Collapsed Double Symmetry Model and Its Decomposition for Square Contingency Tables

Kouji Yamamoto[1], Yuya Matsuda[2] & Sadao Tomizawa[2]

[1] Department of Clinical Epidemiology and Biostatistics, Graduate School of Medicine, Osaka University

[2] Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science

Correspondence: Kouji Yamamoto, Department of Clinical Epidemiology and Biostatistics, Graduate School of Medicine, Osaka University, 2-2, Yamadaoka, Suita, Osaka, Japan. Tel: 81-6-6210-8373. E-mail: yamamoto-k@stat.med.osaka-u.ac.jp

Abstract

For a square contingency table with ordinal categories, there may be a case that one wants to analyze several collapsed tables obtained by combining some adjacent categories of the original table. This paper proposes some new models which indicate double symmetry, quasi double symmetry and marginal double symmetry for the collapsed square tables. It also gives a decomposition of the double symmetry model for collapsed tables. Two kinds of occupational mobility data are analyzed using new models.

Keywords: Collapsed table, decomposition, double symmetry, marginal double symmetry, quasi double symmetry

1. Introduction

Consider square contingency tables having same ordinal row and column classifications, for instance, as Tables 1a and 2a. For these data, a lot of observations tend to fall in (or near) the main diagonal cells. Thus for such data, in many cases, the independence between the row and column variables does not hold. So, we are interested in symmetry instead of independence. Bowker (1948) proposed the symmetry (S) model, which indicates that the probabilities are symmetric with respect to the main diagonal of the table. Wall and Lienert (1976) proposed the point-symmetry (PS) model, which indicates that the probabilities are symmetric with respect to the center point in the table. Tomizawa (1985) proposed the double symmetry (DS) model such that both S and PS models hold. For other models of symmetry, see, for example, Caussinus (1965), Bishop, Fienberg and Holland (1975, Chap.8), Agresti (2013, Chap.11), and Tahata and Tomizawa (2014).

Table 1. Occupational status for Japanese father-daughter pairs; from Hashimoto (2003, p.144).

(a) Original table

Father's class	Daughter's class					Total
	(1)	(2)	(3)	(4)	(5)	
(1)	10	5	38	14	1	68
(2)	2	22	74	31	6	135
(3)	5	8	112	20	9	154
(4)	5	11	94	44	11	165
(5)	9	15	147	26	70	267
Total	31	61	465	135	97	789

(1) Capitalist, (2) New middle, (3) Working, (4) Self-employed and (5) Farming.

(b) Collapsed table ($T^{(1,4)}$ table)

Father's class	Daughter's class			Total
	High	Middle	Low	
High	10	57	1	68
Middle	12	416	26	454
Low	9	188	70	267
Total	31	661	97	789

(c) Collapsed table ($T^{(2,3)}$ table)

Father's class	Daughter's class			Total
	High	Middle	Low	
High	39	112	52	203
Middle	13	112	29	154
Low	40	241	151	432
Total	92	465	232	789

Table 2. The social status of occupations for the husband's father and the wife's father; from Katz (1978). (The upper and lower parenthesized values are the MLEs of expected frequencies under the CoDS and CoMDS models, respectively.)

(a) Original table

Wife's father's status	Husband's father's status				Total
	(1)	(2)	(3)	(4)	
(1)	44	17	4	12	77
	(38.00)	(22.87)	(5.38)	(12.50)	(78.75)
	(38.47)	(20.15)	(4.74)	(15.65)	(79.01)
(2)	10	3	6	2	21
	(7.24)	(3.00)	(6.00)	(2.35)	(18.59)
	(7.69)	(3.00)	(6.00)	(2.17)	(18.86)
(3)	29	7	22	22	80
	(21.01)	(7.00)	(22.00)	(25.90)	(75.91)
	(22.30)	(7.00)	(22.00)	(23.82)	(75.12)
(4)	13	8	21	32	74
	(12.50)	(7.79)	(20.46)	(38.00)	(78.75)
	(10.55)	(8.58)	(22.51)	(37.37)	(79.01)
Total	96	35	53	68	252
	(78.75)	(40.66)	(53.84)	(78.75)	(252.00)
	(79.01)	(38.73)	(55.25)	(79.01)	(252.00)

(1) Professional, technical, kindred white-collar managers, officials, (2) Clerical and sales white collar workers, (3) Craftsmen (blue collar workers), (4) Operatives, service workers, laborers excepting farm workers (blue collar workers)

(b) Collapsed table ($T^{(1,3)}$ table)

Wife's father's status	Husband's father's status			Total
	High	Middle	Low	
High	44	21	12	77
	(38.00)	(28.25)	(12.50)	(78.75)
	(38.47)	(24.89)	(15.65)	(79.01)
Middle	39	38	24	101
	(28.25)	(38.00)	(28.25)	(94.50)
	(29.99)	(38.00)	(25.99)	(93.98)
Low	13	29	32	74
	(12.50)	(28.25)	(38.00)	(78.75)
	(10.55)	(31.09)	(37.37)	(79.01)
Total	96	88	68	252
	(78.75)	(94.50)	(78.75)	(252.00)
	(79.01)	(93.98)	(79.01)	(252.00)

For analyzing the data in Tables 1a and 2a, in some cases, we would like to divide the occupational status into a simple categories, for example, "High", "Middle" and "Low". Table 1b is the collapsed table with "High" category which is

"(1) capitalist" in Table 1a, "Middle" category obtained by combining "(2) new middle", "(3) working" and "(4) self-employed" categories in Table 1a, and "Low" category being "(5) farming" in Table 1a. In other words, Table 1b is obtained by collapsing Table 1a into the 3×3 table by using cut points after the first and fourth rows and after the first and fourth columns. Yamamoto, Tahata and Tomizawa (2012) proposed some S models for collapsed square contingency tables, for example, the collapsed quasi-symmetry model. Yamamoto, Murakami and Tomizawa (2013) proposed some PS models for collapsed square contingency tables, for example, the collapsed quasi point-symmetry model.

Tomizawa (1985) gave the theorem that the DS model holds if and only if both the quasi double symmetry (QDS) and marginal double symmetry (MDS) models hold. See Appendix for the DS, QDS and MDS models.

In the present paper, Section 2 proposes some new models for the collapsed tables and gives the decomposition for the new model. These proposed models are totally different from existing models. Section 3 describes the goodness-of-fit test, and analyzes the data in Tables 1a and 2a using new models. Section 4 provides remarks.

2. New Models and Decomposition

Consider the $r \times r$ contingency table having ordered categories. Let p_{ij} denote the probability that an observation will fall in the (i, j)th cell of the $r \times r$ table ($i = 1, \ldots, r; j = 1, \ldots, r$).

We consider the $[(r-1)/2]$ ways of collapsing an $r \times r$ table into a 3×3 table by using cut points after the hth and $h'(= r - h)$th rows and columns for $h = 1, ..., [(r-1)/2]$, where

$$\left[\frac{r-1}{2}\right] = \begin{cases} \dfrac{r-2}{2} & \text{(when } r \text{ is even)}, \\ \dfrac{r-1}{2} & \text{(when } r \text{ is odd)}. \end{cases}$$

We would like to refer to such a collapsed 3×3 table as the $T^{(h,h')}$ table ($h = 1, ..., [(r-1)/2]$). The $G_{kl}^{(h,h')}$ denotes the cumulative probability in the collapsed $T^{(h,h')}$ table ($k = 1, 2, 3; l = 1, 2, 3; h = 1, ..., [(r-1)/2]$); i.e.,

$$G_{11}^{(h,h')} = \sum_{i=1}^{h} \sum_{j=1}^{h} p_{ij}, \quad G_{12}^{(h,h')} = \sum_{i=1}^{h} \sum_{j=h+1}^{h'} p_{ij}, \quad G_{13}^{(h,h')} = \sum_{i=1}^{h} \sum_{j=h'+1}^{r} p_{ij},$$

$$G_{21}^{(h,h')} = \sum_{i=h+1}^{h'} \sum_{j=1}^{h} p_{ij}, \quad G_{22}^{(h,h')} = \sum_{i=h+1}^{h'} \sum_{j=h+1}^{h'} p_{ij}, \quad G_{23}^{(h,h')} = \sum_{i=h+1}^{h'} \sum_{j=h'+1}^{r} p_{ij},$$

$$G_{31}^{(h,h')} = \sum_{i=h'+1}^{r} \sum_{j=1}^{h} p_{ij}, \quad G_{32}^{(h,h')} = \sum_{i=h'+1}^{r} \sum_{j=h+1}^{h'} p_{ij}, \quad G_{33}^{(h,h')} = \sum_{i=h'+1}^{r} \sum_{j=h'+1}^{r} p_{ij}.$$

We propose the collapsed double symmetry (CoDS), collapsed quasi double symmetry (CoQDS) and collapsed marginal double symmetry (CoMDS) models.

First, we propose the CoDS model defined by

$$G_{kl}^{(h,h')} = G_{lk}^{(h,h')} = G_{k^*l^*}^{(h,h')} = G_{l^*k^*}^{(h,h')}$$

for $k = 1, 2, 3; l = 1, 2, 3$, and $h = 1, \ldots, [(r-1)/2]$, where $k^* = 4 - k$ and $l^* = 4 - l$. This model indicates both S and PS models hold in each collapsed $T^{(h,h')}$ table ($h = 1, ..., [(r-1)/2]$), namely, a structure of DS in each collapsed $T^{(h,h')}$ table. We point out that the CoDS model is totally different from the DS model, because the CoDS model indicates that there are DS structures in *all* collapsed tables. We note that the CoDS model holds if the DS model holds, but the converse does not necessarily hold.

Consider collapsing the r categories of an original table with ordered categories into 3 categories (say, groups A, B and C), by using cut points h and h'. The CoDS model describes that for each collapsed $T^{(h,h')}$ table ($h = 1, ..., [(r-1)/2]$), (i) the probability that both of the row and column values of an observation are in group A equals to the probability that both of the row and column values of it are in group C, (ii) the probability that the row and column values are in groups A and B, respectively, equals to the probability that the row and column values are in groups B and A, respectively, and it also equals to the probability that the row and column values are in groups C and B, respectively, and moreover it equals to the probability that the row and column values are in groups B and C, respectively, and (iii) the probability that the row and column values are in groups A and C, respectively, equals to the probability that the row and column values are in groups C and A, respectively.

Next, we propose the CoQDS model defined by

$$G_{kl}^{(h,h')} = \mu^{(h)}\alpha_k^{(h)}\beta_l^{(h)}\psi_{kl}^{(h)},$$

for $k = 1, 2, 3; l = 1, 2, 3$, and $h = 1, \ldots, [(r-1)/2]$, where

$$\psi_{kl}^{(h)} = \psi_{lk}^{(h)} = \psi_{k^*l^*}^{(h)} = \psi_{l^*k^*}^{(h)}.$$

Note that a special case of the CoQDS model obtained by putting $\{\alpha_k^{(h)} = \beta_k^{(h)}\}$ and $\{\alpha_k^{(h)} = \alpha_{k^*}^{(h)}\}$ is the CoDS model. Also noting that the CoQDS model is identical to the collapsed quasi point-symmetry model in Yamamoto et al. (2013) because the QDS model is identical to the quasi point-symmetry model only when the square contingency table is the 3×3 table. For the quasi point-symmetry model, see Yamamoto et al. (2013).

Finally, we propose the CoMDS model defined by

$$G_{k\cdot}^{(h,h')} = G_{k^*\cdot}^{(h,h')} = G_{\cdot k}^{(h,h')} = G_{\cdot k^*}^{(h,h')},$$

for $k = 1, 2, 3$, and $h = 1, \ldots, [(r-1)/2]$, where

$$G_{k\cdot}^{(h,h')} = \sum_{t=1}^{3} G_{kt}^{(h,h')}, \quad G_{\cdot k}^{(h,h')} = \sum_{s=1}^{3} G_{sk}^{(h,h')}.$$

This model states that there is a structure of MDS in each collapsed $T^{(h,h')}$ table ($h = 1, ..., [(r-1)/2]$). So, the CoMDS model is different from the MDS model. The MDS indicates that the row and column marginal distribution are symmetric and point-symmetric with respect to the midpoint of the row and column categories. Note that the CoMDS model holds if the MDS model holds, but the converse holds only when r is odd.

In a similar manner to the explanation of the CoDS model, the CoMDS model indicates that for each collapsed $T^{(h,h')}$ table ($h = 1, ..., [(r-1)/2]$), (i) the probability that the row value is in group A equals to the probability that the row value is in group C, and it is also equal to the probability that the column value is in group A, and moreover it is equal to the probability that the column value is in group C, and (ii) the probability that the row value is in group B equals to the probability that the column value is in group B; namely, for the original table, the probability that the row variable is h or below is equal to the probability that it is h' or above ($h = 1, ..., [(r-1)/2]$), and it is also equal to the probability that the column variable is h or below, and moreover it is equal to the probability that the column variable is h' or above ($h = 1, ..., [(r-1)/2]$).

Applying the decomposition theorem of the DS model in Tomizawa (1985) (described in Section 1) for collapsed 3×3 table, we can obtain the following theorem:

Theorem 1 *The CoDS model holds if and only if both the CoQDS and CoMDS models hold.*

3. Analysis of Data

3.1 Goodness-of-fit Test

Assume that a random sample of fixed size n is cross-classified according to the categorical variables. The distribution of the cell counts $\{n_{ij}\}$ is then the multinomial distribution specified by the sample size n and the population cell probabilities $\{p_{ij}\}$. We can obtain the maximum likelihood estimates (MLEs) of expected frequencies under the models by using the Newton-Raphson method in the log-likelihood equation. The likelihood-ratio approach to testing models leads to the test statistic

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

where $\widehat{m}_{ij}$ is the MLE of expected frequency m_{ij} under the model. The number of degrees of freedom for the CoDS model is $5(r-2)/2$ when r is even and $5(r-1)/2$ when r is odd, that for the CoQDS model is $r-2$ when r is even and $r-1$ when r is odd, and that for the CoMDS model is $3(r-2)/2$ when r is even and $3(r-1)/2$ when r is odd. Note that the number of degrees of freedom for the CoDS model is equal to the sum of that for the CoQDS model and that of the CoMDS model.

3.2 Analysis of Data in Table 1

Consider the data in Table 1a, taken from Hashimoto (2003, p.144), which describes the cross-classification for father's and his daughter's occupational status categories in Japan. Table 1b is obtained by collapsing Table 1a into the 3×3 table by using cut points after the first and fourth rows and columns. Table 1c is obtained by collapsing Table 1a into the 3×3

table by using cut points after the second and third rows and columns. Table 3 gives the values of the goodness-of-fit test statistic G^2 for the models applied to the data in Table 1a. We see from Table 3 that each of CoQDS and QDS models fits the data in Table 1a well although each of CoDS, CoMDS, DS and MDS models fits those data poorly.

Since the CoDS model fits the data in Table 1a very poorly, we shall explore a reason of the poor fit of the CoDS model by considering the decomposition of the CoDS model. From Theorem 1, we can see that the lack of structure of the CoDS model is caused by the influence of the lack of structure of CoMDS model rather than the CoQDS model.

3.3 Analysis of Data in Table 2

Consider the data in Table 2a, taken from Katz (1978), which describes the cross-classification of the social status of occupations for the husband's father and the wife's father. Table 2b is obtained by collapsing Table 2a into the 3×3 table by using cut points after the first and third rows and columns. We see from Table 3 that each of CoDS, CoQDS, and CoMDS models fits the data in Table 2a well although each of DS, QDS, and MDS models fits those data poorly.

Table 3. Likelihood ratio chi-squared statistic G^2 values for models applied to Tables 1a and 2a.

Models	Table 1a		Table 2a	
	Degree of freedom	G^2	Degree of freedom	G^2
CoDS	10	463.34*	5	8.33
CoQDS	4	6.77	2	1.78
CoMDS	6	445.18*	3	6.44
DS	16	476.87*	10	62.07*
QDS	10	11.04	5	12.75*
MDS	6	445.18*	5	47.14*

* means significant at the 0.05 level.

We denote the MLE of $nG_{kl}^{(h,h')}$ by $\widehat{M}_{kl}^{(h,h')}$ where $n = 252$ (sample size). Under the CoDS model, we obtain

$$\widehat{M}_{11}^{(1,3)} = \widehat{M}_{33}^{(1,3)} = 38.00, \ \widehat{M}_{12}^{(1,3)} = \widehat{M}_{21}^{(1,3)} = \widehat{M}_{32}^{(1,3)} = \widehat{M}_{23}^{(1,3)} = 28.25,$$

and

$$\widehat{M}_{13}^{(1,3)} = \widehat{M}_{31}^{(1,3)} = 12.50.$$

Namely, (i) the probability that the status for the wife's father and the husband's father are both "High" is estimated to be equal to the probability that those are both "Low", (ii) the probability that wife's father's status is "High" and husband's father's status is "Middle" is estimated to be equal to the probability that wife's father's status is "Middle" and husband's father's status is "High", to the probability that wife's father's status is "Low" and husband's father's status is "Middle", and to the probability that wife's father's status is "Middle" and husband's father's status is "Low", and (iii) the probability that wife's father's status is "High" and husband's father's status is "Low" is estimated to be equal to the probability that wife's father's status is "Low" and husband's father's status is "High".

Under the CoMDS model, we obtain

$$\widehat{M}_{1\cdot}^{(1,3)} = \widehat{M}_{\cdot1}^{(1,3)} = \widehat{M}_{3\cdot}^{(1,3)} = \widehat{M}_{\cdot3}^{(1,3)} = 79.01, \ \widehat{M}_{2\cdot}^{(1,3)} = \widehat{M}_{\cdot2}^{(1,3)} = 93.98,$$

where

$$\widehat{M}_{k\cdot}^{(h,h')} = \sum_{t=1}^{3} \widehat{M}_{kt}^{(h,h')}, \ \widehat{M}_{\cdot l}^{(h,h')} = \sum_{s=1}^{3} \widehat{M}_{sl}^{(h,h')}.$$

Namely, (i) the probability that wife's father's status is "High" is estimated to be equal to the probability that wife's father's status is "Low", to the probability that husband's father's status is "High", and to the probability that husband's father's status is "Low", and (ii) the probability that wife's father's status is "Middle" is estimated to be equal to the probability that husband's father's status is "Middle".

4. Remarks

For the analysis of square contingency tables having same ordered row and column classifications, there may be a case that one wants to divide the categories having more than 3 categories into the simpler 3 categories. Then the CoDS, CoQDS and CoMDS models would be useful for seeing the structures of various DS for the $[(r-1)/2]$ ways of collapsed 3×3 tables.

Each of the DS and MDS models is invariant under the combining of adjacent categories as $T^{(h,h')}$ tables ($h = 1, \ldots, [(r - 1)/2]$). Thus, if the DS model (or the MDS model) holds for an original $r \times r$ table, then the DS model (or the MDS model) holds for *all* the collapsed square tables, i.e., the CoDS model (or the CoMDS model) holds; but the converse does not always hold.

However, the QDS model is not invariant under combining of adjacent categories as $T^{(h,h')}$ tables. Therefore, when QDS model holds for an original $r \times r$ table, it is not guaranteed that then the QDS model holds for *all* the collapsed square tables.

Each of proposed models should be applied for the data on an ordinal scale because each model is not invariant under the arbitrary same permutation of the categories of rows and columns.

Acknowledgements

Appendix

For the $r \times r$ contingency table with the cell probabilities $\{p_{ij}\}$, the DS, QDS, and MDS models in Tomizawa (1985) are defined as follows. The DS model is

$$p_{ij} = p_{ji} = p_{i^*j^*} = p_{j^*i^*} \qquad (i = 1, \ldots, r; j = 1, \ldots, r),$$

where $i^* = r + 1 - i$ and $j^* = r + 1 - j$. The QDS model is

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

where

$$\psi_{ij} = \psi_{ji} = \psi_{i^*j^*} = \psi_{j^*i^*}.$$

The MDS model is

$$p_{i\cdot} = p_{\cdot i} = p_{i^*\cdot} = p_{\cdot i^*} \qquad (i = 1, \ldots, r),$$

where

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

References

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). New Jersey: Wiley.

Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: The MIT Press. http://doi.org/10.1007/978-0-387-72806-3

Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association, 43*, 572-574. http://doi.org/10.1080/01621459.1948.10483284

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

Hashimoto, K. (2003). *Class Structure in Contemporary Japan*. Melbourne: Trans Pacific Press. http://doi.org/10.4057/jsr.50.164

Katz, B. M. (1978). Tests for equality of correlated proportions in a polychotomous response design. *Journal of Educational Statistics, 3*, 145-178. http://doi.org/10.2307/1164882

Tahata, K. and Tomizawa, S. (2014). Symmetry and asymmetry models and decompositions of models for contingency tables. *SUT Journal of Mathematics, 50*, 131-165.

Tomizawa, S. (1985). Double symmetry model and its decomposition in a square contingency table. *Journal of the Japan Statistical Society, 15*, 17-23. http://doi.org/10.11329/jjss1970.15.17

Wall, K. D. and Lienert, G. A. (1976). A test for point-symmetry in J-dimensional contingency-cubes. *Biometrical Journal, 18*, 259-264.

Yamamoto, K., Murakami, S. and Tomizawa, S. (2013). Point-symmetry models and decomposition for collapsed square contingency tables. *Journal of Applied Statistics, 40*, 1446-1452. http://doi.org/10.1080/02664763.2013.786028

Yamamoto, K., Tahata, K. and Tomizawa, S. (2012). Some symmetry models for the analysis of collapsed square contingency tables with ordered categories. *Calcutta Statistical Association Bulletin, 64*, 21-36.

Copyrights