

Measure of Departure from Average Cumulative Symmetry for Square Contingency Tables with Ordered Categories

Kouji Tahata, Kouta Miyazawa and Sadao Tomizawa
 Department of Information Sciences, Faculty of Science and Technology,
 Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

Abstract: Problem statement: For square contingency tables with ordered categories, we are interested in considering a structure of weak symmetry when Bowker's symmetry model does not hold and in measuring the degree of departure from weak symmetry. **Approach:** The present study considered the average cumulative symmetry model that has a weaker restriction than the structure of symmetry. It also gave a measure to represent the degree of departure from average cumulative symmetry. When the conditional symmetry and the cumulative linear diagonals-parameter symmetry holds, the proposed measure can measure what degree of departure from the symmetry is. **Results:** The proposed model and the measure were applied and analyzed (1) for the data of 4×4 contingency table of unaided distance vision of 7477 women aged 30-39 employed in Royal Ordnance factories in Britain from 1943-1946 and (2) the data of 4×4 contingency table of the 59 matched pairs using from dose levels of conjugated oestrogen. **Conclusion:** The proposed model is useful when the symmetry model does not hold and the proposed measure is useful for comparing the degree of departure from the weak symmetry model in several tables. Especially the proposed measure is useful to measure the degree of departure from symmetry when the conditional symmetry (the cumulative diagonals-parameter symmetry) model holds.

Key words: Average cumulative symmetry, conditional symmetry, cumulative linear diagonals-parameter symmetry, contingency table, cumulative symmetry

INTRODUCTION

Consider an $r \times r$ square contingency table. Let p_{ij} denote the probability that an observation will fall in the i th row and j th column of the table ($i = 1, \dots, r; j = 1, \dots, r$). The symmetry model is defined by:

$$p_{ij} = p_{ji} \quad (i \neq j)$$

See, for example, Bowker (1948) and Bishop *et al.* (1975). This model states that the probability that an observation will fall in cell (i, j) , $i \neq j$, is equal to the probability that it falls in symmetric cell (j, i) . Let:

$$G_{ij} = \sum_{s=1}^i \sum_{t=j}^r p_{st}, \quad G_{ji} = \sum_{s=j}^r \sum_{t=1}^i p_{st} \quad (i < j)$$

Using the cumulative probabilities $\{G_{ij}, G_{ji}\}$, the symmetry model can be expressed as:

$$G_{ij} = G_{ji} \quad (i < j)$$

When the symmetry model does not hold, we are interested in applying the model that has weaker restriction than the symmetry model. As such a model, for example, there are the marginal homogeneity model (Stuart, 1955) and the quasi symmetry model (Causinus, 1965). However, these models should be applied to the nominal categorical data because these models are invariant under the arbitrary same permutations of categories. So, we are interested in considering the structure of weak symmetry for the square contingency tables with ordinal categories. Moreover, when the structure of weak symmetry does not hold, we are interested in measuring what degree of departure from weak symmetry is.

The purpose of this study is to consider the structure of weak symmetry and to propose the measure to represent the degree of departure from weak symmetry. In present study, we consider the structure of weak symmetry and propose a measure. Moreover we show that it can measure what degree of departure from the symmetry is, when the conditional symmetry model (McCullagh, 1978) and the cumulative linear diagonals-

Corresponding Author: Kouji Tahata, Department of Information Sciences, Faculty of Science and Technology,
 Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

parameter symmetry model (Miyamoto *et al.*, 2004) hold. Two numerical examples are given.

MATERIALS AND METHODS

Average cumulative symmetry and measure: Consider the $r \times r$ table with ordered categories.

Let:

$$\Delta = \sum \sum_{i < j} (G_{ij} + G_{ji})$$

and:

$$G_{ij}^* = \frac{G_{ij}}{\Delta}, \quad G_{ji}^* = \frac{G_{ji}}{\Delta} \quad (i < j)$$

Assuming that $G_{ij} + G_{ji} \neq 0$ ($i < j$), we shall consider a measure defined by:

$$\phi = \frac{4}{\pi} \sum_{i < j} (G_{ij}^* + G_{ji}^*) \left(\theta_{ij} - \frac{\pi}{4} \right)$$

Where:

$$\theta_{ij} = \cos^{-1} \left(\frac{G_{ij}}{\sqrt{(G_{ij})^2 + (G_{ji})^2}} \right)$$

We note that for $i < j$, $(\theta_{ij} - (\pi/4))$ is (1) zero when $G_{ij} = G_{ji}$, (2) negative value when $G_{ij} > G_{ji}$ and (3) positive value when $G_{ij} < G_{ji}$. This measure is expressed as the weighted sum of $(\theta_{ij} - (\pi/4))$. Therefore we shall refer to the structure of $\phi = 0$ as the average cumulative symmetry. Note that if the symmetry holds then the average cumulative symmetry holds, but the converse does not hold.

Now we consider the two types of complete asymmetry. One shall refer to the upper-triangular-asymmetry that is the structure of $G_{ij} > 0$ (then $G_{ji} = 0$) for all $i < j$ and the other shall refer to the lower-triangular-asymmetry that is the structure of $G_{ji} > 0$ (then $G_{ij} = 0$) for all $i < j$. Since the range of θ_{ij} is $0 \leq \theta_{ij} \leq \pi/2$, the measure ϕ lies between -1 and 1. The measure ϕ has characteristic that (i) $\phi = -1$ if and only if there is a structure of upper-triangular-asymmetry and (ii) $\phi = 1$ if and only if there is a structure of lower-triangular-asymmetry.

As the measure ϕ approaches -1, the departure from the average cumulative symmetry becomes greater toward upper-triangular-asymmetry. Also, as the ϕ approaches 1, it becomes greater toward lower-

triangular-asymmetry. Using the ϕ , we can see whether the average cumulative symmetry departs toward the upper-or lower-triangular-asymmetry. Therefore the ϕ would be useful for comparing some contingency tables.

Relationship between the measure and asymmetry models: We consider the relationships between the measure ϕ and asymmetry models. The conditional symmetry model (McCullagh, 1978) is defined by:

$$G_{ij} = \Gamma G_{ji} \quad (i < j)$$

The cumulative linear diagonals-parameter symmetry model (Miyamoto *et al.*, 2004) is defined by:

$$G_{ij} = \Theta^{j-i} G_{ji} \quad (i < j)$$

We note that the conditional symmetry model with $\Gamma = 1$ and the cumulative linear diagonals-parameter symmetry model with $\Theta = 1$ are the symmetry model. First, if there is a structure of conditional symmetry in the table, then the measure ϕ can be expressed as:

$$\phi = \frac{4}{\pi} \cos^{-1} \left(\frac{\Gamma}{\sqrt{\Gamma^2 + 1}} \right) - 1$$

Therefore, $\phi = 0$ if and only if the symmetry model holds, i.e., $\Gamma = 1$, thus $G_{ij} = G_{ji}$ for all $i < j$. Next, if there is a structure of cumulative linear diagonals-parameter symmetry in the table, then the measure ϕ can be expressed as:

$$\phi = \frac{4}{\pi} \sum_{k=1}^{r-1} \left[(\Theta^k + 1) \left(\cos^{-1} \left(\frac{\Theta^k}{\sqrt{\Theta^{2k} + 1}} \right) - \frac{\pi}{4} \right) \sum_{i=1}^{r-k} G_{k+i,i}^* \right]$$

Therefore, $\phi = 0$ if and only if the symmetry model holds, i.e., $\Theta = 1$, thus $G_{ij} = G_{ji}$ for all $i < j$. As the value of Γ (Θ) approaches the infinity, ϕ approaches -1 and as it approaches zero, ϕ approaches 1. Thus, the measure ϕ can represent the degree of departure from the symmetry (not average cumulative symmetry) toward the upper-triangular-asymmetry and toward the lower-triangular-asymmetry, when the conditional symmetry model and the cumulative linear diagonals-parameter symmetry model holds.

Approximate confidence interval for measure: Let n_{ij} denote the observed frequency in the i^{th} row and j^{th}

column of the table ($i = 1, \dots, r, j = 1, \dots, r$). Assuming that a multinomial distribution applies to the $r \times r$ table, we shall consider the approximate variance for estimated measure and large-sample confidence interval for the measure ϕ , using delta method, the descriptions of which are given by, e.g., Agresti (2002). The sample version of ϕ , i.e., $\hat{\phi}$, is given by ϕ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$, where $\hat{p}_{ij} = n_{ij} / n$ and $n = \sum \sum n_{ij}$. Using delta method, $\sqrt{n}(\hat{\phi} - \phi)$ has asymptotically (as $n \rightarrow \infty$) a normal distribution with mean zero and variance:

$$\sigma^2[\phi] = \sum_{i < j} \sum \{p_{ij}(D_{ij})^2 + p_{ji}(D_{ji})^2\}$$

where for $i < j$:

$$D_{ij} = \frac{4}{\pi \Delta} \sum_{k=i}^{j-1} \sum_{l=k+1}^j \left\{ \theta_{kl} - \frac{G_{lk}(G_{kl} + G_{lk})}{(G_{kl})^2 + (G_{lk})^2} \right\} - \frac{(j-i)(j-i+1)(\phi+1)}{2\Delta}$$

$$D_{ji} = \frac{4}{\pi \Delta} \sum_{k=i}^{j-1} \sum_{l=k+1}^j \left\{ \theta_{kl} + \frac{G_{kl}(G_{kl} + G_{lk})}{(G_{kl})^2 + (G_{lk})^2} \right\} - \frac{(j-i)(j-i+1)(\phi+1)}{2\Delta}$$

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

RESULTS

Example 1: Consider the data in Table 1 taken from Stuart (1955). These are data on unaided distance vision of 7477 women aged 30-39 employed in Royal Ordnance factories in Britain from 1943-1946.

We see from Table 3 that for the data in Table 1, the estimated value of measure ϕ is -0.1150 and all values in the confidence interval for ϕ are negative. Therefore, the structure of average cumulative symmetry for the women's right and left eyes departs toward the upper-triangular-asymmetry. Table 4 gives the values of likelihood ratio chi-squared statistic for testing goodness-of-fit of each model. We see from Table 4 that each model of symmetry and average cumulative symmetry fits the data in Table 1 poorly, but the conditional symmetry model and linear

diagonals-parameter symmetry model fit these data well. So we can see from the estimated measure that the degree of departure from symmetry for the vision data in Table 1 is estimated to be 11.5% of the maximum departure toward the upper-triangular-asymmetry which indicates that the right eye is better than her left eye for all women.

Example 2: Consider the data in Table 2 taken from Breslow and Day (1980). These data are obtained from the 59 matched pairs using four dose levels of conjugated estrogen.

We see from Table 3 that for the data in Table 2, the estimated value of measure ϕ is 0.7460 and all values in the confidence interval for ϕ are positive.

Table 1: Unaided distance vision of 7477 women aged 30-39 employed in Royal Ordnance factories in Britain from 1943-1946; from Stuart (1955)

Right eye grade	Left eye grade				Total
	Best (1)	Second (2)	Third (3)	Worst (4)	
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 2: Average does of conjugated estrogen used by cases and matched controls; from Breslow and Day (1980)

Average dose for case (mg day ⁻¹)	Average dose for control (mg day ⁻¹)				Total
	0 (1)	0.1-0.299 (2)	0.3-0.625 (3)	0.626+ (4)	
0 (1)	6	2	3	1	12
0.1-0.299 (2)	9	4	2	1	16
0.3-0.625 (3)	9	2	3	1	15
0.626+ (4)	12	1	2	1	16
Total	36	9	10	4	59

Table 3: The estimates of ϕ , estimated approximate standard errors for ϕ and approximate 95% confidence intervals for ϕ , applied to Table 1 and 2

Applied data	Estimated measure	Standard error	Confidence interval
Table 1	-0.1150	0.0338	(-0.1812, -0.0488)
Table 2	0.7460	0.1061	(0.5380, 0.9540)

Table 4: The values of likelihood ratio chi-squared statistic for the models of Symmetry (S), Average Cumulative Symmetry (ACS), Conditional Symmetry (CS) and Cumulative Linear Diagonals-Parameter Symmetry (CLDPS) applied to Table 1 and 2. The parenthesized values are the numbers of degrees of freedom for the corresponding model

Applied data	S	ACS	CS	CLDPS
Table 1	19.25* (6)	11.29* (1)	7.35 (5)	8.63 (5)
Table 2	19.27*(6)	15.55* (1)	4.56 (5)	9.85 (5)

*: Means significant at the 0.05 level. The values of likelihood ratio chi-squared statistic for the models except the average cumulative symmetry are taken directly from Miyamoto *et al.* (2004)

Therefore, the average cumulative symmetry for the average dose for cases and controls departs toward the lower-triangular-asymmetry. Table 4 gives the values of likelihood ratio chi-squared statistic for testing goodness-of-fit of each model. We see from Table 4 that each model of symmetry and average cumulative symmetry fits the data in Table 2 poorly, but the conditional symmetry model and cumulative linear diagonals-parameter symmetry model fit these data well. So we can see from the estimated measure that the degree of departure from symmetry for the average dose data in Table 2 is estimated to be 74.6% of the maximum departure toward the lower-triangular-asymmetry which indicates that the average dose for case is greater than for control.

DISCUSSION

For square contingency tables with ordered categories, Tomizawa *et al.* (2001) proposed the power-divergence-type measure to represent the degree of departure from symmetry.

Assuming that $\{G_{ij} + G_{ji} > 0\}$, the measure is defined by:

$$\Phi^{(\lambda)} = \frac{1}{2^\lambda - 1} \sum_{i < j} \left\{ G_{ij}^* \left(\left(\frac{G_{ij}^*}{Q_{ij}^*} \right)^\lambda - 1 \right) + G_{ji}^* \left(\left(\frac{G_{ji}^*}{Q_{ji}^*} \right)^\lambda - 1 \right) \right\}$$

($\lambda > -1$)

Where:

$$Q_{ij}^* = Q_{ji}^* = \frac{G_{ij}^* + G_{ji}^*}{2}$$

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

The measure $\Phi^{(\lambda)}$ has three properties, (i) $\Phi^{(\lambda)}$ lies between 0 and 1, (ii) $\Phi^{(\lambda)} = 0$ if and only if the symmetry model holds and (iii) $\Phi^{(\lambda)} = 1$ if and only if the degree of departure from symmetry is maximum; that is, $G_{ij} = 0$ (then $G_{ji} > 0$) for all $i < j$ or $G_{ji} = 0$ (then $G_{ij} > 0$) for all $i < j$.

In the present study, we have considered the two types of complete asymmetry. Nevertheless upper-triangular-asymmetry is just the opposite of lower-triangular-asymmetry, the measure $\Phi^{(\lambda)}$ takes the value 1 in two types of complete asymmetry. The proposed measure ϕ is useful for representing what degree of the departure from the average cumulative symmetry is toward two kinds of complete asymmetry (i.e., the upper-triangular-asymmetry and the lower-triangular-

asymmetry). Also, the measure ϕ can represent the degree of departure from symmetry, when the conditional symmetry and cumulative linear diagonals-parameter symmetry holds. Therefore, when these models holds, the measure ϕ can distinguish these two kinds of complete asymmetry although the measure $\Phi^{(\lambda)}$ in Tomizawa *et al.* (2001) cannot distinguish them.

CONCLUSION

Since the measure ϕ lies between -1 and 1 without the dimension and the sample size, the ϕ would be useful for comparing the degrees of departure from the average cumulative symmetry in several tables.

The measure ϕ should be applied to the ordinal data of square tables with the same row and column classifications because the ϕ is not invariant under arbitrary similar permutations of row and column categories.

We note that (i) the measure $\Phi^{(\lambda)}$ which is proposed by Tomizawa *et al.* (2001) can represent the degree of departure from symmetry and (ii) the measure ϕ can represent the degree of departure from symmetry only when the conditional symmetry (cumulative diagonals-parameter symmetry) model holds.

REFERENCES

Agresti, A., 2002. Categorical Data Analysis. 2nd Edn., Wiley-Interscience, New York, ISBN: 978-0471360933, pp: 734.

Bishop, Y.M.M., S.E. Fienberg and P.W. Holland, 1975. Discrete Multivariate Analysis: Theory and Practice. 1st Edn., The MIT Press, Cambridge, ISBN: 0262021137, pp: 557.

Bowker, A.H., 1948. A test for symmetry in contingency tables. J. Am. Stat. Assoc., 43: 572-574. <http://www.jstor.org/stable/2280710>

Breslow, N.E. and N.E. Day, 1980. Statistical Methods in Cancer Research. 1st Edn., International Agency for Research on Cancer, Lyon, France, ISBN: 13: 9789283201328, pp: 350.

Caussinus, H., 1965. Contribution à l'analyse statistique des tableaux de corrélation. Ann. Fac. Sci. Toulouse, 29: 77-182. http://www.numdam.org/item?id=AFST_1965_4_29__77_0

McCullagh, P., 1978. A class of parametric models for the analysis of square contingency tables with ordered categories. Biometrika, 65: 413-418. DOI: 10.1093/biomet/65.2.413

- Miyamoto, N., W. Ohtsuka and S. Tomizawa, 2004. Linear diagonals-parameter symmetry and quasi-symmetry models for cumulative probabilities in square contingency tables with ordered categories. *Biomet. J.*, 46: 664-674. DOI: 10.1002/bimj.200410066
- Stuart, A., 1955. A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 42: 412-416. DOI: 10.1093/biomet/42.3-4.412
- Tomizawa, S., N. Miyamoto and Y. Hatanaka, 2001. Measure of asymmetry for square contingency tables having ordered categories. *Aust. N. Z. J. Stat.*, 43: 335-349. DOI: 10.1111/1467-842X.00180