SUT Journal of Mathematics
Vol. 51, No. 1 (2015), 99–117

Measure of departure from generalized marginal homogeneity model for square contingency tables with ordered categories

Shuji Ando, Kouji Tahata and Sadao Tomizawa

(Received October 22, 2014; Revised May 13, 2015)

Abstract. For square contingency tables with ordered categories, Tomizawa (1995a) considered the generalized marginal homogeneity model. The present paper proposes a measure which represents the degree of departure from generalized marginal homogeneity. The proposed measure is expressed by the weighted sum of the Cressie-Read power-divergence or Patil-Taillie diversity index. This measure would be useful for measuring the degree of departure from generalized marginal homogeneity toward the maximum departure from it. The relationship between the measure and the further generalized marginal homogeneity model is shown by simulation studies.

AMS 2010 *Mathematics Subject Classification.* 62H17.

Key words and phrases. Kullback-Leibler information, marginal homogeneity, measure, ordered category, power-divergence, Shannon entropy.

§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 ($i = 1, \ldots, R; j = 1, \ldots, R$).

The marginal homogeneity (MH) model (Stuart, 1955) is defined by

$$p_{i\cdot} = p_{\cdot i} \quad \text{for } i = 1, \ldots, R,$$

where $p_{i\cdot} = \sum_{t=1}^{R} p_{it}$ and $p_{\cdot i} = \sum_{s=1}^{R} p_{si}$.

For square contingency tables with ordered categories, Tomizawa (1984) proposed the extended marginal homogeneity (EMH) model, defined by

$$\delta p_{i\cdot}^{-} + p_{ii} + p_{i\cdot}^{+} = \delta p_{\cdot i}^{-} + p_{ii} + p_{\cdot i}^{+} \quad \text{for } i = 1, \ldots, R,$$

where δ is unspecified and

$$p_{i\cdot}^{-}=\sum_{t=1}^{i-1}p_{it},\quad p_{i\cdot}^{+}=\sum_{t=i+1}^{R}p_{it},\quad p_{\cdot i}^{-}=\sum_{s=i+1}^{R}p_{si},\quad p_{\cdot i}^{+}=\sum_{s=1}^{i-1}p_{si}.$$

A special case of this model obtained by putting $\delta=1$ is the MH model.

Let X and Y denote the row and column variables, respectively. Let

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

and

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

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

Tomizawa (1995a) proposed the generalized marginal homogeneity (GMH) model, defined by

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

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.

For square contingency tables with nominal categories, Tomizawa (1995b) proposed two kinds of measures (based on unconditional marginal probabilities and conditional marginal probabilities) to represent the degree of departure from MH, which are expressed by using the Kullback-Leibler information (or the Shannon entropy) and the Pearson's chi-squared type discrepancy (or the Gini concentration). Tomizawa and Makii (2001) considered generalization of Tomizawa's (1995b) measures, which are expressed by using Cressie and Read's (1984) power-divergence (or Patil and Taillie's (1982) diversity index).

For square contingency tables with ordered categories, measures which represent the degree of departure from MH were proposed by Tomizawa, Miyamoto and Ashihara (2003), and Tahata, Iwashita and Tomizawa (2006). Yamamoto, Furuya and Tomizawa (2007) proposed a measure which represents the degree of departure from EMH. Each of these measures is useful to represent what degree the departure from the corresponding model is toward the maximum departure from it. Note that measures in Tomizawa et al.

(2003), in Tahata et al. (2006), and in Yamamoto et al. (2007) are based on the Cressie-Read power-divergence (including the Kullback-Leibler information as a special case) and Patil-Taillie diversity index (including the Shannon entropy as a special case).

We point out that the test statistic (e.g., Pearson's chi-squared statistic or likelihood ratio statistic) is used for testing the goodness-of-fit of the model and the measure is used for measuring the degree of departure from the model. Also, each of above measures is independent of the dimension R and sample size. The measure may be useful for comparing the degrees of departure from the model toward the maximum departure from it in several tables. Therefore, when the GMH model does not hold, we are interested in measuring the degree of departure from GMH (not the MH and the EMH models) toward the maximum departure from it.

The purpose of the present paper is to propose a power-divergence type measure which represents the degree of departure from GMH (although Yamamoto et al. (2007) proposed the power-divergence type measure representing the degree of departure from EMH).

§2. Measure

This section proposes a measure for representing the degree of departure from GMH.

The GMH model may be expressed as

$$c_i = d_i \quad \text{for } \ i = 1, \dots, R-2,$$

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 $i = 1, \dots, R-2$.

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

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

Assume that $c_i + d_i > 0$ for $i = 1, \dots, R-2$, consider a measure defined by

$$\Phi^{(\lambda)} = \frac{\lambda(\lambda+1)}{2(2^\lambda - 1)} \left[I^{(\lambda)}\left(\{c_i\};\{q_i\}\right) + I^{(\lambda)}\left(\{d_i\};\{q_i\}\right) \right] \quad (\lambda > -1). \tag{2.1}$$

Note that (i) $I^{(\lambda)}(\cdot;\cdot)$ is the power-divergence and (ii) if $\lambda \le -1$ in (2.1), 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

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

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

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

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 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

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

We obtain the following theorem.

Theorem 1. *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* $i = 1, \ldots, R-2$*.*

Proof. For each λ (> -1), the minimum value of $H_i^{(\lambda)}$ is 0 when $c_i^* = 0$ (then $d_i^* = 1$) or $d_i^* = 0$ (then $c_i^* = 1$), and the maximum value of it is $(2^\lambda - 1)/(\lambda 2^\lambda)$ (when $\lambda \neq 0$), or $\log 2$ (when $\lambda = 0$) when $c_i^* = d_i^* = \frac{1}{2}$ for $i = 1, \ldots, R-2$. Therefore, we obtain (i) to (iii). □

For analyzing the degree of departure from GMH, we first should check whether or not the GMH model holds by using the test statistic. Then, if it is judged that there is not a structure of GMH, the next step would be to measure the degree of departure from GMH by using the estimated measure $\hat{\Phi}^{(\lambda)}$ (Section 3).

§3. Approximate confidence interval for measure

Let n_{ij} denote the observed frequency in the ith row and jth column of the table ($i = 1, \ldots, R; j = 1, \ldots, R$). 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 $\Phi^{(\lambda)}$, using the delta method, of which descriptions are given by e.g., Bishop, Fienberg and Holland (1975, Sec. 14.6). The sample version of $\Phi^{(\lambda)}$, i.e., $\hat{\Phi}^{(\lambda)}$, is given by $\Phi^{(\lambda)}$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$, where $\hat{p}_{ij} = n_{ij}/n$ and $n = \sum\sum n_{ij}$. Using the delta method, $\sqrt{n}(\hat{\Phi}^{(\lambda)} - \Phi^{(\lambda)})$ has asymptotically (as $n \to \infty$) a normal distribution with mean zero and variance $\sigma^2[\Phi^{(\lambda)}]$. The value of $\sigma^2[\Phi^{(\lambda)}]$ is

$$\sigma^2[\Phi^{(\lambda)}] = \sum\sum_{k<l} \left\{ p_{kl} \left(V_{kl}^{(\lambda)} \right)^2 + p_{lk} \left(W_{lk}^{(\lambda)} \right)^2 \right\},$$

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

$$
\begin{aligned}
V_{kl}^{(\lambda)} &= \frac{2^{\lambda}}{2(2^{\lambda}-1)} \sum_{i=1}^{R-2} \Big\{ (F_{1i(kl)} + F_{3i(kl)}) \left((c_i^*)^{\lambda+1} + (d_i^*)^{\lambda+1} - 1 \right) \\
&\quad + (\lambda+1)(F_{1i(kl)} d_i^* - F_{3i(kl)} c_i^*) \left((c_i^*)^{\lambda} - (d_i^*)^{\lambda} \right) \Big\},
\end{aligned}
$$

$$
\begin{aligned}
W_{lk}^{(\lambda)} &= \frac{2^{\lambda}}{2(2^{\lambda}-1)} \sum_{i=1}^{R-2} \Big\{ (F_{2i(lk)} + F_{4i(lk)}) \left((c_i^*)^{\lambda+1} + (d_i^*)^{\lambda+1} - 1 \right) \\
&\quad + (\lambda+1)(F_{2i(lk)} d_i^* - F_{4i(lk)} c_i^*) \left((c_i^*)^{\lambda} - (d_i^*)^{\lambda} \right) \Big\},
\end{aligned}
$$

and for $\lambda = 0$,

$$
V_{kl}^{(0)} = \frac{1}{2\log 2} \sum_{i=1}^{R-2} \left(F_{1i(kl)} \log c_i^* + F_{3i(kl)} \log d_i^* \right),
$$

$$
W_{lk}^{(0)} = \frac{1}{2\log 2} \sum_{i=1}^{R-2} \left(F_{2i(lk)} \log c_i^* + F_{4i(lk)} \log d_i^* \right),
$$

where

$$
F_{1i(kl)} = \frac{1}{C} \left(\delta(k \le i \le l-1) G_{2(i+1)} - c_i \sum_{j=1}^{R-2} \delta(k \le j \le l-1) G_{2(j+1)} \right),
$$

$$
F_{2i(lk)} = \frac{1}{C} \left(G_{1(i)} \delta(k \le i+1 \le l-1) - c_i \sum_{j=1}^{R-2} G_{1(j)} \delta(k \le j+1 \le l-1) \right),
$$

$$
F_{3i(kl)} = \frac{1}{D} \left(\delta(k \le i+1 \le l-1) G_{2(i)} - d_i \sum_{j=1}^{R-2} \delta(k \le j+1 \le l-1) G_{2(j)} \right),
$$

$$
F_{4i(lk)} = \frac{1}{D} \left(G_{1(i+1)} \delta(k \le i \le l-1) - d_i \sum_{j=1}^{R-2} G_{1(j+1)} \delta(k \le j \le l-1) \right),
$$

and the indicator function $\delta(\cdot)$.

Let $\hat{\sigma}^2[\Phi^{(\lambda)}]$ denote $\sigma^2[\Phi^{(\lambda)}]$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$. Then $\hat{\sigma}[\Phi^{(\lambda)}]/\sqrt{n}$ is an approximate standard error for $\hat{\Phi}^{(\lambda)}$, and $\hat{\Phi}^{(\lambda)} \pm z_{p/2}\hat{\sigma}[\Phi^{(\lambda)}]/\sqrt{n}$ is an approximate $100(1-p)$ percent confidence interval for $\Phi^{(\lambda)}$, where $z_{p/2}$ is the percentage point from the standard normal distribution corresponding to a two-tailed probability equal to p.

We consider comparison between the degrees of departure from GMH in Tables A and B (with sample sizes n_A and n_B). For Tables A and B, the measures $\Phi^{(\lambda)}$ are denoted by $\Phi_A^{(\lambda)}$ and $\Phi_B^{(\lambda)}$, respectively. Then an estimate of the difference between $\Phi_A^{(\lambda)}$ and $\Phi_B^{(\lambda)}$ is given by the sample version difference $\hat{\Phi}_A^{(\lambda)} - \hat{\Phi}_B^{(\lambda)}$. When n_A and n_B are large, this difference has approximately a normal distribution with standard error $\sqrt{\hat{\sigma}^2[\Phi_A^{(\lambda)}]/n_A + \hat{\sigma}^2[\Phi_B^{(\lambda)}]/n_B}$. An approximate large-sample confidence interval for $\Phi_A^{(\lambda)} - \Phi_B^{(\lambda)}$ is $(\hat{\Phi}_A^{(\lambda)} - \hat{\Phi}_B^{(\lambda)}) \pm z_{p/2}\sqrt{\hat{\sigma}^2[\Phi_A^{(\lambda)}]/n_A + \hat{\sigma}^2[\Phi_B^{(\lambda)}]/n_B}$.

§4. Examples

Consider the data in Table 1, taken from Hashimoto (2003, pp. 144-145). These data describe the cross-classification of father's and son's (or daughter's) occupational status categories in Japan which were examined in 1995.

We are interested in applying the GMH model. For the data in Table 1a, this model indicates that for father-son pair, the cumulative probability that the father's occupational status is i or below and his son's occupational status is $i+1$ or above, is $\Delta\Theta^{i-1}$ times higher than the cumulative probability that the father's occupational status is $i+1$ or above and his son's occupational status is i or below, for $i = 1, \ldots, R-1$. When the GMH model is applied to the data in Tables 1a and 1b, the values of likelihood ratio statistic are 16.24 and 169.01, respectively, with 2 degrees of freedom. Therefore, we see that the GMH model fits the data in Tables 1a and 1b poorly.

Next, we shall measure what degree the departure from the GMH model is toward the maximum departure from it, using measure $\Phi^{(\lambda)}$, because it is impossible to measure it by the test statistic. For instance, when $\lambda = 0$, the estimated measures $\hat{\Phi}^{(0)}$ equal 0.0195 and 0.1585 for Tables 1a and 1b, respectively (see Table 2). Thus, the degrees of departure from GMH are estimated to be 1.95 and 15.85 percent of the maximum degree of departure from GMH for Tables 1a and 1b, respectively.

In addition, we shall compare the degrees of departure from GMH in Tables 1a and 1b using the confidence intervals for $\Phi_{1a}^{(\lambda)} - \Phi_{1b}^{(\lambda)}$. For any given λ (> -1), the values in the confidence interval for $\Phi_{1a}^{(\lambda)} - \Phi_{1b}^{(\lambda)}$ are negative (see Table 3). Thus the degree of departure from GMH in Table 1b is greater than that in Table 1a.

§5. Simulation studies

Tahata and Tomizawa (2008) considered the further extension of GMH model (say, EGMH), defined by

$$G_{1(i)} = \Delta \Theta_1^{i-1} \Theta_2^{(i-1)^2} G_{2(i)} \quad \text{for } i = 1, \ldots, R-1.$$

A special case of this model obtained by putting $\Theta_2 = 1$ is equivalent to the GMH model.

In terms of simulation studies, Table 4 gives summary statistics of estimated measure $\hat{\Phi}^{(\lambda)}$ applied to 5×5 square contingency tables of sample size 10000 which are obtained by generating 10000 times by using multinomial distribution random number based on the structure of probability of EGMH model having given parameters. When the GMH model holds, (i.e., EGMH model with $\Theta_2 = 1$), we see from Table 4 that the mean of estimated measure $\hat{\Phi}^{(\lambda)}$ is close to 0. Also, when the GMH model does not hold, (i.e., EGMH model with $\Theta_2 \neq 1$), we see that the mean of estimated measure $\hat{\Phi}^{(\lambda)}$ increases as Θ_2 becomes greater (or smaller) than 1. Although the details are omitted, we can obtain similar results in other given parameters. Therefore, $\hat{\Phi}^{(\lambda)}$ may be appropriate for *measuring* the degree of departure from GMH.

§6. Concluding remarks

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

We shall compare the estimated measure $\hat{\Phi}^{(\lambda)}$ with the test statistic (e.g., likelihood ratio statistic). First, we consider the artificial data in Table 5. When the GMH model is applied to the data in Tables 5a and 5b, the values of likelihood ratio statistic are 9.18 and 91.77, respectively, with 1 degree of freedom. On the other hand, for any fixed λ (> -1), the value of $\hat{\Phi}^{(\lambda)}$ is equal to for Table 5a and for Table 5b (see Table 6). In terms of $\{\hat{c}_i/\hat{d}_i\}$ (see Table 5), it seems natural to conclude that the degree of departure from GMH for Table 5a is equal to that for Table 5b. Therefore, when we want to compare the degrees of departure from GMH in several tables, we should use the estimated measure $\hat{\Phi}^{(\lambda)}$ rather than the test statistic.

Next, we consider the artificial data in Table 7, having same sample size. When the GMH model is applied to the data in Tables 7a and 7b, the values of likelihood ratio statistic are 82.73 and 87.58, respectively, with 1 degree of freedom. On the other hand, for any fixed λ (> -1), the value of $\hat{\Phi}^{(\lambda)}$ is greater for Table 7a than for Table 7b (see Table 8). In terms of $\{\hat{c}_i/\hat{d}_i\}$ (see Table 7), it seems natural to conclude that the degree of departure from

GMH is greater for Table 7a than for Table 7b, because $\{c_i/d_i\}$ are equal to 1 when the GMH model holds. Therefore, $\hat{\Phi}^{(\lambda)}$ would be preferable to the test statistic for measuring the degree of departure from GMH toward the maximum departure from it.

Finally, we point out that the measure $\Phi^{(0)}$ can be expressed as

$$\Phi^{(0)} = \frac{1}{2\log 2}\min_{\{E_i\}}\left[I^{(0)}\left(\{c_i\};\{E_i\}\right) + I^{(0)}\left(\{d_i\};\{E_i\}\right)\right], \tag{6.1}$$

where

$$\sum_{i=1}^{R-2} E_i = 1 \quad \text{and} \quad E_i > 0.$$

We note that q_i in $\Phi^{(\lambda)}$ is the value of E_i such that the sum of Kullback-Leibler (KL) distance (i.e., the KL distance between $\{c_i\}$ and $\{E_i\}$ with a structure of GMH and the KL distance between $\{d_i\}$ and $\{E_i\}$) is a minimum. We note that the readers may be interested in (6.1) with $I^{(0)}$ replaced by the power-divergence $I^{(\lambda)}$; however, it is difficult to obtain the values of $\{E_i\}$ such that the sum of the corresponding power-divergence is a minimum, and difficult to obtain the maximum value of such a measure.

The readers may also be interested in which value of λ is preferred for a given table. However, it would be difficult to discuss it. For the analysis of data, it seems to be important and safe that for measuring the degrees of departure from GMH in several tables, the user calculates the values of $\hat{\Phi}^{(\lambda)}$ for various values of λ and discusses the degree of departure from GMH in terms of them, rather than calculating $\hat{\Phi}^{(\lambda)}$ for only one specified value of λ. However, if the analyst wants to choose one value of λ, the case of $\lambda = 0$, i.e., $\hat{\Phi}^{(0)}$ may be recommended in terms of expression (6.1).

Acknowledgments

The authors would like to thank the editor and an anonymous referee for their helpful suggestions and comments.

References

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

[2] 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.

[3] Hashimoto, K. (2003). *Class Structure in Contemporary Japan*. Melbourne: Trans Pacific Press.

[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. A. C. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. New York: Springer.

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

[7] 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.

[8] Tahata, K. and Tomizawa, S. (2008). Generalized marginal homogeneity model and its relation to marginal equimoments for square contingency tables with ordered categories. *Advances in Data Analysis and Classification*, **2**, 295-311.

[9] 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.

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

[11] Tomizawa, S. (1995b). 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.

[12] 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.

[13] 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.

[14] Yamamoto, K., Furuya, Y. and Tomizawa, S. (2007). Measure of departure from extended marginal homogeneity for square contingency tables with ordered categories. *Revstat*, **5**, 269-283.

Table 1: The cross-classification of father's and son's (or daughter's) occupational status categories in Japan which were examined in 1995 from Hashimoto (2003, pp. 143-144).

(a) For father-son pairs

Father's	Son's status					
status	(1)	(2)	(3)	(4)	(5)	Total
(1)	68	48	36	23	1	176
(2)	33	191	102	33	3	362
(3)	25	147	229	34	2	437
(4)	48	119	146	129	5	447
(5)	40	126	192	82	88	528
Total	214	631	705	301	99	1950

(b) For father-daughter pairs

Father's	Daughter's status					
status	(1)	(2)	(3)	(4)	(5)	Total
(1)	30	8	61	27	1	127
(2)	17	58	186	42	3	306
(3)	18	36	250	44	8	356
(4)	22	35	179	81	8	325
(5)	16	25	207	55	80	383
Total	103	162	883	249	100	1497

Note: Status (1) is Capitalist, (2) New Middle, (3) Working, (4) Self-employed, (5) Farming.

Table 2: Estimate of $\Phi^{(\lambda)}$, approximate standard error for $\hat{\Phi}^{(\lambda)}$, and approximate 95% confidence interval for $\Phi^{(\lambda)}$, applied to Tables 1a and 1b.

(a) For Table 1a

Values of λ	Estimated measure	Standard error	Confidence interval
−0.4	0.0135	0.0069	(0.0001, 0.0270)
0	0.0195	0.0098	(0.0004, 0.0386)
0.6	0.0249	0.0122	(0.0009, 0.0488)
1.0	0.0266	0.0130	(0.0011, 0.0521)
1.6	0.0272	0.0133	(0.0012, 0.0532)

(b) For Table 1b

Values of λ	Estimated measure	Standard error	Confidence interval
−0.4	0.1157	0.0173	(0.0817, 0.1497)
0	0.1585	0.0229	(0.1137, 0.2034)
0.6	0.1928	0.0269	(0.1401, 0.2454)
1.0	0.2027	0.0279	(0.1480, 0.2574)
1.6	0.2058	0.0282	(0.1504, 0.2612)

Table 3: Estimate of difference measure $\Phi_{1a}^{(\lambda)} - \Phi_{1b}^{(\lambda)}$, approximate standard error for $\hat{\Phi}_{1a}^{(\lambda)} - \hat{\Phi}_{1b}^{(\lambda)}$, and approximate 95% confidence interval for $\Phi_{1a}^{(\lambda)} - \Phi_{1b}^{(\lambda)}$, applied to Tables 1a and 1b.

Values of λ	Estimated difference measure	Standard error	Confidence interval
−0.4	−0.1021	0.0187	(−0.1387, −0.0656)
0	−0.1390	0.0249	(−0.1878, −0.0903)
0.6	−0.1679	0.0295	(−0.2258, −0.1100)
1.0	−0.1761	0.0308	(−0.2365, −0.1157)
1.6	−0.1786	0.0312	(−0.2398, −0.1175)

Table 4: Summary statistics of estimated measure $\hat{\Phi}^{(\lambda)}$ applied to 5×5 contingency tables of sample size 10000 which are obtained by generating 10000 times by using multinomial distribution random number based on the structure of probability of EGMH model having given parameters.

(a) $\Delta = 0.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 0.5$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.02451	0.04472	0.05055	0.05073	0.05658	0.08981
0	0.03457	0.06112	0.06862	0.06920	0.07671	0.12150
0.6	0.04329	0.07559	0.08456	0.08533	0.09436	0.14730
1.0	0.04602	0.08018	0.08962	0.09039	0.09990	0.15500
1.6	0.04693	0.08170	0.09131	0.09206	0.10170	0.15740

(b) $\Delta = 0.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 0.67$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.00946	0.02169	0.02481	0.02515	0.02826	0.05653
0	0.01366	0.03108	0.03545	0.03583	0.04018	0.07471
0.6	0.01748	0.03951	0.04488	0.04529	0.05066	0.09354
1.0	0.01873	0.04220	0.04792	0.04830	0.05399	0.09943
1.6	0.01915	0.04309	0.04895	0.04931	0.05512	0.10140

(c) $\Delta = 0.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 1$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.00000	0.00005	0.00013	0.00018	0.00025	0.00179
0	0.00000	0.00008	0.00018	0.00026	0.00037	0.00261
0.6	0.00000	0.00010	0.00024	0.00034	0.00048	0.00336
1.0	0.00000	0.00011	0.00025	0.00036	0.00051	0.00361
1.6	0.00000	0.00011	0.00026	0.00037	0.00053	0.00370

(d) $\Delta = 0.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 1.5$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.03403	0.04252	0.04442	0.04446	0.04636	0.05920
0	0.04886	0.06084	0.06351	0.06356	0.06624	0.08413
0.6	0.06214	0.07711	0.08042	0.08048	0.08381	0.10590
1.0	0.06639	0.08227	0.08579	0.08584	0.08938	0.11270
1.6	0.06782	0.08400	0.08757	0.08763	0.09123	0.11490

(e) $\Delta = 0.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 2$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.07888	0.09460	0.09815	0.09821	0.10160	0.11780
0	0.11070	0.13170	0.13640	0.13640	0.14090	0.16090
0.6	0.13770	0.16250	0.16790	0.16790	0.17300	0.19640
1.0	0.14590	0.17170	0.17730	0.17730	0.18260	0.20700
1.6	0.14850	0.17460	0.18030	0.18030	0.18560	0.21030

Table 4: (Continued.)

(f) $\Delta = 1.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 0.5$

Values of λ	Min	25%	Median	Mean	75%	Max
-0.4	0.03583	0.05503	0.06023	0.06070	0.06602	0.09661
0	0.05157	0.07763	0.08393	0.08430	0.09058	0.13110
0.6	0.06576	0.09711	0.10460	0.10500	0.11240	0.16080
1.0	0.07033	0.10320	0.11110	0.11150	0.11930	0.16990
1.6	0.07188	0.10520	0.11320	0.11370	0.12160	0.17290

(g) $\Delta = 1.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 0.67$

Values of λ	Min	25%	Median	Mean	75%	Max
-0.4	0.01557	0.02547	0.02803	0.02814	0.03066	0.04663
0	0.02258	0.03658	0.04016	0.04028	0.04378	0.06445
0.6	0.02898	0.04655	0.05095	0.05110	0.05544	0.08107
1.0	0.03105	0.04974	0.05441	0.05456	0.05918	0.08630
1.6	0.03176	0.05082	0.05557	0.05572	0.06041	0.08804

(h) $\Delta = 1.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 1$

Values of λ	Min	25%	Median	Mean	75%	Max
-0.4	0.00000	0.00003	0.00007	0.00010	0.00014	0.00104
0	0.00000	0.00004	0.00010	0.00015	0.00020	0.00152
0.6	0.00000	0.00005	0.00013	0.00019	0.00026	0.00196
1.0	0.00000	0.00006	0.00014	0.00020	0.00028	0.00211
1.6	0.00000	0.00006	0.00014	0.00021	0.00029	0.00216

(i) $\Delta = 1.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 1.5$

Values of λ	Min	25%	Median	Mean	75%	Max
-0.4	0.03044	0.03903	0.04083	0.04090	0.04268	0.05305
0	0.04372	0.05586	0.05838	0.05847	0.06096	0.07542
0.6	0.05562	0.07080	0.07393	0.07403	0.07713	0.09495
1.0	0.05943	0.07554	0.07886	0.07897	0.08225	0.10110
1.6	0.06071	0.07713	0.08051	0.08061	0.08395	0.10310

(j) $\Delta = 1.5$, $\Theta_1 = 0.5$ and $\Theta_2 = 2$

Values of λ	Min	25%	Median	Mean	75%	Max
-0.4	0.06882	0.08353	0.08679	0.08692	0.09023	0.10540
0	0.09772	0.11690	0.12120	0.12120	0.12560	0.14370
0.6	0.12290	0.14500	0.14990	0.15010	0.15510	0.17530
1.0	0.13080	0.15360	0.15870	0.15880	0.16410	0.18510
1.6	0.13340	0.15640	0.16150	0.16160	0.16700	0.18830

Table 4: (Continued.)

(k) $\Delta = 0.5$, $\Theta_1 = 1.5$ and $\Theta_2 = 0.5$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.05650	0.07303	0.07676	0.07693	0.08068	0.10040
0	0.07987	0.10260	0.10750	0.10770	0.11250	0.13970
0.6	0.10010	0.12790	0.13370	0.13380	0.13960	0.17280
1.0	0.10640	0.13580	0.14180	0.14190	0.14800	0.18290
1.6	0.10850	0.13840	0.14440	0.14460	0.15070	0.18620

(l) $\Delta = 0.5$, $\Theta_1 = 1.5$ and $\Theta_2 = 0.67$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.02711	0.03493	0.03696	0.03702	0.03896	0.05029
0	0.03902	0.05003	0.05286	0.05293	0.05566	0.07133
0.6	0.04977	0.06347	0.06699	0.06705	0.07044	0.08964
1.0	0.05323	0.06776	0.07148	0.07154	0.07512	0.09536
1.6	0.05440	0.06919	0.07298	0.07304	0.07667	0.09724

(m) $\Delta = 0.5$, $\Theta_1 = 1.5$ and $\Theta_2 = 1$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.00000	0.00002	0.00005	0.00007	0.00010	0.00066
0	0.00000	0.00003	0.00007	0.00011	0.00015	0.00096
0.6	0.00000	0.00004	0.00009	0.00014	0.00019	0.00124
1.0	0.00000	0.00004	0.00010	0.00015	0.00020	0.00133
1.6	0.00000	0.00004	0.00010	0.00015	0.00021	0.00136

(n) $\Delta = 0.5$, $\Theta_1 = 1.5$ and $\Theta_2 = 1.5$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.01972	0.03028	0.03254	0.03266	0.03491	0.04671
0	0.02852	0.04343	0.04655	0.04672	0.04986	0.06602
0.6	0.03654	0.05518	0.05903	0.05921	0.06311	0.08271
1.0	0.03915	0.05893	0.06300	0.06319	0.06730	0.08791
1.6	0.04004	0.06018	0.06432	0.06452	0.06871	0.08962

(o) $\Delta = 0.5$, $\Theta_1 = 1.5$ and $\Theta_2 = 2$

Values of λ	Min	25%	Median	Mean	75%	Max
−0.4	0.04773	0.06406	0.06808	0.06843	0.07240	0.09556
0	0.06799	0.09055	0.09584	0.09604	0.10130	0.12770
0.6	0.08587	0.11340	0.11980	0.11990	0.12620	0.15800
1.0	0.09154	0.12050	0.12720	0.12740	0.13400	0.16800
1.6	0.09343	0.12290	0.12970	0.12980	0.13650	0.17130

Table 5: Artificial data (n is sample size).

(a) $n = 930$

100	100	70	30
10	100	50	30
20	10	100	100
50	40	20	100

Note: $\frac{\hat{c}_1}{\hat{d}_1} = 1.279$, $\frac{\hat{c}_2}{\hat{d}_2} = 0.791$.

(b) $n = 9300$

1000	1000	700	300
100	1000	500	300
200	100	1000	1000
500	400	200	1000

Note: $\frac{\hat{c}_1}{\hat{d}_1} = 1.279$, $\frac{\hat{c}_2}{\hat{d}_2} = 0.791$.

Table 6: The values of $\hat{\Phi}^{(\lambda)}$ applied to Tables 5a and 5b.

Values of λ	Table 5a	Table 5b
-0.4	0.0071	0.0071
0	0.0103	0.0103
0.6	0.0133	0.0133
1.0	0.0143	0.0143
1.6	0.0146	0.0146

Table 7: Artificial data (n is sample size).

(a) $n = 430$

20	100	10	10
10	20	10	10
10	10	20	100
50	10	10	30

Note: $\frac{\hat{c}_1}{\hat{d}_1} = 3.429$, $\frac{\hat{c}_2}{\hat{d}_2} = 0.292$.

(b) $n = 430$

10	100	10	10
10	10	10	10
10	10	10	100
100	10	10	10

Note: $\frac{\hat{c}_1}{\hat{d}_1} = 3.250$, $\frac{\hat{c}_2}{\hat{d}_2} = 0.308$.

Table 8: The values of $\hat{\Phi}^{(\lambda)}$ applied to Tables 7a and 7b.

Values of λ	Table 7a	Table 7b
-0.4	0.1639	0.1517
0	0.2294	0.2129
0.6	0.2842	0.2646
1.0	0.3007	0.2803
1.6	0.3060	0.2853

Shuji Ando
Department of Information Sciences, Faculty of Science and Technology,
Tokyo University of Science
Noda City, Chiba, 278-8510, Japan
E-mail: shuji.ando.0106@gmail.com

Kouji Tahata
Department of Information Sciences, Faculty of Science and Technology,
Tokyo University of Science
Noda City, Chiba, 278-8510, Japan
E-mail: kouji_tahata@is.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