NEW APPROACH TO DISTRIBUTION FREE TESTS IN CONTINGENCY TABLES

T. T. M. NGUYEN

ABSTRACT. For study of two-way contingency tables, in this paper we derive the inverse Fisher information matrix Γ^{-1} under the assumptions that the sample schemes follow multinomial distribution. From the explicit form of Γ^{-1} we clarify mathematical properties of vector $\widehat{Y}$ of components of chi-squared statistic, then propose an approach for testing the hypothesis of independence between two random response variables. The main idea is that based on a unitary transformation, we turn vector $\widehat{Y}$ into vector $\widehat{Z}$ which possesses the same "statistical information" as $\widehat{Y}$ but any statistics based on $\widehat{Z}$, not only chi-squared tests, has distribution free.

1. INTRODUCTION

Let X and Y denote two categorical response variables with $(k+1)$ and $(l+1)$ categories corresponding. So the classifications of subjects on both variables have $M = (k+1)(l+1)$ possible combinations (i, j). As we know, to describe contingency tables, there are several measures of associations between two random variables. For example, (i) assume unrestricted sampling models, the simple model is that the count in the (i, j) cell ν_{ij} is a realization of a Poisson variable with expectation μ_{ij}; (ii) assume the total number of observations is fixed by n, the counts on cells are multinomial distributed; (iii) assume the row totals are fixed, we have independent multinomial distributions for each row.

Our study is considering case (ii) in which the parameters are maximum likelihood estimators. Recall that the distribution of the frequencies ν_{ij} is

$$\mathbb{P}\left\{\nu_{ij} = x_{ij}, \sum_{i=1}^{n} x_{ij} = n\right\} = \frac{n!}{\prod_{i,j} x_{ij}!} \prod_{i,j} p_{ij}^{x_{ij}}.$$

where p_{ij} are the joint distributions of X and Y in cells (i, j). Let indexes i and j be running from 0. Two response random variables X and Y are defined to be independent if

$$H_0 : p_{ij} = a_i b_j \quad \text{for all} \quad i, j$$

in which a_i and b_j are marginal distributions:

$$a_i = \frac{\sum_{j=0}^{l} \nu_{ij}}{n}, \quad b_j = \frac{\sum_{i=0}^{k} \nu_{ij}}{n}.$$

It is obvious that $\sum_{i=0}^{k} a_i = \sum_{j=0}^{l} b_j = 1$. Since it can be said that a_0, b_0 is corresponding linear dependent on $a = (a_1, \cdots, a_k)^T$, $b = (b_1, \cdots, b_l)^T$ the parameter is $\boldsymbol{\theta} = (a^T, b^T)^T \in \mathbb{R}^{\kappa}$ where $\kappa = k + l$ and $a \in \mathbb{R}^k, b \in \mathbb{R}^l$. The log-likelihood function under the null hypothesis is:

$$\ell(\boldsymbol{\theta}) = \log(n!) - \sum_{i,j} \log(\nu_{ij}!) + \sum_{i,j} \nu_{ij} \log a_i + \sum_{i,j} \nu_{ij} \log b_j.$$

Then the maximum likelihood estimators $\hat{\boldsymbol{\theta}} = (\hat{a}^T, \hat{b}^T)^T$ is

$$\hat{a}_i = \frac{\nu_{i+}}{n}, \qquad \hat{b}_j = \frac{\nu_{+j}}{n}, \quad i = 1, \cdots, k,\ j = 1, \cdots, l. \tag{1.1}$$

So the joint distribution is estimated by $\hat{p}_{ij} = \hat{a}_i\hat{b}_j$ and estimated frequencies are $\hat{\nu}_{ij} = n\hat{p}_{ij}$. The most common way to test the statistical independence of X and Y under the null hypothesis H_0 is using chi-squared statistics χ^2 or Likelihood-Ratio Chi-squared G^2 where

$$\chi^2 = \sum_{i,j} \frac{(\nu_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$$

and

$$G^2 = 2\sum_{i,j} \nu_{ij} \log\left(\frac{\nu_{ij}}{\hat{\nu}_{ij}}\right)$$

Here and below, notation $\sum_{i,j}$ means $\sum_{i=0}^{k}\sum_{j=0}^{l}$. It is well known long time ago that for testing statistical independence between X and Y, χ^2 and G^2 has chi-squared distribution with kl degree of freedom (Fisher, 1922).

As in 2013, Khmaladze([3]) introduced a new distribution free goodness of fit tests for discrete distributions, and interestingly, we could apply the same idea for this problem. One would start from considering vector $\widehat{Y}$ of components of chi-squared statistics calculated from estimated parameter then analyze its orthogonal property in order to apply a unitary transformation onto it. The achieved result is that we obtain a transformed vector $\widehat{Z}$ of $\widehat{Y}$ such that, not only chi-squared tests but any statistics based on it has distribution free.

2. Methodology

Let denote the true parameter by $\boldsymbol{\theta}_0$. As normal, the Fisher information matrix is defined by

$$\Gamma = \sum_{i,j} \frac{\dot{p}_{ij}(\boldsymbol{\theta}_0)\dot{p}_{ij}(\boldsymbol{\theta}_0)^T}{p_{ij}(\boldsymbol{\theta}_0)} = \sum_{i,j} \frac{\dot{p}_{ij}(\boldsymbol{\theta}_0)}{\sqrt{p_{ij}(\boldsymbol{\theta}_0)}}\left(\frac{\dot{p}_{ij}(\boldsymbol{\theta}_0)}{\sqrt{p_{ij}(\boldsymbol{\theta}_0)}}\right)^T,$$

where $\dot{p}_{ij}(\boldsymbol{\theta}_0)$ is the vector of partial derivatives of p_{ij} in $\boldsymbol{\theta}_0$.

We will see then the matrix Γ of dimension $\kappa \times \kappa$ is of the form

$$\Gamma = \left(\begin{array}{c|c} \Gamma_a & \mathbf{0} \\ \hline \mathbf{0} & \Gamma_b \end{array}\right). \tag{2.1}$$

In fact, under the null hypothesis H_0, we have

$$\frac{\partial p_{z_1z_2}/\partial a_i}{\sqrt{p_{z_1z_2}}} = \sqrt{\frac{b_{z_2}}{a_{z_1}}}\left[1_{\{z_1=i\}} - 1_{\{z_1=0\}}\right], \quad i = 1, \cdots, k \tag{2.2}$$

and

$$\frac{\partial p_{z_1z_2}/\partial b_j}{\sqrt{p_{z_1z_2}}} = \sqrt{\frac{a_{z_1}}{b_{z_2}}}\left[1_{\{z_2=j\}} - 1_{\{z_2=0\}}\right], \quad j = 1, \cdots, l. \tag{2.3}$$

If $i \neq i'$, then we have

$$\sum_{z_1,z_2} \frac{\partial p_{z_1z_2}/\partial a_i}{\sqrt{p_{z_1z_2}}}\frac{\partial p_{z_1z_2}/\partial a_{i'}}{\sqrt{p_{z_1z_2}}} = \sum_{z_1,z_2} \frac{b_{z_2}}{a_{z_1}} 1_{\{z_1=0\}} = \frac{1}{a_0}.$$

If $i = i'$ then

$$\sum_{z_1,z_2} \frac{\partial p_{z_1z_2}/\partial a_i}{\sqrt{p_{z_1z_2}}}\frac{\partial p_{z_1z_2}/\partial a_{i'}}{\sqrt{p_{z_1z_2}}} = \sum_{z_1,z_2} \frac{b_{z_2}}{a_{z_1}}[1_{\{z_1=0\}} + 1_{\{z_1=i\}}] = \frac{1}{a_i} + \frac{1}{a_0}.$$

For all i, j, it is trivial to see that

$$\sum_{z_1,z_2} \frac{\partial p_{z_1z_2}/\partial a_i}{\sqrt{p_{z_1z_2}}}\frac{\partial p_{z_1z_2}/\partial b_j}{\sqrt{p_{z_1z_2}}} = 0.$$

Hence, if we denote $\mathbf{1}_k = (1, \cdots, 1)^T \in \mathbb{R}^k$ and diagonal matrix

$$D\left(\frac{1}{a}\right) = \begin{pmatrix} \frac{1}{a_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{1}{a_k} \end{pmatrix}$$

and those with similar notation $D(a), D(\sqrt{a}), \ldots$, then we have

$$\Gamma_a = D\left(\frac{1}{a}\right) + \frac{1}{a_0}\mathbf{1}_k\mathbf{1}_k^T.$$

Similarly,

$$\Gamma_b = D\left(\frac{1}{b}\right) + \frac{1}{b_0}\mathbf{1}_l\mathbf{1}_l^T.$$

It is known that a matrix of the above form will have inverse matrix or square root of inverse matrix, etc. of the same form.

Lemma 2.1. *The inverse matrix of the Fisher information matrix in form of blocks is*

$$\Gamma^{-1} = \left(\begin{array}{c|c} \Gamma_a^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \Gamma_b^{-1} \end{array}\right)$$

where

$$\Gamma_a^{-1} = D(a) - aa^T, \qquad \Gamma_b^{-1} = D(b) - bb^T. \tag{2.4}$$

Proof. We only need to show that (2.4) is actually the inverse matrix of Γ_a. In fact,

$$\begin{aligned} \Gamma_a\Gamma_a^{-1} &= \left[D\left(\frac{1}{a}\right) + \frac{1}{a_0}\mathbf{1}_k\mathbf{1}_k^T\right]\left[D(a) - aa^T\right] \\ &= D\left(\frac{1}{a}\right)D(a) + \frac{1}{a_0}\mathbf{1}_k\mathbf{1}_k^T D(a) - D\left(\frac{1}{a}\right)aa^T - \frac{1}{a_0}\mathbf{1}_k\mathbf{1}_k^T aa^T \\ &= \mathbb{I}_k + \frac{1}{a_0}\mathbf{1}_k a^T - \mathbf{1}_k a^T - \frac{1}{a_0}(1-a_0)\mathbf{1}_k a^T \\ &= \mathbb{I}_k, \end{aligned} \tag{2.5}$$

where $\mathbb{I}_k$ is notation for identity matrix of size $k \times k$. At the same time,

$$\begin{aligned} \Gamma_a^{-1}\Gamma_a &= \left[D(a) - aa^T\right]\left[D\left(\frac{1}{a}\right) + \frac{1}{a_0}\mathbf{1}_k\mathbf{1}_k^T\right] \\ &= D(a)D\left(\frac{1}{a}\right) + \frac{1}{a_0}D(a)\mathbf{1}_k\mathbf{1}_k^T - aa^T D\left(\frac{1}{a}\right) - \frac{1}{a_0}aa^T\mathbf{1}_k\mathbf{1}_k^T \\ &= \mathbb{I}_k + \frac{1}{a_0}a\mathbf{1}_k^T - a\mathbf{1}_k^T - \frac{1}{a_0}(1-a_0)a\mathbf{1}_k^T \\ &= \mathbb{I}_k. \end{aligned} \tag{2.6}$$

$\square$

Lemma 2.2. *We have*

$$\Gamma_a^{-1/2} = [\mathbb{I}_k - c_a\sqrt{a}\sqrt{a}^T]D(\sqrt{a}) \tag{2.7}$$

where $c = \frac{1}{1\pm\sqrt{a_0}}$ *in the sense of* $\left(\Gamma_a^{-1/2}\right)^T\Gamma_a^{-1/2} = \Gamma_a^{-1}$.

Proof. We have

$$\left(\Gamma_a^{-1/2}\right)^T = D(\sqrt{a})[\mathbb{I}_k - c_a\sqrt{a}\sqrt{a}^T],$$

so

$$\begin{aligned}\left(\Gamma_a^{-1/2}\right)^T\Gamma_a^{-1/2} &= D(\sqrt{a})[\mathbb{I}_k - c_a\sqrt{a}\sqrt{a}^T][\mathbb{I}_k - c_a\sqrt{a}\sqrt{a}^T]D(\sqrt{a}) \\ &= D(\sqrt{a})\left[\mathbb{I}_k - 2c_a\sqrt{a}\sqrt{a}^T + c_a^2(a-a_0)\sqrt{a}\sqrt{a}^T\right]D(\sqrt{a}) \\ &= D(a) - \left[2c_a - c_a^2(1-a_0)\right]D(\sqrt{a})\sqrt{a}\sqrt{a}^T D(\sqrt{a}) \\ &= D(a) - \left[2c_a - c_a^2(1-a_0)\right]D(\sqrt{a})aa^T. \qquad (2.8)\end{aligned}$$

Hence $\left(\Gamma_a^{-1/2}\right)^T\Gamma_a^{-1/2} = \Gamma_a^{-1} = D(a) - aa^T$ if and only if

$$2c_a - c_a^2(1-a_0) = 1,$$

which implies $c_a = \frac{1}{1\pm\sqrt{a_0}}$. □

It is well known long time ago that under mild assumptions, the maximum likelihood estimator is asymptotically linear

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \Gamma^{-1}\sum_{i,j} Y_{ij}\frac{\dot{p}_{ij}(\boldsymbol{\theta}_0)}{\sqrt{p_{ij}(\boldsymbol{\theta}_0)}} + o_P(1), \qquad (2.9)$$

where

$$Y_{ij} = \frac{\nu_{ij} - np_{ij}(\boldsymbol{\theta}_0)}{\sqrt{np_{ij}(\boldsymbol{\theta}_0)}}.$$

According to Khmaladze([3]), let denote

$$\widehat{Y}_{ij} = \frac{\nu_{ij} - np_{ij}(\hat{\boldsymbol{\theta}})}{\sqrt{np_{ij}(\hat{\boldsymbol{\theta}})}} \qquad (2.10)$$

then we have

$$\widehat{Y}_{ij} = Y_{ij} - \frac{\dot{p}_{ij}(\boldsymbol{\theta}_0)^T}{\sqrt{p_{ij}(\boldsymbol{\theta}_0)}}\Gamma^{-1}\sum_{i,j} Y_{ij}\frac{\dot{p}_{ij}(\boldsymbol{\theta}_0)}{\sqrt{p_{ij}(\boldsymbol{\theta}_0)}} + o_P(1). \qquad (2.11)$$

And define

$$q_{ij} = \Gamma^{-1/2}\frac{\dot{p}_{ij}(\boldsymbol{\theta}_0)}{\sqrt{p_{ij}(\boldsymbol{\theta}_0)}}, \qquad i = 1,\cdots,(k+1),\ j = 1,\cdots,(l+1), \qquad (2.12)$$

where $\Gamma^{-1/2}$ is defined in the sense of $(\Gamma^{-1/2})^T\Gamma^{-1/2} = \Gamma^{-1}$. Then we have convergence in distribution of vector $\widehat{Y} = (\widehat{Y}_{ij}) \in \mathbb{R}^M$:

$$\widehat{Y} = X - \langle X, \sqrt{p}\rangle\sqrt{p} - \langle X, q\rangle q, \qquad (2.13)$$

where $\sqrt{p} = (\sqrt{p_{ij}})$ and $X = (X_1,\cdots,X_M)$ denote a vector of M independent standard normal random variables. It is already claimed in Khmaladze that $\widehat{Y}$ is the projection of vector X orthogonal to subspace generated by $(\kappa+1)$ vectors $\sqrt{p}$ and q. Since the expression of matrix Γ^{-1} is clear, we also can get the explicit form of vectors $q = (q^{(1)},\cdots,q^{(\kappa)})$.

Therefore, from the definition of q in (2.12), replacing what we achieved so far for Γ^{-1} and $\frac{\dot{p}_{ij}(\boldsymbol{\theta}_0)}{\sqrt{p_{ij}(\boldsymbol{\theta}_0)}}$ as in (2.2) and (2.3), each element of vector $q^{(m)}$ with $m \le k$ will be

$$q^{(m)}_{z_1z_2} = \sqrt{\frac{b_{z_2}}{a_{z_1}}}\left[\sqrt{a_m}1_{\{z_1=m\}} - c_a\sqrt{a_m}a_{z_1}\bar{1}_{\{z_1=k+1\}} \pm 1_{\{z_1=k+1\}}\sqrt{a_0}\sqrt{a_m}\right] \qquad (2.14)$$

and for $m \geq k+1$

(2.15)

$$q^{(m)}_{z_1 z_2} = \sqrt{\frac{a_{z_1}}{b_{z_2}}}\left[\sqrt{b_{m-k}}1_{\{z_2=m-k\}} - c_b\sqrt{b_{m-k}}b_{z_2}\bar{1}_{\{z_2=l+1\}} \pm 1_{\{z_2=l+1\}}\sqrt{b_0}\sqrt{b_{m-k}}\right]$$

here by notation $\bar{1}$ define a function $\bar{1}_{\{z=k\}} = 1$ if $z \neq k$ and 0 otherwise.

By the idea suggested in Khmaladze([3]), if we apply a product of a series of unitary operator on $\widehat{Y}$, we will finally get a transformed vector $\widehat{Z}$ which satisfies

$$\widehat{Z} = X - \sum_{k=1}^{\kappa+1}\langle X, r^{(k)}\rangle r^{(k)}, \tag{2.16}$$

that means $\widehat{Z}$ is the projection of X orthogonal to the subspace generated by specified vectors $r^{(k)}$. More specifically, the transformed vector $\widehat{Z}$ is obtained by

$$\widehat{Z} = \left(\prod_{\tau=1}^{\kappa+1} U_{\widetilde{q}_\tau, r_\tau}\right)\widehat{Y}, \tag{2.17}$$

where $\widetilde{q}_\tau$ are retrieved recursively

$$\widetilde{q}_\tau = \left(\prod_{\rho\leq\tau} U_{\widetilde{q}_\rho, r_\rho}\right) q_\tau$$

and the unitary operator is defined as

$$U_{q,r} = I - \frac{1}{1-\langle q, r\rangle}(r-q)(r-q)^T.$$

3. Numerical illustrations

The main aim of this section is to illustrate that the demonstrated methodology works properly and how user can apply this method in practice. Let consider the Kolmogorov-Smirnov statistics

$$KS = \sup_x |F^*(x) - F_0(x)|$$

where $F^*(x)$ is the empirical distribution function based on the sampled data and $F_0(x)$ is the true cumulative distribution function. In the sense of our problem, we can clarifies the KS statistics by

$$KS_1 = \max_{(0,0)\leq(z_1,z_2)\leq(k,l)}\left|\sum_{(i,j)\leq(z_1,z_2)}\widehat{Z}_{ij}\right|$$

or

$$KS_2 = \max_{1\leq k\leq M}\left|\sum_{x=1}^{k}\widehat{Z}_x\right|,$$

where $\widehat{Z}_x$ is written in form of a vector of length M.

Figure 1 shows the cumulative distribution of statistics KS_1 for contingency table of size 7×7. This cumulative distribution function is generated by simulation of 5000 iterations. Choose random parameter $\boldsymbol{\theta}_0$ be the true one. Generate realizations of cell's counts based on this true distribution then estimated $\hat{a}$ and $\hat{b}$ as in (1.1). Calculating vectors $q = (q^{(1)}, \cdots, q^{(\kappa)})$ as in (2.14) and (2.15) but using estimated $\hat{a}$ and $\hat{b}$ instead of the true a and b. From that, we know the limit distribution of vector $\widehat{Y}$ and then apply unitary transformation U recursively as shown in (2.17).

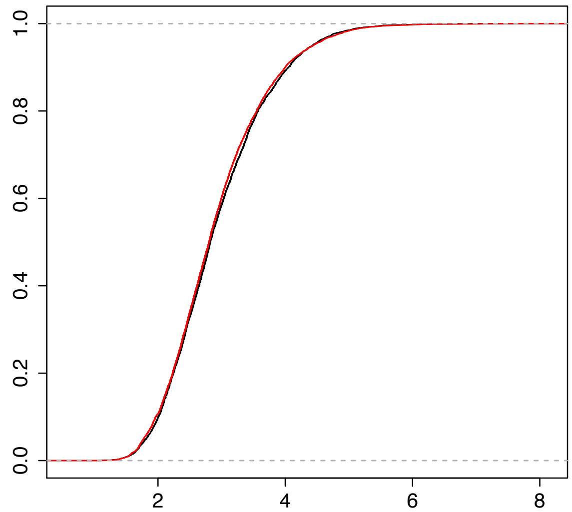

FIGURE 1. *Distribution functions of KS_1 with two different parameters $\boldsymbol{\theta}_0$*

The specific set of vector r here is chosen to be $\sqrt{p}$ and q but in case when $a_i = \frac{1}{k+1}$, $b_j = \frac{1}{l+1}$. For user's convenience, let see the explicit form of r:

$$r^{(0)} = \left(\frac{1}{\sqrt{(k+1)(l+1)}} \right) \tag{3.1}$$

and for $m \leq k$,

$$r^{(m)}_{z_1 z_2} = \frac{\sqrt{l+1}}{k+1} \left[1_{\{z_1 = m\}} - \frac{c_{a0}}{k+1} \bar{1}_{\{z_1=0\}} + \frac{1}{\sqrt{k+1}} 1_{\{z_1=0\}} \right]$$

for $m \geq (k+1)$

$$r^{(m)}_{z_1 z_2} = \frac{\sqrt{k+1}}{l+1} \left[1_{\{z_2 = m-k\}} - \frac{c_{b0}}{l+1} \bar{1}_{\{z_2=0\}} + \frac{1}{\sqrt{l+1}} 1_{\{z_2=0\}} \right],$$

where $c_{a0} = \frac{\sqrt{k+1}}{1 \pm \sqrt{k+1}}$ and $c_{b0} = \frac{\sqrt{l+1}}{1 \pm \sqrt{l+1}}$.

Actually, the simulation result is obtained by using constants c_a, c_b, c_{a0}, c_{b0} with $+$ sign.

Figure 2 shows the graph of cumulative distribution functions of statistics KS_2 for contingency table of size 5×6.

For using this approach in order to test the null hypothesis about statistical independence of X and Y, note that we can always use the estimated parameters in every calculation.

4. DISCUSSION

The attempt to derive Fisher information matrix is in order to make the new distribution free method more transparent. One may think that the transformation is too complicated to apply but in fact, all of the calculation is linear, and can be calculated computationally very easy. The benefit of this approach is that it gives us

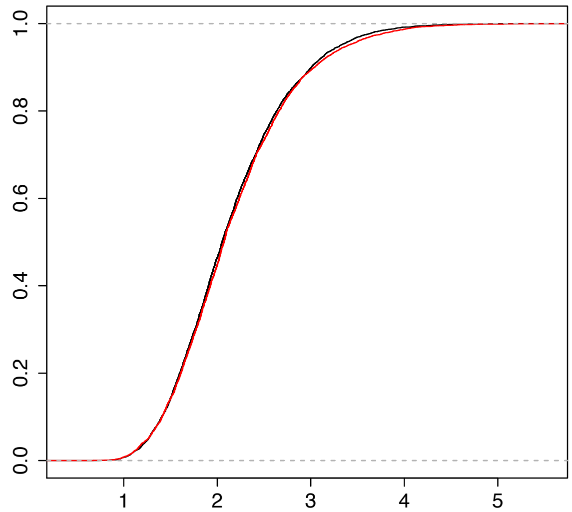

FIGURE 2. *Distribution functions of KS_1 with two different parameters $\boldsymbol{\theta}_0$*

REFERENCES

[1] Fisher, R. A. (1922). On the interpretation of χ^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society,* Vol. 85, No. 1, pp. 87-94.

[2] Palmgren, J. (1981). The Fisher information matrix for log linear models arguing conditionally on observed explanatory variables. *Biometrika*, Vol. 68, No. 2, pp. 563-566.

[3] Khmaladze, E. (2013). Note on distribution free testing for discrete distribution. *Annals of Statistics*. (*Accepted*)

[4] Holt, D., Scott, A. J., Ewings, P. D., (1980). Chi-squared tests with survey data. *J. R. Statist. Soc.* A**143**, 302-320.

SCHOOL OF MATHEMATICS, STATISTICS, AND OPERATIONS RESEARCH, VICTORIA UNIVERSITY OF WELLINGTON, WELLINGTON 6140, NEW ZEALAND

E-mail address: thuong.nguyen@vuw.ac.nz