i}-1}\cdot\frac{e^{\frac{n}{2}t_i}\frac{n}{2}\cdot t_i - (e^{\frac{n}{2}t_i}-1)\cdot 1}{t_i^2} + n\cdot\frac{\pi_i e^{\frac{t_i}{2}}\frac{1}{2}}{\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2}]$$

$$= n\cdot[\frac{-\pi_i e^{\frac{t_i}{2}}\frac{1}{2}\cdot\pi_j e^{\frac{t_j}{2}}\frac{1}{2}}{(\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + 1 - \pi_1 - \pi_2)^2}].$$

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

Theorem 3.6

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* $M_1 \sim FB(n, \pi_1)$.

Proof: The marginal moment-generating function of M_1 is

$$M_1(t) = M(t, 0)$$

$$= \lim_{(t_1, t_2) \to (t, 0)} (\frac{2}{n})^2 (\frac{e^{\frac{n}{2}t_1} - 1}{t_1})(\frac{e^{\frac{n}{2}t_2} - 1}{t_2})[\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + (1 - \pi_1 - \pi_2)]^n$$

$$= (\frac{2}{n})^2 (\frac{e^{\frac{n}{2}t} - 1}{t})[\pi_1 e^{\frac{t}{2}} + \pi_2 e^0 + (1 - \pi_1 - \pi_2)]^n \cdot \lim_{t_2 \to 0}(\frac{e^{\frac{n}{2}t_2} - 1}{t_2})$$

$$= (\frac{2}{n})(\frac{e^{\frac{n}{2}t} - 1}{t})[\pi_1 e^{\frac{t}{2}} + (1 - \pi_1)]^n$$

which is the moment-generating function for $FB(n, \pi_1)$, so that $M_1 \sim FB(n, \pi_1)$. The proof for M_2 is similar.

We now consider a notation for the fuzzy trinomial distribution which will lead to the notation we shall use in the next text for 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 $\pi_3 = 1 - \pi_1 - \pi_2$. Then $M = (M_1, M_2, M_3)$ has joint density function

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

where $K = \{(k_1, k_2, k_3) : k_1 \geq 0, k_2 \geq 0, k_3 \geq 0 and\ k_1 + k_2 + k_3 = n\}$ &

$$(m_1, m_2, m_3) \in S_n^*,$$

S_n^* is denoted by $S_n^* = \{(m_1, m_2, m_3) : m_1 \geq 0, m_2 \geq 0, m_3 \geq 0\ \&\ m_1 + m_2 + m_3 = n\}$.

Also, we let

$$K_n^* = \{(k_1, k_2, k_3) : 2m_i - n < k_i \leq 2m_i, k_i \in N \cup \{0\}\ for i = 1,2,3, \& k_1 + k_2 + k_3 = n\}$$

under the condition $(m_1, m_2, m_3) \in S_n^*$, then we have a relation between K_n^* and S_n^*. When m_1, m_2, m_3 decided, k_1, k_2, k_3 are decided. Under the trinomial theorem, it is straightforward to show that (M_1, M_2, M_3) has joint moment-generating function

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

We note that the joint density function and joint moment-generating function of (M_1, M_2, M_3) are somewhat nicer than they are for (M_1, M_2). Notice also that the density functions of (M_1, M_2) and (M_1, M_2, M_3) are ways of representing 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. In this case, we often say that $M = (M_1, M_2, M_3)$ has a three-dimensional fuzzy multinomial distribution and write $(M_1, M_2, M_3) \sim FM_3(n,\pi_1,\pi_2,\pi_3)$, where $\pi_3 = 1 - \pi_1 - \pi_2$.

Fuzzy multinomial distribution

We have already considered situations that involve two and three random variables. Now, we want to extend it to k random variables.

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

$$f(m_1,\ldots,m_k)=\zeta\begin{cases}(\frac{2}{n})^k\sum\cdots\sum\frac{n!}{k_1!\cdots k_k!}\pi_1^{k_1}\pi_2^{k_2}\cdots\pi_k^{k_k} & if\ (k_1,k_2,\ldots,k_k)\in K_n\\ (\frac{2}{n})^k\sum\cdots\sum\frac{n!}{k_1!\cdots k_k!}\pi_1^{k_1}\pi_2^{k_2}\cdots\pi_k^{k_k} & if\ (k_1,k_2,\ldots,k_k)\in K-K_n\end{cases}\tag{3.4}$$

where $K = \left\{(k_1,k_2,\ldots,k_k) : k_i \geq 0\ for\ i = 1,2,\ldots,k\ and\ \sum_{i=1}^{k} k_i = n\right\}$, $M \in S_n$, and $\zeta = \dim(k-1)$.

On the above, n is a positive integer and the π_i are constants such that

$$\pi_1+\pi_2+\cdots+\pi_k = 1.$$

Moreover,

$$K_n = \left\{ (k_1, k_2, \ldots, k_k) : 2m_i - n < k_i \le 2m_i, k_i \in N \cup \{0\}\, for\, i = 1,2,\ldots,k \ \&\ \sum_{i=1}^{k} k_i = n \right\}.$$

Note that $M_1 + M_2 + \cdots + M_k = n$, and hence, $M_k = n - M_1 - M_2 - \cdots - M_{k-1}$ and $\pi_k = 1 - \pi_1 - \pi_2 - \cdots - \pi_{k-1}$. Note also that

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

and $(M_1, M_2, M_3) \sim FM_3(n, \pi_1, \pi_2, \pi_3) \iff (M_1, M_2) \sim FT(n, (\pi_1, \pi_2))$,

$$M_3 = n - M_1 - M_2.$$

The following theorem summarizes some important facts about the fuzzy multinomial distribution.

Theorem 3.7

a. *The fuzzy multinomial density function in (3.4) is a joint density for all positive integer* n *and* $\pi_1, \cdots, \pi_k$ *such that* $\pi_i \ge 0$ *and* $\pi_1 + \pi_2 + \cdots + \pi_k = 1$.

b. *Let* $M \sim FM_k(n, \pi)$, *where* $M = (M_1, M_2, ..., M_k)$, $\pi = (\pi_1, \pi_2, ..., \pi_k)$,

   $\sum_{i=1}^{k} M_i = n$ *and* $\sum_{i=1}^{k} \pi_i = 1$.

   Then $EM_i = n \cdot \dfrac{1 + 2\pi_i}{4}$, $Var(M_i) = \dfrac{n^2}{48} + \dfrac{n}{4}\pi_i(1 - \pi_i)$, $\mathrm{cov}(M_i, M_j) = -\dfrac{n}{4}\pi_i \pi_j$,

   and the joint moment-generation function is

$$M(t) = \begin{cases} \left(\frac{2}{n}\right)^k \left(\frac{e^{\frac{n}{2}t_1} - 1}{t_1}\right) \cdots \left(\frac{e^{\frac{n}{2}t_k} - 1}{t_k}\right) [\pi_1 e^{\frac{t_1}{2}} + \pi_2 e^{\frac{t_2}{2}} + \cdots + \pi_k e^{\frac{t_k}{2}}]^n & if\ t = 0 \\ 1 & if\ t \ne 0 \end{cases},$$

   where $t = (t_1, \ldots, t_k)$.

c. *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 same proof as theorem 3.5 & 3.6.

The next theorem gives a normal approximation which is often useful.

Theorem 3.8

Let $X_i \sim FB(1, \pi)$ *and* $\overline{X}_n = \sum_{i=1}^{n} \dfrac{x_i}{n} = \dfrac{M}{n}$, *where* $M \sim FB(n, \pi)$ *and* $M = \sum_{i=1}^{n} x_i$.

Suppose that $\mu = E(X_i)$ *is finite and* $\sigma^2 = Var(X_i) < \infty$.

Then

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

Proof: Since $X_i \sim FB(1,\pi)$, we have that

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

Moreover, $\overline{X}_n = \sum_{i=1}^{n}\frac{x_i}{n} = \frac{M}{n}$, where $M \sim FB(n,\pi)$ and

$$\mu = E(M) = n\cdot\frac{1+2\pi}{4}, \quad \sigma^2 = Var(M) = \frac{n^2}{48}+\frac{n}{4}\pi(1-\pi)$$

Hence $\mu = E(\overline{X}_n) = E(\frac{M}{n}) = \frac{1}{n}E(M) = \frac{1}{n}\cdot(n\cdot\frac{1+2\pi}{4}) = \frac{1+2\pi}{4}$ and

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

By the central limit theory, we get that

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

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

We have introduced some new distributions in used of fuzzy theorem. Now, we can use these distributions to derive a very useful test statistic, called fuzzy chi-square test statistic for goodness-of-fit.

3.4 Fuzzy Chi-square Test Statistic for Goodness-of-Fit

In this section, we consider applications of very important chi-square statistic. We begin our study with the same way by considering the basic chi-square statistic, which has only an approximate chi-square distribution. There are many ways to show the χ^2 test for goodness-of-fit, and we will get the same result in any ways. So that, we just only show that in one way.

The l-sample fuzzy multinomial model

Let M_i be the independent l-dimensional random vectors, $M_i \sim FM_k(n_i,\pi_i)$, where the n_i are known integers and the π_i are unknown parameter vectors. We call this model the l-sample fuzzy multinomial model. Our main goal for this model is to test the equality of the π_i. We let M_{ij} be the jth component of M_i, π_{ij} be the jth component of π_i and L_j denote the language variable for $j = 1, 2, \ldots, k$. We can see easily in Table 3.1.

Table 3.1. *The table of membership* M_{ij} *in* L_{ij}

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

Theorem 3.9

A_{ij} is an unbiased estimation of π_{ij} for this model, where $A_{ij} = \dfrac{2M_{ij}}{n_i} - \dfrac{1}{2}$.

Proof: Since $M_i \sim FM_k(n_i, \pi_i)$, which is implied that $M_{ij} \sim FB(n_i, \pi_{ij})$.

So that 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 get that $E[\dfrac{1}{2}(\dfrac{4M_{ij}}{n_i} - 1)] = E(\dfrac{2M_{ij}}{n_i} - \dfrac{1}{2}) = E(A_{ij}) = \pi_{ij}$.

Hence A_{ij} is an unbiased estimation of π_{ij}.

Now, we want to test $H_0: \ \pi_1 = \pi_2 = \ldots = \pi_l$ against $H_1: \ H_0$ is not true.

Under the null hypothesis H_0 that the π_i are all equal, and let $\pi_1 = \pi_2 = \ldots = \pi_l = \pi_0$ 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

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

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

Let

$$\hat{U}_k = \sum_{i=1}^{l}\{\sum_{j=1}^{k-1}\frac{(M_{ij}-\hat{E}_{ij})^2}{\hat{B}_{ij}} + \frac{4[\sum_{j=1}^{k-1}(M_{ij}-\hat{E}_{ij})(1-\frac{n_i^2}{48\hat{B}_{ij}})]^2}{n_i - 4\sum_{j=1}^{k-1}\hat{B}_{ij}(1-\frac{n_i^2}{48\hat{B}_{ij}})^2}\}$$, where $\hat{B}_{ij} = \frac{1}{2}\hat{E}_{ij} - \frac{n_i}{8} + \frac{n_i^2}{48}$.

We call that $\hat{U}_k$ is a fuzzy χ^2 and has $(l-1)(k-1)$ degrees of freedom.

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

In order to prove that $\hat{U}_k \xrightarrow{d} \chi^2(l-1)(k-1)$, we must have a theorem and a lemma.

Theorem 3.10 (Arnold, 1990 [1])

Let $X = (X_1, X_2, \ldots, X_k)$ *have a multivariate normal distribution,* $X \sim N_k(\mu, \Sigma)$, *and* $\Sigma > 0$ *is the variance of* X *, then* $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2(k)$.

Proof: See Arnold, 1990 [1], p.211-212.

Lemma3.11 (Arnold, 1990 [1])

Let A *be a* $q \times q$ *invertible symmetric matrix, let* b *and* c *be* q*-dimensional vectors, and let* $d \neq 0$ *be a number. Then*

$$c'(A-d^{-1}bb')^{-1}c = c'A^{-1}c + \frac{(c'A^{-1}b)^2}{d-b'A^{-1}b}.$$

Proof: Claim that $(A-d^{-1}bb')^{-1} = A^{-1} + (d-b'A^{-1}b)^{-1}A^{-1}bb'A^{-1}$

Since $[A^{-1} + (d-b'A^{-1}b)^{-1}A^{-1}bb'A^{-1}](A-d^{-1}bb')$

$= A^{-1}A + (d-b'A^{-1}b)^{-1}A^{-1}bb'A^{-1}A - A^{-1}d^{-1}bb' - (d-b'A^{-1}b)^{-1}A^{-1}bb'A^{-1}d^{-1}bb'$

$= I + [(d-b'A^{-1}b)^{-1} - d^{-1} - d^{-1}(d-b'A^{-1}b)^{-1}]A^{-1}bb'$

$= I$

We get that $(A-d^{-1}bb')^{-1} = A^{-1} + (d-b'A^{-1}b)^{-1}A^{-1}bb'A^{-1}$

Hence $c'(A-d^{-1}bb')^{-1}c = c'A^{-1}c + c'(d-b'A^{-1}b)^{-1}A^{-1}bb'A^{-1}c$

$$= c'A^{-1}c + \frac{c'A^{-1}bb'A^{-1}c}{d-b'A^{-1}b}$$

$$= c'A^{-1}c + \frac{(c'A^{-1}b)^2}{d-b'A^{-1}b}$$

Theorem3.12 $\hat{U}_k \xrightarrow{d} \chi^2(l-1)(k-1)$.

Proof: Want to show $\hat{U}_k \xrightarrow{d} \chi^2(l-1)(k-1)$, we just only to show that

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

Now, let $M_n = (M_{n1}, M_{n2}, \ldots, M_{nk})'$ and $\pi = (\pi_1, \pi_2, \ldots, \pi_k)'$.
And let $E_n = (E_{n1}, E_{n2}, \ldots, E_{nk})'$ and V be the $K \times K$ matrix whose i th diagonal element is $V_{ii} = \frac{1}{4}\pi_i(1-\pi_i)$ and whose (i, j) th off-diagonal element is $V_{ij} = -\frac{1}{4}\pi_i\pi_j$.

We have know that $Var(X_i) = \frac{1}{48} + \frac{1}{4}\pi_i(1-\pi_i)$, and

$\mathrm{cov}(X_i, X_j) = -\frac{1}{4}\pi_i\pi_j$, for $i \neq j$.

First, we show that $n^{-1}(M_n - E_n)$ is approximately $N_k(0, \frac{V}{n} + \frac{1}{48}I_k)$.

Since $M_n \sim FM_k(n, \pi)$, where $\pi = (\pi_1, \pi_2, \ldots, \pi_k)'$ and $M_n = n\overline{X_n}$.

We have known that $E(\overline{X}_n) = \frac{1+2\pi}{4}$ and $Var(\overline{X}_n) = \frac{V}{n} + \frac{1}{48}I_k$.

Therefore $E_n = EM_n = n \cdot \frac{1+2\pi}{4}$ and $Var(M_n) = n \cdot V + \frac{n^2}{48}I_k$.

By the multinomial central limit theorem, we get that

$$\overline{X}_n - \frac{1+2\pi}{4} \text{ is approximately } N_k(0, \frac{V}{n} + \frac{1}{48}I_k).$$

So that $\dfrac{n\overline{X}_n - n\frac{1+2\pi}{4}}{n}$ is approximately $N_k(0, \frac{V}{n} + \frac{1}{48}I_k)$.

Hence $n^{-1}(M_n - E_n)$ is approximately $N_k(0, \frac{V}{n} + \frac{1}{48}I_k)$.

Since 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, we have that $n^{-1}(M_n^* - E_n^*)$ is approximately $N_{k-1}(0, \frac{V^*}{n} + \frac{1}{48}I_{k-1})$.

By theorem 3.10, we have that

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

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

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

Let $D_{k-1} = \begin{bmatrix} n\pi_1 & 0 & \cdots & 0 \\ 0 & n\pi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n\pi_{k-1} \end{bmatrix}$ and $F_n = (n\pi_1, n\pi_2, \cdots, n\pi_k)'$.

Let $C_n = M_n^* - E_n^*$, $A = \frac{1}{4} D_{k-1} + \frac{n^2}{48} I_{k-1}$, $b = F_n^*$ & $d = 4n$.

By lemma 3.11,

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

$$= (M_n^* - E_n^*)'(\frac{1}{4} D_{k-1} + \frac{n^2}{48} I_{k-1})^{-1}(M_n^* - E_n^*) + \frac{[(M_n^* - E_n^*)'(\frac{1}{4} D_{k-1} + \frac{n^2}{48} I_{k-1})^{-1} F_n^*]^2}{4n - F_n^{*\prime}(\frac{1}{4} D_{k-1} + \frac{n^2}{48} I_{k-1})^{-1} F_n^*}$$

$$= \sum_{i=1}^{k-1} \frac{(M_{ni} - E_{ni})^2}{B_{ni}} + \frac{[4\sum_{i=1}^{k-1}(M_{ni} - E_{ni})(1 - \frac{n^2}{48 B_{ni}})]^2}{4n - 16\sum_{i=1}^{k-1} B_{ni}(1 - \frac{n^2}{48 B_{ni}})^2}$$

$$= \sum_{i=1}^{k-1} \frac{(M_{ni} - E_{ni})^2}{B_{ni}} + \frac{4[\sum_{i=1}^{k-1}(M_{ni} - E_{ni})(1 - \frac{n^2}{48 B_{ni}})]^2}{n - 4\sum_{i=1}^{k-1} B_{ni}(1 - \frac{n^2}{48 B_{ni}})^2}, \text{ where } B_{ni} = \frac{1}{2} E_{ni} - \frac{n}{8} + \frac{n^2}{48}.$$

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

To compute the degrees of freedom in theorem, note that there are $l-1$ degrees of freedom for each of the k populations, so that there are $k(l-1)$ degrees of freedom for the whole model. Under the null hypothesis, we are estimating $l-1$ 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 $k(l-1) - (l-1) = (k-1)(l-1)$.