Statistica Sinica 4(1994), 325-334

TWO KINDS OF MEASURES OF DEPARTURE FROM SYMMETRY IN SQUARE CONTINGENCY TABLES HAVING NOMINAL CATEGORIES

Sadao Tomizawa

Science University of Tokyo

Abstract: Two kinds of measures are proposed to represent the degree of departure from symmetry about the main diagonal of square contingency tables having the same nominal row and column classifications. One measure is expressed by using the Kullback-Leibler information (or the average of conditional Shannon entropy), and the other is expressed by using the Pearson's chi-squared type discrepancy (or the average of conditional Gauss discrepancy). These measures would be useful for comparing the degree of departure from symmetry in several tables.

Key words and phrases: Gauss discrepancy, Kullback-Leibler information, Pearson's chi-squared type discrepancy, Shannon entropy.

1. Introduction

For an $R \times C$ contingency table, many coefficients have been proposed to measure the association between the row and column variables. Examples include Yule's coefficients Q of association and Y of colligation, Pearson's coefficients Φ^2 of mean square contingency and P of contingency, Tschuprow's coefficient T, and Cramér's coefficient V. For details, e.g., see Bishop et al. (1975, Ch.11). These measures represent, in some sense, the degree of departure from independence, i.e., no association. These measures may be useful for comparing several tables.

Next, consider an $R \times R$ square contingency table having the same nominal row and column classifications. For the analysis of such tables, interest would be in the symmetry about the main diagonal of the table rather than the independence between the row and column variables (see Bishop et al. (1975, Ch.8), Agresti (1984, Ch.11)). Let p_{ij} denote the probability that an observation will fall in the ith row and jth column of the table ($i = 1, 2, \ldots, R$; $j = 1, 2, \ldots, R$). The usual symmetry model is defined as

$$p_{ij} = p_{ji} \qquad \text{for } i = 1, 2, \ldots, R;\ j = 1, 2, \ldots, R;\ i \neq j.$$

See Bowker (1948), Bishop et al. (1975, p.282) and Agresti (1984, p.202). (Although the details are omitted, we note that the general topic of symmetry and related work have been considered by many statisticians; see, for example, Tomizawa (1984, 1985, 1987, 1989, 1990, 1992), Havránek and Lienert (1986), Rosenstein (1989), Chino (1990) and Becker (1990).)

The purpose of this paper is to propose measures which represent the degree of departure from symmetry in an $R \times R$ table. The proposed measures would be useful for comparing the degree of departure from symmetry in several tables to which the symmetry model has been fitted.

2. Measures of Departure from Symmetry

Assuming that $\{p_{ij} + p_{ji}\}$ for $i \neq j$ are all positive, consider two kinds of measures defined in the population form by

$$\begin{aligned}\phi_s &= \frac{1}{\delta \log 2} \sum\sum_{i \neq j} p_{ij} \log \frac{2p_{ij}}{p_{ij} + p_{ji}},\\ \psi_s &= \frac{1}{\delta} \sum\sum_{i<j} \frac{(p_{ij} - p_{ji})^2}{p_{ij} + p_{ji}},\end{aligned}$$

where $\delta = \sum\sum_{i \neq j} p_{ij}$ and $0 \log 0 = 0$. Let $p^*_{ij} = p_{ij}/\delta$ and $p^s_{ij} = (p^*_{ij} + p^*_{ji})/2$ for $i = 1, 2, \ldots, R$; $j = 1, 2, \ldots, R$; $i \neq j$. Then ϕ_s and ψ_s may be also expressed as

$$\phi_s = \frac{I(p^*; p^s)}{\log 2}, \qquad \psi_s = D(p^*; p^s),$$

where

$$I(p^*; p^s) = \sum\sum_{i \neq j} p^*_{ij} \log \frac{p^*_{ij}}{p^s_{ij}}, \qquad D(p^*; p^s) = \sum\sum_{i \neq j} \frac{(p^*_{ij} - p^s_{ij})^2}{p^s_{ij}}.$$

Note that $I(p^*; p^s)$ and $D(p^*; p^s)$ are the Kullback-Leibler information and the Pearson's chi-squared type discrepancy, respectively, between $\{p^*_{ij}\}_{i \neq j}$ and $\{p^s_{ij}\}_{i \neq j}$. Let $p^c_{ij} = p_{ij}/(p_{ij} + p_{ji})$ for $i = 1, 2, \ldots, R$; $j = 1, 2, \ldots, R$; $i \neq j$. (Note that $p^c_{ij} + p^c_{ji} = 1$.) Then these measures may be further expressed as

$$\begin{aligned}\phi_s &= 1 - \frac{1}{\log 2} \sum\sum_{i<j} (p^*_{ij} + p^*_{ji}) H_{ij}(p^c),\\ \psi_s &= 2 \sum\sum_{i<j} (p^*_{ij} + p^*_{ji}) \Delta_{ij}(p^c, 1/2),\end{aligned}$$

where

$$\begin{aligned}H_{ij}(p^c) &= -p^c_{ij} \log p^c_{ij} - p^c_{ji} \log p^c_{ji},\\ \Delta_{ij}(p^c, 1/2) &= (p^c_{ij} - 1/2)^2 + (p^c_{ji} - 1/2)^2.\end{aligned}$$

Given the condition that an observation falls in one of the off-diagonal cells of square table, (i) ϕ_s would represent, essentially, the average of Shannon entropy $H_{ij}(p^c)$ on condition that the observation falls in cell (i,j) or (j,i), $i \neq j$; and (ii) ψ_s represents the average of Gauss discrepancy $\Delta_{ij}(p^c, 1/2)$ between $\{p^c_{ij}, p^c_{ji}\}$ and $\{1/2, 1/2\}$ on the condition that the observation falls in cell (i,j) or (j,i), $i \neq j$; (see Linhart and Zucchini (1986, p.18) for Gauss discrepancy).

Now it is easily seen that $I(p^*; p^s)$ must lie between 0 and log 2, and therefore ϕ_s must lie between 0 and 1, and also ψ_s must lie between 0 and 1. Also, (i) there is a structure of symmetry in the $R \times R$ table, i.e., $p_{ij} = p_{ji}$ for $i = 1, 2, \ldots, R$; $j = 1, 2, \ldots, R$; $i \neq j$; if and only if ϕ_s (ψ_s) equals zero, and (ii) there is such a structure that the degree of departure from symmetry is the largest in a sense (say, complete asymmetry), i.e., either $p_{ij} = 0$ or $p_{ji} = 0$ for $i = 1, 2, \ldots, R$; $j = 1, 2, \ldots, R$; $i \neq j$; if and only if ϕ_s (ψ_s) equals 1. According to the Kullback-Leibler information (Pearson's chi-squared type discrepancy) metric, ϕ_s (ψ_s) represents the degree of departure from symmetry, and the degree increases as the value of ϕ_s (ψ_s) increases.

Let n_{ij} denote the observed frequency in the ith row and jth column of the square table ($i = 1, 2, \ldots, R$; $j = 1, 2, \ldots, R$). Assuming that the $\{n_{ij}\}$ result from full multinomial sampling, we consider an approximate standard error and large-sample confidence interval for ϕ_s (ψ_s), 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 ϕ_s (ψ_s), i.e., $\hat{\phi}_s$ ($\hat{\psi}_s$), is given by ϕ_s (ψ_s) 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*, $\sqrt{n}(\hat{\phi}_s - \phi_s)$ and $\sqrt{n}(\hat{\psi}_s - \psi_s)$ have asymptotically (as $n \to \infty$) normal distributions with mean zero and variances

$$\sigma^2_\phi = \left(\sum_{i \neq j}\sum p_{ij}\Omega^2_{ij} - \delta\phi^2_s \right) \Big/ \delta^2$$

and

$$\sigma^2_\psi = \left(\sum_{i \neq j}\sum p_{ij}\Gamma^2_{ij} - \delta\psi^2_s \right) \Big/ \delta^2,$$

respectively, where

$$\Omega_{ij} = \frac{1}{\log 2} \log \frac{2p_{ij}}{p_{ij} + p_{ji}}, \qquad \Gamma_{ij} = \frac{(p_{ij} - p_{ji})(p_{ij} + 3p_{ji})}{(p_{ij} + p_{ji})^2}.$$

Let $\hat{\sigma}^2_\phi$ denote σ^2_ϕ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$. Then $\hat{\sigma}_\phi/\sqrt{n}$ is an estimated standard error for $\hat{\phi}_s$, and $\hat{\phi}_s \pm z_{p/2}\hat{\sigma}_\phi/\sqrt{n}$ is an approximate $100(1-p)$ percent confidence interval for ϕ_s, where $z_{p/2}$ is the percentage point from the standard normal distribution corresponding to a two-tail probability equal to p. In a similar way, an approximate confidence interval for ψ_s is given.

Let G_s^2 (X_s^2) denote the likelihood ratio (Pearson's) chi-squared statistic for testing goodness of fit of the symmetry model, i.e., $G_s^2 = 2\sum\sum_{i\neq j} n_{ij}\log(2n_{ij}/(n_{ij}+n_{ji}))$ and $X_s^2 = \sum\sum_{i<j}(n_{ij}-n_{ji})^2\big/(n_{ij}+n_{ji})$. Note that $\hat{\phi}_s$ ($\hat{\psi}_s$) may then be expressed as $\hat{\phi}_s = G_s^2/n^*$ ($\hat{\psi}_s = X_s^2/n^{**}$), where $n^* = (2\log 2)\sum\sum_{i\neq j} n_{ij}$ and $n^{**} = \sum\sum_{i\neq j} n_{ij}$.

3. Examples

Tables 1a and 1b present data which were earlier analyzed by Andersen (1980, p.328). As reported in Andersen, these data are the results of three consecutive opinion polls held in August 1971, October 1971 and December 1973, which were held in connection with the Danish referendum on whether to join the Common Market or not.

Since the confidence interval for ϕ_s (ψ_s) applied to the data in Table 1a includes zero (see Table 2), this would indicate that there is a structure of symmetry in Table 1a, or if it is not so, that the degree of departure from symmetry is slight. On the other hand, the confidence interval for ϕ_s (ψ_s) applied to the data in Table 1b (see Table 2) would indicate that there is not a structure of symmetry in Table 1b. (See Table 2 for the values of G_s^2 and X_s^2 applied to Table 1.)

4. Note

Consider the artificial data in Table 3. From the values of G_s^2 (X_s^2) in Table 4, the hypothesis of symmetry is accepted for the data in Table 3a (at the 0.05 significance level) but it is rejected for the other data. The value of G_s^2 (X_s^2) for Table 3b is ten times as large as it for Table 3a, but the value of $\hat{\phi}_s$ ($\hat{\psi}_s$) for Table 3b is equal to it for Table 3a. It is easily seen that $\{\hat{p}_{ij}\}$ for Table 3a is equal to $\{\hat{p}_{ij}\}$ for Table 3b. So, it seems natural to consider that the degree of departure from symmetry for Table 3a is the same as for Table 3b. Hence, $\hat{\phi}_s$ ($\hat{\psi}_s$) would be preferable to G_s^2 (X_s^2) for representing the degree of departure from symmetry. (This also would be concluded from the comparison between Tables 3c and 3d and the comparison between Tables 3e and 3f.)

From the value of G_s^2 (X_s^2), the symmetry model fits the data in Table 3c worse than the data in Table 3a, and also it fits the data in Table 3e worse than the data in Table 3c; but it fits the data in Table 3e better than the data in Table 3d. In terms of $\{\hat{p}_{ij}^c\}$, it seems natural to consider that the degree of departure from symmetry for Table 3e is greater than for Table 3c, and the departure for Table 3c is the same as for Table 3d. So, $\hat{\phi}_s$ ($\hat{\psi}_s$) would be preferable to G_s^2 (X_s^2) for comparing the degree of departure from symmetry.

If $\{\hat{p}_{ij}\}$ for two tables are same, the estimated standard error for table with larger sample size is less than that for the other table. In fact, $\{\hat{p}_{ij}\}$ for Tables 3a and 3b are the same; and the estimated standard error for Table 3b is less than that for Table 3a (see Table 4). (A similar result is also obtained from the comparison between Tables 3c and 3d and the comparison between Tables 3e and 3f.)

5. Comments

The symmetry model imposes no restriction on the diagonal cell probabilities $\{p_{ii}\}$. So it seems natural that the measures of degree of departure from symmetry and their ranges do not depend on the diagonal probabilities.

The structure of symmetry based on the probabilities $\{p_{ij}\}$, i.e., $p_{ij} = p_{ji}$ for $i \neq j$, may be also expressed as $p^*_{ij} = p^*_{ji}$ for $i \neq j$, using the conditional probabilities $\{p^*_{ij}\}_{i \neq j}$. In the sample version, G^2_s/n (X^2_s/n) is a measure based on $\{\hat{p}_{ij}\}$, and $\hat{\phi}_s$ $(\hat{\psi}_s)$ (i.e., G^2_s/n^* (X^2_s/n^{**})) is essentially the corresponding measure based on $\{\hat{p}^*_{ij}\}_{i \neq j}$. It may seem, to many readers, that both are reasonable measures for representing the degree of departure from symmetry. However, $\hat{\phi}_s$ $(\hat{\psi}_s)$ rather than G^2_s/n (X^2_s/n) would be useful for comparing the degree of departure from symmetry in several tables. Because the ranges of G^2_s/n and X^2_s/n depend on the diagonal proportions, i.e., $0 \leq (G^2_s/n) \leq (n^*/n)\,[= (2\log 2)(1 - \sum n_{ii}/n)]$ and $0 \leq (X^2_s/n) \leq (n^{**}/n)\,[= 1 - \sum n_{ii}/n]$; but $\hat{\phi}_s$ and $\hat{\psi}_s$ always range between 0 and 1 without depending on the diagonal proportions.

It is known that the symmetry model holds if and only if the quasi-symmetry model and the marginal homogeneity model hold (see Bishop et al. (1975, p.287)). Therefore, for example, when there is marginal homogeneity, the degree of departure from symmetry should be considered by a measure which takes the minimum value when there is symmetry and the maximum value when the degree of departure from symmetry is the largest but there is marginal homogeneity. So, if one wants to see the degree of departure from symmetry under the assumption that there is marginal homogeneity (or quasi-symmetry), the measures ϕ_s and ψ_s would not be suitable. The measures ϕ_s and ψ_s should be used when one wants to see the degree of departure from symmetry under no assumption that there is an extended symmetry (as quasi-symmetry and marginal homogeneity).

The ϕ_s (ψ_s) would be useful when one wants to see the degree of departure from symmetry, using the Kullback-Leibler information (Pearson's chi-squared type discrepancy) or the average of conditional Shannon entropy (the average of conditional Gauss discrepancy), on *condition* that an observation will fall in one of the off-diagonal cells of square table.

The reader may be interested in which of two measures, ϕ_s and ψ_s, is preferred for a given table; however, it seems difficult to discuss it. It seems to be important that for given tables, the analyst calculates both values of $\hat{\phi}_s$ and $\hat{\psi}_s$ and discusses the degree of departure from symmetry in terms of both values.

Finally, we observe that (i) ϕ_s (ψ_s) should be applied to square contingency tables having *nominal* categories because it is invariant under the same arbitrary permutations of row and column categories, (ii) the estimate of the degree of departure from symmetry should be considered in terms of an approximate confidence interval for ϕ_s (ψ_s) and not in terms of $\hat{\phi}_s$ ($\hat{\psi}_s$) itself, and (iii) the asymptotic normal distributions of $\sqrt{n}(\hat{\phi}_s - \phi_s)$ and $\sqrt{n}(\hat{\psi}_s - \psi_s)$, which are described in Section 2, are applicable only when $0 < \phi_s < 1$ and $0 < \psi_s < 1$, respectively; because $\sigma_\phi^2 = 0$ ($\sigma_\psi^2 = 0$) when $\phi_s = 0$ ($\psi_s = 0$) and $\phi_s = 1$ ($\psi_s = 1$), and $\sigma_\phi^2 > 0$ ($\sigma_\psi^2 > 0$) when $0 < \phi_s < 1$ ($0 < \psi_s < 1$).

Acknowledgements

The author is grateful to a referee for the valuable suggestions on improving this paper.

Table 1. The results from two consecutive polls on the question: Do you think Denmark should join the Common Market?

(a) The results of the first and second polls.

Poll I	Poll II			
	Yes	No	Undecided	Total
Yes	176	33	40	249
No	21	94	32	147
Undecided	21	33	43	97
Total	218	160	115	493

(b) The results of the second and third polls.

Poll II	Poll III			
	Yes	No	Undecided	Total
Yes	167	36	15	218
No	19	131	10	160
Undecided	45	50	20	115
Total	231	217	45	493

Source: Andersen (1980, p.328).

Table 2. Estimate of ϕ_s (ψ_s), estimated approximate standard error for $\hat{\phi}_s$ ($\hat{\psi}_s$), approximate 95% confidence interval for ϕ_s (ψ_s) and chi-squared values G_s^2 and X_s^2, applied to Tables 1a and 1b.

(a) Case of ϕ_s:

Applied data	Estimated measure $\hat{\phi}_s$	Standard error $\hat{\sigma}_\phi/\sqrt{n}$	Confidence interval $\hat{\phi}_s \pm z_{0.025}\hat{\sigma}_\phi/\sqrt{n}$
Table 1a	0.035	0.023	(−0.011, 0.081)
Table 1b	0.207	0.053	(0.103, 0.311)

(b) Case of ψ_s:

Applied data	Estimated measure $\hat{\psi}_s$	Standard error $\hat{\sigma}_\psi/\sqrt{n}$	Confidence interval $\hat{\psi}_s \pm z_{0.025}\hat{\sigma}_\psi/\sqrt{n}$
Table 1a	0.048	0.031	(−0.014, 0.109)
Table 1b	0.268	0.064	(0.142, 0.394)

(c) Chi-squared values:

Applied data	Degrees of freedom	Likelihood ratio chi-squared G_s^2	Pearson's chi-squared X_s^2
Table 1a	3	8.7	8.6
Table 1b	3	50.1	46.9

Table 3. Artificial data

(a) $n = 53$ (sample size)

2	7	4	2
5	3	3	3
2	4	4	3
1	3	2	5

(b) $n = 530$

20	70	40	20
50	30	30	30
20	40	40	30
10	30	20	50

(c) $n = 108$

2	15	10	2
5	3	3	4
2	12	4	14
8	16	3	5

(d) $n = 1080$

20	150	100	20
50	30	30	40
20	120	40	140
80	160	30	50

(e) $n = 243$

2	45	22	2
5	3	3	4
2	33	4	36
26	48	3	5

(f) $n = 2430$

20	450	220	20
50	30	30	40
20	330	40	360
260	480	30	50

Table 4. Estimate of ϕ_s (ψ_s), estimated approximate standard error for $\hat{\phi}_s$ ($\hat{\psi}_s$), approximate 95% confidence interval for ϕ_s (ψ_s) and chi-squared values G_s^2 and X_s^2, applied to Table 3.

(a) Case of ϕ_s:

Applied data	Estimated measure $\hat{\phi}_s$	Standard error $\hat{\sigma}_\phi/\sqrt{n}$	Confidence interval $\hat{\phi}_s \pm z_{0.025}\hat{\sigma}_\phi/\sqrt{n}$
Table 3a	0.031	0.047	(−0.062, 0.124)
Table 3b	0.031	0.015	(0.002, 0.061)
Table 3c	0.277	0.082	(0.116, 0.438)
Table 3d	0.277	0.026	(0.226, 0.328)
Table 3e	0.588	0.063	(0.465, 0.712)
Table 3f	0.588	0.020	(0.549, 0.627)

(b) Case of ψ_s:

Applied data	Estimated measure $\hat{\psi}_s$	Standard error $\hat{\sigma}_\psi/\sqrt{n}$	Confidence interval $\hat{\psi}_s \pm z_{0.025}\hat{\sigma}_\psi/\sqrt{n}$
Table 3a	0.043	0.064	(−0.083, 0.169)
Table 3b	0.043	0.020	(0.003, 0.083)
Table 3c	0.358	0.098	(0.165, 0.551)
Table 3d	0.358	0.031	(0.297, 0.419)
Table 3e	0.696	0.061	(0.577, 0.815)
Table 3f	0.696	0.019	(0.658, 0.734)

(c) Chi-squared values:

Applied data	Degrees of freedom	Likelihood ratio chi-squared G_s^2	Pearson's chi-squared X_s^2
Table 3a	6	1.7	1.7
Table 3b	6	17.0	16.8
Table 3c	6	36.1	33.7
Table 3d	6	361.2	336.5
Table 3e	6	186.8	159.4
Table 3f	6	1867.7	1593.9

References

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

Andersen, E. B. (1980). *Discrete Statistical Models With Social Science Applications.* North-Holland, Amsterdam.

Becker, M. P. (1990). Quasisymmetric models for the analysis of square contingency tables. *J. Roy. Statist. Soc. Ser.B* **52**, 369–378.

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

Bowker, A. H. (1948). A test for symmetry in contingency tables. *J. Amer. Statist. Assoc.* **43**, 572–574.

Chino, N. (1990). A generalized inner product model for the analysis of asymmetry. *Behaviormetrika* **27**, 25–46.

Havránek, T. and Lienert, G. A. (1986). Pre-post treatment evaluation by symmetry testing in square contingency table. *Biometrical J.* **28**, 927–935.

Linhart, H. and Zucchini, W. (1986). *Model Selection.* John Wiley, New York.

Rosenstein, R. B. (1989). Tests of symmetry derived as components of Pearson's phi-squared distance measure. *Comm. Statist. Theory Methods* **18**, 1617–1626.

Tomizawa, S. (1984). Three kinds of decompositions for the conditional symmetry model in a square contingency table. *J. Japan Statist. Soc.* **14**, 35–42.

Tomizawa, S. (1985). The decompositions for point-symmetry models in two-way contingency tables. *Biometrical J.* **27**, 895–905.

Tomizawa, S. (1987). Diagonal weighted marginal homogeneity models and decompositions for linear diagonals-parameter symmetry model. *Comm. Statist. Theory Methods* **16**, 477–488.

Tomizawa, S. (1989). Decompositions for conditional symmetry model into palindromic symmetry and modified marginal homogeneity models. *Austral. J. Statist.* **31**, 287–296.

Tomizawa, S. (1990). Another linear diagonals-parameter symmetry model for square contingency tables with ordered categories. *South African Statist. J.* **24**, 117–125.

Tomizawa, S. (1992). Multiplicative models with further restrictions on the usual symmetry model. *Comm. Statist. Theory Methods* **21**, 693–710.

Department of Information Sciences, Faculty of Science & Technology, Science University of Tokyo, Noda City, Chiba 278, Japan.

(Received July 1991; accepted July 1993)