Austrian Journal of Statistics
June 2018, Volume 47, 39–48.
http://www.ajs.or.at/
doi:10.17713/ajs.v47i3.701

Log-normal Distribution Type Symmetry Model for Square Contingency Tables with Ordered Categories

Kiyotaka Iki
Tokyo University of Science

Sadao Tomizawa
Tokyo University of Science

Abstract

For the analysis of square contingency tables with the same row and column ordinal classifications, this article proposes a new model which indicates that the log-ratios of symmetric cell probabilities are proportional to the difference between log-row category and log-column category. The proposed model may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate log-normal distribution. Also, this article gives the decomposition of the symmetry model using the proposed model with the orthogonality of test statistics. Examples are given. The simulation studies based on bivariate log-normal distribution are given.

Keywords: bivariate log-normal distribution, orthogonal decomposition, square contingency table, symmetry.

1. 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 ($i = 1, \dots, R; j = 1, \dots, R$). The symmetry (S) model is defined by

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

where $\psi_{ij} = \psi_{ji}$; see Bowker (1948), Bishop, Fienberg, and Holland (1975, p. 282) and Agresti (2013, p. 426). This model indicates a structure of symmetry of the probabilities $\{p_{ij}\}$ with respect to the main diagonal of the table. Agresti (1983) proposed the linear diagonals-parameter symmetry (LDPS) model, defined by

$$p_{ij} = \alpha^i \beta^j \psi_{ij} \quad (i = 1, \dots, R; j = 1, \dots, R),$$

where $\psi_{ij} = \psi_{ji}$. This model is also expressed as

$$\frac{p_{ij}}{p_{ji}} = \delta^{j-i} \quad (i < j).$$

This indicates that the log-odds that an observation will fall in the (i, j)th cell instead of in the (j, i)th cell, $i < j$, is proportional to the distance $j - i$ from the main diagonal of the table. A special case of this model obtained by putting $\alpha = \beta$ is the S model.

Consider random variables (U, V) having a bivariate normal distribution with means $\mathrm{E}(U) = \mu_1$ and $\mathrm{E}(V) = \mu_2$, variances $\mathrm{Var}(U) = \mathrm{Var}(V) = \sigma^2$, and correlation $\mathrm{Corr}(U, V) = \rho$. Then the bivariate probability density function $f(u, v)$ is

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

It satisfies

$$\frac{f(u,v)}{f(v,u)} = \Delta^{v-u},$$

where

$$\Delta = \exp\left[\frac{\mu_2 - \mu_1}{\sigma^2(1-\rho)}\right].$$

Agresti (1983) described the relationship between the LDPS model and the bivariate normal distribution with equal marginal variances as follows. The $f(u, v)/f(v, u)$ has the form Δ^{v-u}, 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.

Now, we consider a bivariate log-normal distribution. We are interested in proposing a new model, which would be appropriate if it is reasonable to assume an underlying bivariate log-normal distribution.

Section 2 proposes new models. Section 3 gives the decompositions of the S model using the proposed models. Section 4 shows the orthogonality of decompositions given in Section 3. Section 5 gives examples. Section 6 simulates the relationships between the underlying log-normal distribution and proposed models. Section 7 provides some concluding remarks.

2. Log-normal distribution type symmetry models

2.1. Properties of log-normal distribution

Consider random variables (U, V) having a bivariate log-normal distribution, where $U > 0$ and $V > 0$. Define $Z_1 = \log U$ and $Z_2 = \log V$. Then Z_1 and Z_2 have a bivariate normal distribution, where means $\mathrm{E}(Z_1) = \mu_1$ and $\mathrm{E}(Z_2) = \mu_2$, variances $\mathrm{Var}(Z_1) = \sigma_1^2$ and $\mathrm{Var}(Z_2) = \sigma_2^2$, and correlation $\mathrm{Corr}(Z_1, Z_2) = \rho$. The bivariate probability density function $h(u, v)$ is

$$h(u,v) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}uv} \exp\left[-\frac{1}{2(1-\rho^2)}Q(u,v)\right],$$

where

$$Q(u,v) = \left(\frac{\log u - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\log u - \mu_1}{\sigma_1}\right)\left(\frac{\log v - \mu_2}{\sigma_2}\right) + \left(\frac{\log v - \mu_2}{\sigma_2}\right)^2.$$

It satisfies

$$\frac{h(u,v)}{h(v,u)} = \tau_1^{\log v - \log u}\tau_2^{(\log v)^2 - (\log u)^2},$$

where

$$\tau_1 = \exp\left[\frac{1}{1-\rho^2}\left\{-\left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2}\right) - \frac{\rho(\mu_1 - \mu_2)}{\sigma_1\sigma_2}\right\}\right],$$

$$\tau_2 = \exp\left[\frac{1}{2(1-\rho^2)}\left\{\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right\}\right].$$

When Z_1 and Z_2 have equal marginal variances (i.e., $\mathrm{Var}(Z_1) = \mathrm{Var}(Z_2) = \sigma^2$), the probability density function satisfies

$$\frac{h(u,v)}{h(v,u)} = \tau^{\log v - \log u},$$

where

$$\tau = \exp\left[\frac{\mu_2 - \mu_1}{\sigma^2(1-\rho)}\right].$$

2.2. LNS: Log-normal distribution type symmetry

Consider the $R \times R$ square contingency table with ordered categories. We propose a model defined by

$$p_{ij} = \alpha^{\log i}\beta^{\log j}\psi_{ij} \quad (i = 1,\dots,R; j = 1,\dots,R),$$

where $\psi_{ij} = \psi_{ji}$. This model is also expressed as

$$\frac{p_{ij}}{p_{ji}} = \theta^{\log j - \log i} \quad (i < j),$$

where $\theta = \beta/\alpha$. This indicates that the log-odds that an observation will fall in the (i,j)th cell instead of in the (j,i)th cell, $i < j$, is proportional to the difference between $\log j$ and $\log i$. This model may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate log-normal distribution, where marginal variances of logarithms (i.e., of normally distributed variables) are equal. We shall refer to this model as the log-normal distribution type symmetry (LNS) model. A special case of this model obtained by putting $\theta = 1$ is the S model.

2.3. ELNS: Extended log-normal distribution type symmetry

Moreover, we propose another model defined by

$$p_{ij} = \alpha_1^{\log i}\alpha_2^{(\log i)^2}\beta_1^{\log j}\beta_2^{(\log j)^2}\psi_{ij} \quad (i = 1,\dots,R; j = 1,\dots,R),$$

where $\psi_{ij} = \psi_{ji}$. This model is also expressed as

$$\frac{p_{ij}}{p_{ji}} = \theta_1^{\log j - \log i}\theta_2^{(\log j)^2 - (\log i)^2} \quad (i < j),$$

where $\theta_1 = \beta_1/\alpha_1$ and $\theta_2 = \beta_2/\alpha_2$. This indicates that the probability that an observation will fall in the (i,j)th cell, $i < j$, is $\theta_1^{\log j - \log i}\theta_2^{(\log j)^2 - (\log i)^2}$ times higher than the probability that the observation falls in the (j,i)th cell. This model may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate log-normal distribution without equal marginal variances of logarithms. A special case of this model obtained by putting $\theta_1 = \theta_2 = 1$ is the S model. Also, a special case of this model obtained by putting $\theta_2 = 1$ is the LNS model. Thus, we shall refer to this model as the extended log-normal distribution type symmetry (ELNS) model.

3. Decompositions of symmetry model

Let X and Y denote the row and column variables, respectively. Refer to model of equality of log marginal means, i.e., $\mathrm{E}(\log X) = \mathrm{E}(\log Y)$, as the LME model. Also, refer to model of equality of log marginal means and variances, i.e., $\mathrm{E}(\log X) = \mathrm{E}(\log Y)$ and $\mathrm{Var}(\log X) = \mathrm{Var}(\log Y)$, as the LMVE model. We obtain the decompositions for the S model as follows.

Theorem 1. *The S model holds if and only if both the LNS and LME models hold.*

Theorem 2. *The S model holds if and only if both the ELNS and LMVE models hold.*

The proof of Theorem 1 is given in Appendix 1. The proof of Theorem 2 is omitted because it is obtained in a similar way. The LNS model is a special case of the ELNS model. These models both relax symmetry assumptions needed in the S model. To ensure validity of the S model, we have to add some assumptions on moments. Stronger assumptions (equality of the first and second moments, i.e., the LMVE model) have to be added when the ELNS is hold, while weaker assumptions (equality of the first moments, i.e., the LME model being a special case of the LMVE model) have to be added when the LME model hold. In fact, we need some basic assumptions on symmetry (the ELNS model), on moments (the LME model), and one further assumption. It can be either stronger assumption on symmetry ($\theta_2 = 1$, i.e., the LNS model, as described in Theorem 1) or equality of the second moments (i.e., the LMVE model, as described in Theorem 2). Theorems 1 and 2 may be useful for seeing the reason for the poor fit when the S model fits the data poorly.

4. Orthogonality of test statistics

Let n_{ij} denote the observed frequency in the (i, j)th cell of the table $(i = 1, \dots, R; j = 1, \dots, R)$ with $n = \sum\sum n_{ij}$, and let m_{ij} denote the corresponding expected frequency. Assume that $\{n_{ij}\}$ have a multinomial distribution. The maximum likelihood estimates (MLEs) of $\{m_{ij}\}$ under the LNS and ELNS models could be obtained using iterative procedures; for example, see Darroch and Ratcliff (1972). Let $G^2(M)$ denote the likelihood ratio chi-squared statistic for testing goodness-of-fit of model M. The number of degrees of freedom for the LNS model is $(R-2)(R+1)/2$, which is one less than that for the S model. Also, that for the ELNS model is $(R^2 - R - 4)/2$, which is one less than that for the LNS model.

The orthogonality of the test statistics for goodness-of-fit of two models is discussed by, e.g., Lang and Agresti (1994) and Lang (1996). Generally suppose that model M_3 holds if and only if both models M_1 and M_2 hold. As described in Darroch and Silvey (1963), (i) when the test statistic $G^2(M_3)$ is asymptotically equivalent to the sum of $G^2(M_1)$ and $G^2(M_2)$, where $\mathrm{df}(M_3)$ equals the sum of $\mathrm{df}(M_1)$ and $\mathrm{df}(M_2)$, if both M_1 and M_2 are accepted (at the α significance level) with high probability, then M_3 would be accepted; however (ii) when the asymptotic equivalence described above does not hold, such an incompatible situation that both M_1 and M_2 are accepted with high probability but M_3 is rejected with high probability is quite possible. We obtain the following theorems.

Theorem 3. *The test statistic $G^2(S)$ is asymptotically equivalent to the sum of $G^2(LNS)$ and $G^2(LME)$. The number of degrees of freedom for the S model equals the sum of that of the LNS and LME models.*

Theorem 4. *The test statistic $G^2(S)$ is asymptotically equivalent to the sum of $G^2(ELNS)$ and $G^2(LMVE)$. The number of degrees of freedom for the S model equals the sum of that of the ELNS and LMVE models.*

The proof of Theorem 3 is given in Appendix 2. The proof of Theorem 4 is omitted because it is obtained in a similar way.

5. Examples

Example 1. Table 1 taken from Tomizawa (1985) is constructed from the data of the unaided distance vision of 3168 pupils aged 6-12, including about half the girls at elementary schools in Tokyo, Japan examined in June 1984. In Table 1 the row variable is the right eye grade and the column variable is the left eye grade with the categories ordered from Best (1) to Worst (4).

We see from Table 3 that each of the S, LNS, ELNS and LME models fits these data well although the LMVE model fits poorly. We see that the value of the test statistic for the S model is very close to the sum of the values of those for the LNS and LME models. We shall compare the LNS and ELNS models being nested models. For testing hypothesis that the LNS model holds assuming that the ELNS model holds for these data, we can use the likelihood ratio chi-squared statistic $G(LNS|ELNS)$, where $G(LNS|ELNS) = G(LNS) - G(ELNS)$ with one degree of freedom being the difference between the number of degree of freedom for the LNS and ELNS models. Since $G(LNS|ELNS) = 3.90$, this hypothesis is rejected at the 0.05 significance level. Therefore, the ELNS model would be preferable to the LNS model.

Under the ELNS model, the maximum likelihood estimates of θ_1 and θ_2 are $\hat{\theta}_1 = 2.511$ and $\hat{\theta}_2 = 0.519$, respectively. Therefore the probability that a pupil's right eye grade is i and his/her left eye grade is j $(> i)$ is estimated to be $\hat{\theta}_1^{\log j - \log i}\hat{\theta}_2^{(\log j)^2-(\log i)^2}$ times higher than the probability that the pupil's right eye grade is j and his/her left eye grade is i. Moreover, the maximum likelihood estimates of $\{p_{ij}/p_{ji}\}$ are $\hat{p}_{12}/\hat{p}_{21} = 1.381$, $\hat{p}_{13}/\hat{p}_{31} = 1.245$, $\hat{p}_{14}/\hat{p}_{41} = 1.014$, $\hat{p}_{23}/\hat{p}_{32} = 0.901$, $\hat{p}_{24}/\hat{p}_{42} = 0.735$ and $\hat{p}_{34}/\hat{p}_{43} = 0.815$, respectively. Therefore, for the pupils that their right or left eye grade is Best (1), their right eye is estimated to be better than their left eye. On the other hand, for the pupils that both right and left eyes grade is not Best (1), their right eye is estimated to be worse than their left eye.

Table 1: Unaided distance vision of 3168 pupils comprising nearly equal number of boys and girls aged 6-12 at elementary schools in Tokyo, Japan, examined in June 1984; from Tomizawa (1985). (The parenthesized values are MLEs of expected frequencies under the ELNS model.)

Right eye grade	Left eye grade: Best (1)	Second (2)	Third (3)	Worst (4)	Total
Best (1)	2470	126	21	10	2627
	(2470.00)	(128.75)	(17.19)	(11.08)	
Second (2)	96	138	33	5	272
	(93.25)	(138.00)	(35.56)	(5.08)	
Third (3)	10	42	75	15	142
	(13.81)	(39.44)	(75.00)	(13.92)	
Worst (4)	12	7	16	92	127
	(10.92)	(6.92)	(17.08)	(92.00)	
Total	2588	313	145	122	3168

Example 2. Table 2 taken from Tomizawa (1984) is constructed from the data of unaided distance vision of 4746 students aged 18 to about 25, including about 10% of the women of the Faculty of Science and Technology, Tokyo University of Science in Japan examined in April, 1982.

We see from Table 3 that the LNS and ELNS models fit these data well. However, the S, LME and LMVE models fit poorly. We shall compare the LNS and ELNS models being nested models. Since $G(LNS|ELNS) = 1.56$, this hypothesis that the LNS model holds assuming that the ELNS model holds, is accepted at the 0.05 significance level. Therefore, the LNS model would be preferable to the ELNS model, since it is simpler.

Under the LNS model, the maximum likelihood estimate of θ is $\hat{\theta} = 0.757$. Therefore the probability that a student's right eye grade is i and his/her left eye grade is j $(> i)$ is estimated to be $\hat{\theta}^{\log j - \log i}$ times higher than the probability that the student's right eye grade is j and his/her left eye grade is i. In addition, since $\hat{\theta} < 1$, the marginal probability that a student's right eye is i or below $(i = 1, 2, 3)$ is estimated to be less than the marginal probability that the student's left eye is i or below. Thus, under the LNS model, a student's right eye is estimated to be worse than his/her left eye.

We see from Table 3 that the poor fit of the S model is caused by the influence of the lack of structure of the LME (LMVE) model rather than the LNS (ELNS) model.

Table 2: Unaided distance vision of 4746 students in Faculty of Science and Technology, Tokyo University of Science in Japan examined in April, 1982; from Tomizawa (1984). (The parenthesized values are MLEs of expected frequencies under the LNS model.)

Right eye grade	Left eye grade				
	Best (1)	Second (2)	Third (3)	Worst (4)	Total
Best (1)	1291	130	40	22	1483
	(1291.00)	(126.10)	(44.12)	(17.00)	
Second (2)	149	221	114	23	507
	(152.90)	(221.00)	(112.30)	(21.69)	
Third (3)	64	124	660	185	1033
	(59.88)	(125.70)	(660.00)	(208.32)	
Worst (4)	20	25	249	1429	1723
	(25.00)	(26.31)	(225.68)	(1429.00)	
Total	1524	500	1063	1659	4746

Table 3: Likelihood ratio chi-squared values G^2 for models applied to Tables 1 and 2.

Applied models	Table 1		Table 2	
	Degrees of freedom	G^2	Degrees of freedom	G^2
S	6	9.69	6	16.95*
LNS	5	6.71	5	8.55
ELNS	4	2.81	4	6.99
LME	1	2.97	1	8.37*
LMVE	2	6.87*	2	9.90*

* means significant at 0.05 level.

6. Simulation study

Agresti (1983) showed that in terms of a simulation study, the LDPS model gives a good fit (but the S model gives a poor fit) when there is an underlying bivariate normal distribution with equal marginal variances. We now consider the relationship between the LNS (ELNS) model and the joint bivariate log-normal distribution in terms of simulation studies.

Consider random variables (U, V) having a bivariate log-normal distribution, where $U > 0$ and $V > 0$. Define $Z_1 = \log U$ and $Z_2 = \log V$. Then Z_1 and Z_2 have a bivariate normal distribution, where means $\mathrm{E}(Z_1) = \mu_1$ and $\mathrm{E}(Z_2) = \mu_2$, variances $\mathrm{Var}(Z_1) = \sigma_1^2$ and $\mathrm{Var}(Z_2) = \sigma_2^2$, and correlation $\mathrm{Corr}(Z_1, Z_2) = \rho$. As described in Section 2, the ratio of probability density function $h(u, v)/h(v, u)$ has the form $\tau_1^{\log v - \log u} \tau_2^{(\log v)^2 - (\log u)^2}$ for constants τ_1 and τ_2. We consider the 4×4 tables of sample size 5000 formed by using cut points for each variable at $\frac{1}{2}\mathrm{E}(U), \mathrm{E}(U), 2\mathrm{E}(U)$, where $\mathrm{E}(U) = \exp\{\mu_1 + \frac{\sigma_1^2}{2}\}$ from an underlying bivariate log-normal distribution with the conditions $\mu_1 = 0$, $\sigma_1^2 = 1$, $\rho = 0.3$, $\mu_2 = -0.2, -0.1, 0.0, 0.1, 0.2$ and $\sigma_2^2 = 0.8, 1.0, 1.2$. We make 1000 tables under each condition, and test goodness-of-fit of the S, LDPS, LNS and ELNS models at the 0.05 significance level. Table 4 gives the accepted count under each condition.

We see from Table 4 that the ELNS model fits well for each condition, although the LNS model fits well when σ_2^2 is close to 1. The S models fit well when $\mu_1 = \mu_2$ and $\sigma_1^2 \approx \sigma_2^2$.

7. Concluding remarks

The ELNS model may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate log-normal distribution without the equality of marginal variances, although the LNS model may be appropriate if it is reasonable to assume it with equal marginal variances (see Section 6).

Table 4: The counts accepted by the likelihood ratio chi-squared test for models applied to 1000 tables at the 0.05 significance level.

		Models			
μ_2	σ_2^2	S	LDPS	LNS	ELNS
−0.2	0.8	0	216	3	713
−0.2	1.0	0	867	850	919
−0.2	1.2	0	73	519	938
−0.1	0.8	0	87	0	784
−0.1	1.0	13	937	932	947
−0.1	1.2	23	132	328	935
0	0.8	15	33	15	862
0	1.0	940	945	947	939
0	1.2	159	210	151	891
0.1	0.8	0	17	88	923
0.1	1.0	13	942	912	944
0.1	1.2	0	270	32	849
0.2	0.8	0	3	325	929
0.2	1.0	0	911	782	908
0.2	1.2	0	288	1	726

For the orthogonality of test statistic in Theorem 3, we point out that the likelihood ratio chi-squared statistic for testing goodness-of-fit of the S model assuming that the LNS model holds true is $G^2(S) - G^2(LNS)$ and this is asymptotically equivalent to the likelihood ratio chi-squared statistic for testing goodness-of-fit of the LME model, i.e., $G^2(LME)$. We observe from Table 3 that for each of the data in Tables 1 and 2 the value of $G^2(S)$ is very close to the sum of values of $G^2(LNS)$ and $G^2(LME)$, and it is also very close to the sum of values of $G^2(ELNS)$ and $G^2(LMVE)$.

Acknowledgments

The authors would like to thank the referee for their helpful comments.

Appendix 1

Proof of Theorem 1. If the S model holds, then the LNS and LME models hold. Assuming that both the LNS and LME models hold, then we shall show that the S model holds. Let $\{p_{ij}^*\}$ denote the cell probabilities which satisfy both the LNS and LME models. Since the LNS model holds, we see

$$\log p_{ij}^* = (\log i)(\log \alpha) + (\log j)(\log \beta) + \log \psi_{ij} \quad (i = 1, \ldots, R; j = 1, \ldots, R),$$

where $\psi_{ij} = \psi_{ji}$. Let $\{\pi_{ij} = c^{-1}\psi_{ij}\}$ with $c = \sum_{i=1}^R \sum_{j=1}^R \psi_{ij}$. We note that $\sum_{i=1}^R \sum_{i=1}^R \pi_{ij} = 1$ with $0 < \pi_{ij} < 1$. Then, since the $\{p_{ij}^*\}$ satisfy the LNS and LME models, we see

$$\log\left(\frac{p_{ij}^*}{\pi_{ij}}\right) = \log c + (\log i)(\log \alpha) + (\log j)(\log \beta) \quad (i = 1, \ldots, R; j = 1, \ldots, R), \tag{1}$$

and

$$\mu_1^* = \mu_2^*, \tag{2}$$

where $\mu_1^* = \sum_{i=1}^R \sum_{j=1}^R (\log i) p_{ij}^*$ and $\mu_2^* = \sum_{i=1}^R \sum_{j=1}^R (\log j) p_{ij}^*$. Then, we denote μ_1^* $(= \mu_2^*)$ by μ_0.

Consider the arbitrary cell probabilities $\{p_{ij}\}$ satisfying

$$\tilde{\mu}_1 = \tilde{\mu}_2 = \mu_0, \tag{3}$$

where $\tilde{\mu}_1 = \sum_{i=1}^{R} \sum_{j=1}^{R} (\log i) p_{ij}$ and $\tilde{\mu}_2 = \sum_{i=1}^{R} \sum_{j=1}^{R} (\log j) p_{ij}$.
From (1), (2) and (3), we see

$$\sum_{i=1}^{R} \sum_{j=1}^{R} (p_{ij} - p_{ij}^{*}) \log \left(\frac{p_{ij}^{*}}{\pi_{ij}} \right) = 0. \tag{4}$$

Using the equation (4), we obtain

$$K(\{p_{ij}\}, \{\pi_{ij}\}) = K(\{p_{ij}^{*}\}, \{\pi_{ij}\}) + K(\{p_{ij}\}, \{p_{ij}^{*}\}),$$

where

$$K(\{a_{ij}\}, \{b_{ij}\}) = \sum_{i=1}^{R} \sum_{j=1}^{R} a_{ij} \log \left(\frac{a_{ij}}{b_{ij}} \right),$$

and $K(\{a_{ij}\}, \{b_{ij}\})$ is the Kullback-Leibler information between $\{a_{ij}\}$ and $\{b_{ij}\}$. Since $\{\pi_{ij}\}$ being a function $\{p_{ij}^{*}\}$ is fixed, we see

$$\min_{\{p_{ij}\}} K(\{p_{ij}\}, \{\pi_{ij}\}) = K(\{p_{ij}^{*}\}, \{\pi_{ij}\}),$$

and then $\{p_{ij}^{*}\}$ uniquely minimizes $K(\{p_{ij}\}, \{\pi_{ij}\})$; see Bhapkar and Darroch (1990).
Let $\{p_{ij}^{**} = p_{ji}^{*}\}$. Then,

$$\log p_{ij}^{**} = \log p_{ji}^{*} = (\log j)(\log \alpha) + (\log i)(\log \beta) + \log \psi_{ji} \quad (i = 1, \ldots, R; j = 1, \ldots, R), \tag{5}$$

with $\psi_{ij} = \psi_{ji}$. Noting that $\{\pi_{ij} = \pi_{ji}\}$, the equation (5) is also expressed as

$$\log \left(\frac{p_{ij}^{**}}{\pi_{ij}} \right) = \log c + (\log j)(\log \alpha) + (\log i)(\log \beta) \quad (i = 1, \ldots, R; j = 1, \ldots, R), \tag{6}$$

From (2), (3) and (6), we see

$$\sum_{i=1}^{R} \sum_{j=1}^{R} (p_{ij} - p_{ij}^{**}) \log \left(\frac{p_{ij}^{**}}{\pi_{ij}} \right) = 0. \tag{7}$$

Using the equation (7), we obtain

$$K(\{p_{ij}\}, \{\pi_{ij}\}) = K(\{p_{ij}^{**}\}, \{\pi_{ij}\}) + K(\{p_{ij}\}, \{p_{ij}^{**}\}).$$

Since $\{\pi_{ij}\}$ being a function $\{p_{ij}^{**}\}$ is fixed, we see

$$\min_{\{p_{ij}\}} K(\{p_{ij}\}, \{\pi_{ij}\}) = K(\{p_{ij}^{**}\}, \{\pi_{ij}\}),$$

and then $\{p_{ij}^{**}\}$ uniquely minimizes $K(\{p_{ij}\}, \{\pi_{ij}\})$. Therefore, we see $p_{ij}^{*} = p_{ij}^{**}$. Thus, $p_{ij}^{*} = p_{ji}^{*}$. Namely the S model holds. $\square$

Appendix 2

Proof of Theorem 3. Let $p = (p_{11}, \ldots, p_{1R}, \ldots, p_{R1}, \ldots, p_{RR})^t$ denote the $R^2 \times 1$ vector, where A^t denotes the transpose of vector (or matrix) A. Since parameter θ is expressed as the function of $\{p_{ij}\}$ in the LNS model, this model can be written as

$$H_1(p) = 0_{d_1},$$

where

$$H_1(p) = (H_{13}(p), \ldots, H_{1R}(p), H_{23}(p), \ldots, H_{2R}(p), \ldots, H_{R-1,R}(p))^t,$$

with

$$H_{ij}(p) = \log p_{ij} - \log p_{ji} - \frac{\log j - \log i}{\log 2}(\log p_{12} - \log p_{21}),$$

and 0_s denotes the $s \times 1$ vector (or scalar) with all elements zero with $d_1 = (R-2)(R+1)/2$. The LME model is expressed as

$$H_2(p) = 0_{d_2},$$

where

$$H_2(p) = \sum_{i=1}^{R}\sum_{j=1}^{R}(\log i - \log j)p_{ij},$$

and $d_2 = 1$. From Theorem 1, the S model is expressed as

$$H_3(p) = 0_{d_3},$$

where

$$H_3(p) = (H_1(p)^t, H_2(p))^t,$$

and $d_3 = R(R-1)/2$. Let $h_s(p)$ denote the $d_s \times R^2$ matrix of partial derivatives of $H_s(p)$ with respect to p, i.e., $h_s(p) = \partial H_s(p)/\partial p^t$ for $s = 1, 2, 3$. Let $\Sigma(p) = \mathrm{diag}(p) - pp^t$, where $\mathrm{diag}(p)$ denotes a diagonal matrix with ith component of p as ith diagonal component. Let $\hat{p}$ denote estimate of p with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij} = n_{ij}/n\}$. Using the delta method, $\sqrt{n}(H_3(\hat{p}) - H_3(p))$ has asymptotically a normal distribution with mean zero and covariance matrix

$$h_3(p)\Sigma(p)h_3(p)^t = \begin{bmatrix} h_1(p)\Sigma(p)h_1(p)^t & h_1(p)\Sigma(p)h_2(p)^t \\ h_2(p)\Sigma(p)h_1(p)^t & h_2(p)\Sigma(p)h_2(p)^t \end{bmatrix}.$$

We see that all elements of $h_1(p)\Sigma(p)h_2(p)^t$ equal zero. Thus we obtain $\Delta_3(p) = \Delta_1(p)+\Delta_2(p)$, where

$$\Delta_s(p) = H_s(p)^t[h_s(p)\Sigma(p)h_s(p)^t]^{-1}H_s(p). \tag{8}$$

Under each $H_s(p) = 0_{d_s}$ $(s = 1, 2, 3)$, the Wald statistic $W_s = n\Delta_s(\hat{p})$ has asymptotically a chi-squared distribution with d_s degrees of freedom. From equation (8), we see that $W_3 = W_1 + W_2$. From the asymptotic equivalence of the Wald statistic and likelihood ratio statistic, we obtain Theorem 3. □

References

Agresti A (1983). "A Simple Diagonals-parameter Symmetry and Quasi-symmetry Model." *Statistics and Probability Letters*, **1**, 313–316.

Agresti A (2013). *Categorical Data Analysis, 3rd ed.* New Jersey/Wiley.

Bhapkar P, Darroch N (1990). “Marginal Symmetry and Quasi Symmetry of General Order.” *Journal of Multivariate Analysis*, **34**, 173–184.

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

Bowker H (1948). “A Test for Symmetry in Contingency Tables.” *Journal of the American Statistical Association*, **43**, 572–574.

Darroch N, Ratcliff D (1972). “Generalized Iterative Scaling for Log-linear Models.” *Annals of Mathematical Statistics*, **43**, 1470–1480.

Darroch N, Silvey D (1963). “On Testing More than One Hypothesis.” *Annals of Mathematical Statistics*, **34**, 555–567.

Lang B (1996). “On the Partitioning of Goodness-of-fit Statistics for Multivariate Categorical Response Models.” *Journal of the American Statistical Association*, **91**, 1017–1023.

Lang B, Agresti A (1994). “Simultaneously Modeling Joint and Marginal Distributions of Multivariate Categorical Responses.” *Journal of the American Statistical Association*, **89**, 625–632.

Tomizawa S (1984). “Three Kinds of Decompositions for the Conditional Symmetry Model in a Square Contingency Table.” *Journal of the Japan Statistical Society*, **14**, 35–42.

Tomizawa S (1985). “Analysis of Data in Square Contingency Tables with Ordered Categories Using the Conditional Symmetry Model and Its Decomposed Models.” *Environmental Health Perspectives*, **63**, 235–239.

Affiliation:

Kiyotaka Iki
Department of Information Science
Faculty of Science and Technology
Tokyo University of Science
Noda City, Chiba, Japan
E-mail: iki@is.noda.tus.ac.jp

Austrian Journal of Statistics — http://www.ajs.or.at/
published by the Austrian Society of Statistics — http://www.osg.or.at/
Volume 47 — *Submitted:* 2017-04-17
June 2018 — *Accepted:* 2017-08-02