SUT Journal of Mathematics
Vol. 53, No. 2 (2017), 89–105

Scaled double symmetry model and decomposition of double symmetry for square tables

Kouji Tahata, Ryotaro Maeda and Sadao Tomizawa

(Received Feburary 8, 2017; Revised July 26, 2017)

Abstract. For square contingency tables with ordered categories, Tahata and Tomizawa (2010) considered the double linear diagonals-parameter symmetry (DLDPS) model. The present paper proposes the generalized DLDPS model, which implies the structure of both asymmetry with represent to the main diagonal and with respect to the reverse diagonal in the table. Also, the double symmetry (DS) model is separated into the generalized DLDPS model and the moment equality model. Also it is shown that the test statistic for the DS model is asymptotically equivalent to the sum of those for separated models. An example is given.

AMS 2010 *Mathematics Subject Classification.* 62H17.

Key words and phrases. Contingency table, marginal double symmetry, ordered category, orthogonality, quasi-double symmetry.

§1. Introduction

For an $r \times r$ square contingency table with ordered categories, let p_{ij} denote the probability that an observation will fall in the cell in row i and column j ($i = 1, \dots, r; j = 1, \dots, r$). For the analysis of square contingency table, the models, which indicate the structure of symmetry with respect to the main diagonal of the table, have been proposed by many statisticians. For example, Bowker (1948), Stuart (1955), Caussinus (1965), Agresti (1983), and Tomizawa and Tahata (2007). On the other hand, the models, which indicate the structure of point symmetry with respect to the center point of the table, have also been proposed by, for example, Wall and Lienert (1976), Tomizawa (1985a), and Tahata and Tomizawa (2008).

Tomizawa (1985b) considered the double symmetry (DS) model defined by

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

where $i^* = r+1-i$ and $j^* = r+1-j$. This model indicates that the probabilities are symmetric with respect to the main diagonal of the table and also point-symmetric with respect to the center point of the table.

The double linear diagonals-parameter symmetry (DLDPS) model is defined by

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

where $\psi_{ij} = \psi_{ji} = \psi_{i^*j^*} = \psi_{j^*i^*}$ (Tahata and Tomizawa, 2010). This model may be expressed as

$$\frac{p_{ij}}{p_{ji}} = \theta^{j-i} \quad (i < j),$$

and

$$\frac{p_{j^*i^*}}{p_{ij}} = \eta^{r+1-(i+j)} \quad (i + j < r + 1).$$

Namely, this implies both of the structure of the linear diagonals-parameter symmetry (LDPS) model (Agresti, 1983) with respect to the main diagonal of the table and the structure of the LDPS model with respect to the reverse diagonal of the table.

Let X and Y denote the row and column variables, respectively. Tahata and Tomizawa (2010) also pointed out that the DS model holds if and only if both the DLDPS and the double mean equalities (DME) models hold, where the DME model is defined by

$$E(X) = E(Y) = \frac{r+1}{2}.$$

For the analysis of data, when the DLDPS model fits the data poorly, we are interested in applying the extended models of DLDPS model.

In the present paper, Section 2 proposes new models. Section 3 gives the decomposition of the DS model with the models. Section 5 shows relationship between test statistics. Section 6 gives an example.

§2. Scaled double symmetry model

We propose a new model defined by, for a fixed k ($k = 1, \ldots, r-1$),

$$p_{ij} = \left(\prod_{l=1}^{k} \alpha_l^{i^l} \beta_l^{j^l} \right) \psi_{ij} \quad (i = 1, \ldots, r; j = 1, \ldots, r), \tag{2.1}$$

where $\psi_{ij} = \psi_{ji} = \psi_{i^*j^*} = \psi_{j^*i^*}$. We shall refer to this model as the kth scaled double symmetry (SDS$_k$) model. A special case of this model obtained by putting $\{\alpha_l = \beta_l = 1\}$ is the DS model. When $k = 1$, the SDS_1 model is the DLDPS model. When $k = 2$, this model may be expressed as

$$\frac{p_{ij}}{p_{ji}} = \theta_{2a}^{j-i}\theta_{2b}^{j^2-i^2} \quad (i < j),$$

and

$$\frac{p_{j^*i^*}}{p_{ij}} = \eta_{2a}^{r+1-(i+j)}\theta_{2b}^{(j-i)\{r+1-(i+j)\}} \quad (i + j < r + 1),$$

where $\theta_{2a} = \beta_1/\alpha_1$, $\theta_{2b} = \beta_2/\alpha_2$ and $\eta_{2a} = \alpha_1\beta_1(\alpha_2\beta_2)^{r+1}$. Namely, this indicates that (i) there is the structure of the extended LDPS model proposed by Tomizawa (1991), and (ii) there is the structure of asymmetry with respect to the reverse diagonal of the table. A special case of SDS_2 model with $\theta_{2b} = 1$ is the DLDPS model. Also, when $k = 3$, the SDS_3 model may be expressed as

$$\frac{p_{ij}}{p_{ji}} = \theta_{3a}^{j-i}\theta_{3b}^{j^2-i^2}\theta_{3c}^{j^3-i^3} \quad (i < j),$$

and

$$\frac{p_{j^*i^*}}{p_{ij}} = \eta_{3a}^{T_{1(i,j)}}\eta_{3b}^{T_{2(i,j)}}\theta_{3b}^{S_{2(i,j)}}\theta_{3c}^{S_{3(i,j)}} \quad (i + j < r + 1),$$

where

$$\begin{aligned}
T_{1(i,j)} &= r + 1 - (i + j),\\
T_{2(i,j)} &= (r + 1)^3 - \frac{3}{2}(r + 1)^2(i + j) + \frac{3}{2}(r + 1)(j^2 + i^2) - (j^3 + i^3),\\
S_{2(i,j)} &= (j - i)\{r + 1 - (i + j)\},\\
S_{3(i,j)} &= \frac{3}{2}(r + 1)(j - i)\{r + 1 - (i + j)\},
\end{aligned}$$

and $\theta_{3a} = \beta_1/\alpha_1$, $\theta_{3b} = \beta_2/\alpha_2$, $\theta_{3c} = \beta_3/\alpha_3$, $\eta_{3a} = \alpha_1\beta_1(\alpha_2\beta_2)^{r+1}$ and $\eta_{3b} = \alpha_3\beta_3$. Namely, this indicates that (i) there is the structure of the 3rd linear asymmetry model proposed by Tahata and Tomizawa (2011), and (ii) there is the structure of asymmetry with respect to the reverse diagonal of the table. A special case of SDS_3 model with $\eta_{3b} = \theta_{3c} = 1$ is SDS_2 model.

Tomizawa (1985b) also considered the quasi-double symmetry (QDS) model defined by

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

where $\psi_{ij} = \psi_{ji} = \psi_{i^*j^*} = \psi_{j^*i^*}$. Consider the $\log p_{ij}$ for the QDS model and those for the SDS_{r-1} model. Then setting $\log\alpha_i$ by $\sum_{s=0}^{r-1} i^s \log\nu_s$ for $i = 1, \ldots, r$ and $\log\beta_j$ by $\sum_{s=0}^{r-1} j^s \log\xi_s$ for $j = 1, \ldots, r$. We see that two sets $\{\log\alpha_1, \log\alpha_2, \ldots, \log\alpha_r\}$ and $\{\log\nu_0, \log\nu_1, \ldots, \log\nu_{r-1}\}$ are in one-to-one correspondence. Similarly, the set $\{\log\beta_1, \log\beta_2, \ldots, \log\beta_r\}$ and the set $\{\log\xi_0, \log\xi_1, \ldots, \log\xi_{r-1}\}$ are in one to one correspondence. Therefore we pointed out that the SDS_{r-1} model is equivalent to the QDS model.

§3. Decomposition of double symmetry model

For a fixed k ($k = 1, \ldots, r-1$), consider a model defined by

$$E\Big(X^l\Big) = E\Big(Y^l\Big) = E\Big(X^{*l}\Big) = E\Big(Y^{*l}\Big) \quad (l = 1, \ldots, k).$$

Note that $E\Big(X^{*l}\Big) = E\Big((r+1-X)^l\Big)$. We shall refer to this model as the kth double moment equality (DME$_k$) model. When $k = 1$, the DME$_1$ model is equivalent to the DME model and when $k = r-1$, the DME$_{r-1}$ is equivalent to the marginal double symmetry (MDS) model, defined by

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

where $p_{i\cdot} = \sum_{t=1}^{r} p_{it}$, $p_{\cdot i} = \sum_{s=1}^{r} p_{si}$ (Tomizawa, 1985b). This model indicates both of the marginal homogeneity and the marginal point-symmetry. Figure 1 gives the relationships among models.

We obtain the following theorem.

Theorem 3.1. *For a fixed k ($k = 1, \ldots, r-1$), the DS model holds if and only if both the SDS$_k$ and DME$_k$ models hold.*

Proof. For a fixed k, if the DS model holds, then both the SDS$_k$ and DME$_k$ models hold. Assuming that both the SDS$_k$ and DME$_k$ models hold, then we shall show that the DS model holds. Let $\left\{p_{ij}^{(1)}\right\}$ denote the cell probabilities which satisfy both the SDS$_k$ and DME$_k$ models. Since the SDS$_k$ model holds, we see

$$\log p_{ij}^{(1)} = \sum_{l=1}^{k} i^l \log \alpha_l + \sum_{l=1}^{k} j^l \log \beta_l + \log \psi_{ij},$$

where $\psi_{ij} = \psi_{ji} = \psi_{i^*j^*} = \psi_{j^*i^*}$. 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 $0 < \pi_{ij} < 1$. Since the SDS$_k$ and DME$_k$ models hold,

$$\log\left(\frac{p_{ij}^{(1)}}{\pi_{ij}}\right) = \log c + \sum_{l=1}^{k} i^l \log \alpha_l + \sum_{l=1}^{k} j^l \log \beta_l \tag{3.1}$$

and

$$\mu_{1(1)}^{l} = \mu_{2(1)}^{l} = \mu_{1(1)}^{*l} = \mu_{2(1)}^{*l} \quad (l = 1, \ldots, k), \tag{3.2}$$

where

$$\mu_{1(1)}^{l} = \sum_{s=1}^{r}\sum_{t=1}^{r} s^l p_{st}^{(1)}, \quad \mu_{2(1)}^{l} = \sum_{s=1}^{r}\sum_{t=1}^{r} t^l p_{st}^{(1)},$$

$$\mu^{*l}_{1(1)} = \sum_{s=1}^{r}\sum_{t=1}^{r}(r+1-s)^l p^{(1)}_{st}, \quad \mu^{*l}_{2(1)} = \sum_{s=1}^{r}\sum_{t=1}^{r}(r+1-t)^l p^{(1)}_{st}.$$

Then we denote $\mu^l_{1(1)}(=\mu^l_{2(1)}=\mu^{*l}_{1(1)}=\mu^{*l}_{2(1)})$ by μ^l_0.

Consider the arbitrary cell probabilities $\left\{p^{(2)}_{ij}\right\}$ satisfying

$$\mu^l_{1(2)} = \mu^l_{2(2)} = \mu^{*l}_{1(2)} = \mu^{*l}_{2(2)} = \mu^l_0 \quad (l = 1, \ldots, k), \tag{3.3}$$

where

$$\mu^l_{1(2)} = \sum_{s=1}^{r}\sum_{t=1}^{r} s^l p^{(2)}_{st}, \quad \mu^l_{2(2)} = \sum_{s=1}^{r}\sum_{t=1}^{r} t^l p^{(2)}_{st},$$

$$\mu^{*l}_{1(2)} = \sum_{s=1}^{r}\sum_{t=1}^{r}(r+1-s)^l p^{(2)}_{st}, \quad \mu^{*l}_{2(2)} = \sum_{s=1}^{r}\sum_{t=1}^{r}(r+1-t)^l p^{(2)}_{st}.$$

From (3.1), (3.2) and (3.3), we see

$$\sum_{i=1}^{r}\sum_{j=1}^{r}\left(p^{(2)}_{ij} - p^{(1)}_{ij}\right)\log\left(\frac{p^{(1)}_{ij}}{\pi_{ij}}\right) = 0. \tag{3.4}$$

From (3.4) we obtain

$$K\left(p^{(2)}, \pi\right) = K\left(p^{(1)}, \pi\right) + K\left(p^{(2)}, p^{(1)}\right)$$

where for two bivariate discrete probability distribution $\{a_{ij}\}$ and $\{b_{ij}\}$

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

namely, $K(\cdot,\cdot)$ is the Kullback-Leibler information. Note that $K(a,b) \geq 0$ and the equality holds when only $a_{ij} = b_{ij}$. Since π is fixed, we see

$$\min_{p^{(2)}} K\left(p^{(2)}, \pi\right) = K\left(p^{(1)}, \pi\right)$$

and then $\left\{p^{(1)}_{ij}\right\}$ uniquely minimize $K(p^{(2)}, \pi)$ (see Darroch and Ratcliff, 1972).

Let $p^{(3)}_{ij} = p^{(1)}_{ji}$ for $1 \leq i, j \leq r$. Then, noting that $\{\pi_{ij} = \pi_{ji}\}$, we obtain

$$\min_{p^{(2)}} K\left(p^{(2)}, \pi\right) = K\left(p^{(3)}, \pi\right),$$

and then $\left\{p^{(3)}_{ij}\right\}$ uniquely minimize $K(p^{(2)}, \pi)$. Therefore, we see $p^{(1)}_{ij} = p^{(3)}_{ij}$. Thus, $p^{(1)}_{ij} = p^{(1)}_{ji}$ for $1 \leq i, j \leq r$.

Let $p^{(4)}_{ij} = p^{(1)}_{i^*j^*}$ for $1 \le i, j \le r$. Then, noting that $\{\pi_{ij} = \pi_{i^*j^*}\}$ and $\mu^l_0 = \sum_{s=1}^{r}\sum_{t=1}^{r} {s^*}^l p^{(1)}_{s^*t^*} = \sum_{s=1}^{r}\sum_{t=1}^{r} {t^*}^l p^{(1)}_{s^*t^*}$, we obtain

$$\min_{p^{(2)}} K\left(p^{(2)}, \pi\right) = K\left(p^{(4)}, \pi\right).$$

Then $\left\{p^{(4)}_{ij}\right\}$ uniquely minimize $K(p^{(2)}, \pi)$. Therefore, we see $p^{(1)}_{ij} = p^{(4)}_{ij}$. Thus, $p^{(1)}_{ij} = p^{(1)}_{i^*j^*}$ for $1 \le i, j \le r$. Namely the DS model holds. The proof is completed. □

Note that (i) Theorem 3.1 with $k = 1$ is given by Tahata and Tomizawa (2010) and (ii) Theorem 3.1 with $k = r - 1$ is given by Tomizawa (1985b). Also, note that Tahata and Tomizawa (2006) discussed another decomposition of double symmetry.

§4. Goodness-of-fit test

Assume that a multinomial distribution applies to the $r \times r$ table. The maximum likelihood estimates of expected frequencies under each model could be obtained by using the Newton-Raphson method to the log-likelihood equations or using the general iterative procedure for log-linear models.

Let n_{ij} and m_{ij} denote the observed frequency and the expected frequency in the (i, j)th cell, respectively. Also let $\hat{m}_{ij}$ denote the maximum likelihood estimates of m_{ij} under the model. Each model can be tested for goodness-of-fit by, e.g., the likelihood ratio chi-squared statistic of model with the corresponding degrees of freedom (df). The likelihood ratio chi-squared statistic for model M is

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

Table 1 gives the numbers of df for models.

§5. Partition of test statistics

We get the following lemma.

Lemma 5.1. *For a fixed k ($k = 1, \ldots, r-1$), the SDS_k model, which is defined by equation (2.1) in Section 2, can be expressed as (i) when k is odd,*

$$p_{ij} = \left(\prod_{s=1}^{\frac{k+1}{2}} \delta_{2s-1}^{{i^*}^{2s-1} - i^{2s-1}}\right)\left(\prod_{l=1}^{k} \beta_l^{j^l - i^l}\right)\omega_{ij} \quad (i = 1, \ldots, r; j = 1, \ldots, r),$$

where $\omega_{ij} = \omega_{ji} = \omega_{i^*j^*} = \omega_{j^*i^*}$, *and (ii) when* k *is even,*

$$p_{ij} = \left(\prod_{s=1}^{\frac{k}{2}} \delta_{2s-1}^{{i^*}^{2s-1} - i^{2s-1}} \right) \left(\prod_{l=1}^{k} \beta_l^{j^l - i^l} \right) \omega_{ij} \quad (i = 1, \dots, r; j = 1, \dots, r),$$

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

Proof. Consider the case of (i). In equation (2.1), let

$$\alpha_{2m-1} = \beta_{2m-1}^{-1} \prod_{s=m}^{\frac{k+1}{2}} \delta_{2s-1}^{\gamma_{2m-1(2s-1)}},$$

for $m = 1, \dots, \frac{k+1}{2}$, and

$$\alpha_{2m} = \beta_{2m}^{-1} \prod_{s=m+1}^{\frac{k+1}{2}} \delta_{2s-1}^{\gamma_{2m(2s-1)}},$$

for $m = 1, \dots, \frac{k-1}{2}$, where $\gamma_{s(l)}$ are given such that for $l = 2m - 1$ ($m = 1, 2, \dots, \frac{k+1}{2}$),

$${i^*}^l - i^l = (r + 1 - i)^l - i^l = \sum_{s=0}^{l} \gamma_{s(l)} i^s.$$

Also let,

$$\psi_{ij} = \left(\prod_{m=1}^{\frac{k+1}{2}} \delta_{2m-1}^{\gamma_{0(2m-1)}} \right) \omega_{ij}.$$

Therefore the SDS$_k$ model may be expressed as

$$\begin{aligned}
p_{ij} &= \left(\prod_{l=1}^{k} \alpha_l^{i^l} \beta_l^{j^l}\right) \psi_{ij} \\
&= \left(\prod_{m=1}^{\frac{k+1}{2}} \alpha_{2m-1}^{i^{2m-1}}\right) \left(\prod_{n=1}^{\frac{k-1}{2}} \alpha_{2n}^{i^{2n}}\right) \left(\prod_{l=1}^{k} \beta_l^{j^l}\right) \psi_{ij} \\
&= \prod_{m=1}^{\frac{k+1}{2}} \left(\beta_{2m-1}^{-1} \prod_{s=m}^{\frac{k+1}{2}} \delta_{2s-1}^{\gamma_{2m-1(2s-1)}}\right)^{i^{2m-1}} \\
&\quad \times \prod_{n=1}^{\frac{k-1}{2}} \left(\beta_{2n}^{-1} \prod_{s=n+1}^{\frac{k+1}{2}} \delta_{2s-1}^{\gamma_{2n(2s-1)}}\right)^{i^{2n}} \left(\prod_{l=1}^{k} \beta_l^{j^l}\right) \left(\prod_{m=1}^{\frac{k+1}{2}} \delta_{2m-1}^{\gamma_{0(2m-1)}}\right) \omega_{ij} \\
&= \left(\prod_{s=1}^{\frac{k+1}{2}} \delta_{2s-1}^{i^{*2s-1} - i^{2s-1}}\right) \left(\prod_{l=1}^{k} \beta_l^{j^l - i^l}\right) \omega_{ij}.
\end{aligned}$$

Similarly, the case of (ii) is proved although the detail is omitted. The proof is completed. □

We get the following theorem.

Theorem 5.2. *For a fixed k ($k = 1, \ldots, r-1$), the test statistic $G^2(DS)$ is asymptotically equivalent to the sum of $G^2(SDS_k)$ and $G^2(DME_k)$.*

Proof. First, for a fixed k ($k = 1, \ldots, r-1$), consider the case of r being even and k being odd. From Lemma 5.1, the SDS$_k$ model may be expressed as

$$\log p_{ij} = \sum_{l=1}^{k} (j^l - i^l)\epsilon_l + \sum_{s=1}^{\frac{k+1}{2}} (i^{*2s-1} - i^{2s-1})\zeta_{2s-1} + \phi_{ij}, \tag{5.1}$$

where $\phi_{ij} = \phi_{ji} = \phi_{i^*j^*} = \phi_{j^*i^*}$. Let

$$\boldsymbol{p} = (p_{11}, \ldots, p_{1r}, p_{21}, \ldots, p_{2r}, \ldots, p_{r1}, \ldots, p_{rr})^t,$$

$$\boldsymbol{x} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_k, \zeta_1, \zeta_3, \ldots, \zeta_k, \boldsymbol{\gamma})^t,$$

where "t" denotes the transpose, and

$$\boldsymbol{\gamma} = (\phi_{11}, \ldots, \phi_{1r}, \phi_{22}, \ldots, \phi_{2,r-1}, \phi_{33}, \ldots, \phi_{3,r-3}, \ldots, \phi_{\frac{r}{2},\frac{r}{2}}, \phi_{\frac{r}{2},\frac{r}{2}+1})$$

is the $1 \times r(r+2)/4$ vector. Then the SDS$_k$ model is expressed as

$$\log \boldsymbol{p} = \boldsymbol{A}\boldsymbol{x} = (\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots, \boldsymbol{a_k}, \boldsymbol{b_1}, \boldsymbol{b_3}, \ldots, \boldsymbol{b_k}, \boldsymbol{C})\boldsymbol{x},$$

where $\boldsymbol{A}$ is the $r^2 \times K$ matrix with $K = r(r+2)/4 + (3k+1)/2$ and

$$\begin{aligned} \boldsymbol{a_l} &= \boldsymbol{1_r} \otimes \boldsymbol{J_r^l} - \boldsymbol{J_r^l} \otimes \boldsymbol{1_r};\ \text{the } r^2 \times 1 \text{ vector} \quad (l = 1, ..., k), \\ \boldsymbol{b_{2s-1}} &= (\boldsymbol{J^*}_{\boldsymbol{r}}^{\boldsymbol{2s-1}} - \boldsymbol{J_r^{2s-1}}) \otimes \boldsymbol{1_r};\ \text{the } r^2 \times 1 \text{ vector} \quad (s = 1, ..., (k+1)/2), \end{aligned}$$

and $\boldsymbol{C}$ is the $r^2 \times r(r+2)/4$ matrix of 1 or 0 elements determined from (5.1), where $\boldsymbol{1_r}$ is the $r \times 1$ vector of 1 elements, $\boldsymbol{J_r^l} = (1^l, \ldots, r^l)^t$, $\boldsymbol{J^*}_{\boldsymbol{r}}^{\boldsymbol{l}} = (r^l, \ldots, 1^l)^t$, and $\otimes$ denotes the Kronecker product. Note that the matrix $\boldsymbol{A}$ is full column rank which is K. Also note that $\boldsymbol{C}\boldsymbol{1_{r(r+2)/4}} = \boldsymbol{1_{r^2}}$. In a similar manner to Haber (1985), we denote the linear space spanned by the columns of the matrix $\boldsymbol{A}$ by $S(\boldsymbol{A})$ with the dimension K. Let $\boldsymbol{U}$ be an $r^2 \times d_1$ full column rank matrix such that the linear space spanned by the column of $\boldsymbol{U}$, i.e., $S(\boldsymbol{U})$, is the orthogonal complement of the space $S(\boldsymbol{A})$, where $d_1 = r(3r-2)/4 - (3k+1)/2$. Thus, $\boldsymbol{U}^t\boldsymbol{A} = \boldsymbol{O_{d_1,K}}$ where $\boldsymbol{O_{d_1,K}}$ is the $d_1 \times K$ zero matrix. Therefore the SDS$_k$ model is expressed as $\boldsymbol{h_1}(\boldsymbol{p}) = \boldsymbol{0_{d_1}}$ where $\boldsymbol{0_{d_1}}$ is the $d_1 \times 1$ zero vector and $\boldsymbol{h_1}(\boldsymbol{p}) = \boldsymbol{U}^t \log \boldsymbol{p}$. The DME$_k$ model may be expressed as $\boldsymbol{h_2}(\boldsymbol{p}) = \boldsymbol{0_{d_2}}$ where $d_2 = (3k+1)/2$ and $\boldsymbol{h_2}(\boldsymbol{p}) = \boldsymbol{W}\boldsymbol{p}$ with

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{a_1}^t \\ \boldsymbol{a_2}^t \\ \vdots \\ \boldsymbol{a_k}^t \\ \boldsymbol{b_1}^t \\ \boldsymbol{b_3}^t \\ \vdots \\ \boldsymbol{b_k}^t \end{pmatrix};\ \text{the } \frac{3k+1}{2} \times r^2 \text{ matrix.}$$

Thus $\boldsymbol{W}^t$ belongs to the space $S(\boldsymbol{A})$, i.e., $S(\boldsymbol{W}^t) \subset S(\boldsymbol{A})$. Hence $\boldsymbol{W}\boldsymbol{U} = \boldsymbol{O_{d_2,d_1}}$. From Theorem 3.1, the DS model may be expressed as $\boldsymbol{h_3(p)} = \boldsymbol{0_{d_3}}$, where $d_3 = d_1 + d_2 = r(3r-2)/4$ with $\boldsymbol{h_3} = (\boldsymbol{h_1^t}, \boldsymbol{h_2^t})^t$.

Let $\boldsymbol{H_s}$, $s = 1, 2, 3$, denote the $d_s \times r^2$ matrix of partial derivatives of $\boldsymbol{h_s}(\boldsymbol{p})$ with respect to $\boldsymbol{p}$, i.e., $\boldsymbol{H_s} = \partial \boldsymbol{h_s}(\boldsymbol{p})/\partial \boldsymbol{p}^t$. Let $\boldsymbol{\Sigma} = diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^t$, where $diag(\boldsymbol{p})$ denote a diagonal matrix with the ith component of $\boldsymbol{p}$ as the ith diagonal component. Let $\hat{\boldsymbol{p}}$ denote $\boldsymbol{p}$ with p_{ij} replaced by $\hat{p}_{ij}$, where $\hat{p}_{ij} = n_{ij}/n$ with $n = \sum\sum n_{ij}$. Then $\sqrt{n}(\hat{\boldsymbol{p}} - \boldsymbol{p})$ has asymptotically a normal distribution with mean $\boldsymbol{0_{r^2}}$ and covariance matrix $\boldsymbol{\Sigma}$. Using the delta method, $\sqrt{n}(\boldsymbol{h_3}(\hat{\boldsymbol{p}}) - \boldsymbol{h_3}(\boldsymbol{p}))$ has asymptotically a normal distribution with mean $\boldsymbol{0_{d_3}}$ and covariance matrix

$$\boldsymbol{H_3 \Sigma H_3^t} = \begin{bmatrix} \boldsymbol{H_1 \Sigma H_1^t} & \boldsymbol{H_1 \Sigma H_2^t} \\ \boldsymbol{H_2 \Sigma H_1^t} & \boldsymbol{H_2 \Sigma H_2^t} \end{bmatrix}.$$

Since $\boldsymbol{H_1 p} = \boldsymbol{U}^t \boldsymbol{1_{r^2}} = \boldsymbol{0_{d_1}}$, $\boldsymbol{H_1} diag(\boldsymbol{p}) = \boldsymbol{U}^t$ and $\boldsymbol{H_2} = \boldsymbol{W}$, we see

$$\boldsymbol{H_1 \Sigma H_2^t} = \boldsymbol{U}^t \boldsymbol{W}^t = \boldsymbol{O_{d_1,d_2}}.$$

Thus $\Delta_3(\hat{\boldsymbol{p}}) = \Delta_1(\hat{\boldsymbol{p}}) + \Delta_2(\hat{\boldsymbol{p}})$ holds under $\boldsymbol{h_3}(\boldsymbol{p}) = \boldsymbol{0}$, where

$$\Delta_s(\hat{\boldsymbol{p}}) = n\boldsymbol{h_s}(\hat{\boldsymbol{p}})^t \left[\hat{\boldsymbol{H}}_{\boldsymbol{s}}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{H}}_{\boldsymbol{s}}^t\right]^{-1} \boldsymbol{h_s}(\hat{\boldsymbol{p}}).$$

Note that $\hat{\boldsymbol{H}}_{\boldsymbol{s}}$ and $\hat{\boldsymbol{\Sigma}}$ are given by $\boldsymbol{H_s}$ and $\boldsymbol{\Sigma}$ with p_{ij} replaced by $\hat{p}_{ij}$, respectively. Under each $\boldsymbol{h_s}(\boldsymbol{p}) = \boldsymbol{0}_{\boldsymbol{d_s}}$ $(s = 1, 2, 3)$, the statistic $\Delta_s(\hat{\boldsymbol{p}})$ has asymptotically a chi-squared distribution with d_s degrees of freedom. Since above equation holds and the asymptotic equivalence of the Wald statistic and the likelihood ratio statistic (Rao, 1973, Sec.6e.3; Darroch and Silvey, 1963; and Aitchison, 1962), we obtain the theorem. In a similar way, the other cases are proved although the detail is omitted. The proof is completed. □

Note that (i) Theorem 5.2 with $k = 1$ is given by Tahata and Tomizawa (2010) and (ii) Theorem 5.2 with $k = r - 1$ is given by Yamamoto, Takahashi and Tomizawa (2012).

§6. An example

Table 2, taken directly from Goodman (1981), is the cross-classification of father's and his son's occupational status categories in Denmark. These data are analyzed by many statisticians about association and marginal homogeneity. For example, Bishop, Fienberg and Holland (1975, p.100), Bartolucci and Forcina (2002), and Tahata and Yoshimoto (2015).

The main diagonal cells indicate that the pairs of father and his son have same occupational status, and the reverse diagonal cells mean the average of sum of occupational status of father and his son. Therefore, we are interested in applying the model that indicates the structure of symmetry (or asymmetry) with respect to the main diagonal and the reverse main diagonal. Table 3 gives the values of the likelihood ratio chi-square statistics G^2 for models applied to these data. We denote the move to son's level j from his father's level i by "$i \to j$". Namely p_{ij} is the probability of $i \to j$. The DS model fits these data very poorly since the value of G^2 is 893.23 with 16 df.

The SDS_1 (i.e., DLDPS) model fits these data poorly (Table 3). On the other hand, SDS_2, SDS_3 and SDS_4 (QDS) models fit well. We shall consider the hypothesis that the SDS_2 model holds assuming that the SDS_3 model holds. Then we can use the test based on the difference between the likelihood ratio chi-square statistics. This hypothesis is rejected at the 0.05 significance level since the difference between two likelihood ratio chi-square values is 10.77 with 2 df. We also consider another hypothesis that the SDS_3 model holds assuming that the SDS_4 (QDS) model holds. This hypothesis is accepted at the 0.05 significance level since the difference between two likelihood ratio chi-square values is 0.64 with 1 df. Therefore the SDS_3 model would be preferable to each of the SDS_2 and SDS_4 (QDS) models for these data.

Under the SDS_3 model, the MLEs of parameters, θ_{3a}, θ_{3b}, θ_{3c}, η_{3a} and η_{3b} are $\hat{\theta}_{3a} = 0.14$, $\hat{\theta}_{3b} = 1.95$, $\hat{\theta}_{3c} = 0.94$, $\hat{\eta}_{3a} = 0.91$ and $\hat{\eta}_{3b} = 1.03$, respectively. We can infer that the probability of $i \to j$ $(i < j)$ is estimated to be $\theta_{3a}^{j-i}\theta_{3b}^{j^2-i^2}\theta_{3c}^{j^3-i^3}$ times higher than the probability of $j \to i$, and we can infer that the probability of $j^* \to i^*$ $(i + j < 6)$ is estimated to be $\eta_{3a}^{T_{1(i,j)}}\eta_{3b}^{T_{2(i,j)}}\theta_{3b}^{S_{2(i,j)}}\theta_{3c}^{S_{3(i,j)}}$ times higher than the probability of $i \to j$. Namely, (i) the move to son's level j $(j = 2, 3, 4, 5)$ from his father's level 1 (highest) tends to be less than the move to son's level 1 from his father's level j, (ii) the move to son's level j from his father's level i $(1 < i < j)$ tends to be greater than the move to son's level i from his father's level j, and (iii) the move to son's level j from his father's level i $(i+j < 6)$ tends to be less than the move to son's level i^* from his father's level j^*. Namely, the sum of occupational status of father and his son tends to be greater than the average of it (note; the average of the sum of occupational status of father and his son is 6).

From Theorem 3.1 with $k = 2, 3$ and 4, we see that the reason for the poor fit of the DS model is caused by the influence of lack of the DME_k model rather than the SDS_k model.

§7. Concluding remarks

In this paper, we have proposed the SDS_k model. For the analysis of data, when the DLDPS model fits the data poorly, we can apply not only the QDS model, but also the SDS_k $(k = 2, 3, \dots, r-2)$ model, and so we could analyze the data more details. Moreover the models in the SDS_k models are referred to as hierarchical models (Figure 1). Thus it is easy to compare two models because the difference in G^2 values can be used to compare two nested models. Then the conditional test is more powerful than the unconditional test. For example, the hypothesis, that the SDS_2 model holds assuming that the SDS_3 model holds, is equivalent to the hypothesis $\eta_{3b} = \theta_{3c} = 1$. The $G^2(SDS_2|SDS_3)$ statistic has higher power than the $G^2(SDS_2)$ statistic for departures that are described by the SDS_3 model. Therefore, we recommend that (i) all SDS_k models are applied the dataset when the DS model fits the data poorly and (ii) we select the most appropriate model using the hierarchical structure.

In Section 3, we have given the separation of the DS model. When the DS model fits the data poorly, the separation of the DS model (i.e., Theorem 3.1) would be useful for seeing the reason for its poor fit. As seen in the analysis of Table 2, the poor fit of the DS model is caused by the poor fit of the DME_k

model rather than the SDS_k model ($k = 2, 3, 4$).

In Section 5, we have stated that for a fixed k ($k = 1, \ldots, r-1$), the test statistic $G^2(DS)$ is asymptotically equivalent to the sum of $G^2(SDS_k)$ and $G^2(DME_k)$ (i.e., Theorem 5.2). Therefore, for the separation of the DS model into the SDS_k and DME_k models, an incompatible situation, that both the SDS_k and DME_k models are accepted with high probability but the DS model is rejected with high probability, would not arise (see Aitchison, 1962; Darroch and Silvey, 1963). From Theorem 3.1, when the SDS_k model holds, the DME_k model is equivalent to the DS model. Thus, conditional on the SDS_k model, testing the DME_k model is equivalent to testing the DS model. Namely,

$$G^2(DS|SDS_k) = G^2(DS) - G^2(SDS_k) = G^2(DME_k|SDS_k).$$

$G^2(DS) - G^2(SDS_k)$ is asymptotically equivalent to $G^2(DME_k)$ from Theorem 5.2. We can obtain that the conditional test statistic $G^2(DME_k|SDS_k)$ is asymptotically equivalent to the unconditional test statistic $G^2(DME_k)$. Therefore even if the SDS_k model fits the data poorly, the test statistic might work well when the sample size is large.

Acknowledgments

The authors would like to thank an anonymous reviewer for valuable comments and suggestions to improve the quality of the paper.

References

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

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

[3] Bartolucci, F. and Forcina, A. (2002). Extended RC association models allowing for order restrictions and marginal modeling. *Journal of the American Statistical Association*, **97**, 1192-1199.

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

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

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

[7] Darroch, J. N. and Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, **43**, 1470-1480.

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

[9] Goodman, L. A. (1981). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, **68**, 347-355.

[10] Haber, M. (1985). Maximum likelihood methods for linear and log-linear models in categorical data. *Computational Statistics and Data Analysis*, **3**, 1-10.

[11] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. Wiley, New York.

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

[13] Tahata, K. and Tomizawa, S. (2006). Decompositions for extended double symmetry models in square contingency tables with ordered categories. *Journal of the Japan Statistical Society*, **36**, 91-106.

[14] Tahata, K. and Tomizawa, S. (2008). Orthogonal decomposition of point-symmetry for multiway tables. *Advances in Statistical Analysis*, **92**, 255-269.

[15] Tahata, K. and Tomizawa, S. (2010). Double linear diagonals-parameter symmetry and decomposition of double symmetry for square tables. *Statistical Methods and Applications*, **19**, 307-318.

[16] Tahata, K. and Tomizawa, S. (2011). Generalized linear asymmetry model and decomposition of symmetry for multiway contingency tables. *Journal of Biometrics and Biostatistics*, **2**, 1-6.

[17] Tahata, K. and Yoshimoto, T. (2015). Marginal asymmetry model for square contingency tables with ordered categories. *Journal of Applied Statistics*, **42**, 371-379.

[18] Tomizawa, S. (1985a). The decompositions for point symmetry models in two-way contingency tables. *Biometrical Journal*, **27**, 895-905.

[19] Tomizawa, S. (1985b). Double symmetry model and its decomposition in a square contingency table. *Journal of the Japan Statistical Society*, **15**, 17-23.

[20] Tomizawa, S. (1991). An extended linear diagonals-parameter symmetry model for square contingency tables with ordered categories. *Metron*, **49**, 401-409.

[21] Tomizawa, S. and Tahata, K. (2007). The analysis of symmetry and asymmetry: orthogonality of decomposition of symmetry into quasi-symmetry and marginal symmetry for multi-way tables. *Journal de la Société Française de Statistique*, **148**, 3-36.

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

[23] Yamamoto, K., Takahashi, F. and Tomizawa, S. (2012). Double symmetry model and its orthogonal decomposition for multi-way tables. *SUT Journal of Mathematics*, **48**, 83-102.

Table 1: The numbers of degrees of freedom (df) for models.

(a) r; even

Model	df
DS	$\dfrac{r(3r-2)}{4}$
SDS_k	$\begin{cases} \dfrac{r(3r-2)}{4} - \dfrac{3k}{2} & (k;\text{even}) \\ \dfrac{r(3r-2)}{4} - \dfrac{3k+1}{2} & (k;\text{odd}) \end{cases}$
DME_k	$\begin{cases} \dfrac{3k}{2} & (k;\text{even}) \\ \dfrac{3k+1}{2} & (k;\text{odd}) \end{cases}$

(b) r; odd

Model	df
DS	$\dfrac{(r-1)(3r+1)}{4}$
SDS_k	$\begin{cases} \dfrac{(r-1)(3r+1)}{4} - \dfrac{3k}{2} & (k;\text{even}) \\ \dfrac{(r-1)(3r+1)}{4} - \dfrac{3k+1}{2} & (k;\text{odd}) \end{cases}$
DME_k	$\begin{cases} \dfrac{3k}{2} & (k;\text{even}) \\ \dfrac{3k+1}{2} & (k;\text{odd}) \end{cases}$

Table 2: Occupational status for Danish father-son pairs; from Goodman (1981). (The upper, middle and lower parenthesized values are the estimated expected frequencies under the SDS_2, SDS_3 and SDS_4 models, respectively.)

Father's	Son's status					
status	(1)	(2)	(3)	(4)	(5)	Total
(1)	18	17	16	4	2	57
	(16.00)	(24.39)	(20.29)	(7.11)	(4.50)	(72.29)
	(13.73)	(19.08)	(15.90)	(5.91)	(3.80)	(58.42)
	(13.65)	(18.16)	(15.75)	(5.71)	(3.73)	(57.00)
(2)	24	105	109	59	21	318
	(25.39)	(91.74)	(99.33)	(57.38)	(15.66)	(289.50)
	(30.13)	(97.55)	(104.74)	(63.14)	(16.74)	(312.30)
	(31.27)	(97.24)	(108.51)	(63.88)	(17.10)	(318.00)
(3)	23	84	289	217	95	708
	(20.54)	(96.55)	(289.00)	(204.00)	(91.69)	(701.78)
	(22.88)	(95.49)	(289.00)	(216.27)	(92.91)	(716.55)
	(23.01)	(92.09)	(289.00)	(211.90)	(92.00)	(708.00)
(4)	8	49	175	348	198	778
	(6.53)	(50.62)	(185.12)	(361.26)	(211.26)	(814.79)
	(6.63)	(44.86)	(168.50)	(355.45)	(196.86)	(772.30)
	(6.80)	(44.12)	(172.50)	(355.76)	(198.82)	(778.00)
(5)	6	8	69	201	246	530
	(3.50)	(11.70)	(70.48)	(178.96)	(248.00)	(512.64)
	(4.20)	(11.72)	(71.31)	(193.93)	(250.27)	(531.43)
	(4.27)	(11.39)	(72.24)	(191.75)	(250.35)	(530.00)
Total	79	263	658	829	562	2391
	(71.96)	(275.00)	(664.22)	(808.71)	(571.11)	(2391.00)
	(77.57)	(268.70)	(649.45)	(834.70)	(560.58)	(2391.00)
	(79.00)	(263.00)	(658.00)	(829.00)	(562.00)	(2391.00)

Table 3: The values of likelihood ratio chi-squared statistic for the models applied to Table 2.

Models	df	G^2
DS	16	893.23*
SDS_1(DLDPS)	14	24.08*
SDS_2	13	22.24
SDS_3	11	11.47
SDS_4(QDS)	10	10.83
DME_1(DME)	2	832.10*
DME_2	3	840.51*
DME_3	5	861.17*
DME_4(MDS)	6	861.35*

Note: * Significant at 5% level.

$$
\begin{tikzcd}
 & \begin{array}{c}\text{SDS}_1\\ \text{(DLDPS)}\end{array} \arrow[r] & \text{SDS}_2 \arrow[r] & \cdots \arrow[r] & \begin{array}{c}\text{SDS}_{r-1}\\ \text{(QDS)}\end{array} \\
\text{DS} \arrow[ur] \arrow[dr] & & & & \\
 & \begin{array}{c}\text{DME}_{r-1}\\ \text{(MDS)}\end{array} \arrow[r] & \text{DME}_{r-2} \arrow[r] & \cdots \arrow[r] & \begin{array}{c}\text{DME}_1\\ \text{(DME)}\end{array}
\end{tikzcd}
$$

Figure 1: Relationships among models (A $\rightarrow$ B indicates that model A implies model B).

Kouji Tahata
Department of Information Sciences, Tokyo University of Science
Noda City, Chiba, 278-8510, Japan
E-mail: kouji_tahata@is.noda.tus.ac.jp

Ryotaro Maeda
Department of Information Sciences, Tokyo University of Science
Noda City, Chiba, 278-8510, Japan

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