Generalized Linear Asymmetry Model and Decomposition of Symmetry for Multiway Contingency Tables

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
A k-order generalization of the linear diagonals-parameter symmetry model is proposed, and related orthogonal decompositions of the generalization are inspected. Applications to randomized clinical trials are given.

Keywords: Linear diagonals-parameter symmetry; Linear ordinal quasi-symmetry; Marginal equimoment; Orthogonal decomposition; Symmetry

Introduction
Consider an r’t square contingency table with the same row and column classifications. Let $P_{a}$ denote the probability that an observation will fall in the $(i,j)$ th cell of the table $(i=1,…,r; j=1,…,r)$. For the analysis of square contingency tables, one of our interests is whether or not there is a structure of symmetry (or asymmetry) rather than independence in the table. The symmetry ($S^r$) model is defined by

$$P_{a} = \psi_{a},$$

where $\psi_{a} = \psi_{a}$. This indicates that the probability that an observation will fall in the $(i,j)$th cell is equal to the probability that the observation falls in the $(j,i)$th cell. As a model that has the weaker restrictions than the $S^r$ model, Lauritzen [1] considered the quasi-symmetry ($Q_i^r$) model defined by

$$P_{a} = \mu \beta^{\psi_{a}},$$

where $\psi_{a} = \psi_{a}$. A special case of this model with $\{\alpha = \beta\}$ is the $S^r$ model. Also Lauritzen [1] showed a theorem that the $Q_i^r$ model holds if and only if both the $Q_i^r$ and the marginal homogeneity models hold. For the analysis of data, the theorem (say decomposition of the $Q_i^r$ model) will be useful for seeing the reason for the poor fit when the $S^r$ model fits the data poorly.

The $S^r$ and $Q_i^r$ models indicate the structure of symmetry of cell probabilities and odds-ratios, respectively. As a model that indicates the structure of asymmetry of cell probabilities, Agresti [2] considered the linear diagonals-parameter symmetry ($LS^r$) model defined by

$$P_{a} = \mu \alpha^{\psi_{a}},$$

where $\psi_{a} = \psi_{a}$. This model is a special case of $Q_i^r$ model. In this way various symmetry and asymmetry models have been proposed by many statisticians (also see Agresti [3]; Tomizawa and Tahata [4]).

Consider an r’t contingency table with ordered categories. Let $i=(i_1,…,i_k)$ for $i_1=1,…,r$ ($k=1,…,T$), and let $P_{a}$ denote the probability that an observation will fall in the $i$th cell of the table. Let $X_i$ ($k=1,…,T$), denote the $k$th variable. Tahata et al. [5] considered the linear diagonals-parameter symmetry ($LS^r$), and extended $LS^r$ ($ELS^r$) models. The $ELS^r$ model is defined by

$$P_{a} = \mu \prod_{i=1}^{r} \alpha_{a_{i}}^{\psi_{a}},$$

where $\psi_{a} = \psi_{a}$ for any permutation $j=(j_1,…,j_k)$ of $i=(i_1,…,i_k)$. A special case of this model with $\{\alpha_{a_{i}} = 1\}$ is the $LS^r$ model. Also a special case of this model with $\{\alpha_{a_{i1}} = \alpha_{a_{i2}} = 1\}$ is the symmetry ($S^r$) model (Bhapkar and Darroch [6]; Agresti [3]). Note that when $T=2$, the $LS^r$ model is given by Agresti [2].

Tahata et al. [7] considered the $h$th linear ordinal quasi-symmetry ($LO_i^h$) model for fixed $h$ ($h=1,…,T-1$), defined by

$$P_{a} = \mu \prod_{h=1}^{r-1} \beta_{a_{h}}^{\psi_{a}},$$

where $\psi_{a} = \psi_{a}$ for any permutation $j=(j_1,…,j_k)$ of $i=(i_1,…,i_k)$. Note that the $LO_i^h$ model is a special case of the $h$th order quasi-symmetry ($Q_i^h$) model (Bhapkar and Darroch [6]), and $LS^r$ ($ELS^r$) models are special cases of the first order quasi-symmetry ($Q_i^1$) model. Note that the $Q_i^h$ model is defined by $LO_i^h$ with $\{\beta_{h}^{\psi_{a}}\}$ replaced by $\{t_{h_{k}}\} (h=1,…,h).

For the analysis of data, when the $S^r$ model does not hold, one may be interested in applying various asymmetry models; for example, the $LS^r$, $ELS^r$ and $LO_i^h$ models. If these models do not hold, we are interested in applying a more generalized asymmetry model. In addition we are interested in seeing the reason for the poor fit of the $S^r$ model by using the decomposition of the $S^r$ model. Thus the present paper proposes the generalization of the $ELS^r$ model, and gives the orthogonal decomposition of the $S^r$ model.

Generalized linear asymmetry model
Consider a new model defined by, for a fixed $k (k=1,…,r-1)$,

$$P_{a} = \mu \prod_{i=1}^{r} \alpha_{a_{i}}^{\psi_{a}},$$

where $\psi_{a} = \psi_{a}$ for any permutation $j=(j_1,…,j_k)$ of $i=(i_1,…,i_k)$. Without loss
of generality, we may set \( \alpha_{0(i)} = 1 \) \( (i = 1, \ldots, k) \). We shall refer to this model as the \( k \)-th linear asymmetry model (denoted by \( LS^k \)). A special case of the \( LS^k \) model with \( \{ \alpha_{s(i)} = \cdots = \alpha_{v(i)} = 1 \} \), \( s = 1, \ldots, T \), is the \( S^T \) model. Special cases of the \( LS^k \) model with \( k = 1 \) and with \( k = 2 \) are \( LS^T \) (i.e., \( LQ^T \)) and \( ELs^T \) models, respectively.

The \( Q_T^k \) model can be expressed as

\[
P_i = \mu_i \prod_{j=1}^{r} \gamma_{(i,j)},
\]

where \( \gamma_{(i,j)} \) for any permutation \( j = (j_1, \ldots, j_r) \) of \( i = (i_1, \ldots, i_r) \). This model can also be expressed as

\[
P_i = \mu_i \prod_{j=1}^{r} \lambda_{(i,j)},
\]

where \( \lambda_{(i,j)} = \gamma_{(i,j)}/\gamma_{(i,j)} \) and \( \phi_{(i,j)} = \gamma_{(i_1,j_1)} \cdots \gamma_{(i_r,j_r)} \gamma_{(i,j)} \). Then the \( Q_T^k \) model with \( \{ \lambda_{(i,j)} \} \) replaced by \( \{ \alpha_{(i,j)}^{(1)} \times \cdots \times \alpha_{(i,j)}^{(T-1)} \} \) can be expressed as

\[
P_i = \mu_i \prod_{j=1}^{r} \alpha_{(i,j)}^{(1)} \times \cdots \times \prod_{j=1}^{r} \alpha_{(i,j)}^{(T-1)},
\]

where \( \phi_{(i,j)} = \gamma_{(i_1,j_1)} \cdots \gamma_{(i_r,j_r)} \). Namely, this model indicates that the log ratio of symmetric cells is expressed as the polynomial. Note that the \( LS^k \) model is equivalent to the \( Q_T^k \) model. Consider the case of \( T = 2 \). For a fixed \( k (\leq r) \), the \( LS^k \) model can be expressed as

\[
P_i = \mu_i \prod_{j=1}^{r} \gamma_{(i,j)},
\]

where \( \gamma_{(i,j)} \). Under this model, the ratio of \( p_i/p_j \) is

\[
\frac{p_i}{p_j} = \prod_{j=1}^{r} \gamma_{(i,j)} \quad (i, j),
\]

where \( \gamma_{(i,j)} = (\alpha_{(i,j)}/\alpha_{(i,j)}) \). Namely, this model indicates that the log ratio of symmetric cells is expressed as the polynomial. Note that the \( LS^k \) model is equivalent to the \( Q_T^k \) model.

**Decomposition of symmetry model**

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

\[
E(X_i) = \cdots = E(X_i) \quad (i = 1, \ldots, k).
\]

We shall refer to this model as the marginal \( k \)-th moment equality (\( MME^k \)) model. Then we obtain the following theorem.

**Theorem 1**: For the \( r \times 1 \) table \( (T \geq 2) \) and \( k \)-fixed \( k = 1, \ldots, r-1 \), the \( S^T \) model holds if and only if both the \( LS^k \) and \( MME^k \) models hold.

We give the proof in the Appendix 1. Note that special cases of Theorem 1 with \( k = 1 \) and \( k = 2 \) are given by Tahata et al. [5].

Also although the detail is omitted, we can see that the \( MME^k \) model is equivalent to the marginal homogeneity (\( M^h \)) model defined by

\[
P(X_1 = i) = \cdots = P(X_r = i) \quad (i = 1, \ldots, r),
\]

and the \( LS^k \) model is equivalent to the \( Q_T^k \) model. Namely, a special case of Theorem 1 with \( k = r-1 \) is identical to the result given by Bhapkar and Darroch [6].

By the way, the \( MME^k \) model can be expressed as

\[
E(X_1) = \cdots = E(X_r),
\]

and the \( MME^k \) model is equivalent to

\[
E(X_1) = \cdots = E(X_r),
\]

and

\[
V(X_1) = \cdots = V(X_r).
\]

We are also interested in, for \( T \geq 3 \),

\[
E(X_1) = \cdots = E(X_r),
\]

and

\[
\text{Cov}(X_i, X_j) = C \quad (1 \leq i < j \leq T),
\]

where \( C \) is unknown constant. We shall refer to this model as the covariance equality (\( CE^T \)) model. Then, in a similar manner to Theorem 1 and Tahata et al. [7], we can obtain the following theorem.

**Theorem 2**: For the \( r \times 1 \) table \( (T \geq 3) \), the \( S^T \) model holds if and only if both the \( LS^k \) and \( CE^T \) models hold.

The relationships among models are given in (Figure 1).

**Orthogonal decomposition of test statistic**

Let \( n_{i,j,x} \) denote the observed frequency in the \( (i_1, \ldots, i_r) \)-th cell of the \( r \times 1 \) table \( (i_j = 1, \ldots, r; k = 1, \ldots, T) \) with \( n = \sum_{i=1}^{T} \sum_{j=1}^{T} n_{i,j,x} \), and let \( m_{i,j,k} \) denote the corresponding expected frequency. Assume that \( \{ n_{i,j,x} \} \) have a multinomial distribution. The maximum likelihood estimates of expected frequencies \( \{ m_{i,j,k} \} \) under the \( LS^k \), \( MME^k \) and \( CE^T \) models could be obtained using the iterative procedure, for example, the general iterative procedure for log-linear models of Darroch and Ratcliff [8] or using the Newton-Raphson method to the log-likelihood equations.

![Figure 1: Relationships among models (A→B” indicates that model A implies model B).](image)
Each model can be tested for goodness-of-fit by, e.g., the likelihood ratio chi-squared statistic with the corresponding degrees of freedom (df). The numbers of df for the $S^{i}$, $LS^{i}$, and $MME^{i}$ models are $r^{2} - L - N + N_{1}$ and $(T - 1)k$, respectively, where

$$L = \left( \frac{r + T - 1}{T} \right)^{r} T^{(r - 1)^{k}}.$$  

Also the number of df for the $LQ^{i}$ and $CE^{i}$ models are $r^{2} - L - N$ and $N$, respectively, where $N = (T - 2) + T(T - 1)/2$.

Let $G^{2}(M)$ denote the likelihood ratio statistic for testing goodness-of-fit of model $M$. Thus

$$G^{2}(M) = 2 \sum \cdots \sum \sum \frac{m_{i_{1} \cdots i_{n}}}{\hat{m}_{i_{1} \cdots i_{n}}},$$

where $\hat{m}_{i_{1} \cdots i_{n}}$ is the maximum likelihood estimate of expected frequency $m_{i_{1} \cdots i_{n}}$ under model $M$. Then we obtain the following theorem.

**Theorem 3:** For the $r^{2}$ table ($T \geq 2$) and $k$ fixed ($k = 1, \ldots, r - 1$), the test statistic $G^{2}(S^{i})$ is asymptotically equivalent to the sum of $G^{2}(LS^{i})$ and $G^{2}(MME^{i})$.

The proof of Theorem 3 is given in the Appendix 2. In a similar manner to Tahata et al. [5, 7], we can obtain the following theorem.

**Theorem 4:** For the $r^{2}$ table ($T \geq 3$), the test statistic $G^{2}(S^{i})$ is asymptotically equivalent to the sum of $G^{2}(LQ^{i})$ and $G^{2}(CE^{i})$.

Note that special cases of Theorem 3 with $k = 1$ and $k = 2$ are given by Tahata et al. [5].

**Analysis of data**

**Analysis of table 1:** The data in (Table 1), taken from Stuart [9], are constructed from unaided distance vision of 7477 women aged 30-39 employed in Royal Ordnance factories in Britain from 1943 to 1946 (see, e.g., Caussinus [1]; Tomizawa and Tahata [4]).

The $S^{i}$ model fits the data in (Table 1) poorly (see Table 3). By using the decompositions for the $S^{i}$ model, we shall consider the reason why the $S^{i}$ model fits these data poorly. The $LS^{i}$ ($k = 1, 2, 3$) models fit them well, but the $MME^{i}$ ($k = 1, 2, 3$) models fit them poorly. So, for example, we see from Theorem 1 with $k = 1$ (i.e., decomposition of the $S^{i}$ model into the $LS^{i}$ and $MME^{i}$ models) that the poor fit of the $S^{i}$ model is caused by the influence of the lack of structure of the $MME^{i}$ model (rather than the $LS^{i}$ model). From Theorem 1 with $k = 2$ or $k = 3$, we can obtain similar results.

The $LS^{i}$ model may be expressed as

$$\frac{P_{ij}}{p_{ij}} = a_{i1}^{(1)}a_{j1}^{(1)}a_{i2}^{(2)}a_{j2}^{(2)} \quad (1 \leq i < j \leq 4).$$

Therefore, under this model the probability that a woman’s right eye grade is $i$ and her left eye grade is $j$ ($i > j$) is estimated to be $a_{i1}^{(1)}a_{j1}^{(1)}a_{i2}^{(2)}a_{j2}^{(2)}$ times higher than the probability that the woman’s left eye grade is $i$ and her right eye grade is $j$, where $a_{i1}^{(1)} = 1.077$, $a_{i2}^{(2)} = 1.017$, and $a_{j2}^{(2)} = 0.998$. Note that $a_{i2}^{(2)}$ ($i = 1, 2, 3$) are the maximum likelihood estimates of $\alpha_{i2}$. In this similar manner, the interpretations under the $LS^{1}$ and $LS^{2}$ models are obtained although the details are omitted.

According to the test based on the difference between the $G^{2}$ values for the $LS^{i}$ and $LS^{i}$ models, the $LS^{i}$ model may be preferable to the $LS^{i}$ model, and in the similar manner, the $LS^{i}$ model may be preferable to the $LS^{i}$ model.

**Analysis of table 2:** Consider the data in (Table 2) taken from Tahata et al. [5]. These are the results of the treatment group only in randomized clinical trials conducted by a pharmaceutical company in anemic patients with cancer receiving chemotherapy. The response is the patient’s hemoglobin (Hb) concentration at baseline (before treatment) and following 4 weeks and 8 weeks of treatment. (Table 2) shows the $3 \times 3 \times 3$ array of counts of Hb response that is classified as treatment) and following 4 weeks and 8 weeks of treatment. Namely, we are interested in considering the transition of patient’s Hb concentration rather than the interchangeability of evenly spaced points in time with respect to those concentrations. For example, we want to see whether there is an asymmetric transition of those concentrations or not, when the value of those concentration at baseline was given.

We see from (Table 3) that (1) each of the $S^{i}$, $LS^{i}$ ($k = 1, 2$), $MME^{i}$ ($k = 1, 2$), and $CE^{i}$ models fits the data in (Table 2) poorly, however, (2) the $LQ^{i}$ model fits them well.

The $S^{i}$ model fits the data in (Table 2) poorly (see Table 3). By using the decompositions for the $S^{i}$ model, we shall consider the reason why the $S^{i}$ model fits these data poorly. The $LQ^{i}$ model fits them well, but the other models fit them poorly. So, we see from Theorem 2 (i.e., decomposition of the $S^{i}$ model into the $LQ^{i}$ and $CE^{i}$ models) that the poor fit of the $S^{i}$ model is caused by the influence of the lack of
The response categories are (1) ≥10 g/dl, (2) 8–10 g/dl, (3) <8 g/dl.

Table 2: Hemoglobin concentration at baseline, 4 weeks and 8 weeks in carcinoatous anemia patients from a randomized clinical trial; from Tahata et al. [5]. (Parenthesized values are the maximum likelihood estimates of expected frequencies under the $LQ^2_1$ model).

$$
\begin{array}{c|c|c|c|c|c|c}
\hline
\text{Baseline} & \text{4 weeks} & \hline
\text{1} & \text{2} & \text{3} \\
\hline
\text{1} & (1) & (2) & (3) \\
\hline
(1) & 77 & 7 & 1 \\
& (77.00) & (8.53) & (0.22) \\
(2) & 43 & 7 & 0 \\
& (39.80) & (9.35) & (0.13) \\
(3) & 3 & 0 & 0 \\
& (4.72) & (0.60) & (0.02) \\
(1) & 3 & 8 & 1 \\
& (4.67) & (7.69) & (0.74) \\
(2) & 17 & 16 & 5 \\
& (14.96) & (16.00) & (3.30) \\
(3) & 2 & 8 & 1 \\
& (2.80) & (6.42) & (0.70) \\
(1) & 1 & 1 & 1 \\
& (0.07) & (0.41) & (0.59) \\
(2) & 0 & 2 & 3 \\
& (0.33) & (5.28) & (4.03) \\
(3) & 4 & 3 & 3 \\
& (0.39) & (3.27) & (3.00) \\
\hline
\end{array}
$$

The symbol * means significant at 5% level.

In this paper, we have proposed the $k$-order generalization of the linear diagonals-parameter symmetry model that is including the first order quasi-symmetry model, and have given the decomposition of the symmetry model. When the $S_1$ model fits the data poorly, the decomposition of the $S_1$ model (i.e., Theorems 1 and 2) would be useful for seeing the reason for its poor fit. As seen in analysis of (Tables 1, 2), we can see that (1) for the data in (Table 1), the poor fit of the $S_1$ model is caused by the poor fit of the $MME^T_2$ models rather than the $L_1^S$ models, and (2) in the $S_1$ model the poor fit of the $S_1$ model is caused by the $CE^T$ model rather than the $L_1^Q$ model.

In Section 3 we have shown a theorem that the $S_1$ model holds if and only if both the $L_1^S$ and $MME^T_2$ models hold for a fixed $k$ ($k=1,2,3$). Also, we gave the asymptotic equivalence of test statistic for the $S_1$ model in Theorem 3. Thus, for the orthogonal decomposition of the $S_1$ model into the $L_1^S$ and $MME^T_2$ models, an incompatable situation, that both the $L_1^S$ and $MME^T_2$ models are accepted with high probability, but the $S_1$ model is rejected with high probability, would not arise. For the orthogonal decomposition of the $S_1$ model into the $L_1^Q$ and $CE^T$ models, we can also obtain similiar results. (For the details of orthogonal decomposition, see Darroch and Silvey [10]).

The authors would like to thank a referee to give many helpful comments and suggestions.

References

1. Caussinus H (1965) Contribution à l’analyse statistique des tableaux de corrélation. Annales de la Faculté des Sciences de l’Université de Toulouse 29,
77-183.

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

3. Agresti A (2002) Categorical Data Analysis. (2nd edn), Wiley, New York.

4. Tomizawa S, 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.

5. Tahata K, Yamamoto H, Tomizawa S (2008) Orthogonality of decompositions of symmetry into extended symmetry and marginal equiangularity for multi-way tables with ordered categories. Austrian Journal of Statistics 37: 185-194.

6. Bhapkar VP, Darroch JN (1990) Marginal symmetry and quasi symmetry of general order. Journal of Multivariate Analysis 34: 173-184.

7. Tahata K, Yamamoto H, Tomizawa S (2011) Linear ordinal quasi-symmetry model and decomposition of symmetry for multi-way tables. Mathematical Methods of Statistics 20: 158-164.

8. Darroch JN, Ratcliff D (1972) Generalized iterative scaling for log-linear models. Annals of Mathematical Statistics 43: 1470-1480.

9. Stuart A (1953) The estimation and comparison of strengths of association in contingency tables. Biometrika 40: 105-110.

10. Darroch JN, Silvey SD (1963) On testing more than one hypothesis. Annals of Mathematical Statistics 34: 555-567.

Appendix 1

We shall show the proof of Theorem 1 when $T=2$. For a fixed $k$, if the $S$-model holds, then the $LS^1$ and $MME^2$ models hold. Assuming that both the $LS^1$ and $MME^2$ models hold, then we shall show the $S$-model holds. Let $\{p_{ij}\}$ denote the cell probabilities which satisfy both the $LS^1$ and $MME^2$ models. Since the $LS^1$ model holds, we see

$$\log p_{ij} = \log \mu + \sum_k \left( \log \alpha_{ik} + \log \alpha_{jk} \right) + \log \nu_{ij},$$

where $\nu_{ij} = \nu_{ij}$. Let $\nu_{ij} = \nu_{ij}^*/c$ with $c = \sum \nu_{ij}$. Then the $LS^1$- and $MME^2$-models are expressed as

$$\log \frac{p_{ij}}{\nu_{ij}} = \log \mu + \sum_k \left( \log \alpha_{ik} + \log \alpha_{jk} \right),$$

and

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

where $\mu_{ij} = \sum_{k=1}^r s_{ik} \hat{p}_{ik}$ and $\mu_{ij}^* = \sum_{k=1}^r s_{ik} \mu_{ik}$. Then we denote $\mu_{ij}^* = \mu_{ij}^* \hat{p}_{ij}$. Consider the arbitrary cell probabilities $\{p_{ij}\}$ satisfying $\mu_{ij} = \mu_{ij}^* \hat{p}_{ij}$, where $\mu_{ij}^*$ denote $\mu_{ij}^* \hat{p}_{ij}$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$. Let $K\{a_{ij}\};[b_{ij}]\}$ be the Kullback-Leibler information between $\{a_{ij}\}$ and $\{b_{ij}\}$, where

$$K\{a_{ij}\};[b_{ij}]\} = \sum_{i=1}^n \sum_{j=1}^n \left( a_{ij} \log \frac{a_{ij}}{b_{ij}} \right).$$

From above equations, we see

$$\sum_{i=1}^n \sum_{j=1}^n \left( p_{ij} - p_{ij}^* \right) \log \frac{p_{ij}}{\nu_{ij}} = 0.$$

Thus we can obtain $K\{p_{ij}\};[\{\hat{p}_{ij}\}] = K\{\{\hat{p}_{ij}\}\};[\{\hat{p}_{ij}\}]$. Since $\{\hat{p}_{ij}\}$ is fixed, we see $\min_K K\{p_{ij}\};[\{\hat{p}_{ij}\}] = K\{\{\hat{p}_{ij}\}\};[\{\hat{p}_{ij}\}]$, and then $\{\hat{p}_{ij}\}$ uniquely minimize $K\{p_{ij}\};[\{\hat{p}_{ij}\}]$. Therefore, we see $p_{ij}^* = p_{ij}$ for $1 \leq i, j \leq r$. Namely, the $S$-model holds. The proof is completed in case of $T=2$. For the case of $T > 2$, we can obtain by the same way.

Appendix 2

We shall show the proof of Theorem 3 when $T=2$. For a fixed $k$, the $LS^1$ model may be expressed as

$$\log p_{ij} = \sum_{l=1}^r \left( j^l - i^l \right) \beta_l + \phi_{ij} (1 \leq i, j \leq r),$$

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

$$p = (p_1, \ldots, p_r, p_{21}, \ldots, p_{rr}, \ldots, p_{r1}, \ldots, p_{rr}),$$

$$\beta = (\beta_1, \beta_2, \phi),$$

where $\phi$ denotes the transpose and

$$\phi = (\phi_{11}, \phi_{12}, \ldots, \phi_{1r}, \phi_{22}, \ldots, \phi_{rr}),$$

is the $1 \times (r+1)/2$ vector of $\phi_{ij}$ for $1 \leq i \leq j \leq r$. Then the $LS^1$ model is expressed as

$$\log p = \lambda \beta = (\lambda_1, \ldots, \lambda_r, \lambda_{r1}, \ldots, \lambda_{rr}),$$

where $\lambda$ is the $r \times K$ matrix with $K = (r+1)/2 + k$ and $\lambda_{ij} = 1$ if $i \neq j$. Then noting that $\hat{p}_{ij} = \hat{p}_{ij}^*$ for $1 \leq i \leq j \leq r$. Then the $LS^1$ model is expressed as

$$\log \lambda = \lambda \beta = (\lambda_1, \ldots, \lambda_r, \lambda_{r1}, \ldots, \lambda_{rr}).$$

The $MME^2$ model may be expressed as $\lambda_{ij} = 0$ where $\lambda_{ij} = 0\nu_{ij}$ with $\nu_{ij} = 0$.

Thus $W$ belongs to the space $S(X)$, i.e., $S(W) = S(X)$. Hence $\nu = \nu_{ij}$ where $\lambda_{ij} = 0$. From Theorem 1, the $S$-model may be expressed as $\lambda_{ij} = 0$, where $d_i = d_i + d_i = r(r-1)/2$ with $d_i = (k, k')$. Let $H_i\{p\}$, $i = 1, 2, 3$, denote the $d_i \times d_i$ matrix of partial derivatives of $h_i\{p\}$ with respect to $p_i$, i.e., $H_i\{p\} = \partial h_i\{p\}/\partial p_i$. Let $S\{p\} = diag\{p\} - p p'$, where $diag\{p\}$ denotes a diagonal matrix with $i$ th diagonal component. Let $\hat{p}$ denote $p$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$, where $\hat{p}_{ij} = n_{ij}/n$. Then $\hat{p}$ has asymptotically a normal distribution with mean $0$ and covariance matrix $Cov\{p\}$. Using the delta method, $\hat{p} = h_i\{p\} - h_i\{\hat{p}\}$ has asymptotically a normal distribution with mean $0$ and covariance matrix $Cov\{p\}$. We see that $H_i\{p\} \Sigma\{p\} H_j\{p\} = H_i\{p\} \Sigma\{p\} H_j\{p\}$ and $H_i\{p\} \Sigma\{p\} H_j\{p\} = H_i\{p\} \Sigma\{p\} H_j\{p\}$. Therefore we obtain

$$H_i\{p\} \Sigma\{p\} H_j\{p\} = H_i\{p\} \Sigma\{p\} H_j\{p\} = H_i\{p\} \Sigma\{p\} H_j\{p\}.$$
Under each $h_s(r) = 0$ ($s = 1, 2, 3$), the Wald statistic $W_s = n \lambda_s (\hat{r})$ has asymptotically a chi-squared distribution with $d_s$ degrees of freedom. From the asymptotic equivalence of the Wald statistic and likelihood ratio statistic, we obtain Theorem 3 when $T = 2$. The proof is completed. The proof of Theorem 3 when $T > 2$ is omitted because it is obtained in the same way.