Austrian Journal of Statistics
January 2022, Volume 51, 70–82.
http://www.ajs.or.at/
doi:10.17713/ajs.v51i1.1210

Asymmetry Model Using Marginal Ridits for Ordinal Square Contingency Tables

Shuji Ando
Tokyo University of Science

Abstract

This study proposes a new marginal asymmetry model which can infer the relation between marginal ridits of row and column variables for ordinal square contingency tables. When the marginal homogeneity model does not hold, we will apply marginal asymmetry models (e.g., the marginal cumulative logistic and extended marginal homogeneity models). On the other hand, we may measure the degree of departure from the marginal homogeneity model. To measure the degree of that, multiple indexes were proposed. Some of them correspond to the marginal cumulative logistic and extended marginal homogeneity models. The proposed model corresponds to the index, which represents the degree of departure from the MH model, using marginal ridits. We compare the proposed model with the existing marginal asymmetry models and show that the proposed model provides better fit performance than them for real data.

Keywords: marginal homogeneity, marginal ridits, ordered categorical data, symmetry.

1. Introduction

This study focuses on comparing marginal distributions for matched pair of ordered categorical data with same classifications. Denote two ordinal outcomes (X_1, X_2) with C categories. These can be summarized as an $C \times C$ square contingency table. We regard the cell observations n_{ij} $(= np_{ij})$ with joint sample proportions p_{ij} as a multinomial sample with sample size n and parameters π_{ij} for $i, j = 1, \ldots, C$.

The marginal homogeneity (MH) model satisfies the following condition

$$P(X_1 = i) = P(X_2 = i) \quad \text{for } i = 1, \ldots, C,$$

see Stuart (1955). Using the joint probabilities π_{ij}, this model is expressed as

$$\pi_{i+} = \pi_{+i} \quad \text{for } i = 1, \ldots, C,$$

where $\pi_{i+} = \sum_{l=1}^{C} \pi_{il}$ and $\pi_{+i} = \sum_{k=1}^{C} \pi_{ki}$. Let cumulative marginal distributions of X_1 and X_2 denote as

$$F_i^{X_1} = \sum_{k=1}^{i} \pi_{k+} \quad \text{and} \quad F_i^{X_2} = \sum_{l=1}^{i} \pi_{+l} \quad \text{for } i = 1, \ldots, C-1,$$

respectively. The MH model is also expressed as

$$F_i^{X_1} = F_i^{X_2} \quad \text{for } i = 1, \dots, C-1.$$

By considering the difference $F_i^{X_1} - F_i^{X_2}$, the MH model is further expressed as

$$G_{1(i)} = G_{2(i)} \quad \text{for } i = 1, \dots, C-1,$$

where $G_{1(i)} = \sum_{k=1}^{i} \sum_{l=i+1}^{C} \pi_{kl}$ and $G_{2(i)} = \sum_{k=i+1}^{C} \sum_{l=1}^{i} \pi_{kl}$, see, for example, Tahata and Tomizawa (2008).

When the MH model does not hold for data, we will apply the marginal asymmetry models. On the other hand, we may measure the degree of departure from the MH model. We introduce first the marginal asymmetry models, second the relation between the marginal asymmetry models and indexes to measure the degree of departure from the MH model.

As the marginal asymmetry models, the m-additional marginal homogeneity (MH(m)) and m-additional marginal cumulative logistic (MCL(m)) models were proposed by Tahata and Tomizawa (2008) and Kurakami, Tahata, and Tomizawa (2013), respectively. For a given m ($m = 1, \dots, C-1$), the MH(m) model satisfies the following condition

$$G_{1(i)} = \prod_{k=0}^{m-1} \psi_k^{i^k} G_{2(i)} \quad \text{for } i = 1, \dots, C-1. \tag{1}$$

We note that (i) the MH(m) model with $\psi_0 = \cdots = \psi_{m-1} = 1$ is equivalent to the MH model, (ii) the MH(1) model is equivalent to the extended marginal homogeneity model proposed by Tomizawa (1993), and (iii) the MH(2) model is equivalent to the generalized marginal homogeneity model proposed by Tomizawa (1995), see Tahata and Tomizawa (2008). For a given m ($m = 1, \dots, C-1$), the MCL(m) model satisfies the following condition

$$L_i^{X_1} = L_i^{X_2} + \sum_{k=0}^{m-1} i^k \log \psi_k \quad \text{for } i = 1, \dots, C-1,$$

where

$$L_i^{X_1} = \log\left(\frac{F_i^{X_1}}{1 - F_i^{X_1}}\right) \quad \text{and} \quad L_i^{X_2} = \log\left(\frac{F_i^{X_2}}{1 - F_i^{X_2}}\right).$$

The MCL(m) model is also expressed as

$$H_{1(i)} = \prod_{k=0}^{m-1} \psi_k^{i^k} H_{2(i)} \quad \text{for } i = 1, \dots, C-1, \tag{2}$$

where $H_{1(i)} = F_i^{X_1}(1 - F_i^{X_2})$ and $H_{2(i)} = (1 - F_i^{X_1})F_i^{X_2}$, see Kurakami *et al.* (2013). From Equations (1) and (2), we see that the MH(m) and MCL(m) models are expressed as similar multiplicative forms. We note that (i) the MCL(m) with $\psi_0 = \cdots = \psi_{m-1} = 1$ is equivalent to the MH model, (ii) the MCL(1) is equivalent to the marginal cumulative logistic model proposed by McCullagh (1977), and (iii) the MCL(2) is equivalent to the extended marginal cumulative logistic model proposed by Kurakami, Tahata, and Tomizawa (2010).

The MH model is equivalent to the MH(m) and MCL(m) models with $\psi_0 = \cdots = \psi_{m-1} = 1$, however the MH($m$) and MCL($m$) models generally has a different structure from the MH model. Tahata and Tomizawa (2008) considered what a structure is necessary to obtain the MH model in addition to the MH(m) model. Kurakami *et al.* (2013) also considered what a structure is necessary to obtain the MH model in addition to the MCL(m) model.

To measure the degree of departure from the MH model, multiple indexes were proposed, for example, Tomizawa, Miyamoto, and Ashihara (2003), Tahata, Tajima, and Tomizawa (2006), Iki, Tahata, and Tomizawa (2012), and the references therein. The index of Tomizawa *et al.*

(2003) corresponds to the MH(m) model because it is expressed as the function of $\prod_{k=0}^{m-1} \psi_k^{ik}$ under the MH(m) model. The index of Iki *et al.* (2012) corresponds to the MCL(m) model because it is expressed as the function of $\prod_{k=0}^{m-1} \psi_k^{ik}$ under the MCL(m) model. These relations are natural since the parameters $\prod_{k=0}^{m-1} \psi_k^{ik}$ express the degree of departure from the MH model. The index of Tahata *et al.* (2006) is a function of marginal ridits, it does not correspond to the MH(m) and MCL(m) models.

This study proposes a new marginal asymmetry model using marginal ridits that corresponds to the index of Tahata *et al.* (2006). Thus, under the proposed model, the index of Tahata *et al.* (2006) is expressed as the function of asymmetry parameters. We are interested in what a structure is necessary to obtain the MH model in addition to the proposed model, in the similar to Tahata and Tomizawa (2008) and Kurakami *et al.* (2013).

This paper is organized as follows. Section 2 defines the proposed model. Section 3 describes the relation between the proposed model and the index of Tahata *et al.* (2006). Section 4 introduces the relation between the MH and proposed models. Section 5 demonstrates the utility of the proposed model using application to real data. Section 6 extends the proposed model to multi-way contingency tables. Section 7 closes with concluding remarks.

2. Proposed model

The marginal ridits of X_1 and X_2 are denoted by

$$r_i^{X_1} = \sum_{k=1}^{i-1} \pi_{k+} + \frac{1}{2}\pi_{i+} \quad \text{and} \quad r_i^{X_2} = \sum_{l=1}^{i-1} \pi_{l} + \frac{1}{2}\pi_{+i} \quad \text{for } i = 1, \ldots, C,$$

respectively, see Bross (1958) and Agresti (2010, Sec. 8.1.3.). Using the marginal ridits, the MH model is expressed as

$$R_i^{X_1} = R_i^{X_2} \quad \text{for } i = 1, \ldots, C-1, \tag{3}$$

where

$$R_i^{X_1} = \frac{r_i^{X_1}}{r_C^{X_1}} \quad \text{and} \quad R_i^{X_2} = \frac{r_i^{X_2}}{r_C^{X_2}},$$

see Tahata *et al.* (2006).

Assume that the ordered categories of the both row and column variables represent intervals of an underlying continuous distribution. When the underlying distribution is uniform over each interval, the $r_i^{X_1}$ (or $r_i^{X_2}$) would be equal to the probability that the row (or column) value of a randomly selected individual falls below the midpoint of category i for $i = 1, \ldots, C$, see Agresti (1984, p. 168). We are interested in the form of the Equation (3) in addition to the Equations (1) and (2) for ordinal square contingency tables such that are assumed an underlying bivariate continuous distribution.

In a similar way Tahata and Tomizawa (2008) and Kurakami *et al.* (2013), for a given m ($m = 1, \ldots, C-1$), we propose a m-additional marginal ridits (MR(m)) model that satisfies the following condition

$$R_i^{X_1} = \prod_{k=0}^{m-1} \psi_k^{ik} R_i^{X_2} \quad \text{for } i = 1, \ldots, C-1.$$

The MR(m) model indicates that the log-odds, $\log(R_i^{X_1}/R_i^{X_2})$ for $i = 1, \ldots, C-1$, is expressed as the polynomial function of categories i as $\sum_{k=0}^{m-1} i^k \log \psi_k$. The number of degree of freedom for testing the goodness of fit of the MR(m) is $C-1-m$. We note that (i) the MR(m) model with $\psi_0 = \cdots = \psi_{m-1} = 1$ is equivalent to the MH model, (ii) the MR($C-1$) model is saturated, and (iii) there is an inclusion relation as follows:

$$\text{MH} \subseteq \text{MR}(1) \subseteq \text{MR}(2) \subseteq \cdots \subseteq \text{MR}(C-1).$$

Thus, the MR(m) model provides a m-additional parameters class of alternatives to the MH model. Let denote the likelihood ratio test statistic for testing the goodness-of-fit the MR(m) model as $G^2(\text{MR}(m))$. When we test that the values of m-additional parameters equal one (i.e., the MH model holds) assuming that the MR(m) model holds, we use the likelihood ratio statistic $G^2(\text{MH}|\text{MR}(m)) = G^2(\text{MH}) - G^2(\text{MR}(m))$. Under the MH model, the $G^2(\text{MH}|\text{MR}(m))$ has an asymptotic chi-squared distribution with m degrees of freedom. We could obtain an appropriate m (i.e., the appropriate MR(m) model) for the data using these conditional tests.

Under the MR(m) model, $\prod_{k=0}^{m-1} \psi_k^{ik} > 1$ is imply to $r_i^{X_1} > r_i^{X_2}$. Thus, the parameters $\prod_{k=0}^{m-1} \psi_k^{ik}$ of the MR(m) model are useful to infer the relation between marginal ridits of X_1 and X_2.

3. Relation between proposed model and existing index

Tahata *et al.* (2006) proposed an index $\Lambda^{(\lambda)}$ to measure the degree of departure from the MH model. Assume that $\pi_{1+} + \pi_{+1} \neq 0$, the index $\Lambda^{(\lambda)}$ is defined as

$$\Lambda^{(\lambda)} = 1 - \frac{\lambda 2^{\lambda}}{2^{\lambda} - 1} \sum_{i=1}^{C-1} \frac{R_i^{X_1} + R_i^{X_2}}{\sum_{j=1}^{C-1}(R_j^{X_1} + R_j^{X_2})} I_i^{(\lambda)} \quad \text{for } \lambda > -1,$$

where

$$I_i^{(\lambda)} = \frac{1}{\lambda}\left[1 - \left(\frac{R_i^{X_1}}{R_i^{X_1} + R_i^{X_2}}\right)^{\lambda+1} - \left(\frac{R_i^{X_2}}{R_i^{X_1} + R_i^{X_2}}\right)^{\lambda+1}\right].$$

We note that the index $\Lambda^{(\lambda)}$ is a function of the marginal ridits of X_1 and X_2. For each λ, the index $\Lambda^{(\lambda)}$ satisfies the following characteristics: (i) the index $\Lambda^{(\lambda)}$ lies between zero to one, (ii) $\Lambda^{(\lambda)} = 0$ if and only if the MH model holds, (iii) the degree of departure from the MH model is maximum, in the sense that $R_i^{X_1} = 0$ (then $R_i^{X_2} \neq 0$) and $R_i^{X_2} = 0$ (then $R_i^{X_1} \neq 0$) for all $i = 1, \ldots, C-1$, see Tahata *et al.* (2006).

Under the MR(m) model, the index $\Lambda^{(\lambda)}$ is expressed as

$$\Lambda^{(\lambda)} = 1 - \frac{2^{\lambda}}{2^{\lambda} - 1}\left[1 - \left(\frac{\prod_{k=0}^{m-1} \psi_k^{ik}}{\prod_{k=0}^{m-1} \psi_k^{ik} + 1}\right)^{\lambda+1} - \left(\frac{1}{\prod_{k=0}^{m-1} \psi_k^{ik} + 1}\right)^{\lambda+1}\right] \quad \text{for } \lambda > -1. \quad (4)$$

From Equation (4), the index $\Lambda^{(\lambda)}$ corresponds to the MR(m) model because it is expressed as the function of $\prod_{k=0}^{m-1} \psi_k^{ik}$ under the MR(m) model. Figure 1 shows the values of the index $\Lambda^{(\lambda)}$ corresponding to $\prod_{k=0}^{m-1} \psi_k^{ik}$ in the range $0.1 \leq \prod_{k=0}^{m-1} \psi_k^{ik} \leq 10$. From Figure 1, we see that the index $\Lambda^{(\lambda)}$ increases monotonically as $\prod_{k=0}^{m-1} \psi_k^{ik}$ approaches zero or infinity.

4. Relation between models

The MH model is equivalent to the MR(m) model with $\psi_0 = \cdots = \psi_{m-1} = 1$, however the MR($m$) model generally has a different structure from the MH model. We are interested in what a structure is necessary to obtain the MH model in addition to the MR(m) model.

For a given k ($k = 0, 1, \ldots, C-2$), we consider a k-th marginal ridits equality (MRE(k)) model satisfied the following condition

$$\sum_{i=1}^{C-1} i^k R_i^{X_1} = \sum_{i=1}^{C-1} i^k R_i^{X_2}.$$

The number of degree of freedom for testing the goodness of fit of the MRE(k) is one.

We obtain the following theorem:

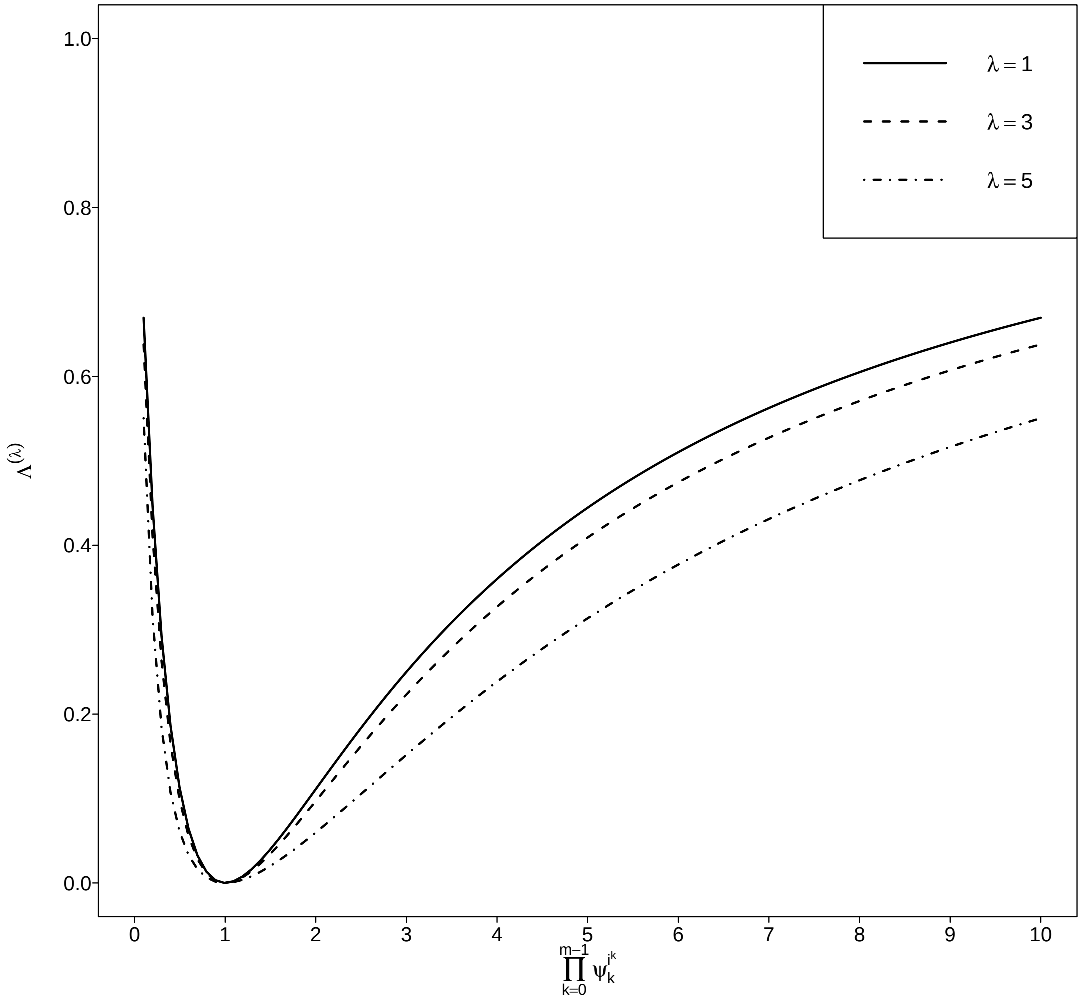

Figure 1: This figure shows the values of the index $\Lambda^{(\lambda)}$ corresponding to $\prod_{k=0}^{m-1} \psi_k^{i^k}$ under the MR(m) model.

Theorem 1. *For a given m ($m = 1, \dots, C-1$), the MH model holds if and only if the MR(m) and MRE(k), for all $k = 0, 1, \dots, m-1$, models hold.*

Proof. For a given m ($m = 1, \dots, C-1$), if the MH model holds, then the MR(m) and MRE(k), for all $k = 1, \dots, m$, models hold. Assuming that the MR(m) and MRE(k), for all $k = 1, \dots, m$, models hold, we show that the MH model hold.
Let

$$\alpha_{1(i)} = \frac{R_i^{X_1}}{\sum_{l=1}^{C-1}(R_l^{X_1} + R_l^{X_2})} \quad \text{and} \quad \alpha_{2(i)} = \frac{R_i^{X_2}}{\sum_{l=1}^{C-1}(R_l^{X_1} + R_l^{X_2})} \quad \text{for } i = 1, \dots, C-1.$$

Using $\alpha_{1(i)}$ and $\alpha_{2(i)}$, the MR(m) model is also expressed as

$$\alpha_{1(i)} = \prod_{k=0}^{m-1} \psi_k^{i^k} \beta_{1(i)} \quad \text{and} \quad \alpha_{2(i)} = \beta_{2(i)} \quad \text{for } i = 1, \dots, C-1,$$

where $\beta_{1(i)}$ and $\beta_{2(i)}$ are parameters that satisfy the restriction $\beta_{1(i)} = \beta_{2(i)}$.
Let π_{ij}^* denote the (i, j) cell probabilities that satisfy the MR(m) and MRE(k), for all $k = 1, \dots, m$, models. And let $\alpha_{1(i)}^*$ and $\alpha_{2(i)}^*$ be $\alpha_{1(i)}$ and $\alpha_{2(i)}$ with π_{ij} replaced by π_{ij}^*. Since the MR(m) model holds,

$$\log \alpha_{1(i)}^* = \sum_{k=0}^{m-1} i^k \log \psi_k + \log \beta_{1(i)} \quad \text{and} \quad \log \alpha_{2(i)}^* = \log \beta_{2(i)} \quad \text{for } i = 1, \dots, C-1.$$

Let $\gamma_{1(i)} = \epsilon^{-1}\beta_{1(i)}$ and $\gamma_{2(i)} = \epsilon^{-1}\beta_{2(i)}$ ($i = 1, \dots, C-1$), where $\epsilon = \sum_{i=1}^{C-1}(\beta_{1(i)} + \beta_{2(i)})$. The MR($m$) model is further expressed as

$$\log\left(\frac{\alpha_{1(i)}^*}{\gamma_{1(i)}}\right) = \sum_{k=0}^{m-1} i^k \log \psi_k + \log \epsilon \quad \text{and} \quad \log\left(\frac{\alpha_{2(i)}^*}{\gamma_{2(i)}}\right) = \log \epsilon \quad \text{for } i = 1, \dots, C-1. \tag{5}$$

Since the MRE(k) models,

$$\sum_{i=1}^{C-1} i^k \alpha_{1(i)}^* = \sum_{i=1}^{C-1} i^k \alpha_{2(i)}^* = \mu^{(k)} \quad \text{for } k = 0, 1, \dots, m-1. \tag{6}$$

Consider arbitrary cell probabilities π_{ij} that satisfy the following condition

$$\sum_{i=1}^{C-1} i^k \alpha_{1(i)} = \sum_{i=1}^{C-1} i^k \alpha_{2(i)} = \mu^{(k)} \quad \text{for } k = 0, 1, \dots, m-1. \tag{7}$$

From Equations (5) to (7),

$$\sum_{i=1}^{C-1} \left(\alpha_{1(i)} - \alpha_{1(i)}^*\right) \log\left(\frac{\alpha_{1(i)}^*}{\gamma_{1(i)}}\right) - \sum_{i=1}^{C-1} \left(\alpha_{2(i)} - \alpha_{2(i)}^*\right) \log\left(\frac{\alpha_{2(i)}^*}{\gamma_{2(i)}}\right) = 0. \tag{8}$$

Consider the Kullback-Leibler information such as follows:

$$K(\alpha, \gamma) = \sum_{i=1}^{C-1} \alpha_{1(i)} \log\left(\frac{\alpha_{1(i)}}{\gamma_{1(i)}}\right) + \sum_{i=1}^{C-1} \alpha_{2(i)} \log\left(\frac{\alpha_{2(i)}}{\gamma_{2(i)}}\right)$$

and

$$K(\alpha^*, \gamma) = \sum_{i=1}^{C-1} \alpha_{1(i)}^* \log\left(\frac{\alpha_{1(i)}^*}{\gamma_{1(i)}}\right) + \sum_{i=1}^{C-1} \alpha_{2(i)}^* \log\left(\frac{\alpha_{2(i)}^*}{\gamma_{2(i)}}\right).$$

From Equation (8),

$$K(\alpha,\gamma) = K(\alpha^*,\gamma) + K(\alpha,\alpha^*),$$

where

$$K(\alpha,\alpha^*) = \sum_{i=1}^{C-1} \alpha_{1(i)} \log\left(\frac{\alpha_{1(i)}}{\alpha^*_{1(i)}}\right) + \sum_{i=1}^{C-1} \alpha_{2(i)} \log\left(\frac{\alpha_{2(i)}}{\alpha^*_{2(i)}}\right).$$

Since γ is fixed,

$$\min_{\alpha} K(\alpha,\gamma) = K(\alpha^*,\gamma),$$

and then $\alpha^*_{1(i)}$ and $\alpha^*_{2(i)}$ $(i = 1,\ldots,C-1)$ uniquely minimize $K(\alpha,\gamma)$, see Darroch and Ratcliff (1972). Thus, we obtain

$$\alpha_{1(i)} = \alpha^*_{1(i)} \quad \text{and} \quad \alpha_{2(i)} = \alpha^*_{2(i)} \quad \text{for } i = 1,\ldots,C-1. \tag{9}$$

Let $\pi^{**}_{ij} = \pi^*_{ji}$ for $i,j = 1,\ldots,C$, and let $\alpha^{**}_{1(i)} = \alpha^*_{2(i)}$ and $\alpha^{**}_{2(i)} = \alpha^*_{1(i)}$ for $i = 1,\ldots,C-1$. In the same manner,

$$\min_{\alpha} K(\alpha,\gamma) = K(\alpha^{**},\gamma),$$

where

$$K(\alpha^{**},\gamma) = \sum_{i=1}^{C-1} \alpha^{**}_{1(i)} \log\left(\frac{\alpha^{**}_{1(i)}}{\gamma_{1(i)}}\right) + \sum_{i=1}^{C-1} \alpha^{**}_{2(i)} \log\left(\frac{\alpha^{**}_{2(i)}}{\gamma_{2(i)}}\right),$$

and then $\alpha^{**}_{1(i)}$ and $\alpha^{**}_{2(i)}$ $(i = 1,\ldots,C-1)$ uniquely minimize $K(\alpha,\gamma)$. Therefore, we obtain

$$\alpha_{1(i)} = \alpha^{**}_{1(i)} = \alpha^*_{2(i)} \quad \text{and} \quad \alpha_{2(i)} = \alpha^{**}_{2(i)} = \alpha^*_{1(i)} \quad \text{for } i = 1,\ldots,C-1. \tag{10}$$

From Equations (9) and (10),

$$\alpha^*_{1(i)} = \alpha^*_{2(i)} \quad \text{for } i = 1,\ldots,C-1.$$

Namely, the MH model holds. The proof is completed. □

5. Application to real data

5.1. Application to occupational status data

Table 1, taken from Agresti (2010, p. 230), relates father's and son's occupational status category for a British sample. We are interested in comparing marginal distributions for these data.

Except for the (n_{13}, n_{31}) combination, the cell observation satisfies $n_{ij} > n_{ji}$ for $i < j$. Moreover, the marginal observations for categories from (1) to (4) are father's greater than son's. Thus, the MH(m), MCL(m) and MR(m) models may be preferable to the MH model.

A quick methods for choosing the best-fitting model among applied models are to use the Akaike information criterion (AIC) and Bayesian information criterion (BIC), which are defined as

$$\text{AIC} = -2 \times (\text{the maximum log likelihood}) + 2 \times (\text{the number of parameters}),$$

$$\text{BIC} = -2 \times (\text{the maximum log likelihood}) + \log n \times (\text{the number of parameters}),$$

see Akaike (1974) and Schwarz (1978). The AIC and BIC give the best-fitting model as the one with the minimum AIC and minimum BIC, respectively. Since only the difference between AICs is required when two models are compared, it is possible to ignore the common constant of AIC, and we may use a modified AIC defined as

$$\text{AIC}^+ = G^2 - 2 \times (\text{the number of degree of freedom}).$$

Table 1: Occupational status for British father–son pairs; taken from Agresti (2010, p. 230). The parenthesized values are the maximum likelihood estimates of expected frequencies under the MR(1) model.

	Son's status					
Father's status	(1)	(2)	(3)	(4)	(5)	Total
(1)	50	45	8	18	8	129
	(50.34)	(39.99)	(7.49)	(16.30)	(7.30)	(121.40)
(2)	28	174	84	154	55	495
	(32.19)	(173.68)	(88.89)	(156.97)	(56.52)	(508.24)
(3)	11	78	110	223	96	518
	(11.92)	(73.93)	(110.20)	(215.62)	(93.54)	(505.22)
(4)	14	150	185	714	447	1510
	(15.74)	(146.87)	(191.81)	(713.63)	(450.36)	(1518.42)
(5)	3	42	72	320	411	848
	(3.35)	(40.83)	(74.08)	(317.47)	(411.00)	(846.72)
Total	106	489	459	1429	1017	3500
	(113.53)	(475.30)	(472.47)	(1419.99)	(1018.72)	

Note: Occupational status is (1) lowest status; (5) highest status.

Similarly, we may use a modified BIC defined as

$$\mathrm{BIC}^{+} = G^2 - \log n \times (\text{the number of degree of freedom}).$$

Thus, for the data, the model with the minimum AIC^{+} (i.e., the minimum AIC) or minimum BIC^{+} (i.e., the minimum BIC) may be the best-fitting model.

Table 2 shows the values of likelihood ratio statistic (G^2), AIC^{+} and BIC^{+} for each model.

Table 2: The values of likelihood ratio statistic (G^2), AIC^{+} and BIC^{+} for each model applied to Table 1

Models	Degree of freedom	G^2	AIC^{+}	BIC^{+}
MH	4	32.80*	24.80	0.16
MH(1)	3	7.30	1.30	−17.18
MH(2)	2	4.25	0.25	−12.07
MH(3)	1	1.04	−0.96	−7.12
MCL(1)	3	9.75*	3.75	−14.75
MCL(2)	2	5.22	1.22	−11.10
MCL(3)	1	0.71	−1.29	−7.45
MR(1)	3	3.11	−2.89	−21.37
MR(2)	2	3.08	−0.92	−13.24
MR(3)	1	3.06	1.06	−5.10

The $*$ means significant at the 0.05 level.

As supposed, the MH model fits these data poorly, and the MH(m), MCL(m) and MR(m) models fits these data well except for the MCL(1) model. According to the test based on the differences between G^2 values for (i) the MR(1) and MR(2) models, (ii) the MR(2) and MR(3) models, the MR(1) model is preferable to the MR(2) and MR(3) models. Moreover, since the MR(1) model has minimum AIC^{+} and BIC^{+} values, it is the best fitted model among applied models.

Under the MR(1) model, the maximum likelihood estimate of ψ_0 is 1.04. Thus, we can infer that the son's occupational status is higher than the father's one by comparing between marginal ridits.

The MRE(0) model fits these data poorly since $G^2 = 6.63$. From Theorem 1, we can consider that the poor fit of the MH model is caused by the influence of the lack of structure of the MRE(0) model rather than the MR(1) model.

5.2. Application to vision data

We consider the data in Table 3, taken from Tomizawa (1985). Table 3 is the data of the of 4746 university students aged 18 to about 25, including about 10% of the women of the Faculty of Science and Technology, Tokyo University of Science examined in 1982. We are interested in comparing marginal distributions for these data.

Table 3: Unaided distance vision of 4746 university students aged 18 to about 25, including about 10% of the women of the Faculty of Science and Technology, Tokyo University of Science examined in 1982; source Tomizawa (1985). The parenthesized values from above are the maximum likelihood estimates of expected frequencies under the MR(1) and MH(1) models, respectively.

	Left eye grade				
Right eye grade	(1)	(2)	(3)	(4)	Total
(1)	1291	130	40	22	1483
	(1291.17)	(134.38)	(40.13)	(22.14)	(1487.82)
	(1291.00)	(130.13)	(38.74)	(21.88)	(1481.75)
(2)	149	221	114	23	507
	(144.24)	(220.88)	(110.69)	(22.40)	(498.21)
	(148.88)	(221.00)	(110.31)	(22.85)	(503.04)
(3)	64	124	660	185	1033
	(63.81)	(127.76)	(660.05)	(185.63)	(1037.24)
	(65.73)	(127.47)	(660.00)	(190.09)	(1043.29)
(4)	20	25	249	1429	1723
	(19.87)	(25.67)	(248.18)	(1429.00)	(1722.72)
	(20.09)	(25.14)	(243.69)	(1429.00)	(1717.92)
Total	1524	500	1063	1659	4746
	(1519.09)	(508.69)	(1059.04)	(1659.18)	
	(1525.70)	(503.73)	(1052.75)	(1663.82)	

Note: Eye grade is (1) highest grade; (4) lowest grade.

Except for the (n_{14}, n_{41}) combination, the cell observation satisfies $n_{ij} < n_{ji}$ for $i < j$. Moreover, the marginal observations for categories from (1) to (3) are left's greater than right's. Thus, the MH(m), MCL(m) and MR(m) models may be preferable to the MH model. Table 4 shows the values of G^2, AIC$^+$ and BIC$^+$ for each model.

As supposed, the MH model fits these data poorly, and the MH(m), MCL(m) and MR(m) models fits these data well. According to the test based on the differences between G^2 values for the MR(1) and MR(2) models, the MR(1) model is preferable to the MR(2) model. Moreover, since the MR(1) model has minimum AIC$^+$ and BIC$^+$ values, it is the best fitted model among applied models.

The reader may point out that there is almost no the differences between AIC$^+$ (or BIC$^+$) values for the MR(1) and MH(1) models. We observe that the MH(m) model is saturated on the main diagonal cells of the $C \times C$ table (see the maximum likelihood estimates of the expected frequencies for the MH(1) model in Table 3). On the other hand, we observe that the MR(m) model is saturated on only the (C, C)th cell of the $C \times C$ table (see the maximum likelihood estimates of the expected frequencies for the MR(1) model in Table 3). As the characteristic of square contingency tables, many observations tend to concentrate the main diagonal cells. In Table 3, 3601 ($= 1291 + 221 + 660 + 1429$) observations fall on the main

Table 4: The values of likelihood ratio statistic (G^2), AIC$^+$ and BIC$^+$ for each model applied to Table 3

Models	Degree of freedom	G^2	AIC$^+$	BIC$^+$
MH	3	11.18*	5.18	−14.22
MH(1)	2	0.56	−3.44	−16.37
MH(2)	1	0.41	−1.59	−8.06
MCL(1)	2	1.41	−2.59	−15.52
MCL(2)	1	1.12	−0.88	−7.35
MR(1)	2	0.55	−3.45	−16.38
MR(2)	1	0.24	−1.76	−8.23

The $*$ means significant at the 0.05 level.

diagonal cells, corresponding to approximately 76% of the sample size. Thus, utilizing the information on the main diagonal cells is important. From these points, when we want to use more observations, we consider that the MR(1) model may be preferred over the MH(1) model.

Under the MR(1) model, the maximum likelihood estimate of ψ_0 is 0.99. Thus, we can infer that the left eye is better than the right eye by comparing between marginal ridits.

6. Extension to multi-way contingency tables

In the previous section, we have dealt in the case of two ordinal outcomes (X_1, X_2) with C categories. In this section, we will deal in the case of multiple ordinal outcome $(X_1, X_2, \ldots, X_T)$ with C categories.

We consider a multi-way C^T contingency table. The MH model satisfies the following condition

$$P(X_1 = i) = P(X_2 = i) = \cdots P(X_T = i) \quad \text{for } i = 1, \ldots, C,$$

see Stuart (1955). Let cumulative marginal distributions of $X_1, X_2, \ldots, X_T$ denote as

$$F_i^{X_t} = P(X_t \leq i) \quad \text{for } i = 1, \ldots, C-1; t = 1, \ldots, T,$$

respectively.

Kurakami *et al.* (2013) proposed the MCL(m) model for a multi-way C^T contingency table with same ordinal classifications. For a given m ($m = 1, \ldots, C-1$), the MCL(m) model satisfies the following condition

$$L_i^{X_1} = L_i^{X_t} + \sum_{k=0}^{m-1} i^k \log \psi_k^{(t)} \quad \text{for } i = 1, \ldots, C-1; t = 2, \ldots, T,$$

where

$$L_i^{X_t} = \log \left(\frac{F_i^{X_t}}{1 - F_i^{X_t}} \right).$$

The MCL(m) model is also expressed as

$$H_{1(i)}^{(t)} = \prod_{k=0}^{m-1} \left(\psi_k^{(t)} \right)^{i^k} H_{2(i)}^{(t)} \quad \text{for } i = 1, \ldots, C-1; t = 2, \ldots, T,$$

where $H_{1(i)}^{(t)} = F_i^{X_1}(1 - F_i^{X_t})$ and $H_{2(i)}^{(t)} = (1 - F_i^{X_1})F_i^{X_t}$, see Kurakami *et al.* (2013).

In a similar way Kurakami *et al.* (2013), for a given m ($m = 1, \ldots, C-1$), we propose the MR(m) model that satisfies the following condition

$$R_i^{X_1} = \prod_{k=0}^{m-1} \left(\psi_k^{(t)}\right)^{i^k} R_i^{X_t} \quad \text{for } i = 1, \ldots, C-1; t = 2, \ldots, T.$$

The MR(m) model indicates that the log-odds, $\log(R_i^{X_1}/R_i^{X_t})$ for $i = 1, \ldots, C-1$, is expressed as the polynomial function of categories i as $\sum_{k=0}^{m-1} i^k \log \psi_k^{(t)}$. The number of degree of freedom for testing the goodness of fit of the MR(m) is $(T-1)(C-1-m)$.

The MH model is equivalent to the MR(m) model with $\psi_0^{(t)} = \cdots = \psi_{m-1}^{(t)} = 1$ for all $t = 2, \ldots, T$, however the MR(m) model generally has a different structure from the MH model. In a similar way the case of two ordinal outcomes, we are interested in what a structure is necessary to obtain the MH model in addition to the MR(m) model.

For a given k ($k = 0, 1, \ldots, C-2$), we consider a k-th marginal ridits equality (MRE(k)) model satisfied the following condition

$$\sum_{i=1}^{C-1} i^k R_i^{X_1} = \sum_{i=1}^{C-1} i^k R_i^{X_2} = \cdots = \sum_{i=1}^{C-1} i^k R_i^{X_T}.$$

The number of degree of freedom for testing the goodness of fit of the MRE(k) is $T-1$.

From Theorem 1, by considering the relationship between the X_1 and X_t for $t = 2, \ldots, T$, we obtain the following theorem:

Theorem 2. *For a given m ($m = 1, \ldots, C-1$), the MH model holds if and only if the MR(m) and MRE(k), for all $k = 0, 1, \ldots, m-1$, models hold.*

7. Concluding remarks

This study proposed the MR(m) model which can compare two marginal ridits for matched pair of ordered categorical data with same classifications. The proposed model was covered from the saturated model to the MH model. We could select an appropriate m (i.e., the appropriate MR(m)) for the data using conditional tests G^2(MH|MR(m)). The MH model is equivalent to the MR(m) model with $\psi_0 = \cdots = \psi_{m-1} = 1$, however the MR($m$) model generally has a different structure from the MH model. We gave the theorem that the MH model holds if and only if the MR(m) and MRE(k), for all $k = 0, 1, \ldots, m-1$, models hold. We demonstrated the utility of the MR(m) model using application to real data.

In Section 6, we dealt the MR(m) corresponding to the multi-way C^T contingency table. However, we have not demonstrated the utility of the MR(m) model using application to real data for the multi-way C^T contingency table. These are a matter to be considered in future studies.

Acknowledgements

The author would like to thank the anonymous reviewers and the editors for careful reading and comments to improve this paper.

References

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

Agresti A (2010). *Analysis of Ordinal Categorical Data*. Second edition. Wiley, Hoboken.

Akaike H (1974). "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control*, **AC-19**, 716–723. doi:10.1109/TAC.1974.1100705.

Bross IDJ (1958). "How to Use Ridit Analysis." *Biometrics*, **14**, 18–38. doi:10.2307/2527727.

Darroch JN, Ratcliff D (1972). "Generalized Iterative Scaling for Log-linear Models." *Annals of Mathematical Statistics*, **43**, 1470–1480.

Iki K, Tahata K, Tomizawa S (2012). "Measure of Departure from Marginal Homogeneity Using Marginal Odds for Multi-way Tables with Ordered Categories." *Journal of Applied Statistics*, **39**, 279–295. doi:10.1080/02664763.2011.586682.

Kurakami H, Tahata K, Tomizawa S (2010). "Extension of the Marginal Cumulative Logistic Model and Decompositions of Marginal Homogeneity for Multi-way Tables." *Journal of Statistics: Advances in Theory and Applications*, **3**, 135–152.

Kurakami H, Tahata K, Tomizawa S (2013). "Generalized Marginal Cumulative Logistic Model for Multi-way Contingency Tables." *SUT Journal of Mathematics*, **49**, 19–32.

McCullagh P (1977). "A Logistic Model for Paired Comparisons with Ordered Categorical Data." *Biometrika*, **64**, 449–453. doi:10.1093/biomet/64.3.449.

Schwarz G (1978). "Estimating the Dimension of a Model." *Annals of Statistics*, **6**, 461–464. doi:10.1214/aos/1176344136.

Stuart A (1955). "A Test for Homogeneity of the Marginal Distributions in a Two-way Classification." *Biometrika*, **42**, 412–416. doi:10.2307/2333387.

Tahata K, Tajima K, Tomizawa S (2006). "A Measure of Asymmetry of Marginal Ridits for Square Contingency Tables with Ordered Categories." *Journal of the Japanese Society of Computational Statistics*, **19**, 69–95. doi:10.5183/jjscs1988.19.69.

Tahata K, Tomizawa S (2008). "Generalized Marginal Homogeneity Model and Its Relation to Marginal Equimoments for Square Contingency Tables with Ordered Categories." *Advances in Data Analysis and Classification*, **2**, 295–311. doi:10.1007/s11634-008-0028-1.

Tomizawa S (1985). "Analysis of Data in Square Contingency Tables with Ordered Categories Using the Conditional Symmetry Model and its Decomposed Models." *Environmental Health Perspectives*, **63**, 235–239. doi:10.1289/ehp.8563235.

Tomizawa S (1993). "Diagonals-parameter Symmetry Model for Cumulative Probabilities in Square Contingency Tables with Ordered Categories." *Biometrics*, **49**, 883–887. doi:10.2307/2532209.

Tomizawa S (1995). "A Generalization of the Marginal Homogeneity Model for Square Contingency Tables with Ordered Categories." *Journal of Educational and Behavioral Statistics*, **20**, 349–360. doi:10.3102/10769986020004349.

Tomizawa S, Miyamoto N, Ashihara N (2003). "Measure of Departure from Marginal Homogeneity for Square Contingency Tables Having Ordered Categories." *Behaviormetrika*, **30**, 173–193. doi:10.2333/bhmk.30.173.

Affiliation:

Shuji Ando
Department of Information and Computer Technology
Faculty of Engineering
Tokyo University of Science
Katsushika-ku, Tokyo, Japan
E-mail: shuji.ando@rs.tus.ac.jp

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 51 *Submitted:* 2020-08-03
January 2022 *Accepted:* 2021-04-23