Behaviormetrika, Vol. 25, No. 2, 95-110
1998

A LOG-BILINEAR MODEL WITH LATENT VARIABLES

Haruhiko Ogasawara*

A factor-analytic model for discrete variables which represent frequencies is developed. The model assumes that the frequencies are generated by the Poisson distribution, in which the logarithm of the parameter of the distribution takes a similar form to the factor-analysis model for continuous variables. The latent variables in the model, which correspond to common factors in factor analysis, are assumed to have independent distributions with fixed parameters or a multivariate normal distribution with unknown correlations. The individual factor scores are integrated out from the model and factor-loadings in the sense of the generalized linear model are obtained by the marginal maximum likelihood method. Numerical examples of non-verbal intelligence tests are given.

1. Introduction

The objective of this paper is to propose a factor-analytic model for two-way frequency tables of subjects by discrete variables. There exist various models which are applied to such two-way tables. One of them is the log-linear model (see e.g., Bishop et al., 1975). However, since the number of subjects may be as large as hundreds, it is not realistic to apply the log-linear model to these situations especially when interaction terms are involved in the model. The log-bilinear model (also called RC association model or log-multiplicative model; Goodman, 1985, 1986, 1991) is the model in which the interaction terms in the log-linear model are parsimoniously structured in factor-analytic forms.

The subjects in the frequency tables are assumed to represest a random sample from a population, which is also a frequently used assumption in the factor-analysis model for continuous variables. In frequency tables of subjects by variables (items, tests etc.), if the number of subjects increases, the number of the parameters concerning individuals (the nuisance parameters) increases proportionally (Clogg, 1986). This paper deals with the estimation of the parameters of interest in such situations by a similar method to that of Bock and Aitkin (1981; see also Bock, Gibbons & Muraki, 1988), though the types of the response functions are different.

The multiplicative Poisson model for two-way frequency tables was developed by Rasch (1960/1980). In this model the frequency (e.g., the number of errors) for the i-th subject and the j-th test is Poisson distributed with the parameter $\lambda_{ij}=\theta_i\delta_j$, where δ_j stands for the difficulty of the j-th test and θ_i denotes the inability of the i-th subject. Though in the original Rasch model subjects were not a random

Key Words and Phrases; Poisson distribution, Rasch model, factor analysis, EM-algorithm, marginal maximum likelihood, latent variables
* Department of Information and Management Science, Otaru University of Commerce, 3-5-21 Midori, Otaru, 047-8501 Japan. E-mail address: hogasa@res.otaru-uc.ac.jp

sample, Rasch (1973) discussed a situation in which subejcts represent a random sample. The Rasch model in which subject parameters are random variables have been proposed in various ways (Böckenholt, 1993; Jansen, 1986; Jansen & Van Duijn, 1993; Ogasawara, 1996; Van Duijn, 1993). However, all of the extended models mentioned above assume a unidimensional latent variable. It is a natural extension to consider several latent variables in the frequency data. The multiple-factor models for dichotomous or polytomous variables have been developed by Christoffersson (1975), Muthén (1987) and Bock and Aitkin (1981). The factor-analytic model with multiple factors for frequency data will be proposed in the following sections.

The Poisson models or more widely item response models can be seen from the viewpoint of the generalized linear model (McCullagh & Nelder, 1989; Mellenbergh, 1994). The generalized linear model assumes that a transformation (the so-called link function) of the expectation of a response variable is a linear function of parameters. For the ordinary factor analysis model for continuous variables, the link function is the identity function. For the logistic model for binary data, the link function is the logit function. For the Poisson model for frequency data, the link function is the logarithm. Since in the item response theory (IRT), the latent varibles representing abilities (factors) are involved, most of the models in IRT may be regarded as factor analysis models in the generalized sense. However, in the case of factor analysis, because latent factors which correspond to the independent variables in the generalized linear model, are generally unknown, the name of (generalized) bilinear model is more appropriate than that of linear model (see Kruskal, 1981 and Choulakian, 1996).

2. Model

Let q be the number of latent orthogonal-factors which represent different abilities. Here we use the terms, factor and factor-loading, in the generalized sense in the previous section. Let the values of the factors for the i-th subject be denoted by the vector $\boldsymbol{\theta}_i=(\theta_{i1},\cdots,\theta_{iq})'$, $(i=1,\cdots,N)$, where N is the number of subjects. Let $\boldsymbol{\delta}_j=(\delta_{j1},\cdots,\delta_{jq})'$, $(j=1,\cdots,p)$ be the vector of factor-loadings for the j-th test, where p is the number of tests. Let the variable representing the count of the i-th subject and the j-th test be denoted by X_{ij}. Then, the probability that the count x_{ij} occurs is assumed to be Poisson distributed with the parameter λ_{ij}. That is,

$$P(X_{ij}=x_{ij}\mid \boldsymbol{\theta}_i,\ \boldsymbol{\delta}_j)=\lambda_{ij}^{x_{ij}}\exp(-\lambda_{ij})/x_{ij}!,\ (i=1,\cdots,N\,;\ j=1,\cdots,p). \tag{1}$$

The Rasch model is the case when q is one and $\lambda_{ij}=\theta_{i1}\delta_{j1}$. When q is greater than one, a natural extension of the Rasch model is

$$\lambda_{ij}=\sum_{k=1}^{q}\theta_{ik}\delta_{jk}=\boldsymbol{\theta}_i'\boldsymbol{\delta}_j. \tag{2}$$

However, in the Poisson distribution the parameter λ_{ij} must be greater than zero. The model of (2) does not necessarily satisfy the condition, unless the regions of θ_{ik} and δ_{jk} are specified appropriately. Thus, we employ a modified model:

$$\lambda_{ij}=\exp(\boldsymbol{\theta}_i'\boldsymbol{\delta}_j) \tag{3}$$

in stead of (2). The models of (2) and (3) mean that the probability of $X_{ij}=x_{ij}$ is determined by the q latent-factors when $\boldsymbol{\delta}_j$ are given. Further we adopt an intercept parmeter μ_j in the following way:

$$\lambda_{ij}=\exp\ (\boldsymbol{\theta}_i'\ \boldsymbol{\delta}_j+\mu_j), \tag{4}$$

where μ_j represents the difficulty of the j-th test. This model is equivalent to the log-bilinear model when the intercept parameters for subjects are missing. Using (4), the equation (1) is described as

$$\begin{aligned} &P\ (X_{ij}=x_{ij}\,|\,\boldsymbol{\theta}_i, \boldsymbol{\delta}_j, \mu_j) \\ &=\exp\{x_{ij}\ (\boldsymbol{\theta}_i'\ \boldsymbol{\delta}_j+\mu_j)\}\exp\{-\exp(\boldsymbol{\theta}_i'\ \boldsymbol{\delta}_j+\mu_j)\}/x_{ij}!, \\ &(i=1, \cdots, N\ ;\ j=1, \cdots, p). \end{aligned} \tag{5}$$

The direct interpretation of $\boldsymbol{\theta}_i$ and $\boldsymbol{\delta}_j$ is possible in the same manner as the log-bilinear model (Goodman, 1985, 1986, 1991):

$$\ln\frac{\lambda_{ij}/\lambda_{ij'}}{\lambda_{i'j}/\lambda_{i'j'}}=(\boldsymbol{\theta}_i-\boldsymbol{\theta}_{i'})'(\boldsymbol{\delta}_j-\boldsymbol{\delta}_{j'}). \tag{6}$$

That is, the difference of two vectors of parameters is proportional to the log-odds-ratio of the rate of occurrence of events.

Let the elements of $\boldsymbol{\theta}_i$ be the random variables which represent the values of the factors for the i-th subject and $h(\boldsymbol{\theta}_i)$ be the density of the factors at $\boldsymbol{\theta}_i$. For the orthogonal model, we assume that

$$h(\boldsymbol{\theta}_i)=\prod_{k=1}^{q} g_k(\theta_{ik}). \tag{7}$$

That is, the elements of $\boldsymbol{\theta}_i$ are supposed to be distributed independently of each other. However, the model of (5) cannot be identified without specifying the locations and dispersions of the factors unless other restrictions are imposed on $\boldsymbol{\delta}_j$ and μ_j. In the following, we assume identical distributions for θ_{ik}, $(k=1, \cdots, q)$, that is, $g_k(\theta_{ik})=g(\theta_{ik})$, $(k=1, \cdots, q)$ in (7). This situation corresponds to the one where the means and variances of latent factors in the factor-analysis model for continuous variables are arbitrary unless more than one population is considered. That is, even if the location and scale for a factor are changed, the equivalent model can be obtained by modifying the associated factor-loadings and intercept parameters appropriately. Consequently, the mean and variance for a factor are usually set to zero and one, respectively. To confirm this result, the mean and variance of X_{ij} are derived in Appendix in the case of the normal model for $h(\boldsymbol{\theta}_i)$.

3. Estimation of parameters for orthogonal model

It is possible to regard $\boldsymbol{\theta}_i$, $(i=1, \cdots, N)$ as unknown and fixed parameters. This approach is not well suited to the case in which interest lies in the characteristics of a relatively small number of tests (or items in IRT), and a relatively large number of subjects are employed to obtain the information about those tests. The number of prameters to be estimated increases with the number of subjects, and complicates an understanding of the asymptotic properties of test parameter estimates. Thus, we use the method of marginal maximum likelihood (Bock & Aitkin, 1981), where $\boldsymbol{\theta}_i$, $(i=1, \cdots, N)$ are integrated out from the likelihood. Let $\boldsymbol{\delta}_j{}^*$ be the vector of "unknown" factor-loadings in the j-th test and let

$$\boldsymbol{\mu}=(\mu_1, \cdots, \mu_p)' \text{ and } \boldsymbol{\delta}=(\boldsymbol{\delta}_1^{*\prime}, \cdots, \boldsymbol{\delta}_p^{*\prime})', \tag{8}$$

then the parameters to be estimated are $\boldsymbol{\mu}$ and $\boldsymbol{\delta}$. Note that $\boldsymbol{\delta}_j$ may include known loadings in addition to those in $\boldsymbol{\delta}_j{}^*$. We set some of the factor-loadings to predetermined values and fix $h(\boldsymbol{\theta}_i)$ so that the unidentifiability problem should be solved. Then, given the values of $\boldsymbol{\theta}_i$, $(i=1, \cdots, N)$, the marginal likelihood L^* with the assumption of local independence, is described as

$$L^*(\boldsymbol{\delta},\boldsymbol{\mu} \mid X)=\prod_{i=1}^{N} \int_{R(\theta_i)} \Big(\prod_{j=1}^{p} P(X_{ij}=x_{ij} \mid \boldsymbol{\theta}_i, \boldsymbol{\delta}_j, \mu_j)\Big)h(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i, \tag{9}$$

where $X=\{x_{ij}\}$. Since it is difficult to obtain algebraically the result of the integration in (9), we use r^q lattice points, $y_{m1} \otimes \cdots \otimes y_{mq}$, $(m_k=1, \cdots, r\,;\ k=1, \cdots, q)$ in the independent q-dimensional density functions. We use the weight $A(y_{m1}) \times \cdots \times A(y_{mq})$ for each lattice point at $y_{m1}, \cdots, y_{mq}$, $(m_k=1, \cdots r,\,;\ k=1, \cdots, q)$ which is constructed to approximate the continuous density function. Let L be an approximation of the marginal likelihood L^*, then

$$L^* \cong L=\prod_{i=1}^{N} \sum_{m1=1}^{r} \cdots \sum_{mq=1}^{r} \Big(\prod_{j=1}^{p} P(X_{ij}=x_{ij} \mid \boldsymbol{y}, \boldsymbol{\delta}_j, \mu_j)\Big)A(y_{m1}) \times \cdots \times A(y_{mq}), \tag{10}$$

where $\boldsymbol{y}=(y_{m1}, \cdots, y_{mq})'$. The estimates of $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ are the values which maximize (10). Taking the logarithm of (10), we have

$$l=\ln L=\sum_{i=1}^{N} \ln P(\boldsymbol{X}_i=\boldsymbol{x}_i \mid \boldsymbol{\delta}, \boldsymbol{\mu}), \tag{11}$$

where $\boldsymbol{X}_i=(X_{i1}, \cdots, X_{ip})'$, $\boldsymbol{x}_i=(x_{i1}, \cdots, x_{ip})'$ and

$$P(\boldsymbol{X}_i=\boldsymbol{x}_i \mid \boldsymbol{\delta}, \boldsymbol{\mu})=\sum_{m1=1}^{r} \cdots \sum_{mq=1}^{r} \Big(\prod_{j=1}^{p} P(X_{ij}=x_{ij} \mid \boldsymbol{y}, \boldsymbol{\delta}_j, \mu_j)\Big) \times A(y_{m1}) \times \cdots \times A(y_{mq}). \tag{12}$$

The maximization of l, which is equivalent to the maximization of L, is performed by a kind of the EM (Expectation-Maximization) algorithm (Bock & Aitkin, 1981; Bock, Gibbons & Muraki, 1988; Dempster et al., 1977; Harwell et al., 1988). The gradient vector for the method is

$$\frac{\partial l}{\partial \delta_{jk}}=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\frac{\partial \ln \prod_{j=1}^{p} P(X_{ij}=x_{ij}\mid \boldsymbol{y}, \boldsymbol{\delta}_j, \mu_j)}{\partial \delta_{jk}}$$
$$\times\frac{(\prod_{j=1}^{p} P(X_{ij}=x_{ij}\mid \boldsymbol{y}, \boldsymbol{\delta}_j, \mu_j))\ A(y_{m1})\times\cdots\times A(y_{mq})}{P(\boldsymbol{X}_i=\boldsymbol{x}_i\mid \boldsymbol{\delta}, \boldsymbol{\mu})},$$
$$(j=1, \cdots, p;\ k=1, \cdots, q).$$

Let $f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu})=\frac{(\prod_{j=1}^{p} P(X_{ij}=x_{ij}\mid \boldsymbol{y}, \boldsymbol{\delta}_j, \mu_j))A(y_{m1})\times\cdots\times A(y_{mq})}{P(\boldsymbol{X}_i=\boldsymbol{x}_i\mid \boldsymbol{\delta}, \boldsymbol{\mu})}$, then $f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu})$ is the posterior probability of $\boldsymbol{y}$, given $\boldsymbol{x}_i$, $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$, where $\boldsymbol{y}$ is the vector of the scores of the latent variables at a lattice point. Using this, the above equation becomes

$$\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\frac{\partial\{x_{ij}(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)-\exp(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)-\ln x_{ij}!\}}{\partial \delta_{jk}}f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu})$$
$$=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\{x_{ij}-\exp(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)\}y_{mk}f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}). \tag{13}$$

Similarly,

$$\frac{\partial l}{\partial \mu_j}=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\{x_{ij}-\exp(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)\}f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}),\quad (j=1, \cdots, p) \tag{14}$$

is obtained.

The equations (13) and (14) can be seen as those in the Poisson regression, if $f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu})$ are regarded as given weights. Though actually the weights $f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu})$ have unknown parameters, $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$, if we regard them as fixed values computed by temporary values of $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$, we have the information matrix as follows:

$$E\left(\frac{-\partial^2 l}{\partial \delta_{jk}\partial \delta_{jk'}}\right)=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)y_{mk}y_{mk'}f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}),$$
$$E\left(\frac{-\partial^2 l}{\partial \delta_{jk}\partial \delta_{j'k'}}\right)=0,\quad (j\neq j'),$$
$$E\left(\frac{-\partial^2 l}{\partial \mu_j\partial \delta_{jk}}\right)=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)y_{mk}f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}), \tag{15}$$
$$E\left(\frac{-\partial^2 l}{\partial \mu_j\partial \delta_{j'k}}\right)=0,\quad (j\neq j'),$$
$$E\left(\frac{-\partial^2 l}{\partial \mu_j^2}\right)=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}),$$
$$E\left(\frac{-\partial^2 l}{\partial \mu_j\partial \mu_{j'}}\right)=0,\quad (j\neq j'),$$
$$(j, j'=1, \cdots, p;\ k, k'=1, \cdots, q).$$

The estimates of $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ are obtained by Fisher's scoring method using (14) and (15) with $f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu})$ as known values which are renewed in each iterative process by updated values of $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$. The iteration can be performed for each test j as follows:

$$\begin{bmatrix} \delta^*{}_j \\ \mu_j \end{bmatrix}_{(i+1)} = \begin{bmatrix} \boldsymbol{\delta}^*{}_j \\ \mu_j \end{bmatrix}_{(i)} + E\left(\frac{-\partial^2 l}{\partial(\boldsymbol{\delta}_j^{*\prime}, \mu_j)'\partial(\boldsymbol{\delta}^{*\prime}{}_j, \mu_j)}\right)_{(i)}^{-1} \left[\frac{\partial l}{\partial(\boldsymbol{\delta}_j^{*\prime}, \mu_j)'}\right]_{(i)} \tag{16}$$

where the subscript (i) indicates the values in the i-th iteration. In the standard EM algorithm, (16) is iterated to obtain the maximum of the likelihood with the fixed weights $f(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu})$. However, the weights can be revised in each iteration before the maximum is attained

In the ordinary scoring method, the inverse of the information matrix computed by the converged values of parameters can be used as an estimator of the asymptotic variance-covariance matrix of the parameters. However, since in (15) the weights $f(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu})$ are regarded as if they were known, (15) cannot be used as the information matrix. Instead, we use the following matrix as an approximation to the exact information matrix (Louis, 1982; Mislevy, 1984, 1985; Mislevy & Sheehan, 1988):

$$\begin{aligned} I = \sum_{i=1}^{N} \Bigg[& \left\{ \sum_{m1=1}^{r} \cdots \sum_{mq=1}^{r} \frac{\partial \ln(\prod_{j=1}^{p} P(X_{ij}=x_{ij} \mid \boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\mu}))}{\partial(\boldsymbol{\delta}', \boldsymbol{\mu}')'} f(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}) \right\} \\ & \times \left\{ \sum_{m1=1}^{r} \cdots \sum_{mq=1}^{r} \frac{\partial \ln(\prod_{j=1}^{p} P(X_{ij}=x_{ij} \mid \boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{\mu}))}{\partial(\boldsymbol{\delta}', \boldsymbol{\mu}')} f(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}) \right\} \Bigg]. \end{aligned} \tag{17}$$

Using this approximation, the estimator of the asymptotic variance-covariance matrix is obtained as follows:

$$C\hat{o}v\{(\hat{\boldsymbol{\delta}}', \hat{\boldsymbol{\mu}}')'\} = I^{-1}. \tag{18}$$

This is a method of estimating the standard errors of the parameters in confirmatory orthogonal factor models. When using a just identified exploratory factor analysis model with the assumption of the mutivariate normal distribution for factors, we can transform the estimated results by factor rotation. In this case, the information matrix (17) should be replaced by the augmented one (see, Silvey, 1975) with appropriate restrictions for parameters (see e.g., Jennrich, 1974; Ogasawara, 1998) in a similar way to that of the rotated solutions in usual factor analysis.

4. Oblique model

When latent factors are correlated, the oblique factor analysis model for Poisson variables is obtained. In the former sections of orthogonal model, no unknown parameter is involved in the density function of $\boldsymbol{\theta}_i$. The distribution types of a factor in the one-factor model are not restricted to be normal: we will have a log-normal distribution and a uniform one in the numerical examples. On the other hand, for the model with correlated factors, we have the parameters for the covariances (correlations) among factors and assume that $\boldsymbol{\theta}_i$ has the multivariate normal distribution.

Suppose that the latent factors have zero means and unit variances, then the density of $\boldsymbol{\theta}_i$ is

$$h(\boldsymbol{\theta}_i \mid \mathrm{P})=\frac{1}{(2\pi)^{q/2}\mid \mathrm{P}\mid^{1/2}}\exp\left(-\frac{1}{2}\boldsymbol{\theta}_i'\mathrm{P}^{-1}\boldsymbol{\theta}_i\right), \tag{19}$$

where $P=\{\rho_{ij}\}$ is the correlation matrix of the factors.

Defining $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ in the similar way to the orthogonal case and $\boldsymbol{\rho}=(\rho_{21}, \rho_{31}, \rho_{32}, \cdots, \rho_{q,(q-1)})'$, the marginal likelihood L^* of $\boldsymbol{\alpha}=(\boldsymbol{\delta}', \boldsymbol{\mu}', \boldsymbol{\rho}')'$ is

$$L^*(\boldsymbol{\alpha} \mid X)=\prod_{i=1}^{N}\int_{R(\theta_i)}(\prod_{j=1}^{p} P(X_{ij}=x_{ij}\mid \boldsymbol{\delta}_j, \mu_j))h(\boldsymbol{\theta}_i\mid \mathrm{P})d\boldsymbol{\theta}_i. \tag{20}$$

Since it is difficult to obtain algebraically the result of integration in (20), the approximation by lattice points is also used as in the orthogonal case. That is,

$$L^*\cong L=\prod_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}(\prod_{j=1}^{p}P(X_{ij}=x_{ij}\mid \boldsymbol{y}, \boldsymbol{\delta}_j, \mu_j))\ A(\boldsymbol{y}\mid \mathrm{P}), \tag{21}$$

where

$$A(\boldsymbol{y}\mid \mathrm{P})=\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)\Big/\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right) \tag{22}$$

and $\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}A(\boldsymbol{y}\mid \mathrm{P})=1$.

The function to be maximized is

$$l=\ln L=\sum_{i=1}^{N}\ln P(\boldsymbol{X}_i=\boldsymbol{x}_i\mid \boldsymbol{\alpha}), \tag{23}$$

where

$$P(\boldsymbol{X}_i=\boldsymbol{x}_i\mid \boldsymbol{\alpha})=\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}(\prod_{j=1}^{p}P(X_{ij}=x_{ij}\mid \boldsymbol{y}, \boldsymbol{\delta}_j, \mu_j))A(\boldsymbol{y}\mid \mathrm{P}). \tag{24}$$

The maximum is obtained by the EM algorithm in a similar way to the orthogonal case. The gradient vector with respect to $\boldsymbol{\delta}$ and $\boldsymbol{\mu}$ is similarly obtained as

$$\frac{\partial l}{\partial \delta_{jk}}=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\{x_{ij}-\exp(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)\}y_{mk}f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\alpha}),$$
$$(j=1, \cdots, p;\ k=1, \cdots, q), \tag{25}$$
$$\frac{\partial l}{\partial \mu_j}=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\{x_{ij}-\exp(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)\}f(\boldsymbol{y}\mid \boldsymbol{x}_i,\boldsymbol{\alpha}),\quad (j=1, \cdots, p).$$

where

$$f(\boldsymbol{y}\mid \boldsymbol{x}_i, \boldsymbol{\alpha})=\frac{(\prod_{j=1}^{p}P(X_{ij}=x_{ij}\mid \boldsymbol{y}, \boldsymbol{\delta}_j, \mu_j))A(\boldsymbol{y}\mid \mathrm{P})}{P(\boldsymbol{X}_i=\boldsymbol{x}_i\mid \boldsymbol{\alpha})} \tag{26}$$

is the posterior probability of $\boldsymbol{y}$ given $\boldsymbol{x}_i$ and $\boldsymbol{\alpha}$. The gradient vector for $\rho_{st}(q\geq s$

$>t\geq 1$) is

$$\frac{\partial l}{\partial\rho_{st}}=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\frac{(\prod_{j=1}^{p}P(X_{ij}=x_{ij}\mid \boldsymbol{y},\boldsymbol{\delta}_j,\mu_j))A(\boldsymbol{y}\mid \mathrm{P})}{P(\boldsymbol{X}_i=\boldsymbol{x}_i\mid\boldsymbol{\alpha})}\times\frac{\partial\ln A(\boldsymbol{y}\mid\mathrm{P})}{\partial\rho_{st}} \tag{27}$$
$$=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\frac{\partial\ln A(\boldsymbol{y}\mid\mathrm{P})}{\partial\rho_{st}}f(\boldsymbol{y}\mid\boldsymbol{x}_i,\boldsymbol{\alpha}).$$

Using

$$\frac{\partial\ln A(\boldsymbol{y}\mid\mathrm{P})}{\partial\rho_{st}}=(\mathrm{P}^{-1}\boldsymbol{y}\boldsymbol{y}'\mathrm{P}^{-1})_{st}-\frac{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}(\mathrm{P}^{-1}\boldsymbol{y}\boldsymbol{y}'\mathrm{P}^{-1})_{st}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)}{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)}$$

and $\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}f(\boldsymbol{y}\mid\boldsymbol{x}_i,\boldsymbol{\alpha})=1$, (27) becomes

$$\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}(\mathrm{P}^{-1}\boldsymbol{y}\boldsymbol{y}'\mathrm{P}^{-1})_{st}\left\{f(\boldsymbol{y}\mid\boldsymbol{x}_i,\boldsymbol{\alpha})-\frac{\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)}{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)}\right\}. \tag{28}$$

Suppose that $f(\boldsymbol{y}\mid\boldsymbol{x}_i,\boldsymbol{\alpha})$ are given weights, the non-zero elements of the information matrix are obtained as follows:

$$E\left(\frac{-\partial^2 l}{\partial(\boldsymbol{\delta}_j^{*\prime},\mu_j)'\partial(\boldsymbol{\delta}_j^{*\prime},\mu_j)}\right),\quad (j=1,\cdots,p),$$
$$=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}E\left(\frac{-\partial^2\ln P(X_{ij}=x_{ij}\mid\boldsymbol{y},\boldsymbol{\delta}_j,\mu_j)}{\partial(\boldsymbol{\delta}_j^{*\prime},\mu_j)'\partial(\boldsymbol{\delta}_j^{*\prime},\mu_j)}\right)f(\boldsymbol{y}\mid\boldsymbol{x}_i,\boldsymbol{\alpha})$$
$$=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp(\boldsymbol{y}'\boldsymbol{\delta}_j+\mu_j)\begin{bmatrix}\boldsymbol{y}\boldsymbol{y}' & \boldsymbol{y}' \\ \boldsymbol{y} & 1\end{bmatrix}f(\boldsymbol{y}\mid\boldsymbol{x}_i,\boldsymbol{\alpha}), \tag{29}$$

$$E\left(\frac{-\partial^2 l}{\partial\rho_{st}\partial\rho_{uv}}\right)=\sum_{i=1}^{N}\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}(\mathrm{P}^{-1}\boldsymbol{y}\boldsymbol{y}'\mathrm{P}^{-1})_{st}\left(\partial\frac{\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)}{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)}\Big/\partial\rho_{uv}\right)$$

$$(q\geq s>t\geq 1;\ q\geq u>v\geq 1)$$

$$=\frac{N\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}(\mathrm{P}^{-1}\boldsymbol{y}\boldsymbol{y}'\mathrm{P}^{-1})_{st}(\mathrm{P}^{-1}\boldsymbol{y}\boldsymbol{y}'\mathrm{P}^{-1})_{uv}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)}{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)} \tag{30}$$

$$-\frac{N\left\{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}(\mathrm{P}^{-1}\boldsymbol{y}\boldsymbol{y}'\mathrm{P}^{-1})_{st}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)\right\}\times\left\{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}(\mathrm{P}^{-1}\boldsymbol{y}\boldsymbol{y}'\mathrm{P}^{-1})_{uv}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)\right\}}{\left\{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\exp\left(-\frac{1}{2}\boldsymbol{y}'\mathrm{P}^{-1}\boldsymbol{y}\right)\right\}^2}.$$

The estimated values of the asymptotic standard errors are obtained from the inverse of

$$
\begin{aligned}
I=\sum_{i=1}^{N}\Bigg[\Bigg\{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\frac{\partial\ln\{(\prod_{j=1}^{p}P(X_{ij}=x_{ij}\,|\,\boldsymbol{y},\boldsymbol{\delta}_j,\mu_j))A(\boldsymbol{y}\,|\,\mathrm{P})\}}{\partial\boldsymbol{\alpha}}f(\boldsymbol{y}\,|\,\boldsymbol{x}_i,\boldsymbol{\alpha})\Bigg\}\\
\times\Bigg\{\sum_{m1=1}^{r}\cdots\sum_{mq=1}^{r}\frac{\partial\ln\{(\prod_{j=1}^{p}P(X_{ij}=x_{ij}\,|\,\boldsymbol{y},\boldsymbol{\delta}_j,\mu_j))A(\boldsymbol{y}\,|\,\mathrm{P})\}}{\partial\boldsymbol{\alpha}'}f(\boldsymbol{y}\,|\,\boldsymbol{x}_i,\boldsymbol{\alpha})\Bigg\}\Bigg] \qquad (31)
\end{aligned}
$$

in the similar way to the orthogonal case.

Table 1

Contents of tests and result of factor analysis for continuous variables

Test	Sub-test	Task	M	Var	Var-M	Product-Moment Correlations				
A/C	A1	Letter search	.13	.25	.12	1.000				
	A2	Finding figures	1.45	2.74	1.29	.127	1.000			
	A3	Symmetric figure I	2.36	5.69	3.33	.128	.317	1.000		
	A4	Rearranging figure	2.22	4.69	2.47	.075	.192	.351	1.000	
	C1	Digit symbol	.83	5.62	4.79	.031	.146	.105	.165	1.000
	C2	Symmetric figure II	1.38	3.37	1.99	.064	.253	.411	.390	.151

Test	Sub-test	Normalized Varimax I (SE)	II (SE)	ψ_j (SE)	AIC
A/C	A1	.06 (.02)	.04 (.05)	.25 (.01)	
	A2	.35 (.42)	2.03 (3.06)	−1.51 (12.1)	
	A3	1.41 (.08)	.37 (.28)	3.56 (.21)	33,357.1
	A4	1.27 (.07)	.12 (.09)	3.06 (.17)	
	C1	.51 (.09)	.20 (.21)	5.32 (.20)	
	C2	1.21 (.06)	.17 (.03)	1.87 (.12)	

Test	Sub-test	Task	M	Var	Var-M	Product-Moment Correlations				
B	B1	Figure comparison	1.15	2.90	1.75	1.000				
	B2	Figure matching	.46	.84	.38	.120	1.000			
	B3	Figure classification	1.03	2.29	1.26	.335	.155	1.000		
	B4	Part and whole	.98	1.60	.62	.221	.148	.276	1.000	
	B5	Construction of square	1.93	3.03	1.10	.257	.174	.318	.346	1.000
	B6	Surface development	3.48	13.02	9.54	.229	.151	.280	.332	.378

Test	Sub-test	Normalized Varimax I (SE)	II (SE)	ψ_j (SE)	AIC
B	B1	.74 (.31)	.41 (.19)	2.19 (.32)	
	B2	.14 (.05)	.20 (.04)	.77 (.03)	
	B3	.96 (.38)	.38 (.12)	1.23 (.63)	33,138.5
	B4	.30 (.06)	.63 (.05)	1.11 (.06)	
	B5	.48 (.10)	.99 (.08)	1.82 (.11)	
	B6	.79 (.15)	2.02 (.15)	8.33 (.52)	

Note: M, Var = Mean and variance of the number of errors; I = The first factor; II = The second factor; ψ_j = Unique variance; SE = Standard error.

5. Numerical examples

Two data sets, Tests A/C ($N=1{,}495$) and Test B ($N=1{,}493$), will be used for numerical examples. These data consist of the numbers of errors in the subtests for the non-verbal intelligence tests for adults. The data were used by Ogasawara (1992, 1996) who showed the appropriateness of the Poisson model for these data. Table 1 contains the contents and descriptive statistics of the discrete variables. Though product-moment correlations are not appropriate for describing the associations among the discrete variables, they are provided as auxiliary information. Table 1 shows also the rotated results of usual factor analysis for continuous variables assuming a two-factor exploratory model, where the likelihood of the parameters using the multivariate normal distribution was maximized instead of ordinary Wishart likelihood for comparison with the AICs (Akaike Information Criterion; Akaike, 1973) of the Poisson models.

Several Poisson models with orthogonal factors have been constructed considering the following points: (1) the number of latent factors (one or two), (2) the intercept parameter μ_j (with or without it), (3) the types of the distributions of $\boldsymbol{\theta}_i$ (uniform, normal or log-normal), and (4) the number of lattice points in each

Table 2
Results of Tests A/C

Model	$a11$	$a12$		$a21$		
	Uni. (5p)	Uni. (5p)		Normalized Varimax Nml. (5p)		
Subtest	I	I	μ_j	I (SE)	II (SE)	μ_j (SE)
A1	−1.12	.50	−2.01	.58 (.066)	.11 (.069)	−2.15 (.077)
A2	.22	.53	.36	.49 (.024)	.20 (.033)	.26 (.025)
A3	.43	.66	.81	.72 (.022)	.04 (.029)	.64 (.026)
A4	.40	.57	.77	.57 (.020)	.17 (.029)	.65 (.023)
C1	.004	1.76	−.95	.40 (.042)	1.93 (.051)	−1.66 (.089)
C2	.24	.90	.17	.87 (.031)	.26 (.042)	−.01 (.036)
AIC	29,289.7	27,002.9		25,827.6		

Model	$a31$			$a32$		
	Oblique Model (5p)			Oblique Model (5p)		
Subtest	I (SE)	II (SE)	μ_j (SE)	I (SE)	II (SE)	μ_j (SE)
A1	.70 (.067)	0 (fixed)	−2.21 (.090)	.65 (.061)	0 (fixed)	−2.20 (.086)
A2	−.03 (.036)	.55 (.031)	.26 (.025)	.45 (.025)	.32 (.044)	.26 (.025)
A3	.22 (.033)	.57 (.030)	.64 (.025)	.70 (.024)	.23 (.054)	.63 (.027)
A4	−.003 (.032)	.60 (.027)	.66 (.022)	.50 (.023)	.34 (.046)	.66 (.023)
C1	−1.95 (.052)	2.13 (.049)	−1.55 (.081)	0 (fixed)	2.03 (.047)	−1.61 (.085)
C2	0 (fixed)	.92 (.029)	−.015 (.035)	.77 (.035)	.50 (.062)	−.007 (.035)
ρ_{12} (SE)	.64 (.023)			−.07 (.087)		
AIC	25,843.1			25,843.6		

Note: Uni. = Uniform distribution; Nrm. = Normal distribution; 5p = 5 points; I = The first factor; II = The second factor; μ_j = Intercept; SE = Standard error.

dimension of latent factors (five or ten).

Table 2 gives the results for Tests A/C. The one-factor models have been obtained by assuming a uniform distribution with mean zero and unit variance which is approximated by five points $\boldsymbol{y}'=(-1.414, -.707, 0, .707, 1.414)$ with corresponding weights$=(.2, .2, .2, .2, .2)$. For the approximated normal distribution for each factor for Model $\alpha 21$, $\boldsymbol{y}'=(-2, -1, 0, 1, 2)$ with weights$=(.054, .244, .403, .244, .054)$ were used. Table 2 also contains the results of confirmatory oblique models, where two factor-loadings were set to zero, which satisfies the identification of the model. The five lattice points $(-2, -1, 0, 1, 2)$ were employed in each dimension of oblique factors with weights computed by (22).

The large difference of the values of the AICs of Model $\alpha 12$ with intercept parameter and Model $\alpha 11$ without it indicates the advantage of introducing the intercept parameter for the data. The estimated values of μ_j seem to correspond to the values of M in Table 1, which can be interpreted as the difficulty of a task. The loadings of the factor I in Model $\alpha 12$ correspond to the variation caused by individual differences ($Var-M$ in Table 1) in the data with overdispersion (the case when $Var > M$; see e.g., Van Duijn, 1993).

For the orthogonal Model $\alpha 21$, the initial factor-loadings with δ_{12} fixed at zero were rotated by the normalized varimax method. The rotated loadings of Model $\alpha 21$ accompany their standard errors, which were obtained by the augmented information matrix for the maximum likelihood estimator with restrictions. The ratios of the loadings of the factor I to those of the factor II are relatively large except for that of Subtest C1. On the other hand, only Subtest C1 has a large loading in the factor II. This result may come from the fact that the value $Var-M$ of Subtest C1 in Table 1 is extremely large compared to M. In the oblique models, $\alpha 31$ and $\alpha 32$, the large values of the loadings for Subtest C1 correspond to those in the orthogonal models.

Table 3
Results of one-factor model for Test B

Model	$\beta 11$ Uni. (5p)	$\beta 12$ Uni. (5p)		$\beta 13$ Uni. (10p)		$\beta 14$ Nrm. (5p)		$\beta 15$ LN (5p)	
Subtest	I	I	μ_j	I	μ_j	I	μ_j	I	μ_j
B1	.14	.75	−.01	.75	−.04	.75	−.10	.70	−.08
B2	−.32	.68	−.90	.68	−.92	.65	−.96	.60	−.95
B3	.09	.75	−.11	.75	−.15	.77	−.22	.70	−.19
B4	.07	.67	−.13	.67	−.15	.66	−.20	.61	−.18
B5	.34	.50	.61	.50	.59	.51	.55	.47	.56
B6	.60	.79	1.08	.79	1.05	.78	.99	.72	1.01
AIC	29,441.8	27,855.8		27,798.9		27,645.4		27,705.5	

Note: Uni.=Uniform distribution; Nrm.=Normal distribution; LN=Log-normal distribution; 5p=5 points; 10p=10 points; I=The first factor; II=The second factor; μ_j=Intercept.

The results of factor analysis assuming the multivariate normal distribution for the observed discrete variables give the large value of AIC for Tests A/C (Table 1). In the rotated factor pattern (Table 1), the second factor has a large loading only for Subtest A2, which is different from the result of the Poisson model $\alpha 21$. The standard errors of the loadings for A2 (Table 1) are also large. Further, the estimated value of the uniqueness for A2 is negative (Heywood case), which indicates inappropriateness of the ordinary factor analysis for the data.

For the usual factor analysis assuming the multivariate normal distribution for response variables, the estimated values of factor loadings depend only on the sample variance-covariance matrix of manifest variables and are independent of

Table 4
Results of two-factor model for Test B

Model / Subtest	$\beta 21$, Nrm. (5p), Normalized Varimax I (SE)	II (SE)	μ_j (SE)
B1	.96 (.033)	.19 (.025)	−.27 (.043)
B2	.58 (.045)	.37 (.044)	−.99 (.046)
B3	.76 (.032)	.37 (.030)	−.28 (.040)
B4	.45 (.039)	.50 (.037)	−.21 (.035)
B5	.34 (.031)	.38 (.027)	.54 (.024)
B6	.18 (.020)	.84 (.020)	.92 (.028)
AIC	27,179.1		
Model / Subtest	**$\beta 31$ Oblique Model (5p) I (SE)**	**II (SE)**	**μ_j (SE)**
B1	.99 (.032)	0 (fixed)	−.28 (.043)
B2	.53 (.049)	.29 (.047)	−.99 (.046)
B3	.73 (.036)	.25 (.034)	−.28 (.039)
B4	.35 (.041)	.44 (.039)	−.21 (.034)
B5	.27 (.031)	.34 (.029)	.54 (.023)
B6	0 (fixed)	.85 (.020)	.92 (.027)
ρ_{12} (SE)	.40 (.042)		
AIC	27,175.8		
Model / Subtest	**$\beta 32$ Oblique Model (5p) I (SE)**	**II (SE)**	**μ_j (SE)**
B1	.93 (.030)	0 (fixed)	−.22 (.040)
B2	.73 (.043)	0 (fixed)	−1.01 (.049)
B3	.88 (.030)	0 (fixed)	−.29 (.040)
B4	.41 (.045)	.37 (.044)	−.21 (.034)
B5	.31 (.034)	.28 (.032)	.54 (.023)
B6	0 (fixed)	.86 (.020)	.92 (.028)
ρ_{12} (SE)	.53 (.034)		
AIC	27,194.3		

Note: Nrm. = Normal distribution; 5p = 5 points; I = The first factor; II = The second factor; μ_j = Intercept; SE = Standard error.

their means. Though the value of *Var-M* in Subtest C1, for which the large loadings were obtained by the Poisson model, is the largest among the subtests, the value *Var* (5.62) is not the largest. The difference of the factor patterns between the Poisson and normal models is emphasized by these data properties.

Table 3 shows the results of one-factor models for Test B. Model $\beta 11$ has no intercept parameter and gives a large value of AIC. Although the number of lattice points in Model $\beta 13$ ($\boldsymbol{y}'=(\pm 1.567, \pm 1.219, \pm .870, \pm .522, \pm .174)$ with weights$=(.1, .1, \cdots, .1)$) is different from that of Model $\beta 12$, the estimated values of the factor-loadings and intercept parameter are similar. In Model $\beta 14$ the standardized normal distribution for the factor is assumed. Model $\beta 15$ employs a skewed distribution with mean zero and unit variance based on the log-normal distribution using the values, ln1, ln2, ln3, ln4 and ln5 ($\boldsymbol{y}'=(-1.354, -.502, .351, 1.203, 2.055)$, with weights$=(.185, .344, .248, .144, .079)$). The estimated values are fairly similar to each other among Models $\beta 12$, $\beta 13$, $\beta 14$ and $\beta 15$.

Table 4 gives the results of the two-factor models. The results of Model $\beta 21$ show that the two subtest-groups (B1, B2, B3) and (B4, B5, B6) are obtained. It is interesting that the two groups also correspond to the grouping of the number-right scores for these subtests which were obtained by the factor analysis for continuous variables (Ogasawara, 1990). The rotated factor-pattern in Table 4 appears stable when compared to the estimated values of the standard errors. Model $\beta 31$ is a just identified oblique model. Model $\beta 32$ is the model in which two additional loadings were set equal to zero, where the increase of AIC from that of Model $\beta 31$ is observed.

In the case of Test B, the result of factor analysis for continuous variables (Table 1) is similar to that of the Poisson model $\beta 21$ with the difference of the pattern for Subtest B2. However, judging from the values of the estimated standard errors of the factor loadings in Table 1, the results of Table 1 appear more unstable than those in Table 4.

6. Discussion

The results of Table 3 show that the estimated values of the parameters do not depend so much upon the assumptions of the types of the distributions of factors, which has also been indicated for the results of IRT models by Bock and Aitkin (1981) and Bartholomew (1988). We conjecture that this is one of the robust characteristics of the marginal maximum likelihood estimation in latent-variable models.

The method for estimating the correlations of latent factors was developed for our Poisson model. But, in the case of ordered categorical variables (dichotomous or polytomous variables), the method for estimating the correlations among factors have been developed by Christoffersson (1975), Muthén (1978) and Lee, Poon, and Bentler (1990, 1992, 1995). These are mostly based on the method of generalized least squares or partitioned maximum likelihood. Our method is based on the full information marginal maximum likelihood, though numerical integration is used. In the case of ordered categories our method may not have practical advantage. However, it is to be noted that in principle our method can be applied to any types of distributions for response variables as are used in the generalized linear model.

Appendix

The marginal mean and variance of the Poisson variable in the orthogonal model

In order to confirm the identifiability of the model of (5), we derive the marginal mean and variance of X_{ij}. If we assume the exchangeability of the orders of integration and summation,

$$\begin{aligned} E(X_{ij}) &= \sum_{x_{ij}=0}^{\infty} \int_{R(\boldsymbol{\theta}_i)} x_{ij} \exp(x_{ij}\lambda_{ij}) \exp\{-\exp(\lambda_{ij})\} \frac{1}{x_{ij}!} h(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i \\ &= \int_{R(\boldsymbol{\theta}_i)} \exp(\lambda_{ij}) \sum_{(x_{ij}-1)=0}^{\infty} \exp\{(x_{ij}-1)\lambda_{ij}\} \exp\{-\exp(\lambda_{ij})\} \frac{1}{(x_{ij}-1)!} h(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i \\ &= \int_{R(\boldsymbol{\theta}_i)} \exp(\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \mu_j) h(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \quad (i=1, \cdots, N;\ j=1, \cdots, p) \qquad \text{(A1)} \end{aligned}$$

is obtained, where $R(\boldsymbol{\theta}_i)$ is the region of integration with respect to $\boldsymbol{\theta}_i$. Similarly, we have

$$E(X_{ij}(X_{ij}-1)) = \int_{R(\boldsymbol{\theta}_i)} \exp\{2(\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \mu_j)\} h(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \quad (i=1, \cdots, N;\ j=1, \cdots, p). \qquad \text{(A2)}$$

As an example of actual expression of $h(\boldsymbol{\theta}_i)$, we derive the results for the case when θ_{ik} is independently and normally distributed as $N(\nu_k, \sigma_k^2)$, $(k=1, \cdots, q)$.
From (A1) and (A2),

$$\begin{aligned} E(X_{ij}) &= \exp(\mu_j) \prod_{k=1}^{q} \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}\sigma_k} \exp\left\{-\frac{(\theta_{ik}-\nu_k)^2}{2\sigma_k^2} + \theta_{ik}\delta_{jk}\right\} d\theta_{ik} \\ &= \exp\left\{\mu_j + \sum_{k=1}^{q} \left(\nu_k \delta_{jk} + \frac{\sigma_k^2 \delta_{jk}^2}{2}\right)\right\}, \quad (i=1, \cdots, N;\ j=1, \cdots, p), \qquad \text{(A3)} \end{aligned}$$

and

$$\begin{aligned} Var(X_{ij}) &= E(X_{ij}(X_{ij}-1)) + E(X_{ij}) - \{E(X_{ij})\}^2 \\ &= \exp(2\mu_j) \prod_{k=1}^{q} \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}\sigma_k} \exp\left\{-\frac{(\theta_{ik}-\nu_k)^2}{2\sigma_k^2} + 2\theta_{ik}\delta_{jk}\right\} d\theta_{ik} \\ &\quad + E(X_{ij}) - \{E(X_{ij})\}^2 \\ &= \exp\{2(\mu_j + \sum_{k=1}^{q} (\nu_k \delta_{jk} + \sigma_k^2 \delta_{jk}^2))\} + \exp\left(\mu_j + \sum_{k=1}^{q} \left(\nu_k \delta_{jk} + \frac{\sigma_k^2 \delta_{jk}^2}{2}\right)\right) \\ &\quad - \exp\left\{2\left(\mu_j + \sum_{k=1}^{q} \left(\nu_k \delta_{jk} + \frac{\sigma_k^2 \delta_{jk}^2}{2}\right)\right)\right\}, \quad (i=1, \cdots, N;\ j=1, \cdots, p), \quad \text{(A4)} \end{aligned}$$

follow. If we consider the reparametrization from (ν_k, σ_k) to (ν_k', σ_k') such that $\nu_k =$

$\dfrac{\nu_k'-a_k}{b_k}$ and $\sigma_k=\dfrac{\sigma_k'}{b_k}$ with arbitrary a_k and $b_k(b_k>0)$, the values of $E(X_{ij})$ and $Var(X_{ij})$ are still unchanged when $\mu_j'=\mu_j-\sum_{k=1}^{q}\dfrac{a_k}{b_k}\delta_{jk}$ and $\delta_{jk}'=\dfrac{\delta_{jk}}{b_k}$ are used as new parameters replacing μ_j and δ_{jk}, respectively. This indicates that the location and dispersion of the distribution of θ_{ik} are indeterminate. Thus, we assume $\nu_k=0$ and $\sigma_k=1$, $(k=1,\cdots,q)$ without loss of generality. For this case, we have

$$E(X_{ij})=\exp\left(\mu_j+\frac{\sum_{k=1}^{q}\delta_{jk}^2}{2}\right)$$

and

$$Var(X_{ij})=\exp\left\{2\left(\mu_j+\sum_{k=1}^{q}\delta_{jk}^2\right)\right\}+\exp\left(\mu_j+\frac{\sum_{k=1}^{q}\delta_{jk}^2}{2}\right)-\exp\left(2\mu_j+\sum_{k=1}^{q}\delta_{jk}^2\right). \qquad (A5)$$

Acknowledgment

Part of the work of this paper was completed when the author was a researcher at Railway Technical Research Institute. The author is indebted to anonymous referees for their comments and suggestions which have led to the improvement of the manuscript.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.). *Second International Symposium on Information Theory* (pp. 267-281). Budapest: Akademiai Kiado.

Bartholomew, D.J. (1988). The sensitivity of latent trait analysis to choice of prior distribution. *British Journal of Mathematical and Statistical Psychology*, **41**, 101-107.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, Massachusetts: MIT Press.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, **46**, 443-459.

Bock, R.D., Gibbons, R., & Muraki, E. (1988). Full information item factor analysis. *Applied Psychological Measurement*, **12**, 261-280.

Böckenholt, U. (1993). Estimating latent distributions in recurrent choice data. *Psychometrika*, **58**, 489-509.

Choulakian, V. (1996). Generalized bilinear models. *Psychometrika*, **61**, 271-283.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Rsychometrika*, **40**, 5-32.

Clogg, C.C. (1986). Discussion of paper by L.A. Goodman (1986) Some useful extension of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review*, **54**, 243-309.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

Goodman, L.A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Annals of Statistics*, **13**, 10-69.

Goodman, L.A. (1986). Some useful extension of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables (with discussion). *International Statistical Review*, **54**, 243-309.

Goodman, L.A. (1991). Measures, models, and graphical displays in the analysis of cross-classified data (with discussion). *Journal of the American Statistical Association*, **86**, 1085-1138.

Harwell, M.R., Baker, F.B., & Zwarts, M. (1988). Item parameter estimation via marginal maximum likelihood and an EM algorithm : A didactic. *Journal of Educational Statistics*, **13**, 243-271.

Jansen, M.G.H. (1986). A Bayesian version of Rasch's multiplicative Poisson model for the number of errors on an achievement test. *Journal of Educational Statistics*, **11**, 147-160.

Jansen, M.G.H., & Van Duijn, M.A.J. (1992). Extensions of Rasch's multiplicative Poisson model. *Psychometrika*, **57**, 405-414.

Jennrich, R.I. (1974). Simplified formulae for standard errors in maximum-lkikelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, **27**, 122-131.

Kruskal, J.B. (1981). Multilinear models for data analysis. *Behaviormetrika*, **10**, 1-20.

Lee, S.Y., Poon, W.Y., & Bentler, P.M. (1990). A three-stage estimation procedure for structural equation models with polytomous variables. *Psychometrika*, **55**, 45-51.

Lee, S.Y., Poon, W.Y., & Bentler, P.M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika*, **57**, 89-105.

Lee, S.Y., Poon, W.Y., & Bentler, P.M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology*, **48**, 339-358.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 226-233.

McCullagh, P., & Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). London : Chapman & Hall.

Mellenbergh, G.J. (1994). Generalized linear item response theory. *Psychological Bulletin*, **115**, 300-307.

Mislevy, R.J. (1984). Estimating latent distributions. *Psychometrika*, **49**, 359-381.

Mislevy, R.J. (1985). Estimation of latent group effects. *Journal of the American Statistical Associations*, **80**, 993-997.

Mislevy, R.J., & Sheehan, K.M. (1989). Information matrices in latent-variable models. *Journal of Educational Statistics*, **14**, 335-359.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, **43**, 551-560.

Ogasawara, H. (1990). Covariance structure model when the factor means and the covariances are functions of the third variable. *Japanese Psychological Research*, **32**, 19-25.

Ogasawara, H. (1992). Models of the number of errors using structured parameters in a generalized Poisson distribution and the Polya-Eggenberger distribution. *Japanese Journal of Behaviormetrics*, **19**(2), 1-13. (in Japanese)

Ogasawara, H. (1996). Rasch's multiplicative Poisson model with covariates. *Psychometrika*, **61**, 73-92.

Ogasawara, H. (1998). Standard errors of several indices for unrotated and rotated factors. *Economic Review, Otaru University of Commerce*, **49**(1), 21-69.

Rasch, G. (1973). Two applications of the multiplicative Poisson models in road accidents statistics. *Bulletin of the International Statistical Institute. Proceedings of the 39th session*, Vienna, 31-43.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago : The University of Chicago Press. (original work published in 1960)

Silvey, S.D. (1975). *Statistical inference*. London : Chapman & Hall.

Van Duijn, M.A.J. (1993). *Mixed models for repeated count data*. Leiden : DSWO Press.

(Received January 1997, Revised September 1998)