Journal of Biometrics & Biostatistics
ISSN: 2155-6180
Saigusa et al., J Biom Biostat 2018, 9:2
DOI: 10.4172/2155-6180.1000395

Research Article

Open Access

Unrestricted Normal Distribution Type Symmetry Model for Square Contingency Tables with Ordered Categories

Yusuke Saigusa[1]*, Shimpei Goda[2], Kouji Yamamoto[3] and Sadao Tomizawa[2]

[1]Department of Biostatistics, Yokohama City University School of Medicine, Yokohama City, Kanagawa 236–0004, Japan
[2]Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Noda City, Chiba 278–8510, Japan
[3]Department of Medical Statistics, Osaka City University, Graduate School of Medicine, Osaka City, Osaka, 545–8585, Japan

Abstract

For square contingency tables with ordered categories, Yamamoto, Nakane and Tomizawa [1] proposed the restricted normal distribution type symmetry (RNDS) model that the cell probability has a similar form as a bivariate normal density with equal marginal means and variances. This paper proposes a new model that has more relaxed constraints based on a bivariate normal density. It also gives a decomposition of the RNDS model into the proposed model and marginal means and variances equality. Moreover it is shown that the test statistic for goodness-of-fit of the RNDS model is asymptotically equivalent to the sum of those for the decomposed models.

Keywords: Bivariate normal distribution; Contingency table; Normal distribution type symmetry; Ordered category

Introduction

For the analysis of square contingency tables with same row and column categories, one of the interest of many statisticians is what structure of symmetry for probabilities in the table there is. Bowker [2] considered the symmetry (S) model which indicates there is a symmetric structure of cell probabilities with respect to the main diagonal of the table. For ordinal square contingency tables, Agresti [3] proposed the linear diagonals-parameter symmetry (LDPS) model which indicates an asymmetric structure such that the ratios of symmetric cell probabilities vary linearly depending on the distance from main diagonal. The LDPS model is appropriate if it is reasonable to assume that there is an underlying continuous distribution which is bivariate normal with equal marginal variances. Tomizawa [4] described the extended LDPS (ELDPS) model which is appropriate assuming that there is an underlying bivariate normal distribution which does not require the equality of marginal variances. The LDPS and ELDPS models have theoretical justifications in the sense that the restrictions of the ratios between symmetric probabilities in each model have similar form as those in the bivariate normal density function [3,4]. Tahata, Yamamoto and Tomizawa [5] considered the normal distribution type symmetry (NDS) model which indicates a cell probability directly has similar form as a bivariate normal density with equal marginal variances. Yamamoto, Nakane and Tomizawa [1] proposed more parsimonious model, described the restricted normal distribution type symmetry (RNDS) model, which indicates a cell probability has similar form as a bivariate normal density with equal marginal means and variances.

First purpose of present paper is to consider a new model which would be appropriate if it is reasonable to assume that there is an underlying bivariate normal distribution even if the marginal means and variances are respectively unequal.

Decompositions of a model are given by several statisticians, for example, Caussinus [6] gave the theorem that the S model holds if and only if both of (1) the quasi-symmetry model which indicates symmetry of the odds ratios [6] and (2) the marginal homogeneity model which indicates the row and column marginal distributions are identical [7] hold. For decompositions of a model, several statisticians described partitioning of goodness-of-fit test statistic for the model. Aitchison [8] discussed the asymptotic separability. Read [9] practically the partitioning of test statistic of model for square contingency tables. Tomizawa and Tahata [10] showed the likelihood ratio statistic for testing the goodness of fit of the S model is asymptotically equivalent to the sum of those for the quasi-symmetry and marginal homogeneity models.

Second purpose of present paper is to give the decomposition of the RNDS model into the proposed model and marginal means and variances equality. In the data analysis, the decomposition may be useful for seeing the reason of the poor fit when the RNDS model fits the data poorly.

Third purpose of present paper is to show that the test statistic for goodness of fit of the RNDS model is asymptotically equivalent to the sum of those for the decomposed models.

In present paper, Section 2 reviews models for square contingency tables. Section 3 proposes a new model based on bivariate normal density. Section 4 gives the decomposition of the RNDS model using the proposed model. Section 5 describes the goodness-of-fit test. Section 6 shows the partitioning of test statistic for the RNDS model. Section 7 gives an example applying the proposed model to contingency table data of decayed teeth. Section 8 gives some numerical simulation studies. Section 9 presents concluding remarks.

Review of Models

Consider an $r \times r$ square contingency table with same ordered categories. Let p_{ij} denote the probability that an observation will fall in the cell in row i and column j (i=1,…,r; j=1,…,r). Bowker [2] described the S model defined by:

***Corresponding author:** Yusuke Saigusa, Department of Biostatistics, Yokohama City University School of Medicine, Yokohama City, Kanagawa 236–0004, Japan, Tel: 0457872311; E-mail: saigusay@yokohama-cu.ac.jp

Received February 14, 2018; **Accepted** February 24, 2018; **Published** March 01, 2018

Citation: Saigusa Y, Goda S, Yamamoto K, Tomizawa S (2018) Unrestricted Normal Distribution Type Symmetry Model for Square Contingency Tables with Ordered Categories. J Biom Biostat 9: 395. doi: 10.4172/2155-6180.1000395

$$p_{ij}=\Psi_{ij}\,(i\neq j),$$

where $\Psi_{ij}=\Psi_{ji}$. Agresti [3] proposed the LDPS model defined by:

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

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

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

Tomizawa (1991) proposed the ELDPS model defined by:

$$p_{ij}=\begin{cases}\delta^{j-i}\gamma^{(j-i)(j+i)/2}\psi_{ij} & (i<j),\\ \psi_{ij} & (i\geq j),\end{cases}$$

where $\Psi_{ij}=\Psi_{ji}$. This indicates that the probability that an observation will fall in the (i,j)th cell is $\delta^{j-i}\gamma^{(j-i)(j+i)/2}$ times higher than the probability that it falls in (j,i)th cell, $i<j$. The ELDPS model is also expressed as:

$$\frac{p_{ij}}{p_{ji}}=\delta^{j-i}\gamma^{(j-i)(j+i)/2}\quad(i<j).$$

Consider the continuous random variables U and V having a joint bivariate normal distribution with means $E(U)=\mu_1$ and $E(V)=\mu_2$, variances $Var(U)=\sigma_1^2$ and $Var(V)=\sigma_2^2$, and correlation $Corr\,(U,V)=\rho$. Then the ratio of joint bivariate normal density $f\,(u,v)$ is:

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

Agresti [3] pointed out that $f(u,v)/(v,u)$ has the form τ^{v-u} for unspecified parameter τ when $\sigma_1^2=\sigma_2^2$, and hence the LDPS model may be appropriate for a square ordinal table if there is an underlying bivariate normal distribution with equal marginal variances. Tomizawa [4] pointed out that the ELDPS model may be appropriate even if marginal variances are unequal.

When $\sigma_1^2=\sigma_2^2=\sigma^2$, the density function $f(u,v)$ is formed by:

$$f(u,v)=ca_1^{(u-v)^2}a_2^{u-v}b_1^{(u+v)^2}b_2^{u+v}$$

where

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

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

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

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

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

Tahata et al. [5] proposed the NDS model defined by:

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

Under the NDS model, $p_{ij}/p_{ji}=\alpha_2^{2(i-j)}$ for i < j. Therefore the NDS model implies the LDPS model. Yamamoto et al. [1] proposed the RNDS model defined by:

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

The RNDS model is the special case of a NDS model obtained by putting $\alpha_2=1$. Under the NDS and RNDS models, the cell probability p_{ij} has a form which is similar to bivariate normal density with equal marginal variances and with equal marginal means and variances, respectively.

Unrestricted Normal Distribution Type Symmetry Model

The joint bivariate normal density $f(u,v)$ may be expressed as:

$$f(u,v)=ca_1^{u^2}a_2^{v^2}b_1^{u}b_2^{v}d^{uv}$$

where

$$c=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{\mu_1^2}{\sigma_1^2}-\frac{2\mu_1\mu_2\rho}{\sigma_1\sigma_2}+\frac{\mu_2^2}{\sigma_2^2}\right)\right)$$

$$a_1=\exp\left(-\frac{1}{2\sigma_1^2(1-\rho^2)}\right)$$

$$a_2=\exp\left(-\frac{1}{2\sigma_2^2(1-\rho^2)}\right)$$

$$b_1=\exp\left(\frac{\mu_1\sigma_2-\mu_2\sigma_1\rho}{\sigma_1^2\sigma_2(1-\rho^2)}\right)$$

$$b_2=\exp\left(\frac{\mu_2\sigma_1-\mu_1\sigma_2\rho}{\sigma_1\sigma_2^2(1-\rho^2)}\right)$$

$$d=\exp\left(\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\right)$$

For an $r\times r$ square table, we shall propose a model defined by:

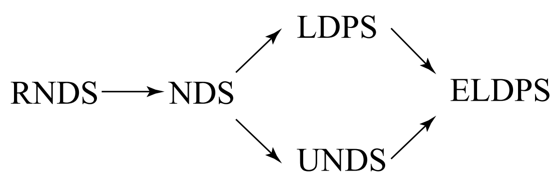
$$p_{ij}=\mu\alpha_1^{i^2}\alpha_2^{j^2}\beta_1^{i}\beta_2^{j}\gamma^{ij}\quad(i=1,\ldots,r;j=1,\ldots,r).$$

We shall refer to this model as the unrestricted normal distribution type symmetry (UNDS) model. We can see that the cell probability p_{ij} has a form similar to the normal density with unequal marginal means and marginal variances under this model. Therefore the UNDS model may be appropriate for an ordinal table if it is possible to assume that there is an underlying bivariate normal distribution which does not require both of the equal marginal means and equal marginal variances. Special cases of the UNDS model obtained, by setting $\alpha_1=\alpha_2$ is the NDS model, and by setting $\alpha_1=\alpha_2$ and $\beta_1=\beta_2$ is the RNDS model. Hence the proposed UNDS model is extension of both the NDS and RNDS models. Under the UNDS model, we see:

$$\frac{p_{ij}}{p_{ji}}=\left(\frac{\alpha_1}{\alpha_2}\right)^{i^2-j^2}\left(\frac{\beta_1}{\beta_2}\right)^{i-j}$$

Therefore the UNDS model implies the ELDPS model.

Citation: Saigusa Y, Goda S, Yamamoto K, Tomizawa S (2018) Unrestricted Normal Distribution Type Symmetry Model for Square Contingency Tables with Ordered Categories. J Biom Biostat 9: 395. doi: 10.4172/2155-6180.1000395



Note: $M_2 \rightarrow M_1$ indicates that model M_2 implies model M_1.

Figure 1: Relationships among models.

Figure 1 represents relationships among the models. Note that $M_2 \rightarrow M_1$ indicates that model M_2 implies model M_1.

Decomposition of Model

Let X_1 and X_2 denote the row and column variables of a table, respectively. Consider the marginal mean and variance equality (MV) model defined by:

$$\mu_1 = \mu_2, \quad \sigma_1^2 = \sigma_2^2,$$

where $\mu_1 = \sum_{i=1}^{r}\sum_{j=1}^{r} i p_{ij}$, $\mu_2 = \sum_{i=1}^{r}\sum_{j=1}^{r} j p_{ij}$, $\sigma_1^2 = \sum_{i=1}^{r}\sum_{j=1}^{r} i^2 p_{ij} - \mu_1^2$ and $\sigma_2^2 = \sum_{i=1}^{r}\sum_{j=1}^{r} j^2 p_{ij} - \mu_2^2$. This indicates that $E(X_1)=E(X_2)$ and $Var(X_1) = Var(X_2)$. This model is also expressed as $E(X_1)=E(X_2)$ and $E(X_1^2) = E(X_2^2)$ Then, we obtain:

Theorem 1

The RNDS model holds if and only if both the UNDS and MV models hold.

Proof. If the RNDS model holds, then the UNDS and MV models hold. Assuming that both the UNDS and MV models hold, we shall show that the RNDS model holds. Let $p^* = \{p_{ij}^*\}$ denote the cell probabilities which satisfy both the UNDS and MV models. Since the UNDS model holds, we see:

$$\log p_{ij}^* = \log\mu + i^2 \log\alpha_1 + j^2 \log\alpha_2 + i\log\beta_1 + j\log\beta_2 + \log\gamma^{ij}$$

Let $\pi_{ij}=\gamma^{ij}/c$ with $c = \sum_{i=1}^{r}\sum_{j=1}^{r}\gamma^{ij}$. We note that $\pi=\{\pi_{ij}\}$ satisfies $\sum_{i=1}^{r}\sum_{j=1}^{r}\pi_{ij} = 1$ with $0 < \pi_{ij} < 1$. Then the UNDS and MV models are expressed as:

$$\log\left(\frac{p_{ij}^*}{\pi_{ij}}\right) = \log\mu + i^2\log\alpha_1 + j^2\log\alpha_2 + i\log\beta_1 + j\log\beta_2 + \log c \tag{1}$$

and

$$\mu_1^* = \mu_2^*, \quad \sigma_1^{*2} = \sigma_2^{*2} \tag{2}$$

where

$$\mu_1^* = \sum_{s=1}^{r}\sum_{t=1}^{r} s p_{st}^*, \qquad \mu_2^* = \sum_{s=1}^{r}\sum_{t=1}^{r} t p_{st}^*$$

$$\sigma_1^{*2} = \sum_{s=1}^{r}\sum_{t=1}^{r} s^2 p_{st}^* - \mu_1^{*2}, \qquad \sigma_2^{*2} = \sum_{s=1}^{r}\sum_{t=1}^{r} t^2 p_{st}^* - \mu_2^{*2}$$

We denote $\mu_1^*(=\mu_2^*)$ by μ_0 and $\sigma_1^{*2}(=\sigma_2^{*2})$ by σ_0^2. Consider the arbitrary cell probability $p=\{p_{ij}\}$ satisfying:

$$\mu_1 = \mu_2 = \mu_0, \qquad \sigma_1^2 = \sigma_2^2 = \sigma_0^2 \tag{3}$$

where,

$$\mu_1 = \sum_{s=1}^{r}\sum_{t=1}^{r} s p_{st}, \qquad \mu_2 = \sum_{s=1}^{r}\sum_{t=1}^{r} t p_{st}$$

$$\sigma_1^2 = \sum_{s=1}^{r}\sum_{t=1}^{r} s^2 p_{st} - \mu_1^2, \qquad \sigma_2^2 = \sum_{s=1}^{r}\sum_{t=1}^{r} t^2 p_{st} - \mu_2^2$$

From eqns. (1), (2) and (3), we see:

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

From eqns. (4) we obtain:

$$K(p,\pi) = K(p,p^*) + K(p^*,\pi)$$

where $K(a,b)$ is the Kullback-Leibler information between $\{a_{ij}\}$ and $\{b_{ij}\}$, defined by:

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

Note that $K(a,b) \geq 0$ and the equality holds when only $a_{ij}=b_{ij}$, for $i,j=1,\ldots,r$. Since $\{\pi_{ij}\}$ is fixed, we see $\min_p K(p,\pi) = K(p^*,\pi)$ and then $\{p_{ij}^*\}$ uniquely minimizes $K(p,\pi)$ [11].

Let $p^{**} = \{p_{ij}^{**}\}$ with $p_{ij}^{**} = p_{ji}^*$ for $1 \leq i,j \leq r$. Then, noting that $\pi_{ij}=\pi_{ji}$, we obtain $\min_p K(p,\pi) = K(p^{**},\pi)$ and then $\{p_{ij}^{**}\}$ uniquely minimizes $K(p,\pi)$. Therefore, we see $p_{ij}^* = p_{ij}^{**}$, namely, $p_{ij}^* = p_{ji}^*$ for $i,j=1,\ldots,r$. Then we see:

$$\frac{p_{ij}^*}{p_{ji}^*} = \left(\frac{\alpha_1}{\alpha_2}\right)^{i^2-j^2}\left(\frac{\beta_1}{\beta_2}\right)^{i-j} = 1 \quad (i,j=1,\ldots,r) \tag{5}$$

and

$$\frac{p_{i+1,j+1}^*}{p_{j+1,i+1}^*} = \left(\frac{\alpha_1}{\alpha_2}\right)^{2(i-j)}\left(\frac{\alpha_1}{\alpha_2}\right)^{i^2-j^2}\left(\frac{\beta_1}{\beta_2}\right)^{i-j} = 1 \quad (i,j=1,\ldots,r-1) \tag{6}$$

From eqns. (5) and (6), we obtain:

$$\left(\frac{\alpha_1}{\alpha_2}\right)^{2(i-j)} = 1 \quad (i,j=1,\ldots,r-1) \tag{7}$$

which leads to $\alpha_1=\alpha_2$. From eqns. (5) and (7), we obtain:

$$\left(\frac{\beta_1}{\beta_2}\right)^{i-j} = 1 \quad (i,j=1,\ldots,r)$$

which leads to $\beta_1=\beta_2$. Therefore the RNDS model holds. The proof is completed.

Goodness of Fit Test

Assume that multinomial distribution applies to the $r \times r$ table. The maximum likelihood estimates of expected frequencies under each model can be obtained by using Newton-Raphson method to the log-likelihood equations or by using iterative procedure, for example, the general iterative procedure for log-linear model of Darroch and Ratcliff [11].

Let n_{ij} denote the observed frequency in the (i,j)th cell of the table.

Citation: Saigusa Y, Goda S, Yamamoto K, Tomizawa S (2018) Unrestricted Normal Distribution Type Symmetry Model for Square Contingency Tables with Ordered Categories. J Biom Biostat 9: 395. doi: 10.4172/2155-6180.1000395

The likelihood ratio chi-square 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 the maximum likelihood estimate of expected frequency m_{ij} under model M. The number of degrees of freedom (df) for the UNDS model is r^2-6. Also the numbers of df for the RNDS, NDS, LDPS, ELDPS and MV models are r^2 - 4, r^2 - 5, (r^2 - r - 2)/2, (r^2 - r - 4)/2 and 2, 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 UNDS model and M_1 is ELDPS model. Let k_1 and k_2 denote the numbers of df for models M_1 and M_2, respectively. Note that $k_1 < k_2$ and G^2 (M_1) ≤ G^2 (M_2). For conditional goodness-of-fit test of 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 model M_1, this test statistic has an asymptotic chisquare distribution with k_2 - k_1 df.

Partition of Test Statistics

We obtain:

Theorem 2

The following asymptotic equivalence holds:

$$G^2(RNDS) \simeq G^2(UNDS) + G^2(MV)$$

The number of df for the RNDS model equals the sum of the numbers of df for the UNDS and MV models.

Proof. The UNDS model may be expressed in a log-linear form:

$$\log p_{ij} = \mu^* + i^2\alpha_1^* + j^2\alpha_2^* + i\beta_1^* + j\beta_2^* + ij\gamma^* \quad (i = 1,\ldots,r; j = 1,\ldots,r)$$

Let

$$p = (p_{11}, p_{12},\ldots, p_{1r}, p_{21},\ldots, p_{rr})^t$$

and

$$\beta = (\mu^*, \alpha_1^*, \alpha_2^*, \beta_1^*, \beta_2^*, \gamma^*)^t$$

where "t" denotes the transpose. Then the UNDS model is expressed as:

$$\log p = X\beta = (1_{r^2}, X_1, X_2, X_3, X_4, X_5)\beta$$

where X is the $r^2 \times 6$ matrix and 1_s is the $s \times 1$ vector of 1 elements,

$$X_1 = (1,1,1,\ldots,1,4,4,4,\ldots,4,\ldots,r^2,r^2,r^2,\ldots,r^2)^t = (1,4,\ldots,r^2)^t \otimes 1_r^t$$

$$X_2 = (1,4,9,\ldots,r^2,1,4,9,\ldots,r^2,\ldots,1,4,9,\ldots,r^2)^t = 1_r^t \otimes (1,4,\ldots,r^2)^t$$

$$X_3 = (1,1,1,\ldots,1,2,2,2,\ldots,2,\ldots,r,r,r,\ldots,r)^t = (1,2,\ldots,r)^t \otimes 1_r^t$$

$$X_4 = (1,2,3,\ldots,r,1,2,3,\ldots,r,\ldots,1,2,3,\ldots,r)^t = 1_r^t \otimes (1,2,\ldots,r)^t$$

$$X_5 = (1,2,3,\ldots,r,2,4,6,\ldots,2r,\ldots,r,2r,3r,\ldots,r^2)^t = (1,2,\ldots,r)^t \otimes (1,2,\ldots,r)^t$$

and ⊗ denotes the Kronecker product. Note that the model matrix X is full column rank and the rank of X is 6. In a similar manner to Haber [12], we denote the linear space spanned by the columns of the matrix X by $S(X)$ with the dimension 6. Let U be an $r^2 \times d_1$ full column rank matrix, where d_1=r^2-6, such that the linear space spanned by the columns of U, $S(U)$, is orthogonal compliment of $S(X)$. Thus, $U^tX = O_{d_1,6}$ which is the $d_1 \times 6$ zero matrix. Therefore the UNDS model is expressed as:

$$h_1(p) = 0_{d_1}$$

where $h_1(p) = U^t\log p$ and 0_s is the $s \times 1$ zero vector. The MV model may be expressed as:

$$h_2(p) = 0_{d_2}$$

where $h_2(p)$=Wp and d_2= 2 and W is the $d_2 \times r^2$ matrix with:

$$W^t = (X_3 - X_4, X_1 - X_2)$$

Therefore the column vectors of W^t belong to space $S(X)$, namely, $S(W^t) \subset S(X)$. Hence $WU = O_{d_2,d_1}$. From Theorem 1, the RNDS model is expressed as:

$$h_3(p) = 0_{d_3}$$

where $h_3(p) = \left(h_1(p)^t, h_2(p)^t\right)^t$ and d_3=d_1 + d_2=r^2 – 4. Note that $h_s(p)$ is the vector of order $d_s \times 1$, s=1,2,3, and d_1,d_2 and d_3 are the numbers of df for testing goodness-of-fit of the UNDS, MV and RNDS models, respectively.

Let $H_s(p)$ denote the $d_s \times r^2$ matrix of partial derivatives of $h_s(p)$ with respect to p^t, i.e.,

$$H_s(p) = \frac{\partial h_s(p)}{\partial p^t} \quad (s = 1,2,3)\cdot$$

Let $\Sigma(p) = diag(p) - pp^t$ where $diag(p)$ denotes a diagonal matrix with ith component of p as i th diagonal component. Then we see:

$$H_1(p)\Sigma(p)H_2(p)^t = U^tW^t = O_{d_1,d_2} \tag{8}$$

Let Δ_s denote $\Delta_s = h_s(p)^t\left[H_s(p)\Sigma(p)H_s(p)^t\right]^{-1}h_s(p)$, s=1,2,3. Namely:

$$\Delta_3 = \begin{pmatrix} h_1(p) \\ h_2(p) \end{pmatrix}^t \begin{pmatrix} H_1(p)\Sigma(p)H_1(p)^t & H_1(p)\Sigma(p)H_2(p)^t \\ H_2(p)\Sigma(p)H_1(p)^t & H_2(p)\Sigma(p)H_2(p)^t \end{pmatrix}^{-1} \begin{pmatrix} h_1(p) \\ h_2(p) \end{pmatrix} \tag{9}$$

From eqns. (8) and (9), we obtain Δ_3=Δ_1 + Δ_2. Under $h_s(p) = 0_{d_s}$, the Wald statistic $W_s = n\Delta_s(\hat{p})$, where $\hat{p} = n_{ij}/n$ and $n = \Sigma\Sigma n_{ij}$, has asymptotically the chi-squared distribution with d_s df, s = 1,2,3. From Δ_3=Δ_1 + Δ_2, W_3=W_1 + W_2. Since the Wald statistic is asymptotically equivalent to the likelihood ratio statistic [8,13], we obtain Theorem 2. The proof is completed.

An Example

The data in Table 1 taken from Tomizawa, Miyamoto and Iwamoto [14] are the decayed teeth data of 363 women aged 18-39, for patients visiting a dental clinic in Sapporo City, Japan, from 2001 to 2005. Table 2 gives the values of likelihood ratio chi-square statistic G^2 for models

	Upper			
Lower	**(1) 0-4**	**(2) 5-8**	**(3) 9+**	**Total**
(1) 0-4	97	62	15	174
	(95.12)	(62.12)	(16.76)	
(2) 5-8	20	63	75	158
	(23.40)	(63.48)	(71.12)	
(3) 9+	2	6	23	31
	(0.48)	(5.40)	(25.12)	
Total	119	131	113	363

Note: The parenthesized values are the maximum likelihood estimates of expected frequencies under the UNDS model.

Table 1: Numbers of decayed teeth data of 363 women aged 18-39, for patients visiting a dental clinic in Sapporo City, Japan, from 2001 to 2005; from Tomizawa, Miyamoto and Iwamoto [14].

Citation: Saigusa Y, Goda S, Yamamoto K, Tomizawa S (2018) Unrestricted Normal Distribution Type Symmetry Model for Square Contingency Tables with Ordered Categories. J Biom Biostat 9: 395. doi: 10.4172/2155-6180.1000395

Applied models	Degrees of freedom	G^2
LDPS	2	11.05*
ELDPS	1	2.66
NDS	4	24.59*
RNDS	5	117.24*
UNDS†	3	3.88
MV	2	89.48*

Note: † means the proposed model.

*means significant at the 0.05 level.

Table 2: Values of likelihood ratio statistic G2 for models applied to the data in Table 1.

	(a) $\mu_2=\mu_1$ and $\sigma_2^2=\sigma_1^2$					
ρ	**Hypothesis**					
	LDPS	**ELDPS**	**NDS**	**RNDS**	**UNDS**	**UNDS|ELDPS**
0.1	948928	931860	945178	943146	945175	888023
0.3	948965	942864	926804	927213	924523	870414
0.5	948412	946063	851006	854888	842392	767789
0.7	944747	942699	604754	619468	583373	470966
	(b) $\mu_2=\mu_1+0.2$ and $\sigma_2^2=\sigma_1^2$					
ρ	**Hypothesis**					
	LDPS	**ELDPS**	**NDS**	**RNDS**	**UNDS**	**UNDS|ELDPS**
0.1	944702	942433	941646	146490	938958	886568
0.3	941935	939529	918831	68252	917840	860676
0.5	937577	935074	842646	11344	833718	758797
0.7	932473	929557	597827	113	576790	472584
	(c) $\mu_2=\mu_1+0.2$ and $\sigma_2^2=1.2\sigma_1^2$					
ρ	**Hypothesis**					
	LDPS	**ELDPS**	**NDS**	**RNDS**	**UNDS**	**UNDS|ELDPS**
0.1	757728	942470	814884	146362	940574	889008
0.3	746636	939991	778780	68213	914640	856961
0.5	725336	935558	663082	13282	822359	744716
0.7	678794	927783	383706	223	541963	434724
	(d) $\mu_2=\mu_1+0.4$ and $\sigma_2^2=1.4\sigma_1^2$					
ρ	**Hypothesis**					
	LDPS	**ELDPS**	**NDS**	**RNDS**	**UNDS**	**UNDS|ELDPS**
0.1	281282	928175	387073	4	925600	867818
0.3	262502	916741	351107	0	891480	837369
0.5	231949	899227	258610	0	782062	706041
0.7	192275	884906	112828	0	492457	407308

Table 3: The frequencies of acceptance (at the 0.05 significance level) of goodnessof-fit test for the LDPS, ELDPS, NDS, RNDS and UNDS models and conditional goodness of fit test for the UNDS model assuming the ELDPS model holds (and the ELDPS model is accepted), denoted UNDS|ELDPS, per 1000000 times for 4×4 tables of sample size 1000 and correlation ρ on some condition.

applied to the data in Table 1.

From Table 2, the LDPS, NDS, RNDS and MV models fit the data in Table 1 poorly. Whereas, the ELDPS and UNDS models fit the data well. In addition, the UNDS model is preferable to the ELDPS model for the data because the conditional goodness-of-fit test of the UNDS model assuming the ELDPS model is not significant at the 0.05 level with the difference between the G^2 values for the UNDS model and the ELDPS model is 1.22 with 3 − 1=2 df.

From Theorem 1, it is inferred that the poor fit of the RNDS model applied to the data in Table 1 is caused by the influence of lack of structure of the MV model which fits the data poor rather than the UNDS model. Thus it is seen that for these data the distribution of the numbers of lower and upper decayed teeth may be similar to a bivariate normal distribution which has unequal marginal means and variances. So the extended parameters a_1(or a_2) and β_1(or β_2) in the UNDS model may improve the fit better than the NDS and RNDS models.

Simulation Studies

The UNDS model may be appropriate for an ordinal square table when it is assumed that there is an underlying bivariate normal distribution even if the marginal means and variances are unequal, respectively. Now we shall consider the simulation studies based on the bivariate normal distribution.

Consider the random vector $Z=(Z_1,Z_2)$ which is distributed as a bivariate normal distribution with means $E(Z_1)=\mu_1$, $E(Z_2)=\mu_2$, variances $Var(Z_1)=\sigma_1^2$, $Var(Z_2)=\sigma_2^2$, and correlation $Corr(Z_1,Z_2)=p$. Suppose that there is an underlying bivariate normal distribution with some conditions and then a table is formed using cut points for each variable at μ_1, $\mu_1 \pm 0.7\sigma_1$. Table 3 gives the frequencies of acceptance (at the 0.05 significance level) of goodness-of-fit test for the LDPS, ELDPS, NDS, RNDS and UNDS models and conditional goodness-of-fit test for the UNDS model assuming the ELDPS model holds (when the ELDPS model is accepted) per 1000000 times for 4×4 tables of sample size 1000 on some conditions given in Table 3. From Table 3a, when the marginal means and variances are respectively equal, each model tends to fit well. From Table 3b, when the marginal means are unequal, each model except for the RNDS model tends to fit well. From Table 3d, when the marginal means and marginal variances are respectively unequal, the UNDS and ELDPS models tend to fit well. From Tables 3a, 3c and 3d, it can be seen that, as the difference of marginal means and marginal variances increases simultaneously, the frequencies of acceptance for the LDPS, NDS and RNDS models decrease, whereas those for the UNDS and ELDPS models do not decrease so much.

From the frequencies of acceptance of conditional goodness-of-fit test for the UNDS model assuming the ELDPS model holds, we see that the UNDS model would be preferable to the ELDPS model when the ρ is small (especially when it is smaller than 0.5).

Concluding Remarks

We have proposed the UNDS model which indicates a cell probability having similar form as a bivariate normal density. The proposed UNDS model may be appropriate for an ordinal square table assuming that there is an underlying bivariate normal distribution even if the marginal means and variances are respectively unequal. We have also shown that with simulation studies.

We have obtained the theorem that the RNDS model holds if and only if both the UNDS and MV models hold. For analysing the data, Theorem 1 may be useful for seeing the reason of the poor fit when the RNDS model fits the data poorly (see section 7).

Acknowledgements

The authors would like to thank referees and the editor for their helpful comments.

References

1. Yamamoto K, Nakane H, Tomizawa S (2016) Symmetry model based on bivariate normal distribution for square contingency tables with ordered categories. J Stat Adv Theory Appl 15: 71-84.
2. Bowker AH (1948) A test for symmetry in contingency tables. J Amer Stat Ass 43: 572-574.
3. Agresti A (1983) A simple diagonals-parameter symmetry and quasi-symmetry model. Stat Prob Let 1: 313-316.
4. Tomizawa S (1991) An extended linear diagonals-parameter symmetry model for square contingency tables with ordered categories. Metron 49: 401-409.
5. Tahata K, Yamamoto K, Tomizawa S (2009) Normal distribution type symmetry model for square contingency tables with ordered categories. The Open Stat Prob J 1: 32-37.
6. Caussinus H (1965) Contribution to the statistical analysis of correlation tables, Annals of the Faculty of Sciences of the University of Toulouse 29: 77-182.
7. Stuart A (1955) A test for homogeneity of the marginal distributions in a two-way classifications. Biometrika 42: 412-416.
8. Aitchison J (1962) Large-sample restricted parametric tests. J Roy Stat Soc 24: 234-250.
9. Read CB (1977) Partitioning chi-square in contingency tables: A teaching approach. Communications in Statistics, Theory and Methods 6: 553-562.
10. Tomizawa S, Tahata K (2007) The analysis of symmetry and asymmetry: Orthogonality of decomposition of symmetry into quasi-symmetry and marginal symmetry for multi-way tables. J Fr Stat Soc 148: 3-36.
11. Darroch JN, Ratcliff D (1972) Generalized iterative scaling for log-linear models. The Annals of Mathematical Statistics 43: 1470-1480.
12. Haber M (1985) Maximum likelihood methods for linear and log-linear models in categorical data. Comput Stat Dat Anal 3: 1-10.
13. Rao CR (1973) Linear Statistical Inference and Its Applications (2ndedn), Wiley, New York.
14. Tomizawa S, Miyamoto N, Iwamoto M (2006) Linear column-parameter symmetry model for square contingency tables: application to decayed teeth data. Biometrical Letters 43: 91-98.