Journal of Statistics: Advances in Theory and Applications
Volume 5, Number 1, 2011, Pages 1-13

DECOMPOSING ASYMMETRY INTO EXTENDED QUASI-SYMMETRY AND MARGINAL HOMOGENEITY FOR CUMULATIVE PROBABILITIES IN SQUARE CONTINGENCY TABLES

KOUJI YAMAMOTO[1], SHUJI ANDO[2] and SADAO TOMIZAWA[2]

[1]Center for Clinical Investigation and Research
Osaka University Hospital
2-15, Yamadaoka, Suita
Osaka, 565-0871
Japan
e-mail: yamamoto-k@hp-crc.med.osaka-u.ac.jp

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

Abstract

For the analysis of square contingency tables with ordered categories, the present paper proposes two kinds of marginal asymmetry models and gives decompositions of models, which have the structure of asymmetry for cumulative probabilities that an observation will fall in row (column) category i or below and column (row) category $j\,(> i)$ or above. The decompositions are obtained by using the extended

2010 Mathematics Subject Classification: 62H17.

Keywords and phrases: cumulative probability, cumulative quasi-symmetry, decomposition, extended marginal homogeneity, linear diagonals-parameter symmetry, square contingency table, symmetry.

Received February 7, 2011

quasi-symmetry, the extended marginal homogeneity, and the marginal asymmetry models, for cumulative probabilities. The new models and decompositions are applied to the father's and his son's occupational mobility data.

1. Introduction

Consider 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 i-th row and j-th column of the table $(i = 1, \dots, R;\ j = 1, \dots, R)$.

Generally, many observations tend to fall in (or near) the main diagonal cells of the square table. Thus, the independence between the row and column classifications is unlikely to hold. So, we are interested in whether or not there is a structure of symmetry, instead of independence, in the table.

For the analysis of such data, we may use models, which represent the structure of symmetry, based on the cell probabilities $\{p_{ij}\}$ (for example, Caussinus [4]; Bishop et al. [2], p. 282; Goodman [5]; Agresti [1]; and Tomizawa [9]). As another approach, we may use models based on the cumulative probabilities $\{G_{ij}\}$, defined by

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

and

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

Tomizawa [10, 11] considered an extended marginal homogeneity (EMH) model, which indicates that the row marginal totals summed by multiplying the probabilities for the cells in the lower left (or upper right) triangle of the square table by a common weight are equal to the column marginal totals summed in the same way. The EMH model may also be expressed as a multiplicative form for $\{G_{ij}\}$, $i \neq j$, as follows:

$$G_{ij} = \begin{cases} \Delta\Psi_{ij} & (j - i = 1), \\ \Psi_{ij} & (j - i \neq 1, 0), \end{cases} \quad p_{ii} = \Psi_{ii},$$

where $\Psi_{ij} = \Psi_{ji}$ for $|j - i| = 1$ (see also Tomizawa [12]). A special case of this model obtained by putting $\Delta = 1$ is equivalent to Stuart's [7] marginal homogeneity (MH) model.

Miyamoto et al. [6] considered two models. One is the cumulative linear diagonals-parameter symmetry (CLDPS) model defined by

$$G_{ij} = \begin{cases} \Theta^{j-i}\Psi_{ij} & (i < j), \\ \Psi_{ij} & (i > j), \end{cases} \quad p_{ii} = \Psi_{ii},$$

where $\Psi_{ij} = \Psi_{ji}$, and the other is the cumulative quasi-symmetry (CQS) model defined by

$$G_{ij} = \alpha_i\beta_j\Psi_{ij} \quad (i = 1, \dots, R;\ j = 1, \dots, R,\ i \neq j), \quad p_{ii} = \Psi_{ii},$$

where $\Psi_{ij} = \Psi_{ji}$. A special case of the CLDPS model obtained by putting $\Theta = 1$ is equivalent to Bowker's [3] symmetry (S) model. The CQS model with $\{G_{ij}\}$ replaced by $\{p_{ij}\}$ is Caussinus's [4] quasi-symmetry (QS) model.

Tomizawa et al. [14] considered the extensions of the CLDPS and CQS models. The cumulative two-ratios-parameter symmetry (C2RPS) model is defined by

$$G_{ij} = \begin{cases} \Gamma\Theta^{j-i}\Psi_{ij} & (i < j), \\ \Psi_{ij} & (i > j), \end{cases} \quad p_{ii} = \Psi_{ii},$$

where $\Psi_{ij} = \Psi_{ji}$, and the cumulative extended quasi-symmetry (CEQS) model is defined by

$$G_{ij} = \alpha_i\beta_j\Psi_{ij} \quad (i = 1, \dots, R;\ j = 1, \dots, R,\ i \neq j), \quad p_{ii} = \Psi_{ii},$$

where $\Psi_{ij} = \gamma\Psi_{ji}$ for $i < j$. Some other models based on the cumulative probabilities have been considered, although the details are omitted (for example, Tomizawa [12]; Tomizawa et al. [13]; Tomizawa and Tahata [15]).

The present paper proposes decompositions of the C2RPS, CLDPS, and S models. Section 2 gives two new models. Section 3 gives the decompositions.

2. Cumulative Symmetry Models

Consider two models defined by

$$G_i^+ = \Delta_1 \sum_{k=i+1}^{R} \Omega^{k-i} G_{ki} \quad (i = 1, \ldots, R-1), \tag{2.1}$$

where

$$G_i^+ = \sum_{k=i+1}^{R} G_{ik},$$

and

$$G_i^- = \Delta_2 \sum_{k=1}^{i-1} \Lambda^{i-k} G_{ki} \quad (i = 2, \ldots, R), \tag{2.2}$$

where

$$G_i^- = \sum_{k=1}^{i-1} G_{ik}.$$

We shall refer to models (2.1) and (2.2) as the cumulative two-weights modified marginal symmetry (C2WM) 1 and 2 models, respectively. The C2WM-1 model indicates that for $R \times R$ table based on cumulative probabilities modified by multiplying the cumulative probability G_{ki} in the lower left triangle of the table by two weights Δ_1 and Ω^{k-i}, the row totals summed in the upper right triangle of the table are equal to the column totals summed in the lower left triangle of the table. The C2WM-2 model also represents a similar structure. Denote the row and column variables by X and Y, respectively. Under the C2WM-1 model, if $\Delta_1 \Omega^l \geq 1$ for $l = 1, \ldots, R-1$, then $\Pr(X < Y) \geq \Pr(X > Y)$. The C2WM-2 model also has a similar property.

We shall refer to the C2WM-1 (C2WM-2) model with $\Delta_1 = 1\,(\Delta_2 = 1)$ as the C1WM-1 (C1WM-2) model.

3. Decompositions of Cumulative Asymmetry Models

In this section, we shall propose decompositions of the C2RPS, CLDPS, and S models.

The CEQS model may be expressed as

$$\frac{G_{ij}}{G_{ji}} = \gamma \frac{\gamma_i}{\gamma_j} \quad (i < j).$$

Note that we may set, e.g., $\gamma_R = 1$ without loss of generality. Also, this model may be expressed as

$$\frac{G_{ij}G_{jk}G_{ki}}{G_{kj}G_{ji}G_{ik}} = \gamma \quad (i < j < k);$$

see Tomizawa et al. [14].

Let

$$D = \frac{\sum_{i<j<k} G_{ij}G_{jk}G_{ki}}{\sum_{i<j<k} G_{kj}G_{ji}G_{ik}}.$$

Then, we consider the model defined by

$$D = \Delta_1, \tag{3.1}$$

where Δ_1 satisfies specified two equations in (2.1). Similarly, we consider the model defined by

$$D = \frac{1}{\Delta_2}, \tag{3.2}$$

where Δ_2 satisfies specified two equations in (2.2). We will refer to models (3.1) and (3.2) as the D-1 and D-2 models, respectively. Especially, when we specify two equations for $i = s$ and $i = t$ in (2.1), we shall denote the D-1 model by D-1(s, t). Similarly, when we specify two equations for $i = m$ and $i = n$ in (2.2), we shall denote the D-2 model by D-2(m, n). Then, we obtain the decomposition of the C2RPS model as follows:

Theorem 1. *For $t = 1, 2$, the C2RPS model holds, if and only if all the CEQS, C2WM-t, and D-t models hold.*

Proof. For $t = 1, 2$, if the C2RPS model holds, then all the CEQS, C2WM-t, and D-t models hold. Assume that the CEQS, C2WM-t, and D-t models hold, and then we shall show that the C2RPS model holds. Consider the case of $t = 1$. From the assumption, we get $D = \gamma$. Thus $\gamma = \Delta_1$.

From the C2WM-1 model, we see

$$G^{+}_{R-1} = G_{R-1, R} = \gamma\Omega G_{R, R-1}.$$

From the CEQS model, we see

$$G_{R-1, R} = \gamma \frac{\gamma_{R-1}}{\gamma_R} G_{R, R-1},$$

with $\gamma_R = 1$ without loss of generality. Thus $\gamma_{R-1} = \Omega$. Also, from the C2WM-1 model, we see

$$G^{+}_{R-2} = G_{R-2, R-1} + G_{R-2, R} = \gamma\Omega G_{R-1, R-2} + \gamma\Omega^2 G_{R, R-2}. \tag{3.3}$$

From the CEQS model, we see

$$G_{R-2, R-1} = \gamma \frac{\gamma_{R-2}}{\gamma_{R-1}} G_{R-1, R-2} = \gamma \frac{\gamma_{R-2}}{\Omega} G_{R-1, R-2},$$

and

$$G_{R-2, R} = \gamma \frac{\gamma_{R-2}}{\gamma_R} G_{R, R-2} = \gamma\gamma_{R-2} G_{R, R-2}.$$

Thus

$$G^{+}_{R-2} = \gamma\gamma_{R-2}\left(\frac{1}{\Omega} G_{R-1, R-2} + G_{R, R-2}\right). \tag{3.4}$$

From (3.3) and (3.4), we obtain $\gamma_{R-2} = \Omega^2$. In a similar way, we obtain $\gamma_i = \Omega^{R-i}$ for $i = 1, \ldots, R$. Therefore, we obtain

$$\frac{G_{ij}}{G_{ji}} = \gamma \frac{\gamma_i}{\gamma_j} = \gamma \Omega^{j-i} \quad (i < j).$$

Namely, we obtain the C2RPS model. Also, the case of $t = 2$ can be proved in a similar way to the case of $t = 1$. The proof is completed.

Next, we obtain the decompositions of the CLDPS model as follows:

Corollary 1. *For $t = 1, 2$, the CLDPS model holds, if and only if both the CQS and C1WM-t models hold.*

We obtain another decomposition of the C2RPS model as follows:

Theorem 2. *The C2RPS model holds, if and only if both the CEQS and EMH models hold.*

Proof. If the C2RPS model holds, then both the CEQS and EMH models hold. Assume that the CEQS and EMH models hold, and then we shall show that the C2RPS model holds. We see

$$\frac{G_{R-1,R}}{G_{R,R-1}} = \gamma \frac{\gamma_{R-1}}{\gamma_R} = \Delta,$$

with $\gamma_R = 1$. Thus,

$$\gamma_{R-1} = \frac{\Delta}{\gamma}.$$

Also, we see

$$\frac{G_{R-2,R-1}}{G_{R-1,R-2}} = \gamma \frac{\gamma_{R-2}}{\gamma_{R-1}} = \Delta.$$

Thus,

$$\gamma_{R-2} = \frac{\Delta}{\gamma} \gamma_{R-1} = \left(\frac{\Delta}{\gamma}\right)^2.$$

In a similar way, we obtain

$$\gamma_i = \left(\frac{\Delta}{\gamma}\right)^{R-i} \quad (i = 1, \ldots, R).$$

Therefore, by putting $\Omega = \Delta / \gamma,$ we obtain

$$\frac{G_{ij}}{G_{ji}} = \gamma \frac{\gamma_i}{\gamma_j} = \gamma\Omega^{j-i} \quad (i < j).$$

Namely, we obtain the C2RPS model. The proof is completed.

By putting $\gamma = 1$ in the proof of Theorem 2, we obtain another decomposition of the CLDPS model as follows:

Corollary 2. *The CLDPS model holds, if and only if both the CQS and EMH models hold.*

By putting $\Theta = 1$ in the CLDPS model and $\Delta = 1$ in the EMH model, we obtain the decomposition of the S model as follows:

Corollary 3. *The S model holds, if and only if both the CQS and MH models hold.*

By the way, Caussinus [4] gave the theorem that the S model holds, if and only if both the QS and MH models hold. We point out that the CQS model is different from the QS model, and thus, Corollary 3 is different from Caussinus's theorem.

4. Goodness-of-fit Test

Let n_{ij} denote the observed frequency in the (i, j)-th cell of the $R \times R$ table $(i = 1, \ldots, R;\ j = 1, \ldots, R).$ Assume that a multinomial distribution applies to the $R \times R$ table. The maximum likelihood estimates of expected frequencies under a model could be obtained by using, e.g., the Newton-Raphson method in the log-likelihood equation.

Each model can be tested for goodness-of-fit by, e.g., the likelihood ratio statistic (denote by G^2) with the corresponding degrees of freedom. The G^2 is given by

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

where $\hat{m}_{ij}$ is the maximum likelihood estimate of expected frequency m_{ij} under the model. The numbers of degrees of freedom for testing goodness-of-fit of the models are given in Table 1.

Table 1. The numbers of degrees of freedom for each model

Models	Degrees of freedom
S	$R(R-1)/2$
CLDPS	$(R-2)(R+1)/2$
C2RPS	$(R^2-R-4)/2$
QS	$(R-2)(R-1)/2$
CQS	$(R-2)(R-1)/2$
CEQS	$R(R-3)/2$
MH	$R-1$
EMH	$R-2$
C1WM-t ($t = 1, 2$)	$R-2$
C2WM-t ($t = 1, 2$)	$R-3$
D-t ($t = 1, 2$)	1

5. An Example

Consider the data in Table 2, taken from Tominaga ([8], p. 131). These data describe the cross-classification of father's and son's occupational status categories in Japan, which were examined in 1955.

Table 2. The cross-classification of father's and son's occupational status categories in Japan, which were examined in 1955 from Tominaga ([8], p. 131). (The upper and lower parenthesized values are maximum likelihood estimates of expected frequencies under the C2WM-1 and the C2WM-2 models, respectively.)

Father's status	Son's status				Total
	(1)	(2)	(3)	(4)	
(1)	80	72	37	19	208
	(80.00)	(69.33)	(37.10)	(18.07)	
	(80.00)	(71.44)	(37.63)	(18.76)	
(2)	44	155	61	31	291
	(44.40)	(155.00)	(66.29)	(33.13)	
	(44.32)	(155.00)	(61.29)	(30.94)	
(3)	26	73	218	45	362
	(26.34)	(71.65)	(218.00)	(41.08)	
	(25.96)	(72.68)	(218.00)	(44.94)	
(4)	69	156	166	614	1005
	(72.26)	(149.53)	(169.82)	(614.00)	
	(69.15)	(155.69)	(166.20)	(614.00)	
Total	219	456	482	709	1866

Note: Status (1) is upper non-manual, (2) lower non-manual, (3) manual, and (4) agriculture.

Table 3 gives the values of likelihood ratio test statistic G^2 for testing goodness-of-fit of each model. The QS, C2WM-1, and C2WM-2 models fit these data very well, however, the other models fit these data poorly.

Table 3. Values of likelihood ratio statistic G^2 for models applied to the data in Table 2

Models	Degrees of freedom	G^2
S	6	205.08*
CLDPS	5	113.53*
C2RPS	4	83.39*
QS	3	0.82
CQS	3	35.11*
CEQS	2	6.17*
MH	3	203.55*
EMH	2	82.68*
C1WM-1	2	98.67*
C1WM-2	2	41.33*
C2WM-1	1	1.63
C2WM-2	1	0.02
D-1(1, 2)	1	32.28*
D-1(1, 3)	1	76.04*
D-1(2, 3)	1	34.53*
D-2(2, 3)	1	36.05*
D-2(2, 4)	1	72.56*
D-2(3, 4)	1	43.89*

* means significant at the 0.05 level.

We see from Theorem 1 that, the poor fit of the C2RPS model is caused by the influence of the lack of structure of the CEQS and D-t models rather than the C2WM-t model for $t = 1, 2$.

Under the C2WM-2 model, the maximum likelihood estimates of Δ_2 and Λ are $\hat{\Delta}_2 = 0.26$ and $\hat{\Lambda} = 4.17$, respectively. Since $\hat{\Delta}_2\hat{\Lambda}^l > 1$ for $l = 1, 2, 3$, it is estimated that $\Pr(X > Y)$ is greater than $\Pr(X < Y)$; namely, it is estimated that the father's status category in a father-son pair tends to be greater than his son's status category.

6. Concluding Remark

In the present paper, we have considered some new models and have given theorems and corollaries with regard to decompositions of the C2RPS, CLDPS, and S models. As seen in example, the decomposition may be useful for exploring the reason for the poor fit when the C2RPS (CLDPS or S) model fits the data poorly.

References

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

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

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

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

[5] L. A. Goodman, Multiplicative models for square contingency tables with ordered categories, Biometrika 66 (1979), 413-418.

[6] 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 (2004), 664-674.

[7] A. Stuart, A test for homogeneity of the marginal distributions in a two-way classification, Biometrika 42 (1955), 412-416.

[8] K. Tominaga, Nippon no Kaisou Kouzou (Japanese Hierarchical Structure). University of Tokyo Press, Tokyo, 1979 (in Japanese).

[9] S. Tomizawa, Decompositions for 2-ratios-parameter symmetry model in square contingency tables with ordered categories, Biometrical Journal 29 (1987a), 45-55.

[10] S. Tomizawa, Diagonal weighted marginal homogeneity models and decompositions for linear diagonals-parameter symmetry model, Communications in Statistics-Theory and Methods 16 (1987b), 477-488.

[11] S. Tomizawa, Decompositions for conditional symmetry model into palindromic symmetry and modified marginal homogeneity models, Australian Journal of Statistics 31 (1989), 287-296.

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

[13] S. Tomizawa, N. Miyamoto and K. Yamamoto, Decomposition for polynomial cumulative symmetry model in square contingency tables with ordered categories, Metron 64 (2006), 303-314.

[14] S. Tomizawa, N. Miyamoto, K. Yamamoto and A. Sugiyama, Extensions of linear diagonals-parameter symmetry and quasi-symmetry models for cumulative probabilities in square contingency tables, Statistica Neerlandica 61 (2007), 273-283.

[15] 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 (2007), 3-36.

■