Journal of Statistics: Advances in Theory and Applications
Volume 15, Number 2, 2016, Pages 71-84
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/jsata_7100121663

SYMMETRY MODEL BASED ON BIVARIATE NORMAL DISTRIBUTION FOR SQUARE CONTINGENCY TABLES WITH ORDERED CATEGORIES

KOUJI YAMAMOTO[1], HIROKI NAKANE[2] and SADAO TOMIZAWA[2]

[1]Department of Clinical Epidemiology and Biostatistics
Graduate School of Medicine
Osaka University
2-2, Yamadaoka, Suita
Osaka, 565-0871
Japan
e-mail: yamamoto-k@stat.med.osaka-u.ac.jp

[2]Department of Information Sciences
Faculty of Science and Technology
Tokyo University of Science
Noda City, Chiba, 278-8510
Japan
e-mail: nhiroki1989@gmail.com
tomizawa@is.noda.tus.ac.jp

Abstract

For square contingency tables with ordered categories, this paper proposes a model that the cell probabilities have a similar structure of bivariate normal density function with equal marginal means and marginal variances. The proposed model has the structure of symmetry, quasi-uniform association, and

2010 Mathematics Subject Classification: 62H17.
Keywords and phrases: bivariate normal distribution, linear diagonals-parameter symmetry model, normal distribution type symmetry model, symmetry model, symmetry plus quasi-uniform association model.
Received April 29, 2016

normal distribution type symmetry models. The simulation studies based on the bivariate normal distribution are given. An example is also given.

1. Introduction

Consider an $r \times r$ square contingency table with the same ordinal 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 $(i = 1, \ldots, r;\ j = 1, \ldots, r)$. Bowker [4] proposed the symmetry (S) model defined by

$$p_{ij} = \psi_{ij} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r),$$

where $\psi_{ij} = \psi_{ji}$ (also see Bishop et al. [3], p. 282). Agresti [1] proposed the linear diagonals-parameter symmetry (LDPS) model defined by

$$p_{ij} = \begin{cases} \theta^{j-i}\psi_{ij} & (i < j), \\ \psi_{ij} & (i \geq j), \end{cases}$$

where $\psi_{ij} = \psi_{ji}$. This indicates that the probability that an observation will fall in the (i, j)-th cell, $i < j$, is θ^{j-i} times higher than the probability that it falls in the (j, i)-th cell. A special case of the LDPS model obtained by putting $\theta = 1$ is the S model.

Let the random vector $\boldsymbol{X} = (X_1, X_2)$ be distributed according to the bivariate normal distribution with means $E(X_1) = \mu_1$, $E(X_2) = \mu_2$, variances $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \sigma^2$, and correlation $\mathrm{Corr}(X_1, X_2) = \rho$, which have the density function

$$f(x_1, x_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sigma^2(1-\rho^2)}\left((x_1-\mu_1)^2 - 2\rho(x_1-\mu_1)(x_2-\mu_2) + (x_2-\mu_2)^2\right)\right]. \quad (1)$$

Then we see

$$\frac{f(x_1, x_2)}{f(x_2, x_1)} = \exp\left(\frac{(x_2 - x_1)(\mu_2 - \mu_1)}{(1-\rho)\sigma^2}\right).$$

Agresti [1] pointed out that the $f(x_1, x_2)/f(x_2, x_1)$ has the form $\theta^{x_2 - x_1}$ for some constant θ, and hence the LDPS model may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal variances.

Yamamoto and Tomizawa [7] proposed the symmetry plus quasi-uniform association (SQU) model defined by

$$p_{ij} = \begin{cases} \mu\delta_i\delta_j\theta^{ij} & (i \neq j), \\ \psi_{ii} & (i = j). \end{cases}$$

Denote the odds ratio for rows i and $j\,(>i)$ and columns s and $t\,(>s)$ by $\theta_{(ij;st)}$; thus, $\theta_{(ij;st)} = (p_{is}p_{jt})/(p_{js}p_{it})$. Using odds ratios, under the SQU model we see

$$\theta_{(ij;st)} = \theta^{(j-i)(t-s)} \quad (i \neq s,\ i \neq t,\ j \neq s,\ j \neq t).$$

This model is a special case of the S model obtained by putting $\psi_{ij} = \delta_i\delta_j\theta^{ij}$ for $i \neq j$.

By the way, the density function (1) further can be expressed as

$$f(x_1, x_2) = c a_1^{(x_1 - x_2)^2} a_2^{x_1 - x_2} b_1^{(x_1 + x_2)^2} b_2^{x_1 + x_2}, \tag{2}$$

where

$$c = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{(\mu_1 - \mu_2)^2}{4\sigma^2(1-\rho)} - \frac{(\mu_1 + \mu_2)^2}{4\sigma^2(1+\rho)}\right),$$

$$a_1 = \exp\left(-\frac{1}{4\sigma^2(1-\rho)}\right),$$

$$a_2 = \exp\left(\frac{\mu_1 - \mu_2}{2\sigma^2(1-\rho)}\right),$$

$$b_1 = \exp\left(-\frac{1}{4\sigma^2(1+\rho)}\right),$$

$$b_2 = \exp\left(\frac{\mu_1 + \mu_2}{2\sigma^2(1+\rho)}\right).$$

So, Tahata et al. [6] proposed the normal distribution type symmetry (NDS) model defined by

$$p_{ij} = \xi\alpha_1^{(i-j)^2}\alpha_2^{i-j}\beta_1^{(i+j)^2}\beta_2^{i+j} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r).$$

This model is a special case of the LDPS model. Tahata et al. [6] pointed out that the $\{p_{ij}\}$ has a similar structure of bivariate normal density function with equal marginal variances, and hence the NDS model may also be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal variances.

Now, if $a_2 = 1$ (i.e., $E(X_1) = E(X_2) = \mu$) for Equation (2), the density function $f(x_1, x_2)$ further can be expressed as

$$f(x_1, x_2) = vs^{x_1^2 + x_2^2}t^{x_1 + x_2}u^{x_1 x_2}, \tag{3}$$

where

$$v = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}}\exp\left(-\frac{\mu^2}{\sigma^2(1+\rho)}\right),$$

$$s = \exp\left(-\frac{1}{2\sigma^2(1-\rho^2)}\right),$$

$$t = \exp\left(\frac{\mu}{\sigma^2(1+\rho)}\right),$$

$$u = \exp\left(\frac{\rho}{\sigma^2(1-\rho^2)}\right).$$

We are now interested in considering a model such that the cell probabilities $\{p_{ij}\}$ have a similar structure of bivariate normal density function with the form of Equation (3).

The purpose of this paper is to propose new models which may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal means and marginal variances. Section 2 describes the new models, Section 3 describes goodness-of-fit test, Section 4 shows some numerical simulations, and Section 5 analyzes the cataracts data using the proposed model.

2. Restricted Normal Distribution Type Symmetry Model

For an $r \times r$ square table, consider a new model defined by

$$p_{ij} = \mu\alpha^{i^2+j^2}\beta^{i+j}\gamma^{ij} \quad (i = 1, \ldots, r;\ j = 1, \ldots, r). \tag{4}$$

This model is a special case of the NDS model obtained by putting $\alpha_2 = 1$. So, we shall refer to model (4) as the restricted normal distribution type symmetry (RNDS) model.

Under the RNDS model, we see

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

and

$$\theta_{(ij;\,st)} = \gamma^{(j-i)(t-s)} \quad (i < j,\ s < t).$$

Therefore, the RNDS model has the structure of the S model and uniform association.

Next, consider the model of quasi RNDS for off-diagonal cells as follows:

$$p_{ij} = \begin{cases} \mu\alpha^{i^2+j^2}\beta^{i+j}\gamma^{ij} & (i \neq j), \\ \psi_{ii} & (i = j). \end{cases}$$

We shall refer to this model as QRNDS model. Note that the RNDS model implies the QRNDS model. Under the QRNDS model, we see

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

and

$$\theta_{(ij;st)} = \gamma^{(j-i)(t-s)} \quad (i \neq s,\ i \neq t,\ j \neq s,\ j \neq t).$$

The QRNDS model is a special case of the SQU model obtained by putting $\delta_i = \alpha^{i^2}\beta^i$ for $i \neq j$. In Figure 1, we show the relationships among the models when $r \geq 4$.

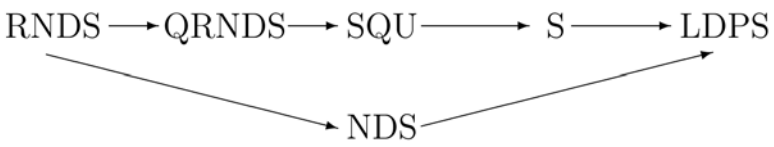

Figure 1. Relationships among the models when $r \geq 4$. Note that $M_2 \rightarrow M_1$ indicates that model M_2 implies model M_1.

The RNDS model may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal means and equal marginal variances because of Equation (3). So, the QRNDS model may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal means and equal marginal variances. We investigate this property in Section 4 in terms of the simulation studies.

3. Goodness-of-fit Test

Let n_{ij} denote the observed frequency in the (i, j)-th cell of the table $(i = 1, \ldots, r;\ j = 1, \ldots, r)$ with $n = \sum\sum n_{ij}$. Assume that a multinomial distribution applies to the $r \times r$ table. The maximum likelihood estimates of expected frequencies under the RNDS and the QRNDS models can be easily obtained by using an iterative procedure, for example, the general iterative procedure for log-linear model of Darroch and Ratcliff [5].

The likelihood ratio statistic for testing the goodness-of-fit of a model symbolized by M is

$$G^2(M) = 2\sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij} \log\left(\frac{n_{ij}}{\hat{m}_{ij}}\right),$$

where $\hat{m}_{ij}$ is maximum likelihood estimate of expected frequency m_{ij} under model M. The numbers of degrees of freedom (df) for the RNDS and the QRNDS models are $r^2 - 4$ and $r^2 - r - 4$, respectively.

Consider two nested models, say M_1 and M_2, such that model M_2 is a special case of model M_1, so when M_2 holds, necessarily M_1 also holds. For example, M_2 is the RNDS model and M_1 is the QRNDS model. Let v_1 and v_2 denote the numbers of df for models M_1 and M_2, respectively. Note that $v_1 < v_2$ and $G^2(M_1) \leq G^2(M_2)$. For testing the hypothesis that model M_2 holds assuming that model M_1 holds, we can use the likelihood ratio statistic $G^2(M_2|M_1)$, where $G^2(M_2|M_1) = G^2(M_2) - G^2(M_1)$. Under the null hypothesis, this test statistic has an asymptotic chi-square distribution with $v_2 - v_1$ df.

4. Simulation Studies

As described in Section 2, the RNDS and the QRNDS models may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal means and equal marginal variances. We shall now consider the simulation studies based on bivariate normal distribution. Consider the random vector $\boldsymbol{Z} = (Z_1, Z_2)$ which is distributed as a bivariate normal distribution with means $E(Z_1) = E(Z_2) = \mu$, variances $\mathrm{Var}(Z_1) = \mathrm{Var}(Z_2) = \sigma^2$, and correlation $\mathrm{Corr}(Z_1, Z_2) = \rho$. Suppose that there is an underlying bivariate normal distribution with some conditions and suppose that a 4×4 table is formed using cut points for each variable at $\mu, \mu \pm 0.6\sigma$. Then in terms of simulation studies, each subtable of Table 1 gives a 4×4 table of sample size 500 or 1000, formed from an underlying bivariate normal distribution with equal marginal means and marginal variances on some correlations.

Table 1. The 4×4 table of sample size 500 or 1000, formed by using cut points for each variable at μ, $\mu \pm 0.6\sigma$, from an underlying bivariate normal distribution with equal marginal means and marginal variances on some conditions

(a) n = 500, ρ = 0.1				(b) n = 500, ρ = 0.3			
48	31	28	36	52	31	26	28
31	32	19	29	31	28	24	26
30	27	23	33	31	20	21	31
29	27	24	53	28	35	29	59
(c) n = 500, ρ = 0.5				(d) n = 500, ρ = 0.7			
76	33	17	18	81	26	17	2
29	39	20	20	33	43	35	13
20	26	22	39	11	28	45	33
17	16	30	78	2	23	37	71
(e) n = 1000, ρ = 0.1				(f) n = 1000, ρ = 0.3			
86	66	65	69	103	59	53	45
67	62	48	54	72	57	50	47
74	48	57	48	41	55	69	60
66	58	58	74	51	62	77	99
(g) n = 1000, ρ = 0.5				(h) n = 1000, ρ = 0.7			
130	75	43	32	168	71	27	14
70	47	49	45	57	68	57	32
59	53	58	72	23	60	81	74
22	45	72	128	11	26	75	156

We see from Table 2 that all models fit well. We consider the test based on the difference between the G^2 values for the RNDS model and each model of QRNDS, SQU, S, NDS, and LDPS models. Since each of the G^2 values based on the difference is not significant at the 0.05 level, the RNDS model is more appropriate for a square contingency table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal means and equal marginal variances.

Table 2. Likelihood ratio statistic G^2 values for (a) the RNDS and the QRNDS models, and (b) the SQU, S, NDS, and LDPS models, applied to the data in Table 1

(a)

Tables	G^2(RNDS) df = 12	G^2(QRNDS) df = 8
Table 1(a)	7.81	3.75
Table 1(b)	5.71	3.37
Table 1(c)	14.62	5.30
Table 1(d)	16.59	8.90
Table 1(e)	6.91	3.34
Table 1(f)	12.05	8.77
Table 1(g)	9.11	6.20
Table 1(h)	15.34	6.46

(b)

Tables	G^2(SQU) df = 7	G^2(S) df = 6	G^2(NDS) df = 11	G^2(LDPS) df = 5
Table 1(a)	3.75	3.72	7.41	3.32
Table 1(b)	3.19	2.20	5.29	1.78
Table 1(c)	5.19	2.94	14.43	2.75
Table 1(d)	8.39	5.95	16.30	5.65
Table 1(e)	2.71	1.75	6.60	1.44
Table 1(f)	8.14	7.63	10.35	5.93
Table 1(g)	5.55	4.71	9.11	4.71
Table 1(h)	6.38	2.92	13.90	1.47

Next, we perform many simulation studies under some conditions. In detail, we count the frequencies of acceptance (at the 0.05 significance level) of the hypothesis that the RNDS model or the QRNDS model holds per 100000 times for 4×4 tables on some conditions. From Table 3, we see that the QRNDS model gives a good fit when it is reasonable to assume an underlying bivariate normal distribution with equal marginal means and equal marginal variances. However, the RNDS model tends to give a poor fit as the correlation becomes larger when the sample size is fixed.

Table 3. The frequencies of acceptance (at the 0.05 significance level) of the hypothesis that the RNDS model or the QRNDS model holds per 100000 times for 4×4 tables on some sample size n and correlation ρ

(a) For the RNDS model

n	ρ	Frequencies
(1) 500	0.1	94575
(2) 500	0.3	93496
(3) 500	0.5	89664
(4) 500	0.7	76321
(5) 1000	0.1	94749
(6) 1000	0.3	92074
(7) 1000	0.5	83146
(8) 1000	0.7	52367

(b) For the QRNDS model

n	ρ	Frequencies
(1) 500	0.1	94729
(2) 500	0.3	94375
(3) 500	0.5	94336
(4) 500	0.7	94061
(5) 1000	0.1	95014
(6) 1000	0.3	94338
(7) 1000	0.5	94209
(8) 1000	0.7	94406

5. An Example

Table 4, taken from Agresti ([2], p. 253) is the data of case-control study investigating a possible relationship between cataracts and the use of head coverings during the summer. Each case reporting to a clinic for cataract care was matched with a control of the same gender and similar age not having a cataract. The row and column categories refer to the frequency with which the subject used head coverings.

Table 4. Case-control study investigating a possible relationship between cataracts and the use of head coverings during the summer (Agresti [2], p. 253). (The parenthesized values are the maximum likelihood estimates of expected frequencies under the QRNDS model)

	Control				
Cataract Case	(1)	(2)	(3)	(4)	
(1)	29	3	3	4	39
	(29.00)	(4.56)	(4.31)	(6.35)	
(2)	5	0	1	1	7
	(4.56)	(0.00)	(1.35)	(1.43)	
(3)	9	0	2	0	11
	(4.31)	(1.35)	(2.00)	(0.50)	
(4)	7	3	1	0	11
	(6.35)	(1.43)	(0.50)	(0.00)	
Total	50	6	7	5	68

Note that (1) is almost or almost always, (2) is frequently, (3) is occasionally, and (4) is never.

From Table 5, we see that each model fits these data well except the RNDS model. Under the QRNDS model, the values of maximum likelihood estimates of parameters are $\hat{\alpha} = 1.249$, $\hat{\beta} = 0.431$, and $\hat{\gamma} = 0.719$. Under the QRNDS model, for example, the probability that using a head covering for a case in a pair is 'always or almost always', and for the control in the pair it is 'never', is estimated to equal the probability that using a head covering for a case in the pair is 'never', and for control in the pair it is 'always or almost always'. Also, for local 2×2 tables that do not contain the cells on the main diagonal, the odds that for a case in a pair using head covering is $s+1$ instead of s is estimated to be $\hat{\gamma} = 0.719$ times when for the control in the pair it is $t+1$ than when it is t. For $i < j$ and $s < t$ with $i \neq s$, $i \neq t$, $j \neq s$, $j \neq t$, the odds that for a case in a pair the using head covering is j instead of i is estimated to be $(0.719)^{(j-i)(t-s)}$ times higher when for the control in the

pair it is t than when it is s. For example, the odds that for a case in a pair the using a head covering is 'never' instead of 'frequently' is estimated to be $0.267\ [\ =(0.719)^4\ ]$ times higher when for control in the pair it is 'occasionally' than when it is 'always or almost always'. Since $\hat{\gamma} < 1$, the use of head coverings may have an effect on preventing cataract.

Table 5. Likelihood ratio statistic G^2 values of models applied to Table 4

Applied models	df	G^2
RNDS	12	21.11*
QRNDS	8	11.67
SQU	7	10.95
S	6	8.29
NDS	11	16.77
LDPS	5	3.96

*Means significant at the 0.05 level.

6. Concluding Remarks

The NDS model may be appropriate for a square table with ordered categories if it is reasonable to assume an underlying bivariate normal distribution with equal marginal variances. On the other hand, the proposed models, i.e., the RNDS and the QRNDS models may be appropriate for a square table with ordered categories if it is reasonable to assume an underlying bivariate normal distribution with equal marginal means and equal marginal variances. In addition from the simulation studies, the QRNDS model is useful for a square contingency table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal means and equal marginal variances. Moreover, the RNDS model is useful for a square contingency table if it is reasonable to assume an underlying bivariate normal distribution with equal marginal means and equal marginal variances when the correlation ρ is not large.

References

[1] A. Agresti, A simple diagonals-parameter symmetry and quasi-symmetry model, Statistics and Probability Letters 1 (1983), 313-316.

[2] A. Agresti, An Introduction to Categorical Data Analysis, New York, Wiley, 1996.

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

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

[5] J. N. Darroch and D. Ratcliff, Generalized iterative scaling for log-linear models, Annals of Mathematical Statistics 43 (1972), 1470-1480.

[6] K. Tahata, K. Yamamoto and S. Tomizawa, Normal distribution type symmetry model for square contingency tables with ordered categories, The Open Statistics and Probability Journal 1 (2009), 32-37.

[7] K. Yamamoto and S. Tomizawa, Symmetry plus quasi uniform association model and its orthogonal decomposition for square contingency tables, Journal of Modern Applied Statistical Methods 9 (2010), 255-262.

■