Outline of Thesis

Modeling and Decompositions of Symmetry for Multi-Way Contingency Tables

(多元分割表における対称性のモデリングと分解)

January, 2020

Takuya Yoshimoto

Contents

Acknowledgments

I greatly appreciate all those who supported me in completing this research. First of all, I would like to express my sincerest gratitude to Professor Sadao Tomizawa, for his research guidance, outstanding advice and encouragement. I would also like to thank Associate Professor Kouji Tahata for his encouragement and mentorship, which was essential to this research. In addition, I am grateful to Assistant Professor Kiyotaka Iki at Nihon University, Dr. Yusuke Saigusa at Yokohama City University for their helpful comments and support.

I would like to appreciate my doctoral committee members, Professor Sadao Tomizawa, Professor Takashi Seo, Professor Makiko Tanaka, Professor Nobuko Miyamoto, and Associate Professor Kouji Tahata, for their time and helpful comments.

I wish to thank my fellow graduate student, Mr. Satoru Shinoda of Taisho Pharmaceutical Co., Ltd., for his encouragement and support. I also wish to express my appreciation to superiors and colleagues of Chugai Pharmaceutical Co., Ltd. for providing me with the research opportunity, and for their encouragement.

Finally, I am thankful to my family for continuing encouragement and support.

Tokyo University of Science — January, 2020

Takuya Yoshimoto

1 Background

Categorical data analysis plays an important role in many fields, from medical and pharmaceutical sciences to social sciences. The need to evaluate data generated in research stimulated statistical methods and techniques for categorical data analysis. A *categorical variable* has a measurement scale consisting of a set of categories. For instance, breast cancer diagnoses based on a mammography use categories of "normal", "benign", "probably benign", "suspicious", and "malignant". Similarly, unaided distance vision is measured as "Best", "Second", "Third", and "Worst". According to the measurement scale, categorical variables are distinguished as *nominal* or *ordinal*. Categories that cannot be ordered in any aspect are nominal. Examples include the presence of a lung cancer with the categories of "yes" and "no" and the geographic region measured as "Asia", "Europe", "North America", and "South America". On the other hand, the categories of with ordinal variables exhibit a natural ordering. An example is an opinion scale with categories of "too little", "about right", and "too much". The variables of interest influences the applicable analytical method. Methods for nominal variables can be used for ordinal variables. In contrast, methods for ordinal variables require ordered categories. Namely, methods designed for ordinal variables do not give the same results under arbitrary permutations of categories, whereas methods designed for nominal variables give the same results under arbitrary permutation of categories. Therefore, methods for ordinal variables should not be applied to data of nominal variables.

Let X_1 and X_2 denote two categorical variables, where X_1 has r levels and X_2 has c levels. rc indicates the possible combinations of outcomes in a rectangular table with r rows for the categories of X_1 and c columns of the categories of X_2. The cells represent the rc possible outcomes. A table in this form where the cells contain frequency counts of outcomes is called a *contingency table*. A contingency table that cross classifies two variables is called a two-way contingency table. A two-way contingency table with r rows and c columns is called an $r \times c$ contingency table. In addition, when T kinds of cross-classifications are constructed from r_i $(i = 1, \ldots, T)$ categories, an $r_1 \times \cdots \times r_T$ contingency table is obtained. This is known as a multi-way (T-way) contingency table.

Table 1.1, which is adapted from Agresti (2013, p.42), shows a two-way contingency table from one of the first studies to link lung cancer and smoking. In this study, patients admitted with lung cancer in the preceding year were queried about their smoking behavior. For each patients, they also recorded the smoking behavior of a noncancer patient of the same gender and similar age (within a five-year grouping) admitted to the same hospital. The *cases* in the first column shows the patients with lung cancer and the *controls* in the second column shows the patients without lung cancer, respectively. For the data in Table 1.1, we are interested in whether the presence of lung cancer is independent of the smoking behavior. If they are independent, the probability that patients with lung cancer (case) and those without lung cancer (control) will fall in the "yes" would be equal.

As a special case of a two-way contingency table, consider a contingency table with the same row and column classifications. Such a table is called a *square contingency table*. A unique characteristic of a square contingency table is that many observations tend to fall in (or near) the main diagonal cells. Table 1.2, which is adapted from Tomizawa (1985), is constructed from the data about unaided distance vision of students aged 18 to about 25 including about 10% women in the Faculty of Science and Technology, Science University of Tokyo in Japan examined in April 1982. The row and column variables are the right and left eye grades, respectively, where the categories are ordered from Best grade (1) to Worst grade (4). For these data, the independence between the row and column is unlikely to hold because many observations fall in the main diagonal cells which indicate that the value of row category is the same as the value of column category. Therefore, instead of evaluating independence, we are interested in whether the right eye grades is symmetric with the left eye grades. Table 1.3 shows the data constructed from the 2016 General Social Survey (Smith *et al.*, 2018) conducted by the National Opinion Research Center at the University of Chicago. These describe the cross classifications of subject's opinions regarding government spending on Education, Environment and Assistance to the poor in 2016. The common response categories are (1) "too little", (2) "about right" and (3) "too much". Table 1.3 is a $3 \times 3 \times 3$ contingency table with the same classifications. Similar to Table 1.2, we are also interested in the symmetry structure of the opinions among items.

Data analysis provides the observed frequencies. However, the probability distribution of the observed frequencies is unknown. The purpose of analyzing a contingency table is to estimate the structure of the unknown probability distribution with a high confidence level from the observed frequencies and to assess the association among categorical variables. Namely, it is important to interpret for the data and propose models that fit the data well. Furthermore, when a model fits the data poorly, it is also important to visualize the reason.

This research deals with methods to analyze contingency tables with the same classifications for multi-way contingency tables. As described above, in analyzing such data it is natural to deal with statistically dependent for categorical response. Consequently, model with various types of symmetry instead of statistically independent models are often used for data tables with the same classifications.

For two-way contingency tables, many models have been proposed, including the symmetry model (Bowker, 1948), quasi-symmetry model (Caussinus, 1965), marginal homogeneity model (Stuart, 1955), marginal cumulative logistic model (McCullagh, 1977), point-symmetry model (Wall and Lienert, 1976), quasi point-symmetry model (Tomizawa, 1985) and marginal point-symmetry model (Tomizawa, 1985). Models for multi-way contingency tables include the complete symmetry model (Bhapkar and Darroch, 1990), quasi-symmetry model (Bhapkar and Darroch, 1990), marginal symmetry model (Bhapkar and Darroch, 1990), marginal cumulative logistic model (Agresti, 2013, p.442), point-symmetry model (Wall and Lienert, 1976), quasi point-symmetry model (Tahata and Tomizawa, 2008) and marginal point-symmetry model (Tahata and Tomizawa, 2008).

Moreover, Agresti (2013, p.440) discussed the decomposition of model. That is, generally suppose that model H_3 implies models H_1 and H_2, model H_3 holds if and only if both models H_1 and H_2 hold. This enables us to see that assuming that model H_1 holds true, the hypothesis that model H_3 holds is equivalent to the hypothesis that the model H_2 holds. For analyzing the data, the decomposition of model should be useful to observe the reason for its poor fit when model H_3 does not fit the data well. Caussinus (1965) and Bhapkar and Darroch (1990) gave a decomposition of the symmetry model for two-way, and multi-way contingency tables, respectively.

This research considers models with a symmetry structure and discusses decompositions for multi-way contingency tables. The aims of this research are (i) to consider a model with an inhomogeneity structure of marginal distribution for general order, and discuss a decomposition of model, and (ii) to consider a generalized model, and consider its property in the information theoretic sense, and discuss a decomposition of model, and (iii) to consider a model with a moment symmetry and homogeneity structure for general order, and present a decomposition of model.

2 Preliminaries

Preliminary material for analysis of contingency table is briefly presented. For an square contingency tables and multi-way contingency tables, various symmetry models, usually applied for contingency table, are introduced. Furthermore, hypothesis testing and model selection are also discussed.

2.1 Joint and Marginal Probability

First, we present the probability for two-way contingency tables. Let X_1 and X_2 denote the row and column variables, respectively, and let $p_{ij} = \Pr(X_1 = i, X_2 = j)$ denote the probability that an observation will fall in the ith row and jth column of the table ($i = 1, \ldots, r;\ j = 1, \ldots, c$). The probability $\{p_{ij}\}$ is the joint distribution of X_1 and X_2. These satisfy $\sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij} = 1$. Let $p_{i\cdot} = \Pr(X_1 = i) = \sum_{t=1}^{c} p_{it}$ and $p_{\cdot j} = \Pr(X_2 = j) = \sum_{s=1}^{r} p_{sj}$. Then $\{p_{i\cdot}\}$ is the marginal distribution of X_1 and $\{p_{\cdot j}\}$ is the marginal distribution of X_2.

Second, we present the probability for three-way contingency tables. For the $r_1 \times r_2 \times r_3$ contingency table, let X_1, X_2, and X_3 denote the first, second, and third variables, respectively, and let $\Pr(X_1 = i, X_2 = j, X_3 = k) = p_{ijk}$, where $i = 1, \ldots, r_1;\ j = 1, \ldots, r_2;\ k = 1, \ldots, r_3$. Let $p_i^{(1)} = p_{i\cdot\cdot} = \Pr(X_1 = i)$, $p_j^{(2)} = p_{\cdot j\cdot} = \Pr(X_2 = j)$ and $p_k^{(3)} = p_{\cdot\cdot k} = \Pr(X_3 = k)$ denote the first-order marginal probability. Note that "$\cdot$" denotes the sum; thus $p_{i\cdot\cdot} = \sum_s \sum_t p_{ist}$, etc. Let $p_{ij}^{(1,2)} = p_{ij\cdot} = \Pr(X_1 = i, X_2 = j)$, $p_{ik}^{(1,3)} = p_{i\cdot k} = \Pr(X_1 = i, X_3 = k)$ and $p_{jk}^{(2,3)} = p_{\cdot jk} = \Pr(X_2 = j, X_3 = k)$ denote the second-order marginal probability.

Finally, we present the probability for multi-way contingency tables. For the $r_1 \times \cdots \times r_T$ contingency table ($T \geq 2$), let $\boldsymbol{i} = (i_1, \ldots, i_T)$ for $i_k = 1, \ldots, r_k$ ($k = 1, \ldots, T$), and let $p_{\boldsymbol{i}}$ denote the probability that an observation will fall in the $\boldsymbol{i}$th cell of the table. In addition, let X_k ($k = 1, \ldots, T$) denote the kth variable. Denote the hth-order ($h = 1, \ldots, T-1$) marginal probability $\Pr(X_{s_1} = i_1, \ldots, X_{s_h} = i_h)$ by $p_{\boldsymbol{i}}^{\boldsymbol{s}_h}$, where $\boldsymbol{s}_h = (s_1, \ldots, s_h)$ and $\boldsymbol{i} = (i_1, \ldots, i_h)$ with $1 \leq s_1 < \cdots < s_h \leq T$ and $i_k = 1, \ldots, r_{s_k}$ ($k = 1, \ldots, h$).

2.2 Various symmetry models

Consider the case of $r_1 = \cdots = r_T$ since each classifications are the same.

(a) Case of $r \times r$ tables

Consider an $r \times r$ square contingency table with the same classifications. The symmetry (S) model is defined by

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

where $\psi_{ij} = \psi_{ji}$ (Bowker, 1948). This indicates the probability that an observation will fall in row category i and column category j is equal to the probability that the observation falls in row category j and column category i. The S model is described by many statisticians; see, for example, Bishop, Fienberg and Holland (1975, p.282), Agresti (2013, p.426) and Kateri (2014, p.236). Also, Bishop, *et al.* (1975, p.282) expressed the S model in terms of a log-linear model as follows:

$$\log p_{ij} = u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad (i = 1, \dots, r;\ j = 1, \dots, r),$$

where $u_{1(i)} = u_{2(j)}$ and $u_{12(ij)} = u_{12(ji)}$.

The quasi-symmetry (QS) model is defined by

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

where $\psi_{ij} = \psi_{ji}$ (Caussinus, 1965). A special case of this model with $\{\alpha_i = \beta_i\}$ is the S model. Denote the odds ratio for row i and j $(> i)$, and column s and t $(> s)$ by $\theta_{(i<j;s<t)}$. Namely $\theta_{(i<j;s<t)} = (p_{is}p_{jt})/(p_{js}p_{it})$. Using the odds ratios, the QS model is further expressed as

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

Therefore the QS model has characterization in terms of symmetry of odds ratios. Also, Bishop, *et al.* (1975, p.286) expressed the QS model in terms of a log-linear model as follows:

$$\log p_{ij} = u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad (i = 1, \dots, r;\ j = 1, \dots, r),$$

where $u_{12(ij)} = u_{12(ji)}$. For details, see, Goodman (1979), Darroch and McCloud (1986), Agresti (2013, p.427) and Kateri (2014, p.238).

The marginal homogeneity (M) model is defined by

$$p_{i\cdot} = p_{\cdot i} \quad (i = 1, \dots, r),$$

where $p_{i\cdot} = \sum_{t=1}^{r} p_{it}$ and $p_{\cdot i} = \sum_{s=1}^{r} p_{si}$ (e.g., Stuart, 1955; Bishop, *et al.*, 1975, p.294). This model states that the row marginal distribution is identical with column marginal distribution.

Let $F_i^{(1)}$ and $F_i^{(2)}$ denotes the marginal cumulative probability, and let $L_i^{(1)}$ and $L_i^{(2)}$ denote the marginal cumulative logit of X_1 and X_2 for $i = 1, \dots, r-1$, $k = 1, 2$; namely, $F_i^{(k)} = \Pr(X_k \le i)$ and $L_i^{(k)} = \text{logit}\left(F_i^{(k)}\right) = \log\left(F_i^{(k)}/\left(1 - F_i^{(k)}\right)\right)$. The marginal cumulative logistic (ML) model is defined by

$$L_i^{(2)} = L_i^{(1)} - \Delta \quad (i = 1, \dots, r-1),$$

see McCullagh (1977). A special case of this model obtained by putting $\Delta = 0$ is the M model (also see Agresti, 2010, p.231). This model states that the odds that X_1 is i or below instead of $i+1$ or above, is $\exp(\Delta)$ times higher than the odds that X_2 is i or below instead of $i+1$ or above, for every $i = 1, \ldots, r-1$.

Caussinus (1965) gave a decomposition of the S model as follows:

Theorem 2.1. *The S model holds if and only if both the QS and M models hold.*

This theorem enables us to see that assuming that the QS model holds true, the hypothesis that the S model holds is equivalent to the hypothesis that the M model holds. For analyzing the data, the decomposition of the S model would be useful for seeing the reason for the poor fit when the S model fits the data poorly.

Refer to model of equality of marginal means, i.e., $E(X_1) = E(X_2)$, as the marginal mean equality (ME) model. Then, Miyamoto, Niibe and Tomizawa (2005) gave a decomposition of the M model as follows:

Theorem 2.2. *The M model holds if and only if both the ML and ME models hold.*

The point-symmetry (PS) model is defined by

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

where $\psi_{ij} = \psi_{i^*j^*}$ and $i^* = r + 1 - i$ (Wall and Lienert, 1976; Tomizawa, 1985). This model states that the probability that an observation will fall in cell (i, j) is equal to the probability that it falls in point symmetric cell (i^*, j^*) with respect to the center point (or cell). The PS model is also expressed in log-linear form

$$\log p_{ij} = u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $u_{1(i)} = u_{1(i^*)}$, $u_{2(j)} = u_{2(j^*)}$ and $u_{12(ij)} = u_{12(i^*j^*)}$.

Tomizawa (1985) considered the quasi point-symmetry (QP) model defined by

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

where $\psi_{ij} = \psi_{i^*j^*}$. A special case of the QP model obtained by putting $\left\{\alpha_{(i)} = \alpha_{(i^*)}\right\}$ and $\left\{\beta_{(j)} = \beta_{(j^*)}\right\}$ is the PS model. Using odds ratios, the QP model is also expressed as

$$\theta_{(i<j;s<t)} = \theta_{(j^*<i^*;t^*<s^*)} \quad (i < j;\ s < t).$$

Therefore the QP model has its characterization in terms of point-symmetry of odds ratios. The QP model is also expressed in log-linear form

$$\log p_{ij} = u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $u_{12(ij)} = u_{12(i^*j^*)}$.

The marginal point-symmetry (MP) model is defined by

$$p_{i\cdot} = p_{i^*\cdot} \quad (i = 1, \dots, r) \quad \text{and} \quad p_{\cdot j} = p_{\cdot j^*} \quad (j = 1, \dots, r).$$

This indicates that the row (column) marginal distributions are point symmetric with respect to the midpoint of the row (column) categories.

Tomizawa (1985) gave a decomposition of the PS model as follows:

Theorem 2.3. *The PS model holds if and only if both the QP and MP models hold.*

Let $p = (p_{ij})$ and $q = (q_{ij})$ be two discrete finite bivariate probability distributions. The f-divergence between p and q is defined as

$$I^C(p : q) = \sum_i \sum_j q_{ij} f\left(\frac{p_{ij}}{q_{ij}}\right),$$

where f is a convex function on $(0, +\infty)$ with $f(1) = 0$. Also, we take $f(0) = \lim_{u\to 0} f(u)$, $0f(0/0) = 0$, and $0f(a/0) = a \lim_{u\to\infty} [f(u)/u]$ (Csiszár and Shields, 2004). Kateri and Papaioannou (1997), Kateri and Agresti (2007), and Saigusa, Tahata and Tomizawa (2015) have described some models of symmetry based on f-divergence, and considered the property of the model in the information theoretic sense. Also see, for example, Gilula, Krieger and Ritov (1988), and Kateri (2018).

(b) Case of r^T tables

Consider an r^T contingency table ($T \geq 2$) with the same classifications. The complete symmetry (S^T) model is defined by

$$p_{\boldsymbol{i}} = p_{\boldsymbol{j}},$$

for any permutation $\boldsymbol{j} = (j_1, \dots, j_T)$ of $\boldsymbol{i} = (i_1, \dots, i_T)$. See Bhapkar and Darroch (1990), Lovison (2000), and Agresti (2013, p.439). The S^T model may be expressed as in a log-linear form

$$\log p_{\boldsymbol{i}} = \lambda_{(\boldsymbol{i})},$$

where $\lambda_{(\boldsymbol{i})} = \lambda_{(\boldsymbol{j})}$ for any permutation $\boldsymbol{j} = (j_1, \dots, j_T)$ of $\boldsymbol{i} = (i_1, \dots, i_T)$.

Bhapkar and Darroch (1990) defined the hth-order ($h = 1, \dots, T-1$) quasi-symmetry (QS_h^T) model, which may be expressed as

$$\log p_{\boldsymbol{i}} = \sum_{k=1}^{T} \lambda_{k(i_k)} + \sum\sum_{1\leq k_1<k_2\leq T} \lambda_{k_1 k_2(i_{k_1}, i_{k_2})} + \cdots + \sum\cdots\sum_{1\leq k_1<\cdots<k_h\leq T} \lambda_{k_1\dots k_h(i_{k_1},\dots,i_{k_h})} + \lambda_{\boldsymbol{i}},$$

where $\lambda_{\boldsymbol{i}} = \lambda_{\boldsymbol{j}}$ for any permutation $\boldsymbol{j} = (j_1, \ldots, j_T)$ of $\boldsymbol{i} = (i_1, \ldots, i_T)$.

For a fixed h ($h = 1, \ldots, T-1$), the hth-order marginal symmetry (M_h^T) model is defined by

$$p_{\boldsymbol{i}}^{\boldsymbol{s}_h} = p_{\boldsymbol{j}}^{\boldsymbol{s}_h} = p_{\boldsymbol{i}}^{\boldsymbol{t}_h},$$

for any permutation $\boldsymbol{j} = (j_1, \ldots, j_h)$ of $\boldsymbol{i} = (i_1, \ldots, i_h)$ and for any $\boldsymbol{s}_h = (s_1, \ldots, s_h)$ and $\boldsymbol{t}_h = (t_1, \ldots, t_h)$ with $1 \leq t_1 < \cdots < t_h \leq T$ and $i_k = 1, \ldots, r$ ($k = 1, \ldots, h$) (Bhapkar and Darroch, 1990; Tomizawa and Tahata, 2007). This model indicates the structure of symmetry and homogeneity of hth-order marginal distribution. For the case of $h = 1$, the M_h^T model is expressed as

$$p_i^{(1)} = \cdots = p_i^{(T)} \quad (i = 1, \ldots, r),$$

where $p_i^{(k)} = \Pr(X_k = i)$. For instance, see Bishop, *et al.* (1975, p.303) and Agresti (2013, p.439). This model indicates the homogeneity structure of first-order marginal distribution.

Let $F_i^{(k)}$ denote the first-order marginal cumulative probability and let $L_i^{(k)}$ denote the first-order marginal cumulative logit of X_k for $i = 1, \ldots, r-1$, $k = 1, \ldots, T$; namely, $F_i^{(k)} = \Pr(X_k \leq i)$ and $L_i^{(k)} = \mathrm{logit}\left(F_i^{(k)}\right) = \log\left(F_i^{(k)} / \left(1 - F_i^{(k)}\right)\right)$. The marginal cumulative logistic (ML^T) model is defined by

$$L_i^{(k)} = L_i^{(1)} - \Delta_k \quad (i = 1, \ldots, r-1;\ k = 1, \ldots, T),$$

where $\Delta_1 = 0$. See e.g., Agresti (2010, p.241) and Agresti (2013, p.442). A special case of this model obtained by putting $\{\Delta_k = 0\}$ is the M_1^T model.

Bhapkar and Darroch (1990) gave the extension of Theorem 2.1 into multi-way contingency tables as follows:

Theorem 2.4. *For an r^T table and fixed h ($h = 1, \ldots, T-1$), the S^T model holds if and only if both the QS_h^T and M_h^T models hold.*

When $T = 2$, this theorem is identical to Theorem 2.1.

The marginal mean equality (ME^T) model can be considered as follows:

$$E(X_1) = \cdots = E(X_T).$$

Tahata, Katakura and Tomizawa (2007) gave a decomposition as follows:

Theorem 2.5. *For an r^T table, the M_1^T model holds if and only if both the ML^T and ME^T models hold.*

When $T = 2$, this theorem is identical to Theorem 2.2.

The point-symmetry (PS^T) model is defined by

$$p_{\boldsymbol{i}} = p_{\boldsymbol{i}^*} \quad \text{for any } \boldsymbol{i},$$

where $\boldsymbol{i}^* = (i_1^*, \ldots, i_T^*)$ for $i_k^* = r + 1 - i_k$ $(k = 1, \ldots, T)$ (Wall and Lienert, 1976).

For a fixed h $(h = 1, \ldots, T-1)$, Tahata and Tomizawa (2008) considered the hth-order quasi point-symmetry (QP_h^T) model defined by

$$\begin{cases} \log p_{\boldsymbol{i}} = \displaystyle\sum_{k=1}^{T} u_{k(i_k)} + \sum\sum_{1 \le k_1 < k_2 \le T} u_{k_1 k_2 (i_{k_1}, i_{k_2})} \\ \qquad\qquad + \cdots + \displaystyle\sum \cdots \sum_{1 \le k_1 < \cdots < k_{T-1} \le T} u_{k_1 \ldots k_{T-1} (i_{k_1}, \ldots, i_{k_{T-1}})} + u_{12\ldots T(\boldsymbol{i})} \text{ for any } \boldsymbol{i}, \\ \text{where } u_{k_1 k_2 \ldots k_l (i_{k_1}, i_{k_2}, \ldots, i_{k_l})} = u_{k_1 k_2 \ldots k_l (i_{k_1}^*, i_{k_2}^*, \ldots, i_{k_l}^*)} \\ \qquad (l = h+1, \ldots, T; 1 \le k_1 < k_2 < \cdots < k_l \le T). \end{cases}$$

Note that the PS^T model can be expressed in a log-linear form as a special case of the QP_h^T model with

$$u_{k_1 k_2 \ldots k_l (i_{k_1}, i_{k_2}, \ldots, i_{k_l})} = u_{k_1 k_2 \ldots k_l (i_{k_1}^*, i_{k_2}^*, \ldots, i_{k_l}^*)} \quad (l = 1, \ldots, T; 1 \le k_1 < k_2 < \cdots < k_l \le T).$$

For a fixed h $(h = 1, \ldots, T-1)$, Tahata and Tomizawa (2008) also considered the hth-order marginal point-symmetry (MP_h^T) model defined by

$$p_{\boldsymbol{i}}^{\boldsymbol{s}_h} = p_{\boldsymbol{i}^*}^{\boldsymbol{s}_h} \quad \text{for any } \boldsymbol{s}_h = (s_1, \ldots, s_h),$$

where $\boldsymbol{i} = (i_1, \ldots, i_h)$ and $\boldsymbol{i}^* = (i_1^*, \ldots, i_h^*)$.

Tahata and Tomizawa (2008) gave the extension of Theorem 2.3 into multi-way contingency tables as follows:

Theorem 2.6. *For an r^T table and fixed h $(h = 1, \ldots, T-1)$, the PS^T model holds if and only if both the QP_h^T and MP_h^T models hold.*

2.3 Hypothesis testing and model selection

Let $n_{i_1 \ldots i_T}$ denote the observed frequency in the $(i_1, \ldots, i_T)$th cell of the r^T table. Assume that a multinomial distribution is applied to the r^T table. The maximum likelihood estimates (MLEs) of the expected frequencies under each model can be obtained by the Newton-Raphson method in the log-likelihood equation. Each model can be tested for the goodness-of-fit using, for example, the likelihood ratio chi-squared statistic (denoted by G^2) with the corresponding degrees of freedom (df). The test statistic G^2 for model H is given by

$$G^2(H) = 2\sum_{i_1=1}^{r} \cdots \sum_{i_T=1}^{r} n_{i_1\ldots i_T} \log\left(\frac{n_{i_1\ldots i_T}}{\hat{m}_{i_1\ldots i_T}}\right),$$

where $\hat{m}_{i_1\ldots i_T}$ is the MLE of expected frequency $m_{i_1\ldots i_T}$ under model H.

We shall consider the comparison between two nested models for analyzing the data. Consider two nested models, say H_1 and H_2, such that model H_1 is a special case of model H_2, if model H_1 holds, then model H_2 also holds. Let v_1 and v_2 denote the df for models H_1 and H_2, respectively. For testing that model H_1 holds assuming that model H_2 holds true, the likelihood ratio statistics is given as $G^2(H_1 \mid H_2) = G^2(H_1) - G^2(H_2)$. Under the null hypothesis, this test statistic has an asymptotic chi-square distribution with $v_1 - v_2$ df.

Akaike (1974) information criterion (AIC) is used to select the preferable model among different models which include non-nested models. For details, see Konishi and Kitagawa (2008). Since the difference between AIC's is only required when two models are compared, it is possible to ignore a common constant of AIC. Thus, the modified AIC is defined as

$$\mathrm{AIC}^{+} = G^2 - 2(\text{number of df}).$$

For analyzing the data, the model with the minimum AIC^{+} is the preferable model.

For the analysis of contingency tables, Lang and Agresti (1994) considered the simultaneous modeling joint and marginal distribution, and Lang (1996) discussed the partitioning of goodness-of-fit statistics. Aitchison (1962) discussed the asymptotic separability of models. Also the similar property of models is described by Darroch and Silvey (1963) and Read (1977). (See also, Tahata and Tomizawa, 2008; Tomizawa and Tahata, 2007). Generally suppose that model H_3 holds if and only if both model H_1 and model H_2 hold. When the test statistic for goodness-of-fit of model H_3 is asymptotically equivalent to the sum of those for model H_1 and model H_2, if both model H_1 and model H_2 are accepted (at the α significance level) with high probability, then model H_3 would be accepted. However, when it does not hold, it is quite possible for an incompatible situation to arise where both model H_1 and model H_2 are accepted but model H_3 is rejected with high probability. Then, for Theorem 2.1, following separability of test statistic holds (Tomizawa and Tahata, 2007).

Theorem 2.7. *For an $r \times r$ table, $G^2(S)$ is asymptotically equivalent to the sum of $G^2(QS)$ and $G^2(M)$.*

Also for Theorems 2.3, 2.4 and 2.6, Tomizawa and Tahata (2007) and Tahata and Tomizawa (2008) gave similar orthogonal decompositions, although the detail is omitted.

3 Outline of Doctoral Thesis

The doctoral thesis consists of five chapters. Chapter 1 overviews the background of this research and provides preparation for Chapters 2 through 4. Chapter 2 introduces a marginal cumulative logistic model of general order and decompositions of the marginal symmetry model. This chapter is based on Yoshimoto, Tahata and Tomizawa (2019). Chapter 3 discusses a quasi point-symmetry model based on f-divergence and orthogonal decompositions of the point-symmetry model. This chapter is based on Yoshimoto, Tahata, Saigusa and Tomizawa (2019). Chapter 4 details a moment symmetry model and decompositions of the complete symmetry model and marginal symmetry model. This chapter is based on Yoshimoto, Tahata, Iki and Tomizawa (2019). Finally, Chapter 5 presents the conclusion of the doctoral thesis. We present here a brief description of the content for Chapters 2 through 4.

3.1 Chapter 2

The ML^T model focuses on the first order marginal distributions, and describes the inhomogeneity structure of first-order marginal distribution. Focusing on the hth-order marginal distribution, we are now interested in the symmetry and inhomogeneity structure of the hth-order marginal distribution.

Chapter 2 (i) proposes a marginal cumulative logistic model for order h, which is the extensions of the ML^T model, (ii) gives the decompositions of the M_h^T model by using proposed model. Chapter 2 also analyzes the data in more details. The decompositions of model may be useful for seeing in more detail the reason for the poor fit when the M_h^T model fits the data poorly. These models and the decompositions are described simply as below.

Consider an r^T table. For order h ($h = 1, \dots, T-1$), denote the hth-order marginal cumulative probability $\Pr(X_{s_1} \leq i_1, \dots, X_{s_h} \leq i_h)$ by $F_{\boldsymbol{i}}^{\boldsymbol{s}_h}$, where $\boldsymbol{s}_h = (s_1, \dots, s_h)$ and $\boldsymbol{i} = (i_1, \dots, i_h)$ with $1 \leq s_1 < \cdots < s_h \leq T$ and $i_k = 1, \dots, r$ ($k = 1, \dots, h$). Then the M_h^T model may be expressed as

$$F_{\boldsymbol{i}}^{\boldsymbol{s}_l} = F_{\boldsymbol{j}}^{\boldsymbol{s}_l} = F_{\boldsymbol{i}}^{\boldsymbol{t}_l} \quad (l = 1, \dots, h),$$

for any permutation $\boldsymbol{j} = (j_1, \dots, j_l)$ of $\boldsymbol{i} = (i_1, \dots, i_l)$, where $i_k = 1, \dots, r-1$ ($k = 1, \dots, l$)

and for any $\boldsymbol{s}_l = (s_1, \ldots, s_l)$ and $\boldsymbol{t}_l = (t_1, \ldots, t_l)$. Focusing on the hth-order marginal distribution, we consider the symmetry and inhomogeneity structure based on the logits of $\left\{F_{\boldsymbol{i}}^{\boldsymbol{s}_h}\right\}$. Then, for a fixed h $(h = 1, \ldots, T-1)$, consider a model defined by

$$L_{\boldsymbol{i}}^{\boldsymbol{s}_h} = L_{\boldsymbol{i}}^{\boldsymbol{l}_h} - \Delta_{\boldsymbol{s}_h} \quad \text{and} \quad L_{\boldsymbol{i}}^{\boldsymbol{s}_h} = L_{\boldsymbol{j}}^{\boldsymbol{s}_h},$$

for any permutation $\boldsymbol{j} = (j_1, \ldots, j_h)$ of $\boldsymbol{i} = (i_1, \ldots, i_h)$ and $\boldsymbol{l}_h = (1, \ldots, h)$, where $i_k = 1, \ldots, r-1$ $(k = 1, \ldots, h)$ and for any $\boldsymbol{s}_h = (s_1, \ldots, s_h)$ with $1 \leq s_1 < \cdots < s_h \leq T$,

$$L_{\boldsymbol{i}}^{\boldsymbol{s}_h} = \operatorname{logit}\left(F_{\boldsymbol{i}}^{\boldsymbol{s}_h}\right) = \log\left(\frac{F_{\boldsymbol{i}}^{\boldsymbol{s}_h}}{1 - F_{\boldsymbol{i}}^{\boldsymbol{s}_h}}\right),$$

where $\Delta_{\boldsymbol{l}_h} = 0$. We shall refer to this model as the hth-order marginal cumulative logistic (ML_h^T) model. When $h = 1$, the ML_h^T model is identical to the ML^T model.

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

$$E\left(X_{s_1} \cdots X_{s_h}\right) = E\left(X_1 \cdots X_h\right),$$

for $1 \leq s_1 < \cdots < s_h \leq T$. We shall refer to this model as the hth-order marginal moment equality (ME_h^T) model. Although the details and the proof are omitted, we obtain the following theorem and corollary.

Theorem 3.1. *For a fixed h $(h = 1, \ldots, T-1)$, the M_h^T model holds if and only if all the ML_h^T, ME_h^T, and M_{h-1}^T models hold, where the M_0^T model indicates the saturated model.*

Corollary 3.1. *The M_{T-1}^T model holds if and only if all the ML_h^T models for $h = 1, \ldots, T-1$ and all the ME_h^T models for $h = 1, \ldots, T-1$ hold.*

The details of the proof for theorems and examples are given in Chapter 2.

3.2 Chapter 3

Chapter 3 proposes a quasi point-symmetry (QP[f]) model based on f-divergence, and discusses the decomposition of the PS model for two-way, three-way and multi-way contingency tables, respectively. Proposed model is a generalized model in the sense that it includes the QP model. These models and the decompositions are described simply as below.

First, consider an $r \times r$ table. Let f be a twice-differentiable and strictly convex function, and let $F(u) = df(u)/du$. Let $p_{ij}^{PS} = (p_{ij} + p_{i^*j^*})/2$ for $i = 1, \ldots, r$ and $j = 1, \ldots, r$, where $i^* = r + 1 - i$. Consider the quasi point-symmetry (QP[f]) model based on f-divergence defined by

$$p_{ij} = p_{ij}^{PS} F^{-1}(u_{1(i)} + u_{2(j)} + u_{12(ij)}) \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $u_{12(ij)} = u_{12(i^*j^*)}$. The QP[$f$] model can also be expressed as

$$\theta^{[f]}_{(i<j;k<l)} = \theta^{[f]}_{(j^*<i^*;l^*<k^*)} \quad (i<j;\ k<l),$$

where

$$\theta^{[f]}_{(i<j;k<l)} = F\left(\frac{p_{ik}}{p^{PS}_{ik}}\right) + F\left(\frac{p_{jl}}{p^{PS}_{jl}}\right) - F\left(\frac{p_{jk}}{p^{PS}_{jk}}\right) - F\left(\frac{p_{il}}{p^{PS}_{il}}\right).$$

Therefore the QP[f] model has its characterization in terms of point-symmetry of $\theta^{[f]}_{(i<j;k<l)}$. The QP[f] model with $f(u) = u \log u,\ u > 0$ is the QP model. We obtain the following theorems.

Theorem 3.2. *The QP[f] model for $\{p_{ij}\}$ minimizes the f-divergence between $\{p_{ij}\}$ and $\{p^{PS}_{ij}\}$ with the structure of the PS model under the condition that row marginals $\{p_{i\cdot}\}$, column marginals $\{p_{\cdot j}\}$ and sums $\{p_{ij} + p_{i^*j^*}\}$ are given.*

Theorem 3.3. *The PS model holds if and only if both the QP[f] and MP models hold.*

Second, consider an $r \times r \times r$ contingency table. Let f be a twice-differentiable and strictly convex function, and let $F(u) = df(u)/du$. Let $p^{PS}_{ijk} = (p_{ijk} + p_{i^*j^*k^*})/2$ for $1 \le i,j,k \le r$, where $i^* = r+1-i$. First, consider the first-order quasi point-symmetry (QP[f]$^3_1$) model based on f-divergence defined by

$$\begin{cases} p_{ijk} = p^{PS}_{ijk} F^{-1}\big(u_{1(i)} + u_{2(j)} + u_{3(k)} \\ \qquad\qquad + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}\big) \quad (1 \le i,j,k \le r), \\ \text{where for } 1 \le s < t \le 3, \\ u_{st(ij)} = u_{st(i^*j^*)}, \quad u_{123(ijk)} = u_{123(i^*j^*k^*)}. \end{cases}$$

The QP[f]$^3_1$ model can also be expressed as

$$\theta^{[f]}_{(i;j_1<j_2;k_1<k_2)} = \theta^{[f]}_{(i^*;j_2^*<j_1^*;k_2^*<k_1^*)}\ (1 \le i \le r;\ 1 \le j_1 < j_2 \le r;\ 1 \le k_1 < k_2 \le r),$$

$$\theta^{[f]}_{(i_1<i_2;j;k_1<k_2)} = \theta^{[f]}_{(i_2^*<i_1^*;j^*;k_2^*<k_1^*)}\ (1 \le i_1 < i_2 \le r;\ 1 \le j \le r;\ 1 \le k_1 < k_2 \le r),$$

and

$$\theta^{[f]}_{(i_1<i_2;j_1<j_2;k)} = \theta^{[f]}_{(i_2^*<i_1^*;j_2^*<j_1^*;k^*)}\ (1 \le i_1 < i_2 \le r;\ 1 \le j_1 < j_2 \le r;\ 1 \le k \le r),$$

where

$$\theta^{[f]}_{(i;j_1<j_2;k_1<k_2)} = F\left(\frac{p_{ij_1k_1}}{p^{PS}_{ij_1k_1}}\right) + F\left(\frac{p_{ij_2k_2}}{p^{PS}_{ij_2k_2}}\right) - F\left(\frac{p_{ij_2k_1}}{p^{PS}_{ij_2k_1}}\right) - F\left(\frac{p_{ij_1k_2}}{p^{PS}_{ij_1k_2}}\right),$$

$$\theta^{[f]}_{(i_1<i_2;j;k_1<k_2)} = F\left(\frac{p_{i_1jk_1}}{p^{PS}_{i_1jk_1}}\right) + F\left(\frac{p_{i_2jk_2}}{p^{PS}_{i_2jk_2}}\right) - F\left(\frac{p_{i_2jk_1}}{p^{PS}_{i_2jk_1}}\right) - F\left(\frac{p_{i_1jk_2}}{p^{PS}_{i_1jk_2}}\right),$$

$$\theta^{[f]}_{(i_1<i_2;j_1<j_2;k)} = F\left(\frac{p_{i_1j_1k}}{p^{PS}_{i_1j_1k}}\right) + F\left(\frac{p_{i_2j_2k}}{p^{PS}_{i_2j_2k}}\right) - F\left(\frac{p_{i_2j_1k}}{p^{PS}_{i_2j_1k}}\right) - F\left(\frac{p_{i_1j_2k}}{p^{PS}_{i_1j_2k}}\right).$$

The $\mathrm{QP}[f]_1^3$ model with $f(u) = u \log u,\ u > 0$ is equivalent to the first-order quasi point-symmetry (QP_1^3) model.

Next, consider the second-order quasi point-symmetry ($\mathrm{QP}[f]_2^3$) model based on f-divergence defined by

$$\begin{cases} p_{ijk} = p^{PS}_{ijk} F^{-1}\big(u_{1(i)} + u_{2(j)} + u_{3(k)} \\ \qquad\qquad + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}\big) \quad (1 \le i, j, k \le r), \\ \text{where } u_{123(ijk)} = u_{123(i^*j^*k^*)}. \end{cases}$$

The $\mathrm{QP}[f]_2^3$ model can also be expressed as

$$\theta^{[f]}_{(i_2;j_1<j_2;k_1<k_2)} - \theta^{[f]}_{(i_1;j_1<j_2;k_1<k_2)} = \theta^{[f]}_{(i_2^*;j_2^*<j_1^*;k_2^*<k_1^*)} - \theta^{[f]}_{(i_1^*;j_2^*<j_1^*;k_2^*<k_1^*)},$$

or

$$\theta^{[f]}_{(i_1<i_2;j_2;k_1<k_2)} - \theta^{[f]}_{(i_1<i_2;j_1;k_1<k_2)} = \theta^{[f]}_{(i_2^*<i_1^*;j_2^*;k_2^*<k_1^*)} - \theta^{[f]}_{(i_2^*<i_1^*;j_1^*;k_2^*<k_1^*)},$$

or

$$\theta^{[f]}_{(i_1<i_2;j_1<j_2;k_2)} - \theta^{[f]}_{(i_1<i_2;j_1<j_2;k_1)} = \theta^{[f]}_{(i_2^*<i_1^*;j_2^*<j_1^*;k_2^*)} - \theta^{[f]}_{(i_2^*<i_1^*;j_2^*<j_1^*;k_1^*)},$$

for $1 \le i_1 < i_2 \le r$; $1 \le j_1 < j_2 \le r$ and $1 \le k_1 < k_2 \le r$. The $\mathrm{QP}[f]_2^3$ model with $f(u) = u \log u, u > 0$ is equivalent to the second-order quasi point-symmetry (QP_2^3) model. We obtain the following theorems.

Theorem 3.4. *The $QP[f]_1^3$ model for $\{p_{ijk}\}$ minimizes the f-divergence between $\{p_{ijk}\}$ and $\{p^{PS}_{ijk}\}$ with the structure of the PS^3 model under the condition that $\{p_{i\cdot\cdot}\}$, $\{p_{\cdot j\cdot}\}$, $\{p_{\cdot\cdot k}\}$, $\{p_{ij\cdot} + p_{i^*j^*\cdot}\}$, $\{p_{i\cdot k} + p_{i^*\cdot k^*}\}$, $\{p_{\cdot jk} + p_{\cdot j^*k^*}\}$ and $\{p_{ijk} + p_{i^*j^*k^*}\}$ are given.*

Theorem 3.5. *The $QP[f]_2^3$ model for $\{p_{ijk}\}$ minimizes the f-divergence between $\{p_{ijk}\}$ and $\{p^{PS}_{ijk}\}$ with the structure of the PS^3 model under the condition that $\{p_{i\cdot\cdot}\}$, $\{p_{\cdot j\cdot}\}$, $\{p_{\cdot\cdot k}\}$, $\{p_{ij\cdot}\}$, $\{p_{i\cdot k}\}$, $\{p_{\cdot jk}\}$ and $\{p_{ijk} + p_{i^*j^*k^*}\}$ are given.*

Theorem 3.6. *For an $r \times r \times r$ table and h fixed ($h = 1, 2$), the PS^3 model holds if and only if both the $QP[f]_h^3$ and MP_h^3 models hold.*

Finally, consider an r^T table. Let f be a twice-differentiable and strictly convex function, and let $F(u) = df(u)/du$. Let $p_{\boldsymbol{i}}^{PS} = (p_{\boldsymbol{i}} + p_{\boldsymbol{i}^*})/2$ for any $\boldsymbol{i} = (i_1, \ldots, i_T)$, where $i_k^* = r + 1 - i_k$ $(k = 1, \ldots, T)$. For a fixed h $(h = 1, \ldots, T-1)$, consider the hth-order quasi point-symmetry ($\mathrm{QP}[f]_h^T$) model based on f-divergence defined by

$$\begin{cases} p_{\boldsymbol{i}} = p_{\boldsymbol{i}}^{PS} F^{-1}\left(\displaystyle\sum_{k=1}^{T} u_{k(i_k)} + \sum\sum_{1 \le k_1 < k_2 \le T} u_{k_1 k_2 (i_{k_1}, i_{k_2})} \right. \\ \qquad \left. + \cdots + \displaystyle\sum \cdots \sum_{1 \le k_1 < \cdots < k_{T-1} \le T} u_{k_1 \ldots k_{T-1}(i_{k_1}, \ldots, i_{k_{T-1}})} + u_{12\ldots T(\boldsymbol{i})} \right) \text{ for any } \boldsymbol{i}, \\ \text{where } u_{k_1 k_2 \ldots k_l (i_{k_1}, i_{k_2}, \ldots, i_{k_l})} = u_{k_1 k_2 \ldots k_l (i_{k_1}^*, i_{k_2}^*, \ldots, i_{k_l}^*)} \\ \qquad (l = h+1, \ldots, T; 1 \le k_1 < k_2 < \cdots < k_l \le T). \end{cases}$$

When we set $f(u) = u \log u$, $u > 0$, the $\mathrm{QP}[f]_h^T$ model is equivalent to the QP_h^T model. We obtain the following theorems.

Theorem 3.7. *For an r^T table and h fixed $(h = 1, \ldots, T-1)$, the $QP[f]_h^T$ model for $\{p_{\boldsymbol{i}}\}$ minimizes the f-divergence between $\{p_{\boldsymbol{i}}\}$ and $\left\{p_{\boldsymbol{i}}^{PS}\right\}$ with the structure of the PS^T model under the condition that $\left\{p_{\boldsymbol{i}_k}^{\boldsymbol{s}_k}\right\}$ for any $\boldsymbol{s}_k = (s_1, \ldots, s_k)$ and $\boldsymbol{i}_k = (i_1, \ldots, i_k)$, $k = 1, \ldots, h$, are given, and $\left\{p_{\boldsymbol{i}_l}^{\boldsymbol{s}_l} + p_{\boldsymbol{i}_l^*}^{\boldsymbol{s}_l}\right\}$ for any $\boldsymbol{s}_l = (s_1, \ldots, s_l)$, $\boldsymbol{i}_l = (i_1, \ldots, i_l)$ and $\boldsymbol{i}_l^* = (i_1^*, \ldots, i_l^*)$, $l = h+1, \ldots, T$, are given.*

Theorem 3.8. *For an r^T table and h fixed $(h = 1, \ldots, T-1)$, the PS^T model holds if and only if both the $QP[f]_h^T$ and MP_h^T models hold.*

Moreover, Chapter 3 shows that under the PS model, the likelihood ratio statistic for testing goodness-of-fit of the PS model is asymptotically equivalent to the sum of those for testing the QP[f] and MP models for two-way, three-way and multi-way contingency tables, respectively. Although the details and the proof of Theorems 3.9, 3.10 and 3.11 are omitted, we obtain the following theorems.

Theorem 3.9. *For an $r \times r$ table, $G^2(PS)$ is asymptotically equivalent to the sum of $G^2(QP[f])$ and $G^2(MP)$ under the PS model.*

Theorem 3.10. *For an $r \times r \times r$ table and a fixed h $(h = 1, 2)$, $G^2(PS^3)$ is asymptotically equivalent to the sum of $G^2(QP[f]_h^3)$ and $G^2(MP_h^3)$ under the PS^3 model.*

Theorem 3.11. *For an r^T table and a fixed h $(h = 1, \ldots, T-1)$, $G^2(PS^T)$ is asymptotically equivalent to the sum of $G^2(QP[f]_h^T)$ and $G^2(MP_h^T)$ under the PS^T model.*

The details of the proof for theorems and examples are given in Chapter 3.

3.3 Chapter 4

Chapter 4 proposes the moment symmetry model, and gives decompositions of symmetry for two-way, three-way and multi-way tables, respectively. These models and decompositions are described simply as below.

First, for an $r \times r$ tables, consider random variables

$$\boldsymbol{X}^{(1)} = (X_1^{(1)}, X_2^{(1)}, \ldots, X_r^{(1)})^t, \quad \boldsymbol{X}^{(2)} = (X_1^{(2)}, X_2^{(2)}, \ldots, X_r^{(2)})^t,$$

where $X_i^{(1)} = 0, 1$, $X_j^{(2)} = 0, 1$ for $i = 1, \ldots, r$; $j = 1, \ldots, r$, $\sum_{i=1}^{r} X_i^{(1)} = \sum_{j=1}^{r} X_j^{(2)} = 1$, and "$t$" denotes the transpose. $\boldsymbol{\delta}_i$ is the $r \times 1$ vector where the ith factor is set to 1 and all others are set to 0. Define the probability distribution of $(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)})$ by

$$\Pr(\boldsymbol{X}^{(1)} = \boldsymbol{\delta}_i, \boldsymbol{X}^{(2)} = \boldsymbol{\delta}_j) = \Pr(X_1 = i, X_2 = j) = p_{ij},$$

for $i = 1, \ldots, r$; $j = 1, \ldots, r$. Consider the second-order moment symmetry (MOS_2) model, which is expressed as

$$Cov(X_i^{(1)}, X_j^{(2)}) = Cov(X_j^{(1)}, X_i^{(2)}) \quad \text{for } 1 \leq i < j \leq r,$$

where

$$Cov(X_i^{(1)}, X_j^{(2)}) = E\left[(X_i^{(1)} - \mu_i^{(1)})(X_j^{(2)} - \mu_j^{(2)})\right] = p_{ij} - p_{i\cdot}p_{\cdot j},$$
$$\boldsymbol{\mu}_1 = E(\boldsymbol{X}^{(1)}) = (p_{1\cdot}, \ldots, p_{r\cdot})^t, \quad \boldsymbol{\mu}_2 = E(\boldsymbol{X}^{(2)}) = (p_{\cdot 1}, \ldots, p_{\cdot r})^t.$$

The MOS_2 model indicates the symmetry of the second-order moments about the means and leads to the following theorem.

Theorem 3.12. *For an $r \times r$ table, the S model holds if and only if both the MOS_2 and M models hold.*

Second, for an $r \times r \times r$ tables, consider random variables $\boldsymbol{X}^{(1)}$, $\boldsymbol{X}^{(2)}$, and $\boldsymbol{X}^{(3)}$ with

$$\Pr(\boldsymbol{X}^{(1)} = \boldsymbol{\delta}_i, \boldsymbol{X}^{(2)} = \boldsymbol{\delta}_j, \boldsymbol{X}^{(3)} = \boldsymbol{\delta}_k) = \Pr(X_1 = i, X_2 = j, X_3 = k) = p_{ijk}.$$

Then

$$\Pr(\boldsymbol{X}^{(s)} = \boldsymbol{\delta}_i, \boldsymbol{X}^{(t)} = \boldsymbol{\delta}_j) = \Pr(X_s = i, X_t = j) = p_{ij}^{(s,t)} \quad (1 \leq s < t \leq 3),$$

and

$$\Pr(\boldsymbol{X}^{(u)} = \boldsymbol{\delta}_i) = \Pr(X_u = i) = p_i^{(u)} \quad (u = 1, 2, 3).$$

Then, consider the second-order moment symmetry (MOS_2^3) model, which is given by

$$Cov(X_i^{(s)}, X_j^{(t)}) = Cov(X_j^{(s)}, X_i^{(t)}) = Cov(X_i^{(k)}, X_j^{(l)}),$$

where

$$Cov(X_i^{(s)}, X_j^{(t)}) = E\left[(X_i^{(s)} - \mu_i^{(s)})(X_j^{(t)} - \mu_j^{(t)})\right] = p_{ij}^{(s,t)} - p_i^{(s)} p_j^{(t)},$$

for $1 \leq s < t \leq 3$, $1 \leq i, j \leq r$, and $1 \leq k < l \leq 3$; $(s,t) \neq (k,l)$. This model indicates the symmetry and homogeneity of the second-order moments about the means. Because we are also interested in the structure of third-order moments about the means, consider the third-order moment symmetry (MOS_3^3) model, which is expressed as

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

where

$$\mu_{ijk} = E\left[(X_i^{(1)} - \mu_i^{(1)})(X_j^{(2)} - \mu_j^{(2)})(X_k^{(3)} - \mu_k^{(3)})\right].$$

Note that

$$\mu_{ijk} = p_{ijk} - p_i^{(1)} p_{jk}^{(2,3)} - p_j^{(2)} p_{ik}^{(1,3)} - p_k^{(3)} p_{ij}^{(1,2)} + 2 p_i^{(1)} p_j^{(2)} p_k^{(3)}.$$

We obtain the following two theorems and corollary.

Theorem 3.13. *For an $r \times r \times r$ table, the M_2^3 model holds if and only if both the MOS_2^3 and M_1^3 models hold.*

Theorem 3.14. *For an $r \times r \times r$ table, the S^3 model holds if and only if both the MOS_3^3 and M_2^3 models hold.*

Corollary 3.2. *For an $r \times r \times r$ table, the S^3 model holds if and only if all the MOS_3^3, MOS_2^3, and M_1^3 models hold.*

Finally, for an r^T tables, consider random variables $\boldsymbol{X}^{(1)}$, $\boldsymbol{X}^{(2)}$, ..., $\boldsymbol{X}^{(T)}$ with

$$\Pr(\boldsymbol{X}^{(1)} = \boldsymbol{\delta}_{i_1}, \ldots, \boldsymbol{X}^{(T)} = \boldsymbol{\delta}_{i_T}) = \Pr(X_1 = i_1, \ldots, X_T = i_T) = p_{\boldsymbol{i}},$$

where $\boldsymbol{i} = (i_1, \ldots, i_T)$.

Denote the hth-order $(h = 2, \ldots, T)$ moment,

$$E\left[(X_{i_1}^{(s_1)} - \mu_{i_1}^{(s_1)})(X_{i_2}^{(s_2)} - \mu_{i_2}^{(s_2)}) \cdots (X_{i_h}^{(s_h)} - \mu_{i_h}^{(s_h)})\right]$$

by $\mu_{\boldsymbol{i}}^{\boldsymbol{s}_h}$, where $\boldsymbol{s}_h = (s_1, \ldots, s_h)$ and $\boldsymbol{i} = (i_1, \ldots, i_h)$ with $1 \leq s_1 < \cdots < s_h \leq T$ and $i_k = 1, \ldots, r$ $(k = 1, \ldots, h)$.

For a fixed h ($h = 2, \dots, T$), consider the hth-order moment symmetry (MOS_h^T) model, which is given by

$$\mu_{\boldsymbol{i}}^{\boldsymbol{s}_h} = \mu_{\boldsymbol{j}}^{\boldsymbol{s}_h} = \mu_{\boldsymbol{i}}^{\boldsymbol{t}_h},$$

for any permutation $\boldsymbol{j} = (j_1, \dots, j_h)$ of $\boldsymbol{i} = (i_1, \dots, i_h)$ and for any $\boldsymbol{s}_h = (s_1, \dots, s_h)$ and $\boldsymbol{t}_h = (t_1, \dots, t_h)$.

We obtain following theorem and corollary.

Theorem 3.15. *For an r^T table and a fixed h ($h = 2, \dots, T$), the M_h^T model holds if and only if both the MOS_h^T and M_{h-1}^T models hold.*

Corollary 3.3. *For an r^T table, the S^T model holds if and only if all the MOS_T^T, MOS_{T-1}^T, ..., MOS_2^T, and M_1^T models hold.*

The details of the proof for theorems and examples are given in Chapter 4.

Table 1.1: Cross-classification of smoking by lung cancer; adapted from Agresti (2013, p.42). Note that n_{ij} denote the observed frequency in the (i, j) cell of the table ($i = 1, 2$; $j = 1, 2$), and $n_{\cdot 1}$ and $n_{\cdot 2}$ denote the marginal observed frequencies, i.e., $n_{\cdot j} = \sum_{i=1}^{2} n_{ij}$.

	Lung Cancer	
Smoker	Cases	Controls
Yes	n_{11}	n_{12}
No	n_{21}	n_{22}
Total	$n_{\cdot 1}$	$n_{\cdot 2}$

Table 1.2: Unaided distance vision of students aged 18 to about 25 including about 10% women in Faculty of Science and Technology, Science University of Tokyo in Japan examined in April 1982; adapted from Tomizawa (1985). Note that n_{ij} denotes the observed frequency in the (i, j) cell of the table ($i = 1, \ldots, 4$; $j = 1, \ldots, 4$), $n_{i\cdot}$ and $n_{\cdot j}$ denote the row and column marginal observed frequencies, i.e., $n_{i\cdot} = \sum_{j=1}^{4} n_{ij}$ and $n_{\cdot j} = \sum_{i=1}^{4} n_{ij}$, and n denotes the total number of observed frequencies.

Right eye	Left eye grade				Total
grade	Best(1)	Second(2)	Third(3)	Worst(4)	
Best(1)	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1\cdot}$
Second(2)	n_{21}	n_{22}	n_{23}	n_{24}	$n_{2\cdot}$
Third(3)	n_{31}	n_{32}	n_{33}	n_{34}	$n_{3\cdot}$
Worst(4)	n_{41}	n_{42}	n_{43}	n_{44}	$n_{4\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$n_{\cdot 4}$	n

Table 1.3: Opinions about government spending in 2016 from the 2016 General Social Survey; constructed from Smith *et al.* (2018). Note that n_{ijk} denotes the observed frequency in the (i, j, k) cell of the table ($i = 1, \ldots, 3$; $j = 1, \ldots, 3$; $k = 1, \ldots, 3$).

		Assistance to the poor		
Education	Environment	(1) too little	(2) about right	(3) too much
(1) too little	(1) too little	n_{111}	n_{112}	n_{113}
(1) too little	(2) about right	n_{121}	n_{122}	n_{123}
(1) too little	(3) too much	n_{131}	n_{132}	n_{133}
(2) about right	(1) too little	n_{211}	n_{212}	n_{213}
(2) about right	(2) about right	n_{221}	n_{222}	n_{223}
(2) about right	(3) too much	n_{231}	n_{232}	n_{233}
(3) too much	(1) too little	n_{311}	n_{312}	n_{313}
(3) too much	(2) about right	n_{321}	n_{322}	n_{323}
(3) too much	(3) too much	n_{331}	n_{332}	n_{333}

References

[1] Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, 2nd edition. Wiley, New Jersey.

[2] Agresti, A. (2013). *Categorical Data Analysis*, 3rd edition. Wiley, New Jersey.

[3] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **AC-19**, 716-723.

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

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

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

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

[8] Caussinus, H. (1965). Contribution *à* l'analyse statistique des tableaux de corr*é*lation. *Annales de la Faculté des Sciences de l'Université de Toulouse*, **29**, 77-182.

[9] Csiszár, I. and Shields, P. C. (2004). *Information Theory and Statistics: A Tutorial.* Now Publishers, Hanover, Massachusetts.

[10] Darroch, J. N. and McCloud, P. I. (1986). Category distinguishability and observer agreement. *Australian Journal of Statistics*, **28**, 371-388.

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

[12] Gilula, Z., Krieger, A. M. and Ritov, Y. (1988). Ordinal association in contingency tables: some interpretive aspects. *Journal of the American Statistical Association*, **83**, 540-545.

[13] Goodman, L. A. (1979). Simple models for the analysis of association in cross classifications having ordered categories. *Journal of the American Statistical Association*, **74**, 537-552.

[14] Kateri, M. (2014). *Contingency Table Analysis: Methods and Implementation Using R*. Birkhäuser/Springer, New York.

[15] Kateri, M. (2018). ϕ-divergence in contingency table analysis. *Entropy*, **20**, 1-12.

[16] Kateri, M. and Agresti. A. (2007). A class of ordinal quasi-symmetry models for square contingency tables. *Statistics and Probability Letters*, **77**, 598-603.

[17] Kateri, M. and Papaioannou, T. (1997). Asymmetry models for contingency tables. *Journal of the American Statistical Association*, **92**, 1124-1131.

[18] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling.* Springer, New York.

[19] Lang, J. B. (1996). On the partitioning of goodness-of-fit statistics for multivariate categorical response models. *Journal of the American Statistical Association*, **91**, 1017-1023.

[20] Lang, J. B. and Agresti, A. (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association*, **89**, 625-632.

[21] Lovison, G. (2000). Generalized symmetry models for hypercubic concordance tables. *International Statistical Review*, **68**, 323-338.

[22] McCullagh, P. (1977). A logistic model for paired comparisons with ordered categorical data. *Biometrika*, **64**, 449-453.

[23] Miyamoto, N., Niibe, K. and Tomizawa, S. (2005). Decompositions of marginal homogeneity model using cumulative logistic models for square contingency tables with ordered categories. *Austrian Journal of Statistics*, **34**, 361-373.

[24] Read, C. B. (1977). Partitioning chi-square in contingency tables: a teaching approach. *Communications in Statistics-Theory and Methods*, **6**, 553-562.

[25] Saigusa, Y., Tahata, K. and Tomizawa, S. (2015). Orthogonal decomposition of symmetry model using the ordinal quasi-symmetry model based on f-divergence for square contingency tables. *Statistics and Probability Letters*, **101**, 33-37.

[26] Smith, T. W., Davern, M., Freese, J. and Hout, M. (2018). *General Social Surveys, 1972-2016* [machine-readable data file] /Principal Investigator, Smith, T. W.; Co-Principal Investigator, Davern, M.; Co-Principal Investigator, Freese, J; Co-Principal Investigator, Hout, M.; Sponsored by National Science Foundation. –NORC ed.– Chicago: NORC at the University of Chicago [producer]; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut [distributor].

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

[28] Tahata, K., Katakura, S. and Tomizawa, S. (2007). Decompositions of marginal homogeneity model using cumulative logistic models for multi-way contingency tables. *Revstat: Statistical Journal*, **5**, 163-176.

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

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

[31] 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.

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

[33] Yoshimoto, T., Tahata, K., Iki, K. and Tomizawa, S. (2019) Moment symmetry models and decompositions of symmetry for multi-way contingency tables. *Calcutta Statistical Association Bulletin*, **71**, 83-98.

[34] Yoshimoto, T., Tahata, K., Saigusa, Y. and Tomizawa, S. (2019) Quasi point-symmetry models based on f-divergence and decomposition of point-symmetry for multi-way contingency tables. *SUT Journal of Mathematics*, **55**, 109-137.

[35] Yoshimoto, T., Tahata, K. and Tomizawa, S. (2019) Marginal cumulative logistic model of general order for multi-way contingency tables. *Austrian Journal of Statistics*, to appear.