THE PROOF OF INDEPENDENCE OF MARGINALS FOR $k \times l$ CONTINGENCY TABLE

Hisao NAGAO

Received June 16, 2009

Abstract. The purpose of this note is to extend Fisher's exact test in 2×2 table to $k \times l$ table. Also a simple numerical example is given in 2×3 table.

1. Preparation At first, we consider the following 2×2 table.

$\mathcal{A}$ \ $\mathcal{B}$	B	$\bar{B}$	total
A	N_{11}	N_{12}	$N_{1\cdot}$
$\bar{A}$	N_{21}	N_{22}	$N_{2\cdot}$
total	$N_{\cdot 1}$	$N_{\cdot 2}$	n

Here we denote A and B are two events and $\bar{A}$ and $\bar{B}$ denote the complement of A and B, respectively. We repeat n times experiment in which $A \cap B, A \cap \bar{B}, \bar{A} \cap B$ and $\bar{A} \cap \bar{B}$ occur. The N_{11} denotes the number which $A \cap B$ occurs. The other N_{12}, N_{21} and N_{22} are defined similarly. On the other hand, $N_{1\cdot}$ is sum of N_{11} and N_{12}. $N_{2\cdot}$ is sum of N_{21} and N_{22}. That is, $N_{1\cdot}$ denotes the number the event A occurs and $N_{2\cdot}$ is the number for $\bar{A}$. Similarly $N_{\cdot 1}$ and $N_{\cdot 2}$ are the number for B and $\bar{B}$, respectively.
We consider the testing hypothesis problem whether the events A and B are independent. When n is small, there is Fisher's exact test in 2×2 table. To caluculate it, we must show that the random numbers $N_{1\cdot}$ and $N_{\cdot 1}$ are independent. As far as the author knows, only one book giving its proof is Kitagawa and Inaba's 「統計学通論」written in Japanese and published by Kyoritsu-Pub. Co. (1970). Their proof is not so readable, but unclear, rather seems to be not collect. We give the proof for a more general case in the next section.

2. $k \times l$ table We consider the generality of 2×2 table. Let Ω be a whole event. Let $A_1, \cdots, A_k$ be a partition $\mathcal{A}$ of Ω and $B_1, \cdots, B_l$ be another partition $\mathcal{B}$ of Ω. We repeat n times experiment. N_{ij} denotes the number which $A_i \cap B_j$ appears in this experiment. Then we have the following $k \times l$ contigency table.

2000 *Mathematics Subject Classification.* 62-07.
Key words and phrases. Independence, contingency table, Fisher's exact test, probability generating function .

$\mathcal{B}$ / $\mathcal{A}$	B_1	$\cdots$	B_j	$\cdots$	B_l	total
A_1	N_{11}	$\cdots$	$\cdots$	$\cdots$	N_{1l}	$N_{1\cdot}$
$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$	$\vdots$
A_i	$\vdots$		N_{ij}		$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$	$\vdots$
A_k	N_{k1}	$\cdots$	$\cdots$	$\cdots$	N_{kl}	$N_{k\cdot}$
total	$N_{\cdot 1}$	$\cdots$	$\cdots$	$\cdots$	$N_{\cdot l}$	n

$N_{i\cdot} = \sum_{j=1}^{l} N_{ij}$ and $N_{\cdot j} = \sum_{i=1}^{k} N_{ij}$. Here $N_{i\cdot}$ represents the number of A_i's occuring and $N_{\cdot j}$ represents the number of B_j. We assume that $\Pr(A_i \cap B_j) = \Pr(A_i)\Pr(B_j)$ ($i = 1, \cdots, k; j = 1, \cdots, l$). Then we shall show that $(N_{1\cdot}, \cdots, N_{k\cdot})$ and $(N_{\cdot 1}, \cdots, N_{\cdot l})$ are independent. Before giving the proof, we put $p_{i\cdot} = \Pr(A_i), p_{\cdot j}$ $= \Pr(B_j)$ and $p_{ij} = \Pr(A_i \cap B_j)$. Let $a_1, \cdots, a_k$ and $b_1, \cdots, b_l$ be positive numbers. We consider the probability generating function of $(N_{1\cdot}, \cdots, N_{k\cdot})$ and $(N_{\cdot 1}, \cdots, N_{\cdot l})$.

$$\mathrm{E}(a_1^{N_{1\cdot}} \cdots a_k^{N_{k\cdot}} b_1^{N_{\cdot 1}} \cdots b_l^{N_{\cdot l}}) = \mathrm{E}((a_1 b_1)^{N_{11}} (a_1 b_2)^{N_{12}} \cdots (a_k b_l)^{N_{kl}})$$

$$= (a_1 b_1 p_{11} + a_1 b_2 p_{12} + \cdots + a_k b_l p_{kl})^n$$

$$= (a_1 b_1 p_{1\cdot} p_{\cdot 1} + a_1 b_2 p_{1\cdot} p_{\cdot 2} + \cdots + a_k b_l p_{k\cdot} p_{\cdot l})^n$$

$$= (a_1 p_{1\cdot} + a_2 p_{2\cdot} + \cdots + a_k p_{k\cdot})^n (b_1 p_{\cdot 1} + b_2 p_{\cdot 2} + \cdots + b_l p_{\cdot l})^n$$

$$= \sum \frac{n!}{n_{1\cdot}! \cdots n_{k\cdot}!} (p_{1\cdot} a_1)^{n_{1\cdot}} \cdots (p_{k\cdot} a_k)^{n_{k\cdot}} \times \sum \frac{n!}{n_{\cdot 1}! \cdots n_{\cdot l}!} (p_{\cdot 1} b_1)^{n_{\cdot 1}} \cdots (p_{\cdot l} b_l)^{n_{\cdot l}}$$

Therefore we have

$$\Pr(N_{1\cdot} = n_{1\cdot}, \cdots, N_{k\cdot} = n_{k\cdot}, N_{\cdot 1} = n_{\cdot 1}, \cdots, N_{\cdot l} = n_{\cdot l})$$

$$= \Pr(N_{1\cdot} = n_{1\cdot}, \cdots, N_{k\cdot} = n_{k\cdot}) \Pr(N_{\cdot 1} = n_{\cdot 1}, \cdots, N_{\cdot l} = n_{\cdot l})$$

According to the above formula, we have

$$\Pr(N_{1\cdot} = n_{1\cdot}, \cdots, N_{k\cdot} = n_{k\cdot}) = \frac{n!}{n_{1\cdot}! \cdots n_{k\cdot}!} (p_{1\cdot})^{n_{1\cdot}} \cdots (p_{k\cdot})^{n_{k\cdot}}$$

and

$$\Pr(N_{\cdot 1} = n_{\cdot 1}, \cdots, N_{\cdot l} = n_{\cdot l}) = \frac{n!}{n_{\cdot 1}! \cdots n_{\cdot l}!} (p_{\cdot 1})^{n_{\cdot 1}} \cdots (p_{\cdot l})^{n_{\cdot l}}$$

Thus we have the following theorem.

Theorem. If $\Pr(A_i \cap B_j) = \Pr(A_i)\Pr(B_j)$ ($i = 1, \cdots, k; j = 1, \cdots, l$), we have $(N_{1\cdot}, \cdots, N_{k\cdot})$ and $(N_{\cdot 1}, \cdots, N_{\cdot l})$ are independent and the former is multinomial distribution with $(p_{1\cdot}, \cdots, p_{k\cdot})$ and the latter is the same distribution with $(p_{\cdot 1}, \cdots, p_{\cdot l})$.

To get the genelalized result for $k \times l$ table of exact test in 2×2 table, using this theorem, we have

$$\Pr(N_{ij} = n_{ij}, i = 1, \cdots, k; j = 1, \cdots, l | N_{i\cdot} = n_{i\cdot}, i = 1, \cdots, k, N_{\cdot j} = n_{\cdot j}, j = 1, \cdots, l)$$

$$= \frac{(\prod_{i=1}^{k} n_{i\cdot}!)(\prod_{j=1}^{l} n_{\cdot j}!)}{n!(\prod_{i=1}^{k} \prod_{j=1}^{l} n_{ij}!)}$$

3. Numerical example Here we suppose that the following data are given.

$\mathcal{A}$ \ $\mathcal{B}$	B_1	B_2	B_3	total
A_1	6	3	1	10
A_2	1	1	3	5
total	7	4	4	15

These data are artificial numbers. We consider whether A's and B's are independent. If these are independent, each value estimated of each cell is as follows;

$\mathcal{A}$ \ $\mathcal{B}$	B_1	B_2	B_3	total
A_1	14/3	8/3	8/3	10
A_2	7/3	4/3	4/3	5
total	7	4	4	15

According to Fisher's consideration, we have the following form.

$\mathcal{A}$ \ $\mathcal{B}$	B_1	B_2	B_3
A_1	+	+	−
A_2	−	−	+

For example, we denote + at $A_1 \cap B_1$, because of $6 - 14/3 > 0$. Similarly we denote − at $A_2 \cap B_1$, because of $1 - 7/3 < 0$. Thus we get the above table.
By generalizing the consideration of Fisher's 2×2 table, we obtain the following three types as rejection region $\mathcal{R}$.

$$\begin{pmatrix} 6 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix} \quad \begin{pmatrix} 6 & 4 & 0 \\ 1 & 0 & 4 \end{pmatrix} \quad \begin{pmatrix} 7 & 3 & 0 \\ 0 & 1 & 4 \end{pmatrix}$$

From the result of the above theorem,

$$\Pr(\mathcal{R}) = \frac{10!5!7!4!4!}{15!}\left(\frac{1}{6!3!3!} + \frac{1}{6!4!4!} + \frac{1}{7!4!3!}\right)$$
$$\approx 3.83\%.$$

When the significance level is 5%, the hypothethis is rejected.

Hisao Nagao
5-1-25-506, Senrioka
Settsu,Osaka 566-0001
Japan
e-mail address: nagao@sb.dcns.ne.jp