International Journal of Pure and Applied Mathematics
Volume 50 No. 1 2009, 137-143

A GENERALIZATION OF MEASURE OF DEPARTURE FROM UNIFORM ASSOCIATION IN TWO-WAY CONTINGENCY TABLES

Nobuko Miyamoto[1 §], Takeshi Fukushi[2], Sadao Tomizawa[3]
[1,2,3]Department of Information Sciences
Faculty of Science and Technology
Tokyo University of Science
2641, Yamazaki, Noda City, Chiba, 278-8510, JAPAN
[1]e-mail: miyamoto@is.noda.tus.ac.jp
[3]e-mail: tomizawa@is.noda.tus.ac.jp

Abstract: For a two-way contingency table with ordered categories, Tomizawa [8] considered a measure to represent the degree of departure from uniform association. The present paper proposes a generalization of this measure. The proposed measure is expressed by using Patil and Taillie's [5] diversity index. A special case of the proposed measure includes Tomizawa's measure. The measure would be useful for comparing the degree of departure from uniform association in several tables.

AMS Subject Classification: 62H17
Key Words: diversity index, local odds ratio, uniform association

1. Introduction

For an $r \times c$ contingency table with ordered categories, let p_{ij} denote the probability that an observation will fall in the i-th row and j-th column of the table ($i = 1, 2, \cdots, r; j = 1, 2, \cdots, c$).

The uniform association model is defined by

$$p_{ij} = \alpha_i \beta_j \theta^{ij} \quad \text{for } i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, c;$$

see Goodman [3, 4]. For the 2×2 subtables formed from adjacent rows i and

Received: December 20, 2008

§Correspondence author

$i+1$, and adjacent columns j and $j+1$ in the $r \times c$ table, let θ_{ij} denote the corresponding local odds ratio ($i = 1, 2, \cdots, r-1$; $j = 1, 2, \cdots, c-1$) based on the probabilities. Thus $\theta_{ij} = (p_{ij}p_{i+1,j+1})/(p_{i+1,j}p_{i,j+1})$. The uniform association model may be expressed as

$$\theta_{ij} = \theta \quad \text{for } i = 1, 2, \cdots, r-1;\ j = 1, 2, \cdots, c-1.$$

A special case of this model obtained by putting $\theta = 1$ is the independence model (the null association model).

Tomizawa [8] proposed the Shannon entropy type measure ϕ_u which represents the degree of departure from uniform association. Let

$$\nu = \sum_{i=1}^{r-1}\sum_{j=1}^{c-1}\theta_{ij},\ \ \theta_{ij}^{*} = \frac{\theta_{ij}}{\nu}.$$

The measure ϕ_u is expressed as

$$\phi_u = 1 - \frac{H(\theta^*)}{\log(r-1)(c-1)},$$

where

$$H(\theta^*) = -\sum_{i=1}^{r-1}\sum_{j=1}^{c-1}\theta_{ij}^{*}\log\theta_{ij}^{*}.$$

Note that $H(\theta^*)$ is the Shannon entropy for $\{\theta_{ij}^{*}\}$.

Further, ϕ_u may be expressed as

$$\phi_u = \frac{I(\theta^*;\theta^{\mathrm{u}})}{\log(r-1)(c-1)},$$

where

$$I(\theta^*;\theta^{\mathrm{u}}\ ) = \sum_{i=1}^{r-1}\sum_{j=1}^{c-1}\theta_{ij}^{*}\log\left(\frac{\theta_{ij}^{*}}{\theta^{\mathrm{u}}}\right),$$

$$\theta^{\mathrm{u}} = \frac{1}{(r-1)(c-1)}.$$

Note that $I(\theta^*;\theta^{\mathrm{u}})$ is the Kullback-Leibler information between $\{\theta_{ij}^{*}\}$ and $\{\theta^{\mathrm{u}}\}$.

The purpose of this paper is to propose a generalization of Tomizawa's measure ϕ_u using the Patil and Taillie's [5] diversity index. The proposed measure represents the degree of departure from uniform association, and includes the measure ϕ_u in a special case. It would be useful for comparing the degree of departure from uniform association in several tables.

2. Generalized Measure of Departure from Uniform Association

Assume that $\{p_{ij}\}$ are positive. Consider a measure defined by

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

where

$$H^{(\lambda)}(\theta^*) = \frac{1}{\lambda}\left(1 - \sum_{i=1}^{r-1}\sum_{j=1}^{c-1}(\theta_{ij}^*)^{\lambda+1}\right),$$

$$C^{(\lambda)} = \frac{1}{\lambda}\left(1 - \left(\frac{1}{(r-1)(c-1)}\right)^{\lambda}\right),$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$. Note that $H^{(\lambda)}(\theta^*)$ is the Patil and Taillie's [5] diversity index of degree λ for $\{\theta_{ij}^*\}$ which includes the Shannon entropy (when $\lambda = 0$) and the Gini concentration (when $\lambda = 1$) in special cases. The measure $\phi_u^{(\lambda)}$ is a generalization of Tomizawa's [8] measure.

Further, $\phi_u^{(\lambda)}$ may be expressed as

$$\phi_u^{(\lambda)} = \lambda(\lambda+1)\frac{(\theta^{\mathrm{u}})^{\lambda}}{1-(\theta^{\mathrm{u}})^{\lambda}} I^{(\lambda)}(\{\theta_{ij}^*\};\{\theta^{\mathrm{u}}\}),$$

where

$$I^{(\lambda)}(\cdot;\cdot) = \frac{1}{\lambda(\lambda+1)}\sum_{i=1}^{r-1}\sum_{j=1}^{c-1}\theta_{ij}^*\left\{\left(\frac{\theta_{ij}^*}{\theta^{\mathrm{u}}}\right)^{\lambda} - 1\right\}.$$

and the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$. Note that $I^{(\lambda)}(\{\theta_{ij}^*\};\{\theta^{\mathrm{u}}\})$ is the power divergence between $\{\theta_{ij}^*\}$ and $\{\theta^{\mathrm{u}}\}$. The index $H^{(\lambda)}(\theta^*)$ must lie between 0 and $C^{(\lambda)}$ but it cannot attain the lower limit of 0 since $\{p_{ij}\}$ are positive. Thus, $\phi_u^{(\lambda)}$ must lie between 0 and 1 but it cannot attain the upper limit of 1. Also, for each λ ($\lambda > -1$), there is an uniform association in the $r \times c$ table if and only if $\phi_u^{(\lambda)}$ equals zero. Thus, according to this metric, the degree of departure from uniform association increases as the value of $\phi_u^{(\lambda)}$ increases.

Let n_{ij} denote the observed frequency in the i-th row and j-th column of the table ($i = 1, 2, \cdots, r; j = 1, 2, \cdots, c$). We assume that a multinomial distribution applies to the $r \times c$ table. We shall consider an approximate standard error and large-sample confidence interval for $\phi_u^{(\lambda)}$ using the delta method, of which descriptions are given by Bishop et al [2, Section 14.6] and Agresti [1, Appendix C, p. 185]. The sample version of $\phi_u^{(\lambda)}$, i.e. $\hat{\phi}_u^{(\lambda)}$, is given by $\phi_u^{(\lambda)}$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$, where $\hat{p}_{ij} = n_{ij}/n$ and $n = \sum\sum n_{ij}$.

Using the delta method, we obtain the following theorem.

Theorem 1. *$\sqrt{n}(\hat{\phi}_u^{(\lambda)} - \phi_u^{(\lambda)})$ has asymptotically (as $n \to \infty$) a normal distribution with mean zero and variance*

$$\sigma^2 = \left(\frac{\lambda+1}{\lambda\nu^{\lambda+1}C^{(\lambda)}}\right)^2 \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(L_{ij}^{(\lambda)}\right)^2}{p_{ij}},$$

where

$$L_{ij}^{(\lambda)} = \frac{\delta^{(\lambda)}}{\nu}(\theta_{i-1,j-1} - \theta_{i-1,j} - \theta_{i,j-1} + \theta_{ij}) + \theta_{i-1,j-1}^{\lambda+1} - \theta_{i-1,j}^{\lambda+1} - \theta_{i,j-1}^{\lambda+1} + \theta_{ij}^{\lambda+1},$$

$$\delta^{(\lambda)} = -\sum_{i=1}^{r-1}\sum_{j=1}^{c-1} \theta_{ij}^{\lambda+1},$$

$$\theta_{s0} = \theta_{0t} = \theta_{rt} = \theta_{sc} = 0 \text{ for } s = 1, 2, \cdots, r;\ t = 1, 2, \cdots, c.$$

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

3. Examples

Consider the data in Tables 1 and 2. Table 1 taken from Stuart [6] is constructed from the data of the unaided distance vision of 7477 women aged 30-39 employed in Royal Ordinance factories in Britain from 1943 to 1946. Table 2 taken from Tomizawa [7] is constructed from the data of the unaided distance vision of 4746 students aged 18 to about 25, including about 10 percent of the women of Faculty of Science and Technology, Science University of Tokyo in Japan examined in April, 1982. In Tables 1 and 2, the row variable is the right eye grade and the column variable is the left eye grade with the categories ordered from the best grade (1) to the worst grade (4).

Because the confidence intervals for $\phi_u^{(\lambda)}$ applied to the data in Tables 1 and 2 do not include zero for all λ (see Table 3), this would indicate that there is not a structure of uniform association in each table.

When the degrees of departure from uniform association in Tables 1 and

Right eye grade	Left eye grade Best (1)	Second (2)	Third (3)	Worst (4)	Total
Best (1)	1520	266	124	66	1976
Second (2)	234	1512	432	78	2256
Third (3)	117	362	1772	205	2456
Worst (4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

Table 1: Unaided distance vision of 7477 women aged 30-39 employed in Royal Ordnance factories from 1943 to 1946; from Stuart [6]

Right eye grade	Left eye grade Best (1)	Second (2)	Third (3)	Worst (4)	Total
Best (1)	1291	130	40	22	1483
Second (2)	149	221	114	23	507
Third (3)	64	124	660	185	1033
Worst (4)	20	25	249	1429	1723
Total	1524	500	1063	1659	4746

Table 2: Unaided distance vision of 4746 students aged 18 to about 25 including about 10 % women in Faculty of Science and Technology, Science University of Tokyo in Japan examined in April 1982; from Tomizawa [7]

2 are compared using the estimates of $\phi_u^{(\lambda)}$, the degree of departure in Table 1 would be stronger than that in Table 2. This is because, for any given λ (> -1), the estimates of $\phi_u^{(\lambda)}$ applied to the data in Table 1 are greater than the corresponding estimates of $\phi_u^{(\lambda)}$ applied to the data in Table 2.

4. Concluding Remarks

The measure $\phi_u^{(\lambda)}$ always ranges between 0 and 1 independent of dimension r and c, and sample size n. Therefore, $\phi_u^{(\lambda)}$ would be useful for comparing the degrees of departure from uniform association in several tables.

(a) For Table 1

Values of λ	Estimated measure	Standard error	Confidence interval
-0.8	0.258	0.008	(0.241, 0.274)
-0.4	0.438	0.011	(0.417, 0.459)
0	0.433	0.011	(0.412, 0.454)
0.4	0.369	0.012	(0.346, 0.391)
1.0	0.256	0.013	(0.230, 0.281)
1.6	0.163	0.013	(0.138, 0.188)

(b) For Table 2

Values of λ	Estimated measure	Standard error	Confidence interval
-0.8	0.149	0.011	(0.127, 0.171)
-0.4	0.282	0.019	(0.245, 0.318)
0	0.296	0.019	(0.260, 0.333)
0.4	0.260	0.017	(0.227, 0.292)
1.0	0.179	0.013	(0.153, 0.206)
1.6	0.110	0.011	(0.089, 0.132)

Table 3: The estimates of $\phi_u^{(\lambda)}$, estimated approximate standard errors for $\hat{\phi}_u^{(\lambda)}$, and approximate 95% confidence intervals for $\phi_u^{(\lambda)}$, applied to Tables 1, and 2

The reader may be interested in which value of λ is preferred for a given table. However, it seems difficult to discuss this. It seems to be important and safe that, for given tables, the analyst calculates the values of $\hat{\phi}_u^{(\lambda)}$ for various values of λ and discusses the degree of departure from uniform association in terms of them. Because for comparing the degrees of departure from uniform association in several tables, say, table A and table B, it may arise that $\hat{\phi}_u^{(\lambda_1)}$ for a fixed λ_1 is greater for table A than for table B, but $\hat{\phi}_u^{(\lambda_2)}$ for a fixed λ_2 ($\neq \lambda_1$) is less for table A than for table B, though we have not yet obtained such a case.

Acknowledgements

We would like to thank the referees.

References

[1] A. Agresti, *Analysis of Ordinal Categorical Data*, Wiley, New York (1984).

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

[3] L.A. Goodman, Simple models for the analysis of association in cross-classification having ordered categories, *Journal of the American Statistical Association*, **74** (1979), 537-552.

[4] L.A. Goodman, Association models and canonical correlation in the analysis of cross-classification having ordered categories, *Journal of the American Statistical Association*, **76** (1981), 320-334.

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

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

[7] S. Tomizawa, Analysis of data in square contingency tables with ordered categories using the conditional asymmetry model and its decomposed models, *Environmental Health Perspectives*, **63** (1985), 235-239.

[8] S. Tomizawa, Shannon entropy type measure of departure from uniform association in cross-classifications having ordered categories, *Statistics and Probability Letters*, **11** (1991), 547-550.