Entropy **2008**, *10*, 776-785; DOI: 10.3390/e10040776

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Generalized Measure of Departure from No Three-Factor Interaction Model for $2 \times 2 \times K$ Contingency Tables

Kouji Yamamoto, Yohei Ban and Sadao Tomizawa ⋆

Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Noda City, Chiba, 278-8510, Japan.
E-mails: yamamoto_kouji@ymail.plala.or.jp; tomizawa@is.noda.tus.ac.jp

⋆ Author to whom correspondence should be addressed.

Received: 31 October 2008 / Accepted: 16 December 2008 / Published: 22 December 2008

Abstract: For $2 \times 2 \times K$ contingency tables, Tomizawa considered a Shannon entropy type measure to represent the degree of departure from a log-linear model of no three-factor interaction (the NOTFI model). This paper proposes a generalization of Tomizawa's measure for $2 \times 2 \times K$ tables. The measure proposed is expressed by using Patil-Taillie diversity index or Cressie-Read power-divergence. A special case of the proposed measure includes Tomizawa's measure. The proposed measure would be useful for comparing the degrees of departure from the NOTFI model in several tables.

Keywords: Diversity index; odds-ratio; power-divergence.

---

1. Introduction

For the $I \times J \times K$ contingency table, let p_{ijk} denote the probability that an observation will fall in the cell (i, j, k) of the table $(i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K)$. One can express $\log p_{ijk}$ as

$$\log p_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}, \tag{1}$$

where

$$\sum_i u_{s(i)} = 0 \quad (s = 1, 2, 3),$$
$$\sum_i u_{st(ij)} = \sum_j u_{st(ij)} = 0 \quad (1 \leq s < t \leq 3),$$

$$\sum_i u_{123(ijk)} = \sum_j u_{123(ijk)} = \sum_k u_{123(ijk)} = 0;$$

see, e.g., Bishop, Fienberg and Holland [1, Chap. 2]. Let $l_{ijk} = \log p_{ijk}$. The u-term in (1) are, for example,

$$u = \frac{l_{\cdot\cdot\cdot}}{IJK} \quad \text{(overall mean)},$$
$$u_{1(i)} = \frac{l_{i\cdot\cdot}}{JK} - \frac{l_{\cdot\cdot\cdot}}{IJK} \quad \text{(main effect of variable 1)},$$
$$u_{12(ij)} = \frac{l_{ij\cdot}}{K} - \left(\frac{l_{i\cdot\cdot}}{JK} + \frac{l_{\cdot j\cdot}}{IK}\right) + \frac{l_{\cdot\cdot\cdot}}{IJK} \quad \text{(two-factor effect between variables 1 and 2)},$$

and

$$u_{123(ijk)} = l_{ijk} - (u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)})$$
$$\text{(three-factor effect (interaction))},$$

where

$$l_{\cdot\cdot\cdot} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} l_{ijk}, \quad l_{i\cdot\cdot} = \sum_{j=1}^{J}\sum_{k=1}^{K} l_{ijk}, \quad l_{ij\cdot} = \sum_{k=1}^{K} l_{ijk}, \quad l_{\cdot j\cdot} = \sum_{i=1}^{I}\sum_{k=1}^{K} l_{ijk};$$

see, e.g., Bishop et al. [1, Chap. 2].

We obtain the well-known four models by setting the parameters in (1) as

$$\text{(i) } u_{12(ij)} = u_{13(ik)} = u_{23(jk)} = u_{123(ijk)} = 0,$$
$$\text{(ii) } u_{13(ik)} = u_{23(jk)} = u_{123(ijk)} = 0,$$
$$\text{(iii) } u_{13(ik)} = u_{123(ijk)} = 0,$$
$$\text{(iv) } u_{123\ (ijk)} = 0,$$

for all i, j, k. Model (1) imposed restriction (iv) is usually referred to as the no three-factor interaction (NOTFI) model (or no second-order interaction model). Model (1) imposed restrictions (i), (ii), (iii) and (iv) also can be expressed as

$$\mathrm{H}_1 : p_{ijk} = p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k},$$
$$\mathrm{H}_2 : p_{ijk} = p_{ij\cdot}p_{\cdot\cdot k},$$
$$\mathrm{H}_3 : p_{ijk} = \frac{p_{ij\cdot}p_{\cdot jk}}{p_{\cdot j\cdot}},$$
$$\mathrm{H}_4 : \theta_{ij(1)} = \cdots = \theta_{ij(K)},$$

respectively, where

$$p_{i\cdot\cdot} = \sum_j\sum_k p_{ijk}, \quad p_{\cdot j\cdot} = \sum_i\sum_k p_{ijk}, \quad p_{\cdot\cdot k} = \sum_i\sum_j p_{ijk},$$
$$p_{ij\cdot} = \sum_k p_{ijk}, \quad p_{\cdot jk} = \sum_i p_{ijk},$$
$$\theta_{ij(t)} = \frac{p_{ijt}p_{i+1,j+1,t}}{p_{i,j+1,t}p_{i+1,j,t}};$$

see, e.g., Fienberg [2, Chap. 3]. When none of models $\mathrm{H}_1,\ \mathrm{H}_2,\ \mathrm{H}_3$ and H_4 holds, namely, when model H_4 does not hold, we are interested in seeing the degree of departure from model H_4, i.e., the degree of non-uniformity of odds-ratios $\{\theta_{ij(t)}\}$.

For the $2 \times 2 \times K$ contingency table, Tomizawa [3] considered a measure which represents the degree of departure from the NOTFI model. The measure is expressed by using the Shannon entropy (see Appendix).

By the way, Patil and Taillie [4] considered the diversity index, which includes the Shannon entropy in a special case. We are interested in a measure of departure from the NOTFI model, based on the diversity index.

The purpose of this paper is to propose a generalization of Tomizawa's measure for the $2 \times 2 \times K$ table. The proposed measure includes Tomizawa's measure in a special case. The measure would be useful for comparing the degrees of departure from the NOTFI model in several tables.

2. A generalization of measure

Consider the $2 \times 2 \times K$ contingency table. The NOTFI model is expressed as

$$\theta_1 = \theta_2 = \cdots = \theta_K,$$

where

$$\theta_t = \frac{p_{11t}p_{22t}}{p_{12t}p_{21t}}.$$

This shows that the K odds-ratios are identical. Let

$$D = \sum_{k=1}^{K} \theta_k, \quad \theta_t^* = \frac{\theta_t}{D},$$

for $t = 1, \ldots, K$.

Assuming that the $\{p_{ijk}\}$ are positive, consider a measure to represent the degree of departure from the NOTFI model, defined by

$$\varphi^{(\lambda)} = 1 - \frac{H^{(\lambda)}(\theta^*)}{C^{(\lambda)}}, \quad \text{for } \lambda > -1 \tag{2}$$

where

$$\begin{aligned} H^{(\lambda)}(\theta^*) &= \frac{1}{\lambda}\left(1 - \sum_{t=1}^{K}(\theta_t^*)^{\lambda+1}\right), \\ C^{(\lambda)} &= \frac{1}{\lambda}\left[1 - \left(\frac{1}{K}\right)^{\lambda}\right], \end{aligned}$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$, where λ is a real value that is chosen by the user. Thus, $\varphi^{(0)}$ is equal to φ in Appendix. Note that $\varphi^{(0)}$ in equation (2) is the same as Tomizawa's measure. Also, note that $H^{(\lambda)}(\theta^*)$ is Patil and Taillie's diversity index of degree λ for $\{\theta_t^*\}$, which includes the Shannon entropy (when $\lambda = 0$) in a special case.

The measure $\varphi^{(\lambda)}$ may be expressed as

$$\varphi^{(\lambda)} = \frac{\lambda+1}{K^{\lambda}C^{(\lambda)}} I^{(\lambda)}\left(\{\theta_t^*\};\left\{\frac{1}{K}\right\}\right),$$

where

$$I^{(\lambda)}(\cdot;\cdot) = \frac{1}{\lambda(\lambda+1)}\sum_{t=1}^{K}\theta_t^*\left[\left(\frac{\theta_t^*}{1/K}\right)^{\lambda}-1\right].$$

Note that $I^{(\lambda)}(\{\theta_t^*\};\{\frac{1}{K}\})$ is the power-divergence between $\{\theta_t^*\}$ and $\{\frac{1}{K}\}$. For more details of the power-divergence $I^{(\lambda)}(\cdot;\cdot)$, see Cressie and Read [5], and Read and Cressie [6, p. 15].

The $H^{(\lambda)}(\theta^*)$ must lie between 0 and $C^{(\lambda)}$ but it cannot attain the lower limit of 0 in terms of the assumption that the $\{p_{ijk}\}$ are positive. Thus the measure $\varphi^{(\lambda)}$ must lie between 0 and 1, but it cannot attain the upper limit of 1. Now it is easily seen that the NOTFI model holds if and only if the measure $\varphi^{(\lambda)}$ is equal to zero. According to the diversity index or the power-divergence, $\varphi^{(\lambda)}$ represents the degree of departure from NOTFI model, and the degree increases as the value of $\varphi^{(\lambda)}$ increases.

3. Approximate confidence interval for measure

Let n_{ijk} denote the observed frequency in the cell (i,j,k) of the $2\times 2\times K$ table ($i=1,2; j=1,2; k=1,\ldots,K$). Assuming that $\{n_{ijk}\}$ result from full multinomial sampling, we shall consider an approximate standard error and large-sample confidence interval of measure $\varphi^{(\lambda)}$, using the delta method of which descriptions are given by, for example, Bishop et al. [1, Sec. 14.6]. The sample version of measure $\varphi^{(\lambda)}$, i.e., $\hat{\varphi}^{(\lambda)}$, is given by $\varphi^{(\lambda)}$ with $\{p_{ijk}\}$ replaced by $\{\hat{p}_{ijk}\}$, where $\hat{p}_{ijk}=n_{ijk}/n$ and $n=\sum\sum\sum n_{ijk}$. Using the delta method, $\sqrt{n}(\hat{\varphi}^{(\lambda)}-\varphi^{(\lambda)})$ has asymptotically (as $n\to\infty$) a normal distribution with mean zero and variance

$$\begin{aligned}\sigma^2[\varphi^{(\lambda)}] &= \left(\frac{\lambda+1}{\lambda C^{(\lambda)}D^{\lambda+2}}\right)^2 \\ &\quad\times\sum_{t=1}^{K}\theta_t^2\left(D\theta_t^{\lambda}-\sum_{k=1}^{K}\theta_k^{\lambda+1}\right)^2\left(\frac{1}{p_{11t}}+\frac{1}{p_{12t}}+\frac{1}{p_{21t}}+\frac{1}{p_{22t}}\right).\end{aligned}$$

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

4. Examples

Table 1 taken from Agresti [7, p. 68] refers to the effect of passive smoking on lung cancer. It summarizes results of case-control studies from three countries among nonsmoking women married to smokers. For these data, the estimated odds-ratios between having passive smoking and lung cancer in Japan, Great Britain, and United States are 0.66, 0.63, and 0.76, respectively.

Let X, Y and Z denote the first, second and third variables, respectively. For Table 2 which is the $2\times 2\times 3$ artificial data, the estimated odds-ratios between variables X and Y at each level of Z are 7.50, 0.33, and 1.33.

Table 1. The results of case-control studies from three countries among non-smoking women married to smokers; from Agresti [7, p. 68].

Country	Spouse Smoked	Cases	Controls
Japan	No	21	82
	Yes	73	188
Great Britain	No	5	16
	Yes	19	38
United States	No	71	249
	Yes	137	363

Table 2. Artificial data (n is sample size).

$n = 300$

		Y	
Z	X	(1)	(2)
(1)	(1)	50	20
	(2)	10	30
(2)	(1)	10	30
	(2)	20	20
(3)	(1)	20	20
	(2)	30	40

Because the confidence intervals for $\varphi^{(\lambda)}$ applied to the data in Table 1 include zero for all λ (see Table 3a), this would indicate that there is a structure of NOTFI model in Table 1; or, if this is not the case, then it indicates that the degree of departure from NOTFI model is slight. In contrast, since the confidence intervals for $\varphi^{(\lambda)}$ applied to the data in Table 2 do not include zero for all λ (see Table 3b), this would indicate that there is not a structure of NOTFI model in Table 2.

When the degrees of departure from NOTFI model in Tables 1 and 2 are compared using the confidence intervals for $\varphi^{(\lambda)}$, the degree of departure in Table 2 would be greater than that in Table 1. This is because, for any given λ (> -1), the values in the confidence interval for $\varphi^{(\lambda)}$ applied to the data in Table 2 are greater than the values in the corresponding confidence interval for $\varphi^{(\lambda)}$ applied to the data in Table 1. We note that in Table 3a the confidence interval for $\varphi^{(\lambda)}$ includes the negative values and this is natural because $\hat{\varphi}^{(\lambda)}$ has asymptotically a normal distribution.

$Note$: Let $W^{(\lambda)}$ denote the power-divergence statistic for testing goodness-of-fit of the NOTFI model with $K-1$ degrees of freedom, i.e.,

$$W^{(\lambda)} = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{2} \sum_{j=1}^{2} \sum_{k=1}^{K} n_{ijk} \left[\left(\frac{n_{ijk}}{\hat{m}_{ijk}} \right)^{\lambda} - 1 \right], \quad \text{for } -\infty < \lambda < \infty$$

where $\hat{m}_{ijk}$ is the maximum likelihood estimate of the expected frequency m_{ijk} under the NOTFI model and the values at $\lambda = -1$ and $\lambda = 0$ are taken to be the limits as $\lambda \to -1$ and as $\lambda \to 0$, respectively. For the details of power-divergence test statistic, see Cressie and Read [5], and Read and Cressie [6, p. 15]. In particular, note that $W^{(0)}$ and $W^{(1)}$ are the likelihood ratio and Pearson chi-squared statistics, respectively. Table 4 gives the values of $W^{(\lambda)}$ applied to the data in Tables 1 and 2. Therefore, the NOTFI model fits the data in Table 1 well, but it does not fit the data in Table 2 well.

Table 3. Estimates of $\varphi^{(\lambda)}$, estimated approximate standard error for $\hat{\varphi}^{(\lambda)}$, approximate 95% confidence interval for $\varphi^{(\lambda)}$, applied to Tables 1 and 2.

(a) For Table 1

Values of λ	Estimated measure	Standard error	Confidence interval
-0.4	0.002	0.012	(-0.021, 0.025)
0	0.003	0.016	(-0.028, 0.034)
0.6	0.003	0.018	(-0.031, 0.038)
1.0	0.003	0.017	(-0.031, 0.037)
1.6	0.003	0.015	(-0.027, 0.032)

(b) For Table 2

Values of λ	Estimated measure	Standard error	Confidence interval
-0.4	0.388	0.124	(0.145, 0.630)
0	0.486	0.149	(0.194, 0.777)
0.6	0.536	0.166	(0.211, 0.861)
1.0	0.538	0.172	(0.200, 0.876)
1.6	0.517	0.180	(0.165, 0.869)

Table 4. Values of power-divergence statistic $W^{(\lambda)}$ (with 2 degrees of freedom) for testing goodness-of-fit of the NOTFI model, applied to Tables 1 and 2.

Values of λ	For Table 1	For Table 2
-0.4	0.240	24.889
0	0.240	24.462
0.6	0.239	24.056
1.0	0.238	23.933
1.6	0.237	23.957

5. Remark

Consider the case of $K = 2$, i.e., $2 \times 2 \times 2$ contingency table. Then the measure $\varphi^{(\lambda)}$ can be simply expressed as

$$\varphi^{(\lambda)} = \begin{cases} 1 - \dfrac{1}{\lambda C^{(\lambda)}}\left(1 - \frac{r^{\lambda+1}+1}{(1+r)^{\lambda+1}}\right), & \text{for } \lambda > -1;\ \lambda \neq 0, \\ 1 - \dfrac{1}{(\log 2)(1+r)}\left((1+r)\log(1+r) - r\log r\right), & \text{for } \lambda = 0, \end{cases}$$

where

$$r = \frac{\theta_1}{\theta_2} = \frac{p_{111}p_{221}p_{122}p_{212}}{p_{121}p_{211}p_{112}p_{222}}.$$

In addition, the approximate variance of $\sqrt{n}(\hat{\varphi}^{(\lambda)} - \varphi^{(\lambda)})$, which was given in Section 3, can be simply expressed as

$$\sigma^2[\varphi^{(\lambda)}] = \left(\frac{\lambda+1}{\lambda C^{(\lambda)}}\right)^2 \left(\frac{r^{\lambda+1} - r}{(1+r)^{\lambda+2}}\right)^2 \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} \frac{1}{p_{ijk}}.$$

Note that $\sigma^2[\varphi^{(\lambda)}] = 0$ when $r = 1$. Now, three kinds of expressions of r are obtained as

$$\begin{aligned} r &= \left(\frac{p_{111}p_{221}}{p_{121}p_{211}}\right) / \left(\frac{p_{112}p_{222}}{p_{122}p_{212}}\right) \\ &= \left(\frac{p_{111}p_{212}}{p_{112}p_{211}}\right) / \left(\frac{p_{121}p_{222}}{p_{122}p_{221}}\right) \\ &= \left(\frac{p_{111}p_{122}}{p_{112}p_{121}}\right) / \left(\frac{p_{211}p_{222}}{p_{212}p_{221}}\right). \end{aligned}$$

Therefore, the measure $\varphi^{(\lambda)}$, which represents the degree of departure from the equality of odds-ratio between variables X and Y at each level of variable Z, also represents the degree of departure from the equality of odds-ratio between X and Z at each level of Y and further represents it between Y and Z at each of X.

6. Concluding Remarks

The measure $\hat{\varphi}^{(\lambda)}$ would be useful for comparing the degrees of departure from the NOTFI model in several tables.

Table 5. (a), (b) Artificial data (n is sample size).

(a) $n = 315$

		Y	
Z	X	(1)	(2)
(1)	(1)	25	20
	(2)	25	40
(2)	(1)	45	15
	(2)	30	30
(3)	(1)	30	20
	(2)	20	15

(b) $n = 1575$

		Y	
Z	X	(1)	(2)
(1)	(1)	125	100
	(2)	125	200
(2)	(1)	225	75
	(2)	150	150
(3)	(1)	150	100
	(2)	100	75

Consider the artificial data in Tables 5a and 5b. For Table 5a, the estimated odds-ratios between variables X and Y at each level of Z are 2.00, 3.00, and 1.13. All values of observed frequencies in

Table 6. Values of $\hat{\varphi}^{(\lambda)}$ applied to Tables 5a and 5b.

Values of λ	For Table 5a	For Table 5b
-0.4	0.050	0.050
0	0.066	0.066
0.6	0.073	0.073
1.0	0.070	0.070
1.6	0.061	0.061

Table 7. Values of power-divergence statistic $W^{(\lambda)}$ (with 2 degrees of freedom) for testing goodness-of-fit of the NOTFI model, applied to Tables 5a and 5b.

Values of λ	For Table 5a	For Table 5b
-0.4	2.734	13.669
0	2.730	13.648
0.6	2.726	13.630
1.0	2.726	13.628
1.6	2.727	13.637

Table 5a multiplied by 5 equal the values in Table 5b. Thus, it is natural that the estimated odds-ratios between variables X and Y at each level of Z for Table 5b are equal to those for Table 5a. Therefore, the value of $\hat{\varphi}^{(\lambda)}$ (for every λ) for Table 5a is identical with that for Table 5b (see Table 6). However the value of $W^{(\lambda)}$ is greater for Table 5b than for Table 5a (see Table 7). Therefore the measure $\hat{\varphi}^{(\lambda)}$ rather than test statistic $W^{(\lambda)}$ would be useful for comparing the degrees of departure from the NOTFI model in several tables.

The $W^{(\lambda)}$ is also an information measure on the cell probability scale, and moreover $W^{(\lambda)}/n$ seems to be a reasonable measure of departure from the NOTFI model (though it is not a function of odds-ratios $\{\theta_i\}$, $i = 1, \ldots, K$). However, $\hat{\varphi}^{(\lambda)}$ rather than $W^{(\lambda)}/n$ would be useful for comparing the degrees of departure from the NOTFI model in several tables. This is because $\hat{\varphi}^{(\lambda)}$ is always in the range between 0 and 1, but $W^{(\lambda)}/n$ is not; namely, $\hat{\varphi}^{(\lambda)}$ can measure the degree of departure toward the maximum departure from uniformity of odds-ratios $\{\theta_i\}$, $i = 1, \ldots, K$; but the $W^{(\lambda)}/n$ cannot measure it.

The readers may be interested in which value of λ is preferred for a given table. However, in comparing tables, it seems difficult to discuss this. For example, consider the artificial data in Tables 8a and 8b. We see from Table 8c that the value of $\hat{\varphi}^{(0)}$ is greater for Table 8a than for Table 8b, but the value of $\hat{\varphi}^{(1)}$ is less for Table 8a than for Table 8b. So, for these cases, it may be impossible to decide (by using $\hat{\varphi}^{(\lambda)}$) whether the degree of departure from the NOTFI model is greater for Table 8a or for Table 8b. But generally, for the comparison between two tables, it would be possible to draw a conclusion if $\hat{\varphi}^{(\lambda)}$ (for every λ) is always greater (or always less) for one table than for the other table. Thus, it seems

Table 8. (a), (b) Artificial data (n is sample size) and (c) corresponding values of $\hat{\varphi}^{(\lambda)}$ applied to Tables 8a and 8b.

(a) $n = 291$

		Y	
Z	X	(1)	(2)
(1)	(1)	27	9
	(2)	10	16
(2)	(1)	14	35
	(2)	31	45
(3)	(1)	28	18
	(2)	13	45

(b) $n = 291$

		Y	
Z	X	(1)	(2)
(1)	(1)	22	23
	(2)	30	16
(2)	(1)	20	18
	(2)	22	43
(3)	(1)	11	21
	(2)	26	39

(c) Values of $\hat{\varphi}^{(\lambda)}$

Values of λ	For Table 8a	For Table 8b
-0.4	0.186	0.126
0	0.213	0.170
0.6	0.200	0.197
1.0	0.178*	0.198
1.6	0.140*	0.183

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

to be important that which value of λ is preferred for a given table, the analyst calculates the value of $\hat{\varphi}^{(\lambda)}$ for various values of λ and discusses the degree of departure from the NOTFI model in terms of $\hat{\varphi}^{(\lambda)}$ values. It may seem to readers that when the odds-ratios of Table 8a vary more widely (relatively in ratio) than those of Table 8b, the $\varphi^{(\lambda)}$ values in Table 8c may vary with a pattern; namely, they are large for Table 8a for smaller values of λ, but the other way round when λ is greater than certain value less than 1. However, we cannot prove that the case holds. It may be dangerous to compare the degrees of departure from the NOTFI model in several tables in terms of only Tomizawa's [3] measure, i.e., $\hat{\varphi}^{(0)}$; because it may arise that for two tables (say, table A and table B), $\hat{\varphi}^{(0)}$ is greater for table A than for table B, however, $\hat{\varphi}^{(\lambda_1)}$ with some $\lambda_1 (\neq 0)$ is less for table A than for table B.

The measure $\hat{\varphi}^{(\lambda)}$ would be useful when one wants to measure how far the odds-ratios $\{\theta_t\}$ are directly distant from the uniformity, although $W^{(\lambda)}/n$ may be useful when one wants to measure how far the estimated cell probability distribution with the structure of NOTFI is distant from the sample cell probability distribution.

The readers may be interested in extending the measure $\varphi^{(\lambda)}$ to a $2 \times 3 \times K$ table or $I \times J \times K$ table; however, it may be difficult to consider a single-valued measure to represent the degree of departure from no three-factor interaction.

Appendix

For the $2 \times 2 \times K$ contingency table, a measure of departure from the NOTFI model by Tomizawa [3] is given as follows:

$$\varphi = 1 - \frac{H(\theta^*)}{\log K},$$

where

$$H(\theta^*) \;=\; -\sum_{t=1}^{K} \theta_t^* \log \theta_t^*$$

and $\{\theta_t^*\}$ are defined in Section 2.

Acknowledgements

The authors would like to thank two referees for the helpful comments.

References and Notes

1. Bishop, Y. M. M.; Fienberg, S. E.; Holland, P. W. *Discrete Multivariate Analysis: Theory and Practice*; The MIT Press: Cambridge, Massachusetts, 1975. 1., 3.
2. Fienberg, S. E. *The Analysis of Cross-Classified Categorical Data*, 2nd Ed.; The MIT Press, Cambridge, Massachusetts, 1980. 1.
3. Tomizawa, S. A measure of departure from no three-factor interaction model in a $2 \times 2 \times k$ contingency table. *J. Stat. Res.* **1993**, *27*, 1–8. 1., 6., 6.
4. Patil, G. P.; Taillie, C. Diversity as a concept and its measurement. *J. Am. Stat. Assoc.* **1982**, *77*, 548–561. 1.
5. Cressie, N. A. C.; Read, T. R. C. Multinomial goodness-of-fit tests. *J. Royal Stat. Soc., Series B* **1984**, *46*, 440–464. 2., 4.
6. Read, T. R. C.; Cressie, N. A. C. *Goodness-of-Fit Statistics for Discrete Multivariate Data*: Springer: New York, NY, 1988. 2., 4.
7. Agresti, A. *An Introduction to Categorical Data Analysis*; Wiley: New York, NY, 1996. 4., 1