DEPARTURE FROM QUASI-UNIFORM ASSOCIATION MODEL IN ORDERED SQUARE CONTINGENCY TABLES

Serpil AKTAS ALTUNAY*, Department of Statistics, Faculty of Science, Hacettepe University, Turkey, spxl@hacettepe.edu.tr
(https://orcid.org/0000-0003-3364-6388)
Ayfer Ezgi YILMAZ, Department of Statistics, Faculty of Science, Hacettepe University, Turkey, ezgiyilmaz@hacettepe.edu.tr
(https://orcid.org/0000-0002-6214-8014)

Abstract

The Uniform Association (UA) and Quasi-Uniform Association (QUA) models are used for analysis of two-way square contingency tables with ordered categories. When the QUA model does not fit the data, one wants to measure the degree of departure from the quasi-uniform association model. The degree of the departure from the model is calculated by using the power divergence statistics. In this study, as an alternative to measures in the literature, a measure to estimate the degree of departure from the QUA model is suggested. This suggested measure allows us to compare the several R×R tables in terms of the departure from the model. The measures are compared by the simulation study and discussed by three unaided distance vision data.
Keywords: Ordinal categories, Square contingency tables, Uniform association model, Quasi-Uniform association model, Kullback-Leiber divergence

SIRALANABİLİR KARESEL OLUMSALLIK TABLOLARINDA YARI TEKDÜZE İLİŞKİ MODELİNDEN SAPMA

Özet

Tekdüze İlişki ve Yarı Tekdüze İlişki modelleri, sıralanabilir düzeylere sahip iki yönlü karesel olumsallık tablolarının çözümlemesinde kullanılır. Yarı tekdüze ilişki modeline uyum sağlanmadığı durumda modelden ayrılışın derecesi ölçülmek istenebilir. Modelden ayrılış miktarları güç sapma istatistikleri kullanılarak hesaplanır. Bu çalışmada, literatürde yer alan ölçme kriterlerine alternatif olarak, yarı tekdüze ilişki modelinden ayrılış derecesini tahmin etmek amacıyla, yeni bir ölçme kriteri önerilmiştir. Önerilen ölçme kriteri, farklı R×R tabloların modelden sapma miktarlarını karşılaştırma imkanı da sağlamaktadır. Ölçme kriterleri benzetim çalışması karşılaştırılmış ve üç yardımsız uzak görüş verisi ile karşılaştırılmıştır.
Anahtar Kelimeler: Sıralanabilir düzeyler, Karesel olumsallık tabloları, Tekdüze ilişki modeli, Yarı tekdüze ilişki modeli, Kullback-Leiber sapması

Cite

1. Association Models

The contingency tables which have the same number of row and column categories are called square contingency tables or $R×R$ tables [1,2]. The association models exhibit the association between row and column variables. Let X be the first variable and Y be the second variable of an $R×R$ table and n denotes the sample size. n_{ij} denotes the cell frequencies, and μ_{ij} denotes the expected frequencies ($i = 1, …, R$; $j = 1, …, R$).

For the $R \times R$ table, let p_{ij} denotes the cell probabilities, then the QUA model is defined as,

$$p_{ij} = \begin{cases} \mu\alpha_i\beta_j\Theta^{ij} & i \neq j \\ \psi_{ij} & i = j \end{cases}, \qquad i,j = 1,2,…,R \qquad (1)$$

when $\Theta = 1$, this model refers to the quasi-independence (QI) model, and the cells on the main diagonal are ignored in the QUA model. The QUA model has $R(R - 3)$ degrees of freedom [3].

$\theta_{(i<j;s<t)}$ denotes the odds ratio calculated from the expected frequencies of the QUA model.

$$\theta_{(i<j;s<t)} = \frac{\mu_{is}\mu_{jt}}{\mu_{js}\mu_{it}}.$$

The QUA model can be also expressed as,

$$log\ \theta_{(i<j;s<t)} = (j - i)(t - s)\phi,$$

where $1 \leq i < j \leq R$, $1 \leq s < t \leq R$, $i \neq s$, $i \neq t$, $j \neq s$, and $j \neq t$ [4]. When the QUA model does not fit the data, the degree of departure from the QUA model may be estimated.

2. Departure from the QUA Model

The power divergence of two probability distributions is [5]

$$I^\lambda(p_{ij}, p_{ij}^*) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^{R-1} \sum_{j=1}^{R-1} p_{ij} \left[\left(\frac{P_{ij}}{p_{ij}^*} \right)^\lambda - 1 \right], \quad (2)$$

where λ is an arbitrary value chosen by the user $(-\infty < \lambda < \infty)$.

A degree of departure from symmetry (Φ^λ) is [6,7]

$$\Phi^\lambda = \frac{\lambda(\lambda+1)}{2(2^\lambda-1)} I^\lambda(p_{ij}^*, p_{ij}^s). \quad (3)$$

where

$$I^\lambda(p_{ij}^*, p_{ij}^s) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^{R} \sum_{\substack{j=1 \\ i \neq j}}^{R} p_{ij}^* \left[\left(\frac{p_{ij}^*}{p_{ij}^s} \right)^\lambda - 1 \right], \quad (4)$$

$$p_{ij}^* = \frac{p_{ij}}{\delta} \quad (i \neq j), \ \delta = \sum_{i=1}^{R} \sum_{\substack{j=1 \\ i \neq j}}^{R} p_{ij}, \text{ and}$$

$$p_{ij}^s = \frac{(p_{ij}^* + p_{ji}^*)}{2} \quad (i \neq j).$$

Here, $I^\lambda(p_{ij}^*, p_{ij}^s)$ is the power divergence between p_{ij}^* and p_{ij}^s, given in Equation (2).

Similarly, Miyamoto et al. [8] suggested a measure to estimate the degree of departure from diagonals-parameter symmetry.

The Jensen–Shannon divergence (JSD) based on Kullback–Leibler divergence [9-12] is used to measure the similarity between two probability distributions.

Let $K(P, Q)$ be Kullback–Leibler divergence, then the Jensen–Shannon divergence is written in terms of Kullback-Leibler divergence is

$$JSD = \frac{1}{2} [K(P,M) + K(Q,M)], \quad (5)$$

where M is the midpoint, $M = \frac{1}{2}(P + Q)$.

Jeffries-Matusita distance is (J-M)

$$J - M = \sqrt{\sum_{i=1}^{R} \sum_{j=1}^{R} \left(\sqrt{p_{ij}} - \sqrt{p_{ij}^*} \right)^2}. \quad (6)$$

J-M distance is more reliable than Euclidean distance in terms of classification and unlike the Euclidean distance, the J-M distance can be generalized to any distribution [13,14].

Yamamoto et al. [15] suggested a measure to estimate the degree of departure from symmetry with the assumption of $(p_{ij} + p_{ji} > 0)$ where $i \neq j$. The measure is

$$J - M1 = \sqrt{\frac{2+\sqrt{2}}{2} \sum_{i=1}^{R} \sum_{j=1}^{R} \left(\sqrt{p_{ij}^*} - \sqrt{p_{ij}^s} \right)^2}. \quad (7)$$

We shall generalize Matusita distance for a measure to estimate the degree of departure from the QUA model in Section 3.

3. Proposed Measure for Departure from the QUA Model

In this section, we suggested a modification for Matisuta measure to estimate the degree of departure from the QUA in ordered square contingency tables. Modified Matisuta can be defined as

$$MM = \left[-\frac{\sqrt{3}}{2} \sqrt{\sum_{i=1}^{R} \sum_{j=1}^{R} \left(\sqrt{p_{ij}} - \sqrt{p_{ij}^*} \right)^2} + \frac{\sum_{m=1}^{M} \lambda_m^2}{2(R+2)} \right], \quad (8)$$

where, $\sum_{m=1}^{R-1} \lambda_m^2$ is the total inertia and λ_m is the eigenvalue of each dimension. The total inertia of a table is a measure of how much variation there is in the table [16]. Thus adding the total variation to Matusita distance could be an alternative distance.

MM always lies between 0 and 1. If the $MM \cong 0$ this means that the degree of departure from the QUA model is the smallest.

4. Simulation Study

A simulation study is implemented to compare the four methods including JSD-Midpoint, JSD, J-M, and MM. We generate random data from the multinomial distribution with marginal probabilities [0.25, 0.15, 0.30, 0.30] for R=4 with sample size n=500; and [0.20, 0.15, 0.25, 0.25, 0.15] for R=5 with sample size n = 100. The number of replication is assigned as 1000 times.

Table 1 gives the mean values of the likelihood ratio chi-squared statistic for simulated tables. The QUA model fits the data well in both scenerios. Therefore, we expect the degree of departure of QUA is a very small number, almost around zero.

Table 1. Mean likelihood ratio statistics for simulated tables.

Scenario	df	Likelihood Ratio	*p*-value
R=4, n=500	4	6.037	>0.05
R=5, n=100	10	9.217	>0.05

The degree of departure from QUA is estimated as a very small number of the maximum degree of departure from QUA (see Table 2). The smallest value is calculated for the MM. This indicates that the QUA model fits data well in accordance with the results in Table 1.

Table 2. Mean degree of departure of QUA values for simulated tables.

Methods	Value for R=4, n=500	Value for R=5, n=100
JSD-Midpoint	0.00199	0.00179
JSD	0.00397	0.00364
J-M	0.01746	0.01556
MM	0.00013	0.00011

Using the mesures of departure from the QUA model, we calculated the degree of departure from the QUA values assigning different λ values (see Table 3).

Table 3. Mean degree of departure of QUA values for various λ values for simulated tables.

Values of λ	Estimated measure for R=4, n=500	Estimated measure for R=5, n=100
-0.5	0.00109	0.00120
0.0	0.00414	0.00376
0.5	0.00612	0.00511
1.0	0.00724	0.00673
1.5	0.00773	0.00692
2.0	0.00773	0.00654
2.5	0.00741	0.00657
3.0	0.00688	0.00701

We can say that, all λ values are close to zero and this means that the QUA model fits data well. The degree of departure from the QUA model for λ=-0.5 gives the smallest values.

5. Numerical Examples

Three real-life datasets are utilized to illustrate the performance of the measures. The unaided distance vision data [17] in Table 4 is the cross-tabulation of the right eye and left eye's visions of 7477 women aged 30 to 39 employed in Royal Ordnance factories in Britain from 1943 to 1946. The visions of eyes are categorized into 4 ordered categories from the Best (1) to the Worst (4). Similarly, the data in Table 5 and Table 6 are the classifications of 3242 men in Britain [18] and 4746 students in Japan [19], respectively.

Table 4. Unaided distance vision data of 7477 women in Britain.

Right eye grade	Left eye grade				Total
	(1)	(2)	(3)	(4)	
(1) Best	1520	266	124	66	1976
(2) Second	234	1512	432	78	2256
(3) Third	117	363	1772	205	2456
(4) Worst	36	82	179	492	789
Total	1907	2222	2507	841	7477

Table 5. Unaided distance vision data of 3242 men in Britain.

Right eye grade	Left eye grade				Total
	(1)	(2)	(3)	(4)	
(1) Best	821	112	85	35	1053
(2) Second	116	494	145	27	782
(3) Third	72	151	583	87	893
(4) Worst	43	34	106	331	514
Total	1052	791	919	480	3242

Table 6. Unaided distance vision data of 4746 students in Japan.

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

The design matrix is defined as below to implement the QUA model. The entries in the nine columns correspond to the quasi-independence parameter and the last column is the uniform association parameter.

$$
\log \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{14} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{24} \\ \mu_{31} \\ \mu_{32} \\ \mu_{33} \\ \mu_{34} \\ \mu_{41} \\ \mu_{41} \\ \mu_{43} \\ \mu_{44} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 5 & 2 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 5 & 3 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 & 5 & 4 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 5 & 2 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 2 & 4 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 5 & 6 \\ 1 & 0 & 1 & 0 & -1 & -1 & -1 & 5 & 8 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 5 & 3 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 & 6 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 3 & 9 \\ 1 & 0 & 0 & 1 & -1 & -1 & -1 & 5 & 12 \\ 1 & -1 & -1 & -1 & 1 & 0 & 0 & 5 & 4 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & 5 & 8 \\ 1 & -1 & -1 & -1 & 0 & 0 & 1 & 5 & 12 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 & 4 & 16 \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda_1^X \\ \lambda_2^X \\ \lambda_3^X \\ \lambda_1^Y \\ \lambda_2^Y \\ \lambda_3^Y \\ \Theta \\ \beta \end{bmatrix}
$$

QUA model is applied to Tables 4-6 and the results are summarized in Table 7.

Table 7. Likelihood ratio statistics under the QUA model.

Tables	*df*	Likelihood Ratio Statistic
Table 4	4	61.3082*
Table 5	4	36.5191*
Table 6	4	43.3568*

*Significant at %5 level.

Expected frequencies are obtained for the QUA model for each table (see Table 8, Table 9, Table 10). QUA model does not fit any of the tables. Thus, the degree of departure from the model can be calculated.

Table 8. Expected frequencies under QUA model for Table 4.

Right eye grade	Left eye grade				Total
	(1)	(2)	(3)	(4)	
(1) Best	1520	256.747	164.297	34.956	1976
(2) Second	229.340	1512	395.542	119.118	2256
(3) Third	132.355	356.719	1772	194.926	2456
(4) Worst	25.305	96.534	175.161	492	789
Total	1907	2222	2507	841	7477

Table 9. Expected frequencies under QUA model for Table 5.

Right eye grade	Left eye grade				Total
	(1)	(2)	(3)	(4)	
(1) Best	821	110.709	97.593	23.698	1053
(2) Second	108.641	494	136.280	43.078	782
(3) Third	94.006	133.770	583	82.224	893
(4) Worst	28.353	52.521	102.127	331	514
Total	1052	791	919	480	3242

Table 10. Expected frequencies under QUA model for Table 6.

Right eye grade	Left eye grade				Total
	(1)	(2)	(3)	(4)	
(1) Best	1291	124.715	59.157	8.129	1483
(2) Second	147.213	221	105.886	32.9	507
(3) Third	73.133	110.896	660	188.971	1033
(4) Worst	12.654	43.389	237.957	1429	1723
Total	1524	500	1063	1659	4746

For the various lambda values in Equation (4), estimated measures of the degree of departure from the QUA model are illustrated in Table 11. The smallest values are obtained for $\lambda = -0.5$ and the largest for $\lambda = 2.5$ except for Table 6.

Table 11. Estimated measures under different lambda values.

Values of λ	Estimated measure		
	Table 4	Table 5	Table 6
-0.5	0.00172	0.00244	0.00189
0.0	0.00295	0.00406	0.00329
0.5	0.00378	0.00504	0.00434
1.0	0.00431	0.00555	0.00516
1.5	0.00461	0.00569	0.00584
2.0	0.00474	0.00559	0.00646
2.5	0.00475	0.00532	0.00707
3.0	0.00466	0.00495	0.00773

The measure of departure from the QUA model based on Matusita distance, JDS, and the proposed measures are given in Table 12.

Table 12. Measure of departure from QUA model based on Matusita Distance, JSD, and proposed measure.

Tables	J-M	JSD	JDS-midpoint	MM
Table 4	0.04496	0.06341	0.07735	0.05135
Table 5	0.05352	0.05994	0.07385	0.04392
Table 6	0.04707	0.06244	0.07647	0.04946

The the smallest value based on Matusita distance is obtained for Table 4. Hence, we may say that the largest degree of departure according to the MM measure from the QUA model is for Table 4.

Unlike the Matusita Distance, JDS gives the smallest values for Table 5 and Table 4 has the maximum degree of departure from QUA.

More accurate results occur for the proposed measure. The smallest MM value is obtained for Table 5 and the greatest value is obtained for Table 4. The degree of departure from QUA is estimated to be 5.1 percent of the maximum degree of departure from QUA for Table 4, 4.4 percent for Table 5 and, 4.9 percent for Table 6, respectively.

6. Concluding Remarks

The Uniform Association (UA) and Quasi-Uniform Association (QUA) models are the log-linear models used for the analysis of ordered square contingency tables. When the models do not fit the data, the researchers wish to know how far the data is from the underlying model. In the literature, there are different measures to estimate the degree of departure. In this study, we proposed a new measure to estimate the degree of departure from the QUA model in ordered square contingency tables by considering the measure of association.

By using these measures, it is also possible to compare different models or data sets in terms of their departure from the QUA model. Departure measures range between 0 and 1, and their values are not affected by the number of categories or sample size. Yamamoto et. al. [17] pointed out that these measures are useful when the degree the departure from an underlying model is wanted to be calculated, whereas the test statistic can not. It is seen that the proposed MM measure gives more accurate results. The simulation study results and the real data examples showed that MM measure is very useful us to measure the degree of departure from the QUA model. The limitation is that it cannot be used for R<4.

We used three well-known data sets in the literature. The analysis results showed that, when the degrees of departure from QUA in Tables 4-6 are compared, it is clear that the degree of departure in Table 4 is greater than that in Table 5 and Table 6 based on the proposed measure.

7. References

[1] Altun, G., "Quasi Local Odds Symmetry Model for Square Contingency Table with Ordinal Categories", J*ournal of Statistical Computation and Simulation*, 89(15), 2899-2913, 2019.

[2] Altun, G. and Aktaş, S., "Measures of departure from marginal homogeneity model in square contingency tables", *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 11(2), 93-108, 2018.

[3] Goodman, L.A., "Simple Models for the Analysis of Association in Cross-Classifications having Ordered Categories", *Journal of the American Statistical Association*, 74, 537-552, 1979.

[4] Tomizawa, S., "The Decomposition of a Quasi-uniform Association Model and Its Application to the Analysis of British Occupational Mobility Table Data", *Behaviormetrika*, 17(27), 115-121, 1990.

[5] Kullback, S., *Information Theory and Statistics*, Dover Publications, New York, 1996.

[6] Tomizawa, S., "Two Kinds of Measures of Departure from Symmetry in Square Contingency Tables having Nominal Categories", *Statistica Sinica*, 4, 325-334, 1994.

[7] Tomizawa, S., Seo, T. and Yamamoto, H., "Power-divergence-type Measure of Departure from Symmetry for Square Contingency Tables that have Nominal Categories", *Journal of Applied Statistics*, 25(3), 387-398, 1998.

[8] Miyamoto, N., Kato, T. and Tomizawa, S., "A Measure of Departure from Diagonals-parameter Symmetry based on Association Measure for Square Contingency Tables", *Journal of Statistics: Advances in Theory and Application*, 3(1), 1-15, 2010.

[9] Cressie, N.A.C. and Read, T.R.C., "Multinomial Goodness-of-fit Tests", *Journal of the Royal Statistical Society Series*, B46, 440–464, 1984.

[10] Patil, G.P. and Taillie, C., "Diversity as a Concept and its Measurement", J*ournal of the American Statistical Association*, 77, 548-561, 1982.

[11] Read, T.R.C. and Cressie, N., *Goodness-of-fit Statistics for Discrete Multivariate Data*, Springer, New York, 1988.

[12] Sgarro, A., "Informational Divergence and the Dissimilarity of Probability Distributions", *Calcolo*, 18(3), 293–302, 1981.

[13] Matusita, K., "On the Estimation by the Minimum Distance Method", *Annals of the Institute of Statistical Mathematics*, 5, 59-65, 1954.

[14] Matusita, K., "Decision Rules, based on the Distance, for Problems of Fit, Two Samples, and Estimation", *The Annals of Mathematical Statistics*, 26(4), 631-640, 1955.

[15] Yamamoto, K., Miyamoto, N., Tsuboi, H. and Tomizawa, S., "Distance Measure of Departure from Symmetry for Square Contingency Tables with Nominal Categories", *International Journal of Pure and Applied Mathematics*, 48:4, 483-489, 2008.

[16] Beh, E.J., "Simple Correspondence Analysis: A Bibliographic Review", *International Statistical Review*, 72(2), 257-284, 2004.

[17] Stuart, A., "A Test for Homogeneity of the Marginal Distributions in a Two-way Classification", *Biometrika*, 42, 412-416, 1955.

[18] Stuart, A., "The Estimation and Comparison of Strengths of Association in Contingency Tables", *Biometrika*, 40, 105-110, 1953.

[19] Tomizawa, S., "Three Kinds of Decompositions for the Conditional Symmetry Model in a Square Contingency Table", *Journal of the Japan Statistical Society*, 14(1), 35-42, 1984.