Decompositions of Marginal Homogeneity Model Using Cumulative Logistic Models for Square Contingency Tables with Ordered Categories

Nobuko Miyamoto, Kouji Niibe, and Sadao Tomizawa Tokyo University of Science, Japan

Abstract: For square contingency tables with ordered categories, Agresti (1984, 2002) considered the marginal cumulative logistic (ML) model, which is an extension of the marginal homogeneity (MH) model. The ML model depends on the probabilities on the main diagonal of the table. This paper (1) proposes the conditional marginal cumulative logistic (CML) model which does not depend on the probabilities on the main diagonal, and (2) decomposes the MH model into the ML (CML) model and the model which indicates the equality of row and column marginal means. Examples are given.

Zusammenfassung: Für quadratische Kontingenztafeln mit geordneten Kategorien betrachtete Agresti (1984, 2002) ein marginales kumulatives logistisches Modell (ML), welches eine Erweiterung des marginalen Homogenitätsmodells (MH) ist. Das ML Modell hängt von den Wahrscheinlichkeiten auf der Hauptdiagonalen der Tafel ab. In diesem Aufsatz wird (1) das konditionale marginale kumulative logistische (CML) Modell empfohlen, das nicht von den Wahrscheinlichkeiten auf der Hauptdiagonalen abhängt, und (2) das MH Modell zerlegt in das ML (CML) Modell und ein Modell, welches die Gleichheit von marginalen Zeilen- und Spaltenmittel anzeigt. Beispiele werden angeführt.

Keywords: Decomposition, Marginal Cumulative Logistic Model, Marginal Homogeneity, Marginal Mean Equivalence, Square Contingency Table.

1 Introduction

For the $R \times R$ square contingency table with ordered categories, let p_{ij} denote the probability that an observation will fall in the cell in row i, i = 1, ..., R, and column j, j = 1, ..., R, and let X and Y denote the row and column variables, respectively. First, consider the marginal homogeneity (MH) model defined by

$$\Pr(X = i) = \Pr(Y = i) \text{ for } i = 1, ..., R,$$

that is

$$p_{i} = p_{\cdot i} \quad \text{for } i = 1, \dots, R,$$
 (1)

where

$$p_{i\cdot} = \sum_{k=1}^{R} p_{ik}, \qquad p_{\cdot i} = \sum_{k=1}^{R} p_{ki}.$$

This model indicates that the row marginal distribution is the same as the column marginal distribution. For this model see also Stuart (1955), Bishop et al. (1975, p. 294), Agresti (1984, p. 207), and Tomizawa (1991, 1993b, 1998).

Let F_i^X and F_i^Y denote the marginal cumulative probabilities of X and Y, respectively. These are $F_i^X = \Pr(X \le i) = \sum_{k=1}^i p_k$ and $F_i^Y = \Pr(Y \le i) = \sum_{k=1}^i p_k$, for $i = 1, \ldots, R - 1$. Then the MH model may also be expressed as

$$F_i^X = F_i^Y$$
 for $i = 1, ..., R - 1$

Let L_i^X and L_i^Y denote the marginal cumulative logit of X and Y, respectively. These are given as

$$L_i^X = \text{logit}\left[\Pr(X \le i)\right] = \log\left[\frac{\Pr(X \le i)}{1 - \Pr(X \le i)}\right] ,$$

and

$$L_i^Y = \text{logit}\left[\Pr(Y \le i)\right] = \log\left[\frac{\Pr(Y \le i)}{1 - \Pr(Y \le i)}\right]$$

for i = 1, ..., R - 1. Then the MH model may be further expressed as

$$L_i^X = L_i^Y$$
 for $i = 1, ..., R - 1$.

As an extension of the MH model, Agresti (1984, p. 205), Agresti (2002, p. 420) considered the marginal cumulative logistic (ML) model defined by

$$L_i^X = L_i^Y + \Delta \quad \text{for } i = 1, \dots, R - 1.$$
 (2)

This model indicates that one marginal distribution is a location shift of the other marginal distribution on a logistic scale. So, this model states that the odds that X is i or below instead of i + 1 or above, is $\exp(\Delta)$ times higher than the odds that Y is i or below instead of i + 1 or above, for i = 1, ..., R - 1. If $\Delta > 0$, X rather than Y tends to be i or below instead of i + 1 or above, for i = 1, ..., R - 1. If $\Delta > 0$, X rather than Y tends to be i or below instead of i + 1 or above, for i = 1, ..., R - 1. The ML model may also be expressed as

$$F_i^X = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)}$$
 and $F_i^Y = \frac{\exp(\theta_i - \Delta)}{1 + \exp(\theta_i - \Delta)}$,

for i = 1, ..., R - 1. A special case of this model obtained by putting $\Delta = 0$ is the MH model. Therefore, the MH model implies the ML model, but the converse does not hold. Hence, we are now interested in what restrictions need to be imposed on the ML model in order that the MH model holds.

The MH model defined by (1) essentially does not depend on the probabilities $\{p_{ii}\}$ on the main diagonal of the table. Thus, the MH model may also be expressed as

$$\Pr(X = i | X \neq Y) = \Pr(Y = i | X \neq Y) \quad \text{for } i = 1, \dots, R,$$

that is

$$p_{i}^{c} = p_{.i}^{c} \quad \text{for } i = 1, \dots, R,$$
(3)

where

$$p_{i\cdot}^c = \frac{p_{i\cdot} - p_{ii}}{\delta} = \Pr(X = i | X \neq Y), \qquad \delta = \sum_{s=1}^R \sum_{\substack{t=1\\t\neq s}}^R p_{st} = \Pr(X \neq Y),$$
$$p_{\cdot i}^c = \frac{p_{\cdot i} - p_{ii}}{\delta} = \Pr(Y = i | X \neq Y).$$

Let $F_i^{X(c)}$ and $F_i^{Y(c)}$ denote the conditional marginal cumulative probabilities of X and Y given that $X \neq Y$, i.e.

$$F_i^{X(c)} = \Pr(X \le i | X \ne Y) = \sum_{k=1}^i p_{k.}^c$$

$$F_i^{Y(c)} = \Pr(Y \le i | X \ne Y) = \sum_{k=1}^i p_{\cdot k}^c,$$

for i = 1, ..., R - 1. Then the MH model may be further expressed as $F_i^{X(c)} = F_i^{Y(c)}$. We are also interested in proposing another ML model defined by (2) with $\{F_i^{X(c)}\}$ and $\{F_i^{Y(c)}\}$ instead of $\{F_i^X\}$ and $\{F_i^Y\}$. In addition, we are interested in decomposing the MH model using such another ML model.

Consider the data in Tables 1 to 3. Table 1 is taken from Stuart (1955) and contains data on distance visions. These data have been analyzed by many statisticians, including Stuart (1955), Bishop et al. (1975, p. 284), McCullagh (1978), Goodman (1979), Agresti (1984, p. 215), Tomizawa (1993a), and Miyamoto et al. (2004).

Table 1: Unaided distance vision of 7477 women aged 30-39 employed in Royal Ordnance factories from 1943 to 1946. The upper and lower parenthesized values are the MLEs of expected frequencies under the ML and CML models, respectively.

	Left eye grade				
Right eye grade	Best (1)	Second (2)	Third (3)	Worst (4)	Total
Best (1)	1520	266	124	66	1976
	(1519.45)	(269.57)	(125.40)	(65.37)	
	(1520.00)	(264.65)	(123.69)	(65.92)	
Second (2)	234	1512	432	78	2256
	(230.86)	(1511.98)	(431.09)	(76.26)	
	(235.31)	(1512.00)	(433.09)	(78.30)	
Third (3)	117	362	1772	205	2456
	(115.67)	(362.76)	(1772.00)	(200.83)	
	(117.34)	(361.02)	(1772.00)	(205.24)	
Worst (4)	36	82	179	492	789
	(36.37)	(83.99)	(182.96)	(492.44)	
	(36.05)	(81.66)	(178.72)	(492.00)	
Total	1907	2222	2507	841	7477

	Number			
Number of				
Lambs 1953	0	1	2	Total
0	58	52	1	111
	(58.00)	(50.83)	(0.96)	
1	26	58	3	87
	(26.61)	(58.00)	(2.93)	
2	8	12	9	29
	(8.39)	(12.28)	(9.00)	
Total	92	122	13	227

Table 2: Merino ewes according to number of lambs born in consecutive years. The parenthesized values are the MLEs of expected frequencies under the ME model.

Table 3: The results from two Danish polls on the question: Do you think Denmark should join the Common Market? The upper and lower parenthesized values are the MLEs of expected frequencies under the ML and CML models, respectively.

	Poll II (October 1971)			
Poll I				_
(August 1971)	Yes (1)	Undecided (2)	No (3)	Total
Yes (1)	176	40	33	249
	(176.64)	(36.27)	(33.52)	
	(176.00)	(35.21)	(32.91)	
Undecided (2)	21	43	33	97
	(23.56)	(43.11)	(37.53)	
	(25.19)	(43.00)	(38.06)	
No (3)	21	32	94	147
	(20.60)	(28.43)	(93.32)	
	(21.10)	(27.53)	(94.00)	
Total	218	115	160	493

Table 2 is taken from Tallis (1962) and constructed from the cross-classified data of Merino ewes according to the numbers of lambs born in consecutive years, 1952 and 1953 (also see Bishop et al., 1975, p. 288).

Table 3 is taken from Andersen (1980, p. 328) and includes the results of two consecutive opinion polls held in August and October 1971, in connection with the Danish referendum on whether to join the Common Market or not.

For analyzing the data of square tables with ordered categories like Tables 1, 2 and 3, the various models of symmetry or asymmetry, like e.g., symmetry, quasi-symmetry, MH, and conditional symmetry models are applied (see, e.g., Tomizawa, 1993a). These models do not depend on the main diagonal cell probabilities $\{p_{ii}\}$; namely, these indicate

the structure of symmetry or asymmetry on condition that an observation will fall in one of off-diagonal cells of the table. So, many statisticians may be interested in the structure of symmetry or asymmetry for off-diagonal cell probabilities $\{p_{ij}\}, i \neq j$. We are also interested in another ML model defined by (2) with $\{F_i^{X(c)}\}$ and $\{F_i^{Y(c)}\}$, which indicates the structure of marginal inhomogeneity on condition that an observation will fall in one of off-diagonal cells of the table.

In addition, the readers may be interested in, for example, (1) for the data in Table 1, seeing how the right eye of a woman is symmetric or asymmetric to her left eye on condition that the her right eye grade is different from her left eye grade, and (2) for the data in Table 3, seeing, when the opinion in poll II of an individual changed from in poll I, how the individual changed the attitude. So, we are again interested in models which indicate the structure of marginal inhomogeneity on condition that an observation will fall in one of off-diagonal cells of the table.

When the MH model does not hold for the data, many statisticians may be interested in applying various models of symmetry or asymmetry instead of the MH model, and in selecting a best-fitting model by using, e.g., the AIC-criterion (see Sakamoto et al., 1986). In addition, we are then interested in the reason why the MH model does not hold. Although it would be impossible to see the reason by the AIC-criterion, the decomposition of the MH model into some models may be useful to see the reason, according to which model of decomposed models fits well and which model of them fits poorly (see Section 3 for details).

Generally, consider a decomposition of model, say, M_1 , such that model M_1 holds if and only if both models M_2 and M_3 hold. When models M_1 and M_2 fit the data poorly and model M_3 fits the data well, we can then understand that the poor fit of model M_1 is caused by the lack of structure of model M_2 rather than the structure of model M_3 . Thus, the decomposition of model M_1 (especially, into two models) would be useful to see the reason for the poor fit of model M_1 .

Denote the likelihood ratio chi-squared statistic for testing the goodness-of-fit of model M by $G^2(M)$. For testing if model M_1 holds assuming that model M_2 holds, the likelihood ratio statistic is given as $G^2(M_1|M_2) = G^2(M_1) - G^2(M_2) \ge 0$. Then the decomposition of model M_1 into models M_2 and M_3 also has the advantages that (1) when model M_2 holds, the conditional statistic $G^2(M_1|M_2)$ is more powerful than the unconditional statistic $G^2(M_1)$ (see, e.g., Agresti, 1984, p. 82), and (2) from the decomposition of model M_1 , it is possible to see what the structure that model M_1 holds assuming that model M_2 holds indicates, i.e., that indicates the structure that model M_3 holds.

The purpose of this paper is (1) to propose another ML model based on $\{F_i^{X(c)}\}\$ and $\{F_i^{Y(c)}\}\$, and (2) to decompose the MH model using the two kinds of ML models. The decompositions may be useful for seeing the reason for the poor fit when the MH model fits the data poorly.

2 Decompositions of the Marginal Homogeneity Model

We shall consider two kinds of decompositions of the MH model.

2.1 A decomposition of the MH model using the ML model

Consider a model defined as

$$\sum_{i=1}^{R} i p_{i\cdot} = \sum_{i=1}^{R} i p_{\cdot i} , \qquad (4)$$

i.e., E(X) = E(Y). This indicates that the mean of the row variable X equals the mean of the column variable Y. Note that the MH model implies model (4).

Consider a specified monotonic function g(k) satisfying $g(1) \le g(2) \le \cdots \le g(R)$ or $g(1) \ge g(2) \ge \cdots \ge g(R)$, where at least one strict inequality holds. Using the function g(k), model (4) is generalized as

$$\sum_{i=1}^{R} g(i)p_{i\cdot} = \sum_{i=1}^{R} g(i)p_{\cdot i}, \qquad (5)$$

i.e., E(g(X)) = E(g(Y)). We shall refer to (5) as the marginal mean equivalence (ME) model. This indicates that the mean of g(X) is equal to the mean of g(Y).

The $\{g(k)\}$ may be considered as the ordered scores $\{u_k\}$ assigned to both row and column categories if it is possible to assign the scores; namely, $g(k) = u_k$ satisfying $u_1 \leq u_2 \leq \cdots \leq u_R$ or $u_1 \geq u_2 \geq \cdots \geq u_R$. Particular, when the scores are equalinterval; that is, when $u_2 - u_1 = u_3 - u_2 = \cdots = u_R - u_{R-1}$, then the ME model with $g(k) = u_k$ is equivalent to the model (4), i.e, the ME model with g(k) = k. Therefore, model (5) applied on examples in Section 3 may also be considered as the ME model with g(k) = k. (We note that when the scores are not equal-interval, it seems difficult to obtain an intuitive interpretation. However, in many cases, the equal-interval scores seems to be used.) We now obtain the following theorem.

Theorem 1: The MH model holds if and only if both the ML and ME models hold.

Proof. If the MH model holds, then both the ML and ME models hold. Therefore, we assume that both the ML and ME models hold, and then we show that the MH model holds. We have

$$\begin{split} \mathcal{E}(g(X)) &= \sum_{k=1}^{R} g(k) p_{k\cdot} = g(1) + \sum_{k=2}^{R} \left(d_k \sum_{i=k}^{R} p_{i\cdot} \right) = g(1) + \sum_{k=2}^{R} d_k \left(1 - F_{k-1}^X \right) \\ &= g(R) - \sum_{k=2}^{R} d_k F_{k-1}^X \,, \end{split}$$

where

$$d_k = g(k) - g(k-1) \,.$$

Similarly, we have

$$E(g(Y)) = g(R) - \sum_{k=2}^{R} d_k F_{k-1}^Y.$$

This yields

$$E(g(Y)) - E(g(X)) = \sum_{k=2}^{R} d_k \left(F_{k-1}^X - F_{k-1}^Y \right) .$$
(6)

Since the ML and ME models hold, we obtain

$$\sum_{k=2}^{R} d_k \left(\frac{\exp(\theta_{k-1})}{1 + \exp(\theta_{k-1})} - \frac{\exp(\theta_{k-1} - \Delta)}{1 + \exp(\theta_{k-1} - \Delta)} \right) = 0.$$

Thus

$$(1 - \exp(-\Delta)) \sum_{k=2}^{R} d_k \frac{\exp(\theta_{k-1})}{(1 + \exp(\theta_{k-1}))(1 + \exp(\theta_{k-1} - \Delta))} = 0.$$

Then

$$\sum_{k=2}^{R} d_k \frac{\exp(\theta_{k-1})}{(1 + \exp(\theta_{k-1}))(1 + \exp(\theta_{k-1} - \Delta))} \neq 0$$

because $d_k \ge 0$ or $d_k \le 0$ for all k = 2, ..., R, with at least one of the d_k s being not equal to zero. Therefore we obtain $\Delta = 0$, i.e., the MH model holds. The proof is completed.

2.2 A decomposition of the MH model using the CML model

Consider a model defined by

$$L_i^{X(c)} = L_i^{Y(c)} + \Delta^* \text{ for } i = 1, \dots, R-1,$$

where

$$\begin{split} L_i^{X(c)} &= \text{logit}\left[\Pr(X \le i | X \ne Y)\right] = \log\left[\frac{\Pr(X \le i | X \ne Y)}{1 - \Pr(X \le i | X \ne Y)}\right] \\ L_i^{Y(c)} &= \text{logit}\left[\Pr(Y \le i | X \ne Y)\right] = \log\left[\frac{\Pr(Y \le i | X \ne Y)}{1 - \Pr(Y \le i | X \ne Y)}\right] \,. \end{split}$$

We shall refer to this model as the conditional marginal cumulative logistic (CML) model. The CML model indicates that given that an observation falls in one of the off-diagonal cells of the table, one conditional marginal distribution is a location shift of the other conditional marginal distribution on a logistic scale. Therefore, this model states that given that an observation falls in one off-diagonal cell, the odds that X is *i* or below instead of i + 1 or above, is $\exp(\Delta^*)$ times higher than the odds that Y is *i* or below instead of i + 1 or above, for i = 1, ..., R - 1. If $\Delta^* > 0$, X rather than Y tends to be *i* or below instead of i + 1 or above. We now obtain the following theorem.

Theorem 2: The MH model holds if and only if both the CML and ME models hold.

We omit the proof because it can be obtained in a similar way as the proof of Theorem 1.

Let n_{ij} denote the observed frequency in the *i*th row and *j*th column of the $R \times R$ table with $n = \sum \sum n_{ij}$, and let m_{ij} denote the corresponding expected frequency. We assume that $\{n_{ij}\}$ have a multinomial distribution. The maximum likelihood estimates

(MLEs) of the expected frequencies under each model can be easily obtained using a Newton-Raphson method to solve the likelihood equation (see the Appendix, e.g., for the CML model). The numbers of degrees of freedom (df) for testing the goodness-of-fit of the MH, ML (CML), and ME models are R - 1, R - 2, and 1, respectively.

3 Examples

We analyze the data in Tables 1 to 3 using the decompositions of the MH model. Table 4 presents the values of the likelihood ratio statistic G^2 for the models applied to these data.

	Table 1		Ta	Table 2		Table 3	
Models	df	G^2	df	G^2	df	G^2	
MH	3	11.99^{*}	2	18.65^{*}	2	8.49*	
ML	2	0.39	1	18.55^{*}	1	1.69	
CML	2	0.03	1	18.59^{*}	1	2.76	
ME	1	11.98^{*}	1	0.07	1	5.69^{*}	

Table 4: Likelihood ratio values G^2 for models applied to the data in Tables 1, 2 and 3.

Note, g(k) for the ME model are the equal-interval scores. * means significant at the 0.05 level.

3.1 Unaided vision data

When the MH model is applied to the data in Table 1, this model fits poorly, yielding $G^2(MH) = 11.99$ with 3 df. However, the ML (CML) model, which is one of the decomposed models for the MH model, fits the data well where $G^2(ML) = 0.39$ and $G^2(CML) = 0.03$ with 2 df. On the other hand, the ME model with g(k) = k, k = 1, 2, 3, 4, which is the other of the decomposed models, fits these data very poorly, yielding $G^2(ME) = 11.98$ with 1 df.

Since the ML (CML) model fits very well, we consider the hypothesis that the MH model holds under the assumption that the ML (CML) model holds, i.e., the hypothesis that $\Delta = 0$ ($\Delta^* = 0$) under the assumption, which is also equivalent to the hypothesis that the ME model holds, i.e., the mean of right eye grade is equal to the mean of left eye grade (i.e., E(X) = E(Y)), under the same assumption, by Theorem 1 (or Theorem 2). The difference between the G^2 values for the MH and ML (CML) models is $G^2(MH|ML) = G^2(MH) - G^2(ML) = 11.60$ ($G^2(MH|CML) = G^2(MH) - G^2(CML) = 11.96$) with 3 - 2 = 1 df, and thus this hypothesis is rejected at the 0.05 significance level. This shows very strong evidence of $\Delta \neq 0$ ($\Delta^* \neq 0$), i.e., $\exp(\Delta) \neq 1$ ($\exp(\Delta^*) \neq 1$) in the ML (CML) model, i.e., very strong evidence of $E(X) \neq E(Y)$ under the ML (CML) model. Therefore the ML (CML) model is preferable to the MH model for these data.

Under the ML model we get $\exp(\Delta) = 1.06$, or $\Delta = 0.05$ with a standard error of 0.02. Note that $\exp(\Delta)$ is estimated to be greater than 1 (from the test described above).

Thus, under the ML model the odds that a woman's right eye grade is i or below instead of i + 1 or above, for i = 1, 2, 3, is estimated to be 1.06 times higher than the odds that her left eye grade is i or below instead of i + 1 or above. Since $\exp(\hat{\Delta}) > 1$, a woman's right eye grade rather than her left eye grade tends to be i or below instead of i + 1 or above. Thus, a woman's right eye tends to be better than her left eye.

Under the CML model the MLE is $\exp(\Delta^*) = 1.21$, or $\Delta^* = 0.19$ with a standard error of 0.06. Thus, from the CML model follows that given that a woman's right eye grade is different from her left eye grade, the odds that the woman's right eye grade is *i* or below instead of i + 1 or above is estimated to be 1.21 times higher than the odds that her left eye grade is *i* or below instead of i + 1 or above. Since $\exp(\hat{\Delta}^*) > 1$, the woman's right eye grade rather than her left eye grade tends to be *i* or below instead of i + 1 or above. Therefore, when a woman's right eye grade is different from her left eye grade, her right eye tends to be better than her left eye.

3.2 Ewe Data

Consider the data in Table 2. When the MH model is applied to these data, the model fits poorly and gives $G^2(MH) = 18.65$ with 2 df. In addition, the ML (CML) model fits also poorly and yield $G^2(ML) = 18.55$ and $G^2(CML) = 18.59$ with 1 df. However, the ME model with g(k) = k, k = 0, 1, 2, fits very well and gives $G^2(ME) = 0.07$ with 1 df. Therefore, we consider the hypothesis that the MH model holds assuming that the ME model also holds. Because of $G^2(MH|ME) = G^2(MH) - G^2(ME) = 18.58$ with 2 - 1 = 1 df, this hypothesis is rejected at the 0.05 significance level, and hence the ME model is preferable to the MH model.

The ME model would provide that the mean for the number of lambs born to a ewe in 1953 is equal to the mean for the number in 1952. However, the distribution for the number of lambs in 1953 is different from the distribution for the number in 1952, because of the poor fit of the MH model. Note that under the ME model, the mean for the number of lambs born to a ewe in 1953 (in 1952) is estimated to be 0.65.

3.3 Danish Opinion Polls Data

From Table 4 we see that the MH model fits the data in Table 3 poorly. However, the ML (CML) model fits well and yields $G^2(ML) = 1.69$ and $G^2(CML) = 2.76$ with 1 df. The ME model with g(k) = k, k = 1, 2, 3, fits poorly (see Table 4).

Consider the hypothesis that the MH model holds under the assumption that the ML (CML) model holds. Thus, we test the hypothesis that $\Delta = 0$ ($\Delta^* = 0$) under the assumption. Because $G^2(MH|ML) = G^2(MH) - G^2(ML) = 6.80$ and $G^2(MH|CML) = G^2(MH) - G^2(CML) = 5.73$ with 1 df, we reject the hypothesis at the 0.05 level. This shows strong evidence of $\Delta \neq 0$ ($\Delta^* \neq 0$) in the ML (CML) model, i.e., evidence of $E(X) \neq E(Y)$. Therefore, the ML (CML) model is preferable to the MH model here.

Under the ML model we get $\exp(\hat{\Delta}) = 1.23$, i.e., $\hat{\Delta} = 0.21$ with a standard error of 0.08. Thus, the ML model provides that (1) the odds that the opinion of an individual is 'yes' instead of 'not yes (i.e., undecided or no)' is estimated to be 1.23 times higher in poll I than in poll II, and (2) the odds that the opinion of the individual is 'not no'

instead of 'no' is 1.23 times higher in poll I than in poll II. The interpretation (2) may also be described as the odds that the opinion of the individual is 'no' instead of 'not no' is estimated to be 1.23 times higher in poll II than in poll I. Therefore, (i) the tendency that the opinion of an individual is 'yes' is stronger in poll I than in poll II, and (ii) the tendency that it is 'no' is stronger in poll II than in poll I.

Next, under the CML model we get $\exp(\hat{\Delta}^*) = 1.76$, or $\hat{\Delta}^* = 0.56$ with a standard error of 0.24. Thus, the CML model provides that (1) under the condition that the opinion in poll II of an individual changed from in poll I, the odds that the opinion is 'yes' instead of 'not yes' is estimated to be 1.76 times higher in poll I than in poll II, and (2) on the same condition, the odds that the opinion of the individual is 'no' instead of 'not no' is estimated to be 1.76 times higher in poll I. Therefore, on condition that the opinion in poll II of an individual changed from in poll I. Therefore, on condition that the opinion of the individual is 'yes' is stronger in poll I than in poll I, (i) the tendency that the opinion of the individual is 'yes' is stronger in poll I than in poll II, and (ii) the tendency that it is 'no' is stronger in poll II than in poll I.

4 Concluding Remarks

The decompositions of the MH model is useful in order to see the reason for its poor fit.

Indeed, for the data in Table 1, the poor fit of the MH model is caused by the poor fit of the ME model rather than the ML (or CML) model, i.e., by the reason that the mean of a woman's right eye grade is different from the mean of the woman's left eye grade. Note that under the ML model, which fits these data very well, the MLE of the mean of the woman's right eye grade is 2.28 and that of the woman's left eye grade is 2.30. Under the CML model, which also fits very well, the MLE of the mean of the woman's right eye grade is 2.27 and that of the woman's left eye grade is 2.31. Conversely, for the data in Table 2, the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model is caused by the poor fit of the MH model.

The MH model as the CML and ME model they all do not depend on the probabilities $\{p_{ii}\}$ on the main diagonal of the table, but the ML model depends on them. Notice that the estimated expected frequencies on the main diagonal cells under the ML model are different from the observed frequencies on the main diagonal (see Tables 1 and 3). Thus, if the MH model does not hold and if we want to see the reason, especially, the reason why the equalities of the conditional row cumulative probabilities $\{F_i^{X(c)}\}$ and the conditional column cumulative probabilities $\{F_i^{Y(c)}\}$ do not hold, the analyst may be interested in inferring the structure of off-diagonal probabilities $\{p_{ij}\}, i \neq j$ and not the main diagonal probabilities $\{p_{ii}\}$. In this case, the decomposition of the MH model into the CML and ME models may be preferable to that into the ML and ME models when the analyst wants to see the reason why the equalities of $\{F_i^{X(c)}\}$ and $\{F_i^{Y(c)}\}$, or the equalities of $\{p_{i\cdot}^c\}$ and $\{p_{i\cdot}^c\}$, do not hold.

However, the MH model indicates the equalities of the row cumulative probabilities $\{F_i^X\}$ and the column cumulative probabilities $\{F_i^Y\}$, which include the probabilities $\{p_{ii}\}$ on the main diagonal. Therefore, if the MH model does not hold and if we want to see the reason why the equalities of $\{F_i^X\}$ and $\{F_i^Y\}$ do not hold, the analyst may also be

interested in inferring the structure of $\{F_i^X\}$ and $\{F_i^Y\}$. Then, the decomposition of the MH model into the ML and ME models may be preferable to that into the CML and ME models when the analyst wants to see the reason why the equalities of $\{F_i^X\}$ and $\{F_i^Y\}$, or the equalities of $\{p_{i\cdot}\}$ and $\{p_{\cdot i}\}$, do not hold, because each of the ML and ME models can be expressed as the function of $\{F_i^X\}$ and $\{F_i^Y\}$.

For the data in Table 3, the readers may be interested in seeing, given the opinion in poll II of an individual changed from in poll I, how the individual changed its attitude. Then the CML model rather than the ML model is useful.

The decompositions of the MH model described here should be considered for ordinal categorical data, because each of the decomposed models is not invariant under the same arbitrary permutations of the row and column categories (although the MH model is invariant under them).

The reader may also be interested in whether it is possible to decompose the test statistic for the MH model. For the decompositions of the MH model described here, it is not guaranteed that the test statistic for the MH model is asymptotically equivalent to the sum of the test statistics for the decomposed models, although the number of df for the MH model is equal to the sum of those for the decomposed models. However, the decomposition of the MH model would be useful for seeing the reason for the poor fit when the MH model fits the data poorly.

Acknowledgements

The authors would like to thank the editor and a referee for their many valuable comments.

References

Agresti, A. (1984). Analysis of Ordinal Categorical Data. New York: Wiley.

- Agresti, A. (2002). Categorical Data Analysis (2nd ed.). New York: Wiley.
- Andersen, E. B. (1980). *Discrete Statistical Models with Social Science Applications*. Amsterdam: North-Holland.
- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Massachusetts: The MIT Press.
- Goodman, L. A. (1979). Multiplicative models for square contingency tables with ordered categories. *Biometrika*, 66, 413-418.
- McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, 65, 413-418.
- Miyamoto, N., Ohtsuka, W., and Tomizawa, S. (2004). Linear diagonals-parameter symmety and quasi-symmety models for cumulative probabilities in square contingency tables with ordered categories. *Biometrical Journal*, *46*, 664-674.
- Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). Akaike Information Criterion Statistics. Dordrecht: D. Reidel Publishing Company.
- Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 42, 412-416.
- Tallis, G. M. (1962). The maximum likelihood estimation of correlation from contingency tables. *Biometrics*, *18*, 342-353.

- Tomizawa, S. (1991). Decomposing the marginal homogeneity model into two models for square contingency tables with ordered categories. *Calcutta Statistical Association Bulletin*, 41, 201-207.
- Tomizawa, S. (1993a). Diagonals-parameter symmetry model for cumulative probabilities in square contingency tables with ordered categories. *Biometrics*, 49, 883-887.
- Tomizawa, S. (1993b). A decomposition of the marginal homogeneity model for square contingency tables with ordered categories. *Calcutta Statistical Association Bulletin*, 43, 123-125.
- Tomizawa, S. (1998). A decomposition of the marginal homogeneity model into three models for square contingency tables with ordered categories. *Sankhyā*, *Series B*, 60, 293-300.

Appendix

We consider the MLEs of the expected frequencies $\{m_{ij}\}$ under the CML model. Those under the MH, ML, and ME models can be obtained in the similar manner, although those are omitted here.

To obtain these MLEs under the CML model, we must maximize the Lagrangian

$$L = \sum_{i=1}^{R} \sum_{j=1}^{R} n_{ij} \log p_{ij} - \mu \left(\sum_{i=1}^{R} \sum_{j=1}^{R} p_{ij} - 1 \right) - \sum_{i=1}^{R-1} \lambda_i \left(F_i^{X(c)} \left(1 - F_i^{Y(c)} \right) - \exp(\Delta^*) \left(1 - F_i^{X(c)} \right) F_i^{Y(c)} \right)$$

with respect to $\{p_{ij}\}, \mu, \{\lambda_i\}$, and Δ^* . Setting the partial derivatives of L equal to zero, we obtain the equations

$$m_{ij} = n_{ij} - p_{ij}\delta^{-1} \sum_{u=1}^{R-1} I_{(i \neq j)}$$

$$\times \left[I_{(u \geq i)} I_{(u \geq j)} \lambda_u \left\{ \left(1 - F_u^{Y(c)} \right) - \exp(\Delta^*) \left(1 - F_u^{X(c)} \right) \right\} \right.$$

$$+ I_{(u < i)} I_{(u \geq j)} \lambda_u \left\{ -\exp(\Delta^*) \left(F_u^{Y(c)} + \left(1 - F_u^{X(c)} \right) \right) \right\}$$

$$+ I_{(u \geq i)} I_{(u < j)} \lambda_u \left\{ \left(1 - F_u^{Y(c)} \right) + F_u^{X(c)} \right\}$$

$$+ I_{(u < i)} I_{(u < j)} \lambda_u \left\{ F_u^{X(c)} - \exp(\Delta^*) F_u^{Y(c)} \right\} \right], \quad \text{for } i = 1, \dots, R; \ j = 1, \dots, R,$$

as well as

$$\sum_{i=1}^{R-1} \lambda_i \left(1 - F_i^{X(c)} \right) F_i^{Y(c)} = 0 \,,$$

and

$$F_i^{X(c)}\left(1 - F_i^{Y(c)}\right) = \exp(\Delta^*)\left(1 - F_i^{X(c)}\right)F_i^{Y(c)}, \quad \text{for } i = 1, \dots, R-1,$$

where $m_{ij} = np_{ij}$ and $I_{(\cdot)}$ is the indicator function. Using the Newton-Raphson method, we can solve the equations with respect to $\{p_{ij}\}, \{\lambda_i\}$ and Δ^* . Therefore, we can obtain the MLEs of $\{m_{ij}\}$ and Δ^* under the CML model.

Author's addresses:

Dr. Nobuko Miyamoto Department of Information Sciences Faculty of Science and Technology Tokyo University of Science Noda City, Chiba, 278-8510 Japan E-mail: miyamoto@is.noda.tus.ac.jp Kouji Niibe Graduate School of Science and Technology Tokyo University of Science Noda City, Chiba, 278-8510 Japan Prof. Dr. Sadao Tomizawa Department of Information Sciences Faculty of Science and Technology

Tokyo University of Science Noda City, Chiba, 278-8510 Japan E-mail: tomizawa@is.noda.tus.ac.jp