Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/050101

Decomposition of Diagonals-Parameter Symmetry Model Using the Diagonal Uniform Association Symmetry Model for Square Contingency Tables

Kouji Tahata, Nobuko Miyamoto and Sadao Tomizawa*

Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

Received: 12 Jan. 2017, Revised: 5 Dec. 2017, Accepted: 10 Dec. 2017
Published online: 1 Jan. 2018

Abstract: For square contingency tables with ordered categories, the present paper gives the theorem that the diagonals-parameter symmetry model holds if and only if two models hold such that the diagonal common uniform association symmetry model and the model of equality of concordance and discordance for two diagonals with the same distance from the main diagonal in the table. An example is shown.

Keywords: Concordance, decomposition, diagonals-parameter symmetry, diagonal uniform association symmetry, discordance, model, square contingency table

1 Introduction

For 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 ith row and jth column of the table ($i = 1,\dots,r; j = 1,\dots,r$). Goodman (1979) considered the diagonals-parameter symmetry (DPS) model defined by

$$p_{ij} = \begin{cases} \delta_{j-i}\psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

where $\psi_{ij} = \psi_{ji}$. Special cases of this model obtained by putting $\delta_1 = \cdots = \delta_{r-1} = 1$ and $\delta_1 = \cdots = \delta_{r-1}$ are the symmetry model (Bowker, 1948; Bishop, Fienberg and Holland, 1975, p.282) and the conditional symmetry model (McCullagh, 1978), respectively.

Tomizawa (1991) considered the diagonal uniform association symmetry (DUS) model defined by

$$p_{ij} = \begin{cases} \delta_{j-i}\phi_{j-i}^{i-1}\psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

where $\psi_{ij} = \psi_{ji}$. A special case of this model obtained by putting $\phi_1 = \cdots = \phi_{r-2} = 1$ is the DPS model. Tomizawa (1991) also considered the diagonal common uniform association symmetry (DCUS) model defined by

$$p_{ij} = \begin{cases} \delta_{j-i}\phi^{i-1}\psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

where $\psi_{ij} = \psi_{ji}$. This model is a special case of the DUS model with $\phi_1 = \cdots = \phi_{r-2} (= \phi)$. Note that Tomizawa and Miyamoto (2007) considered the DUS and DCUS models for the cumulative probabilities instead of the cell probabilities $\{p_{ij}\}$; although the details are omitted.

* Corresponding author e-mail: kouji_tahata@is.noda.tus.ac.jp

Let T_k denote the $2\times(r-k)$ table constructed using two diagonals that are k units from the main diagonal for $k=1,\ldots,r-2$. Thus, the first row of T_k is $(p_{1,1+k},p_{2,2+k},\ldots,p_{r-k,r})$ and the second row is $(p_{1+k,1},p_{2+k,2},\ldots,p_{r,r-k})$; see Agresti (1984, p.202). For T_k table ($k=1,\ldots,r-2$), consider the odds-ratio θ_{ij} as follows;

$$\theta_{ij}=\frac{p_{ij}p_{j+1,i+1}}{p_{ji}p_{i+1,j+1}},$$

where $|i-j|=k$. Using these odds-ratios, the DPS model may be expressed as

$$\theta_{ij}=1 \quad (i=1,\ldots,r-2; j=i+1,\ldots,r-1).$$

The DCUS model may be expressed as

$$\theta_{ij}=\phi^{-1} \quad (i=1,\ldots,r-2; j=i+1,\ldots,r-1).$$

Let X and Y denote the row and column variables, respectively. Let for $k=1,\ldots,r-2$,

$$U_k=\begin{cases}1 \text{ (when } X-Y=-k),\\ 2 \text{ (when } X-Y=k),\end{cases}$$

and

$$V_k=X+Y \quad (\text{when } |X-Y|=k).$$

Tahata and Tomizawa (2009) gave the following theorem:

The DPS model holds if and only if the DUS model holds and the covariance (or correlation) of U_k and V_k are zero for all $k=1,\ldots,r-2$.

Note that Tomizawa, Miyamoto and Sakurai (2008) and Tahata, Miyamoto and Tomizawa (2008) showed that for two-way contingency tables the independence model holds if and only if the uniform association model holds and the Pearson's correlation coefficient ρ (Kendall's τ_b or Spearman's ρ_s) equals zero; although the details are omitted.

The DPS model implies the DCUS model, however, the converse does not hold. Therefore we are now interested in what structure of the probabilities is necessary for obtaining the DPS model when the DCUS model holds.

The present paper gives a decomposition of the DPS model using the DCUS model (in Section 2).

2 Decomposition of diagonals-parameter symmetry model

As described in Section 1, the DPS model implies the DCUS model, however, the converse does not hold. For T_k table, we now consider the structure to obtain the DPS model when the DCUS model holds.

Let

$$C=\sum_{s=1}^{r-2}\sum_{t=s+1}^{r-1}p_{st}p_{t+1,s+1},$$

and

$$D=\sum_{s=1}^{r-2}\sum_{t=s+1}^{r-1}p_{ts}p_{s+1,t+1}.$$

For a randomly selected pair of observations, (1) $2p_{st}p_{t+1,s+1}$ with $t-s=k$ ($k=1,\ldots,r-2$), is the probability of concordance in table T_k such that the member that ranks in the second row (i.e., $Y-X=-k\ (<0)$) rather than in the first row (i.e., $Y-X=k\ (>0)$) in table T_k ranks in column $s+t+2$ (i.e., $X+Y=s+t+2$) rather than in column $s+t$ (i.e., $X+Y=s+t$) in T_k, and (2) $2p_{ts}p_{s+1,t+1}$ with $t-s=k$ ($k=1,\ldots,r-2$), is the probability of discordance such that the member that ranks in the second row rather than in the first row in table T_k ranks in column $s+t$ rather than in column $s+t+2$ in T_k. Therefore, C and D indicate the probability of concordance and that of discordance for all tables $\{T_k\}$, $k=1,\ldots,r-2$.

We shall consider the model of equality of concordance and discordance for all tables $\{T_k\}$, by

$$C=D.$$

We shall denote this model by ECD. Then we obtain the following theorem.

Theorem 1. *The DPS model holds if and only if both the DCUS model and the ECD model hold.*

Proof. Assume that the DPS model holds. Then the DCUS model holds. Also

$$C = \sum_{s=1}^{r-2}\sum_{t=s+1}^{r-1} p_{st}p_{t+1,s+1}$$
$$= \sum_{s=1}^{r-2}\sum_{t=s+1}^{r-1} \delta_{t-s}p_{ts}p_{t+1,s+1}$$

and

$$D = \sum_{s=1}^{r-2}\sum_{t=s+1}^{r-1} p_{ts}p_{s+1,t+1}$$
$$= \sum_{s=1}^{r-2}\sum_{t=s+1}^{r-1} p_{ts}\delta_{t-s}p_{t+1,s+1}.$$

Therefore the ECD model holds. Assuming that both the DCUS model and the ECD model hold, then we shall show that the DPS model holds. Since the DCUS model holds, we see

$$C = \sum_{s=1}^{r-2}\sum_{t=s+1}^{r-1} \delta_{t-s}\phi^{s-1}p_{ts}p_{t+1,s+1}$$

and

$$D = \phi\sum_{s=1}^{r-2}\sum_{t=s+1}^{r-1} \delta_{t-s}\phi^{s-1}p_{ts}p_{t+1,s+1}.$$

Since the ECD model holds, we obtain $\phi = 1$. Thus the DPS model holds. The proof is completed.

Let x_{ij} denote the observed frequency in the (i, j) cell ($i = 1,\dots,r; j = 1,\dots,r$). Assume that a multinomial distribution applies to the $r\times r$ table. Let G^2 denote the likelihood ratio chi-squared statistic for testing goodness-of-fit of model 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 expected frequency m_{ij} under the model. The numbers of degrees of freedom (df) for testing the DPS, DCUS, and ECD models are $(r-2)(r-1)/2$, $r(r-3)/2$, and 1, respectively. Note that the df for the DPS model equals to the sum of df for the DCUS model and that for the ECD model.

3 An example

Table 1 is the data, taken from Mullins and Sites (1984), which relate mother's education to father's education for a sample of eminent black Americans (also see Tomizawa and Miyamoto, 2007).

Table 1: Cross-classification of mother's and father's education for a sample of eminent black Americans: from Mullins and Sites (1984).

Mother's	Father's education				
education	(1)	(2)	(3)	(4)	Total
(1)	81	3	9	11	104
(2)	14	8	9	6	37
(3)	43	7	43	18	111
(4)	21	6	24	87	138
Total	159	24	85	122	390

The DPS model fits these data very poorly yielding the likelihood ratio chi-squared value $G^2 = 10.96$ with 3 df. Also the DCUS model does not fit these data so well yielding $G^2 = 6.00$ with 2 df, which gives almost p-value 0.05. However, the ECD model fits these data well yielding $G^2 = 2.75$ with 1 df. From Theorem 1, we can see that the poor fit of the DPS model is caused by the influence of the lack of structure of the DCUS model rather than the ECD model.

4 Concluding remarks

We have given a theorem (Theorem 1) that the DPS model holds if and only if both the DCUS model and the ECD model hold. For a given data, when the DPS model fits poorly, Theorem 1 would be useful for seeing the reason why the DPS model fits the data poorly; namely, which of the lack of structure of the DCUS model and that of the ECD model influences strongly.

Acknowledgement

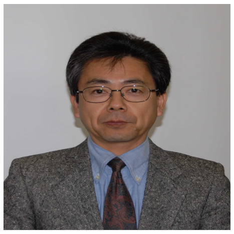
The authors would like to thank the editor and two referees for their helpful comments and suggestions.

References

[1] Agresti, A. (1984). *Analysis of Ordinal Categorical Data.* Wiley, New York.
[2] Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* The MIT Press, Cambridge, Massachusetts.
[3] Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, **43**, 572-574.
[4] Goodman, L. A. (1979). Multiplicative models for square contingency tables with ordered categories. *Biometrika*, **66**, 413-418.
[5] McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, **65**, 413-418.
[6] Mullins, E. I. and Sites, P. (1984). The origins of contemporary eminent black americans: a three-generation analysis of social origin. *American Sociological Review*, **49**, 672-685.
[7] Tahata, K. and Tomizawa, S. (2009). Decomposition of diagonals-parameter symmetry model and orthogonality of its test statistic for square contingency tables with ordered categories. *Journal of Statistical Theory and Applications*, **8**, 406-415.
[8] Tahata, K., Miyamoto, N. and Tomizawa, S. (2008). Decomposition of independence using Pearson, Kendall and Spearman's correlations and association model for two-way classifications. *Far East Journal of Theoretical Statistics*, **25**, 273-283.
[9] Tomizawa, S. (1991). Diagonal uniform association symmetry model for square contingency tables with ordered categories. *The New Zealand Statistician*, **26**, 10-17.
[10] Tomizawa, S. and Miyamoto, N. (2007). Diagonal uniform association symmetry models for cumulative probabilities in square tables. *Advances in Statistical Analysis*, **91**, 269-278.
[11] Tomizawa, S., Miyamoto, N. and Sakurai, M. (2008). Decomposition of independence model and separability of its test statistic for two-way contingency tables with ordered categories. *Advances and Applications in Statistics*, **8**, 209-218.

Kouji Tahata is Junior Associate Professor of Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science. He received the Ph. D. (Science) at Tokyo University of Science. His research interests are in the areas of mathematical statistics and categorical data analysis, especially, analysis of contingency tables.

Nobuko Miyamoto is Associate Professor of Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science. She received the PhD degree in Policy and Planning Sciences at University of Tsukuba. Her research interests are in the areas of mathematical statistics and design of experiments.



Sadao Tomizawa is Professor of Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science. He received the Doctoral degree in the science at Tokyo University of Science. His research interests are in the areas of mathematical statistics, multivariate statistical analysis, especially, analysis of contingency tables.