EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 6, No. 3, 2013, 299-306
ISSN 1307-5543 – www.ejpam.com

Decomposition of Symmetry Model into Three Models for Cumulative Probabilities in Square Contingency Tables

Kouji Tahata[1,*], Kouji Yamamoto[2], Sadao Tomizawa[1]

[1] *Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Noda City, Chiba, 278-8510, Japan*
[2] *Department of Medical Innovation, Osaka University Hospital, 2-15, Yamadaoka, Suita, Osaka, 565-0871, Japan*

Abstract. For square contingency tables with ordered categories, we decompose the symmetry model into three models for cumulative probabilities. Three models are the cumulative two ratios-parameter symmetry, the global symmetry, and the marginal means equality models. An example is given.

2010 Mathematics Subject Classifications: 62H17

Key Words and Phrases: Decomposition, Global symmetry, Marginal mean, Square contingency table, Symmetry

---

1. Introduction

For an $r \times r$ square contingency table with the same row and column classifications, let p_{ij} denote the probability that an observation will fall in the ith row and jth column of the table ($i = 1, \ldots, r; j = 1, \ldots, r$). Bowker [3] considered the symmetry (S) model defined by

$$p_{ij} = p_{ji} \quad (i \neq j);$$

see [2, p.282]. Caussinus [4] considered the quasi-symmetry (QS) model defined by

$$p_{ij} = \mu \alpha_i \beta_j \psi_{ij} \quad (i = 1, \ldots, r; j = 1, \ldots, r),$$

where $\psi_{ij} = \psi_{ji}$. A special case of QS model obtained by putting $\{\alpha_i = \beta_i\}$ is the S model. The marginal homogeneity (MH) model is defined by

$$p_{i\cdot} = p_{\cdot i} \quad (i = 1, \ldots, r),$$

*Corresponding author.

Email addresses: kouji_tahata@is.noda.tus.ac.jp (K. Tahata), yamamoto-k@hp-crc.med.osaka-u.ac.jp (K. Yamamoto), tomizawa@is.noda.tus.ac.jp (S. Tomizawa)

where

$$p_{i\cdot} = \sum_{t=1}^{r} p_{it}, \quad p_{\cdot i} = \sum_{s=1}^{r} p_{si};$$

see [8]. Caussinus [4] gave the following theorem.

Theorem 1. *The S model holds if and only if both the QS and MH models hold.*

Tomizawa [11] considered the two ratios-parameter symmetry (2RPS) model defined by

$$\frac{p_{ij}}{p_{ji}} = \gamma\phi^{j-i} \quad (i < j).$$

Special cases of 2RPS model obtained by putting $\phi = 1$ and $\gamma = 1$ are McCullagh's [6] conditional symmetry (CS) and Agresti's [1] linear diagonals-parameter symmetry (LDPS) models, respectively.

Define the global symmetry (GS) model by

$$\delta_U = \delta_L,$$

where

$$\delta_U = \sum\sum_{i<j} p_{ij}, \quad \delta_L = \sum\sum_{i>j} p_{ij}.$$

Let X and Y denote the row and column variables, respectively. Define the marginal means equality (ME) model by

$$E(X) = E(Y),$$

where

$$E(X) = \sum_{i=1}^{r} i p_{i\cdot}, \quad E(Y) = \sum_{i=1}^{r} i p_{\cdot i}.$$

Yamamoto, Iwashita and Tomizawa [15], and Tahata, Yamamoto and Tomizawa [10] gave the following theorem.

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

Tahata and Tomizawa [9] gave the theorem as follows.

Theorem 3. *The S model holds if and only if all the 2RPS, GS and ME models hold.*

Let

$$G_{ij} = \sum_{s=1}^{i} \sum_{t=j}^{r} p_{st} \quad (i < j),$$

and

$$G_{ij} = \sum_{s=i}^{r} \sum_{t=1}^{j} p_{st} \quad (i > j).$$

Note that the S model may be expressed as

$$G_{ij} = G_{ji} \quad (i \neq j).$$

The MH model may be expressed as

$$G_{i,i+1} = G_{i+1,i} \quad (i = 1, \dots, r-1).$$

Miyamoto, Ohtsuka and Tomizawa [7] considered the cumulative quasi-symmetry (CQS) model defined by

$$G_{ij} = \mu \xi_i \eta_j \Psi_{ij} \quad (i \neq j),$$

where $\Psi_{ij} = \Psi_{ji}$. Yamamoto, Ando and Tomizawa [16] gave the following theorem.

Theorem 4. *The S model holds if and only if both the CQS and MH models hold.*

Miyamoto et al. [7] also considered the cumulative linear diagonals-parameter symmetry (CLDPS) model defined by

$$\frac{G_{ij}}{G_{ji}} = \Theta^{j-i} \quad (i < j).$$

Yamamoto and Tomizawa [17] gave the theorem as follows.

Theorem 5. *The S model holds if and only if both the CLDPS and ME models hold.*

Tomizawa, Miyamoto, Yamamoto and Sugiyama [13] considered the cumulative two ratios-parameter symmetry (C2RPS) model defined by

$$\frac{G_{ij}}{G_{ji}} = \Gamma \Theta^{j-i} \quad (i < j).$$

We are now interested in whether or not Theorem 3 with the 2RPS model replaced by the C2RPS model holds.

The purpose of this paper is to decompose the S model into three models, i.e., the C2RPS, the GS, and the ME models.

2. New Decomposition of Symmetry

We can obtain a new decomposition of the symmetry model as follows.

Theorem 6. *The S model holds if and only if all the C2RPS, GS, and ME models hold.*

Proof. If the S model holds, then all the C2RPS, GS, and ME models hold. Assume that the C2RPS, GS, and ME models hold, and then we shall show that the S model holds. We see

$$E(X) = \sum_{i=1}^{r} i p_i.$$

$$
\begin{aligned}
&= \sum_{s=1}^{r} \sum_{t=s}^{r} p_{t\cdot} \\
&= \sum_{s=1}^{r} (1 - F_{s-1}^{X}) \\
&= r - \sum_{i=1}^{r-1} F_i^X,
\end{aligned}
$$

where $F_i^X = P(X \le i)$. Similarly we see

$$E(Y) = r - \sum_{i=1}^{r-1} F_i^Y,$$

where $F_i^Y = P(Y \le i)$. Thus we see

$$
\begin{aligned}
E(Y) - E(X) &= \sum_{i=1}^{r-1} F_i^X - \sum_{i=1}^{r-1} F_i^Y \\
&= \sum_{i=1}^{r-1} G_{i,i+1} - \sum_{i=1}^{r-1} G_{i+1,i}.
\end{aligned}
$$

From the ME model, we see

$$\sum_{i=1}^{r-1} G_{i,i+1} = \sum_{i=1}^{r-1} G_{i+1,i}. \tag{1}$$

From the C2RPS model, we obtain

$$\sum_{i=1}^{r-1} G_{i,i+1} = \Gamma\Theta \sum_{i=1}^{r-1} G_{i+1,i}.$$

From (1) we see $\Gamma = \Theta^{-1}$. Thus

$$\frac{G_{ij}}{G_{ji}} = \Theta^{j-i-1} \quad (i < j). \tag{2}$$

We can see that

$$\sum_{i=1}^{r-1} G_{i,i+1} = \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} (j-i) p_{ij},$$

and

$$\sum_{i=1}^{r-1} G_{i+1,i} = \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} (j-i) p_{ji}.$$

Also we can see that

$$\sum_{i=1}^{r-2} G_{i,i+2} = \sum_{i=1}^{r-2} \sum_{j=i+2}^{r} (j-i-1)p_{ij},$$

and

$$\sum_{i=1}^{r-2} G_{i+2,i} = \sum_{i=1}^{r-2} \sum_{j=i+2}^{r} (j-i-1)p_{ji}.$$

Therefore we can see that

$$\delta_U = \sum_{i=1}^{r-1} G_{i,i+1} - \sum_{i=1}^{r-2} G_{i,i+2},$$

and

$$\delta_L = \sum_{i=1}^{r-1} G_{i+1,i} - \sum_{i=1}^{r-2} G_{i+2,i}.$$

From the GS model (i.e., $\delta_U = \delta_L$) and from (1), we can obtain

$$\sum_{i=1}^{r-2} G_{i,i+2} = \sum_{i=1}^{r-2} G_{i+2,i}.$$

From (2) we obtain

$$\sum_{i=1}^{r-2} \Theta G_{i+2,i} = \sum_{i=1}^{r-2} G_{i+2,i}.$$

Thus $\Theta = 1$, i.e., the S model holds. The proof is completed. □

3. Goodness-of-fit Test

Let x_{ij} denote the observed frequency in the ith row and jth column of the $r \times r$ table ($i = 1, \dots, r; j = 1, \dots, r$), with $N = \sum\sum x_{ij}$. Let m_{ij} denote the corresponding expected frequency. Assuming that $\{x_{ij}\}$ have a multinomial distribution, the maximum likelihood estimates of expected frequencies $\{m_{ij}\}$ under each model could be obtained, for example, using the Newton-Raphson method to the log-likelihood equations. The goodness-of-fit of each model can be tested by, e.g., the likelihood ratio chi-squared statistic G^2 with the corresponding degrees of freedom, defined by

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

where $\hat{m}_{ij}$ is the maximum likelihood estimate of m_{ij} under the model. The numbers of degrees of freedom for each model are omitted here; however, when $r = 4$, those are given in Table 2.

4. Example

Consider the vision data in Table 1. The row variable X is the right eye grade and the column variable Y is the left eye grade. The categories are ordered from best (1) to worst (4). These data have been analyzed by many statisticians, including Stuart [8], Bishop et al. [2, p.284], McCullagh [6], Goodman [5], Tomizawa [12], Miyamoto et al. [7], and Tomizawa and Tahata [14].

Table 1: Unaided distance vision of 7,477 women aged 30-39 employed in Royal Ordnance factories in Britain from 1943 to 1946; from [8].

Right eye	Left eye grade				
grade	Best (1)	Second (2)	Third (3)	Worst (4)	Total
Best (1)	1520	266	124	66	1976
Second (2)	234	1512	432	78	2256
Third (3)	117	362	1772	205	2456
Worst (4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

Table 2 gives the values of the likelihood ratio chi-squared statistic G^2 for each model. The S model fits the data in Table 1 poorly. We can see from Theorem 1 that the poor fit of the S model is caused by the influence of the lack of structure of the MH model rather than that of the QS model.

Table 2: Likelihood ratio chi-square values for models applied to the data in Table 1. ($*$ means significant at the 0.05 level.)

Applied models	Degrees of freedom	Likelihood ratio chi-square
S	6	19.25^*
QS	3	7.27
MH	3	11.99^*
CS	5	7.35
LDPS	5	7.28
2RPS	4	6.83
GS	1	11.90^*
ME	1	11.98^*
CLDPS	5	8.63
C2RPS	4	6.26
CQS	3	8.43^*

We can also see from Theorem 2 (Theorem 3) that the poor fit of the S model is caused by the influence of the lack of structure of the ME model (the GS and ME models) rather than that of the LDPS (the 2RPS) model.

The S model also indicates the structure of symmetry of cumulative probabilities $\{G_{ij}\}$, $i \neq j$, instead of the cell probabilities $\{p_{ij}\}$, $i \neq j$. Therefore we shall next consider the reason why the S model fits the data in Table 1 poorly using the models which describe the structure of cumulative probabilities.

We see from Theorem 4 that the poor fit of the S model is caused by the influence of the lack of structures of both the CQS and MH models.

We also see from Theorem 5 that the poor fit of the S model is caused by the influence of the lack of structure of the ME model rather than the CLDPS model.

In more details, we can see from Theorem 6 that the poor fit of the S model is caused by the influence of the lack of structures of the GS and ME models rather than the C2RPS model.

5. Concluding Remarks

Theorems 1 through 6 would be useful for seeing the reason for poor fit of the S model when the S model fits the data poorly.

When we are interested in the structure of symmetry of cumulative probabilities $\{G_{ij}\}$, $i \neq j$, instead of the cell probabilities $\{p_{ij}\}$, $i \neq j$, Theorems 4, 5 and 6 would be useful. Especially, Theorem 6 rather than Theorem 5 would be useful for seeing in more details the reason for poor fit of the S model when the S model fits the data poorly.

References

[1] A Agresti. A simple diagonals-parameter symmetry and quasi-symmetry model. *Statistics and Probability Letters*, 1:313–316, 1983.

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

[3] A H Bowker. A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 43:572–574, 1948.

[4] H Caussinus. Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse*, 29:77–182, 1965.

[5] L A Goodman. Multiplicative models for square contingency tables with ordered categories. *Biometrika*, 66:413–418, 1979.

[6] P McCullagh. A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, 65:413–418, 1978.

[7] N Miyamoto, W Ohtsuka, and S Tomizawa. Linear diagonals-parameter symmetry and quasi-symmetry models for cumulative probabilities in square contingency tables with ordered categories. *Biometrical Journal*, 46:664–674, 2004.

[8] A Stuart. A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 42:412–416, 1955.

[9] K Tahata and S Tomizawa. Decomposition of symmetry using two-ratios-parameter symmetry model and orthogonality for square contingency tables. *Journal of Statistics: Advances in Theory and Applications*, 1:19–33, 2009.

[10] K Tahata, H Yamamoto, and S Tomizawa. Orthogonality of decompositions of symmetry into extended symmetry and marginal equimoment for multi-way tables with ordered categories. *Austrian Journal of Statistics*, 37:185–194, 2008.

[11] S Tomizawa. Decompositions for 2-ratios-parameter symmetry model in square contingency tables with ordered categories. *Biometrical Journal*, 29:45–55, 1987.

[12] S Tomizawa. Diagonals-parameter symmetry model for cumulative probabilities in square contingency tables with ordered categories. *Biometrics*, 49:883–887, 1993.

[13] S Tomizawa, N Miyamoto, K Yamamoto, and A Sugiyama. Extensions of linear diagonal-parameter symmetry and quasi-symmetry models for cumulative probabilities in square contingency tables. *Statistica Neerlandica*, 61:273–283, 2007.

[14] S Tomizawa and K Tahata. The analysis of symmetry and asymmetry: Orthogonality of decomposition of symmetry into quasi-symmetry and marginal symmetry for multi-way tables. *Journal de la Société Française de Statistique*, 148:3–36, 2007.

[15] H Yamamoto, T Iwashita, and S Tomizawa. Decomposition of symmetry into ordinal quasi-symmetry and marginal equimoment for multi-way tables. *Austrian Journal of Statistics*, 36:291–306, 2007.

[16] K Yamamoto, S Ando, and S Tomizawa. Decomposing asymmetry into extended quasi-symmetry and marginal homogeneity for cumulative probabilities in square contingency tables. *Journal of Statistics: Advances in Theory and Applications*, 5:1–13, 2011.

[17] K Yamamoto and S Tomizawa. Analysis of unaided vision data using new decomposition of symmetry. *American Medical Journal*, 3:37–42, 2012.