REVSTAT – Statistical Journal
Volume 5, Number 3, November 2007, 269–283

# MEASURE OF DEPARTURE FROM EXTENDED MARGINAL HOMOGENEITY FOR SQUARE CONTINGENCY TABLES WITH ORDERED CATEGORIES

Authors: Kouji Yamamoto
– Graduate School of Sciences and Technology,
Tokyo University of Science,
Noda City, Chiba, 278-8510, Japan
yamamoto_kouji@ymail.plala.or.jp

Yukari Furuya
– Graduate School of Sciences and Technology,
Tokyo University of Science,
Noda City, Chiba, 278-8510, Japan

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

Received: April 2007 Revised: September 2007 Accepted: October 2007

Abstract:

- For the analysis of square contingency tables, Tomizawa and Makii (2001), and Tomizawa, Miyamoto and Ashihara (2003) considered the measures to represent the degree of departure from marginal homogeneity (MH). Tomizawa (1984) considered an extended marginal homogeneity (EMH) model for square tables with ordered categories. This paper proposes a measure to represent the degree of departure from EMH. The measure proposed is expressed by using the Cressie and Read's (1984) power-divergence or Patil and Taillie's (1982) diversity index. The measure would be useful for comparing the degree of departure from EMH in several tables. Examples are given.

Key-Words:

- *cumulative probability; Kullback–Leibler information; marginal homogeneity; measure; ordered category; power-divergence; Shannon entropy; square contingency table.*

AMS Subject Classification:

- 62H17.

## 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 $i$-th row and $j$-th column of the table ($i=1,...,R$; $j=1,...,R$), and let $X$ and $Y$ denote the row and column variables, respectively. The marginal homogeneity (MH) model is defined by

$$\Pr(X=i) = \Pr(Y=i) \qquad \text{for} \;\; i=1,...,R \;,$$

namely

$$p_{i\cdot} = p_{\cdot i} \qquad \text{for} \;\; i=1,...,R \;,$$

where $p_{i\cdot} = \sum_{k=1}^{R} p_{ik}$ and $p_{\cdot i} = \sum_{k=1}^{R} p_{ki}$ (see, e.g., Stuart, 1955; Bhapkar, 1966; Bishop, Fienberg and Holland, 1975, p. 294).

Let

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

and

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

for $i=1,...,R-1$. By considering the difference between the cumulative marginal probabilities, $F_i^X - F_i^Y$ for $i=1,...,R-1$, where $F_i^X = \Pr(X\le i)$ and $F_i^Y = \Pr(Y\le i)$, we see that the MH model may also be expressed as

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

Namely, this 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 $i=1,...,R-1$.

Tomizawa (1984, 1995a) considered the extended marginal homogeneity (EMH) model defined by

$$p_{i\cdot}^{(\delta)} = p_{\cdot i}^{(\delta)} \qquad \text{for} \;\; i=1,...,R \;,$$

where the parameter $\delta$ is unspecified and

$$p_{i\cdot}^{(\delta)} = \delta \sum_{t=1}^{i-1} p_{it} + \sum_{t=i}^{R} p_{it} \;, \qquad p_{\cdot i}^{(\delta)} = \sum_{s=1}^{i} p_{si} + \delta \sum_{s=i+1}^{R} p_{si} \;.$$

Consider the artificial probabilities in Table 1. We see that the EMH model holds with $\delta=2$ in Table 1. The EMH model may also be expressed as

$$G_{1(i)} = \delta\, G_{2(i)} \qquad \text{for} \;\; i=1,...,R-1 \;.$$

**Table 1**: Artificial probabilities having the structure of EMH with $\delta = 2$.

| | | | |
|---|---|---|---|
| 0.04 | 0.02 | 0.04 | 0.26 |
| 0.01 | 0.03 | 0.08 | 0.16 |
| 0.02 | 0.04 | 0.02 | 0.04 |
| 0.13 | 0.08 | 0.02 | 0.01 |

A special case of this model obtained by putting $\delta = 1$ is the MH model. 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 $\delta$ times higher than the cumulative probability that the observation falls in column category $i$ or below and row category $i+1$ or above for $i = 1, ..., R-1$. The EMH model may further be expressed as

$$(1.1) \qquad G^*_{1(i)} = G^*_{2(i)} \qquad \text{for} \quad i = 1, ..., R-1 \ ,$$

where

$$G^*_{1(i)} = G_{1(i)}/G_1 \ , \qquad G^*_{2(i)} = G_{2(i)}/G_2 \ ,$$

$$G_1 = \sum_{i=1}^{R-1} G_{1(i)} \ , \qquad G_2 = \sum_{i=1}^{R-1} G_{2(i)} \ .$$

Namely the EMH model indicates that there is a structure of symmetry between $\{G^*_{1(i)}\}$ and $\{G^*_{2(i)}\}$ for $i = 1, ..., R-1$.

For square contingency tables with *nominal* categories, Tomizawa (1995b) considered two kinds of measures to represent the degree of departure from MH, which are expressed by using the Shannon entropy and Gini concentration. Tomizawa and Makii (2001) considered a generalization of Tomizawa measures, which is expressed by using the Cressie and Read's (1984) power-divergence (or Patil and Taillie's (1982) diversity index). For square contingency tables with *ordered* categories, Tomizawa, Miyamoto and Ashihara (2003) considered a measure to represent the degree of departure from MH.

When the MH model does not hold, these measures would be useful for measuring the degree of departure from MH. When the EMH model does not hold, we are now interested in measuring the degree of departure from EMH (instead of that from MH).

The purpose of this paper is to propose a power-divergence type measure which represents the degree of departure from EMH for square contingency tables with ordered categories. In Section 2 we propose such a measure which is expressed as a function of $\{G^*_{1(i)}\}$ and $\{G^*_{2(i)}\}$. It would be useful for *comparing* the degree of departure from EMH in several tables with ordered categories.

## 2. MEASURE OF DEPARTURE FROM EXTENDED MARGINAL HOMOGENEITY

Assume that $G_1 > 0$, $G_2 > 0$ and $G_{1(i)} + G_{2(i)} > 0$ for $i = 1, ..., R-1$. Let

$$C_i = \frac{G^*_{1(i)} + G^*_{2(i)}}{2} \qquad \text{for} \quad i = 1, ..., R-1 \; .$$

Note that $\sum_{i=1}^{R-1} C_i = 1$. To represent the degree of departure from EMH, consider a measure defined by

$$\Gamma^{(\lambda)}_{\mathrm{EM}} = \frac{\lambda(\lambda+1)}{2\,(2^\lambda - 1)} \Big[ I^{(\lambda)}\Big(\big\{G^*_{1(i)}\big\}; \big\{C_i\big\}\Big) + I^{(\lambda)}\Big(\big\{G^*_{2(i)}\big\}; \big\{C_i\big\}\Big)\Big] \qquad \text{for} \quad \lambda > -1 \; ,$$

where

$$I^{(\lambda)}\big(\{a_i\}; \{b_i\}\big) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^{R-1} a_i \left[ \left(\frac{a_i}{b_i}\right)^{\lambda} - 1 \right] ,$$

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

$$\Gamma^{(0)}_{\mathrm{EM}} = \frac{1}{2 \log 2} \Big[ I^{(0)}\Big(\big\{G^*_{1(i)}\big\}; \big\{C_i\big\}\Big) + I^{(0)}\Big(\big\{G^*_{2(i)}\big\}; \big\{C_i\big\}\Big)\Big] , \tag{2.1}$$

where

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

The $I^{(\lambda)}\big(\{a_i\}; \{b_i\}\big)$ is the power-divergence between $\{a_i\}$ and $\{b_i\}$, and especially $I^{(0)}\big(\{a_i\}; \{b_i\}\big)$ is the Kullback–Leibler information (KL) between them. For more details of the power-divergence $I^{(\lambda)}(\cdot;\cdot)$, see Cressie and Read (1984), and Read and Cressie (1988, p.15). Note that a real value $\lambda$ is chosen by user.

Let

$$G^c_{1(i)} = \frac{G^*_{1(i)}}{G^*_{1(i)} + G^*_{2(i)}} \; , \quad G^c_{2(i)} = \frac{G^*_{2(i)}}{G^*_{1(i)} + G^*_{2(i)}} \qquad \text{for} \quad i = 1, ..., R-1 \; .$$

Note that $\big\{G^c_{1(i)} + G^c_{2(i)} = 1\big\}$. The EMH model can be expressed as

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

Then the measure $\Gamma^{(\lambda)}_{\mathrm{EM}}$ may be expressed as

$$\Gamma^{(\lambda)}_{\mathrm{EM}} = \frac{\lambda(\lambda+1)}{2^\lambda - 1} \sum_{i=1}^{R-1} C_i \, I^{(\lambda)}_i \left( \Big\{G^c_{1(i)}, G^c_{2(i)}\Big\} ; \left\{ \frac{1}{2}, \frac{1}{2} \right\} \right) \qquad \text{for} \quad \lambda > -1 \; , \tag{2.2}$$

where

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

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

$$\Gamma^{(0)}_{\mathrm{EM}} \;=\; \frac{1}{\log 2}\sum_{i=1}^{R-1} C_i\, I_i^{(0)}\left(\left\{G^c_{1(i)}, G^c_{2(i)}\right\};\left\{\frac{1}{2},\frac{1}{2}\right\}\right),$$

where

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

Therefore, $\Gamma^{(\lambda)}_{\mathrm{EM}}$ in equation (2.2) would represent, essentially, the weighted sum of the power-divergence $I_i^{(\lambda)}\big(\{G^c_{1(i)}, G^c_{2(i)}\};\{1/2,1/2\}\big)$.

Moreover, $\Gamma^{(\lambda)}_{\mathrm{EM}}$ may be expressed as

$$\Gamma^{(\lambda)}_{\mathrm{EM}} \;=\; 1-\frac{\lambda\,2^{\lambda}}{2^{\lambda}-1}\sum_{i=1}^{R-1} C_i\, H_i^{(\lambda)}\left(\left\{G^c_{1(i)}, G^c_{2(i)}\right\}\right) \qquad \text{for } \lambda>-1, \tag{2.3}$$

where

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

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

$$\Gamma^{(0)}_{\mathrm{EM}} \;=\; 1-\frac{1}{\log 2}\sum_{i=1}^{R-1} C_i\, H_i^{(0)}\left(\left\{G^c_{1(i)}, G^c_{2(i)}\right\}\right),$$

where

$$H_i^{(0)}(\cdot) \;=\; -G^c_{1(i)}\log G^c_{1(i)} - G^c_{2(i)}\log G^c_{2(i)}\,.$$

Note that $H_i^{(\lambda)}\big(\{G^c_{1(i)}, G^c_{2(i)}\}\big)$ is Patil and Taillie's (1982) diversity index for $\{G^c_{1(i)}, G^c_{2(i)}\}$, which includes the Shannon entropy (when $\lambda=0$) and the Gini concentration (when $\lambda=1$) in special cases. Therefore, $\Gamma^{(\lambda)}_{\mathrm{EM}}$ in equation (2.3) would represent essentially the weighted sum of the diversity index $H_i^{(\lambda)}\big(\{G^c_{1(i)}, G^c_{2(i)}\}\big)$.

Noting that $I_i^{(\lambda)}\big(\{G^c_{1(i)}, G^c_{2(i)}\};\{1/2,1/2\}\big)\ge 0$ and $H_i^{(\lambda)}\big(\{G^c_{1(i)}, G^c_{2(i)}\}\big)\ge 0$, we see that the measure $\Gamma^{(\lambda)}_{\mathrm{EM}}$ must lie between 0 and 1. Also, for each $\lambda$ $(>-1)$,

- **(i)** there is a structure of EMH in the $R\times R$ table if and only if $\Gamma^{(\lambda)}_{\mathrm{EM}}=0$, and
- **(ii)** the degree of departure from EMH is the largest in the sense that $G^c_{1(i)}=0$ (then $G^c_{2(i)}=1$) or $G^c_{2(i)}=0$ (then $G^c_{1(i)}=1$) [namely, $G^*_{1(i)}=0$ (then $G^*_{2(i)}>0$) or $G^*_{2(i)}=0$ (then $G^*_{1(i)}>0$)] for $i=1,\dots,R-1$; if and only if $\Gamma^{(\lambda)}_{\mathrm{EM}}=1$.

Note that $\Gamma^{(\lambda)}_{\mathrm{EM}}=1$ indicates that $G^*_{1(i)}/G^*_{2(i)}=\infty$ for some $i$ and $G^*_{1(i)}/G^*_{2(i)}=0$ for the other $i$, and therefore it seems appropriate to consider that then the degree of departure from EMH (i.e., from $G^*_{1(i)}/G^*_{2(i)}=1$ for $i=1,...,R-1$) is largest. In addition, according to the weighted sum of the power-divergence or the weighted sum of the Patil and Taillie's diversity index, the degree increases as the value of $\Gamma^{(\lambda)}_{\mathrm{EM}}$ increases.

## 3. APPROXIMATE CONFIDENCE INTERVAL FOR MEASURE

Let $n_{ij}$ denote the observed frequency in the $i$-th row and $j$-th column of the table ($i=1,...,R$; $j=1,...,R$). 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 $\Gamma^{(\lambda)}_{\mathrm{EM}}$ using the delta method, descriptions of which are given by Bishop *et al.* (1975, Sec. 14.6) and Agresti (1990, Sec. 12.1). The sample version of $\Gamma^{(\lambda)}_{\mathrm{EM}}$, i.e., $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$, is given by $\Gamma^{(\lambda)}_{\mathrm{EM}}$ 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, $\sqrt{n}\big(\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}-\Gamma^{(\lambda)}_{\mathrm{EM}}\big)$ has asymptotically (as $n\to\infty$) a normal distribution with mean zero and variance,

$$\sigma^2\Big[\Gamma^{(\lambda)}_{\mathrm{EM}}\Big] = \sum_{k=1}^{R-1}\sum_{l=k+1}^{R}\Big[p_{kl}\big(\omega^{(\lambda)}_{1(kl)}\big)^2+p_{lk}\big(\omega^{(\lambda)}_{2(kl)}\big)^2\Big] ,$$

where for $\lambda>-1$, $\lambda\neq 0$; $t=1,2$,

$$\omega^{(\lambda)}_{t(kl)} = \frac{2^\lambda}{2\,(2^\lambda-1)\,G_t}\left[\sum_{i=k}^{l-1}\Delta^{(\lambda)}_{t(i)}-(l-k)\sum_{i=1}^{R-1}G^*_{t(i)}\,\Delta^{(\lambda)}_{t(i)}\right] ,$$

$$\begin{aligned}\Delta^{(\lambda)}_{1(i)} &= \big(G^c_{1(i)}\big)^\lambda+\lambda\Big\{\big(G^c_{1(i)}\big)^\lambda-\big(G^c_{2(i)}\big)^\lambda\Big\}\,G^c_{2(i)} ,\\ \Delta^{(\lambda)}_{2(i)} &= \big(G^c_{2(i)}\big)^\lambda+\lambda\Big\{\big(G^c_{2(i)}\big)^\lambda-\big(G^c_{1(i)}\big)^\lambda\Big\}\,G^c_{1(i)} ;\end{aligned}$$

and for $\lambda=0$; $t=1,2$,

$$\omega^{(0)}_{t(kl)} = \frac{1}{2\,(\log 2)\,G_t}\left[\sum_{i=k}^{l-1}\log\big(G^c_{t(i)}\big)-(l-k)\sum_{i=1}^{R-1}G^*_{t(i)}\log\big(G^c_{t(i)}\big)\right] .$$

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

## 4. EXAMPLES

Consider the data in Table 2, 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.

**Table 2**: Occupational status for Japanese father-son pairs; from Tominaga (1979, p. 53).

| Father's status | Son's status | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | |
| (a) Examined in 1955 | | | | | | | | | |
| (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 | | | | | | | | | |
| (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 | | | | | | | | | |
| (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 |

*Note*: Status (1) is Professional, (2) Managers, (3) Clerical, (4) Sales, (5) Skilled manual, (6) Semiskilled manual, (7) Unskilled manual and (8) Farmers.

Since the confidence intervals for $\Gamma_{\mathrm{EM}}^{(\lambda)}$ applied to the data in Tables 2a, 2b and 2c do not include zero for all $\lambda$ (see Table 3), these would indicate that there is not a structure of EMH in each table.

**Table 3**: Estimate of $\Gamma_{\mathrm{EM}}^{(\lambda)}$, estimated approximate standard error for $\hat{\Gamma}_{\mathrm{EM}}^{(\lambda)}$, and approximate 95% confidence interval for $\Gamma_{\mathrm{EM}}^{(\lambda)}$, applied to Tables 2a, 2b and 2c.

| Values of $\lambda$ | Estimated measure | Standard error | Confidence interval |
|---|---|---|---|
| (a) For Table 2a | | | |
| −0.5 | 0.017 | 0.004 | (0.009, 0.024) |
| 0 | 0.028 | 0.006 | (0.016, 0.040) |
| 0.5 | 0.035 | 0.008 | (0.019, 0.050) |
| 1.0 | 0.038 | 0.009 | (0.021, 0.055) |
| 1.5 | 0.039 | 0.009 | (0.022, 0.056) |
| 2.0 | 0.038 | 0.009 | (0.021, 0.055) |
| 2.5 | 0.036 | 0.008 | (0.020, 0.052) |
| (b) For Table 2b | | | |
| −0.5 | 0.043 | 0.006 | (0.031, 0.055) |
| 0 | 0.070 | 0.009 | (0.051, 0.088) |
| 0.5 | 0.085 | 0.011 | (0.063, 0.107) |
| 1.0 | 0.093 | 0.012 | (0.069, 0.116) |
| 1.5 | 0.095 | 0.012 | (0.071, 0.118) |
| 2.0 | 0.093 | 0.012 | (0.069, 0.116) |
| 2.5 | 0.088 | 0.012 | (0.066, 0.111) |
| (c) For Table 2c | | | |
| −0.5 | 0.053 | 0.007 | (0.040, 0.066) |
| 0 | 0.086 | 0.010 | (0.066, 0.106) |
| 0.5 | 0.105 | 0.012 | (0.081, 0.129) |
| 1.0 | 0.114 | 0.013 | (0.089, 0.139) |
| 1.5 | 0.116 | 0.013 | (0.091, 0.142) |
| 2.0 | 0.114 | 0.013 | (0.089, 0.139) |
| 2.5 | 0.109 | 0.012 | (0.084, 0.133) |

When the degrees of departure from EMH in Tables 2a, 2b and 2c are compared using the confidence interval for $\Gamma_{\mathrm{EM}}^{(\lambda)}$, it is greater in Tables 2b and 2c than in Table 2a. However, the comparison between Tables 2b and 2c may be impossible, because the values in the confidence interval for Table 2b are not always greater than the values in the confidence interval for Table 2c.

We shall investigate the degree of departure from EMH in more details. For instance, when $\lambda = 1$, the estimated measure $\hat{\Gamma}_{\mathrm{EM}}^{(1)}$ equals 0.038 for Table 2a, 0.093 for Table 2b, and 0.114 for Table 2c (see Table 3). Thus,

**(i)** for Table 2a, the degree of departure from EMH is estimated to be 3.8 percent of the maximum degree of departure from EMH,

**(ii)** for Table 2b, it is estimated to be 9.3 percent of the maximum degree of departure from EMH,

and

**(iii)** for Table 2c, it is estimated to be 11.4 percent of the maximum degree of departure from EMH.

*Note*: Let $W^{(\lambda)}$ $(-\infty < \lambda < \infty)$ denote the power-divergence statistic for testing goodness-of-fit of the EMH model with $R-2$ degrees of freedom. [See Appendix for $W^{(\lambda)}$, and see Cressie and Read (1984) and Read and Cressie (1988, p.15) for details of the power-divergence test statistic.] In particular, $W^{(0)}$ and $W^{(1)}$ are the likelihood ratio and Pearson's chi-squared statistics, respectively. Table 4 gives the values of $W^{(\lambda)}$ applied to the data in Tables 2a, 2b, and 2c. These data fit the EMH model very poorly.

**Table 4**: The values of power-divergence statistic $W^{(\lambda)}$ (with 6 degrees of freedom) for testing goodness-of-fit of the EMH model, applied to Tables 2a, 2b and 2c.

| Values of $\lambda$ | For Table 2a | For Table 2b | For Table 2c |
|---|---|---|---|
| $-0.5$ | 118.52 | 300.36 | 333.41 |
| 0 | 116.76 | 231.58 | 280.73 |
| 0.5 | 117.38 | 200.39 | 252.56 |
| 1.0 | 120.39 | 186.77 | 239.04 |
| 1.5 | 125.95 | 183.14 | 235.42 |
| 2.0 | 134.42 | 186.48 | 239.54 |
| 2.5 | 146.33 | 195.69 | 250.66 |

## 5. CONCLUDING REMARKS

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

Consider the artificial data in Table 5. Table 6 gives the values of $W^{(\lambda)}$ (with 2 degrees of freedom) for testing goodness-of-fit of the EMH model applied to these data. Compare the values of $W^{(\lambda)}$ for Tables 5a and 5b. From $W^{(\lambda)}$ with any fixed $\lambda$, we see that the EMH model fits the data in Table 5a worse than the data in Table 5b (see Table 6). In contrast, for any fixed $\lambda$ $(> -1)$, the value of $\hat{\Gamma}_{\mathrm{EM}}^{(\lambda)}$ is less for Table 5a than for Table 5b (see Table 7). In terms of $\hat{G}_{1(i)}/\hat{G}_{2(i)}$, $i=1,2,3$ (see Table 5), it seems natural to conclude that the degree of departure from EMH is less for Table 5a than for Table 5b. Therefore $\hat{\Gamma}_{\mathrm{EM}}^{(\lambda)}$

**Table 5**: Artificial data.

(a) $n = 2829$

| | (1) | (2) | (3) | (4) | Total |
|---|---|---|---|---|---|
| (1) | 187 | 330 | 70 | 20 | 607 |
| (2) | 30 | 178 | 60 | 40 | 308 |
| (3) | 50 | 100 | 898 | 60 | 1108 |
| (4) | 70 | 20 | 10 | 706 | 806 |
| Total | 337 | 628 | 1038 | 826 | 2829 |

*Note*: $\dfrac{\hat{G}_{1(1)}}{\hat{G}_{2(1)}} = 2.80, \quad \dfrac{\hat{G}_{1(2)}}{\hat{G}_{2(2)}} = 0.79, \quad \dfrac{\hat{G}_{1(3)}}{\hat{G}_{2(3)}} = 1.20\,.$

(b) $n = 2654$

| | (1) | (2) | (3) | (4) | Total |
|---|---|---|---|---|---|
| (1) | 687 | 80 | 10 | 5 | 782 |
| (2) | 5 | 178 | 5 | 12 | 200 |
| (3) | 5 | 25 | 898 | 13 | 941 |
| (4) | 10 | 8 | 7 | 706 | 731 |
| Total | 707 | 291 | 920 | 736 | 2654 |

*Note*: $\dfrac{\hat{G}_{1(1)}}{\hat{G}_{2(1)}} = 4.75, \quad \dfrac{\hat{G}_{1(2)}}{\hat{G}_{2(2)}} = 0.67, \quad \dfrac{\hat{G}_{1(3)}}{\hat{G}_{2(3)}} = 1.20\,.$

(b) $n = 429$

| | (1) | (2) | (3) | (4) | Total |
|---|---|---|---|---|---|
| (1) | 68 | 80 | 10 | 5 | 163 |
| (2) | 5 | 17 | 5 | 12 | 39 |
| (3) | 5 | 25 | 89 | 13 | 132 |
| (4) | 10 | 8 | 7 | 70 | 95 |
| Total | 88 | 130 | 111 | 100 | 429 |

*Note*: $\dfrac{\hat{G}_{1(1)}}{\hat{G}_{2(1)}} = 4.75, \quad \dfrac{\hat{G}_{1(2)}}{\hat{G}_{2(2)}} = 0.67, \quad \dfrac{\hat{G}_{1(3)}}{\hat{G}_{2(3)}} = 1.20\,.$

may be preferable to $W^{(\lambda)}$ for *comparing* the degree of departure from EMH in several tables. It may seem, to many readers, that $W^{(\lambda)}/n$ (for a given $\lambda$) is also a reasonable measure for representing the degree of departure from EMH. However, it does not seem to us that $W^{(\lambda)}/n$ is a reasonable measure. For example, consider the artificial data in Tables 5b and 5c. The values of $W^{(\lambda)}/n$ are, for example, when $\lambda = 0$ ($\lambda = 1$), $W^{(0)}/n = 0.024$ ($W^{(1)}/n = 0.022$) for Table 5b, and $W^{(0)}/n = 0.147$ ($W^{(1)}/n = 0.138$) for Table 5c. Therefore the value of $W^{(\lambda)}/n$ is less for Table 5b than for Table 5c. On the other side, for any fixed $\lambda$ ($> -1$), the value of $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$ for Table 5b is theoretically identical to that for Table 5c (see Table 7). In addition, $\hat{G}_{1(i)}/\hat{G}_{2(i)}$, $i=1,2,3$, for Table 5b is identical to that for Table 5c (see Table 5). So it seems natural to conclude that the degree of departure from EMH for Table 5b is equal to that for Table 5c. Therefore $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$ may also be preferable to $W^{(\lambda)}/n$ for *comparing* the degree of departure from EMH in several tables.

**Table 6**: The values of $W^{(\lambda)}$ (with 2 degrees of freedom) for testing goodness-of-fit of the EMH model, applied to Tables 5a, 5b and 5c.

| Values of $\lambda$ | For Table 5a | For Table 5b | For Table 5c |
|---|---|---|---|
| −0.5 | 194.43 | 69.35 | 69.35 |
| 0 | 182.76 | 63.25 | 63.25 |
| 0.5 | 175.25 | 60.04 | 60.04 |
| 1.0 | 171.21 | 59.04 | 59.04 |
| 1.5 | 170.17 | 59.93 | 59.93 |
| 2.0 | 171.89 | 62.61 | 62.61 |
| 2.5 | 176.28 | 67.17 | 67.17 |

**Table 7**: The values of $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$ applied to Tables 5a, 5b and 5c.

| Values of $\lambda$ | For Table 5a | For Table 5b | For Table 5c |
|---|---|---|---|
| −0.5 | 0.034 | 0.076 | 0.076 |
| 0 | 0.057 | 0.125 | 0.125 |
| 0.5 | 0.071 | 0.153 | 0.153 |
| 1.0 | 0.078 | 0.167 | 0.167 |
| 1.5 | 0.080 | 0.171 | 0.171 |
| 2.0 | 0.078 | 0.167 | 0.167 |
| 2.5 | 0.074 | 0.159 | 0.159 |

Since the EMH model is expressed as equation (1.1), we are interested in measuring how far $\left\{G^*_{1(i)}\right\}$ and $\left\{G^*_{2(i)}\right\}$ are distant from those with an EMH structure when the EMH model does not hold. The measure $\Gamma^{(\lambda)}_{\mathrm{EM}}$ is a function of $\left\{G^*_{1(i)}\right\}$ and $\left\{G^*_{2(i)}\right\}$. Since equation (1.1), it seems natural that the measure is expressed as a function of $\left\{G^*_{1(i)}\right\}$ and $\left\{G^*_{2(i)}\right\}$.

For the measure $\Gamma^{(\lambda)}_{\mathrm{EM}}$, the analyst may be interested in which value of $\lambda$ 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 EMH in several tables, the analyst calculates the values of $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$ for various values of $\lambda$ and discusses the degree of departure from EMH in terms of them (rather than calculating $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$ for *only* one specified value of $\lambda$).

**Table 8**: Artificial data.

(a) $n = 1895$ (sample size)

| | | | |
|---|---|---|---|
| 374 | 602 | 170 | 64 |
| 18 | 255 | 139 | 71 |
| 4 | 23 | 42 | 55 |
| 2 | 6 | 17 | 53 |

(b) $n = 5397$

| | | | |
|---|---|---|---|
| 81 | 444 | 632 | 726 |
| 646 | 178 | 498 | 6 |
| 787 | 288 | 68 | 762 |
| 72 | 105 | 17 | 87 |

Consider the artificial data in Tables 8a and 8b. Then we see from Table 9 that the value of $\hat{\Gamma}^{(0)}_{\mathrm{EM}}$ is less for Table 8a than for Table 8b, but the value of $\hat{\Gamma}^{(1)}_{\mathrm{EM}}$ is greater for Table 8a than for Table 8b. However, the differences are very slight in these cases. So, for these cases, we may conclude (by using $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$) that the departure from EMH for Table 8a is similar to that for Table 8b. But generally, for the comparison between two tables, it would be possible to conclude for which of two tables the departure from the EMH is greater if $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$ (for every $\lambda$) is always greater (or always less) for one table than for the other table.

**Table 9**: The values of $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$ applied to Tables 8a and 8b.

| Values of $\lambda$ | For Table 8a | For Table 8b |
|---|---|---|
| $-0.5^*$ | 0.040 | 0.042 |
| $0^*$ | 0.066 | 0.068 |
| $0.5^*$ | 0.081 | 0.082 |
| 1.0 | 0.089 | 0.088 |
| 1.5 | 0.091 | 0.090 |
| 2.0 | 0.089 | 0.088 |
| 2.5 | 0.084 | 0.084 |

$*$ indicates that $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$ is less for Table 8a than for Table 8b.

By the way, it is easily seen that the measure $\Gamma^{(0)}_{\mathrm{EM}}$ with equation (2.1) can be expressed as

$$(5.1) \qquad \Gamma^{(0)}_{\mathrm{EM}} = \frac{1}{2\log 2} \min_{\{D_i\}} \left[ I^{(0)}\Big(\{G^*_{1(i)}\}; \{D_i\}\Big) + I^{(0)}\Big(\{G^*_{2(i)}\}; \{D_i\}\Big) \right],$$

where $\sum_{i=1}^{R-1} D_i = 1$ and $D_i > 0$. Therefore we point out that $C_i$ in $\Gamma^{(\lambda)}_{\mathrm{EM}}$ is the value of $D_i$ such that the sum of two KL distances (i.e., the KL distance between

$\{G^*_{1(i)}\}$ and $\{D_i\}$ with an EMH structure and the KL distance between $\{G^*_{2(i)}\}$ and $\{D_i\}$) is a minimum. [Note that the readers may also be interested in equation (5.1) with $I^{(0)}(\cdot;\cdot)$ replaced by the power-divergence $I^{(\lambda)}(\cdot;\cdot)$; however, it is difficult to obtain the value of $D_i$ such that the corresponding two power-divergence is a minimum, and also difficult to obtain the maximum value of such a measure.]

Finally, we observe that

**(i)** the measure should be applied to contingency tables with *ordered* categories because it is not invariant under the same arbitrary permutations of row and column categories except the reverse order,

**(ii)** $\Gamma^{(\lambda)}_{\mathrm{EM}}$ should be used when there is not a structure of EMH in square tables,

**(iii)** $\hat{\Gamma}^{(\lambda)}_{\mathrm{EM}}$ cannot be used for testing goodness-of-fit of the EMH model (though $W^{(\lambda)}$ can be used),

and

**(iv)** the value of $\Gamma^{(1)}_{\mathrm{EM}}$ is theoretically equal to the value of $\Gamma^{(2)}_{\mathrm{EM}}$.

## APPENDIX

The power-divergence statistic for testing goodness-of-fit of the EMH model is given by

$$W^{(\lambda)} = 2\,n\,I^{(\lambda)}\Big(\{\hat{p}_{ij}\};\{\hat{p}^{\mathrm{ML}}_{ij}\}\Big) \qquad \text{for} \quad -\infty<\lambda<\infty \ ,$$

where

$$I^{(\lambda)}(\cdot;\cdot) = \frac{1}{\lambda(\lambda+1)}\sum_{i=1}^{R}\sum_{j=1}^{R}\hat{p}_{ij}\left[\left(\frac{\hat{p}_{ij}}{\hat{p}^{\mathrm{ML}}_{ij}}\right)^{\lambda}-1\right], \qquad \hat{p}_{ij} = \frac{n_{ij}}{n} \ ,$$

and $\hat{p}^{\mathrm{ML}}_{ij}$ is the maximum likelihood estimate of $p_{ij}$ under the EMH model, where the values at $\lambda=-1$ and $\lambda=0$ are taken to be the limit as $\lambda\to-1$ and $\lambda\to 0$, respectively. The number of degrees of freedom is $R-2$.

## ACKNOWLEDGMENTS

The authors thank a referee and the associate editor for their helpful comments.

## REFERENCES

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

[2] Bhapkar, V.P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data, *Journal of the American Statistical Association*, **61**, 228–235.

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

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

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

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

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

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

[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). 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.

[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.