SUT Journal of Mathematics
Vol. 42, No. 1 (2006), 7–29

Measure of departure from symmetry of cumulative marginal probabilities for square contingency tables with ordered categories

Kouji Tahata, Toshiya Iwashita and Sadao Tomizawa

(Received December 16, 2005)

Dedicated to Professor Minoru Siotani on his 80th birthday

Abstract. For the analysis of square contingency tables with ordered categories, Tomizawa, Miyamoto and Ashihara (2003) considered the measure which represents the degree of departure from the marginal homogeneity (MH) model and does not depend on the diagonal probabilities in the table. This paper proposes another measure which represents the degree of departure from MH and depends on the diagonal probabilities. The measure proposed is expressed by using the Cressie-Read power-divergence or Patil-Taillie diversity index, which is applied for the cumulative marginal probabilities that an observation will fall in row (column) category i or below [or in row (column) category $i+1$ or above]. The measure is useful for seeing how far the cumulative marginal probabilities are distant from those with a MH structure, and for comparing the degree of departure from MH in several tables. Examples are given.

AMS 2000 *Mathematics Subject Classification.* 62H17.

Key words and phrases. Category ordering, cumulative marginal probabilities, marginal homogeneity, measure, power-divergence, square contingency table.

§1. Introduction

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 ($i = 1, 2, \dots, R; j = 1, 2, \dots, R$), and let X and Y denote the row and column variables, respectively. The marginal homogeneity (MH) model is defined by

$$\Pr(X = i) = \Pr(Y = i) \qquad \text{for} \quad i = 1, 2, \dots, R,$$

namely

$$p_{i\cdot} = p_{\cdot i} \qquad \text{for} \quad i = 1, 2, \dots, R,$$

where $p_{i\cdot} = \sum_{t=1}^{R} p_{it}$ and $p_{\cdot i} = \sum_{s=1}^{R} p_{si}$ (Stuart, 1955). This model indicates that the row marginal distribution is identical with the column marginal distribution. This model may be expressed as

$$\Pr(X = i | X \neq Y) = \Pr(Y = i | X \neq Y) \qquad \text{for} \quad i = 1, 2, \ldots, R,$$

namely

$$p^c_{i\cdot} = p^c_{\cdot i} \qquad \text{for} \quad i = 1, 2, \ldots, R,$$

where

$$p^c_{i\cdot} = (p_{i\cdot} - p_{ii})/\delta, \; p^c_{\cdot i} = (p_{\cdot i} - p_{ii})/\delta \text{ and } \delta = \sum_{i \neq j}\sum p_{ij}.$$

This states that the conditional row marginal distribution is identical with the conditional column marginal distribution, given that an observation will fall in one of the off-diagonal cells of the table.

Let F^X_i and F^Y_i denote the cumulative marginal probabilities of X and Y, respectively; those are $F^X_i = \Pr(X \leq i) = \sum_{k=1}^{i} p_{k\cdot}$ and $F^Y_i = \Pr(Y \leq i) = \sum_{k=1}^{i} p_{\cdot k}$ for $i = 1, 2, \ldots, R-1$. Then the MH model may also be expressed as

$$F^X_i = F^Y_i \qquad \text{for} \quad i = 1, 2, \ldots, R-1.$$

This states that the row cumulative marginal distribution is identical with the column cumulative marginal distribution. Then, by considering the difference between the cumulative marginal probabilities, $F^X_i - F^Y_i$ for $i = 1, 2, \ldots, R-1$, we see that the MH model may further be expressed as

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

where

$$G_{1(i)} = \sum_{s=1}^{i} \sum_{t=i+1}^{R} p_{st} = \Pr(X \leq i, Y \geq i+1),$$

and

$$G_{2(i)} = \sum_{s=i+1}^{R} \sum_{t=1}^{i} p_{st} = \Pr(X \geq i+1, Y \leq i).$$

Namely, this model 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 square contingency tables with *nominal* categories, Tomizawa (1995) proposed the measure to represent the degree of departure from MH, which are expressed by using the Kullback-Leibler information (or the Shannon entropy) and the Pearson χ^2-type discrepancy (or the Gini concentration); namely, (i) two kinds of measures (denoted by $\Psi^{(0)}$ and $\Psi^{(1)}$) being functions of $\{p_{i\cdot}\}$ and $\{p_{\cdot i}\}$, and (ii) two kinds of measures (denoted by $\Phi^{(0)}$ and $\Phi^{(1)}$) being functions of $\{p^c_{i\cdot}\}$ and $\{p^c_{\cdot i}\}$. Tomizawa and Makii (2001) considered a generalization of Tomizawa's (1995) measures, which is expressed by using Cressie and Read's (1984) power-divergence (or Patil and Taillie's (1982) diversity index); the measures are denoted by $\Psi^{(\lambda)}$ and $\Phi^{(\lambda)}$, $\lambda > -1$, though the details are omitted here. Note that the measure $\Psi^{(\lambda)}$ depends on the diagonal probabilities in the table and the measure $\Phi^{(\lambda)}$ does not depend on the diagonal probabilities. The measures $\Psi^{(\lambda)}$ and $\Phi^{(\lambda)}$ are applied to *nominal* data because those are invariant under arbitrary similar permutations of row and column categories.

For square contingency tables with *ordered* categories, Tomizawa, Miyamoto and Ashihara (2003) proposed the measure to represent the degree of departure from MH. The measure (denoted by $\Gamma^{(\lambda)}$) is a function of the cumulative probabilities $\{G_{1(i)}\}$ and $\{G_{2(i)}\}$, and it is not invariant under arbitrary similar permutations of row and column categories except the reverse order. The measure $\Gamma^{(\lambda)}$ does not depend on the diagonal probabilities.

So we are also interested in a measure (1) which is a function of the cumulative marginal probabilities $\{F_i^X\}$ and $\{F_i^Y\}$, (2) which depends on the diagonal probabilities, and (3) which is applied to the ordinal data; because (i) the MH model indicates that $\{F_i^X\}$ is identical with $\{F_i^Y\}$, (ii) F_i^X (F_i^Y) depend on the diagonal probabilities, and (iii) F_i^X (F_i^Y) are meaningful for the ordinal data.

The purpose of this paper is to propose a measure which represents the degree of departure from MH for square contingency tables with *ordered* categories. The measure proposed is a function of the cumulative marginal probabilities $\{F_i^X\}$ and $\{F_i^Y\}$, and depends on the diagonal probabilities. The measure is applied to square tables with *ordered* categories. It would be useful for seeing how far the cumulative marginal probabilities are distant from those with a MH structure and for comparing the degree of departure from MH in several tables.

§2. Measure of departure from marginal homogeneity

In Sections 2.1 and 2.2, we shall define the two kinds of submeasures to represent the degree of departure from MH (denoted by $\Omega^{(\lambda)}_{M1}$ and $\Omega^{(\lambda)}_{M2}$). In Section 2.3, we shall define the complete measure which represents the degree of departure from MH (denoted by $\Omega^{(\lambda)}_{M}$).

2.1. Submeasure I

For the $R \times R$ square contingency table with ordered categories, let

$$\Delta_1 = \sum_{i=1}^{R-1} (F_i^X + F_i^Y),$$

and

$$F_{1(i)}^* = \frac{F_i^X}{\Delta_1},\ F_{2(i)}^* = \frac{F_i^Y}{\Delta_1},\ Q_{1(i)}^* = \frac{1}{2}(F_{1(i)}^* + F_{2(i)}^*) \qquad \text{for} \quad i = 1, 2, \ldots, R-1.$$

We see that $\{F_{1(i)}^* = F_{2(i)}^* = Q_{1(i)}^*\}$ when the MH model holds. Note that $\sum_{i=1}^{R-1}(F_{1(i)}^* + F_{2(i)}^*) = 1$ and $\sum_{i=1}^{R-1}(2Q_{1(i)}^*) = 1$. Assume that $F_1^X + F_1^Y \neq 0$ (thus, $F_i^X + F_i^Y \neq 0$ for $i = 1, 2, \ldots, R-1$). Consider the submeasure defined by

$$\Omega_{M1}^{(\lambda)} = \frac{\lambda(\lambda+1)}{2^\lambda - 1} I^{(\lambda)} \left(\{F_{1(i)}^*, F_{2(i)}^*\}; \{Q_{1(i)}^*, Q_{1(i)}^*\} \right) \qquad \text{for} \quad \lambda > -1,$$

where

$$I^{(\lambda)}(\cdot;\cdot) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^{R-1} \left[F_{1(i)}^* \left\{ \left(\frac{F_{1(i)}^*}{Q_{1(i)}^*} \right)^\lambda - 1 \right\} + F_{2(i)}^* \left\{ \left(\frac{F_{2(i)}^*}{Q_{1(i)}^*} \right)^\lambda - 1 \right\} \right],$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$. Thus,

$$\Omega_{M1}^{(0)} = \frac{1}{\log 2} I^{(0)} \left(\{F_{1(i)}^*, F_{2(i)}^*\}; \{Q_{1(i)}^*, Q_{1(i)}^*\} \right),$$

where

$$I^{(0)}(\cdot;\cdot) = \sum_{i=1}^{R-1} \left[F_{1(i)}^* \log \left(\frac{F_{1(i)}^*}{Q_{1(i)}^*} \right) + F_{2(i)}^* \log \left(\frac{F_{2(i)}^*}{Q_{1(i)}^*} \right) \right].$$

The $I^{(\lambda)}(\{F_{1(i)}^*, F_{2(i)}^*\}; \{Q_{1(i)}^*, Q_{1(i)}^*\})$ is the power-divergence between $\{F_{1(i)}^*, F_{2(i)}^*\}$ and $\{Q_{1(i)}^*, Q_{1(i)}^*\}$, $i = 1, 2, \ldots, R-1$, and especially, $I^{(0)}(\cdot;\cdot)$ is the Kullback-Leibler information between them. For more details of the power-divergence, see Cressie and Read (1984), and Read and Cressie (1988, p.15). We see that $I^{(\lambda)}(\cdot;\cdot) = 0$ when the MH model holds. Note that a real value λ is chosen by the user.

Let

$$F_{1(i)}^c = \frac{F_i^X}{F_i^X + F_i^Y},\ F_{2(i)}^c = \frac{F_i^Y}{F_i^X + F_i^Y} \qquad \text{for} \quad i = 1, 2, \ldots, R-1.$$

Then $F^c_{1(i)}$ indicates the ratio of the probability that the value of X for an observation is i or below to the sum of the probability that the value of X is i or below and the probability that the value of Y is i or below, and $F^c_{2(i)}$ in a similar manner. Noting that $\{F^c_{1(i)} + F^c_{2(i)} = 1\}$, the MH model may be expressed as

$$F^c_{1(i)} = F^c_{2(i)} \left(= \frac{1}{2}\right) \qquad \text{for} \quad i = 1, 2, \ldots, R-1.$$

So, the MH model also states that the ratio of the probability that the value of X for an observation is i or below to the sum of the probability that the value of X is i or below and the probability that the value of Y is i or below, is equal to the ratio of the probability that the value of Y for the observation is i or below to the same sum of the probabilities. Then the measure $\Omega^{(\lambda)}_{M1}$ may also be expressed as

$$\Omega^{(\lambda)}_{M1} = \frac{\lambda(\lambda+1)}{2^\lambda - 1} \sum_{i=1}^{R-1} (F^*_{1(i)} + F^*_{2(i)}) I^{(\lambda)}_i \left(\left\{ F^c_{1(i)}, F^c_{2(i)} \right\}; \left\{ \frac{1}{2}, \frac{1}{2} \right\} \right),$$

for $\lambda > -1$, where

$$I^{(\lambda)}_i(\cdot;\cdot) = \frac{1}{\lambda(\lambda+1)} \left[F^c_{1(i)} \left\{ \left(\frac{F^c_{1(i)}}{1/2} \right)^\lambda - 1 \right\} + F^c_{2(i)} \left\{ \left(\frac{F^c_{2(i)}}{1/2} \right)^\lambda - 1 \right\} \right],$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$. Thus

$$\Omega^{(0)}_{M1} = \frac{1}{\log 2} \sum_{i=1}^{R-1} (F^*_{1(i)} + F^*_{2(i)}) I^{(0)}_i \left(\left\{ F^c_{1(i)}, F^c_{2(i)} \right\}; \left\{ \frac{1}{2}, \frac{1}{2} \right\} \right),$$

where

$$I^{(0)}_i(\cdot;\cdot) = F^c_{1(i)} \log \left(\frac{F^c_{1(i)}}{1/2} \right) + F^c_{2(i)} \log \left(\frac{F^c_{2(i)}}{1/2} \right).$$

Therefore, for each λ, the $\Omega^{(\lambda)}_{M1}$ would represent essentially the weighted sum of the power-divergence $I^{(\lambda)}_i(\{F^c_{1(i)}, F^c_{2(i)}\}; \{\frac{1}{2}, \frac{1}{2}\})$. The $I^{(\lambda)}_i(\cdot;\cdot)$ indicates how far the $\{F^c_{1(i)}, F^c_{2(i)}\}$ is distant from those with an MH structure, i.e., from $\{\frac{1}{2}, \frac{1}{2}\}$.

Furthermore, the measure $\Omega^{(\lambda)}_{M1}$ may be expressed as

$$\Omega^{(\lambda)}_{M1} = 1 - \frac{\lambda 2^\lambda}{2^\lambda - 1} \sum_{i=1}^{R-1} (F^*_{1(i)} + F^*_{2(i)}) H^{(\lambda)}_i(\{F^c_{1(i)}, F^c_{2(i)}\}) \qquad \text{for} \quad \lambda > -1,$$

where

$$H_i^{(\lambda)}(\cdot) = \frac{1}{\lambda}\left[1 - (F_{1(i)}^c)^{\lambda+1} - (F_{2(i)}^c)^{\lambda+1}\right],$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$. Thus

$$\Omega_{M1}^{(0)} = 1 - \frac{1}{\log 2}\sum_{i=1}^{R-1}(F_{1(i)}^* + F_{2(i)}^*)H_i^{(0)}(\{F_{1(i)}^c, F_{2(i)}^c\}),$$

where

$$H_i^{(0)}(\cdot) = -F_{1(i)}^c \log F_{1(i)}^c - F_{2(i)}^c \log F_{2(i)}^c.$$

The $H_i^{(\lambda)}(\{F_{1(i)}^c, F_{2(i)}^c\})$ is the Patil and Taillie's (1982) diversity index of degree-λ for $\{F_{1(i)}^c, F_{2(i)}^c\}$, which includes the Shannon entropy when $\lambda = 0$. The measure $\Omega_{M1}^{(\lambda)}$ represents essentially the weighted sum of the diversity index $H_i^{(\lambda)}(\{F_{1(i)}^c, F_{2(i)}^c\})$.

Noting that for each λ, the minimum value of $H_i^{(\lambda)}(\{F_{1(i)}^c, F_{2(i)}^c\})$ is 0 when $F_{1(i)}^c = 0$ (then $F_{2(i)}^c = 1$) or $F_{2(i)}^c = 0$ (then $F_{1(i)}^c = 1$), and the maximum value of it is $(2^\lambda - 1)/(\lambda 2^\lambda)$ (if $\lambda \neq 0$), $\log 2$ (if $\lambda = 0$), when $F_{1(i)}^c = F_{2(i)}^c$, we see that the measure $\Omega_{M1}^{(\lambda)}$ must lie between 0 and 1. Also for each λ (> -1), (i) there is a structure of MH in the $R \times R$ table (i.e., $F_{1(i)}^c = F_{2(i)}^c = 1/2$ (thus $F_i^X = F_i^Y$), for all $i = 1, 2, \ldots, R-1$) if and only if $\Omega_{M1}^{(\lambda)} = 0$, and (ii) the degree of departure from MH is the largest, in the sense that $F_{1(i)}^c = 0$ (then $F_{2(i)}^c = 1$) or $F_{2(i)}^c = 0$ (then $F_{1(i)}^c = 1$) [i.e., $F_i^X = 0$ (then $F_i^Y \neq 0$) or $F_i^Y = 0$ (then $F_i^X \neq 0$)] for all $i = 1, 2, \ldots, R-1$, if and only if $\Omega_{M1}^{(\lambda)} = 1$ (namely, the ratio of the probability that the value of X for an observation is i or below to the sum of the probability that the value of X is i or below and the probability that the value of Y is i or below, is equal to 0 or 1 for all $i = 1, 2, \ldots, R-1$).

According to the weighted sum of the power-divergence or the weighted sum of the Patil-Taillie diversity index, $\Omega_{M1}^{(\lambda)}$ represents the degree of the departure from MH, and the degree increases as the value of $\Omega_{M1}^{(\lambda)}$ increases.

2.2. Submeasure II

Let S_i^X and S_i^Y denote the reverse cumulative marginal probabilities of X and Y, respectively, defined by $S_i^X = \Pr(X \geq i+1) = \sum_{k=i+1}^{R} p_{k\cdot}$ and $S_i^Y = \Pr(Y \geq i+1) = \sum_{k=i+1}^{R} p_{\cdot k}$ for $i = 1, 2, \ldots, R-1$. These are the cumulative marginal probabilities which are taken in reverse order of categories; thus,

$$S_i^X = 1 - F_i^X, \; S_i^Y = 1 - F_i^Y \qquad \text{for} \quad i = 1, 2, \ldots, R-1.$$

Then the MH model may further be expressed as

$$S_i^X = S_i^Y \qquad \text{for} \quad i = 1, 2, \dots, R-1.$$

Let

$$\Delta_2 = \sum_{i=1}^{R-1} (S_i^X + S_i^Y),$$

and

$$S_{1(i)}^* = \frac{S_i^X}{\Delta_2}, \; S_{2(i)}^* = \frac{S_i^Y}{\Delta_2}, \; Q_{2(i)}^* = \frac{1}{2}(S_{1(i)}^* + S_{2(i)}^*) \qquad \text{for} \quad i = 1, 2, \dots, R-1.$$

We see that $\{S_{1(i)}^* = S_{2(i)}^* = Q_{2(i)}^*\}$ when the MH model holds. Note that $\sum_{i=1}^{R-1}(S_{1(i)}^* + S_{2(i)}^*) = 1$ and $\sum_{i=1}^{R-1}(2Q_{2(i)}^*) = 1$. Assuming that $S_{R-1}^X + S_{R-1}^Y \neq 0$ (thus $S_i^X + S_i^Y \neq 0$ for $i = 1, 2, \dots, R-1$), we shall define the submeasure $\Omega_{M2}^{(\lambda)}$ (for $\lambda > -1$), which represents the degree of departure from MH, by $\Omega_{M1}^{(\lambda)}$ with $\{F_{1(i)}^*\}$, $\{F_{2(i)}^*\}$, and $\{Q_{1(i)}^*\}$ replaced by $\{S_{1(i)}^*\}$, $\{S_{2(i)}^*\}$, and $\{Q_{2(i)}^*\}$, respectively.

2.3. Measure for marginal homogeneity

We shall define the complete measure which represents the degree of departure from MH.

Assume that $F_1^X + F_1^Y \neq 0$ and $S_{R-1}^X + S_{R-1}^Y \neq 0$ (thus $F_i^X + F_i^Y \neq 0$ and $S_i^X + S_i^Y \neq 0$ for $i = 1, 2, \cdots, R-1$). Consider a measure defined by

$$\Omega_M^{(\lambda)} = \frac{1}{2}\left(\Omega_{M1}^{(\lambda)} + \Omega_{M2}^{(\lambda)}\right) \qquad \text{for} \quad \lambda > -1,$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$. Thus

$$\Omega_M^{(0)} = \frac{1}{2}\left(\Omega_{M1}^{(0)} + \Omega_{M2}^{(0)}\right).$$

We obtain the following theorem although the proof is omitted.

Theorem 1. *For each λ,*

(i) $0 \leq \Omega_M^{(\lambda)} \leq 1$,

(ii) $\Omega_M^{(\lambda)} = 0$ *if and only if there is a structure of MH in the $R \times R$ table,*

(iii) $\Omega_M^{(\lambda)} = 1$ *if and only if the degree of departure from MH is the largest, in the sense that $F_i^X = 0$ (then $S_i^X = 1$) and $F_i^Y = 1$ (then $S_i^Y = 0$), or $F_i^X = 1$ (then $S_i^X = 0$) and $F_i^Y = 0$ (then $S_i^Y = 1$), for arbitrary cut point i ($i = 1, 2, \dots, R-1$).*

We point out that $\Omega_M^{(\lambda)} = 1$ indicates that the cell probability p_{R1} is 1 and other cell probabilities are 0 or the cell probability p_{1R} is 1 and other cell probabilities are 0. Thus, $\Omega_M^{(\lambda)} = 1$ indicates that $p_{R\cdot} = 1$ and $p_{\cdot 1} = 1$ (thus $p_{1\cdot} = \cdots = p_{R-1\cdot} = 0$ and $p_{\cdot 2} = \cdots = p_{\cdot R} = 0$) or $p_{1\cdot} = 1$ and $p_{\cdot R} = 1$ (thus $p_{2\cdot} = \cdots = p_{R\cdot} = 0$ and $p_{\cdot 1} = \cdots = p_{\cdot R-1} = 0$). So, this indicates that $\Pr(X \le i) = 0$ and $\Pr(Y \le i) = 1$ for $i = 1, 2, \ldots, R-1$, or $\Pr(X \le i) = 1$ and $\Pr(Y \le i) = 0$ for $i = 1, 2, \ldots, R-1$.

§3. Approximate confidence interval for measure

Let n_{ij} denote the observed frequency in the ith row and jth column of the table ($i = 1, 2, \ldots, R; j = 1, 2, \ldots, R$). Assuming that a multinomial distribution applies to the $R \times R$ table, we shall consider an approximate standard error and large-sample confidence interval for the measure $\Omega_M^{(\lambda)}$, using the delta method, as described by Bishop, Fienberg and Holland (1975, Section 14.6) and Agresti (1990, Section 12.1). The sample version of $\Omega_M^{(\lambda)}$, i.e., $\widehat{\Omega}_M^{(\lambda)}$, is given by $\Omega_M^{(\lambda)}$ with $\{p_{ij}\}$ replaced by $\{\widehat{p}_{ij}\}$, where $\widehat{p}_{ij} = n_{ij}/n$ and $n = \sum\sum n_{ij}$. Using the delta method, we obtain the following theorem.

Theorem 2. $\sqrt{n}(\widehat{\Omega}_M^{(\lambda)} - \Omega_M^{(\lambda)})$ *has asymptotically a normal distribution with mean zero and variance* $\sigma^2[\widehat{\Omega}_M^{(\lambda)}]$*, where* $\sigma^2[\widehat{\Omega}_M^{(\lambda)}]$ *is given in Appendix.*

We note that the asymptotic distribution of $\sqrt{n}(\widehat{\Omega}_M^{(\lambda)} - \Omega_M^{(\lambda)})$ is not applicable when $\Omega_M^{(\lambda)} = 0$ and $\Omega_M^{(\lambda)} = 1$ because then $\sigma^2[\widehat{\Omega}_M^{(\lambda)}] = 0$. Let $\widehat{\sigma}^2[\widehat{\Omega}_M^{(\lambda)}]$ denote $\sigma^2[\widehat{\Omega}_M^{(\lambda)}]$ with $\{p_{ij}\}$ replaced by $\{\widehat{p}_{ij}\}$. Then $\widehat{\sigma}[\widehat{\Omega}_M^{(\lambda)}]/\sqrt{n}$ is an estimated approximate standard error for $\widehat{\Omega}_M^{(\lambda)}$, and $\widehat{\Omega}_M^{(\lambda)} \pm z_{p/2}\widehat{\sigma}[\widehat{\Omega}_M^{(\lambda)}]/\sqrt{n}$ is an approximate $100(1-p)$ percent confidence interval for $\Omega_M^{(\lambda)}$, where $z_{p/2}$ is the percentage point from the standard normal distribution corresponding to a two-tail probability equal to p.

§4. Comparison between measures

First, we shall compare the measures $\Omega_M^{(\lambda)}$ and $\Psi^{(\lambda)}$ ($\Phi^{(\lambda)}$) (see Tomizawa and Makii (2001) for $\Psi^{(\lambda)}$ ($\Phi^{(\lambda)}$)). Consider the artificial data in Table 1a, and their modified data in Table 1b, which are obtained by interchanging categories 1, 2, and 3. Then we can see from Table 2 that for each λ, (i) the values of $\widehat{\Psi}^{(\lambda)}$ ($\widehat{\Phi}^{(\lambda)}$) for Table 1a are theoretically equal to the corresponding values for Table 1b, but (ii) the value of $\widehat{\Omega}_M^{(\lambda)}$ is greater for Table 1a than for Table 1b.

Generally, (i) the measure $\Omega_M^{(\lambda)}$ is not invariant under arbitrary similar permutations of row and column categories (except the reverse order), but (ii) the measure $\Psi^{(\lambda)}$ ($\Phi^{(\lambda)}$) is invariant under them. If the data in Tables 1a and 1b are on a *nominal* scale, then it would be natural to conclude that the degree of departure from MH for Table 1a is equal to that for Table 1b because the pairs of counts in the marginal same row and column categories of the tables are the same for Tables 1a and 1b. On the other hand, if the data in Tables 1a and 1b are on an *ordinal* scale and if we want to utilize the information about the category ordering, then it seems natural to conclude that the degree of departure from MH is different between Tables 1a and 1b and it is greater for Table 1a rather than for Table 1b, because from the comparison between Tables 1c and 1d (also from that between Tables 1e and 1f), it seems that the degree of departure from MH (i.e., from $F_i^X = F_i^Y$ and $S_i^X = S_i^Y$ for $i = 1, 2, 3$) is different between Tables 1a and 1b and the degree is greater for Table 1a rather than for Table 1b.

Therefore we conclude that it is suitable to use the measure $\Psi^{(\lambda)}$ ($\Phi^{(\lambda)}$) for analyzing the data on a *nominal* scale and also it may be possible to use $\Psi^{(\lambda)}$ ($\Phi^{(\lambda)}$) for analyzing the data on an *ordinal* scale since it only requires a categorical scale. When used for analyzing the data on an *ordinal* scale, however, $\Psi^{(\lambda)}$ ($\Phi^{(\lambda)}$) does not use the information about the category ordering. Therefore, for the data on an *ordinal* scale, the measure $\Omega_M^{(\lambda)}$ rather than $\Psi^{(\lambda)}$ ($\Phi^{(\lambda)}$) should be used when one wants to use the information about that ordering.

We note that it is dangerous to use the measure $\Omega_M^{(\lambda)}$ for analyzing the data on a *nominal* scale because the $\Omega_M^{(\lambda)}$ is not invariant under arbitrary similar permutations of row and column categories.

Secondly, we shall compare the measures $\Omega_M^{(\lambda)}$ and $\Gamma^{(\lambda)}$ (see Tomizawa, Miyamoto and Ashihara (2003) for $\Gamma^{(\lambda)}$). Consider the artificial data in Table 3. The values of observations for the off-diagonal cells are the same for Tables 3a, 3b and 3c. Thus it is easily seen that $\{\widehat{G}_{1(i)}\}$ and $\{\widehat{G}_{2(i)}\}$ for Table 3a are equal to those for Table 3b and 3c, but $\{\widehat{F}_i^X\}$ and $\{\widehat{F}_i^Y\}$ for Table 3a are not equal to those for Table 3b and 3c. From Table 3d, we see that the values of $\widehat{\Gamma}^{(\lambda)}$ are the same for Tables 3a, 3b, and 3c, but the values of $\widehat{\Omega}_M^{(\lambda)}$ are not the same for those data. In addition, from Tables 3a, 3b, 3c, and 3d, we see that the value of $\widehat{\Omega}_M^{(\lambda)}$ becomes closer to the value of $\widehat{\Gamma}^{(\lambda)}$ as the observed proportions on the main diagonal decrease. So, it seems that the values of $\widehat{\Omega}_M^{(\lambda)}$ and $\widehat{\Gamma}^{(\lambda)}$ are markedly different when the observed proportions on the main diagonal are great. Because the measure $\Gamma^{(\lambda)}$ does not depend on the main diagonal probabilities but the measure $\Omega_M^{(\lambda)}$ depends on the main diagonal probabilities.

The measure $\Omega_M^{(\lambda)}$ is useful for seeing how far the cumulative marginal probabilities $\{F_i^X\}$ and $\{F_i^Y\}$ are distant from those with the MH structure (though the measure $\Gamma^{(\lambda)}$ is useful for seeing how far the cumulative probabilities $\{G_{1(i)}\}$ and $\{G_{2(i)}\}$ are distant from those with the MH structure).

Moreover, we compare the cases of $\Omega_M^{(\lambda)} = 1$ and $\Gamma^{(\lambda)} = 1$. As shown in Section 2, $\Omega_M^{(\lambda)} = 1$ indicates that the degree of asymmetry is the largest in the sense that $F_i^X = 0$ (then $S_i^X = 1$) and $F_i^Y = 1$ (then $S_i^Y = 0$), or $F_i^X = 1$ (then $S_i^X = 0$) and $F_i^Y = 0$ (then $S_i^Y = 1$), for arbitrary cut point i ($i = 1, 2, \ldots, R-1$). On the other hand, $\Gamma^{(\lambda)} = 1$ indicates that the degree of asymmetry is the largest in the sense that $G_{1(i)}^c = 0$ (then $G_{2(i)}^c = 1$), or $G_{2(i)}^c = 0$ (then $G_{1(i)}^c = 1$) for all $i = 1, 2, \ldots, R-1$, where $G_{1(i)}^c = G_{1(i)}/(G_{1(i)} + G_{2(i)})$ and $G_{2(i)}^c = G_{2(i)}/(G_{1(i)} + G_{2(i)})$ (assuming that $G_{1(i)} + G_{2(i)} \neq 0$). The definition of the maximum departure from MH for the measure $\Omega_M^{(\lambda)}$ depends on the main diagonal probabilities. However, the definition of that for the measure $\Gamma^{(\lambda)}$ does not depend on them. Since $\{F_i^X\}$ and $\{F_i^Y\}$ depend on the main diagonal probabilities, when we want to utilize the information on the main diagonal cells, the measure $\Omega_M^{(\lambda)}$ (rather than $\Gamma^{(\lambda)}$) is useful.

§5. Examples

Consider the data in Table 4, taken from Tominaga (1979, p.53). These data describe the cross-classification of father's and son's occupational status categories in Japan which were examined in 1955, 1965 and 1975.

Since the confidence intervals for $\Omega_M^{(\lambda)}$ applied to the data in each of Tables 4a, 4b and 4c do not include zero for all λ (see Table 5), these would indicate that there is not a structure of MH in each table.

When the degree of departure from MH in Tables 4a, 4b and 4c are compared using the confidence interval for $\Omega_M^{(\lambda)}$, it would be greater for Tables 4b and 4c than for Table 4a.

We denote the power-divergence statistic for testing goodness-of-fit of the MH model with $R - 1 = 7$ degrees of freedom by $W_M^{(\lambda)}$. See Cressie and Read (1984) and Read and Cressie (1988, p.15) for details of the power-divergence test statistic. In particular, $W_M^{(0)}$ and $W_M^{(1)}$ are the likelihood ratio and the Pearson's chi-squared statistics, respectively. Table 6 gives the values of $W_M^{(\lambda)}$ applied to the data in Tables 4a, 4b and 4c. The data in each table fit the MH model very poorly.

§6. Discussion

The measure $\Omega_M^{(\lambda)}$ always ranges between 0 and 1 independent of the dimension R and sample size n. Therefore, $\Omega_M^{(\lambda)}$ may be useful for *comparing* the degree of departure from MH in several tables.

As described in Section 2.3, the measure $\Omega_M^{(\lambda)}$ would be useful when we want to see with single summary measure how degree the departure from MH is toward the complete marginal asymmetry of cumulative marginal probabilities. We defined the complete marginal asymmetry, namely, the case of $\Omega_M^{(\lambda)} = 1$, as $\Pr(X \le i) = 0$ and $\Pr(Y \le i) = 1$ for $i = 1, 2, \ldots, R-1$, or $\Pr(X \le i) = 1$ and $\Pr(Y \le i) = 0$ for $i = 1, 2, \ldots, R-1$. This seems natural as the definition of the maximum departure from MH for the data on an *ordinal* scale.

We point out that when one wants to *compare* the degrees of departure from the MH model in several tables, it may be dangerous to use the test statistic such as $W_M^{(\lambda)}$ because it may arise that the value of $\widehat{\Omega}_M^{(\lambda)}$ is greater for table A than for table B, but the value of test statistic is less for table A than for table B. For example, consider the artificial data in Tables 7a and 7b. Then we see from Tables 7c and 7d that, for each λ, the value of $\widehat{\Omega}_M^{(\lambda)}$ is greater for Table 7a than for Table 7b, but the value of $W_M^{(\lambda)}$ is less for Table 7a than for Table 7b. So, like these cases, it would be dangerous to use the test statistic for *comparing* the degrees of departure from the MH model in several tables.

In addition, for several tables, using the measure $\widehat{\Omega}_M^{(\lambda)}$ we can compare how degree the departure from MH is toward the complete marginal asymmetry (defined above), however, using the test statistic $W_M^{(\lambda)}$ we cannot do it.

For analyzing the degree of departure from MH, we first should check whether or not the MH model holds by using a test statistic, such as $W_M^{(\lambda)}$. Then, if it is judged that there is not a structure of MH, the next step would be to measure the degree of departure from MH by using $\widehat{\Omega}_M^{(\lambda)}$. However, if it is judged that there is a structure of MH in the table by the test statistic, then it may be not meaningful to use the measure $\widehat{\Omega}_M^{(\lambda)}$.

Furthermore, we point out that when $\lambda = 0$, the submeasure $\Omega_{M1}^{(0)}$ in the measure $\Omega_M^{(0)}$ can be expressed as

$$\Omega_{M1}^{(0)} = \frac{1}{\log 2} \min_{\{C_{1(i)}, C_{2(i)}\}} I^{(0)} \left(\{F_{1(i)}^*, F_{2(i)}^*\}; \{C_{1(i)}, C_{2(i)}\} \right), \tag{6.1}$$

where

$$I^{(0)}(\cdot;\cdot) = \sum_{i=1}^{R-1} \left[F_{1(i)}^* \log \left(\frac{F_{1(i)}^*}{C_{1(i)}} \right) + F_{2(i)}^* \log \left(\frac{F_{2(i)}^*}{C_{2(i)}} \right) \right],$$

$$C_{1(i)} = C_{2(i)},\ C_{1(i)} \geq 0,\ C_{2(i)} \geq 0,\ \sum_{i=1}^{R-1} \left(C_{1(i)} + C_{2(i)}\right) = 1.$$

Namely, $\Omega_{M1}^{(0)}$ indicates the minimum Kullback-Leibler information between $\{F^*_{1(i)}, F^*_{2(i)}\}$ and $\{C_{1(i)}, C_{2(i)}\}$ with the structure of MH. We note that $\{C_{1(i)}(= C_{2(i)})\}$ minimize $I^{(0)}(\cdot;\cdot)$ in (6.1) when $\{C_{1(i)} = (F^*_{1(i)} + F^*_{2(i)})/2 = Q^*_{1(i)}\}$. In a similar way, the submeasure $\Omega_{M2}^{(0)}$ in the measure $\Omega_{M}^{(0)}$ is expressed. Note that the reader may also be interested in (6.1) with $I^{(0)}(\cdot;\cdot)$ replaced by the power-divergence $I^{(\lambda)}(\cdot;\cdot)$; however, it would be difficult to obtain the value of $\{C_{1(i)}, C_{2(i)}\}$ such that the corresponding power-divergence is a minimum, and also difficult to obtain the maximum value of such a measure.

For the measure $\Omega_{M}^{(\lambda)}$, the analyst may be interested in which value of λ is preferred for a given table. However, it seems difficult to discuss this. It seems to be important and safe that for *comparing* the degrees of departure from MH in several tables, the analyst calculates the values of $\widehat{\Omega}_{M}^{(\lambda)}$ for various values of λ and discusses the degree of departure from MH in terms of them. For example, consider the artificial data in Tables 8a and 8b. Then we see from Table 8c that the value of $\widehat{\Omega}_{M}^{(0)}$ is less for Table 8a than for Table 8b, but the value of $\widehat{\Omega}_{M}^{(1)}$ is greater for Table 8a than for Table 8b (though the differences are slight in these cases). So, for these cases, it may be impossible to decide (by using $\widehat{\Omega}_{M}^{(\lambda)}$) whether the degree of departure from MH is greater for Table 8a or for Table 8b. But generally, for the comparison between two tables, it would be possible to draw a conclusion if $\widehat{\Omega}_{M}^{(\lambda)}$ is always greater (or always less) for one table than for the other table. If the analyst wants to set importance on the interpretation of the measure, the case of $\lambda = 0$, i.e., $\widehat{\Omega}_{M}^{(0)}$ may be recommended in terms of expression (6.1).

Finally we observe that (i) the estimate of the degree of departure from MH should be considered in terms of an approximate confidence interval for the measure $\Omega_{M}^{(\lambda)}$ and not in terms of $\widehat{\Omega}_{M}^{(\lambda)}$ itself, (ii) the measure $\Omega_{M}^{(\lambda)}$ would be useful for describing relative magnitudes (of departure from MH) rather than absolute magnitudes, (iii) $\Omega_{M}^{(\lambda)}$ cannot be used for testing goodness-of-fit of the MH model, and (iv) $\Omega_{M}^{(2)}$ is theoretically equal to $\Omega_{M}^{(1)}$, though the test statistic $W_{M}^{(2)}$ is not always equal to $W_{M}^{(1)}$ (see Table 7).

Acknowledgements

The authors would like to thank a referee for the helpful comments.

Appendix

Using the delta method, $\sqrt{n}(\widehat{\Omega}_M^{(\lambda)} - \Omega_M^{(\lambda)})$ has the asymptotic variance $\sigma^2[\widehat{\Omega}_M^{(\lambda)}]$ as follows:

$$\sigma^2[\widehat{\Omega}_M^{(\lambda)}] = \frac{1}{4}\sum_{i=1}^{R}\sum_{j=1}^{R}\left(w_{ij}^{(\lambda)} + v_{ij}^{(\lambda)}\right)^2 p_{ij},$$

where for $\lambda > -1$ and $\lambda \neq 0$,

$$\begin{aligned} w_{ij}^{(\lambda)} &= \frac{2^\lambda}{\Delta_1(2^\lambda - 1)}\left[\sum_{k=1}^{R-1}\left\{I(i \le k)(F_{1(k)}^c)^\lambda + I(j \le k)(F_{2(k)}^c)^\lambda\right.\right. \\ &\quad + \lambda(F_{1(k)}^c)^\lambda(I(i \le k) - F_{1(k)}^c(I(i \le k) + I(j \le k))) \\ &\quad \left. + \lambda(F_{2(k)}^c)^\lambda(I(j \le k) - F_{2(k)}^c(I(i \le k) + I(j \le k)))\right\} \\ &\quad \left. - (2R - (i+j))\frac{(2^\lambda - 1)\Omega_{M1}^{(\lambda)} + 1}{2^\lambda}\right], \end{aligned}$$

$$\begin{aligned} v_{ij}^{(\lambda)} &= \frac{2^\lambda}{\Delta_2(2^\lambda - 1)}\left[\sum_{k=1}^{R-1}\left\{I(i > k)(S_{1(k)}^c)^\lambda + I(j > k)(S_{2(k)}^c)^\lambda\right.\right. \\ &\quad + \lambda(S_{1(k)}^c)^\lambda(I(i > k) - S_{1(k)}^c(I(i > k) + I(j > k))) \\ &\quad \left. + \lambda(S_{2(k)}^c)^\lambda(I(j > k) - S_{2(k)}^c(I(i > k) + I(j > k)))\right\} \\ &\quad \left. - ((i+j) - 2)\frac{(2^\lambda - 1)\Omega_{M2}^{(\lambda)} + 1}{2^\lambda}\right], \end{aligned}$$

and where for $\lambda = 0$,

$$\begin{aligned} w_{ij}^{(0)} &= \frac{1}{\Delta_1 \log 2}\left[\sum_{k=1}^{R-1}\left\{I(i \le k)\log(F_{1(k)}^c) + I(j \le k)\log(F_{2(k)}^c)\right\}\right. \\ &\quad \left. - (2R - (i+j))(\log 2)(\Omega_{M1}^{(0)} - 1)\right], \end{aligned}$$

$$\begin{aligned} v_{ij}^{(0)} &= \frac{1}{\Delta_2 \log 2}\left[\sum_{k=1}^{R-1}\left\{I(i > k)\log(S_{1(k)}^c) + I(j > k)\log(S_{2(k)}^c)\right\}\right. \\ &\quad \left. - ((i+j) - 2)(\log 2)(\Omega_{M2}^{(0)} - 1)\right], \\ S_{1(k)}^c &= \frac{S_k^X}{S_k^X + S_k^Y},\ S_{2(k)}^c = \frac{S_k^Y}{S_k^X + S_k^Y}, \end{aligned}$$

and $I(\cdot)$ is the indicator function, $I(\cdot) = 1$ if true, 0 if not.

References

[1] Agresti, A. (1990). *Categorical Data Analysis.* John Wiley, New York.

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

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

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

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

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

[7] Tominaga, K. (1979). *Nippon no Kaisou Kouzou (Japanese Hierarchical Structure).* University of Tokyo Press, Tokyo, (in Japanese).

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

[9] 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.

[10] 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.

Table 1: Artificial data (Tables 1a and 1b) and the corresponding values of $\{n\widehat{F}_i^X\}$, $\{n\widehat{F}_i^Y\}$, $\{n\widehat{S}_i^X\}$ and $\{n\widehat{S}_i^Y\}$ (n is sample size)

(a) $n = 1539$

	Y				
X	(1)	(2)	(3)	(4)	Total
(1)	200	170	150	90	610
(2)	11	180	109	60	360
(3)	25	4	160	230	419
(4)	4	5	1	140	150
Total	240	359	420	520	1539

(b) $n = 1539$

	Y				
X	(1)	(2)	(3)	(4)	Total
(1)	180	109	11	60	360
(2)	4	160	25	230	419
(3)	170	150	200	90	610
(4)	5	1	4	140	150
Total	359	420	240	520	1539

(c) Values of $\{n\widehat{F}_i^X\}$ and $\{n\widehat{F}_i^Y\}$ for Table 1a

i	1	2	3
$n\widehat{F}_i^X$	610	970	1389
$n\widehat{F}_i^Y$	240	599	1019

(d) Values of $\{n\widehat{F}_i^X\}$ and $\{n\widehat{F}_i^Y\}$ for Table 1b

i	1	2	3
$n\widehat{F}_i^X$	360	779	1389
$n\widehat{F}_i^Y$	359	779	1019

(e) Values of $\{n\widehat{S}_i^X\}$ and $\{n\widehat{S}_i^Y\}$ for Table 1a

i	1	2	3
$n\widehat{S}_i^X$	929	569	150
$n\widehat{S}_i^Y$	1299	940	520

(f) Values of $\{n\widehat{S}_i^X\}$ and $\{n\widehat{S}_i^Y\}$ for Table 1b

i	1	2	3
$n\widehat{S}_i^X$	1179	760	150
$n\widehat{S}_i^Y$	1180	760	520

Table 2: Values of $\widehat{\Omega}_M^{(\lambda)}$, $\widehat{\Psi}^{(\lambda)}$ and $\widehat{\Phi}^{(\lambda)}$ applied to Tables 1a and 1b

Values of λ	For Table 1a			For Table 1b		
	$\widehat{\Omega}_M^{(\lambda)}$	$\widehat{\Psi}^{(\lambda)}$	$\widehat{\Phi}^{(\lambda)}$	$\widehat{\Omega}_M^{(\lambda)}$	$\widehat{\Psi}^{(\lambda)}$	$\widehat{\Phi}^{(\lambda)}$
0	0.054	0.090	0.337	0.022	0.090	0.337
0.6	0.068	0.112	0.373	0.027	0.112	0.373
1	0.072	0.119	0.381	0.029	0.119	0.381
1.8	0.073	0.120	0.383	0.029	0.120	0.383

Table 3: Artificial data (Tables 3a, 3b and 3c) and the corresponding values of $\widehat{\Omega}_M^{(\lambda)}$ and $\widehat{\Gamma}^{(\lambda)}$ (n is sample size)

(a) $n = 7022$

	(1)	(2)	(3)	(4)	Total
(1)	1032	2	8	60	1102
(2)	2	2304	8	58	2372
(3)	3	4	982	46	1035
(4)	4	5	4	2500	2513
Total	1041	2315	1002	2664	7022

(b) $n = 878$

	(1)	(2)	(3)	(4)	Total
(1)	102	2	8	60	172
(2)	2	230	8	58	298
(3)	3	4	92	46	145
(4)	4	5	4	250	263
Total	111	241	112	414	878

(c) $n = 268$

	(1)	(2)	(3)	(4)	Total
(1)	12	2	8	60	82
(2)	2	18	8	58	86
(3)	3	4	12	46	65
(4)	4	5	4	22	35
Total	21	29	32	186	268

(d) Values of $\widehat{\Omega}_M^{(\lambda)}$ and $\widehat{\Gamma}^{(\lambda)}$

	For Table 3a		For Table 3b		For Table 3c	
λ	$\widehat{\Omega}_M^{(\lambda)}$	$\widehat{\Gamma}^{(\lambda)}$	$\widehat{\Omega}_M^{(\lambda)}$	$\widehat{\Gamma}^{(\lambda)}$	$\widehat{\Omega}_M^{(\lambda)}$	$\widehat{\Gamma}^{(\lambda)}$
0	0.0002	0.5544	0.0145	0.5544	0.1648	0.5544
0.6	0.0003	0.6404	0.0187	0.6404	0.2038	0.6404
1	0.0003	0.6619	0.0200	0.6619	0.2155	0.6619
1.8	0.0003	0.6656	0.0203	0.6656	0.2179	0.6656

Table 4: Occupational status for Japanese father-son pairs; from Tominaga (1979, p.53)

(a) Examined in 1955

	Son's status								
Father's status	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total
(1)	36	4	14	7	8	2	3	8	82
(2)	20	20	27	24	11	11	2	11	126
(3)	9	6	23	12	9	5	3	16	83
(4)	15	14	39	81	17	16	11	15	208
(5)	6	7	22	13	72	20	6	13	159
(6)	3	2	5	12	18	19	9	7	75
(7)	5	3	10	11	21	15	38	25	128
(8)	39	30	76	80	69	52	45	614	1005
Total	133	86	216	240	225	140	117	709	1866

(b) Examined in 1965

	Son's status								
Father's status	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total
(1)	27	10	16	3	6	6	1	2	71
(2)	15	38	30	20	8	4	3	7	125
(3)	13	17	32	17	7	16	6	5	113
(4)	12	36	40	132	22	30	13	6	291
(5)	8	22	38	41	91	42	22	9	273
(6)	2	2	7	12	13	16	3	2	57
(7)	3	2	11	11	13	26	30	6	102
(8)	38	44	95	101	132	114	60	309	893
Total	118	171	269	337	292	254	138	346	1925

(c) Examined in 1975

	Son's status								
Father's status	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total
(1)	44	18	28	8	6	8	1	5	118
(2)	15	50	45	20	18	17	4	7	176
(3)	18	25	47	30	24	18	5	7	174
(4)	16	27	53	77	40	29	9	6	257
(5)	18	25	42	31	122	43	17	13	311
(6)	12	15	21	15	36	33	3	8	143
(7)	3	5	8	7	26	21	9	3	82
(8)	44	65	114	92	184	195	58	325	1077
Total	170	230	358	280	456	364	106	374	2338

Table 5: Estimate of $\Omega_M^{(\lambda)}$, estimated approximate standard error for $\widehat{\Omega}_M^{(\lambda)}$, and approximate 95% confidence interval for $\Omega_M^{(\lambda)}$, applied to Tables 4a, 4b and 4c

(a) For Table 4a

Values of λ	Estimated measure	Standard error	Confidence interval
-0.4	0.008	0.001	(0.006, 0.011)
0	0.012	0.002	(0.008, 0.016)
0.6	0.016	0.002	(0.011, 0.020)
1	0.017	0.003	(0.012, 0.022)
1.4	0.017	0.003	(0.012, 0.022)
2	0.017	0.003	(0.012, 0.022)

(b) For Table 4b

Values of λ	Estimated measure	Standard error	Confidence interval
-0.4	0.019	0.002	(0.015, 0.022)
0	0.027	0.003	(0.022, 0.032)
0.6	0.035	0.003	(0.029, 0.041)
1	0.037	0.003	(0.031, 0.044)
1.4	0.038	0.003	(0.031, 0.045)
2	0.037	0.003	(0.031, 0.044)

(c) For Table 4c

Values of λ	Estimated measure	Standard error	Confidence interval
-0.4	0.021	0.002	(0.018, 0.024)
0	0.030	0.002	(0.025, 0.034)
0.6	0.038	0.003	(0.032, 0.044)
1	0.041	0.003	(0.034, 0.047)
1.4	0.042	0.003	(0.035, 0.048)
2	0.041	0.003	(0.034, 0.047)

Table 6: Values of power-divergence test statistic $W_M^{(\lambda)}$ (with 7 degrees of freedom), applied to Tables 4a, 4b and 4c

λ	For Table 4a	For Table 4b	For Table 4c
-0.2	270.21	700.11	822.08
0	260.89	636.53	763.18
0.2	253.13	589.32	717.64
0.6	241.59	527.51	656.03
1	234.43	493.65	622.41
1.8	230.40	474.66	610.05

Table 7: Artificial data (Tables 7a and 7b) and the corresponding values of $\widehat{\Omega}_M^{(\lambda)}$ and the test statistic $W_M^{(\lambda)}$ (n is sample size)

(a) $n = 612$

	(1)	(2)	(3)	(4)	Total
(1)	30	20	15	141	206
(2)	20	60	96	15	191
(3)	10	95	15	20	140
(4)	15	15	15	30	75
Total	75	190	141	206	612

(b) $n = 612$

	(1)	(2)	(3)	(4)	Total
(1)	30	20	15	141	206
(2)	10	95	15	20	140
(3)	20	60	96	15	191
(4)	15	15	15	30	75
Total	75	190	141	206	612

(c) Value of $\widehat{\Omega}_M^{(\lambda)}$

λ	For Table 7a	For Table 7b
-0.2	0.038	0.031
0	0.044	0.036
0.6	0.055	0.046
1	0.059	0.049
1.4	0.060	0.050
2	0.059	0.049

(d) Value of $W_M^{(\lambda)}$

λ	For Table 7a	For Table 7b
-0.2	107.44	132.26
0	104.54	128.56
0.6	99.03	121.14
1	97.55	118.74
1.4	97.51	118.04
2	99.84	119.81

Table 8: Artificial data (Tables 8a and 8b) and the corresponding values of $\widehat{\Omega}_M^{(\lambda)}$ (n is sample size)

(a) $n = 585$

	(1)	(2)	(3)	(4)	Total
(1)	17	71	114	290	492
(2)	15	1	15	7	38
(3)	7	12	6	8	33
(4)	5	7	4	6	22
Total	44	91	139	311	585

(b) $n = 791$

	(1)	(2)	(3)	(4)	Total
(1)	67	71	250	310	698
(2)	15	3	9	7	34
(3)	7	12	6	8	33
(4)	5	7	4	10	26
Total	94	93	269	335	791

(c) Value of $\widehat{\Omega}_M^{(\lambda)}$

λ	For Table 8a	For Table 8b
−0.2	0.3454	0.3462
0	0.3856	0.3862
0.2	0.4156	0.4158
0.6*	0.4535	0.4530
1*	0.4716	0.4707
1.6*	0.4769	0.4758

* indicates that $\widehat{\Omega}_M^{(\lambda)}$ is greater for Table 8a than for Table 8b.

Kouji Tahata
Department of Information Sciences, Faculty of Science and Technology
Tokyo University of Science
Noda City, Chiba, 278-8510, Japan
E-mail: `tahata@jul.rikadai.jp`

Toshiya Iwashita
Department of Liberal Arts, Faculty of Science and Technology
Tokyo University of Science
Noda City, Chiba, 278-8510, Japan
E-mail: `iwashita@rs.noda.tus.ac.jp`

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