American Journal of Mathematics and Statistics 2013, 3(3): 95-98
DOI: 10.5923/j.ajms.20130303.01

Marginal Asymmetry Measure Based on Entropy for Square Contingency Tables with Ordered Categories

Kouji Tahata*, Takuya Yoshimoto, Sadao Tomizawa

Department of Information Sciences, Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

Abstract For the analysis of square contingency tables, Tomizawa, Miyamoto and Ashihara (2003) considered a measure to represent the degree of departure from marginal homogeneity. The measure lies between 0 and 1, and it takes the minimum value when the marginal homogeneity holds and the maximum value when one of two symmetric cumulative probabilities for any category is zero. This paper proposes improvement of the measure so that the degree of departure from marginal homogeneity can attain the maximum value even when the cumulative probabilities are not zero. The proposed measure would be useful for representing the degree of departure from marginal homogeneity, especially when some asymmetry models hold as the extended marginal homogeneity model or the conditional symmetry model. Examples are given.

Keywords Kullback-Leibler information, Measure, Power-divergence, Shannon entropy

1. Introduction

Consider an $R \times R$ square contingency table with the same row and column classifications. Let p_{ij} denote the probability that an observation will fall in the i th row and j th column of the table ($i = 1,\dots,R; j = 1,\dots,R$), and let X and Y denote the row and column variables, respectively. The marginal homogeneity (MH) model is defined by

$$\Pr(X = i) = \Pr(Y = i) \quad \text{for } i = 1,\dots,R,$$

namely

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

where $p_{i\cdot} = \sum_{k=1}^{R} p_{ik}$ and $p_{\cdot i} = \sum_{k=1}^{R} p_{ki}$ (see, for example, Stuart, 1955; Bishop, Fienberg and Holland, 1975, p.293). Let

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

and

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

for $i = 1,\dots,R-1$. By considering the difference between the $\Pr(X \le i)$ and $\Pr(Y \le i)$, the MH model also be expressed as

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

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 $i = 1,\dots,R-1$. When the MH model does not hold, we are interested in measuring the degree of departure from MH.

For square contingency tables with ordered categories, Tomizawa, Miyamoto and Ashihara (2003) proposed the measure (denoted by $\gamma^{(\lambda)}$ in Section 2) to represent the degree of departure from MH. The measure $\gamma^{(\lambda)}$ ranges between 0 and 1. Also, (i) $\gamma^{(\lambda)} = 0$ if and only if the MH model holds, and (ii) $\gamma^{(\lambda)} = 1$ if and only if the degree of departure from MH is a maximum; that is, $G_{1(i)} = 0$ (then $G_{2(i)} > 0$) or $G_{2(i)} = 0$ (then $G_{1(i)} > 0$) for all $i = 1,\dots,R-1$.

However, for the analysis of square contingency tables, all cell probabilities $\{p_{ij}\}$ are positive in many cases. Thus, the measure $\gamma^{(\lambda)}$ may be unsuitable for such data, because the measure $\gamma^{(\lambda)}$ cannot attain the maximum value. So, we are now interested in the measure to represent the degree of departure from MH such that it can attain the maximum value even when each of cell probabilities $\{p_{ij}\}$ is not zero. Yamamoto, Masumura and Tomizawa (2011) considered such a measure for *nominal* square table. We are now interested in proposing such a measure for *ordinal* square table.

The purpose of this paper is to consider an improvement of measure for square contingency tables with *ordered*

---

* Corresponding author:
kouji_tahata@is.noda.tus.ac.jp (Kouji Tahata)
Published online at http://journal.sapub.org/ajms

categories when all cell probabilities $\{p_{ij}\}$ are positive.

2. Improved Measure for Marginal Homogeneity

Consider an $R \times R$ table with *ordered* categories. Assume that $\{G_{1(i)} + G_{2(i)}\}$ are positive. Let

$$G_{1(i)}^{c} = \frac{G_{1(i)}}{G_{1(i)} + G_{2(i)}}, \quad G_{2(i)}^{c} = \frac{G_{2(i)}}{G_{1(i)} + G_{2(i)}},$$

for $i = 1, \ldots, R-1$; and let

$$\Delta = \sum_{i=1}^{R-1} (G_{1(i)} + G_{2(i)}).$$

For a specified d which satisfies $0.5 < d \le 1$ and $1 - d \le G_{1(i)}^{c} \le d$ for all $i = 1, \ldots, R-1$, consider a measure defined by

$$\Gamma^{(\lambda)}(d) = \frac{1}{C^{(\lambda)}(d)} \gamma^{(\lambda)} \quad (\lambda > -1),$$

where

$$C^{(\lambda)}(d) = 1 - \frac{\lambda 2^{\lambda}}{2^{\lambda} - 1} H^{(\lambda)}(d),$$

$$H^{(\lambda)}(d) = \frac{1}{\lambda}(1 - d^{\lambda+1} - (1-d)^{\lambda+1}),$$

$$\gamma^{(\lambda)} = 1 - \frac{\lambda 2^{\lambda}}{2^{\lambda} - 1} \sum_{i=1}^{R-1} \left(\frac{G_{1(i)} + G_{2(i)}}{\Delta} \right) H_i^{(\lambda)},$$

with

$$H_i^{(\lambda)} = \frac{1}{\lambda}[1 - (G_{1(i)}^{c})^{\lambda+1} - (G_{2(i)}^{c})^{\lambda+1}],$$

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

$$\Gamma^{(0)}(d) = \frac{1}{C^{(0)}(d)} \gamma^{(0)},$$

where

$$C^{(0)}(d) = 1 - \frac{1}{\log 2} H^{(0)}(d),$$

$$H^{(0)}(d) = -d \log d - (1-d)\log(1-d),$$

$$\gamma^{(0)} = 1 - \frac{1}{\log 2} \sum_{i=1}^{R-1} \left(\frac{G_{1(i)} + G_{2(i)}}{\Delta} \right) H_i^{(0)},$$

with

$$H_i^{(0)} = -G_{1(i)}^{c} \log G_{1(i)}^{c} - G_{2(i)}^{c} \log G_{2(i)}^{c}.$$

Note that $H_i^{(\lambda)}$ is the diversity index proposed by Patil and Taillie (1982), which includes the Shannon entropy when $\lambda = 0$. When $d = 1$, then $C^{(\lambda)}(1) = 1$. So, $\Gamma^{(\lambda)}(1)$ is identical to the measure $\gamma^{(\lambda)}$ given by Tomizawa et al. (2003).

Since $1 - d \le G_{1(i)}^{c} \le d$, the minimum value of $H_i^{(\lambda)}$ is $(1 - d^{\lambda+1} - (1-d)^{\lambda+1})/\lambda$ (if $\lambda \ne 0$) or $-d \log d - (1-d) \log(1-d)$ (if $\lambda = 0$), and the maximum value of it is $(2^{\lambda} - 1)/(\lambda 2^{\lambda})$ (if $\lambda \ne 0$) or $\log 2$ (if $\lambda = 0$) when $G_{1(i)}^{c} = G_{2(i)}^{c} = 1/2$ for all $i = 1, \ldots, R-1$. So, when $d \ne 1$, $\gamma^{(\lambda)}$ cannot attain the value 1. The proposed measure $\Gamma^{(\lambda)}(d)$ with $d \ne 1$ is modified by using modification coefficient $1/C^{(\lambda)}(d)$ such that the measure $\Gamma^{(\lambda)}(d)$ can attain the value 1. If all $\{p_{ij}\}$ are positive, then d must be taken as $d < 1$.

Moreover, for each $\lambda\ (> -1)$ and a fixed d, the measure $\Gamma^{(\lambda)}(d)$ has characteristics that (i) $\Gamma^{(\lambda)}(d)$ must lie between 0 and 1, (ii) $\Gamma^{(\lambda)}(d) = 0$ if and only if the MH model holds, i.e., $G_{1(i)}^{c} = G_{2(i)}^{c} = 1/2$ for all $i = 1, \ldots, R-1$, and (iii) $\Gamma^{(\lambda)}(d) = 1$ if and only if the degree of departure from MH is the largest in the sense that $G_{1(i)}^{c} = 1 - d$ or $G_{2(i)}^{c} = 1 - d$ for all $i = 1, \ldots, R-1$.

The measure also may be expressed as, for $\lambda > -1$

$$\Gamma^{(\lambda)}(d) = \frac{1}{C^{(\lambda)}(d)} \left[\frac{\lambda(\lambda+1)}{2^{\lambda} - 1} \sum_{i=1}^{R-1} \left(\frac{G_{1(i)} + G_{2(i)}}{\Delta} \right) I_i^{(\lambda)} \right],$$

where

$$I_i^{(\lambda)} = \frac{1}{\lambda(\lambda+1)} \left[G_{1(i)}^{c} \left\{ \left(\frac{G_{1(i)}^{c}}{1/2} \right)^{\lambda} - 1 \right\} + G_{2(i)}^{c} \left\{ \left(\frac{G_{2(i)}^{c}}{1/2} \right)^{\lambda} - 1 \right\} \right],$$

especially

$$I_i^{(0)} = G_{1(i)}^{c} \log \left(\frac{G_{1(i)}^{c}}{1/2} \right) + G_{2(i)}^{c} \log \left(\frac{G_{2(i)}^{c}}{1/2} \right).$$

Note that $I_i^{(\lambda)}$ is the power-divergence between $\{G_{1(i)}^{c}, G_{2(i)}^{c}\}$ and $\{1/2, 1/2\}$ (Cressie and Read, 1984) which includes the Kullback-Leibler information when $\lambda = 0$.

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 ($i = 1, \ldots, R; j = 1, \ldots, R$). Assume that a multinomial distribution applies to the $R \times R$ table. The sample version of $\Gamma^{(\lambda)}(d)$, i.e., $\hat{\Gamma}^{(\lambda)}(d)$, is given by $\Gamma^{(\lambda)}(d)$ 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 (Bishop et al., 1975, Sec. 14.6), $\sqrt{n}\left(\hat{\Gamma}^{(\lambda)}(d) - \Gamma^{(\lambda)}(d)\right)$ has asymptotically (as $n \to \infty$) a normal distribution with mean zero and variance

$$\sigma^2[\Gamma^{(\lambda)}(d)] = \frac{V^{(\lambda)}}{(C^{(\lambda)}(d))^2},$$

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

$$V^{(\lambda)} = \left[\frac{1}{\Delta^2(2^\lambda - 1)^2}\right] \sum_{s=1}^{R-1} \sum_{t=s+1}^{R} \left[p_{st}(A_{st}^{(\lambda)})^2 + p_{ts}(B_{ts}^{(\lambda)})^2 \right],$$

with

$$A_{st}^{(\lambda)} = \sum_{i=s}^{t-1} \Big[2^\lambda \left\{ (G_{1(i)}^c)^\lambda + \lambda \left((G_{1(i)}^c)^\lambda - (G_{2(i)}^c)^\lambda \right) G_{2(i)}^c \right\} - \left\{ (2^\lambda - 1)\gamma^{(\lambda)} + 1 \right\} \Big],$$

$$B_{ts}^{(\lambda)} = \sum_{i=s}^{t-1} \Big[2^\lambda \left\{ (G_{2(i)}^c)^\lambda + \lambda \left((G_{2(i)}^c)^\lambda - (G_{1(i)}^c)^\lambda \right) G_{1(i)}^c \right\} - \left\{ (2^\lambda - 1)\gamma^{(\lambda)} + 1 \right\} \Big],$$

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

Let $\hat{\sigma}^2[\Gamma^{(\lambda)}(d)]$ denote $\sigma^2[\Gamma^{(\lambda)}(d)]$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$. Using this result, the estimated approximate confidence interval for the measure $\Gamma^{(\lambda)}(d)$ is obtained.

4. Examples

Consider the data in Table 1, taken from Andersen (1997, p.226). These data show the forecasts for production and prices for the coming three year periods given by experts in July 1956 and the actual production figures for production and prices in May 1959 given from Danish factories.

For these data, the cell probabilities $\{p_{ij}\}$ are theoretically positive (not zero). Thus, it may be irrelevance to use the measure $\Gamma^{(\lambda)}(d)$ with $d = 1$. So we should use the measure $\Gamma^{(\lambda)}(d)$ with $d < 1$ (for example, $d = 0.999$) so that the measure can attain the maximum value 1.

Table 1. Results from the forecasts for production and prices and the actual production figures for production and prices (Andersen, 1997, p.226)

(a) Prices

	Actual 1959			
Forecast 1956	Higher (1)	No change (2)	Lower (3)	Total
Higher (1)	209	169	6	384
No change (2)	190	3073	184	3447
Lower (3)	3	62	81	146
Total	402	3304	271	3977

(b) Production

	Actual 1959			
Forecast 1956	Higher (1)	No change (2)	Lower (3)	Total
Higher (1)	532	394	69	995
No change (2)	447	1727	334	2508
Lower (3)	39	230	231	500
Total	1018	2351	634	4003

Table 2. When $d = 0.999$, the estimates of $\Gamma^{(\lambda)}(d)$, estimated approximate standard error for $\hat{\Gamma}^{(\lambda)}(d)$, and approximate 95% confidence interval for $\Gamma^{(\lambda)}(d)$, applied to Tables 1a and 1b

(a) For Table 1a

λ	$\hat{\Gamma}^{(\lambda)}(d)$	Standard error	Confidence interval
−0.6	0.044	0.011	(0.023, 0.065)
0.0	0.076	0.018	(0.041, 0.111)
0.6	0.094	0.022	(0.052, 0.137)
1.0	0.100	0.023	(0.056, 0.144)
1.6	0.102	0.023	(0.057, 0.147)

(b) For Table 1b

λ	$\hat{\Gamma}^{(\lambda)}(d)$	Standard error	Confidence interval
−0.6	0.007	0.003	(0.002, 0.012)
0.0	0.012	0.005	(0.003, 0.021)
0.6	0.016	0.006	(0.004, 0.027)
1.0	0.017	0.006	(0.005, 0.029)
1.6	0.017	0.006	(0.005, 0.030)

If we set $d = 0.999$ and $\lambda = 1$, the estimated measure $\hat{\Gamma}^{(\lambda)}(d)$ is 0.100 for Table 1a and 0.017 for Table 1b from Tables 2a and 2b. Thus, (i) for Table 1a, the degree of departure from MH is estimated to be 10.0 percent of the maximum degree of departure from MH and (ii) for Table 1b, it is estimated to be 1.7 percent of the maximum. Furthermore, we see from Tables 2a and 2b that the degree of departure from MH is greater for Table 1a than for Table 1b because the values in the confidence intervals for $\Gamma^{(\lambda)}(d)$ are greater for Table 1a than for Table 1b.

5. Discussion

Consider the extended MH (EMH) model defined by

$$G_{1(i)} = \tau G_{2(i)} \quad \text{for } i = 1, \ldots, R-1;$$

also see Tahata and Tomizawa (2008). A special case of EMH model obtained by putting $\tau = 1$ is the MH model. When the EMH model holds, the proposed measure $\Gamma^{(\lambda)}(d)$ is expressed as

$$\Gamma^{(\lambda)}(d) = \frac{1}{C^{(\lambda)}(d)} \left(1 - \frac{\lambda 2^\lambda}{2^\lambda - 1} H^{(\lambda)} \right), \tag{1}$$

where

$$H^{(\lambda)} = \frac{1}{\lambda} \left[1 - \left(\frac{\tau}{1+\tau} \right)^{\lambda+1} - \left(\frac{1}{1+\tau} \right)^{\lambda+1} \right].$$

For λ fixed and d fixed, $\Gamma^{(\lambda)}(d)$ increases as $\tau (> 1)$ increases (or as $\tau (< 1)$ decreases). Especially, when $d = 1$, $\Gamma^{(\lambda)}(1)$ is identical to $\gamma^{(\lambda)}$ proposed by Tomizawa et al. (2003). When the EMH model holds, $\gamma^{(\lambda)}$ approaches 1 as τ approaches infinity or zero. However, when the EMH model holds, $\gamma^{(\lambda)}$ cannot attain 1 because then $\{G_{1(i)}^c > 0\}$ and $\{G_{2(i)}^c > 0\}$, namely there is not the structure of

$\left\{G_{1(i)}^{c}=0 \text{ or } G_{2(i)}^{c}=0\right\}$ being the condition of $\gamma^{(\lambda)}=1$. The measure $\Gamma^{(\lambda)}(d)$ with $d<1$ can attain the maximum value 1 even if $\left\{G_{1(i)}^{c}>0\right\}$ and $\left\{G_{2(i)}^{c}>0\right\}$ for all $i=1,\dots,R-1$. Therefore, the measure $\Gamma^{(\lambda)}(d)$ with $d<1$ rather than $\gamma^{(\lambda)}$ may be appropriate when the EMH model holds. Also since the probabilities $\left\{p_{ij}\right\}$ are positive (not zero), the measure $\Gamma^{(\lambda)}(d)$ with $d<1$ rather than $\gamma^{(\lambda)}$ would be appropriate to represent the degree of departure from the MH toward the structure of maximum departure from MH which can be defined actually.

The conditional symmetry (CS) model (McCullagh, 1978) is defined by

$$p_{ij}=\theta p_{ji} \quad \text{for } i<j.$$

A special case of this model obtained by putting $\theta=1$ is the symmetry model (Bowker, 1948). If the symmetry model holds, then the MH model holds. Also if the CS model holds, then the EMH model holds. Therefore when the CS model holds, the measure $\Gamma^{(\lambda)}(d)$ is expressed by (1) with τ replaced by θ. Thus by the similar reason, when the CS model holds, the measure $\Gamma^{(\lambda)}(d)$ with $d<1$ rather than $\gamma^{(\lambda)}$ would be appropriate.

REFERENCES

[1] E. B. Andersen, Introduction to the Statistical Analysis of Categorical Data. Berlin, Germany: Springer, 1997.

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

[3] Bowker, A. H., 1948, A test for symmetry in contingency tables. Journal of the American Statistical Association, 43, 572-574.

[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] McCullagh, P., 1978, A class of parametric models for the analysis of square contingency tables with ordered categories. Biometrika, 65, 413-418.

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

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

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

[9] Tomizawa, S., Miyamoto, N. and Ashihara, N., 2003, Measure of departure from marginal homogeneity for square contingency tables having ordered categories. Behaviormetrika, 30, 173-193.

[10] Yamamoto, K., Masumura, K. and Tomizawa, S., 2011, Improved measures of departure from marginal homogeneity for square contingency tables with nominal categories. Oriental Journal of Statistical Methods, Theory and Applications, 1, 29-40.