

Decomposition of Symmetry into Ordinal Quasi-Symmetry and Marginal Equipoment for Multi-way Tables

Hideharu Yamamoto^{1,2}, Toshiya Iwashita¹ and Sadao Tomizawa¹

¹ Dept. of Information Sciences, Tokyo University of Science Japan

² Chugai Clinical Research Center Co., Ltd., Tokyo, Japan

Abstract: For the analysis of square contingency tables with ordered categories, Agresti (1983) introduced the linear diagonals-parameter symmetry (LDPS) model. Tomizawa (1991) considered an extended LDPS (ELDPS) model, which has one more parameter than the LDPS model. These models are special cases of Caussinus (1965) quasi-symmetry (QS) model. Caussinus showed that the symmetry (S) model is equivalent to the QS model and the marginal homogeneity (MH) model holding simultaneously. For square tables with ordered categories, Agresti (2002, p.430) gave a decomposition for the S model into the ordinal quasi-symmetry and MH models. This paper proposes some decompositions which are different from Caussinus' and Agresti's decompositions. It gives (i) two kinds of decomposition theorems of the S model for two-way tables, (ii) extended models corresponding to the LDPS and ELDPS, and the generalized model further for multi-way tables, and (iii) three kinds of decomposition theorems of the S model into their models and marginal equipoment models for multi-way tables. The proposed decompositions may be useful if it is reasonable to assume the underlying multivariate normal distribution.

Zusammenfassung: Zur Analyse quadratischer Kontingenztafeln mit geordneten Kategorien führte Agresti (1983) das lineare Diagonal-Parameter Symmetrie (LDPS) Modell ein. Tomizawa (1991) betrachtete ein erweitertes LDPS (ELDPS) Modell, das um einen Parameter mehr hat als das LDPS Modell. Diese Modelle sind Spezialfälle des Quasi-Symmetrie (QS) Modells von Caussinus (1965). Caussinus zeigte, dass das Symmetrie (S) Modell äquivalent dem QS Modell ist und dass das marginale Homogenitäts- (MH) Modell dann auch hält. Für quadratische Tafeln mit geordneten Kategorien gab Agresti (2002, p.430) eine Zerlegung des S Modells in das ordinale Quasi-Symmetrie und das MH Modell an. Wir schlagen Zerlegungen vor, die sich von jenen in Caussinus und Agresti unterscheiden. Wir liefern (i) zwei Arten Zerlegungssätze des S Modells für zwei-weg Tafeln, (ii) erweiterte Modelle entsprechend dem LDPS und ELDPS, das generalisierte Modell für mehr-weg Tafeln, and (iii) drei Arten Zerlegungssätze des S Modells in deren Modelle und marginal Equipoment Modelle für mehr-weg Tafeln. Die vorgeschlagenen Zerlegungen könnten nützlich sein, falls die Annahme einer zugrunde liegenden multivariaten Normalverteilung begründet ist.

Keywords: Linear Diagonals-Parameter Symmetry, Marginal Homogeneity, Square Contingency Tables, Underlying Normal Distribution.

1 Introduction

Suppose that an $R \times R$ square contingency table has the same categories in the row classification as in the column classification. Let X_1 and X_2 denote the row and column variables, respectively, and let p_{ij} denote the probability that an observation will fall in the i th row and j th column of the table ($i, j = 1, \dots, R$). Thus, $\Pr(X_1 = i, X_2 = j) = p_{ij}$. The symmetry (S) model is defined as

$$p_{ij} = \psi_{ij}, \quad i, j = 1, \dots, R,$$

where $\psi_{ij} = \psi_{ji}$ (Bowker, 1948; Bishop, Fienberg, and Holland, 1975, p.282). This indicates that the probability that an observation will fall in the (i, j) cell, $i \neq j$, is equal to the probability that the observation falls in the symmetric (j, i) cell.

Caussinus (1965) considered the quasi-symmetry (QS) model, defined by

$$p_{ij} = \mu\alpha_i\beta_j\psi_{ij}, \quad i, j = 1, \dots, R,$$

where $\psi_{ij} = \psi_{ji}$. A special case of this model with $\{\alpha_i = \beta_i\}$ is the S model. Denote the odds ratio for rows i and j ($> i$) and columns s and t ($> s$) by $\theta_{(i<j;s<t)}$. Thus $\theta_{(i<j;s<t)} = (p_{is}p_{jt})/(p_{js}p_{it})$. Using the odds ratios, the QS model is further expressed as

$$\theta_{(i<j;s<t)} = \theta_{(s<t;i<j)}, \quad 1 \leq i < j \leq R; 1 \leq s < t \leq R.$$

Therefore, the QS model has characterization in terms of symmetry of odds ratio. For the QS model, also see, e.g., Bishop et al. (1975, p.286), Goodman (1979a), Darroch and McCloud (1986), and Agresti (2002, p.425).

The marginal homogeneity (MH) model is defined by

$$p_{i.} = p_{.i}, \quad i = 1, \dots, R,$$

where $p_{i.} = \sum_{t=1}^R p_{it}$ and $p_{.i} = \sum_{s=1}^R p_{si}$ (Stuart, 1955, Bishop et al., 1975, p.293). This indicates that the row marginal distribution is identical with the column marginal distribution.

For square tables with ordered categories, Agresti (1984, p.203) proposed the linear diagonals-parameter symmetry (LDPS) model defined by

$$p_{ij} = \begin{cases} \delta^{j-i}\psi_{ij} & \text{if } i < j, \\ \psi_{ij} & \text{if } i \geq j, \end{cases}$$

where $\psi_{ij} = \psi_{ji}$. A special case of this model obtained by putting $\delta = 1$ is the S model. Note that the LDPS model is a special case of the diagonals-parameter symmetry model of Goodman (1979b). The LDPS model may be also expressed as

$$p_{ij} = \delta^{(j-i)/2}\phi_{ij}, \quad i, j = 1, \dots, R,$$

or

$$p_{ij} = \alpha^i\beta^j\phi_{ij}, \quad i, j = 1, \dots, R,$$

where $\phi_{ij} = \phi_{ji}$. Moreover, it may be expressed as

$$\frac{p_{ij}}{p_{ji}} = \delta^{j-i}, \quad i < j.$$

This indicates that the probability that an observation will fall in the (i, j) cell, $i < j$, is δ^{j-i} times higher than the probability that the observation falls in the (j, i) cell. Moreover, Agresti (2002, p.429) considered the ordinal quasi-symmetry (OQS) model defined by

$$p_{ij} = \begin{cases} \delta^{u_j - u_i} \psi_{ij} & \text{if } i < j, \\ \psi_{ij} & \text{if } i \geq j, \end{cases}$$

where $\psi_{ij} = \psi_{ji}$ and $u_1 \leq \dots \leq u_R$ denote the ordered scores which assigned for both the rows and columns. Note that the OQS model with integer scores $\{u_i = i\}$ is identical to the LDPS model.

Tomizawa (1991) considered a model defined by

$$p_{ij} = \begin{cases} \delta^{j-i} \gamma^{(j-i)(j+i)/2} \psi_{ij} & \text{if } i < j, \\ \psi_{ij} & \text{if } i \geq j, \end{cases}$$

where $\psi_{ij} = \psi_{ji}$. A special case of this model obtained by putting $\gamma = 1$ is the LDPS model; namely, this is an extended LDPS (ELDPS) model.

Consider now random variables U and V having a joint bivariate normal distribution with means $E(U) = \mu_1$ and $E(V) = \mu_2$, variances $\text{var}(U) = \sigma_1^2$ and $\text{var}(V) = \sigma_2^2$, and correlation $\text{cor}(U, V) = \rho$. Then the joint bivariate normal density $f(u, v)$ satisfies

$$\frac{f(u, v)}{f(v, u)} = \exp \left[-\frac{u - v}{2(1 - \rho^2)} \left\{ \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right) (u + v) - 2 \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2} \right) - \frac{2\rho(\mu_1 - \mu_2)}{\sigma_1\sigma_2} \right\} \right].$$

Agresti (1983) described the relationship between the LDPS model and the joint bivariate normal distribution as follows. When $\sigma_1^2 = \sigma_2^2$, the $f(u, v)/f(v, u)$ has the form ξ^{v-u} for some constant ξ , and hence the LDPS model may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal variances. Tomizawa (1991) described that the ELDPS model rather than the LDPS model would be appropriate if it is reasonable to assume an underlying bivariate normal distribution which does not require the equality of marginal variances.

Caussinus (1965) gave the theorem that the S model holds if and only if both the QS and MH models hold for square contingency tables. Bishop et al. (1975, p.287) and Bhapkar and Darroch (1990) gave the decompositions for the S model for three-way tables and for multi-way tables, respectively. Agresti (2002, p.429) showed that the S model holds if and only if both the OQS and MH models hold. Note that the LDPS (OQS) and ELDPS models are special cases of the QS model. Since the OQS model has restrictions stronger than the QS model, we are interested in decomposing the S model into a model with weaker restrictions instead of the MH model.

In this paper we propose the other decompositions for the S model and give some extended models for the multi-way tables. Section 2 proposes two kinds of decomposition theorems of the S model for two-way tables. Sections 3 and 4 propose the extended models corresponding to the LDPS and ELDPS models, and the generalized model further for multi-way tables, and give some decomposition theorems of the S model.

2 Decompositions of Symmetry into Ordinal Quasi-Symmetry and Marginal Equipoment

Define the monotonic function as $0 < g(1) < \dots < g(R)$ [or $0 > g(1) > \dots > g(R)$], where the function is specified. Consider the marginal mean equality (ME) model defined by

$$\mu_1 = \mu_2,$$

where $\mu_t = E(g(X_t))$. This indicates that the mean of $g(X_1)$ is equal to the mean of $g(X_2)$. We shall consider the decompositions for the S model as follows:

Theorem 2.1 The S model holds if and only if both the LDPS and ME models hold.

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

$$\log p_{ij}^* = i \log \alpha + j \log \beta + \log \phi_{ij},$$

where $\phi_{ij} = \phi_{ji}$. Let $\pi_{ij} = c^{-1} \phi_{ij}$ with $c = \sum_{i=1}^R \sum_{j=1}^R \phi_{ij}$. We note that $\sum_{i=1}^R \sum_{j=1}^R \pi_{ij} = 1$ with $0 < \pi_{ij} < 1$. Then the LDPS and ME models are expressed as

$$\log \left(\frac{p_{ij}^*}{\pi_{ij}} \right) = \log c + i \log \alpha + j \log \beta, \quad (1)$$

and

$$\mu_1^* = \mu_2^*, \quad (2)$$

where $\mu_1^* = \sum_{s=1}^R sp_{s\cdot}^* = \sum_{s=1}^R \sum_{t=1}^R sp_{st}^*$ and $\mu_2^* = \sum_{s=1}^R sp_{\cdot s}^* = \sum_{s=1}^R \sum_{t=1}^R sp_{ts}^*$. Then, we denote $\mu_1^* (= \mu_2^*)$ by μ_0 .

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

$$\mu_1 = \mu_2 = \mu_0, \quad (3)$$

where $\mu_1 = \sum_{s=1}^R sp_{s\cdot} = \sum_{s=1}^R \sum_{t=1}^R sp_{st}$ and $\mu_2 = \sum_{s=1}^R sp_{\cdot s} = \sum_{s=1}^R \sum_{t=1}^R sp_{ts}$.

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

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

Let

$$K(p, \pi) = \sum_{i=1}^R \sum_{j=1}^R p_{ij} \log \left(\frac{p_{ij}}{\pi_{ij}} \right),$$

and

$$K(p^*, \pi) = \sum_{i=1}^R \sum_{j=1}^R p_{ij}^* \log \left(\frac{p_{ij}^*}{\pi_{ij}} \right).$$

Note that $K(\cdot, \cdot)$ is the Kullback-Leibler information. From (4) we obtain

$$K(p, \pi) = K(p^*, \pi) + K(p, p^*),$$

where

$$K(p, p^*) = \sum_{i=1}^R \sum_{j=1}^R p_{ij} \log \left(\frac{p_{ij}}{p_{ij}^*} \right).$$

Since π is fixed, we see

$$\min_p K(p, \pi) = K(p^*, \pi),$$

and then $\{p_{ij}^*\}$ uniquely minimizes $K(p, \pi)$ (see Darroch and Ratciff (1972); Darroch and Speed (1983); Bhapkar and Darroch (1990)).

Let $p_{ij}^{**} = p_{ji}^*$ for $1 \leq i, j \leq R$. Then, noting that $\{\pi_{ij} = \pi_{ji}\}$, we obtain

$$\min_p K(p, \pi) = K(p^{**}, \pi),$$

where

$$K(p^{**}, \pi) = \sum_{i=1}^R \sum_{j=1}^R p_{ij}^{**} \log \left(\frac{p_{ij}^{**}}{\pi_{ij}} \right),$$

and then $\{p_{ij}^{**}\}$ uniquely minimizes $K(p, \pi)$. Therefore, we see $p_{ij}^* = p_{ij}^{**}$. Thus, $p_{ij}^* = p_{ji}^*$ for $1 \leq i, j \leq R$. Namely the S model holds. The proof is completed. \square

Next, consider the marginal variance equality (VE) model defined by

$$\sigma_1^2 = \sigma_2^2,$$

where $\sigma_t^2 = \text{var}(g(X_t))$. This indicates that the variance of $g(X_1)$ is equal to the variance of $g(X_2)$. We shall consider the other decomposition for the S model as follows.

Theorem 2.2 The S model holds if and only if all the ELDPS, ME and VE models hold.

The proof is omitted because it is obtained in a similar way to the proof of Theorem 2.1. Theorems 2.1 and 2.2 may be useful for seeing the reason for the poor fit when the S model fits the data poorly.

3 Extension to Three-way Tables

We shall extend the LDPS and ELDPS models to three-way tables and consider a generalized model. Furthermore we shall give the some decomposition theorems of the S model for three-way tables.

3.1 Models

For an $R \times R \times R$ contingency table, let X_1 , X_2 , and X_3 denote the first, second, and third variable, respectively, and let p_{ijk} denote the probability that an observation will fall in the (i, j, k) cell of the table for $1 \leq i, j, k \leq R$. The symmetry model is defined by

$$p_{ijk} = \psi_{ijk}, \quad 1 \leq i, j, k \leq R,$$

where $\psi_{ijk} = \psi_{ikj} = \psi_{jik} = \psi_{jki} = \psi_{kij} = \psi_{kji}$ (Bishop et al., 1975, p.301). We shall denote this model by S-3.

First, consider a model defined by

$$p_{ijk} = \alpha_1^i \alpha_2^j \alpha_3^k \psi_{ijk}, \quad 1 \leq i, j, k \leq R,$$

where $\psi_{ijk} = \psi_{ikj} = \psi_{jik} = \psi_{jki} = \psi_{kij} = \psi_{kji}$. Without loss of generality we may set, e.g., $\alpha_3 = 1$. This model may be also expressed as

$$p_{ijk}/p_{lmn} = \alpha_1^{i-l} \alpha_2^{j-m} \alpha_3^{k-n},$$

where (l, m, n) is any permutation of (i, j, k) . It is easily seen that this model is an extension of the LDPS model to three-way tables. We shall denote this model by LDPS-3. For example, when X_3 is constant, $p_{ijk}/p_{jik} = (\alpha_2/\alpha_1)^{j-i}$, namely, the more the difference between X_1 and X_2 is large, the more the LDPS-3 model shifts from symmetry greatly exponentially.

Consider now three variables U , V and W having a joint normal distribution with means $E(U) = \mu_1$, $E(V) = \mu_2$, $E(W) = \mu_3$, variances $\text{var}(U) = \text{var}(V) = \text{var}(W) = \sigma^2$, and correlations $\text{cor}(U, V) = \text{cor}(U, W) = \text{cor}(V, W) = \rho$. Denote the probability density function by $f(u, v, w)$. Then the $f(u, v, w)/f(w, u, v)$ has the form $\xi_1^{w-u} \xi_2^{u-v} \xi_3^{v-w}$ for some constants ξ_1 , ξ_2 , and ξ_3 . Hence if it is reasonable to assume this underlying three-variate normal distribution, the LDPS-3 model may be appropriate for an ordinal three-way table (see Section 7).

Secondly, consider a model defined by

$$p_{ijk} = \alpha_1^i \alpha_2^j \alpha_3^k \beta_1^{i^2} \beta_2^{j^2} \beta_3^{k^2} \psi_{ijk}, \quad 1 \leq i, j, k \leq R,$$

where $\psi_{ijk} = \psi_{ikj} = \psi_{jik} = \psi_{jki} = \psi_{kij} = \psi_{kji}$. Without loss of generality we may set, e.g., $\alpha_3 = \beta_3 = 1$. It is easily seen that this model is an extension of the ELDPS model to three-way tables because for two-way tables this model indicates that the p_{ij}/p_{ji} has the form $\delta^{j-i} \gamma^{j^2-i^2}$ for some constants δ and γ . We shall denote this model by ELDPS-3. If it is reasonable to assume an underlying three-variate normal distribution which does not require the equality of marginal variances, then the ELDPS-3 model rather than the LDPS-3 model may be appropriate for an ordinal three-way table (see Section 7).

Finally, consider a model defined by

$$p_{ijk} = \alpha_1^i \alpha_2^j \alpha_3^k \beta_1^{i^2} \beta_2^{j^2} \beta_3^{k^2} \gamma_{12}^{ij} \gamma_{13}^{ik} \gamma_{23}^{jk} \psi_{ijk}, \quad 1 \leq i, j, k \leq R,$$

where $\psi_{ijk} = \psi_{ikj} = \psi_{jik} = \psi_{jki} = \psi_{kij} = \psi_{kji}$. Without loss of generality we may set, e.g., $\alpha_3 = \beta_3 = \gamma_{23} = 1$. We shall denote this model by GLDPS-3. A special case of this

model obtained by putting $\gamma_{12} = \gamma_{13} = \gamma_{23} = 1$ is the ELDPS-3 model; namely, this is an extension of the LDPS-3 model. If it is reasonable to assume an underlying more general three-variate normal distribution which does not require the equality of marginal variances and the equality of correlations, then the GLDPS-3 model rather than the ELDPS-3 model may be appropriate for an ordinal three-way table (see Section 7).

3.2 Decompositions for the Symmetry Model

Using the monotonic function as $0 < g(1) < \dots < g(R)$ [or $0 > g(1) > \dots > g(R)$], where this function is specified, first, consider the marginal mean equality (ME-3) model defined by

$$\mu_1 = \mu_2 = \mu_3,$$

where $\mu_t = E(g(X_t))$.

Secondly, consider the marginal variance equality (VE-3) model defined by

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2,$$

where $\sigma_t^2 = \text{var}(g(X_t))$.

Finally, consider the correlation equality (CE-3) model defined by

$$\rho_{12} = \rho_{13} = \rho_{23},$$

where ρ_{st} is the correlation between $g(X_s)$ and $g(X_t)$. We obtain the following theorems.

Theorem 3.1 The S-3 model holds if and only if both the LDPS-3 and ME-3 models hold.

Theorem 3.2 The S-3 model holds if and only if all the ELDPS-3, ME-3 and VE-3 models hold.

Theorem 3.3 The S-3 model holds if and only if all the GLDPS-3, ME-3, VE-3 and CE-3 models hold.

The proofs of these theorems are omitted because these are obtained in a similar way to the proof of Theorem 2.1.

4 Extension to Multi-Way Tables

We extend the models and decompositions in Section 3 to multi-way tables. For an R^T contingency table, let $p_{i_1 \dots i_T}$ denote the probability that an observation falls in the (i_1, \dots, i_T) cell of the table ($i_t = 1, \dots, R; t = 1, \dots, T$).

First, the symmetry model (Bhappkar and Darroch, 1990; Agresti, 2002, p.440) is defined by

$$p_{i_1 \dots i_T} = p_{j_1 \dots j_T},$$

for $i_t = 1, \dots, R; t = 1, \dots, T$, and $(j_1, \dots, j_T) \in D(i_1, \dots, i_T)$, where

$$D(i_1, \dots, i_T) = \left\{ (j_1, \dots, j_T) \mid (j_1, \dots, j_T) \text{ is any permutation of } (i_1, \dots, i_T) \right\}.$$

We shall denote this model by S - T . In particular, when $T = 3$, the S - T model is defined as

$$p_{ijk} = p_{ikj} = p_{jik} = p_{jki} = p_{kij} = p_{kji} \quad \text{for } 1 \leq i, j, k \leq R.$$

Secondly, we now consider a model defined by

$$p_{i_1 \dots i_T} = \left(\prod_{s=1}^T \alpha_s^{i_s} \right) \psi_{i_1 \dots i_T},$$

where $\psi_{i_1 \dots i_T} = \psi_{j_1 \dots j_T}$ with $(j_1, \dots, j_T) \in D(i_1, \dots, i_T)$. Note that we may set, e.g., $\alpha_T = 1$. We shall denote this model by LDPS- T .

Thirdly, we consider a model defined by

$$p_{i_1 \dots i_T} = \left(\prod_{s=1}^T \alpha_s^{i_s} \right) \left(\prod_{t=1}^T \beta_t^{i_t^2} \right) \psi_{i_1 \dots i_T},$$

where $\psi_{i_1 \dots i_T} = \psi_{j_1 \dots j_T}$ with $(j_1, \dots, j_T) \in D(i_1, \dots, i_T)$. Note that we may set, e.g., $\alpha_T = \beta_T = 1$. We shall denote this model by ELDPS- T .

Lastly, we consider a model defined by

$$p_{i_1 \dots i_T} = \left(\prod_{s=1}^T \alpha_s^{i_s} \right) \left(\prod_{t=1}^T \beta_t^{i_t^2} \right) \left(\prod_{s=1}^{T-1} \prod_{t=s+1}^T \gamma_{st}^{i_s i_t} \right) \psi_{i_1 \dots i_T},$$

where $\psi_{i_1 \dots i_T} = \psi_{j_1 \dots j_T}$, with $(j_1, \dots, j_T) \in D(i_1, \dots, i_T)$. Note that we may set, e.g., $\alpha_T = \beta_T = \gamma_{T-1, T} = 1$. We shall denote this model by GLDPS- T . When $T = 2$, this model is identical to the ELDPS model. Thus, the GLDPS- T model is defined when $T \geq 3$.

Note that Bishop et al. (1975, p.303) defined the QS model for three-way tables, and Bhapkar and Darroch (1990) defined the h th-order ($1 \leq h < T$) QS model for multi-way R^T tables (also see Agresti, 2002, p.440, for the first order QS model). We note that the LDPS- T and ELDPS- T models are special cases of the first order QS model and the GLDPS- T model is a special case of the second order QS model.

Denote the ME, VE and CE models for R^T tables by ME- T , VE- T and CE- T , respectively. Then we obtain the following decomposition theorems of the S - T model for R^T tables.

Theorem 4.1 The S - T model holds if and only if both the LDPS- T and ME- T models hold.

Theorem 4.2 The S - T model holds if and only if all the ELDPS- T , ME- T and VE- T models hold.

Theorem 4.3 The S - T model holds if and only if all the GLDPS- T , ME- T , VE- T and CE- T models hold.

The proofs of these theorems are omitted because they are obtained in similar ways to the proof of Theorem 2.1.

Table 1: Numbers of degrees of freedom (df) for models applied to the R^T table ($T \geq 2$), where the GLDPS- T model is defined when $T \geq 3$.

Models	df
S- T	$R^T - K$
LDPS- T	$R^T - K - (T - 1)$
ELDPS- T	$R^T - K - 2(T - 1)$
GLDPS- T	$R^T - K - (T^2 + 3T - 6)/2$
ME- T	$T - 1$
VE- T	$T - 1$
CE- T	$(T^2 - T - 2)/2$

$$\text{Note: } K = \binom{R+T-1}{T} = \frac{(R+T-1)!}{(R-1)!T!}$$

5 Goodness-of-Fit Test

Assume that a multinomial distribution applies to the R^T table. The maximum likelihood estimates of expected frequencies under each model could be obtained using the Newton-Raphson method to the log-likelihood equations or using the iterative procedures, for example, the general iterative procedure for log-linear models of Darroch and Ratciff (1972).

Each model can be tested for goodness-of-fit by, e.g., the likelihood ratio chi-square statistic (denoted by G^2) with the corresponding degrees of freedom (df). Note that e.g., for square tables, G^2 is

$$G^2 = 2 \sum_i \sum_j n_{ij} \log \left(\frac{n_{ij}}{\hat{m}_{ij}} \right),$$

where n_{ij} is the observed frequency in the (i, j) th cell, and \hat{m}_{ij} is the maximum likelihood estimate of expected frequency m_{ij} under the given model. The numbers of df for models are given in Table 1. Note that the number of df for the S- T model is equal to the sum of those for the decomposed models.

6 Examples

6.1 Example 1

Table 2 taken directly from Agresti (1984, p.206) is the father's and son's occupational mobility data in Britain. These data have been analyzed by some statisticians including Bishop et al. (1975, p.100), Goodman (1981, 1984), Agresti (1984, pp.205-206), and Tomizawa (1990a, 1990b, 1990c, 1991).

Table 3 gives the values of the likelihood ratio statistic G^2 for models applied to these data. The S model fits the data in Table 2 very poorly since the value of G^2 is 37.5 ($p < 0.001$) with 10 df. The LDPS model does not fit these data so well yielding $G^2 = 17.1$ ($p = 0.047$) with 9 df. However the ELDPS model fits these data well yielding $G^2 = 11.1$

Table 2: Occupational status for British father-son pairs; from Agresti (1984, p.206). The parenthesized values are the maximum likelihood estimates of expected frequencies under the ELDPS model.

Father's status	Son's status					Total
	(1)	(2)	(3)	(4)	(5)	
(1)	50 (50)	45 (35.19)	8 (9.32)	18 (16.80)	8 (6.44)	129
(2)	28 (37.81)	174 (174)	84 (82.36)	154 (165.08)	55 (58.47)	495
(3)	11 (9.68)	78 (79.64)	110 (110)	223 (218.15)	96 (99.91)	518
(4)	14 (15.20)	150 (138.92)	185 (189.85)	714 (714)	447 (430.14)	1510
(5)	3 (4.56)	42 (38.53)	72 (68.09)	320 (336.86)	411 (411)	848
Total	106	489	459	1429	1017	3500

Table 3: Likelihood ratio chi-square values G^2 for models applied to the data in Table 2.

Models	df	G^2	p -value
S	10	37.5	< 0.001
LDPS	9	17.1	0.047
ELDPS	8	11.1	0.194
ME	1	20.3	< 0.001
VE	1	0.13	0.721

($p = 0.194$) with 8 df. Using Theorems 2.1 and 2.2, we shall consider the reason why the S model fits these data poorly.

The VE model with $g(k) = k$, $k = 1, \dots, 5$, fits the data in Table 2 very well, but the ME model with $g(k) = k$ fits these data poorly (see Table 3). Therefore it is seen from Theorem 2.2 that for these data, the poor fit of the S model is caused by the influence of the poor fit of the ME model rather than the ELDPS and VE models because the ELDPS and VE models fit these data well.

6.2 Example 2

The data in Table 4 give 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 and 8 weeks of treatment. Table 4 shows the $3 \times 3 \times 3$ array of counts of Hb response that is classified as ≥ 10 g/dl, $8 - 10$ g/dl, and < 8 g/dl.

The S-3 model fits these data in Table 4 very poorly, yielding $G^2 = 76.2$ ($p < 0.001$) with 17 df (Table 5). By using the decompositions for the S-3 model, we shall consider the reason why the S-3 model fits these data poorly.

Table 4: Hemoglobin concentration at baseline, 4 weeks and 8 weeks in carcinomatous anemia patients from a randomized clinical trial. The parenthesized values are the maximum likelihood estimates of expected frequencies under the GLDPS-3 model.

Baseline	4 weeks	8 weeks		
		$\geq 10\text{g/dl}$	$8 - 10\text{g/dl}$	$< 8\text{g/dl}$
$\geq 10\text{g/dl}$	$\geq 10\text{g/dl}$	77 (77)	7 (8.32)	1 (0.33)
$8 - 10\text{g/dl}$	$\geq 10\text{g/dl}$	43 (39.68)	7 (7.53)	0 (0.12)
$< 8\text{g/dl}$	$\geq 10\text{g/dl}$	3 (4.64)	0 (0.37)	0 (0.02)
$\geq 10\text{g/dl}$	$8 - 10\text{g/dl}$	3 (5.00)	8 (7.55)	1 (0.99)
$8 - 10\text{g/dl}$	$8 - 10\text{g/dl}$	17 (16.92)	16 (16)	5 (4.75)
$< 8\text{g/dl}$	$8 - 10\text{g/dl}$	3 (3.11)	8 (6.68)	1 (0.99)
$\geq 10\text{g/dl}$	$< 8\text{g/dl}$	1 (0.04)	1 (0.20)	1 (0.58)
$8 - 10\text{g/dl}$	$< 8\text{g/dl}$	0 (0.21)	2 (3.56)	3 (4.22)
$< 8\text{g/dl}$	$< 8\text{g/dl}$	0 (0.40)	4 (2.79)	3 (3)

Table 5: Likelihood ratio chi-square values G^2 for models applied to the data in Table 4.

Models	df	G^2	p -value
S-3	17	76.2	< 0.001
LDPS-3	15	41.6	< 0.001
ELDPS-3	13	35.5	< 0.001
GLDPS-3	11	13.7	0.252
ME-3	2	23.8	< 0.001
VE-3	2	1.72	0.424
CE-3	2	22.7	< 0.001

Each of the GLDPS-3 and VE-3 models with $g(k) = k$ fits the data in Table 4 very well, but the LDPS-3, ELDPS-3, ME-3 and CE-3 models fit these data poorly (see Table 5). From Theorem 3.2, the poor fit of the S-3 model is caused by the influence of the poor fits of both the ELDPS-3 and ME-3 models with $g(k) = k$ (rather than the VE-3 model). Also, from Theorem 3.3, the poor fit of the S-3 model is caused by the influence of the poor fits of both the ME-3 and CE-3 models with $g(k) = k$ (rather than the GLDPS-3 and VE-3 models).

Table 6: Stationary two-step transitions in a panel study of potential voters in Erie County, Ohio, 1940 (from Bishop et al., 1975, p.305).

	Time $t - 2$	Time $t - 1$	Time t		
			R	U	D
R	R	R	557	16	6
U	R	R	71	11	1
D	R	R	18	5	0
R	U	U	17	21	5
U	U	U	62	346	54
D	U	U	4	24	10
R	D	D	3	0	8
U	D	D	6	6	63
D	D	D	9	22	435

(R: Republican, U: Undecided, D: Democrat)

Table 7: Likelihood ratio chi-square values G^2 for models applied to the data in Table 6.

Models	df	G^2	p -value
S-3	17	229.8	< 0.001
LDPS-3	15	221.2	< 0.001
ELDPS-3	13	110.3	< 0.001
GLDPS-3	11	43.1	< 0.001
ME-3	2	6.58	0.037
VE-3	2	73.2	< 0.001
CE-3	2	70.8	< 0.001

6.3 Example 3

The data in Table 6, taken directly from Bishop et al. (1975, p.305), give the $3 \times 3 \times 3$ array of counts of stationary two-step transitions in the panel survey of potential voters in Erie County, Ohio, 1940, which summarize the voting intentions of the 1940 presidential elections. Although the voter's supportive political party was classified into Republican, Democrat, and Undecided, we regard the voters with 'Undecided' as the middle class which could not decide Republican or Democrat, and give an order like Republican, Undecided, and Democrat.

The S-3 model fits these data poorly, yielding $G^2 = 229.8$ with 17 df (Table 7). By using the decompositions for the S-3 model, we shall consider the reason why the S-3 model fits these data poorly.

The ME-3 model does not fit the data in Table 6 very well since the value of G^2 is 6.58 ($p < 0.05$) with 2 df, but it fits much better than any other models (Table 7). In terms of the various decompositions theorems, we can see that the poor fit of the S-3 model may be caused by the influence of the more poor fits of the other models rather than the ME-3 model.

7 Concluding Remarks

Consider the $R \times R$ table. As described in Section 1, Caussinus (1965) gave the theorem that the S model holds if and only if both the QS and MH models hold. Agresti (2002, p.430) described that the S model holds if and only if both the OQS and MH models hold for ordered categories. The OQS model is a special case of the QS model and has restrictions stronger than the QS model. Therefore it seems natural to decompose the S model into the ME model (instead of the MH model) which has restrictions weaker than the MH model. Especially, the decompositions for the S model into the ME or VE model may be useful if it is reasonable to assume an underlying bivariate normal distribution when the S model fits poorly.

We assume that there is an underlying three-dimensional normal distribution and that an $R \times R \times R$ table is formed using cutpoints for each variable. Consider random variables U , V and W having a joint three-dimensional normal distribution with means $E(U) = \mu_1$, $E(V) = \mu_2$, and $E(W) = \mu_3$, variances $\text{var}(U) = \sigma_1^2$, $\text{var}(V) = \sigma_2^2$, and $\text{var}(W) = \sigma_3^2$, and correlations $\text{cor}(U, V) = \rho_{12}$, $\text{cor}(U, W) = \rho_{13}$, and $\text{cor}(V, W) = \rho_{23}$. Table 8 shows the likelihood ratio chi-square values for models applied to the simulated $4 \times 4 \times 4$ tables of sample size 10000, formed by using cutpoints for each variable at μ_2 , $\mu_1 \pm 0.6\sigma_1$, for an underlying three-dimensional normal distribution with the conditions (a) mean vector $(\mu_1, \mu_2, \mu_3) = (0, 0, 0)$, variance vector $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1, 1)$, and correlation vector $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.2, 0.2, 0.2)$, (b) mean vector $(-0.1, 0, 0.1)$, variance vector $(1, 1, 1)$, and correlation vector $(0.2, 0.2, 0.2)$, (c) mean vector $(-0.1, 0, 0.1)$, variance vector $(1, 1.2, 1.4)$, and correlation vector $(0.2, 0.2, 0.2)$, and (d) mean vector $(-0.1, 0, 0.1)$, variance vector $(1, 1.2, 1.4)$, and correlation vector $(0.4, 0.2, 0.3)$. We can see from Table 8 that (i) all models fit well if there is an underlying three-dimensional normal distribution with equal marginal means, variances and correlations (Table 8a), (ii) the S-3 and ME-3 models fit very poorly if there is an underlying three-dimensional normal distribution with *different* marginal means (Table 8b), (iii) the S-3, LDPS-3, ME-3, and VE-3 models fit very poorly if with *different* marginal means and variances (Table 8c), and (iv) the GLDPS-3 only fits well if with *different* marginal means, variances, and correlations (Table 8d).

Generally, the LDPS- T model may be appropriate for an ordinal table if it is reasonable to assume an underlying T -dimensional normal distribution with equal marginal variances and correlations, the ELDPS- T model (rather than the LDPS- T model) may be appropriate if it is reasonable to assume an underlying normal distribution which does not require the equality of marginal variances but requires the equality of correlations, and the GLDPS- T model (rather than the ELDPS- T model) may be appropriate if it is reasonable to assume an underlying normal distribution which does not require the equality of marginal variances and that of correlations.

Agresti (2002, p.440) proposed the OQS model (the first order QS model) with ordered scores and described that the S- T model holds if and only if both the OQS and (the first order) MH models hold for R^T tables (also see Bhapkar and Darroch, 1990). The OQS model with integer scores $\{u_i = i\}$ is identical to the LDPS- T model for R^T tables, but we proposed some models which are extensions of the LDPS- T model and gave the decomposition theorems of the S- T model in this paper. As seen in Examples, the

Table 8: The values of likelihood ratio statistic G^2 for $4 \times 4 \times 4$ tables of sample size 10000, formed by using cutpoints for each variable at μ_2 , $\mu_1 \pm 0.6\sigma_1$, from different underlying three-variate normal distributions.

(a) $(\mu_1, \mu_2, \mu_3) = (0, 0, 0)$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1, 1)$, $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.2, 0.2, 0.2)$			
Models	df	G^2	p -value
S-3	44	43.5	0.493
LDPS-3	42	41.4	0.497
ELDPS-3	40	40.1	0.465
GLDPS-3	38	38.5	0.449
ME-3	2	2.09	0.352
VE-3	2	1.33	0.515
CE-3	2	1.80	0.406

(b) $(\mu_1, \mu_2, \mu_3) = (-0.1, 0, 0.1)$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1, 1)$, $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.2, 0.2, 0.2)$			
Models	df	G^2	p -value
S-3	44	221.1	< 0.001
LDPS-3	42	33.2	0.833
ELDPS-3	40	31.8	0.820
GLDPS-3	38	31.1	0.780
ME-3	2	186.6	< 0.001
VE-3	2	1.13	0.569
CE-3	2	1.11	0.575

(c) $(\mu_1, \mu_2, \mu_3) = (-0.1, 0, 0.1)$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1.2, 1.4)$, $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.2, 0.2, 0.2)$			
Models	df	G^2	p -value
S-3	44	257.9	< 0.001
LDPS-3	42	100.1	< 0.001
ELDPS-3	40	37.7	0.573
GLDPS-3	38	36.1	0.560
ME-3	2	157.3	< 0.001
VE-3	2	66.6	< 0.001
CE-3	2	0.92	0.633

(d) $(\mu_1, \mu_2, \mu_3) = (-0.1, 0, 0.1)$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 1.2, 1.4)$, $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.4, 0.2, 0.3)$			
Models	df	G^2	p -value
S-3	44	509.4	< 0.001
LDPS-3	42	342.3	< 0.001
ELDPS-3	40	275.6	< 0.001
GLDPS-3	38	46.2	0.169
ME-3	2	149.9	< 0.001
VE-3	2	71.2	< 0.001
CE-3	2	225.4	< 0.001

decompositions for the S-T model would be useful for seeing the reason for the poor fit when the S-T model fits the data poorly. Moreover, the decomposition for the S-T model into more (three or four) models rather than into two models would be useful for seeing in more details the reason for the poor fit when the S-T model fits the data poorly.

Because the S model can be decomposed in at least two ways, one may be interested in which decomposition should one apply. For square tables, from Theorems 2.1 and 2.2, the S model is decomposed into (1) the LDPS and ME models and (2) the ELDPs, ME, and VE models. However, the LDPS model is not equivalent to the ELDPs and VE models holding simultaneously. Therefore both decompositions should be applied for analyzing the data.

It may seem to readers that in Examples the decomposed model (e.g., the LDPS and ME models) are tested after the S model is rejected, and the test of the S model can therefore be seen as a preliminary test. However the decomposed models should be applied even if the S model is accepted. Assuming that the LDPS model holds true, the hypothesis that the S model holds, i.e., $\delta = 1$ in the LDPS model, can be tested by the difference between the G^2 values for the S and LDPS models. Even if the S model fits the data well, the structure of complete symmetry may not exist for the data. For the ordinal data, then we are also interested in seeing the structure of asymmetry, e.g., the structure of the LDPS model. The estimate of parameter δ in the LDPS model would be useful for making inferences such as that X_1 is stochastically less than X_2 or vice versa according as the estimated δ is greater than 1 (or less than 1). So, for the ordinal data, the LDPS model would be useful even when the S model fits the data well. The ME and VE models would be useful for seeing the structure of the marginal distributions.

It also may seem that the decision procedure consists of a sequence of likelihood ratio tests, and these might be a simultaneous testing problem. However, when we want to see which model of the decomposed models has the more poor fit (e.g., by p -values), we would not need the adjustment of the individual significance levels. If we want to judge whether or not the S model holds by judging whether or not each of decomposed models holds at the given significance level, we had better adjust the individual significance level.

Acknowledgements

The authors would like to thank a referee 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.
- 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: MIT Press.

- Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 43, 572-574.
- 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 McCloud, P. I. (1986). Category distinguishability and observer agreement. *Australian Journal of Statistics*, 28, 371-388.
- Darroch, J. N., and Ratciff, D. (1972). Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics*, 43, 1470-1480.
- Darroch, J. N., and Speed, T. P. (1983). Additive and multiplicative models and interactions. *Annals of Statistics*, 11, 724-738.
- Goodman, L. A. (1979a). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74, 537-552.
- Goodman, L. A. (1979b). Multiplicative models for square contingency tables with ordered categories. *Biometrika*, 66, 413-418.
- Goodman, L. A. (1981). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, 68, 347-355.
- Goodman, L. A. (1984). *The Analysis of Cross-Classified Data Having Ordered Categories*. Cambridge: Harvard University Press.
- Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 42, 412-416.
- Tomizawa, S. (1990a). Another linear diagonals-parameter symmetry model for square contingency tables with ordered categories. *South African Statistical Journal*, 24, 117-125.
- Tomizawa, S. (1990b). The decomposition of a quasi-uniform association model and its application to the analysis of British occupational mobility table data. *Behaviormetrika*, 27, 115-121.
- Tomizawa, S. (1990c). Polynomial diagonals-parameter symmetry model for square contingency tables with ordered categories. *Statistica*, 50, 171-178.
- Tomizawa, S. (1991). An extended linear diagonals-parameter symmetry model for square contingency tables with ordered categories. *Metron*, 49, 401-409.

Authors' Address:

Hideharu Yamamoto, Toshiya Iwashita, and Sadao Tomizawa
Department of Information Sciences
Sciences Faculty of Science and Technology
Tokyo University of Science Japan
Noda City
Chiba
278-8510 Japan

E-mail: yamamotohdh@chugai-pharm.co.jp and tomizawa@is.noda.tus.ac.jp