Journal of Statistical Theory and Applications
Volume 11, Number 1, 2012, pp. 9-22 ISSN 1538-7887

On Quasi-Symmetry Based on Ridit for Analysis of Square Contingency Tables

Kiyotaka Iki [*], Kouji Tahata [†] and Sadao Tomizawa [‡]
Department of Information Sciences, Faculty of Science and Technology,
Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

Abstract

For square contingency tables with the same row and column ordinal classifications, we propose a model of quasi-symmetry using the row and column marginal ridits scores. Using the proposed model, the model of equality of marginal mean ridits and the model of equality of marginal variance ridits, we give a theorem such that the symmetry model holds if and only if all these models hold. Moreover, we show that the likelihood ratio statistic for testing goodness-of-fit of the symmetry model is asymptotically equivalent to the sum of those for the decomposed models. The proposed model is illustrated with an application to occupational status data.

Key words: Ordinal data, Orthogonal decomposition, Quasi-symmetry, Ridit, Square contingency table, Symmetry.

AMS 2010 Subject Classifications: 62H17.

[*]E-mail: kiyotaka_iki@ybb.ne.jp
[†]E-mail: kouji_tahata@is.noda.tus.ac.jp
[‡]E-mail: tomizawa@is.noda.tus.ac.jp

1. Introduction

For an $R \times R$ square contingency table with the same row and column ordinal classifications, let p_{ij} denote the probability that an observation will fall in the ith row and jth column of the table ($i = 1, \ldots, R;\ j = 1, \ldots, R$). Bowker (1948) considered the symmetry (S) model, defined by

$$p_{ij} = \psi_{ij} \quad (i = 1, \ldots, R;\ j = 1, \ldots, R),$$

where $\psi_{ij} = \psi_{ji}$ (see also Bishop, Fienberg and Holland, 1975, p. 282). Caussinus (1965) considered the quasi-symmetry (QS) model, defined by

$$p_{ij} = \alpha_i \beta_j \psi_{ij} \quad (i = 1, \ldots, R;\ j = 1, \ldots, R),$$

where $\psi_{ij} = \psi_{ji}$ (see also Bishop et al., 1975, p. 286). Agresti (1983) proposed the linear diagonals-parameter symmetry (LDPS) model, defined by

$$p_{ij} = \alpha^i \beta^j \psi_{ij} \quad (i = 1, \ldots, R;\ j = 1, \ldots, R),$$

where $\psi_{ij} = \psi_{ji}$. Tomizawa (1991) proposed the extended linear diagonals-parameter symmetry (ELDPS) model, defined by

$$p_{ij} = \alpha^i \beta^j \gamma^{i^2} \delta^{j^2} \psi_{ij} \quad (i = 1, \ldots, R;\ j = 1, \ldots, R),$$

where $\psi_{ij} = \psi_{ji}$. A special case of this model obtained by putting $\gamma = \delta$ is the LDPS model. The LDPS and ELDPS models are also special cases of the QS model. We note that the ELDPS model may be expressed as

$$\log \left(\frac{p_{ij}}{p_{ji}} \right) = (j - i)\theta + (j - i)(j + i)\eta \quad (i < j).$$

Let $u_1 < \cdots < u_R$ denote the ordered known scores assigned to both the rows and columns of same classifications. The generalized LDPS model with $\{u_i\}$ instead of $\{i\}$ is the ordinal quasi-symmetry model (Agresti, 2002, p. 429).

Let X and Y denote the row and column variables, respectively. Let

$$r_i^X = \sum_{k=1}^{i-1} p_{k\cdot} + \frac{p_{i\cdot}}{2}, \quad \text{and} \quad r_i^Y = \sum_{l=1}^{i-1} p_{\cdot l} + \frac{p_{\cdot i}}{2},$$

where $p_{i\cdot} = \sum_{t=1}^{R} p_{it}$ and $p_{\cdot i} = \sum_{s=1}^{R} p_{si}$, for $i = 1, \ldots, R$. The $\{r_i^X\}$ and $\{r_i^Y\}$ are the marginal ridits; see Bross (1958). When we cannot assign known scores $\{u_i\}$ for analyzing given data, Iki, Tahata and Tomizawa (2009) considered the ridit score type quasi-symmetry (RQS) model, defined by

$$p_{ij} = \alpha^{v_i}\beta^{v_j}\psi_{ij} \quad (i = 1, \ldots, R;\ j = 1, \ldots, R),$$

where $\psi_{ij} = \psi_{ji}$ and $v_i = (r_i^X + r_i^Y)/2$, for $i = 1, \ldots, R$. Note that $\{v_i\}$ are unknown scores. A special case of this model obtained by putting $\alpha = \beta$ is the S model. The RQS model is the LDPS model with $\{i\}$ replaced by the ridit scores $\{v_i\}$. The RQS model may be expressed as

$$\log\left(\frac{p_{ij}}{p_{ji}}\right) = (v_j - v_i)\theta \quad (i < j).$$

This indicates that the log odds that an observation will fall in the (i, j)th cell instead of the (j, i)th cell, $i < j$, is proportional to the difference between the ridit scores, $v_j - v_i$.

Consider the data in Table 1, taken from Tominaga (1979, p. 131). These data describe the cross-classification of father's and son's occupational status categories in Japan which were examined in 1955. For these data, it may not be appropriate to assign equal interval scores or known scores to the category (i.e., the occupational status). Therefore we are interested in applying the model based on unknown scores (e.g., the RQS model) to these data. The RQS model also may be expressed as

$$R_{ij}(\{p_{st}\}) = \theta \quad (i < j),$$

where

$$R_{ij}(\{p_{st}\}) = \frac{1}{v_j - v_i}\log\left(\frac{p_{ij}}{p_{ji}}\right).$$

This indicates that the ratio of the log odds that a father-son pair will fall in the (i, j)th cell instead of the (j, i)th cell, $i < j$, to the difference between the father's and his son's status scores, is constant. Let n_{ij} denote the observed frequency in the (i, j)th cell of the square table with $n = \sum\sum n_{ij}$. Denote the observed proportion in the (i, j)th cell by $\tilde{p}_{ij}$, i.e., $\tilde{p}_{ij} = n_{ij}/n$. Table 2 shows the values of $\{v_k\}$ and $\{R_{ij}(\{p_{st}\})\}$ with $\{p_{st}\}$ replaced by $\{\tilde{p}_{st}\}$. We see from Table 2 that the values of $\{R_{ij}(\{\tilde{p}_{st}\})\}$ are not constant. Thus, for the data in Table 1, it is unlikely that the log odds that the occupational status categories for the father's and his son's pair are i and j $(i < j)$, respectively, instead of j and i, is proportional to the difference between the occupational status scores. Indeed, the RQS model fits these data poorly (see Table 3 and Section 5). We see from Table 2 that (1) the values of $R_{ij}(\{\tilde{p}_{st}\})$, $i < j$, decrease as j increases when i is fixed, (2) they decrease as i increases when j is fixed, and (3) they decrease as $i + j$

increases in parallel with the main-diagonal. Therefore for these data, the $\{R_{ij}(\{\tilde{p}_{st}\})\}$ may exhibit a function of the sum of the ridit scores (i.e., $\{v_j + v_i\}$), especially, a linear function of $\{v_j + v_i\}$. We are interested in proposing a model in which $R_{ij}(\{p_{st}\})$ are a linear function of $v_j + v_i$, $i < j$, instead of constant, namely, in proposing the ELDPS model with $\{i\}$ replaced by the ridit scores $\{v_i\}$.

Section 2 proposes a new model using the ridit scores. Section 3 gives a decomposition of the S model using the proposed model. Section 4 shows that the likelihood ratio statistic for testing goodness-of-fit of the S model is asymptotically equivalent to the sum of those for the decomposed models.

2. An extended ridit score type quasi-symmetry model

For an $R \times R$ contingency table with the same row and column ordinal classifications, consider a model defined by

$$p_{ij} = \alpha^{v_i}\beta^{v_j}\gamma^{v_i^2}\delta^{v_j^2}\psi_{ij} \quad (i = 1, \ldots, R;\ j = 1, \ldots, R),$$

where $\psi_{ij} = \psi_{ji}$. Note that $v_1 < \cdots < v_R$ are unknown scores defined in Section 1. A special case of this model obtained by putting $\gamma = \delta$ is the RQS model. This model is an extension of the RQS model. We shall refer to this model as the extended ridit score type quasi-symmetry (ERQS) model. A special case of the ERQS model obtained by putting $\alpha = \beta$ and $\gamma = \delta$ is the S model. The ERQS model implies the QS model. For the 3×3 contingency table, the ERQS model is identical to the QS model although the detail is omitted.

The ERQS model may be expressed as

$$\log\left(\frac{p_{ij}}{p_{ji}}\right) = (v_j - v_i)\theta + (v_j - v_i)(v_j + v_i)\eta \quad (i < j),$$

where $\theta = \log(\beta/\alpha)$ and $\eta = \log(\delta/\gamma)$. This indicates that the probability that an observation will fall in the (i, j)th cell, $i < j$, is $\exp\{(v_j - v_i)\theta + (v_j^2 - v_i^2)\eta\}$ times higher than the probability that the observation falls in the (j, i)th cell. Under the model, when $i = 1$, we see

$$\log\left(\frac{p_{1j}}{p_{j1}}\right) = \xi + \theta v_j + \eta v_j^2 \quad (j = 2, \ldots, R).$$

Therefore the log odds that an observation will fall in the $(1, j)$th cell, instead of the $(j, 1)$th cell, is a polynomial of degree 2 of score v_j. When $\eta = 0$, the ERQS model is identical to the RQS model, which indicates that the log odds, $\log(p_{1j}/p_{j1})$, is the linear function of score v_j. Thus, under the RQS model, the odds, p_{1j}/p_{j1}, increases (or decreases) exponentially as j increases; however, under the ERQS model, the odds does not always increase (or decrease) as j increase. For the data in Table 1, under the RQS model, the odds that a father's status category is less than the son's (i.e., p_{1j}/p_{j1}) will increase (or decrease) exponentially as the difference between the status scores for the pair, $v_j - v_1$, increases (i.e., as v_j increases ($j = 2, 3, 4$)). However, under the ERQS model, the odds does not always increase (or decrease) as the difference between the status scores for the pair increases. For analyzing the data, it may be possible to see such the structure in terms of the ERQS model although it is impossible in terms of the RQS model.

Under the ERQS model, if $\theta > 0$ and $\eta > 0$, then $\{p_{ij} > p_{ji}\}$ for all $i < j$, thus, if $\theta > 0$ and $\eta > 0$, then $\{F_i^X > F_i^Y\}$, where $F_i^X = \sum_{k=1}^{i} p_{k\cdot}$ and $F_i^Y = \sum_{k=1}^{i} p_{\cdot k}$, for all $i = 1, \ldots, R-1$. Therefore the parameters θ and η in the ERQS model may be useful for making inferences such as that X is stochastically less than Y or vice versa. On the other hand, under the ERQS model, if $\theta > 0$ and $\eta < 0$ (or $\theta < 0$ and $\eta > 0$), then it is likely that there is the structure of $\{p_{ij} > p_{ji}\}$ for some $i < j$ and $\{p_{kl} < p_{lk}\}$ for some $k < l$. Thus, for the data having such the structure, the ERQS model rather than the RQS model would be appropriate.

3. Decompositions of symmetry model

The mean ridit for the distribution of Y when the distribution of X is the identified distribution for calculating the ridits is $R_X(Y) = \sum_j r_j^X p_{\cdot j}$. Similarly, we have $R_Y(X) = \sum_i r_i^Y p_{i\cdot}$, and also $R_X(X) = R_Y(Y) = 0.5$, where $R_X(X) = \sum_i r_i^X p_{i\cdot}$, and $R_Y(Y) = \sum_j r_j^Y p_{\cdot j}$. Then, we shall refer to $R_X(Y) = R_Y(X)$ as the marginal mean ridits equality (MR) model (see Agresti, 1984, p. 209). The MR model also may be expressed as $\mu_1 = \mu_2$, where $\mu_1 = \sum_i v_i p_{i\cdot}$, and $\mu_2 = \sum_i v_i p_{\cdot i}$. Agresti (1984) considered the comparison between the marginal distributions using

$$\begin{aligned} \tau &= \sum_{i<j}\sum p_{i\cdot}p_{\cdot j} - \sum_{i>j}\sum p_{i\cdot}p_{\cdot j} \\ &= \mathrm{P}(X < Y) - \mathrm{P}(X > Y), \end{aligned}$$

where X is selected at random from the row marginal distribution and Y is selected independently at random from the column marginal distribution. We note that the marginal homogeneity (MH) model $\{$i.e., $p_{i\cdot} = p_{\cdot i}\}$ implies $\tau = 0$, and $\tau = 0$ is equivalent to the MR model, i.e., $R_X(Y) = R_Y(X)$.

Consider a model defined by $\sigma_1^2 = \sigma_2^2$, where $\sigma_1^2 = \sum_i (v_i - \mu_1)^2 p_{i\cdot}$, and $\sigma_2^2 = \sum_i (v_i - \mu_2)^2 p_{\cdot i}$. We shall refer to this model as the marginal variance ridits equality (VR) model. We obtain the decomposition of the S model as follows:

Theorem 1 : *The S model holds if and only if all the ERQS, MR and VR models hold.*

Proof : If the S model holds, then the ERQS, MR and VR models hold. Assuming that all the ERQS, MR and VR models hold, then we shall show that the S model holds. Let $\{p_{ij}^*\}$ denote the cell probabilities which satisfy all the ERQS, MR and VR models. Since the ERQS model holds, we see

$$\log p_{ij}^* = v_i^* \log \alpha + v_j^* \log \beta + (v_i^*)^2 \log \gamma + (v_j^*)^2 \log \delta + \log \psi_{ij},$$

where $\psi_{ij} = \psi_{ji}$ and v_i^* denote v_i with $\{p_{st}\}$ replaced by $\{p_{st}^*\}$. Let $\pi_{ij} = c^{-1}\psi_{ij}$ with $c = \sum_{i=1}^R \sum_{j=1}^R \psi_{ij}$. We note that $\sum_{i=1}^R \sum_{j=1}^R \pi_{ij} = 1$ with $0 < \pi_{ij} < 1$. Then, since $\{p_{ij}^*\}$ satisfy the ERQS, MR and VR models, we see

$$\log\left(\frac{p_{ij}^*}{\pi_{ij}}\right) = \log c + v_i^* \log \alpha + v_j^* \log \beta + (v_i^*)^2 \log \gamma + (v_j^*)^2 \log \delta, \tag{1}$$

and

$$\mu_1^* = \mu_2^*, \quad \text{and} \quad \nu_1^* = \nu_2^*, \tag{2}$$

where $\mu_1^* = \sum_{s=1}^R v_s^* p_{s\cdot}^*$, $\mu_2^* = \sum_{s=1}^R v_s^* p_{\cdot s}^*$, $\nu_1^* = \sum_{s=1}^R (v_s^*)^2 p_{s\cdot}^*$, $\nu_2^* = \sum_{s=1}^R (v_s^*)^2 p_{\cdot s}^*$, and $p_{s\cdot}^*$ and $p_{\cdot s}^*$ denote $p_{s\cdot}$ and $p_{\cdot s}$, respectively, with $\{p_{st}\}$ replaced by $\{p_{st}^*\}$. Then, we denote μ_1^* $(= \mu_2^*)$ by μ_0 and ν_1^* $(= \nu_2^*)$ by ν_0.

Consider the arbitrary cell probabilities $\{p_{ij}\}$ satisfying

$$\tilde{\mu}_1 = \tilde{\mu}_2 = \mu_0, \quad \text{and} \quad \tilde{\nu}_1 = \tilde{\nu}_2 = \nu_0, \tag{3}$$

where $\tilde{\mu}_1 = \sum_{s=1}^R v_s^* p_{s\cdot}$, $\tilde{\mu}_2 = \sum_{s=1}^R v_s^* p_{\cdot s}$, $\tilde{\nu}_1 = \sum_{s=1}^R (v_s^*)^2 p_{s\cdot}$ and $\tilde{\nu}_2 = \sum_{s=1}^R (v_s^*)^2 p_{\cdot s}$.

From (1), (2) and (3), we see

$$\sum_{i=1}^R \sum_{j=1}^R (p_{ij} - p_{ij}^*) \log\left(\frac{p_{ij}^*}{\pi_{ij}}\right) = 0. \tag{4}$$

Using the equation (4), we obtain

$$K(\{p_{ij}\}, \{\pi_{ij}\}) = K(\{p_{ij}^*\}, \{\pi_{ij}\}) + K(\{p_{ij}\}, \{p_{ij}^*\}),$$

where

$$K(\{a_{ij}\},\{b_{ij}\}) = \sum_{i=1}^{R}\sum_{j=1}^{R} a_{ij}\log\left(\frac{a_{ij}}{b_{ij}}\right),$$

and $K(\{a_{ij}\},\{b_{ij}\})$ is the Kullback-Leibler information between $\{a_{ij}\}$ and $\{b_{ij}\}$. Since $\{\pi_{ij}\}$ being a function of $\{p^*_{ij}\}$ is fixed, we see

$$\min_{\{p_{ij}\}} K(\{p_{ij}\},\{\pi_{ij}\}) = K(\{p^*_{ij}\},\{\pi_{ij}\}),$$

and then $\{p^*_{ij}\}$ uniquely minimizes $K(\{p_{ij}\},\{\pi_{ij}\})$ (see Bhapkar and Darroch, 1990).

Let $p^{**}_{ij} = p^*_{ji}$ for $1 \le i,j \le R$. Then

$$\log p^{**}_{ij} = \log p^*_{ji} = v^*_j \log\alpha + v^*_i \log\beta + (v^*_j)^2\log\gamma + (v^*_i)^2\log\delta + \log\psi_{ji}, \tag{5}$$

with $\psi_{ij} = \psi_{ji}$. Noting that $\{\pi_{ij} = \pi_{ji}\}$, the equation (5) is also expressed as

$$\log\left(\frac{p^{**}_{ij}}{\pi_{ij}}\right) = \log c + v^*_j \log\alpha + v^*_i\log\beta + (v^*_j)^2\log\gamma + (v^*_i)^2\log\delta. \tag{6}$$

From (2), (3) and (6), we see

$$\sum_{i=1}^{R}\sum_{j=1}^{R}(p_{ij} - p^{**}_{ij})\log\left(\frac{p^{**}_{ij}}{\pi_{ij}}\right) = 0. \tag{7}$$

Using the equation (7), we obtain

$$K(\{p_{ij}\},\{\pi_{ij}\}) = K(\{p^{**}_{ij}\},\{\pi_{ij}\}) + K(\{p_{ij}\},\{p^{**}_{ij}\}).$$

Since $\{\pi_{ij}\}$ being a function of $\{p^{**}_{ij}\}$ is fixed, we see

$$\min_{\{p_{ij}\}} K(\{p_{ij}\},\{\pi_{ij}\}) = K(\{p^{**}_{ij}\},\{\pi_{ij}\}),$$

and then $\{p^{**}_{ij}\}$ uniquely minimizes $K(\{p_{ij}\},\{\pi_{ij}\})$. Therefore, we see $\{p^*_{ij} = p^{**}_{ij}\}$. Thus, $\{p^*_{ij} = p^*_{ji}\}$. Namely the S model holds. The proof is completed.

We also obtain the following theorem.

Theorem 2 : *The S model holds if and only if both the RQS and MR models hold.*

The proof of Theorem 2 is omitted because it is obtained in a similar way. Theorems 1 and 2 may be useful for seeing the reason for the poor fit when the S model fits the data poorly.

4. Orthogonality of test statistic for symmetry model

Aitchison (1962) discussed the asymptotic separability, which is equivalent to the orthogonality in Read (1977) and the independence in Darroch and Silvey (1963), of the test statistics for goodness-of-fit of two models. Assume that the observed frequencies have a multinomial distribution. Let $G^2(M)$ denote the likelihood ratio statistic for testing goodness-of-fit of model M, and let df(M) be the number of degree of freedom.

Generally suppose that model M_3 holds if and only if both models M_1 and M_2 hold. As described in Darroch and Silvey (1963), (i) when the test statistic $G^2(M_3)$ is asymptotically equivalent to the sum of $G^2(M_1)$ and $G^2(M_2)$, where df(M_3) equals the sum of df(M_1) and df(M_2), if both M_1 and M_2 are accepted (at the α significance level) with high probability, then M_3 would be accepted; however (ii) when the asymptotic equivalence described above does not hold, such an incompatible situation that both M_1 and M_2 are accepted with high probability but M_3 is rejected with high probability is quite possible. Therefore, we shall show that the orthogonal decompositions of the S model hold for Theorems 1 and 2.

For the decomposition of the S model into three models in Theorem 1, we see that the orthogonality of test statistics does not hold although the detail is omitted. We shall modify the decomposed three models into two models. Consider a model defined by $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$, which combines two constraints of marginal mean ridits equality and marginal variance ridits equality. We shall refer to this model as the marginal mean and variance ridits equality (MVR) model. Then, we obtain the following theorem.

Theorem 3 : *The test statistic $G^2(S)$ is asymptotically equivalent to the sum of $G^2(ERQS)$ and $G^2(MVR)$, where df(S) equals the sum of df(ERQS) and df(MVR).*

Proof : Let $p = (p_{11}, \ldots, p_{1R}, \ldots, p_{R1}, \ldots, p_{RR})^T$ denote the $R^2 \times 1$ vector, where A^T denotes the transpose of matrix (or vector) A. The MVR model is expressed as

$$h_1(p) = 0_{d_1}$$

where

$$h_1(p) = \Big(\sum_{i=1}^{R} v_i(p_{i\cdot} - p_{\cdot i}), \sum_{i=1}^{R} v_i^2(p_{i\cdot} - p_{\cdot i}) \Big)^T,$$

and 0_s denotes the $s \times 1$ vector with all elements zero, and $d_1 = 2$. Since parameters θ and η

are expressed as the function of $\{p_{ij}\}$ in the ERQS model, this model can be written as

$$h_2(p) = 0_{d_2},$$

where

$$h_2(p) = (h_{14}(p), \ldots, h_{1R}(p), h_{23}(p), \ldots, h_{2R}(p), \ldots, h_{R-1,R}(p))^T,$$

with

$$\begin{aligned} h_{st}(p) &= (v_3 - v_2)(v_3 - v_1)(v_2 - v_1)\log(p_{st}/p_{ts}) \\ &\quad + \{(v_t - v_3) + (v_s - v_1)\}(v_t - v_s)(v_3 - v_1)\log(p_{12}/p_{21}) \\ &\quad - \{(v_t - v_2) + (v_s - v_1)\}(v_t - v_s)(v_2 - v_1)\log(p_{13}/p_{31}), \end{aligned}$$

and $d_2 = (R^2 - R - 4)/2$. From Theorem 1, the S model is expressed as

$$h_3(p) = 0_{d_3},$$

where

$$h_3(p) = (h_1(p)^T, h_2(p)^T)^T,$$

and $d_3 = R(R-1)/2$. Let $H_s(p)$, for $s = 1, 2, 3$, denote the $d_s \times R^2$ matrix of partial derivatives of $h_s(p)$ with respect to p, i.e., $H_s(p) = \partial h_s(p)/\partial p^T$. Let $\Sigma(p) = \mathrm{diag}(p) - pp^T$, where $\mathrm{diag}(p)$ denotes a diagonal matrix with ith component of p as ith diagonal component. Let $\tilde{p}$ denote p with $\{p_{ij}\}$ replaced by $\{\tilde{p}_{ij}\}$. Using the delta method, $\sqrt{n}(h_3(\tilde{p}) - h_3(p))$ has asymptotically a normal distribution with mean zero and covariance matrix

$$H_3(p)\Sigma(p)H_3(p)^T = \begin{bmatrix} H_1(p)\Sigma(p)H_1(p)^T & H_1(p)\Sigma(p)H_2(p)^T \\ H_2(p)\Sigma(p)H_1(p)^T & H_2(p)\Sigma(p)H_2(p)^T \end{bmatrix}.$$

We see that all elements of $H_1(p)\Sigma(p)H_2(p)^T$ equal zero under the S model. Thus we obtain $\Delta_3(p) = \Delta_1(p) + \Delta_2(p)$, where

$$\Delta_s(p) = h_s(p)^T[H_s(p)\Sigma(p)H_s(p)^T]^{-1}h_s(p),$$

for $s = 1, 2, 3$. Therefore, from Rao (1973, Sec. 6e. 3), Darroch and Silvey (1963) and Aitchison (1962), we obtain Theorem 3. The proof is completed.

We also obtain the following theorem.

Theorem 4 : *The test statistic $G^2(S)$ is asymptotically equivalent to the sum of $G^2(RQS)$ and $G^2(MR)$, where df(S) equals the sum of df(RQS) and df(MR).*

The proof of Theorem 4 is omitted because it is obtained in a similar way.

The maximum likelihood estimates of expected frequencies under the ERQS, MR, VR and MVR models could be obtained using the Newton-Raphson method in the log-likelihood equation. The number of degrees of freedom for testing goodness-of-fit of the ERQS model is $(R^2 - R - 4)/2$, which is one less than that for the RQS model.

5. Example

Consider the data in Table 1 again. We see from Table 3 that the S model fits these data poorly. Thus, we shall consider the models which indicate the structures of asymmetry. From Table 3 we see that the RQS and LDPS models fit these data poorly, however, the ERQS and ELDPS models fit these data well. Since it may not be appropriate to assign equal interval scores or known scores to the occupational status categories, the ERQS model would be preferable to the ELDPS model for these data.

Also the QS model fits these data well. We shall test the hypothesis that the ERQS model holds assuming that the QS model holds. Since the difference between the G^2 values of the ERQS and QS models is 1.29 which has an asymptotic chi-squared distribution with one degree of freedom under the hypothesis, there is not significant difference at the 0.05 level. Therefore, the ERQS model would be preferable to the QS model for these data.

From Theorem 1 we see that the poor fit of the S model is caused by the influence of the lack of the MR model rather than the ERQS and VR models. We note that for these data the value of the test statistic for the S model is very close to the sum of the values of those for the ERQS and MVR models.

Under the ERQS model, the maximum likelihood estimates of θ and η are $\hat{\theta} = 4.043$ and $\hat{\eta} = -6.903$. Since the value of $\hat{\eta}$ is not close to 0, this indicates that the RQS model fits these data poorly. Under the ERQS model, the maximum likelihood estimates of $\{v_k\}$ are $\hat{v}_1 = 0.057$, $\hat{v}_2 = 0.215$, $\hat{v}_3 = 0.429$ and $\hat{v}_4 = 0.771$. We note that the maximum likelihood estimates of unknown scores $\{v_k\}$ are very close to the values of $\{v_k\}$ based on the observed proportion $\{\tilde{p}_{st}\}$ in Table 2.

Under the ERQS model, the probability that the status category for the father in a pair is i and that for his son is j ($> i$), is estimated to be $\exp\{(\hat{v}_j - \hat{v}_i)\hat{\theta} + (\hat{v}_j^2 - \hat{v}_i^2)\hat{\eta}\}$ times higher than the probability that the status category for the father is j and that for his son is i. When $i = 1$, the odds that a father's status category is less than his son's, p_{1j}/p_{j1}, is estimated to be $\exp\{\hat{\xi} + \hat{\theta}\hat{v}_j + \hat{\eta}\hat{v}_j^2\}$ with $\hat{\xi} = -0.209$. Since $\hat{p}_{12}/\hat{p}_{21} = 1.41$, $\hat{p}_{13}/\hat{p}_{31} = 1.29$, and $\hat{p}_{14}/\hat{p}_{41} = 0.30$, the odds that a father's status category is less than his son's, is estimated to decrease (not exponentially) as the difference between the status scores for the pair increases. Similarly, when $i = 2$, since $\hat{p}_{23}/\hat{p}_{32} = 0.92$, and $\hat{p}_{24}/\hat{p}_{42} = 0.22$, the odds that a father's status category is less than his son's, p_{2j}/p_{j2}, is estimated to decrease as the differences between the status scores for the pair increase.

Moreover, from the maximum likelihood estimates of $\{p_{ij}/p_{ji}\}$, $i < j$, under the ERQS model, it is inferred that for these data there is the structure of $\{p_{ij} > p_{ji}\}$ for some $i < j$ and $\{p_{kl} < p_{lk}\}$ for some $k < l$ in these data. We note that these interpretation cannot be obtained under the RQS model.

6. Concluding remarks

The ERQS model rather than the ELDPS model may be appropriate when we cannot assign the integer scores to the categories of the square table.

We point out that for the decomposition in Theorem 1, $G^2(S)$ is not asymptotically equivalent to the sum of $G^2(ERQS)$, $G^2(MR)$ and $G^2(VR)$ because the sum of $G^2(MR)$ and $G^2(VR)$ is not asymptotically equivalent to $G^2(MVR)$, however, the $G^2(S)$ is asymptotically equivalent to the sum of $G^2(ERQS)$ and $G^2(MVR)$ (see Theorem 3). Therefore, for the decomposition of the S model into the ERQS and MVR models, the incompatible situation, described in Section 3, would not arize for analyzing the data.

Acknowledgements

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

References

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

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

Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed. New York: Wiley.

Aitchison, J. (1962). Large-sample restricted parametric tests. *Journal of the Royal Statistical Society, Series B*, **24**, 234-250.

Bhapkar, V. P., and Darroch, J. N. (1990). Marginal symmetry and quasi symmetry of general order. *Journal of Multivariate Analysis*, **34**, 173-184.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge: The MIT Press.

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

Bross, I. D. J. (1958). How to use ridit analysis. *Biometrics*, **14**, 18-38.

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.

Darroch, J. N., and Silvey, S. D. (1963). On testing more than one hypothesis. *Annals of Mathematical Statistics*, **34**, 555-567.

Iki, K., Tahata, K., and Tomizawa, S. (2009). Ridit score type quasi-symmetry and decomposition of symmetry for square contingency tables with ordered categories. *Austrian Journal of Statistics*, **38**, 183-192.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley.

Read, C. B. (1977). Partitioning chi-square in contingency table: A teaching approach. *Communications in Statistics-Theory and Methods*, **6**, 553-562.

Tominaga, K. (1979). *Nippon no Kaisou kouzou (Japanese Hierarchical Structure).* Tokyo: University of Tokyo Press (in Japanese).

Tomizawa, S. (1991). An extended linear diagonals-parameter symmetry model for square contingency tables with ordered categories. *Metron: International Journal of Statistics*, **49**, 401-409.

Table 1

Occupational status for Japanese father-son pairs which were examined in 1955; from Tominaga (1979, p. 131). (The parenthesized values are the maximum likelihood estimates of expected frequencies under the ERQS model.)

	Son's status				
Father's status	(1)	(2)	(3)	(4)	Total
(1)	80	72	37	19	208
	(80.30)	(68.08)	(35.55)	(20.47)	(204.40)
(2)	44	155	61	31	291
	(48.41)	(155.73)	(64.24)	(33.18)	(301.56)
(3)	26	73	218	45	362
	(27.53)	(69.98)	(217.71)	(40.02)	(355.24)
(4)	69	156	166	614	1005
	(67.59)	(154.04)	(170.59)	(612.58)	(1004.80)
Total	219	456	482	709	1866
	(223.83)	(447.83)	(488.09)	(706.25)	(1866)

Table 2

Values of $\{v_k\}$ and $\{R_{ij}(\{p_{st}\})\}$ with $\{p_{st}\}$ replaced by the observed proportion $\{\tilde{p}_{st} = n_{st}/n\}$ for the data in Table 1.

(a) Values of $\{v_k\}$ with $\{\tilde{p}_{st}\}$

$k = 1$	2	3	4
0.057	0.214	0.428	0.770

(b) Values of $\{R_{ij}(\tilde{p}_{st})\}$ for $i < j$

	$j = 2$	3	4
$i = 1$	3.131	0.952	−1.808
2		−0.843	−2.907
3			−3.809

Table 3

Likelihood ratio chi-squared values G^2 for models applied to the data in Table 1.

Applied models	Degree of freedom	G^2
S	6	205.08*
QS	3	0.82
RQS	5	60.42*
ERQS	4	2.11
LDPS	5	97.13*
ELDPS	4	2.27
MR	1	142.25*
VR	1	3.32
MVR	2	202.09*

* means significant at the 0.05 level.