Journal of Mathematics and Statistics

Original Research Paper

Geometric Mean Type Measure of Marginal Homogeneity for Square Contingency Tables with Ordered Categories

Tomoyuki Nakagawa, Toshiki Takei, Aki Ishii and Sadao Tomizawa

Department of Information Sciences, Tokyo University of Science, Chiba, Japan

Article history
Received: 15-06-2020
Revised: 25-07-2020
Accepted: 10-08-2020

Corresponding Authors:
Tomoyuki Nakagawa
Department of Information Sciences, Tokyo University of Science, Chiba, Japan
Email: t_nakagawa@rs.tus.ac.jp

Abstract: For square contingency tables, some studies have developed the weighted arithmetic mean type measure to represent the degree of departure from the marginal homogeneity. The present paper proposes (1) the cumulative partial marginal homogeneity model which has the weaker restriction than the marginal homogeneity model and (2) the measure to represent the degree of departure from the proposed model. The measure is expressed as a weighted geometric mean of the diversity index. Finally, numerical studies are presented.

Keywords: Geometric Mean, Marginal Homogeneity, Measure, Square Contingency Table

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 ith row and jth column of the table $(i = 1,...,r : j = 1,...,r)$. Stuart (1955) gave the Marginal Homogeneity (MH) model defined by:

$$p_{i\cdot} = p_{\cdot i} \ \textit{for all}\ i = 1,...,r,$$

where, $p_{i\cdot} = \sum_{t=1}^{r} p_{it}$ and $p_{\cdot i} = \sum_{s=1}^{r} p_{si}$. The MH model indicates that the row marginal distribution is identical to the column marginal distribution. Saigusa *et al.* (2020) proposed the Partial Marginal Homogeneity (PMH) model which has weaker restriction than the MH model as follows:

$$p_{i\cdot} = p_{\cdot i} \ \textit{for at least one}\ i\,(i = 1,...,r).$$

The PMH model indicates that the row marginal distribution is identical to the column marginal distribution for at least one i. In addition, Saigusa *et al.* (2020) also proposed the measure to represent the degree of departure from the PMH model. Assuming $p_{i\cdot} + p_{\cdot i} \neq 0$ for all $i = 1,\ldots, r$, the measure is defined by:

$$\Phi^{(\lambda)} = \prod_{i=1}^{r}\left(\gamma_i^{(\lambda)}\right)^{\pi_i} \ (\lambda > -1),$$

where, $\pi_i = (p_{i\cdot} + p_{\cdot i})/2$, $p_{1(i)} = p_{i\cdot}/(p_{i\cdot} + p_{\cdot i})$ and $p_{2(i)} = p_{\cdot i}/(p_{i\cdot} + p_{\cdot i})$ and:

$$\gamma_i^{(\lambda)} = 1 - \frac{\lambda 2^{\lambda}}{2^{\lambda} - 1} I_i^{(\lambda)},$$

$$I_i^{(\lambda)} = \frac{1}{\lambda}\left\{1 - \left(p_{1(i)}\right)^{\lambda+1} - \left(p_{2(i)}\right)^{\lambda+1}\right\}.$$

The value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$. It should be noted that $\Phi^{(\lambda)}$ is expressed the weighted geometric mean of $\left\{\gamma_i^{(\lambda)}\right\}$. We also remark that the $I_i^{(\lambda)}$ is the diversity index given by (Patil and Taillie, 1982). $I_i^{(0)}$ is identical to the Shannon entropy.

Over the past few years, such partial structure and geometric mean type measure have been developed by many studies (Saigusa *et al*., 2016; 2019).

Let X and Y denote the row and column variables, respectively. By considering the difference between cumulative probabilities $G_{1(i)}$ and $G_{2(i)}$, the MH model is also expressed as:

$$G_{1(i)} = G_{2(i)} \ \textit{for all}\ i = 1,...,r-1,$$

where the cumulative probabilities are defined as follows:

$$G_{1(i)} = P(X \le i, Y \ge i+1) = \sum_{s=1}^{i}\sum_{t=i+1}^{r} p_{st},$$

$$G_{2(i)} = P(X \ge i+1, Y \le i) = \sum_{s=i+1}^{r}\sum_{t=1}^{i} p_{st}.$$

Tomizawa *et al*. (2003) proposed the measure to represent the degree of departure from the MH model.

Assuming $G_{1(i)} + G_{2(i)} \neq 0$ for $i = 1,..., r - 1$, we put $G^c_{1(i)} = G_{1(i)}/(G_{1(i)} + G_{2(i)})$, $G^c_{2(i)} = G_{2(i)}/(G_{1(i)} + G_{2(i)})$, $G^*_{1(i)} = G_{1(i)}/\Delta$, $G^*_{2(i)} = G_{2(i)}/\Delta$ and $\Delta = \sum_{i=1}^{r-1}\left(G_{1(i)} + G_{2(i)}\right)$. The measure is defined by:

$$\Gamma_M^{(\lambda)} = \sum_{i=1}^{r-1}\left(G^*_{1(i)} + G^*_{2(i)}\right)\omega_i^{(\lambda)} \,(\lambda > -1),$$

where:

$$\begin{aligned}\omega_i^{(\lambda)} &= 1 - \frac{\lambda 2^\lambda}{2^\lambda - 1} H_i^{(\lambda)},\\ H_i^{(\lambda)} &= \frac{1}{\lambda}\left\{1 - \left(G^c_{1(i)}\right)^{\lambda+1} - \left(G^c_{2(i)}\right)^{\lambda+1}\right\}.\end{aligned} \quad (1)$$

Note that the value at $\lambda = 0$ is taken to be the limit as $\lambda \to 0$, that is:

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

$\Gamma_M^{(\lambda)}$ is the weighted arithmetic mean type measure of $\left\{\omega_i^{(\lambda)}\right\}$.

Tomizawa (2001) gave the measure to represent the degree of departure from the MH model for square contingency tables with nominal categories. As described above, (Tomizawa *et al*., 2003) gave the measure $\Gamma_M^{(\lambda)}$ from the MH model for those with ordered categories. Saigusa *et al*. (2020) proposed the measure $\Phi^{(\lambda)}$ for the PMH model for square tables with nominal categories. In this study, we are interested in the partial homogeneity of cumulative marginal probabilities $G_{1(i)}$ and $G_{2(i)}$ for square tables with ordered categories.

This paper is organized as follows. Section 2 proposes a new model with respect to the partial homogeneity of cumulative marginal probabilities $G_{1(i)}$ and $G_{2(i)}$ and a measure to represent the degree of departure from the new model. Section 3 derives approximate confidence interval of the measure. Section 4 applies the measure to artificial examples and real data.

Model and Measure

In this section, we propose a new model which has the structure of the cumulative partial marginal homogeneity for an $r \times r$ contingency table with ordered categories. In addition, we also propose the geometric mean type measure to represent the degree of departure from the new model.

New Model

A new model is proposed as:

$$G_{1(i)} = G_{2(i)} \text{ for at least one } i \; (i = 1,...,r-1).$$

Table 1: Artificial cell probability tables

(a) Table 1a

	(1)	(2)	(3)	(4)	Total
(1)	0.04	0.07	0.03	0.01	0.15
(2)	0.05	0.05	0.12	0.13	0.35
(3)	0.03	0.17	0.06	0.04	0.30
(4)	0.03	0.06	0.09	0.02	0.20
Total	0.15	0.35	0.30	0.20	1.00

(b) Table 1b

	(1)	(2)	(3)	(4)	Total
(1)	0.04	0.07	0.03	0.01	0.15
(2)	0.05	0.05	0.02	0.17	0.29
(3)	0.03	0.10	0.04	0.05	0.22
(4)	0.03	0.09	0.03	0.19	0.34
Total	0.15	0.31	0.12	0.42	1.00

Table 2: Cumulative probability tables for Table 1

(a) Cumulative probabilities for Table 1a

	$i = 1$	2	3
G1*(i)*	0.11	0.29	0.18
G2*(i)*	0.11	0.29	0.18

(b) Cumulative probabilities for Table 1b

	$i = 1$	2	3
G1*(i)*	0.11	0.23	0.23
G2*(i)*	0.11	0.25	0.15

We refer to this model as the Cumulative Partial Marginal Homogeneity (CPMH) model herein. It should be noted that the CPMH model has a different structure from the MH and PMH models. It is easy to see that the MH model has the structure of CPMH. Consider the artificial probability in Table 1 and the marginal cumulative probability for in Table 2. Tables 2a and 2b give the cumulative probability calculated from Tables 1a and 1b, respectively. Table 1a has the structure of MH and CPMH models, while Table 1b does not have that of the MH model. Therefore the CPMH model is not equivalent to the MH model.

Measure

Assume that $G_{1(i)} + G_{2(i)} \neq 0$ for $i = 1,\cdots, r - 1$. The measure to represent the degree of departure from the CPMH model is proposed as follows:

$$\tau_M^{(\lambda)} = \prod_{i=1}^{r-1}\left(\omega_i^{(\lambda)}\right)^{\left(G^*_{1(i)} + G^*_{2(i)}\right)} (\lambda > -1),$$

is defined by (1). Note that λ is a real value chosen by users. The measure holds the following properties. For any $\lambda > -1$:

(i) The measure $\tau_M^{(\lambda)}$ must lie between 0 and 1

(ii) $\tau_M^{(\lambda)} = 0$ if and only if the probability table has the structure of CPMH, that is, $G_{1(i)} = G_{2(i)}$ for at least one i

(iii) $\tau_M^{(\lambda)} = 1$ if and only if the probability table has the structure of complete marginal inhomogeneity in the sense that $G_{1(i)} = 0$ (then $G_{2(i)} \neq 0$) or $G_{2(i)} = 0$ (then $G_{1(i)} \neq 0$) for $i = 1,\cdots, r-1$.

It should be noted that the measure $\tau_M^{(\lambda)}$ is expressed as the weighted geometric mean of the diversity index whereas $\Gamma_M^{(\lambda)}$ is the weighted arithmetic mean.

Approximate Confidence Interval

In this section, n_{ij} denotes the observed frequency in the ith row and jth column of the table ($i = 1,\cdots, r$; $j = 1,\cdots, r$) and $n = \sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}$. Let the $r \times r$ table $\{n_{ij}\}$ be distributed as the multinomial distribution with probabilities $\{p_{ij}\}$. It is a common occurrence that the probabilities $\{p_{ij}\}$ are unknown. Therefore, we need to estimate $\{p_{ij}\}$ and derive the approximate standard error and the large-sample confidence interval of $\tau_M^{(\lambda)}$. The sample version of $\tau_M^{(\lambda)}$, denoted by $\hat{\tau}_M^{(\lambda)}$, is given by $\tau_M^{(\lambda)}$ with p_{ij} replaced by $\hat{p}_{ij}$ where $\hat{p}_{ij} = n_{ij}/n$. Using the delta method (Agresti, 2003), $\sqrt{n}\left(\hat{\tau}_M^{(\lambda)} - \tau_M^{(\lambda)}\right)$ asymptotically has (as $n \to \infty$) a normal distribution with mean zero and variance σ^2, where:

$$\sigma^2 = \left(\frac{\tau_M^{(\lambda)}}{\Delta}\right)^2 \left\{ \sum_{i=1}^{r}\sum_{\substack{j=1\\ i\neq j}}^{r} p_{ij}\left(\beta_{ij}^{(\lambda)}\right)^2 - \left(\sum_{i=1}^{r}\sum_{\substack{j=1\\ i\neq j}}^{r} p_{ij}\beta_{ij}^{(\lambda)}\right)^2 \right\},$$

with, $\beta_{ij}^{(0)} = \lim_{\lambda\to 0} \beta_{ij}^{(\lambda)}$ and:

$$A_{12}(t) = \frac{2^{\lambda}(\lambda+1)G_{2(t)}^{c}}{\left(2^{\lambda}-1\right)\omega_t^{(\lambda)}}\left(\left(G_{1(t)}^{c}\right)^{\lambda} - \left(G_{2(t)}^{c}\right)^{\lambda}\right) + \log\omega_t^{(\lambda)},$$

$$A_{21}(t) = \frac{2^{\lambda}(\lambda+1)G_{1(t)}^{c}}{\left(2^{\lambda}-1\right)\omega_t^{(\lambda)}}\left(\left(G_{2(t)}^{c}\right)^{\lambda} - \left(G_{1(t)}^{c}\right)^{\lambda}\right) + \log\omega_t^{(\lambda)},$$

$$\beta_{ij}^{(\lambda)} = \begin{cases} \sum_{t=i}^{j-1} A_{12}(t) - (j-i)\log\tau_M^{(\lambda)} \; (i<1) \\ \sum_{t=j}^{i-1} A_{21}(t) - (i-j)\log\tau_M^{(\lambda)} \; (i>1) \end{cases},$$

for $\lambda > -1$. It should be noted that the asymptotic normal distribution of $\sqrt{n}\left(\hat{\tau}_M^{(\lambda)} - \tau_M^{(\lambda)}\right)$ is applicable only when $0 < \tau_M^{(\lambda)} < 1$ because $\sigma^2 = 0$ when $\tau_M^{(\lambda)} = 0$ and $\tau_M^{(\lambda)} = 1$. Let $\hat{\sigma}^2$ denote σ^2 with p_{ij} replaced by $\hat{p}_{ij}$. Then $\hat{\sigma}/\sqrt{n}$ is the estimated approximate standard error of $\hat{\tau}_M^{(\lambda)}$ and the approximate 100(1-α)% confidence interval of $\tau_M^{(\lambda)}$ is given by $\hat{\tau}_M^{(\lambda)} \pm z_{\alpha/2}\hat{\sigma}/\sqrt{n}$, where z_α is the 1- α quantile of the standard normal distribution.

Numerical Studies

This section presents the results of applying the model and the measure to some examples and real data.

Artificial Examples

Consider the 4×4 artificial cell and cumulative probability tables given in Tables 3 and 4. Table 3a has a structure of CPMH with $G_{1(1)} = G_{2(1)} = 0.030$. For each Tables 4b-4e, the values of $G_{1(i)}$ and $G_{2(i)}$ for i = 2 and 3 equal the corresponding values of $G_{1(i)}$ and $G_{2(i)}$. The ratio $G_{1(1)}/G_{2(1)}$ varies for Tables 4a to 4e: 1.0 for Table 4a, 2.0 for Table 4b, 3.0 for Table 4c, 4.0 for Table 4d and 5.0 for Table 4e. Thus it is natural to consider that the degree of departure from the CPMH model increases in the order of Tables 4a to 4e. Table 4f shows complete asymmetry in the sense that cell probabilities in the upper right triangle are all zeros.

Table 5 gives the values of measure $\tau_M^{(\lambda)}$ applied to each of Tables 4a to 4f. We see that (i) the value of $\tau_M^{(\lambda)}$ for Table 4a equals zero, (ii) for any fixed λ, the value of $\tau_M^{(\lambda)}$ increases as the ratio $G_{1(1)}/G_{2(1)}$ increases and (iii) the value of $\tau_M^{(\lambda)}$ for Table 4f equals 1. Therefore the measure $\tau_M^{(\lambda)}$ would be appropriate to represent the degree of departure from the CPMH model.

Real Data

Consider the data in Tables 6a and 6b taken from (Tominaga, 1979). These data describe the cross-classifications of father's and his son's occupational status categories in Japan, which were examined in 1955 and in 1965, respectively. We regard as the smaller category number means the higher status herein. We are interested in whether there is the structure of CPMH in each table. Table 7 gives the estimated values of the measure $\tau_M^{(\lambda)}$ applied to Tables 6a and 6b.

Table 3: Artificial cell probabilities
(a)

	(1)	(2)	(3)	(4)	Total
(1)	0.160	0.017	0.008	0.005	0.190
(2)	0.002	0.160	0.012	0.005	0.179
(3)	0.022	0.220	0.159	0.020	0.421
(4)	0.006	0.023	0.021	0.160	0.210
Total	0.190	0.420	0.200	0.190	1.000

(b)

	(1)	(2)	(3)	(4)	Total
(1)	0.152	0.047	0.008	0.005	0.212
(2)	0.002	0.152	0.012	0.005	0.171
(3)	0.022	0.220	0.152	0.020	0.414
(4)	0.006	0.023	0.021	0.153	0.203
Total	0.182	0.442	0.193	0.183	1.000

(c)

	(1)	(2)	(3)	(4)	Total
(1)	0.145	0.077	0.008	0.005	0.235
(2)	0.002	0.145	0.012	0.005	0.164
(3)	0.022	0.220	0.145	0.020	0.407
(4)	0.006	0.023	0.021	0.144	0.194
Total	0.175	0.465	0.186	0.174	1.000

(d)

	(1)	(2)	(3)	(4)	Total
(1)	0.137	0.107	0.008	0.005	0.257
(2)	0.002	0.137	0.012	0.005	0.156
(3)	0.022	0.220	0.137	0.020	0.399
(4)	0.006	0.023	0.021	0.138	0.188
Total	0.167	0.487	0.178	0.168	1.000

(e)

	(1)	(2)	(3)	(4)	Total
(1)	0.130	0.137	0.008	0.005	0.280
(2)	0.002	0.130	0.012	0.005	0.149
(3)	0.022	0.220	0.130	0.020	0.392
(4)	0.006	0.023	0.021	0.129	0.179
Total	0.160	0.510	0.171	0.159	1.000

(f)

	(1)	(2)	(3)	(4)	Total
(1)	0.177	0	0	0	0.177
(2)	0.002	0.177	0	0	0.179
(3)	0.022	0.220	0.176	0	0.418
(4)	0.006	0.023	0.021	0.176	0.226
Total	0.207	0.420	0.197	0.176	1.000

Table 4: Cumulative probabilities calculated from Tables 3a-f

(a) For Table 3a

	$i = 1$	2	3
$G_{1(i)}$	0.030	0.030	0.030
$G_{2(i)}$	0.030	0.271	0.050

(b) For Table 3b

	$i = 1$	2	3
$G_{1(i)}$	0.060	0.030	0.030
$G_{2(i)}$	0.030	0.271	0.050

(c) For Table 3c

	$i = 1$	2	3
$G_{1(i)}$	0.090	0.030	0.030
$G_{2(i)}$	0.030	0.271	0.050

(d) For Table 3d

	$i = 1$	2	3
$G_{1(i)}$	0.120	0.030	0.030
$G_{2(i)}$	0.030	0.271	0.050

(e) For Table 3e

	$i = 1$	2	3
$G_{1(i)}$	0.150	0.030	0.030
$G_{2(i)}$	0.030	0.271	0.050

(f) For Table 3f

	$i = 1$	2	3
$G_{1(i)}$	0	0	0
$G_{2(i)}$	0.030	0.271	0.050

Table 5: Values of $\tau_M^{(\lambda)}$ for Tables 3a-f values of $\tau_M^{(\lambda)}$

	Values of λ		
Applied tables	-0.5	0	1.0
Table 3a	0	0	0
Table 3b	0.159	0.245	0.309
Table 3c	0.182	0.280	0.353
Table 3d	0.199	0.306	0.384
Table 3e	0.215	0.328	0.409
Table 3f	1	1	1

Table 6: Cross-classifications of Japanese father's and his son's occupational status (a) in 1955 and (b) in 1965 (Tominaga (1979))
(a) in 1955

Father's status	Son's status (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total
(1)	36	4	14	7	8	2	3	8	82
(2)	20	20	27	24	11	11	2	11	126
(3)	9	6	23	12	9	5	3	16	83
(4)	15	14	39	81	17	16	11	15	208
(5)	6	7	22	13	72	20	6	13	159
(6)	3	2	5	12	18	19	9	7	75
(7)	5	3	10	11	21	15	38	25	128
(8)	39	30	76	80	69	52	45	614	1005
Total	133	86	216	240	225	140	117	709	1866

(b) in 1965

Father's status	Son's status (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total
(1)	27	10	16	3	6	6	1	2	71
(2)	15	38	30	20	8	4	3	7	125
(3)	13	17	32	17	7	16	6	5	113
(4)	12	36	40	132	22	30	13	6	291
(5)	8	22	38	41	91	42	22	9	273
(6)	2	2	7	12	13	16	3	2	57
(7)	3	2	11	11	13	26	30	6	102
(8)	38	44	95	101	132	114	60	309	893
Total	118	171	269	337	292	254	138	346	1925

Note: Status (1) is Professional; (2) Managers; (3) Clerical; (4) Sales; (5) Skilled manual; (6) Semiskilled manual; (7) Unskilled manual; and (8) Farmers

Table 7: Estimates of $\tau_M^{(\lambda)}$, estimated approximate Standard Errors (SE) of $\tau_M^{(\lambda)}$ and approximate 95% Confidence Intervals (CI) of $\tau_M^{(\lambda)}$, applied to Tables 6a and 6b

(a) For Table 6a

λ	$\tau_M^{(\lambda)}$	SE	CI
-0.5	0.063	0.024	(0.016, 0.110)
0	0.102	0.039	(0.027, 0.178)
1	0.136	0.051	(0.036, 0.236)

(b) For Table 6b

λ	$\tau_M^{(\lambda)}$	SE	CI
-0.5	0.162	0.018	(0.127, 0.198)
0	0.253	0.027	(0.201, 0.306)
1	0.323	0.032	(0.260, 0.386)

We shall compare the values of measure which represents the degree of departure from the CPMH model for Tables 6a and 6b. From Table 7, the values in the confidence interval of $\tau_M^{(\lambda)}$ are greater for Table 6b than for Table 6a. Therefore, it is inferred that the degree of departure from the CPMH model for father-son pairs is larger in 1965 than in 1955.

Concluding Remarks

For an $r \times r$ square contingency table with ordered categories, we have proposed the CPMH model which has weaker restriction than that of the MH model. We also have proposed the measure to represent the degree of departure from the CPMH model. The proposed measure $\tau_M^{(\lambda)}$ indicates how far the marginal distributions are distant from those with the structure of CPMH. The measure would be useful for comparing the degrees of departure from the CPMH model among the several tables (as seen in Section 4). The measure $\tau_M^{(\lambda)}$ is appropriate for the square table with ordered categories because the value of $\tau_M^{(\lambda)}$ is not invariant under arbitrary same permutation of the row and column categories.

Acknowledgement

The authors would like to thank the two reviewers for their valuable comments and suggestions to improve the quality of this article.

Author's Contributions

All authors contributed to the all of the study and the writing of this manuscript.

Ethics

This paper is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Agresti, A. (2003). Categorical data analysis (Vol. 482). John Wiley & Sons.

Patil, G. P., & Taillie, C. (1982). Diversity as a concept and its measurement. Journal of the American statistical Association, 77(379), 548-561.

Saigusa, Y., Kubo, Y., Tahata, K., & Tomizawa, S. (2020). A measure of departure from partial marginal homogeneity for square contingency tables. Journal of Statistics Applications and Probability Letters, 7(1), 1–7.

Saigusa, Y., Tahata, K., & Tomizawa, S. (2016). Measure of departure from partial symmetry for square contingency tables. Journal of Mathematics and Statistics, 12(3), 152-156.

Saigusa, Y., Takami, M., Ishii, A., Nakagawa, T., & Tomizawa, S. (2019). Measure for departure from cumulative partial symmetry for square contingency tables with ordered categories. Journal of Statistics: Advances in Theory and Applications, 21(1), 53–70.

Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. Biometrika, 42(3/4), 412-416.

Tominaga, K. (1979). Nippon no kaisou kouzou (Japanese hierarchical structure). University of Tokyo press, Tokyo (in Japanese).

Tomizawa, S. (2001). Generalized measures of departure from marginal homogeneity for contingency tables with nominal categories. Journal of Statistical Research, 35, 1-24.

Tomizawa, S., Miyamoto, N., & Ashihara, N. (2003). Measure of departure from marginal homogeneity for square contingency tables having ordered categories. Behaviormetrika, 30(2), 173-193.