Computational Statistics and Data Analysis 78 (2014) 135–140

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Model based on skew normal distribution for square contingency tables with ordinal categories

Kouji Yamamoto [a,*], Hidetoshi Murakami [b]

[a] Department of Clinical Epidemiology and Biostatistics, Graduate School of Medicine, Osaka University, 2-2, Yamadaoka, Suita, Osaka, 565-0871, Japan

[b] Department of Mathematics, National Defense Academy, 1-10-20 Hashirimizu, Yokosuka, Kanagawa, 239-8686, Japan

ARTICLE INFO

Article history:
Received 23 July 2013
Received in revised form 19 December 2013
Accepted 10 April 2014
Available online 26 April 2014

Keywords:
Linear diagonals-parameter symmetry
Normal distribution type symmetry
Ordinal category

ABSTRACT

For the analysis of square contingency tables with ordinal categories, Tahata, Yamamoto and Tomizawa (2009) considered the normal distribution type symmetry model, which may be appropriate if it is reasonable to assume an underlying bivariate normal distribution with equal marginal variances. The present paper proposes a new model which may be appropriate for a square ordinal table if it is reasonable to assume an underlying bivariate skew normal distribution with equal marginal variances. Simulations are used to investigate the fitting of new model for bivariate skew normal distribution. The decayed teeth data are analyzed by using the new model.

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, j)th cell of the table $(i = 1, \ldots, r; j = 1, \ldots, r)$.

Agresti (1983) proposed the linear diagonals-parameter symmetry model, namely LDPS, as follows:

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

A special case of this model obtained by putting $\theta = 1$ is the symmetry model (see, e.g., Bowker, 1948; Bishop et al., 1975, p. 282).

Let the random variable $\boldsymbol{X} = (U, V)^T$ be distributed according to the bivariate normal distribution with means $E(U) = \mu_1$ and $E(V) = \mu_2$, variances $\mathrm{Var}(U) = \mathrm{Var}(V) = \sigma^2$, and correlation $\mathrm{Corr}(U, V) = \rho$. The density function is expressed as

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

Then it satisfies

$$\frac{f(u, v)}{f(v, u)} = \exp\left(\frac{(v-u)(\mu_2-\mu_1)}{(1-\rho)\sigma^2}\right) \quad (u < v).$$

Agresti (1983) described that $f(u, v)/f(v, u)$ has the form θ^{v-u} 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 (see also Tomizawa, 1991; Yamamoto et al., 2007; Yamamoto et al., 2008; Tahata and Tomizawa, 2010).

* Corresponding author.
E-mail addresses: yamamoto-k@stat.med.osaka-u.ac.jp (K. Yamamoto), murakami@gug.math.chuo-u.ac.jp (H. Murakami).

http://dx.doi.org/10.1016/j.csda.2014.04.007

In addition, the normal density $f(u, v)$ can be expressed as

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

where

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

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

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

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

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

Then, the normal distribution type symmetry (NDS) model was proposed by Tahata et al. (2009) as follows:

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

This model is a special case of the LDPS model. Tahata et al. (2009) pointed out that the $\{p_{ij}\}$ has a similar structure to the bivariate normal density 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.

The skew normal distribution (e.g., Azzalini and Dalla Valle, 1996) is well known to be an extended distribution of the normal distribution. A standard bivariate skew normal density function $g_2(\boldsymbol{u})$ is given as

$$g_2(\boldsymbol{u}) = 2\phi_2(\boldsymbol{u})\Phi(\boldsymbol{\gamma}'\boldsymbol{u}) \quad (\boldsymbol{u} \in R^2), \tag{1}$$

where $\phi_2(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the probability density function of the $N_2(\boldsymbol{0}, I_2)$ distribution and cumulative distribution function of the standard normal distribution, and $\boldsymbol{\gamma} \in R^2$ is the skewness parameter (see Baghfalaki and Ganjali, 2011). More generally, for an $N_2(\boldsymbol{\mu}, \Omega)$ distribution, we denote the equivalent functions by $\phi_2(\cdot; \boldsymbol{\mu}, \Omega)$. As a location-scale extension of (1), the bivariate skew normal density $h(\boldsymbol{v})$ is given as

$$h(\boldsymbol{v}) = 2\phi_2(\boldsymbol{v}; \boldsymbol{\mu}, \Omega)\Phi(\boldsymbol{\gamma}'\Omega^{-1/2}(\boldsymbol{v}-\boldsymbol{\mu})) \quad (\boldsymbol{v} \in R^2), \tag{2}$$

with the mean vector $\boldsymbol{\mu}$, the covariance matrix $\boldsymbol{\Omega}$, and the skew vector $\boldsymbol{\gamma}$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \boldsymbol{\Omega} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad \boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}.$$

When $\boldsymbol{\gamma} = \boldsymbol{0}$, the density (2) is equivalent to the density of the $N_2(\boldsymbol{\mu}, \Omega)$ distribution. Then, we are interested in considering a model which has a similar structure of bivariate skew normal density of (2).

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

2. New model and test

We consider a new model based on the skew normal distribution for square contingency tables with ordinal categories. In Section 2.1, we propose a skew normal distribution type symmetry model, namely SNDS. In Section 2.2, we describe a goodness-of-fit test for the model SNDS.

2.1. Skew normal distribution type symmetry model

We propose a new model for square contingency tables with ordinal categories as follows:

$$p_{ij} = 2\xi\alpha_1^{(i-j)^2} \alpha_2^{i-j} \beta_1^{(i+j)^2} \beta_2^{i+j} \Phi(i\lambda_1 + j\lambda_2) \quad (i = 1, \ldots, r; j = 1, \ldots, r),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard univariate normal distribution, and $\lambda_1, \lambda_2 \in \boldsymbol{R}$. Then we shall refer to this model as the skew normal distribution type symmetry (SNDS) model. It is easily seen that the SNDS model is an extension of the NDS model.