Journal of Statistics: Advances in Theory and Applications
Volume 17, Number 1, 2017, Pages 33-42
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/jsata_7100121776

ON DECOMPOSITION OF POINT-SYMMETRY FOR SQUARE CONTINGENCY TABLES WITH ORDERED CATEGORIES

HIROYUKI KURAKAMI, KEIGO NEGISHI and SADAO TOMIZAWA

Department of Information Sciences
Faculty of Science and Technology
Tokyo University of Science
Noda City, Chiba, 278-8510
Japan
e-mail: kurakami@rs.tus.ac.jp
tomizawa@is.noda.tus.ac.jp

Abstract

For square contingency tables with ordered categories, this paper proposes the model that has the structure of point-symmetry, and show the decomposition of the proposed model. Also the orthogonal decomposition of test statistic for proposed model is shown.

1. Introduction

Consider an $r \times r$ square contingency table with the same ordinal row and column classifications. Let p_{ij} denote the probability that an observation will fall in the i-th row and j-th column of the table

2010 Mathematics Subject Classification: 62H17.
Keywords and phrases: contingency table, decomposition, orthogonal decomposition, point-symmetry.
Received February 16, 2017

$(i = 1, \ldots, r;\ j = 1, \ldots, r)$. The symmetry (S) model (Bowker [2]) is defined by

$$p_{ij} = \phi_{ij} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $\phi_{ij} = \phi_{ji}$ (also see Bishop et al. [1], p. 282; Caussinus [3]). The conditional symmetry (CS) model (McCullagh [5]) is defined by

$$p_{ij} = \begin{cases} \theta\phi_{ij} & (i < j), \\ \phi_{ij} & (i \geq j), \end{cases}$$

where $\phi_{ij} = \phi_{ji}$ (also see Read [7]). Let X and Y denote the row and column variables, respectively. The CS model indicates that

$$P(X = i, Y = j \mid X < Y) = P(X = j, Y = i \mid X > Y),$$

for $i < j$. The global symmetry (GS) model (Read [7]) is defined by

$$\sum_{i<j} p_{ij} = \sum_{i>j} p_{ij} \quad (\text{i.e., } P(X < Y) = P(X > Y)).$$

Read [7] gave the following theorem:

Theorem 1. *The S model holds if and only if the CS and GS models hold.*

Consider the marginal mean equality (ME) model, i.e., $E(X) = E(Y)$. Noting that the ME model may be expressed as

$$\sum_{i<j} (j - i) p_{ij} = \sum_{i>j} (i - j) p_{ij}.$$

In the similar way to Theorem 1, the following theorem can be obtained:

Theorem 2. *The S model holds if and only if the CS and ME models hold.*

Wall and Lienert [11] considered the point-symmetry (PS) model defined by

$$p_{ij} = \phi_{ij} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $\phi_{ij} = \phi_{i^* j^*}$ with $i^* = r + 1 - i$ and $j^* = r + 1 - j$ (also see Tomizawa [9]; Tahata and Tomizawa [8]). Tomizawa [10] considered the conditional point-symmetry (CPS) model defined by

$$p_{ij} = \begin{cases} \beta\phi_{ij} & (i + j < r + 1), \\ \phi_{i^* j^*} & (i + j > r + 1), \end{cases}$$

where $\phi_{ij} = \phi_{i^* j^*}$. The CPS model indicates that

$$P(X = i, Y = j \mid X + Y < r + 1) = P(X = i^*, Y = j^* \mid X + Y > r + 1).$$

Tomizawa and Tahata [8] proposed the quasi point-symmetry model and the marginal point-symmetry model, and considered the decomposition of the PS model using these models. We are now interested in considering the decompositions of the PS model in the similar manner to Theorems 1 and 2.

The present paper proposes another PS model, gives the decomposition of proposed model and shows that the test statistic for the proposed model is equal to the sum of those for decomposed model.

2. Decomposition of Point-Symmetry Model

First, we consider another point symmetry model (APS) as follows:

$$p_{ij} = \phi_{ij} \quad (i = 1, \dots, r;\ j = 1, \dots, r;\ i + j \neq r + 1),$$

where $\phi_{ij} = \phi_{i^* j^*}$ with $i^* = r + 1 - i$ and $j^* = r + 1 - j$. We note that a special case of the CPS model obtained by putting $\beta = 1$ is the APS model. Also, we note that the APS model is less restrictive than the PS model with respect to excluding the restriction of reverse diagonal elements from the PS model.

Next consider the reverse global symmetry (RGS) model as follows:

$$\Delta_U = \Delta_L,$$

where

$$\Delta_U = \sum_{i+j<r+1}\sum p_{ij} \quad [= P(X+Y<r+1)],$$

$$\Delta_L = \sum_{i+j>r+1}\sum p_{ij} \quad [= P(X+Y>r+1)].$$

We obtain the following theorem:

Theorem 3. *The APS model holds if and only if the CPS and RGS models hold.*

Proof. If the APS model holds, the CPS and RGS models hold. Assuming that the CPS and RGS models hold, we shall show that the APS model holds. Since the CPS model holds, we see

$$\Delta_U = \sum_{i+j<r+1}\sum p_{ij}$$

$$= \beta \sum_{i+j<r+1}\sum p_{i^* j^*}.$$

By putting $i^* = s$ and $j^* = t$, we see

$$\Delta_U = \beta \sum_{s+t>r+1}\sum p_{st}$$

$$= \beta \Delta_L.$$

Since the RGS model holds, we obtain $\beta = 1$. Thus the APS model holds. The proof is completed.

Consider the model, which indicates the point-symmetry of mean of sum of X and Y, as follows;

$$E(X+Y) = E(X^* + Y^*).$$

We shall denote this model by SMPS. The model may be expressed as

$$E(X+Y) = r+1,$$

or

$$E(X) = E(Y^*).$$

This model also indicates that $E(X)$ is point symmetry to $E(Y)$ since $E(X) + E(Y) = r + 1$. We also obtain the following theorem:

Theorem 4. *The APS model holds if and only if the CPS and SMPS models hold.*

Proof. Assume that the APS model holds. Then the CPS model holds. Also we see

$$E(X - Y^*) = \sum_{i+j<r+1}\sum (i - j^*)p_{ij} + \sum_{i+j=r+1}\sum (i - j^*)p_{ij} + \sum_{i+j>r+1}\sum (i - j^*)p_{ij}.$$

The second term of right side equals zero since $j^* = r + 1 - j$. In the first term of right side, by putting $i^* = s$ and $j^* = t$, we obtain

$$E(X - Y^*) = -\sum_{s+t>r+1}\sum (s - t^*)p_{s^*t^*} + \sum_{i+j>r+1}\sum (i - j^*)p_{ij}.$$

Since $p_{s^*t^*} = p_{st}$, we obtain $E(X) = E(Y^*)$. Thus the SMPS model holds.

Conversely, assuming that both the CPS and SMPS models hold, then we shall show that the APS model holds. Since the CPS model holds, we obtain

$$E(X - Y^*) = \beta \sum_{i+j<r+1}\sum (i - j^*)p_{i^*j^*} + \sum_{i+j>r+1}\sum (i - j^*)p_{ij}.$$

By putting $i^* = s$ and $j^* = t$, we obtain

$$E(X - Y^*) = -\beta \sum_{s+t>r+1}\sum (s - t^*)p_{st} + \sum_{i+j>r+1}\sum (i - j^*)p_{ij}.$$

Since the SMPS model holds, i.e., $E(X - Y^*) = 0$, we obtain $\beta = 1$. Thus the APS model holds. The proof is completed.

3. Test Statistic and its Orthogonality

Let x_{ij} denote the observed frequency in the (i, j) cell $(i = 1, \ldots, r;$ $j = 1, \ldots, r)$. Assume that a multinomial distribution applies to the $r \times r$ table. Let $G^2(M)$ denote the likelihood ratio chi-squared statistic for testing goodness-of-fit of model M defined by

$$G^2(M) = 2\sum_{i=1}^{r}\sum_{j=1}^{r} x_{ij} \log\left(\frac{x_{ij}}{\hat{m}_{ij}}\right),$$

where $\hat{m}_{ij}$ is the maximum likelihood estimate (MLE) of expected frequency m_{ij} under the model M. The number of degrees of freedom (df) for testing the APS model is $r(r-1)/2$. The number of df for testing the CPS model is one less than that for the APS model. The number of df for testing each of RGS and SMPS models is one.

The MLEs of $\{m_{ij}\}$ under the APS, CPS, and RGS models are expressed as the closed-forms as follows:

(a) The MLE of m_{ij} under the APS model is

$$\hat{m}_{ij} = \begin{cases} \dfrac{x_{ij} + x_{i^* j^*}}{2} & (i + j \neq r + 1), \\ x_{ij} & (i + j = r + 1). \end{cases}$$

(b) The MLE of m_{ij} under the CPS model is

$$\hat{m}_{ij} = \begin{cases} \dfrac{B}{B + C}(x_{ij} + x_{i^* j^*}) & (i + j < r + 1), \\ \dfrac{C}{B + C}(x_{ij} + x_{i^* j^*}) & (i + j > r + 1), \\ x_{ij} & (i + j = r + 1), \end{cases}$$

where

$$B = \sum\sum_{i+j<r+1} x_{ij}, \quad C = \sum\sum_{i+j>r+1} x_{ij}.$$

(c) The MLE of m_{ij} under the RGS model is

$$\hat{m}_{ij} = \begin{cases} \dfrac{B+C}{2B} x_{ij} & (i+j < r+1), \\ \dfrac{B+C}{2C} x_{ij} & (i+j > r+1), \\ x_{ij} & (i+j = r+1). \end{cases}$$

Note that the MLE of m_{ij} under the SMPS model can be obtained using, e.g., the Newton-Raphson method to the log-likelihood equation, although the detail is omitted. From the MLEs of APS, CPS, and RGS, we can obtained the following theorem:

Theorem 5. $G^2(APS) = G^2(CPS) + G^2(RGS)$.

4. Examples

4.1. Example 1

Consider the data in Table 1 that shows cross-classification of mother's and father's education for a sample of eminent black Americans, taken from Mullins and Sites [6]. From Table 3, we see that the APS and CPS models fit the data in Table 1 poorly. However, both the RGS and SMPS models fit these data well. Therefore, it is seen from Theorem 3 that the poor fit of the APS model is caused by the influence of the lack of structure of the CPS model rather than the RGS model. Also, from Theorem 4, we can obtain the similar inference. Under the RGS model, we can estimate that the probability of couple with child with higher education is close to the probability of couple with child with inferior education.

Table 1. The origins of contemporary eminent black Americans (Mullins and Sites [6]). The upper and lower parenthesized values are the MLEs of expected frequencies under the RGS and SMPS models, respectively

	Father's Education				
Mother's Education	(1)	(2)	(3)	(4)	Total
(1)	81	3	9	11	104
	(87.665)	(3.247)	(9.741)	(11.000)	
	(90.120)	(3.217)	(9.314)	(11.000)	
(2)	14	8	9	6	37
	(15.152)	(8.658)	(9.000)	(5.576)	
	(15.013)	(8.279)	(9.000)	(5.804)	
(3)	43	7	43	18	111
	(46.538)	(7.000)	(39.962)	(16.728)	
	(44.501)	(7.000)	(41.597)	(16.863)	
(4)	21	6	24	87	138
	(21.000)	(5.576)	(22.304)	(80.853)	
	(21.000)	(5.804)	(22.483)	(79.005)	
Total	159	24	85	122	390

(1): 8th grade or less; (2) Part high school; (3) High school; (4) College.

4.2. Example 2

Consider the data in Table 2 that shows the Conneticut blood pressure survey, taken from Freeman [4]. From Table 3, we see that the APS, RGS, and SMPS models fit the data in Table 2 poorly. However, the CPS model fits these data well. Therefore, it is seen from Theorem 3 that the poor fit of the APS model is caused by the influence of the lack of structure of the RGS model rather than the CPS model. Also from the Theorem 4, it is seen that the poor fit of the APS model is caused by the influence of lack of SMPS model. Under the CPS model, the MLE of β is 12.398. Thus, under the CPS model, we can estimate that, for example, the probability that a initial survey is borderline and follow-up survey is normal is estimated to be 12.398 times higher than the probability that a initial survey is borderline and follow-up survey is elevated.

Table 2. Distribution initial and follow-up blood pressure according to hypertension status: WHO definitions, taken from Freeman [4]. The parenthesized values are the MLEs of expected frequencies under the CPS model

	Follow-up Survey			
Initial Survey	Normal	Borderline	Elevated	Total
Normal	105	9	3	117
	(103.642)	(10.179)	(3.000)	
Borderline	10	12	1	23
	(10.179)	(12.000)	(0.821)	
Elevated	3	2	7	12
	(3.000)	(0.821)	(8.358)	
Total	118	23	11	152

Table 3. Likelihood ratio statistic G^2 values for models applied to the data in Tables 1 and 2

Applied	For Table 1		For Table 2	
Models	df	G^2	df	G^2
APS	6	77.796*	3	116.261*
CPS	5	75.818*	2	1.637
RGS	1	1.978	1	114.624*
SMPS	1	2.130	1	113.265*

*means significant at the 0.05 level.

5. Concluding Remarks

In Theorems 3 and 4, we have proposed the two kinds of decompositions of the APS model (i) into the CPS and RGS models and (ii) into the CPS and SMPS models. These decompositions may be useful for seeing the reason for the poor fit when the APS model fits the data poorly.

In addition, we point out that the likelihood ratio chi-squared statistic for testing goodness-of-fit of the APS model assuming that the CPS model holds true, is $G^2(APS) - G^2(CPS)$ and this is equal to the likelihood ratio chisquared statistic for testing goodness-of-fit of the RGS model, i.e., $G^2(RGS)$ (from Theorem 5).

References

[1] Y. M. M. Bishop, S. E. Fienberg and P. W. Holland, Discrete Multivariate Analysis: Theory and Practice, The MIT Press, Cambridge, Massachusetts, 1975.

[2] A. H. Bowker, A test for symmetry in contingency tables, Journal of the American Statistical Assosiation 43 (1948), 572-574.

[3] H. Caussinus, Contribution à l'analyse statistique des tableaux de correlation, Annales de la Faculté des Sciences de l'Université de Toulouse, 29 (1965), 77-182.

[4] D. H. Freeman, The analysis of twice classified data, American Statistical Association: Proceedings of the Social Statistics Section (1981), 178-182.

[5] P. McCullagh, A class of parametric models for the analysis of square contingency tables with ordered categories, Biometrika 65 (1978), 413-418.

[6] E. L. Mullins and P. Sites, The origins of contemporary eminent black Americans: A three generations analysis of social origin, American Sociological Review 49 (1984), 672-685.

[7] C. B. Read, Partitioning chi-square in contingency tables: A teaching approach, Communications in Statistics-Theory and Method 6 (1977), 553-562.

[8] K. Tahata and S. Tomizawa, Orthogonal decompositions of point-symmetry for multiway tables, Advances in Statistical Analysis 92 (2008), 255-269.

[9] S. Tomizawa, The decompositions for point-symmetry in two-way contingency tables, Biometrical Journal 27 (1985), 895-905.

[10] S. Tomizawa, Four kinds of symmetry models and their decompositions in a square contingency table with ordered categories, Biometrica Journal 28 (1986), 387-393.

[11] K. D. Wall and G. A. Lienert, A test of point-symmetry in J-dimensional contingency tables, Biometrical Journal 18 (1976), 259-264.

■