Original Article

A Measure of Departure from Symmetry for Multi-way Contingency Tables with Nominal Categories

Hideharu Yamamoto*

* Department of Information Sciences, Faculty of Science and Technology,
Tokyo University of Science, Noda City, Chiba, 278-8510, Japan;
Statistical Group Clinical Data Analysis DEPT.,
CHUGAI PHARMACEUTICAL CO.,LTD., Tokyo, 104-8301, Japan
e-mail: yamamotohdh@chugai-pharma.co.jp

For square contingency tables with nominal categories, Tomizawa (1994) and Tomizawa, Seo and Yamamoto (1998) considered measures that reflect the degree of departure from symmetry. This paper proposes a generalized measure for T -way ($T \geq 3$) tables. The proposed measure is expressed by using the power divergence of Cressie and Read (1984) or the diversity index of Patil and Taillie (1982). The measure could be useful for comparing the degrees of departure from symmetry in several multi-way tables. Some examples of analysis using biomedical data are shown.

Key words: Biomedical data analysis, Kullback-Leibler information, Measure, Power divergence, Shannon entropy.

1. Introduction

Consider an $R \times R$ square contingency table that has the same nominal 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 ($i = 1, 2, \dots, R$; $j = 1, 2, \dots, R$). Symmetry is defined as

$$p_{ij} = p_{ji} \quad \text{for } i = 1, 2, \dots, R; j = 1, 2, \dots, R$$

(see, for example, Bowker, 1948; Bishop, Fienberg, and Holland, 1975, p.282). When symmetry does not hold, we may be interested in whether there is other structure, such as extended symmetry (e.g., quasi-symmetry, or marginal homogeneity; see Bishop *et al.*, 1975, Chapter 8) or for ordinal data some particular form of asymmetry (e.g., conditional symmetry, diagonals-parameter symmetry, and so on; see Agresti, 1984, p.202). When there is asymmetry in a square contingency table, we are interested in measuring the degree of departure from symmetry. For square contingency tables with nominal categories, Tomizawa (1994), and Tomizawa, Seo and Yamamoto (1998) considered measures that represent the degree of departure from symmetry

(see Appendix A). Note that Tomizawa, Miyamoto and Hatanaka (2002) considered a measure for square contingency tables with ordinal categories.

Symmetry may also be considered for multi-way R^T ($T \geq 3$) contingency tables (see Bishop *et al.*, 1975, p.300, and Agresti, 2002, p.440). We can test whether there is symmetry in a multi-way contingency table using the likelihood ratio chi-squared statistic or Pearson's chi-squared statistic. However, since these test statistics describe absolute magnitudes, they do not represent efficiently how far is the degree of departure from symmetry, and they may not be suitable for comparing the degrees of asymmetry in several tables (see also Section 6). We are therefore interested in a measure that represents the degree of departure from symmetry in a multi-way table.

The purpose of this paper is to propose a power-divergence-type measure, which represents the degree of departure from symmetry in multi-way contingency tables with nominal categories. In Section 2, we propose the measure for the case of a three-way table, and in Section 3, for the case of T -way ($T \geq 4$) tables. It should be useful for comparing the degrees of departure from symmetry toward the maximum degree of asymmetry in several tables.

2. Measure for three-way tables

2.1 Measure

Consider an $R \times R \times R$ contingency table having nominal categories. Let p_{ijk} denote the probability that an observation will fall in the (i, j, k) cell of the table for $1 \leq i, j, k \leq R$. Symmetry (S) is defined as

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

It is easily seen that S holds if and only if both S_1 and S_2 hold, where S_1 is defined as

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

and S_2 is defined as

$$p_{iij} = p_{iji} = p_{jii} \quad \text{for } i \neq j.$$

Note that for the $2 \times 2 \times 2$ table, S_2 model can be expressed as $p_{112} = p_{121} = p_{211}$ and $p_{221} = p_{212} = p_{122}$.

Let

$$E_1 = \{(i, j, k) \mid 1 \leq i, j, k \leq R; i \neq j, j \neq k, k \neq i\},$$

$$E_2 = \{(i, j, k) \mid 1 \leq i, j, k \leq R; i = j \neq k \text{ or } k = i \neq j \text{ or } j = k \neq i\},$$

and

$$\delta_t = \sum_{(i,j,k) \in E_t} \sum_{(i,j,k) \in E_t} p_{ijk} \quad \text{for } t = 1, 2.$$

Define the conditional probabilities

$$p_{ijk[t]}^* = \frac{p_{ijk}}{\delta_t} \quad \text{for } (i, j, k) \in E_t, \text{ and } t = 1, 2.$$

Note that $p_{ijk[t]}^*$ indicates the probability that an observation will fall in cell (i, j, k) , on the condition that it falls in one of the cells in E_t .

Consider the following measure, which represents the degree of departure from the S :

$$\Phi^{(\lambda)} = \frac{\delta_1 \Phi_1^{(\lambda)} + \delta_2 \Phi_2^{(\lambda)}}{\delta_1 + \delta_2} \quad \text{for } \lambda > -1 \ (\lambda \neq 0),$$

with

$$\Phi_t^{(\lambda)} = \frac{\lambda(\lambda+1)}{\left(\frac{6}{t!}\right)^\lambda - 1} I_t^{(\lambda)}(\{p_{ijk[t]}^*\}; \{p_{ijk[t]}^s\}),$$

where

$$I_t^{(\lambda)}(\cdot; \cdot) = \frac{1}{\lambda(\lambda+1)} \sum_{(i,j,k) \in E_t} \sum \sum p_{ijk[t]}^* \left[\left(\frac{p_{ijk[t]}^*}{p_{ijk[t]}^s} \right)^\lambda - 1 \right],$$

$$p_{ijk[t]}^s = \frac{1}{6} (p_{ijk[t]}^* + p_{ikj[t]}^* + p_{jik[t]}^* + p_{jki[t]}^* + p_{kij[t]}^* + p_{kji[t]}^*),$$

and where the value at $\lambda = 0$ is taken to be the continuous limit as $\lambda \rightarrow 0$. Thus, for $t = 1, 2$,

$$\Phi_t^{(0)} = \frac{1}{\log\left(\frac{6}{t!}\right)} \sum_{(i,j,k) \in E_t} \sum \sum p_{ijk[t]}^* \log \left(\frac{p_{ijk[t]}^*}{p_{ijk[t]}^s} \right).$$

This measure, $\Phi_t^{(0)}$, is expressed using the Kullback-Leibler information (or the Shannon entropy). We note that $\Phi^{(\lambda)}$ would represent the average of $\{\Phi_1^{(\lambda)}\}$ and $\{\Phi_2^{(\lambda)}\}$ on condition that an observation falls in one of the off-main-diagonal cells of the $R \times R \times R$ table. In addition, note that for $t = 1, 2$, $I_t^{(\lambda)}(\{p_{ijk[t]}^*\}; \{p_{ijk[t]}^s\})$ is the power divergence between the two distributions $\{p_{ijk[t]}^*\}$ and $\{p_{ijk[t]}^s\}$. (For details of power divergence, see Cressie and Read, 1984, and Read and Cressie, 1988, p.15.)

For $(i, j, k) \in E_1$, let

$$p_{ijk[1]}^c = \frac{p_{ijk}}{p_{ijk} + p_{ikj} + p_{jik} + p_{jki} + p_{kij} + p_{kji}},$$

and for $(i, j, k) \in E_2$,

$$p_{ijk[2]}^c = \frac{2p_{ijk}}{p_{ijk} + p_{ikj} + p_{jik} + p_{jki} + p_{kij} + p_{kji}}.$$

For example, for the $4 \times 4 \times 4$ table,

$$p_{123[1]}^c = \frac{p_{123}}{p_{123} + p_{132} + p_{213} + p_{231} + p_{312} + p_{321}},$$

and

$$p_{224[2]}^c = \frac{p_{224}}{p_{224} + p_{242} + p_{422}}.$$

In addition, let

$$D(i, j, k) = \{(l, m, n) \mid (l, m, n) \text{ is any permutation of } (i, j, k)\}.$$

For example, for the case of a $4 \times 4 \times 4$ table,

$$D(1, 2, 3) = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\},$$

and

$$D(1, 2, 2) = \{(1, 2, 2), (2, 1, 2), (2, 2, 1)\}.$$

Note that $p_{ijk[t]}^c$, $t = 1, 2$, indicates the probability that an observation will fall in cell (i, j, k) , on the condition that the observation falls in one of the cells in $D(i, j, k)$ of the $R \times R \times R$ table. Then, $\{\Phi_t^{(\lambda)}\}$, $t = 1, 2$, may be expressed as,

$$\begin{aligned} \Phi_1^{(\lambda)} &= \frac{\lambda(\lambda+1)}{6^\lambda - 1} \sum_{\substack{1 \leq i < j < k \leq R \\ (i, j, k) \in E_1}} (p_{ijk[1]}^* + p_{ikj[1]}^* + p_{jik[1]}^* + p_{jki[1]}^* + p_{kij[1]}^* + p_{kji[1]}^*) \\ &\quad \times I_{ijk[1]}^{(\lambda)}(\{p_{lmn[1]}^c\}; \{\frac{1}{6}\}), \end{aligned}$$

where

$$I_{ijk[1]}^{(\lambda)}(\cdot; \cdot) = \frac{1}{\lambda(\lambda+1)} \sum_{(l, m, n) \in D(i, j, k)} \sum \sum \sum p_{lmn[1]}^c \left[\left(\frac{p_{lmn[1]}^c}{\frac{1}{6}} \right)^\lambda - 1 \right],$$

and

$$\begin{aligned} \Phi_2^{(\lambda)} &= \frac{\lambda(\lambda+1)}{2(3^\lambda - 1)} \sum_{\substack{1 \leq i \leq j \leq k \leq R \\ (i, j, k) \in E_2}} (p_{ijk[2]}^* + p_{ikj[2]}^* + p_{jik[2]}^* + p_{jki[2]}^* + p_{kij[2]}^* + p_{kji[2]}^*) \\ &\quad \times I_{ijk[2]}^{(\lambda)}(\{p_{lmn[2]}^c\}; \{\frac{1}{3}\}), \end{aligned}$$

where

$$I_{ijk[2]}^{(\lambda)}(\cdot; \cdot) = \frac{1}{\lambda(\lambda+1)} \sum_{(l, m, n) \in D(i, j, k)} \sum \sum \sum p_{lmn[2]}^c \left[\left(\frac{p_{lmn[2]}^c}{\frac{1}{3}} \right)^\lambda - 1 \right].$$

Moreover, these may be expressed as

$$\begin{aligned} \Phi_1^{(\lambda)} &= \sum_{\substack{1 \leq i < j < k \leq R \\ (i, j, k) \in E_1}} \sum \sum \sum (p_{ijk[1]}^* + p_{ikj[1]}^* + p_{jik[1]}^* + p_{jki[1]}^* + p_{kij[1]}^* + p_{kji[1]}^*) \\ &\quad \times \left[1 - \frac{\lambda 6^\lambda}{6^\lambda - 1} H_{ijk[1]}^{(\lambda)}(\{p_{lmn[1]}^c\}) \right], \end{aligned}$$

and

$$\begin{aligned} \Phi_2^{(\lambda)} &= \frac{1}{2} \sum_{\substack{1 \leq i \leq j \leq k \leq R \\ (i, j, k) \in E_2}} \sum \sum \sum (p_{ijk[2]}^* + p_{ikj[2]}^* + p_{jik[2]}^* + p_{jki[2]}^* + p_{kij[2]}^* + p_{kji[2]}^*) \\ &\quad \times \left[1 - \frac{\lambda 3^\lambda}{3^\lambda - 1} H_{ijk[2]}^{(\lambda)}(\{p_{lmn[2]}^c\}) \right], \end{aligned}$$

where

$$H_{ijk[t]}^{(\lambda)}(\cdot) = \frac{1}{\lambda} \left[1 - \sum_{(l,m,n) \in D(i,j,k)} \sum_{(l,m,n) \in D(i,j,k)} \sum_{(l,m,n) \in D(i,j,k)} (p_{lmn[t]}^c)^{\lambda+1} \right].$$

For $t = 1, 2$, $\Phi_t^{(\lambda)}$ must lie between 0 and 1, and therefore $\Phi^{(\lambda)}$ must lie between 0 and 1. We note that $0 \leq \min(\Phi_1^{(\lambda)}, \Phi_2^{(\lambda)}) \leq \Phi^{(\lambda)} \leq \max(\Phi_1^{(\lambda)}, \Phi_2^{(\lambda)}) \leq 1$. Also, (i) the $R \times R \times R$ table has structure of the form S if and only if $\Phi_1^{(\lambda)} = 0$ and $\Phi_2^{(\lambda)} = 0$, i.e., $\Phi^{(\lambda)} = 0$, and (ii) the degree of departure from S is largest in the sense that, for $(l, m, n) \in D(i, j, k)$, $i, j, k = 1, 2, \dots, R$,

$$p_{lmn}^c = \begin{cases} 1 & \text{for } (l, m, n) = (l_0, m_0, n_0) \text{ with a cell } (l_0, m_0, n_0) \in D(i, j, k), \\ 0 & \text{for } (l, m, n) \neq (l_0, m_0, n_0), \end{cases}$$

if and only if $\Phi_1^{(\lambda)} = 1$ and $\Phi_2^{(\lambda)} = 1$, i.e., $\Phi^{(\lambda)} = 1$. The degree of departure from S increases as the value of $\Phi^{(\lambda)}$ increases.

2.2 Measure for each symmetrically located cells

When the symmetry assumption does not hold, we are interested in finding which symmetrically located cells of the $R \times R \times R$ table contribute most to the departure from symmetry. The larger the dimension of the table becomes, the more important this may be. For each (i, j, k) with $1 \leq i \leq j \leq k \leq R$, we assume that $\sum \sum \sum_{(l,m,n) \in D(i,j,k)} p_{lmn} > 0$. Then consider the measure defined, for each $(i, j, k) \in E_1$ and $1 \leq i < j < k \leq R$, by

$$\Phi_{ijk}^{(\lambda)} = \frac{\lambda(\lambda+1)}{6\lambda-1} I_{ijk[1]}^{(\lambda)}(\{p_{lmn[1]}^c\}; \{\frac{1}{6}\}) \quad \text{for } \lambda > -1,$$

and for each $(i, j, k) \in E_2$ and $1 \leq i \leq j \leq k \leq R$, by

$$\Phi_{ijk}^{(\lambda)} = \frac{\lambda(\lambda+1)}{3\lambda-1} I_{ijk[2]}^{(\lambda)}(\{p_{lmn[2]}^c\}; \{\frac{1}{3}\}) \quad \text{for } \lambda > -1,$$

where the value at $\lambda = 0$ is taken to be the limit as $\lambda \rightarrow 0$.

Moreover, for each $(i, j, k) \in E_1$ and $1 \leq i < j < k \leq R$, $\Phi_{ijk}^{(\lambda)}$ may be expressed as

$$\Phi_{ijk}^{(\lambda)} = 1 - \frac{\lambda 6^\lambda}{6\lambda-1} H_{ijk[1]}^{(\lambda)}(\{p_{lmn[1]}^c\}) \quad \text{for } \lambda > -1,$$

and for each $(i, j, k) \in E_2$ and $1 \leq i \leq j \leq k \leq R$, as

$$\Phi_{ijk}^{(\lambda)} = 1 - \frac{\lambda 3^\lambda}{3\lambda-1} H_{ijk[2]}^{(\lambda)}(\{p_{lmn[2]}^c\}) \quad \text{for } \lambda > -1.$$

It is seen that each measure must lie between 0 and 1. Also, for each $\lambda (> -1)$, and for fixed i, j , and k , (i) $p_{ijk} = p_{ikj} = p_{jik} = p_{jki} = p_{kij} = p_{kji}$ holds if and only if $\Phi_{ijk}^{(\lambda)} = 0$, and (ii) for $(l, m, n) \in D(i, j, k)$,

$$p_{lmn}^c = \begin{cases} 1 & \text{for } (l, m, n) = (l_0, m_0, n_0) \text{ with a cell } (l_0, m_0, n_0) \in D(i, j, k), \\ 0 & \text{for } (l, m, n) \neq (l_0, m_0, n_0), \end{cases}$$

if and only if $\Phi_{ijk}^{(\lambda)} = 1$.

3. Extension to multi-way tables

We extend the measure proposed in Section 2 to multi-way tables. Consider an R^T table ($T \geq 3$) having nominal categories. Let $p_{i_1 i_2 \dots i_T}$ denote the probability that an observation falls in the $(i_1, i_2, \dots, i_T)$ cell of the table ($i_t = 1, 2, \dots, R; t = 1, 2, \dots, T$). Symmetry (S) (Agresti, 2002, p.439) is defined by

$$p_{i_1 i_2 \dots i_T} = p_{j_1 j_2 \dots j_T}$$

for $i_t = 1, 2, \dots, R; t = 1, 2, \dots, T$, and $(j_1, j_2, \dots, j_T) \in D(i_1, i_2, \dots, i_T)$, where

$$D(i_1, i_2, \dots, i_T) = \{(j_1, j_2, \dots, j_T) \mid (j_1, j_2, \dots, j_T) \text{ is any permutation of } (i_1, i_2, \dots, i_T)\}.$$

In particular, when $T = 3$, S is defined as

$$p_{ijk} = p_{ikj} = p_{jik} = p_{jki} = p_{kij} = p_{kji} \quad (1 \leq i, j, k \leq R).$$

Let

$$E_{(r_1, r_2, \dots, r_T)} = \{(i_1, i_2, \dots, i_T) \mid i_t = 1, 2, \dots, R; t = 1, 2, \dots, T, \text{ where the } r_1\text{'s elements of } (i_1, i_2, \dots, i_T) \text{ are equal, } r_2\text{'s elements are equal, } \dots, \text{ and } r_T\text{'s elements are equal}\},$$

for $(r_1, r_2, \dots, r_T)$ which satisfies

$$r_1 + r_2 + \dots + r_T = T \text{ and } T - 1 \geq r_1 \geq r_2 \geq \dots \geq r_T \geq 0.$$

Note that

$$E_{(1, 1, \dots, 1)} = \{(i_1, i_2, \dots, i_T) \mid i_t = 1, 2, \dots, R; t = 1, 2, \dots, T, \text{ where all elements of } (i_1, i_2, \dots, i_T) \text{ are different}\}.$$

For example, in the case of a 4^4 table, $E_{(r_1, r_2, \dots, r_T)}$ can be divided into 4 groups as follows,

$$\begin{aligned} E_{(3, 1, 0, 0)} &= \{(1, 1, 1, 2), (1, 1, 1, 3), \dots, (4, 4, 4, 3)\}; \\ E_{(2, 2, 0, 0)} &= \{(1, 1, 2, 2), (1, 1, 3, 3), \dots, (4, 4, 3, 3)\}; \\ E_{(2, 1, 1, 0)} &= \{(1, 1, 2, 3), (1, 1, 2, 4), \dots, (4, 4, 3, 2)\}; \\ E_{(1, 1, 1, 1)} &= \{(1, 2, 3, 4), (1, 2, 4, 3), \dots, (4, 3, 2, 1)\}. \end{aligned}$$

It is easily seen that S holds if and only if all $S_{(r_1, r_2, \dots, r_T)}$ hold, where the $S_{(r_1, r_2, \dots, r_T)}$ are defined as

$$p_{i_1 i_2 \dots i_T} = \Gamma_{r_1, r_2, \dots, r_T}$$

for $(i_1, i_2, \dots, i_T) \in E_{(r_1, r_2, \dots, r_T)}$ and $\Gamma_{r_1, r_2, \dots, r_T}$ being constant.

Consider a measure of departure from S defined as

$$\Phi^{(\lambda)} = \frac{1}{\delta} \sum_{(r_1, r_2, \dots, r_T) \in S^*} \sum \dots \sum \delta_{r_1 r_2 \dots r_T} \Phi_{r_1 r_2 \dots r_T}^{(\lambda)} \quad \text{for } \lambda > -1,$$

where

$$\begin{aligned} \delta_{r_1 r_2 \dots r_T} &= \sum_{(i_1, i_2, \dots, i_T) \in E_{(r_1, r_2, \dots, r_T)}} \sum \dots \sum p_{i_1 i_2 \dots i_T}, \\ \delta &= \sum_{(r_1, r_2, \dots, r_T) \in S^*} \sum \dots \sum \delta_{r_1 r_2 \dots r_T} \left[1 - \sum_{k=1}^T p_{k k \dots k} \right], \\ S^* &= \{(r_1, r_2, \dots, r_T) \mid r_1 + r_2 + \dots + r_T = T, \quad T-1 \geq r_1 \geq r_2 \geq \dots \geq r_T \geq 0\}, \end{aligned}$$

and $\Phi_{r_1 r_2 \dots r_T}^{(\lambda)}$ is a measure of departure from $S_{(r_1, r_2, \dots, r_T)}$ (defined below), and the value at $\lambda = 0$ is taken to be the continuous limit as $\lambda \rightarrow 0$. Therefore $\Phi^{(\lambda)}$ would represent the average of $\{\Phi_{r_1 r_2 \dots r_T}^{(\lambda)}\}$ on condition that an observation falls in one of the off-main-diagonal cells of the R^T table. For $(r_1, r_2, \dots, r_T) \in S^*$, $\Phi_{r_1 r_2 \dots r_T}^{(\lambda)}$ in $\Phi^{(\lambda)}$ is defined by

$$\Phi_{r_1 r_2 \dots r_T}^{(\lambda)} = \frac{\lambda(\lambda+1)}{({}^T C_{r_1 r_2 \dots r_T})^\lambda - 1} I_{r_1 r_2 \dots r_T}^{(\lambda)} (\{p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^*\}; \{p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^s\}),$$

where

$$I_{r_1 r_2 \dots r_T}^{(\lambda)}(\cdot; \cdot) = \frac{1}{\lambda(\lambda+1)} \sum_{(i_1, i_2, \dots, i_T) \in E_{(r_1, r_2, \dots, r_T)}} \sum \dots \sum p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^* \left[\left(\frac{p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^*}{p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^s} \right)^\lambda - 1 \right],$$

$$\begin{aligned} p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^* &= p_{i_1 i_2 \dots i_T} / \delta_{r_1 r_2 \dots r_T}, \\ p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^s &= \sum_{(j_1, j_2, \dots, j_T) \in D(i_1, i_2, \dots, i_T)} \sum \dots \sum p_{j_1 j_2 \dots j_T [r_1 r_2 \dots r_T]}^* / ({}^T C_{r_1 r_2 \dots r_T}), \\ {}^T C_{r_1 r_2 \dots r_T} &= \frac{T!}{r_1! r_2! \dots r_T!}. \end{aligned}$$

When $\lambda = 0$, this is,

$$\Phi_{r_1 r_2 \dots r_T}^{(0)} = \frac{1}{\log({}^T C_{r_1 r_2 \dots r_T})} \sum_{(i_1, i_2, \dots, i_T) \in E_{(r_1, r_2, \dots, r_T)}} \sum \dots \sum p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^* \log \left(\frac{p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^*}{p_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^s} \right).$$

These may also be expressed as, for $(r_1, r_2, \dots, r_T) \in S^*$,

$$\begin{aligned} \Phi_{r_1 r_2 \dots r_T}^{(\lambda)} &= \frac{\lambda(\lambda+1)}{({}^T C_{r_1 r_2 \dots r_T})^\lambda - 1} \sum_{\substack{1 \leq i_1 \leq i_2 \leq \dots \leq i_T \leq R \\ (i_1, i_2, \dots, i_T) \in E_{(r_1, r_2, \dots, r_T)}}} \sum \dots \sum \left(\sum_{(j_1, j_2, \dots, j_T) \in D(i_1, i_2, \dots, i_T)} \sum \dots \sum p_{j_1 j_2 \dots j_T [r_1 r_2 \dots r_T]}^* \right) \\ &\quad \times I_{i_1 i_2 \dots i_T}^{(\lambda)} \left(\{p_{u_1 u_2 \dots u_T}^c\}; \left\{ \frac{1}{{}^T C_{r_1 r_2 \dots r_T}} \right\} \right), \end{aligned}$$

where

$$I_{i_1 i_2 \dots i_T}^{(\lambda)}(\cdot; \cdot) = \frac{1}{\lambda(\lambda+1)} \sum_{(u_1, u_2, \dots, u_T) \in D(i_1, i_2, \dots, i_T)} \sum \dots \sum p_{u_1 u_2 \dots u_T}^c \left[\left(\frac{p_{u_1 u_2 \dots u_T}^c}{\frac{1}{T C_{r_1 r_2 \dots r_T}}} \right)^\lambda - 1 \right],$$

$$p_{u_1 u_2 \dots u_T}^c = p_{u_1 u_2 \dots u_T} / \sum_{(j_1, j_2, \dots, j_T) \in D(u_1, u_2, \dots, u_T)} p_{j_1 j_2 \dots j_T}.$$

Moreover, for $(r_1, r_2, \dots, r_T) \in S^*$, these may be expressed as

$$\Phi_{r_1 r_2 \dots r_T}^{(\lambda)} = \sum_{\substack{1 \leq i_1 \leq i_2 \leq \dots \leq i_T \leq R \\ (i_1, i_2, \dots, i_T) \in E_{(r_1, r_2, \dots, r_T)}}} \sum \dots \sum \left(\sum_{(j_1, j_2, \dots, j_T) \in D(i_1, i_2, \dots, i_T)} p_{j_1 j_2 \dots j_T}^* [r_1 r_2 \dots r_T] \right) \\ \times \left[1 - \frac{\lambda (T C_{r_1 r_2 \dots r_T})^\lambda}{(T C_{r_1 r_2 \dots r_T})^\lambda - 1} H_{i_1 i_2 \dots i_T}^{(\lambda)}(\{p_{u_1 u_2 \dots u_T}^c\}) \right],$$

where

$$H_{i_1 i_2 \dots i_T}^{(\lambda)}(\cdot) = \frac{1}{\lambda} \left[1 - \sum_{(u_1, u_2, \dots, u_T) \in D(i_1, i_2, \dots, i_T)} (p_{u_1 u_2 \dots u_T}^c)^{\lambda+1} \right].$$

Note that $\Phi^{(\lambda)}$ with $T = 3$ is analogous to $\Phi^{(\lambda)}$ considered in Section 2.

For $(r_1, r_2, \dots, r_T) \in S^*$, $\Phi_{r_1 r_2 \dots r_T}^{(\lambda)}$ must lie between 0 and 1, and therefore $\Phi^{(\lambda)}$ must lie between 0 and 1. Also, (i) the R^T table has structure S if and only if $\Phi_{r_1 r_2 \dots r_T}^{(\lambda)} = 0$ for all $(r_1, r_2, \dots, r_T) \in S^*$, i.e., $\Phi^{(\lambda)} = 0$, and (ii) the degree of departure from S is largest in the sense that, for $(u_1, u_2, \dots, u_T) \in D(i_1, i_2, \dots, i_T)$; $i_t = 1, 2, \dots, R$; $t = 1, 2, \dots, T$,

$$p_{u_1 u_2 \dots u_T}^c = \begin{cases} 1 & \text{for } (u_1, u_2, \dots, u_T) = (u_1^0, u_2^0, \dots, u_T^0) \\ & \text{with a cell } (u_1^0, u_2^0, \dots, u_T^0) \in D(i_1, i_2, \dots, i_T), \\ 0 & \text{for } (u_1, u_2, \dots, u_T) \neq (u_1^0, u_2^0, \dots, u_T^0), \end{cases}$$

if and only if $\Phi_{r_1 r_2 \dots r_T}^{(\lambda)} = 1$ for all $(r_1, r_2, \dots, r_T) \in S^*$ i.e., $\Phi_{r_1 r_2 \dots r_T}^{(\lambda)} = 1$.

4. Approximate confidence intervals for measures

Let $n_{i_1 i_2 \dots i_T}$ denote the observed frequency in the $(i_1, i_2, \dots, i_T)$ cell of the R^T table ($i_t = 1, 2, \dots, R$; $t = 1, 2, \dots, T$) with $T \geq 3$. Assuming that the $\{n_{i_1 i_2 \dots i_T}\}$ result from full multinomial sampling, we consider an approximate standard error and large-sample confidence interval for $\Phi^{(\lambda)}$, using the delta method, descriptions of which are given by, Bishop *et al.* (1975, Sec.14.6) and Agresti (1984, p.185, Appendix C). The sample version of $\Phi^{(\lambda)}$, $\hat{\Phi}^{(\lambda)}$, is given by $\Phi^{(\lambda)}$ with $\{p_{i_1 i_2 \dots i_T}\}$ replaced by $\{\hat{p}_{i_1 i_2 \dots i_T}\}$, where $\hat{p}_{i_1 i_2 \dots i_T} = n_{i_1 i_2 \dots i_T} / n$ and $n = \sum \sum \dots \sum n_{i_1 i_2 \dots i_T}$. Using the delta method, $\sqrt{n}(\hat{\Phi}^{(\lambda)} - \Phi^{(\lambda)})$ has asymptotically (as $n \rightarrow \infty$) a normal distribution with mean zero and variance $\sigma^2[\Phi^{(\lambda)}]$, where

$$\sigma^2[\Phi^{(\lambda)}] = \frac{1}{\delta^2} \left[\sum_{not \ i_1=i_2=\dots=i_T} \sum \dots \sum p_{i_1 i_2 \dots i_T} \left(\Delta_{i_1 i_2 \dots i_T}^{(\lambda)} \right)^2 - \delta \left(\Phi^{(\lambda)} \right)^2 \right] \quad \text{for } \lambda > -1; \lambda \neq 0,$$

with

$$\Delta_{i_1 i_2 \dots i_T}^{(\lambda)} = \frac{1}{\left(g(i_1, i_2, \dots, i_T)\right)^\lambda - 1} \left[(\lambda + 1) \left(g(i_1, i_2, \dots, i_T) p_{i_1 i_2 \dots i_T}^c\right)^\lambda - 1 \right. \\ \left. - \frac{\lambda p_{i_1 i_2 \dots i_T}^c}{p_{i_1 i_2 \dots i_T}} \sum_{(j_1, j_2, \dots, j_T) \in D(i_1, i_2, \dots, i_T)} p_{j_1 j_2 \dots j_T} \left(g(i_1, i_2, \dots, i_T) p_{j_1 j_2 \dots j_T}^c\right)^\lambda \right],$$

where

$$g(i_1, i_2, \dots, i_T) = {}_T C_{r_1 r_2 \dots r_T} \text{ with } (r_1, r_2, \dots, r_T) \text{ satisfying } (i_1, i_2, \dots, i_T) \in E_{(r_1, r_2, \dots, r_T)},$$

and for $\lambda = 0$,

$$\sigma^2[\Phi^{(0)}] = \frac{1}{\delta^2} \left[\sum_{not\ i_1=i_2=\dots=i_T} \sum \dots \sum p_{i_1 i_2 \dots i_T} \left(\Delta_{i_1, i_2, \dots, i_T}^{(0)}\right)^2 - \delta \left(\Phi^{(0)}\right)^2 \right],$$

with

$$\Delta_{i_1 i_2 \dots i_T}^{(0)} = 1 + \frac{\log(p_{i_1 i_2 \dots i_T}^c)}{\log(g(i_1, i_2, \dots, i_T))}.$$

We note that $\sigma^2[\Phi^{(0)}]$ is equal to the continuous limit of $\sigma^2[\Phi^{(\lambda)}]$ as $\lambda \rightarrow 0$.

Similarly, $\sqrt{n} \left(\hat{\Phi}_{i_1 i_2 \dots i_T}^{(\lambda)} - \Phi_{i_1 i_2 \dots i_T}^{(\lambda)} \right)$ has asymptotically (as $n \rightarrow \infty$) a normal distribution with mean zero and variance, $\sigma^2[\Phi_{i_1 i_2 \dots i_T}^{(\lambda)}]$, which is given in Appendix B.

Let $\hat{\sigma}^2[\Phi^{(\lambda)}]$ denote $\sigma^2[\Phi^{(\lambda)}]$ with $\{p_{i_1 i_2 \dots i_T}\}$ replaced by $\{\hat{p}_{i_1 i_2 \dots i_T}\}$. Then $\hat{\sigma}[\Phi^{(\lambda)}]/\sqrt{n}$ is an estimated standard error of $\hat{\Phi}^{(\lambda)}$, and $\hat{\Phi}^{(\lambda)} \pm z_{p/2} \hat{\sigma}[\Phi^{(\lambda)}]/\sqrt{n}$ is an approximate 100(1 - p)% confidence interval for $\hat{\Phi}^{(\lambda)}$, where $z_{p/2}$ is the percentage point from the standard normal distribution corresponding to a two-tail probability equal to p .

5. Examples

Example 1:

The data in Table 1 give the results of a multinational (13 countries) randomized clinical trial conducted by a pharmaceutical company (2001-2003) comparing two treatments (A and B) in patients infected with hepatitis B virus (HBV). Patients enrolled in the study, all of whom had raised blood levels of ALT, a specific liver enzyme serving as a marker for liver inflammation, were treated for 48 weeks with either A or B. They were then observed for a further 24 weeks with no treatment. The treatment was considered effective if ALT level fell to normal by the end of the follow-up period. Table 1 shows the $2 \times 2 \times 2$ array of counts of ALT response (normal or abnormal) at 48 weeks (end of the treatment, EOT), 60 weeks and 72 weeks (end of the follow-up, EOF) for each of the treatments A and B. Table 2 gives the estimates of the measure $\Phi^{(\lambda)}$, the estimated approximate standard error for $\hat{\Phi}^{(\lambda)}$ and the approximate 95% confidence interval for $\Phi^{(\lambda)}$.

Table 1. Serum ALT normal response at 48 weeks (EOT), 60 weeks and 72 weeks (EOF) in HBV patients from a randomized clinical trial.

(a) Treatment A				
Response at 72 weeks	Response at 48 weeks			
	Yes		No	
	Response at 60 weeks		Response at 60 weeks	
	Yes	No	Yes	No
Yes	44	29	0	7
No	13	46	1	41

(b) Treatment B				
Response at 72 weeks	Response at 48 weeks			
	Yes		No	
	Response at 60 weeks		Response at 60 weeks	
	Yes	No	Yes	No
Yes	43	21	23	20
No	8	15	12	37

(ALT response is a normalization of the serum ALT value: Yes, normal; No, abnormal).

Table 2. Estimates of $\Phi^{(\lambda)}$, estimated approximate standard errors for $\hat{\Phi}^{(\lambda)}$ and approximate 95% confidence intervals for $\Phi^{(\lambda)}$, applied to Table 1.

Group	Value of λ	Estimated measure	Standard error	Confidence interval
Treatment A	-0.4	0.467	0.055	(0.358, 0.576)
	0	0.510	0.060	(0.392, 0.628)
	0.6	0.516	0.067	(0.385, 0.648)
	1.0	0.502	0.071	(0.363, 0.642)
	1.4	0.482	0.074	(0.336, 0.628)
	2.0	0.444	0.078	(0.291, 0.597)
	2.4	0.417	0.080	(0.261, 0.573)
Treatment B	-0.4	0.039	0.023	(-0.006, 0.083)
	0	0.050	0.028	(-0.005, 0.104)
	0.6	0.052	0.028	(-0.003, 0.108)
	1.0	0.049	0.026	(-0.002, 0.101)
	1.4	0.044	0.023	(-0.002, 0.090)
	2.0	0.035	0.018	(-0.002, 0.071)
	2.4	0.029	0.015	(-0.002, 0.059)

Because the confidence intervals for $\Phi^{(\lambda)}$ applied to the data for treatment B in Table 1 include zero for any λ (> -1) (see Table 2), this would indicate that there is either symmetry with treatment B or the degree of departure from symmetry is slight. In contrast, the confidence interval of $\Phi^{(\lambda)}$ applied to the data for treatment A (see Table 2) indicate that there is asymmetry with treatment A.

In addition, when the degrees of departure from symmetry for each the treatments A and

B are compared using the confidence intervals for $\Phi^{(\lambda)}$, the degree of departure with treatment A would be stronger than that with treatment B. This is because, for any given $\lambda (> -1)$, the values in the confidence interval for $\Phi^{(\lambda)}$ applied to the data for treatment A are greater than the values in the corresponding confidence interval for $\Phi^{(\lambda)}$ applied to the data for treatment B.

We see from Table 1 that (i) the combination of observations such that only one of three responses is “Yes”, i.e., (Yes, No, No), (No, Yes, No), and (No, No, Yes), is (46, 1, 7) with treatment A and (15, 12, 20) with treatment B, and (ii) the combination of observations such that only one of three responses is “No”, i.e., (No, Yes, Yes), (Yes, No, Yes), and (Yes, Yes, No), is (0, 29, 13) with treatment A and (23, 21, 8) with treatment B. Therefore it seems natural to conclude that the degree of departure from symmetry is greater with treatment A than with treatment B, in terms of the measure $\hat{\Phi}^{(\lambda)}$.

Example 2:

Consider the data in Table 3, which are from Grizzle, Starmer and Koch (1969). Table 3 comes from a crossover study in which 46 subjects used each of three drugs for treatment of a chronic condition at three times. The response measured the reaction as favorable or unfavorable. The $2 \times 2 \times 2$ table gives the (favorable, unfavorable) classification for reaction to drug A in the first dimension, drug B in the second, and drug C in the third.

Since the confidence intervals for $\Phi^{(\lambda)}$ applied to the data in Table 3 do not include zero for any $\lambda (> -1)$ (see Table 4), this would indicate that there is asymmetry in the data of Table 3.

We shall further analyze the data of Table 3 using $\Phi_{ijk}^{(\lambda)}$ with $(i, j, k) = (1, 1, 2), (1, 2, 2)$. Table 5 gives the estimates of $\Phi_{ijk}^{(\lambda)}$, the estimated approximate standard errors for $\hat{\Phi}_{ijk}^{(\lambda)}$ and the approximate 95% confidence intervals for $\Phi_{ijk}^{(\lambda)}$. We see from Table 3 that (i) the combination of observations such that only one of the three responses for drugs A, B and C is “Unfavorable”, i.e., (Favorable, Favorable, Unfavorable), (Favorable, Unfavorable, Favorable), (Unfavorable, Favorable, Favorable), is (16, 2, 2), and (ii) the combination of observations such that only one of the three responses for drugs A, B and C is “Favorable”, i.e., (Favorable, Unfavorable, Unfavorable), (Unfavorable, Favorable, Unfavorable), (Unfavorable, Unfavorable, Favorable), is (4, 4, 6). Then we see from Table 5 that the value of the measure $\Phi_{112}^{(\lambda)}$ for the combination of

Table 3. Responses to three drugs in a crossover study.

Response to C	Response to A			
	(1) Favorable		(2) Unfavorable	
	Response to B		Response to B	
	(1) Favorable	(2) Unfavorable	(1) Favorable	(2) Unfavorable
(1) Favorable	6	2	2	6
(2) Unfavorable	16	4	4	6

Taken from Grizzle *et al.* (1969).

Table 4. Estimates of $\Phi^{(\lambda)}$, estimated approximate standard errors for $\hat{\Phi}^{(\lambda)}$ and approximate 95% confidence intervals for $\Phi^{(\lambda)}$, applied to the data of Table 3.

Value of λ	Estimated measure	Standard error	Confidence interval
-0.4	0.192	0.086	(0.022, 0.361)
0	0.254	0.107	(0.044, 0.463)
0.6	0.292	0.118	(0.061, 0.523)
1.0	0.297	0.120	(0.062, 0.531)
1.4	0.291	0.120	(0.056, 0.527)
2.0	0.273	0.119	(0.039, 0.507)
2.4	0.258	0.118	(0.025, 0.490)

Table 5. Estimates of $\Phi_{ijk}^{(\lambda)}, (i, j, k) = (1, 1, 2), (1, 2, 2)$, estimated approximate standard errors for $\hat{\Phi}_{ijk}^{(\lambda)}$ and approximate 95% confidence interval for $\Phi_{ijk}^{(\lambda)}$, applied to the data of Table 3.

$\Phi_{ijk}^{(\lambda)}$	Value of λ	Estimated measure	Standard error	Confidence interval
$\Phi_{112}^{(\lambda)}$	-0.4	0.317	0.138	(0.046, 0.587)
	0	0.419	0.169	(0.087, 0.750)
	0.6	0.482	0.185	(0.120, 0.844)
	1.0	0.490	0.188	(0.122, 0.858)
	1.4	0.482	0.189	(0.111, 0.853)
	2.0	0.453	0.190	(0.081, 0.826)
	2.4	0.428	0.190	(0.055, 0.801)
$\Phi_{122}^{(\lambda)}$	-0.4	0.013	0.036	(-0.057, 0.083)
	0	0.018	0.049	(-0.078, 0.114)
	0.6	0.021	0.057	(-0.091, 0.132)
	1.0	0.020	0.057	(-0.091, 0.132)
	1.4	0.019	0.053	(-0.086, 0.124)
	2.0	0.016	0.046	(-0.073, 0.105)
	2.4	0.014	0.039	(-0.064, 0.091)

observations such that only one of the three responses is “Unfavorable” is greater than the value of measure $\Phi_{122}^{(\lambda)}$ for the combination of responses such that only one of the three responses is “favorable”. Therefore, judging which of the drugs A, B and C is most unfavorable may be easier than judging which of them is the most favorable. Indeed, the value of observation such that the combination of responses for drugs (A, B, C) is (Favorable, Favorable, Unfavorable), 16, is the greatest.

6. Concluding remarks

The measure $\hat{\Phi}^{(\lambda)}$ could be useful when we want to see whether there is symmetry, and when we want to compare the degrees of asymmetry in categorical response data that many studies observe for each subject repeatedly, at several times or under various conditions, as shown in

Example 1. In addition, the measure $\hat{\Phi}^{(\lambda)}$ may be useful to compare three (or more) drugs in a crossover study in which each subject uses each of those drugs at several times, as shown in Example 2. In the case of a crossover study, when symmetry does not hold, we are interested in finding which symmetrically located cells contribute most to the departure from symmetry. By examining the internal structure of symmetry, we can know which of the drug combination responses are the strongest. However, we cannot obtain such information merely by examining the degree of departure from marginal homogeneity.

We assume that there is an underlying three-dimensional normal distribution with equal variances but different means for the first, second and third variables, and that an $R \times R \times R$ table is formed using cutpoints for each variable. Consider random variables Z_1, Z_2 and Z_3 having a joint three-dimensional normal distribution with means $E(Z_1) = \mu_1$, $E(Z_2) = \mu_2$ and $E(Z_3) = \mu_3$, variances $\text{var}(Z_1) = \text{var}(Z_2) = \text{var}(Z_3) = \sigma^2$, and correlation $\text{corr}(Z_1, Z_2) = \text{corr}(Z_1, Z_3) = \text{corr}(Z_2, Z_3) = \rho$, namely, with intraclass correlation structure. Tables 6a and 7a give the simulated $4 \times 4 \times 4$ tables of sample size 10000, formed by using cutpoints for each variable at $\mu_1, \mu_1 \pm 0.6\sigma$, for an underlying three-dimensional normal distribution with $\sigma^2 = 1$ and $\rho = 0, 0.3, 0.6, 0.9$. Tables 6b and 7b give the corresponding values of $\hat{\Phi}^{(\lambda)}$. We can see from Tables 6b and 7b that (i) for any fixed λ , $\hat{\Phi}^{(\lambda)}$ in Table 6a increase as ρ increases, (ii) for any fixed λ , $\hat{\Phi}^{(\lambda)}$ in Table 7a also increase as ρ increases, and (iii) for any fixed ρ and λ , $\hat{\Phi}^{(\lambda)}$ is greater for Table 7a with $\mu_2 - \mu_1 = 0.6$ than for Table 6b with $\mu_2 - \mu_1 = 0.4$. As $\rho (\geq 0)$ approaches 1, the three-way table with an underlying three-dimensional normal distribution (with equality of variances) tends to have zero observations in one or the other of the off-diagonals (see Tables 6a and 7a). Therefore, it seems natural to assume that as ρ approaches 1, the degree of asymmetry for the observed three-way contingency tables tends to approach the maximum. So, $\hat{\Phi}^{(\lambda)}$ may be useful for measuring the degree of asymmetry in a three-way table formed from an underlying three-way normal distribution with equal variances and equal correlation.

We point out that generally the confidence intervals of $\Phi^{(\lambda)}$ might be expected to perform less well for small (large) values of $\Phi^{(\lambda)}$, so that the lower (upper) bound is near 0 (1) or even beyond the range $[0, 1]$ of $\Phi^{(\lambda)}$. Consider the following simulation studies. Table 8 gives the artificial $3 \times 3 \times 3$ tables of sample size $n = 50$ (Table 8b), $n = 100$ (Table 8c), $n = 500$ (Table 8d), and $n = 1000$ (Table 8e), which are obtained from the probability distribution with $\Phi^{(\lambda)}$ being near 0 such that $\Phi^{(-0.4)} = 0.048$, $\Phi^{(0)} = 0.059$, $\Phi^{(0.6)} = 0.059$, $\Phi^{(1)} = 0.054$, $\Phi^{(2.2)} = 0.031$, and $\Phi^{(3)} = 0.020$, assumed in Table 8a. Table 9 gives the values of $\hat{\Phi}^{(\lambda)}$ and the estimated 95% confidence intervals for each table. From Table 8 we see that (i) $\hat{\Phi}^{(\lambda)}$ approaches $\Phi^{(\lambda)}$ for all λ as the sample size n increases, and (ii) when $n = 100$ and $\lambda = 3$, the lower confidence bound for $\Phi^{(\lambda)}$ is beyond 0. So, generally, when $\Phi^{(\lambda)}$ is near 0 or 1 and when n is small, the confidence interval for $\Phi^{(\lambda)}$ may extend beyond the range $[0, 1]$.

For testing the goodness-of-fit of a model, the analyst usually uses the likelihood ratio chi-

Table 6. (a) The $4 \times 4 \times 4$ tables of sample size 10000, formed by using cutpoints for each variable at $\mu_1, \mu_1 \pm 0.6\sigma$, from an underlying three-variate normal distribution with the conditions $\mu_2 = \mu_1 + 0.4, \mu_3 = \mu_1 + 0.8, \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1 (= \sigma^2)$ and $\rho = 0, 0.3, 0.6, 0.9$, and (b) the values of $\hat{\Phi}^{(\lambda)}$.

(i) $\rho = 0$															
1				2				3				4			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
28	54	78	261	41	46	99	320	50	73	116	384	102	152	228	671
23	47	87	215	33	60	100	250	33	73	97	349	66	119	214	595
26	55	79	204	25	58	83	229	43	67	124	331	65	128	217	532
36	61	86	219	47	60	84	289	55	88	119	370	100	141	249	666
(ii) $\rho = 0.3$															
1				2				3				4			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
146	118	200	224	76	112	175	255	75	101	156	294	68	111	169	423
58	73	111	151	40	82	133	228	38	81	118	343	31	89	187	598
32	55	65	132	24	68	84	215	32	54	109	361	28	95	192	720
17	35	47	93	22	45	73	215	28	58	91	399	36	97	192	1222
(iii) $\rho = 0.6$															
1				2				3				4			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
332	311	240	115	119	197	237	180	67	128	144	217	23	61	107	225
54	97	114	86	42	108	181	236	28	78	192	386	8	68	151	532
17	27	57	55	16	57	97	209	9	54	146	400	9	37	187	889
2	6	19	30	3	17	44	95	3	36	85	358	4	29	152	1787
(iv) $\rho = 0.9$															
1				2				3				4			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
727	531	156	9	79	364	414	64	3	65	168	100	0	1	18	4
7	57	63	7	9	140	440	174	1	51	442	591	0	2	75	302
0	0	4	2	0	14	70	81	0	6	174	642	0	2	64	1207
0	0	0	0	0	0	0	9	0	0	8	123	0	0	14	2516

Table 6. (continued) (b) the values of $\hat{\Phi}^{(\lambda)}$ for Table 6a.

λ	$\rho = 0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.9$
-0.4	0.088	0.129	0.212	0.587
0	0.111	0.162	0.259	0.663
0.6	0.116	0.170	0.271	0.692
1.0	0.109	0.162	0.259	0.686
2.2	0.075	0.117	0.198	0.629

squared statistic (denoted by G^2) or Pearson's chi-squared statistic (denote by X^2). Cressie and Read (1984) and Read and Cressie (1988) proposed the power-divergence statistic, which includes G^2 and X^2 as special cases. Let $W^{(\lambda)}$ denote the power-divergence statistic for testing goodness-of-fit of the symmetry model for a three-way contingency table (see Appendix C for

Table 7. (a) The $4 \times 4 \times 4$ tables of sample size 10000, formed by using cutpoints for each variable at $\mu_1, \mu_1 \pm 0.6\sigma$, from an underlying 3-variate normal distribution with the conditions $\mu_2 = \mu_1 + 0.6, \mu_3 = \mu_1 + 1.2, \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1 (= \sigma^2)$ and $\rho = 0, 0.3, 0.6, 0.9$, and (b) the values of $\hat{\Phi}^{(\lambda)}$.

(i) $\rho = 0$															
1				2				3				4			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
8	31	46	227	12	31	53	308	24	35	82	473	63	99	213	998
8	20	45	204	12	22	54	268	17	39	86	430	25	80	190	861
8	20	45	182	8	30	68	252	15	36	67	368	38	91	195	843
11	28	41	201	19	37	75	302	26	56	78	425	61	101	203	1006
(ii) $\rho = 0.3$															
1				2				3				4			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
59	85	112	252	43	60	140	347	42	76	126	388	40	80	183	670
17	42	73	156	17	41	79	269	15	53	94	411	16	52	156	870
12	17	46	117	10	29	73	226	8	37	73	370	13	44	165	1026
5	13	26	88	9	19	48	202	9	33	67	397	13	55	160	1526
(iii) $\rho = 0.6$															
1				2				3				4			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
174	178	246	189	62	130	224	293	40	95	170	356	10	40	114	382
17	44	82	97	11	61	125	276	10	39	138	493	5	18	121	824
3	13	25	57	3	20	68	190	3	23	84	425	2	19	96	1235
1	1	6	22	0	4	24	91	1	11	36	302	0	12	74	2085
(iv) $\rho = 0.9$															
1				2				3				4			
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
331	464	261	34	23	218	533	238	2	35	202	283	0	2	17	60
1	10	28	12	1	16	223	259	0	12	196	946	0	1	34	622
0	0	0	1	0	1	15	53	0	0	36	537	0	0	10	1613
0	0	0	0	0	0	0	0	0	0	0	66	0	0	0	2604

Table 7. (continued) (b) the values of $\hat{\Phi}^{(\lambda)}$ for Table 7a.

λ	$\rho = 0$	$\rho = 0.3$	$\rho = 0.6$	$\rho = 0.9$
-0.4	0.183	0.253	0.403	0.812
0	0.228	0.307	0.470	0.864
0.6	0.239	0.320	0.488	0.885
1.0	0.228	0.307	0.475	0.884
2.2	0.171	0.241	0.403	0.855

details). Note that $W^{(0)}$ and $W^{(1)}$ are identical to G^2 and X^2 , respectively.

Consider two tables, A and B. We are interested in which of them better conforms to the symmetry assumption. When one wants to compare the degrees of departure from symmetry in several tables, it may be dangerous to use the test statistic $W^{(\lambda)}$ because, for fixed λ , the value

Table 8. Simulation studies. Table 8a assumes the probability distribution $\{p_{ijk}\}$ with $\Phi^{(-0.4)} = 0.048$, $\Phi^{(0)} = 0.059$, $\Phi^{(0.6)} = 0.059$, $\Phi^{(1)} = 0.054$, $\Phi^{(2.2)} = 0.031$, and $\Phi^{(3)} = 0.020$. Tables 8b, 8c, 8d and 8e are the artificial $3 \times 3 \times 3$ tables of sample sizes $n = 50, 100, 500, 1000$ obtained from the $\{p_{ijk}\}$ by simulation.

(a) Values of $\{p_{ijk}\}$								
1			2			3		
1	2	3	1	2	3	1	2	3
0.05	0.02	0.03	0.04	0.04	0.03	0.01	0.07	0.03
0.06	0.03	0.02	0.06	0.05	0.01	0.06	0.03	0.04
0.04	0.06	0.03	0.02	0.03	0.02	0.03	0.04	0.05

(b) $n = 50$								
1			2			3		
1	2	3	1	2	3	1	2	3
1	0	2	0	4	3	0	5	4
3	0	2	1	3	0	3	3	4
1	4	0	0	2	2	0	2	1

(c) $n = 100$								
1			2			3		
1	2	3	1	2	3	1	2	3
4	2	2	3	5	3	1	8	5
3	2	4	4	5	0	10	3	5
7	4	1	1	4	4	3	3	4

(d) $n = 500$								
1			2			3		
1	2	3	1	2	3	1	2	3
26	14	11	19	17	17	6	34	12
31	19	12	23	17	8	37	16	25
19	29	11	9	16	10	17	19	26

(e) $n = 1000$								
1			2			3		
1	2	3	1	2	3	1	2	3
44	21	21	44	40	37	15	85	24
54	32	20	50	38	13	68	39	50
39	49	34	21	27	15	35	38	47

of measure $\hat{\Phi}^{(\lambda)}$ may be less for Table A than for Table B, yet the value of $W^{(\lambda)}$ greater for Table A than for Table B. For example, consider the artificial $2 \times 2 \times 2$ tables in Tables 10a and 10b. We see that (i) for Table 10a, the sample proportions for the combination of cells (1,1,2) and (1,2,2) are (0.4, 0.15, 0.05) and (0.1, 0.1, 0.1), respectively, and (ii) for Table 10b they are (0.2, 0.1, 0.1) and (0.1, 0.1, 0.1), respectively. Therefore it seems natural to conclude that the degree of departure from symmetry is greater for Table 10a than for Table 10b. We see from Table 11 that the values of $\hat{\Phi}^{(\lambda)}$ are greater for Table 10a than for Table 10b, but the values of $W^{(\lambda)}$ are less for Table 10a than for Table 10b. Therefore, the measure $\hat{\Phi}^{(\lambda)}$ may be useful for comparing the degree of departure from symmetry in several multi-way tables.

Table 9. Estimates of $\Phi^{(\lambda)}$, estimated approximate standard errors for $\hat{\Phi}^{(\lambda)}$ and approximate 95% confidence intervals for $\Phi^{(\lambda)}$, applied to the data of Tables 8b, 8c, 8d and 8e.

Table (sample size n)	Value of λ	Estimated measure	Standard error	Confidence interval
Table 8b ($n = 50$)	-0.4	0.360	0.055	(0.252, 0.468)
	0	0.345	0.062	(0.224, 0.465)
	0.6	0.319	0.069	(0.184, 0.453)
	1.0	0.301	0.071	(0.162, 0.441)
	2.2	0.256	0.073	(0.112, 0.400)
	3.0	0.233	0.073	(0.091, 0.376)
Table 8c ($n = 100$)	-0.4	0.111	0.033	(0.046, 0.176)
	0	0.125	0.039	(0.048, 0.201)
	0.6	0.121	0.041	(0.041, 0.200)
	1.0	0.110	0.039	(0.033, 0.187)
	2.2	0.071	0.033	(0.007, 0.136)
	3.0	0.050	0.028	(-0.005, 0.106)
Table 8d ($n = 500$)	-0.4	0.040	0.010	(0.020, 0.060)
	0	0.049	0.013	(0.025, 0.074)
	0.6	0.050	0.013	(0.024, 0.075)
	1.0	0.045	0.012	(0.021, 0.069)
	2.2	0.026	0.009	(0.009, 0.044)
	3.0	0.017	0.006	(0.005, 0.029)
Table 8e ($n = 1000$)	-0.4	0.048	0.008	(0.032, 0.063)
	0	0.059	0.010	(0.040, 0.078)
	0.6	0.059	0.010	(0.040, 0.079)
	1.0	0.054	0.009	(0.036, 0.072)
	2.2	0.032	0.006	(0.019, 0.044)
	3.0	0.020	0.005	(0.011, 0.029)

Note: $\Phi^{(-0.4)} = 0.048$, $\Phi^{(0)} = 0.059$, $\Phi^{(0.6)} = 0.059$, $\Phi^{(1)} = 0.054$, $\Phi^{(2.2)} = 0.031$, $\Phi^{(3)} = 0.020$ for Table 8a.

Table 10. Artificial $2 \times 2 \times 2$ data for sample sizes (n) of 100 and 1000.

(a) $n = 100$					(b) $n = 1000$				
i					i				
k	1		2		k	1		2	
	j		j			j		j	
	1	2	1	2		1	2	1	2
1	5	15	5	10	1	150	100	100	100
2	40	10	10	5	2	200	100	100	150

The reader may be interested in which value of λ is preferred for a given table, but we cannot provide clear guidance. It seems to be important that, for given tables, the analyst calculate the values of $\hat{\Phi}^{(\lambda)}$ for various values of λ and assess the degree of departure from symmetry in terms of $\hat{\Phi}^{(\lambda)}$. In addition, if the analyst wants to see with a *minimum* distance measure how distant the cell probabilities are from those under symmetry, the case of $\lambda = 0$, *i.e.*, $\hat{\Phi}^{(0)}$ (see Section 3), may

Table 11. Estimates $\hat{\Phi}^{(\lambda)}$ and values of the power-divergence statistic $W^{(\lambda)}$ (with 4 degrees of freedom) for testing goodness-of-fit of the symmetry model, applied to the data of Tables 10a and 10b.

(a) $n = 100$			(b) $n = 1000$		
λ	$\hat{\Phi}^{(\lambda)}$	$W^{(\lambda)}$	λ	$\hat{\Phi}^{(\lambda)}$	$W^{(\lambda)}$
-0.4	0.130	34.59	-0.4	0.022	46.27
0	0.167	32.96	0	0.031	47.11
0.6	0.167	32.14	0.6	0.036	48.70
1.0	0.181	32.50	1.0	0.036	50.00
2.2	0.144	37.50	2.2	0.027	55.18

be recommended, because $\{\hat{p}_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^s\}$ is the value of $\{\hat{p}_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^{*es}\}$ such that the Kullback-Leibler distance $I_{r_1 r_2 \dots r_T}^{(0)}(\{\hat{p}_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^*\}; \{\hat{p}_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^{*es}\})$ for $(r_1, r_2, \dots, r_T) \in S^*$, which indicates the distance between the sample conditional distribution $\{\hat{p}_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^*\}$ and the estimated conditional distribution $\{\hat{p}_{i_1 i_2 \dots i_T [r_1 r_2 \dots r_T]}^{*es}\}$ under symmetry, is a *minimum*.

Finally, we observe that for the R^T table, the asymptotic normal distribution of $\sqrt{n}(\hat{\Phi}^{(\lambda)} - \Phi^{(\lambda)})$ using the delta method may not be applicable in the case of sparse data having small cell counts. However, it is beyond the scope of this paper, the purpose of which is to propose a measure that represents the degree of departure from symmetry for multi-way tables. We intend to conduct further research to address the impact of sparse tables.

Acknowledgements

The author would like to express his sincere thanks to Professor S. Tomizawa and Dr. N. Miyamoto, Tokyo University of Science, for their many helpful comments and encouragement. Special thanks are also due to the two referees and the editor for comments on an earlier version of this paper.

REFERENCES

- Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. New York: Wiley.
- Agresti, A. (2002). *Categorical Data Analysis (2nd edition)*. New York: Wiley.
- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.
- Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, **43**, 572-574.
- Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, **46**, 440-464.
- Grizzle, J. E., C. F. Starmer, and G. G. Koch. (1969). Analysis of categorical data by linear models. *Biometrics*, **25**, 489-504.

- Patil, G. P. and Taillie, C. (1982). Diversity as a concept and its measurement. *Journal of the American Statistical Association*, **77**, 548-561.
- Read, T. R. C. and Cressie, N. (1988). *Goodness-of-fit Statistics for Discrete Multivariate Data*. New York: Springer-Verlag.
- Tomizawa, S. (1994). Two kinds of measures of departure from symmetry in square contingency tables having nominal categories. *Statistica Sinica*, **4**, 325-334.
- Tomizawa, S., Seo, T., and Yamamoto, H. (1998). Power-divergence-type measure of departure from symmetry for square contingency tables that have nominal categories. *Journal of Applied Statistics*, **25**, 387-398.
- Tomizawa, S., Miyamoto, N., and Hatanaka, Y. (2002). Measure of asymmetry for square contingency tables having ordered categories. *Australian and New Zealand Journal of Statistics*, **43**, 335-349.

Appendix

A

For an $R \times R$ table, the measure of departure from symmetry, which is considered by Tomizawa, Seo and Yamamoto (1998), is given as follows: assuming that $\{p_{ij} + p_{ji} \neq 0\}$,

$$\Phi^{(\lambda)} = \frac{\lambda(\lambda+1)}{2^\lambda - 1} I^{(\lambda)}(\{p_{ij}^*\}; \{p_{ij}^s\}) \quad \text{for } \lambda > -1,$$

where $\delta = \sum \sum_{i \neq j} p_{ij}$, $p_{ij}^* = p_{ij}/\delta$ and $p_{ij}^s = (p_{ij}^* + p_{ji}^*)/2$,

$$I^{(\lambda)}(\cdot; \cdot) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^R \sum_{\substack{j=1 \\ j \neq i}}^R p_{ij}^* \left[\left(\frac{p_{ij}^*}{p_{ij}^s} \right)^\lambda - 1 \right],$$

and the value at $\lambda = 0$ is taken to be the continuous limit as $\lambda \rightarrow 0$. Note that in special cases $\Phi^{(0)}$ and $\Phi^{(1)}$ are identical to the measures given by Tomizawa (1994). In addition, we note that $I^{(\lambda)}(\{p_{ij}^*\}; \{p_{ij}^s\})$ is the power-divergence discrepancy between two conditional distributions $\{p_{ij}^*\}$ and $\{p_{ij}^s\}$, $i \neq j$.

B

Consider the R^T table with $T \geq 3$. Using the delta method, for $1 \leq i_1 \leq i_2 \leq \dots \leq i_T \leq R$, $\sqrt{n}(\hat{\Phi}_{i_1 i_2 \dots i_T}^{(\lambda)} - \Phi_{i_1 i_2 \dots i_T}^{(\lambda)})$ has asymptotically (as $n \rightarrow \infty$) a normal distribution with mean zero and variance $\sigma^2[\Phi_{i_1 i_2 \dots i_T}^{(\lambda)}]$, where for $\lambda > -1$ ($\lambda \neq 0$),

$$\begin{aligned} \sigma^2[\Phi_{i_1 i_2 \dots i_T}^{(\lambda)}] = & \left[\frac{(\lambda+1)(g(i_1, i_2, \dots, i_T))^\lambda}{(g(i_1, i_2, \dots, i_T))^\lambda - 1} \right]^2 \frac{1}{\sum_{(v_1, v_2, \dots, v_T) \in D(i_1, i_2, \dots, i_T)} \sum \dots \sum p_{v_1 v_2 \dots v_T}} \\ & \times \sum_{(j_1, j_2, \dots, j_T) \in D(i_1, i_2, \dots, i_T)} \sum \dots \sum p_{j_1 j_2 \dots j_T}^c \left[(p_{j_1 j_2 \dots j_T}^c)^\lambda \right. \\ & \left. - \sum_{(u_1, u_2, \dots, u_T) \in D(i_1, i_2, \dots, i_T)} \sum \dots \sum (p_{u_1 u_2 \dots u_T}^c)^{\lambda+1} \right]^2, \end{aligned}$$

and for $\lambda = 0$,

$$\begin{aligned} \sigma^2[\Phi_{i_1 i_2 \dots i_T}^{(0)}] = & \left[\frac{1}{\log(g(i_1, i_2, \dots, i_T))} \right]^2 \frac{1}{\sum_{(v_1, v_2, \dots, v_T) \in D(i_1, i_2, \dots, i_T)} \sum \dots \sum p_{v_1 v_2 \dots v_T}} \\ & \times \sum_{(j_1, j_2, \dots, j_T) \in D(i_1, i_2, \dots, i_T)} \sum \dots \sum p_{j_1 j_2 \dots j_T}^c \left[\log(p_{j_1 j_2 \dots j_T}^c) \right. \\ & \left. - \sum_{(u_1, u_2, \dots, u_T) \in D(i_1, i_2, \dots, i_T)} \sum \dots \sum p_{u_1 u_2 \dots u_T}^c \log(p_{u_1 u_2 \dots u_T}^c) \right]^2. \end{aligned}$$

Note that $\sigma^2[\Phi_{i_1 i_2 \dots i_T}^{(0)}]$ is equal to the limit of $\sigma^2[\Phi_{i_1 i_2 \dots i_T}^{(\lambda)}]$ as $\lambda \rightarrow 0$.

C

For an $R \times R \times R$ table, the power-divergence chi-squared statistic for testing goodness-of-fit of the symmetry model is given, for $-\infty < \lambda < \infty$, by

$$W^{(\lambda)} = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^R \sum_{j=1}^R \sum_{k=1}^R n_{ijk} \left[\left(\frac{6n_{ijk}}{n_{ijk} + n_{ikj} + n_{jik} + n_{jki} + n_{kij} + n_{kji}} \right)^\lambda - 1 \right],$$

where the values at $\lambda = -1$ and $\lambda = 0$ are taken to be the limits as $\lambda \rightarrow -1$ and $\lambda \rightarrow 0$, respectively. The number of degrees of freedom is $R(R-1)(5R+2)/6$.