Outline of Thesis

Model and Measure of Symmetry for Ordinal Square Contingency Tables

(順序正方分割表における対称性のモデルと尺度)

Shuji Ando

Department of Information Sciences, Faculty of Science and Technology,
Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

October, 2015

Contents

Chapter 1

Introduction

The development of methods for analyzing categorical data that began in early the twentieth century has continued in recent years. Also, the development of the statistical analysis methods for categorical data has been used in many areas, such as biomedicine, epidemiology, education, social sciences, marketing, and quality control.

The simple categorical variables in one that has just two categories. Such variables, for which the two categories are often given the generic labels as "yes-no" or "success-failure", are called binary variables. When a categorical variables has more than two categories, we distinguish between two types of categorical scales. Variables having categories without a natural ordering are said to measured on a nominal scales and are called nominal variables. Example is favorite type of music ("classical", "country", "folk", "jazz", "rock"). On the other hand, many categorical variables have natural ordering. Such variables are said to be measured on an ordinal scale and are called ordinal categories. Examples are eye grade ("best", "second", "third", "worst") and value level ("upper", "middle", "lower").

In an example of two-way contingency tables, Table 1 comes from one of the link between lung cancer and smoking. For the data in Table 1, we are interested in whether there is a structure of independence. Thus, we are interested in whether onset of lung cancer is independent on the smoking behavior. For the data in Table 1, when there is a structure of independence, the probability that the patients of cases will fall in category "yes" is equal to the probability that the patients of controls will fall in category "yes".

We next consider contingency tables with the same row and column classifications such as Table 2. Such tables are called square contingency tables. The data in Table 2 are constructed from unaided distance vision. Generally, in contingency tables, we are interested in whether the independence between the row and column classifications holds or not. However, for the analysis of square contingency tables, we are interested in whether the row classifications are symmetric with the column classifications or not, instead of the independence. For square contingency tables, the independence between the row and column is unlikely to hold, because many observations fall in main diagonal cells which indicate that the value of row category is the same as the value of column category.

This thesis deals especially with the methods for analyzing square contingency tables. In the analysis of contingency tables, we can only obtain observed frequencies. The purpose of analyzing contingency tables is to estimate the unknown probability distribution with high confidence and easy interpretation. It is necessary to introduce the statistical models that fit the data well and provide easy interpretation. For the structure of symmetry, Bowker (1948) proposed the symmetry model. The extensions of the symmetry model were proposed, for example, the marginal homogeneity model (Stuart, 1955), the quasi-symmetry model (Caussinus, 1965). For the structure of asymmetry, the conditional symmetry model (McCullagh, 1978), the diagonals-parameter symmetry model (Goodman, 1979) and the linear diagonals-parameter symmetry model (Agresti, 1983) were proposed.

Caussinus (1965) gave the decomposition of the symmetry model such that the symmetry model holds if and only if both the quasi-symmetry and the marginal homogeneity models hold. We see from this decomposition which assuming that the quasi-symmetry model holds true, the hypothesis that the symmetry model holds is equivalent to the hypothesis that the marginal homogeneity model holds. The decompositions of the symmetry model are given (see Section 1.1.2).

For the analysis of contingency tables, when a model does not hold, we are interested in applying the extended models, analyzing the residual, and also measuring the degree of departure from the corresponding model. For contingency tables, various measures of association have been proposed (see Agresti, 2013, Sec. 2.4). For square contingency tables, Tomizawa (1994), Tomizawa, Seo and Yamamoto (1998), Tomizawa, Miyamoto and Hatanaka (2001) and Tahata, Yamamoto, Nagatani and Tomizawa (2009) consider measures which represent the degree of departure from the symmetry model.

This chapter describes the background of the research, and describes outline for the Chapter 2 through 4 (although the outline of thesis is omitted the contents of Chapter 2 thorough 4). Chapter 2 proposes an extended asymmetry model based on quasi-symmetry, and gives the decomposition of the symmetry model using the proposed model. Chapter 3 presents a measure which represents the degree of departure from average marginal homogeneity for square contingency tables with ordered categories. Chapter 4 shows a measure which represents the degree of departure from generalized marginal homogeneity for square contingency tables with ordered categories.

1.1 Analysis of square contingency tables

1.1.1 Symmetry and asymmetry models

Consider an $R \times R$ square contingency table with the same row and column classifications. Let p_{ij} denote the probability that an observation will fall in the ith row and jth column of the table ($i = 1, \dots, R; j = 1, \dots, R$).

Bowker (1948) considered the symmetry (S) model, defined by

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

where $\psi_{ij} = \psi_{ji}$. This indicates that the probability that an observation will fall in the (i, j) cell is equal to the probability that the observation falls in the symmetric (j, i) cell. Namely, this describes a structure of symmetry of the cell probabilities $\{p_{ij}\}$ with respect to the main diagonal of the table. For the data in Table 2, this model indicates that the probability that a man's/woman's right eye grade is i and her left eye grade is j $(> i)$ is equal to the probability that the man's/woman's right eye grade is j and her left eye grade is i.

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

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

where $\psi_{ij} = \psi_{ji}$. A special case of this model obtained by putting $\{\alpha_i = \beta_i\}$ is the S model. Denote the odds ratio for rows i and j $(> i)$, and columns s and t $(> s)$ by $\theta_{(i<j;s<t)}$, where

$$\theta_{(i<j;s<t)} = \frac{p_{is}p_{jt}}{p_{js}p_{it}}.$$

Using the odds ratio $\theta_{(i<j;s<t)}$, the QS model is also expressed as

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

Thus, the QS model has characterization in terms of symmetry of odds ratios.

Let X and Y denote the row and column variables, respectively. Stuart (1955) considered the marginal homogeneity (MH) model, defined by

$$\Pr(X = i) = \Pr(Y = i) \quad (i = 1, \ldots, R),$$

namely

$$p_{i\cdot} = p_{\cdot i} \quad (i = 1, \ldots, R),$$

where $p_{i\cdot} = \sum_{k=1}^{R} p_{ik}$ and $p_{\cdot i} = \sum_{k=1}^{R} p_{ki}$. The MH model indicated that the row marginal distribution is identical to the column marginal distribution. Note that the S model implies the MH model. For the data in Table 2, this model indicates that the probability that a man's/woman's right eye grade X is i is equal to the probability that the man's/woman's left eye grade Y is i.

For square contingency tables with ordered categories, McCullagh (1978) considered the conditional symmetry (CS) model.

$$p_{ij} = \begin{cases} \delta\psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

where $\psi_{ij} = \psi_{ji}$. A special case of this model obtained by putting $\delta = 1$ is the S model. For the data in Table 2, this model indicates that the probability that a woman's right eye grade is i and her left eye grade is j $(> i)$ is δ times higher than the probability that the woman's right eye grade is j and her left eye grade is i. If $\delta > 1$, a man's/woman's right eye is better than her left eye, and if $\delta < 1$, a man's/woman's left eye is better than her right eye. Also, the CS model indicates that the conditional probability $\Pr(X = i, Y = j | X < Y)$ is equal to the conditional probability $\Pr(X = j, Y = i | X > Y)$.

Agresti (1983) considered the linear diagonals-parameter symmetry (LDPS) model, defined by

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

where $\psi_{ij} = \psi_{ji}$. 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.

When we can assign the ordered known scores $u_1 < \cdots < u_R$, the ordinal quasi-symmetry (OQS) model is defined by

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

where $\psi_{ij} = \psi_{ji}$, see Agresti (2013, p. 431). Note that the OQS model with $\{u_i = i\}$ is identical to the LDPS model.

Let

$$r_i^X = \sum_{k=1}^{i-1} p_{k\cdot} + \frac{p_{i\cdot}}{2}, \quad r_i^Y = \sum_{l=1}^{i-1} p_{\cdot l} + \frac{p_{\cdot i}}{2} \quad (i = 1, \ldots, R).$$

The $\{r_i^X\}$ and $\{r_i^Y\}$ are the marginal ridits; see Bross (1958). When it may be difficult to assign known scores $\{u_i\}$ for the given data, Iki, Tahata and Tomizawa (2009) considered the ridit score type quasi-symmetry (RQS) model, defined by

$$p_{ij} = \begin{cases} \delta^{v_j - v_i}\psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

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

As the extension of the LDPS model, Tomizawa (1987) considered the two-ratios-parameter symmetry (2PS) model, defined by

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

where $\psi_{ij} = \psi_{ji}$. A special case of this model obtained by putting $\tau = 1$ is the LDPS model. Also, the 2PS model with $\delta = 1$ is the CS model. Further, the 2PS model with $\tau = \delta = 1$ is the S model. This indicates that the probability that an observation will fall in the (i, j) cell, $i < j$, is $\tau\delta^{j-i}$ times higher than the probability that the observation falls in the (j, i) cell.

1.1.2 Decomposition of symmetry model

This section shows the decomposition of the S model.

If the S model holds, the MH model holds; but the converse does not hold. We are interested in what the structure is necessary for obtaining the S model, in addition to the MH model.

Caussinus (1965) gave the following theorem for a decomposition of the S model using the MH model.

Theorem 1: *The S model holds if and only if both the QS and MH models hold.*

When the S model fits the data poorly, this theorem would be useful for seeing which of the lack of QS model, the lack of MH model influences stronger. Also, we see from this decomposition that assuming that the QS model holds true, the hypothesis that the S model holds is equivalent to the hypothesis that the MH model holds.

We are next interested in what the structure is necessary for obtaining the S model, in addition to the asymmetry (e.g., CS, LDPS, RQS and 2PS) model.

Consider the global symmetry (GS) model defined by

$$\sum_{i<j}\sum p_{ij} = \sum_{i>j}\sum p_{ij};$$

see Read (1977).

Read (1977) gave the following theorem for a decomposition of the S model using the CS model.

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

The marginal mean equality (ME) model can be considered as follow:

$$\sum_{i=1}^{R} ip_{i\cdot} = \sum_{j=1}^{R} jp_{\cdot j}.$$

This model indicates that the mean of row variable X equals to the mean of column variable Y, i.e, $E(X) = E(Y)$. The mean ridit for the distribution of Y when the distribution of X is the identified distribution for calculating the ridits is

$$R_X(Y) = \sum_{j=1}^{R} r_j^X p_{\cdot j}.$$

Similarly, we have

$$R_Y(X) = \sum_{i=1}^{R} r_i^Y p_{i\cdot}.$$

We shall refer to the structure of $R_X(Y) = R_Y(X)$ as the marginal mean ridits equality (MR) model (see Agresti, 1984, p. 209).

Yamamoto, Iwashita and Tomizawa (2007) gave the following theorem for the decomposition of the S model using the LDPS model.

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

Also, Iki et al. (2009) showed the following therem for the decomposition of the S model using the RQS model.

Theorem 4: *The S model holds if and only if both the RQS and MR models hold.*

In addition, Tahata and Tomizawa (2009) gave the following theorem for the decomposition of the S model using the 2PS model.

Theorem 5: *The S model holds if and only if all of the 2PS, ME and GS models hold.*

1.1.3 Goodness-of-fit test

Let n_{ij} denote the observed frequency in the (i, j) cell of the table ($i = 1, \ldots, R; j = 1, \ldots, R$). Assume that a multinomial distribution applies to the $R \times R$ table. The maximum likelihood estimates of expected frequencies under the model could be obtained using, e.g., the Newton-Raphson method in the log-likelihood equation.

Each model can be tested for goodness-of-fit by the power-divergence statistic with the corresponding degrees of freedom. The power-divergence statistic is defined by

$$I^{(\lambda)} = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{R} \sum_{j=1}^{R} \left[n_{ij} \left\{ \left(\frac{n_{ij}}{\widehat{m}_{ij}} \right)^{\lambda} - 1 \right\} \right] \quad (-\infty < \lambda < \infty),$$

where $\widehat{m}_{ij}$ is the maximum likelihood estimate of expected frequency m_{ij} under the model, the values at $\lambda = 0$ and $\lambda = -1$ are taken to be the continuous limits as $\lambda \to 0$ and $\lambda \to -1$, respectively. For instance, the power-divergence statistic includes the likelihood ratio chi-square statistic (denoted by G^2) when $\lambda = 0$, and Pearson's chi-squared statistic (denoted by X^2) when $\lambda = 1$. For more details of the power-divergence statistics $I^{(\lambda)}$, see Cressie and Read (1984), and Read and Cressie (1988, p. 15).

The test statistic G^2 is given by

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

The test statistic X^2 is given by

$$X^2 = \sum_{i=1}^{R} \sum_{j=1}^{R} \frac{(n_{ij} - \widehat{m}_{ij})^2}{\widehat{m}_{ij}}.$$

Under the model, these statistics have asymptotically a central chi-squared distribution with the corresponding degrees of freedom.

1.1.4 Measures of departure from symmetry and asymmetry models

For the analysis of square contingency tables, when the model does not hold, we are interested in applying the extended model. Also, we are interested in measuring the degree of departure from the model. We point out that the test statistic (e.g., Pearson's chi-squared statistic or likelihood ratio chi-squared statistic) is used for testing the goodness-of-fit of the model, however, the test statistic would not be useful for comparing the degree of departure from the model in several tables. Because the test statistic depends on the dimension R and sample size. Therefore, we are interested in measuring or comparing the degree of departure from the model in several tables using the analysis method of independent of the dimension R and sample size.

For square contingency tables with nominal categories, Tomizawa (1994) considered a measure (denoted by ϕ) which represents the degree of departure from symmetry. Assume that $\{p_{ij} + p_{ji} > 0\}$. The measure ϕ is expressed as follows:

$$\phi = \frac{1}{\log 2} I_S,$$

where

$$I_S = \sum\sum_{i \neq j} a_{ij} \log\left(\frac{a_{ij}}{b_{ij}}\right),$$

with

$$\delta = \sum\sum_{i \neq j} p_{ij}, \quad a_{ij} = \frac{p_{ij}}{\delta}, \quad b_{ij} = \frac{a_{ij} + a_{ji}}{2}.$$

Note that I_S is the Kullback-Leibler information between $\{a_{ij}\}$ and $\{b_{ij}\}$. Note that (1) the measure ϕ lies between 0 and 1, (2) $\phi = 0$ if and only if the S model holds, and (3) $\phi = 1$ if and only if the degree of departure from symmetry is the largest, in the sense that $p_{ij} = 0$ (then $p_{ji} > 0$) or $p_{ji} = 0$ (then $p_{ij} > 0$) for $i \neq j$. Tomizawa et al. (1998) considered a generalization of ϕ using Cressie-Read power-divergence between $\{a_{ij}\}$ and $\{b_{ij}\}$. Measures which represent the degree of departure from marginal homogeneity were proposed by Tomizawa (1995a) and Tomizawa and Makii (2001).

For square contingency tables with ordered categories, measures which represent the degree of departure from symmetry were proposed by Tomizawa et al. (2001), Tahata et al. (2009), and Tahata, Miyazawa and Tomizawa (2010). Also, measures which represent the degree of departure from marginal homogeneity were proposed by Tomizawa, Miyamoto and Ashihara (2003) and Tahata, Iwashita and Tomizawa (2006). Although the detail is omitted, for various models, measures which represent the degree of departure from the corresponding model were proposed.

1.2 Outline of Chapters

We point out that (1) Chapter 2 is constructed from Tahata, Ando and Tomizawa (2011), (2) Chapter 3 is constructed from Yamamoto, Ando and Tomizawa (2011), and (3) Chapter 4 is constructed from Ando, Tahata and Tomizawa (2015).

Chapter 2 (1) proposes the ridit score type two-parameters symmetry (R2PS) model, which is the 2PS model with $\{i\}$ replaced by $\{v_i\}$, and (2) gives a decomposition of the S model using the R2PS model.

For an $R \times R$ square contingency table with ordered categories, consider the R2PS model defined by

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

where $\phi_{ij} = \phi_{ji}$ and

$$\delta_{ij} = \begin{cases} \gamma & (i < j), \\ 1 & (i \geq j). \end{cases}$$

Note that $v_1 < \cdots < v_R$ are unknown scores based on the ridits, defined in Section 1.1.1. The model may be expressed as

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

This indicates that the probability that an observation will fall in the cell (i, j), $i < j$, is $\gamma \theta^{v_j - v_i}$ times higher than the probability that the observation falls in the cell (j, i). Special cases of this model obtained by putting $\gamma = 1$ and $\{v_i = i\}$ are the RQS and 2PS models, respectively. The R2PS model with $\gamma = \theta = 1$ is the S model.

Under the R2PS model, if $\gamma > 1$ and $\theta > 1$, then $p_{ij} > p_{ji}$ for all $i < j$, thus, $F_i^X > F_i^Y$ for all $i = 1, \ldots, R-1$, where $F_i^X = \sum_{k=1}^{i} p_{k\cdot}$ and $F_i^Y = \sum_{l=1}^{i} p_{\cdot l}$. Also, if $\gamma < 1$ and $\theta < 1$, then $p_{ij} < p_{ji}$ for all $i < j$, thus, $F_i^X < F_i^Y$ for all $i = 1, \ldots, R-1$. Therefore the parameters γ and θ in the R2PS model may be useful for making inferences such as that X is stochastically less than Y or vice versa. Moreover, under the R2PS model, if $\gamma > 1$ and $\theta < 1$ (or $\gamma < 1$ and $\theta > 1$), then it is likely that some pairs of symmetric cells will have the structure of $p_{ij} > p_{ji}$ $(i < j)$ and others will have that of $p_{kl} < p_{lk}$ $(k < l)$ depending on the deference between ridit scores. Thus, for the data having the above structure, the R2PS model may be appropriate although the RQS model cannot have the above structure.

Next, in Chapter 2, we obtain the following theorem for a decomposition of the S model using the R2PS model:

Theorem 6: *The S model holds if and only if all of the R2PS, MR and GS models hold.*

See Chapter 2 for the detail of the proof of this theorem.

Chapter 3 proposes a measure which represents the degree of departure from average marginal homogeneity for square contingency tables with ordered categories.

Consider an $R \times R$ table with ordered categories. Let

$$G_{1(i)} = \sum_{s=1}^{i} \sum_{t=i+1}^{R} p_{st} \quad [= \Pr(X \le i, Y \ge i+1)],$$

and

$$G_{2(i)} = \sum_{s=i+1}^{R} \sum_{t=1}^{i} p_{st} \quad [= \Pr(X \ge i+1, Y \le i)],$$

for $i = 1, \ldots, R-1$. Then, by considering the difference between the cumulative marginal probabilities, $F_i^X - F_i^Y$ for $i = 1, \ldots, R-1$, we see that the MH model may also be expressed as

$$G_{1(i)} = G_{2(i)} \quad (i = 1, \ldots, R-1).$$

Namely, this model also states that the cumulative probability that an observation will fall in row category i or below and column category $i+1$ or above is equal to the cumulative probability that the observation falls in column category i or below and row category $i+1$ or above for $i = 1, \ldots, R-1$.

Let

$$\Delta = \sum_{i=1}^{R-1} (G_{1(i)} + G_{2(i)}),$$

and for $i = 1, \ldots, R-1$,

$$G_{1(i)}^{*} = \frac{G_{1(i)}}{\Delta}, \quad G_{2(i)}^{*} = \frac{G_{2(i)}}{\Delta}.$$

Assuming that $\{G_{1(i)} + G_{2(i)} \neq 0\}$, Chapter 3 considers the measure defined by

$$\Psi = \frac{4}{\pi} \sum_{i=1}^{R-1} \left(G_{1(i)}^{*} + G_{2(i)}^{*} \right) \left(\theta_i - \frac{\pi}{4} \right),$$

where

$$\theta_i = \cos^{-1}\left(\frac{G_{1(i)}}{\sqrt{G_{1(i)}^2 + G_{2(i)}^2}}\right).$$

Noting that the range of θ_i is $0 \leq \theta_i \leq \pi/2$, we see that the measure Ψ lies between -1 and 1. The measure Ψ has characteristics that (i) $\Psi = -1$ if and only if $G_{2(i)} = 0$ (then $F_i^X > F_i^Y$) for all $i = 1, \ldots, R-1$, [marginal inhomogeneity with all probabilities zero of lower left triangle (say, L-marginal inhomogeneity)], and (ii) $\Psi = 1$ if and only if $G_{1(i)} = 0$ (then $F_i^X < F_i^Y$) for all $i = 1, \ldots, R-1$, [marginal inhomogeneity with all probabilities zero of upper right triangle (say, U-marginal inhomogeneity)]. In addition, $\Psi = 0$ indicates that the weighted average of $\{\theta_i - \frac{\pi}{4}\}$ equals zero. Thus when $\Psi = 0$, we shall refer to this structure as the average marginal homogeneity. We note that if the marginal homogeneity holds then the average marginal homogeneity holds, but the converse does not hold.

Therefore, using the measure Ψ, we can see whether the average marginal homogeneity departs toward the L-marginal inhomogeneity or toward the U-marginal inhomogeneity. As the measure Ψ approaches -1, the departure from the average marginal homogeneity becomes greater toward the L-marginal inhomogeneity. While as the Ψ approaches 1, it becomes greater toward the U-marginal inhomogeneity.

Next, we consider the relationship between the measure Ψ and the extended marginal homogeneity (EMH) model. The EMH model considered by Tomizawa (1984), is defined by

$$G_{1(i)} = \tau G_{2(i)} \quad (i = 1, \ldots, R-1).$$

A special case of this model obtained by putting $\tau = 1$ is the MH model. If the EMH model holds true, then the measure Ψ can be expressed as

$$\Psi = \frac{4}{\pi}\cos^{-1}\left(\frac{\tau}{\sqrt{\tau^2+1}}\right) - 1.$$

Therefore, $\Psi = 0$ if and only if the marginal homogeneity model holds, i.e., $\tau = 1$, thus $G_{1(i)} = G_{2(i)}$ for $i = 1, \ldots, R-1$. As the value of τ approaches the infinity, the measure Ψ approaches -1. While as the value of τ approaches zero, Ψ approaches 1. Thus when the extended marginal homogeneity model holds in a table, the measure Ψ represents the degree of departure from marginal homogeneity toward the L-marginal inhomogeneity or toward the U-marginal inhomogeneity.

Chapter 4 proposes a measure which represents the degree of departure from generalized marginal homogeneity (GHM) for square contingency tables with ordered categories.

Tomizawa (1995b) proposed the GMH model, defined by

$$G_{1(i)} = \Delta\Theta^{i-1} G_{2(i)} \quad (i = 1, \ldots, R-1).$$

This model indicates that the cumulative probability that an observation will fall in row category i or below and column category $i+1$ or above, is $\Delta\Theta^{i-1}$ times higher than the cumulative probability that the observation falls in column category i or below and row category $i+1$ or above for $i = 1, \ldots, R-1$. A special case of this model obtained by putting $\Theta = 1$ is equivalent to the EMH model. Also the GMH model with $\Delta = \Theta = 1$ is equivalent to the MH model. Moreover, the GMH model may be expressed as

$$c_i = d_i \quad (i = 1, \ldots, R-2),$$

where

$$c_i = \frac{G_{1(i)} G_{2(i+1)}}{C}, \quad d_i = \frac{G_{1(i+1)} G_{2(i)}}{D},$$

$$C = \sum_{i=1}^{R-2} G_{1(i)} G_{2(i+1)}, \quad D = \sum_{i=1}^{R-2} G_{1(i+1)} G_{2(i)},$$

with $C > 0$ and $D > 0$. Namely the GMH model indicates that there is a structure of homogeneity between $\{c_i\}$ and $\{d_i\}$ for $i = 1, \ldots, R-2$.

The power-divergence between two discrete probability distributions $\{a_i\}$ and $\{q_i\}$ for $i = 1, \ldots, R-2$, is defined by

$$I^{(\lambda)}\left(\{a_i\};\{q_i\}\right) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^{R-2} \left[a_i \left\{ \left(\frac{a_i}{q_i} \right)^{\lambda} - 1 \right\} \right] \quad (-\infty < \lambda < \infty),$$

where the values at $\lambda = 0$ and $\lambda = -1$ are taken to be the continuous limits as $\lambda \to 0$ and $\lambda \to -1$, respectively. For instance, the power-divergence includes the Kullback-Leibler information when $\lambda = 0$, and the Pearson's chi-squared type discrepancy when $\lambda = 1$. For more details of the power-divergence $I^{(\lambda)}(\cdot;\cdot)$, see Cressie and Read (1984), and Read and Cressie (1988, p. 15).

Let

$$q_i = \frac{c_i + d_i}{2} \quad \text{for } i = 1, \ldots, R-2.$$

Assume that $c_i + d_i > 0$ for $i = 1, \ldots, R-2$, Chapter 4 considers the measure defined by

$$\Phi^{(\lambda)} = \frac{\lambda(\lambda+1)}{2(2^{\lambda}-1)} \left[I^{(\lambda)}\left(\{c_i\};\{q_i\}\right) + I^{(\lambda)}\left(\{d_i\};\{q_i\}\right) \right] \quad (\lambda > -1). \qquad (1.1)$$

Note that (i) $I^{(\lambda)}(\cdot;\cdot)$ is the power-divergence and (ii) if $\lambda \leq -1$ in (1.1), then $\Phi^{(\lambda)}$ becomes diverging. Also note that a real value λ is chosen by user. When $\lambda = 0$, we see

$$\Phi^{(0)} = \frac{1}{2\log 2} \left[I^{(0)}\left(\{c_i\};\{q_i\}\right) + I^{(0)}\left(\{d_i\};\{q_i\}\right) \right],$$

where

$$I^{(0)}\left(\{a_i\};\{q_i\}\right) = \sum_{i=1}^{R-2} a_i \log\left(\frac{a_i}{q_i}\right).$$

Note that $I^{(0)}(\cdot;\cdot)$ is the Kullback-Leibler information. When $\lambda = 1$, we see

$$\Phi^{(1)} = I^{(1)}\left(\{c_i\};\{q_i\}\right) + I^{(1)}\left(\{d_i\};\{q_i\}\right),$$

where

$$I^{(1)}\left(\{a_i\};\{q_i\}\right) = \frac{1}{2} \sum_{i=1}^{R-2} \frac{(a_i - q_i)^2}{q_i}.$$

Note that $I^{(1)}(\cdot;\cdot)$ is Pearson's chi-squared type discrepancy.

Let

$$c_i^* = \frac{c_i}{c_i + d_i}, \quad d_i^* = \frac{d_i}{c_i + d_i} \quad \text{for } i = 1, \ldots, R-2.$$

Note that $\{c_i^* + d_i^* = 1\}$. The GMH model can be expressed as

$$c_i^* = d_i^* \left(= \frac{1}{2} \right) \quad \text{for } i = 1, \ldots, R-2.$$

Then the measure $\Phi^{(\lambda)}$ may be expressed as

$$\Phi^{(\lambda)} = \frac{\lambda(\lambda+1)}{2(2^\lambda - 1)} \sum_{i=1}^{R-2} (c_i + d_i)\, I_i^{(\lambda)} \quad (\lambda > -1),$$

where

$$I_i^{(\lambda)} = \frac{1}{\lambda(\lambda+1)} \left[c_i^* \left\{ \left(\frac{c_i^*}{1/2} \right)^\lambda - 1 \right\} + d_i^* \left\{ \left(\frac{d_i^*}{1/2} \right)^\lambda - 1 \right\} \right].$$

Therefore, the measure $\Phi^{(\lambda)}$ would represent the weight sum of the power-divergence $I_i^{(\lambda)}$. When $\lambda = 0$, we see

$$\Phi^{(0)} = \frac{1}{2 \log 2} \sum_{i=1}^{R-2} (c_i + d_i)\, I_i^{(0)},$$

where

$$I_i^{(0)} = \left[c_i^* \log \left(\frac{c_i^*}{1/2} \right) + d_i^* \log \left(\frac{d_i^*}{1/2} \right) \right].$$

When $\lambda = 1$, we see

$$\Phi^{(1)} = \sum_{i=1}^{R-2} (c_i + d_i)\, I_i^{(1)},$$

where

$$I_i^{(1)} = \frac{1}{2} \left[\frac{\left(c_i^* - \frac{1}{2}\right)^2}{\frac{1}{2}} + \frac{\left(d_i^* - \frac{1}{2}\right)^2}{\frac{1}{2}} \right].$$

For $\lambda > -1$, the Patil-Taillie's (1982) diversity index of degree λ for $\{c_i^*, d_i^*\}$, is defined by

$$H_i^{(\lambda)} = \frac{1}{\lambda} \left[1 - (c_i^*)^{\lambda+1} - (d_i^*)^{\lambda+1} \right],$$

where the value at $\lambda = 0$ is taken to be the continuous limit as $\lambda \to 0$. For instance, the diversity index includes the Shannon entropy when $\lambda = 0$, and the Gini concentration when $\lambda = 1$.

Moreover, the $\Phi^{(\lambda)}$ may also be expressed by using the diversity index as

$$\Phi^{(\lambda)} = 1 - \frac{\lambda 2^{\lambda}}{2(2^{\lambda}-1)} \sum_{i=1}^{R-2} (c_i + d_i)\, H_i^{(\lambda)} \quad (\lambda > -1).$$

Therefore, the measure $\Phi^{(\lambda)}$ would represent the weight sum of the diversity index $H_i^{(\lambda)}$. When $\lambda = 0$, we see

$$\Phi^{(0)} = 1 - \frac{1}{2\log 2} \sum_{i=1}^{R-2} (c_i + d_i)\, H_i^{(0)},$$

where

$$H_i^{(0)} = -c_i^* \log c_i^* - d_i^* \log d_i^*.$$

When $\lambda = 1$, we see

$$\Phi^{(1)} = 1 - \sum_{i=1}^{R-2} (c_i + d_i)\, H_i^{(1)},$$

where

$$H_i^{(1)} = 1 - (c_i^*)^2 - (d_i^*)^2 .$$

In Chapter 4, we obtain the following theorem.

Theorem 7: *For each* λ (> -1),
(i) *the measure* $\Phi^{(\lambda)}$ *lies between 0 and 1,*
(ii) $\Phi^{(\lambda)} = 0$ *if and only if the GMH model holds, namely* $\{c_i^* = d_i^* = \frac{1}{2}\}$, *and*
(iii) $\Phi^{(\lambda)} = 1$ *if and only if the degree of departure from GMH is the largest in the sense that* $c_i^* = 0$ *(then* $d_i^* = 1$*) or* $d_i^* = 0$ *(then* $c_i^* = 1$*) [namely,* $c_i = 0$ *(then* $d_i > 0$*) or* $d_i = 0$ *(then* $c_i > 0$*)] for* $i = 1, \ldots, R-2$.

See Chapter 4 for the detail of the proof of this theorem.

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.* Wiley, New York.

Agresti, A. (2013). *Categorical Data Analysis*, 3rd ed. Wiley, New York.

Ando, S., Tahata, K. and Tomizawa, S. (2015). Measure of departure from generalized marginal homogeneity model for square contingency tables with ordered categories. *SUT Journal of Mathematics*, **51**, 99-117.

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

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

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

Cressie, N. A. C. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Ser. B*, **46**, 440-464.

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.

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

McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, **65**, 413-418.

Patil, G. P. and Taillie, C. (1982). Diversity as a concept and its measurement. *Journal of the American Statistical Association*, **77**, 548-561.

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

Read, T. R. C. and Cressie, N. A. C. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. New York: Springer.

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

Tahata, K., Ando, S. and Tomizawa, S. (2011). Ridit score type asymmetry model and decomposition of symmetry for square contingency tables. *Model Assisted Statistics and Applications*, **6**, 279-286.

Tahata, K., Iwashita, T. and Tomizawa, S. (2006). Measure of departure from symmetry of cumulative marginal probabilities for square contingency tables with ordered categories. *SUT Journal of Mathematics*, **42**, 7-29.

Tahata, K., Miyazawa, K. and Tomizawa, S. (2010). Measure of departure from average cumulative symmetry for square contingency tables with ordered categories. *American Journal of Biostatistics*, **1**, 62-66.

Tahata, K., and Tomizawa, S. (2009). Decomposition of symmetry using two-ratios-parameter symmetry model and orthogonality for square contingency tables. *Journal of Statistics: Advances in Theory and Applications*, **1**, 19-33.

Tahata, K., Yamamoto, K., Nagatani, N. and Tomizawa, S. (2009). A measure of departure from average symmetry for square contingency tables with ordered categories. *Austrian Journal of Statistics*, **38**, 101-108.

Tomizawa, S. (1984). Three kinds of decompositions for the conditional symmetry model in a square contingency table. *Journal of the Japan Statistical Society*, **14**, 35-42.

Tomizawa, S. (1987). Decompositions for 2-ratios-parameter symmetry model in square contingency tables with ordered categories. *Biometrical Journal*, **29**, 45-55.

Tomizawa, S. (1994). Two kinds of measures of departure from symmetry in square contingency tables having nominal categories. *Statistica Sinica*, **4**, 325-334.

Tomizawa, S. (1995a). Measure of departure from marginal homogeneity for contingency tables with nominal categories. *Journal of the Royal Statistical Society, Ser. D; The Statistician*, **44**, 425-439.

Tomizawa, S. (1995b). A generalization of the marginal homogeneity model for square contingency tables with ordered categories. *Journal of Educational and Behavioral Statistics*, **20**, 349-360.

Tomizawa, S., Seo, T. and Yamamoto, H. (1998). Power-divergence-type measure of departure from symmetry for square contingency tables that have nominal categories. *Journal of Applied Statistics*, **25**, 387-398.

Tomizawa, S. and Makii, T. (2001). Generalized measures of departure from marginal homogeneity for contingency tables with nominal categories. *Journal of Statistical Research*, **35**, 1-24.

Tomizawa, S., Miyamoto, N. and Ashihara, N. (2003). Measure of departure from marginal homogeneity for square contingency tables having ordered categories. *Behaviormetrika*, **30**, 173-193.

Tomizawa, S., Miyamoto, N. and Hatanaka, Y. (2001). Measure of asymmetry for square contingency tables having ordered categories. *The Australian and New Zealand Journal of Statistics*, **43**, 335-349.

Yamamoto, K., Ando, S. and Tomizawa, S. (2011). A measure of departure from average marginal homogeneity for square contingency tables with ordered categories. *Revstat*, **9**, 115-126.

Yamamoto, H., Iwashita, T., and Tomizawa, S. (2007). Decomposition of symmetry into ordinal quasi-symmetry and marginal equimoment for multi-way tables. *Austrian Journal of Statistics*, **36**, 291-306.

Table 1

Cross-classification of smoking by lung cancer. 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 frequency (namely, $n_{\cdot j} = \sum_{i=1}^{2} n_{ij}$).

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

Table 2

Cross-classification of the unaided vision data. Note that n_{ij} denote the observed frequency in the (i, j) cell of the table $(i = 1, \ldots, 4; j = 1, \ldots, 4)$, $n_{\cdot 1}$ and $n_{\cdot 2}$ denote the marginal observed frequency (namely, $n_{i\cdot} = \sum_{j=1}^{4} n_{ij}$ and $n_{\cdot j} = \sum_{i=1}^{4} n_{ij}$), and N denote sample size.

Right eye grade	Left eye grade				Total
	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

Acknowledgments

I would like to deeply my sincere thanks to Professor S. Tomizawa, for his research guidance, many advices and very encouragement. He was very kind in carefully reading the article and giving many useful suggestions to improve the paper. In addition, I am grateful to Dr. K. Tahata of Tokyo University of Science, Dr. K. Yamamoto of Osaka University, Dr. K. Iki of Tokyo University of Science and Dr. H. Kurakami of Tokyo University of Science for their helpful comments and supports.