Behaviormetrika, Vol. 25, No. 1, 1-12
1998

A FACTOR ANALYSIS MODEL FOR A MIXTURE OF VARIOUS TYPES OF VARIABLES[1)]

Haruhiko Ogasawara*

A factor analysis model is proposed for the case of a mixture of various types of discrete and continuous manifest variables. It is indicated that the likelihood of parameters can be described for a mixture of different types of distributions by assuming local independence. For estimation of the parameters of interest, the method of marginal maximum likelihood is used, where scores of latent factors are integrated out from the likelihood. A kind of the EM algorithm is utilized for optimization. As an example, the case of a mixture of normal, binomial and Poisson distributions is provided.

1. Introduction

Factor analysis was developed as a latent variable model for continuous variables. Theories of statistical inference for factor analysis have been provided, in which usually multivariate normal distribution is assumed for latent factors (e.g., Lawley & Maxwell, 1971).

The aim of factor analysis resides in that the relationships among manifest variables are to some extent explained by latent variables. From this purpose, manifest variables should not be limited to normally distributed variables. However, while the covariance matrix of normally distributed variables constitutes a sufficient statistic for the parameters in a usual factor analysis model, the situation is more complicated in the case of non-normal or discrete variables. In the latter case, individual theories have been developed.

For example, the IRT (item response theory) model is a kind of factor analysis model. This corresponds to the fact that it was formerly called latent trait model (Lord & Novick, 1968). Christoffersson (1975) used the term, factor analysis, in the title of the paper concerning dichotomous variables. Though a large part of the works of IRT has been concerned with dichotomous variables with unidimensional assumption, the theories and methods for graded responses (Samejima, 1969), multifactors (Christoffersson, 1975; Muthên, 1978) and nominal responses (Bock, 1972) have been developed. These also can be regarded as factor analysis models in the sense that latent variables are assumed behind manifest variables in a factor-analytic form.

One of the problems in treating various types of manifest variables is how to

---

Key Words and Phrases; factor analysis, marginal maximum likelihood method, EM algorithm, communality, Poisson distribution

* Otaru University of Commerce, 3-5-21, Midori, Otaru, 047-8501 Japan, e-mail: hogasa@otaru-uc.ac.jp

1) Part of this work was completed when the author was a researcher at Railway Technical Research Institute.

deal with a mixture of them. There have been two types of approaches. One is Muthén's method (Muthén, 1983, 1984; Muthén & Sattora, 1995). This method decomposes the process of estimating parameters into stages; firstly, thresholds and related parameters are estimated and secondly, structural parameters (i.e., factor loadings) are estimated by the generalized least squares (GLS). For this method, the computer program LISCOMP has been provided.

The second approach is based on the estimation of polychoric and polyserial correlations (Lee & Poon, 1986, 1987; Poon & Lee, 1987; Jöreskog, 1994) which are generalizations of tetrachoric and biserial correlations. Structural parameters are estimated simultaneously with thresholds or successively by the method of maximum likelihood (ML) or GLS (Lee, Poon & Bentler, 1990a, 1990b, 1992).

In these methods, there is a basic notion of thresholds. Using Muthén's terminology, the link of thresholds and corresponding manifest variables is "outer measurement model", and the model which explains the variation of "outer" continuous variables by assuming latent factors is "inner measurement model". Hence, in these models, only ordered categorical variables are considered among discrete variables.

In this paper, a method which deals with other types of variables including ordered categories is proposed.

2. General description of model

One of the situations where various types of continuous manifest variables coexist with various types of discrete manifest variables is the one in which subjects take different types of psychological tests. For instance, a data matrix may consist of measurements of response times, dichotomous scores (failure or success), polytomous scores (rating scales) and nominal responses (choices of alternatives) in addition to ordinary continuous scores. These scores can be explained partly by assuming a common latent trait, which is empirically accepted as a general ability. The assumption of several distinct abilities depending on the types of measurements may also be reasonable apart from the multifactor theory of intelligence.

Let x_{ij} be the score of the i-th ($i=1, \cdots, N$) subject for the j-th ($j=1, \cdots, p$) continuous or discrete variable. Let $\boldsymbol{\theta}_i=(\theta_{i1}, \cdots, \theta_{iq})'$ denote the scores of q latent variables for the i-th subject. Let $\boldsymbol{\delta}_j=(\delta_{j1}, \cdots, \delta_{jm_j})'$ be the vector of the m_j parameters which determine the distribution of x_{ij} given $\boldsymbol{\theta}_i$. When x_{ij} is a score of a continuous variable, let the density function of x_{ij} be denoted by $f_j(x_{ij} \mid \boldsymbol{\theta}_i, \boldsymbol{\delta}_j)$. When x_{ij} is a score of a discrete variable, the probability function of x_{ij} is written as $P_j(x_{ij} \mid \boldsymbol{\theta}_i, \boldsymbol{\delta}_j)$.

It is assumed that $\boldsymbol{\theta}_i$, ($i=1, \cdots, N$) are distributed identically and independently of each other. In addition, $\theta_{i1}, \cdots, \theta_{iq}$ are assumed to be distributed identically and independently of each other. In these distributions, the parameters in the density function of $\boldsymbol{\theta}_i$, $h(\boldsymbol{\theta}_i)$, are assumed to be known and to have fixed values. These assumptions are not requisites for identification of the model which will be proposed. However, since location and dispersion of latent variables are arbitrary in most of cases and orthogonality of latent variables frequently brings about simplicity of description, the above assumptions are adopted. In ordinary factor analysis models, the elements of $\boldsymbol{\theta}_i$ correspond to factor scores which are not of great interest except in the case we are interested in individual subjects. In this paper emphasis is more on estimation of factor loadings than individual factor scores. Consequently, $\boldsymbol{\theta}_i$, $(i=1, \cdots, N)$ will be treated as nuisance parameters.

While the relationships among manifest variables are explained by $\boldsymbol{\theta}_i$ which are common to $f_j(\cdot)$ and $P_j(\cdot)$, the manifest variables are supposed to be independent of each other given $\boldsymbol{\theta}_i$ (local independence). Therefore, the likelihood of $\boldsymbol{\delta}=(\boldsymbol{\delta}_1', \cdots, \boldsymbol{\delta}_p')'$ based on the i-th observation, L_i^*, is described as:

$$L_i^*(\boldsymbol{\delta} \mid \boldsymbol{x}_i, \boldsymbol{\theta}_i)=\{\prod_{j=1}^{p_1} f_j(x_{ij} \mid \boldsymbol{\theta}_i, \boldsymbol{\delta}_j)\}\{\prod_{j=p_1+1}^{p} P_j(x_{ij} \mid \boldsymbol{\theta}_i, \boldsymbol{\delta}_j)\}, \tag{1}$$

where $\boldsymbol{x}_i=(x_{i1}, \cdots, x_{ip})'$; p_1 continuous variables are supposed to be in the first part of the manifest variables and $p_2(=p-p_1)$ discrete variables are located in the second part. Though, probably same types of probability (density) functions may be in equation (1), they are not differentiated.

The likelihood of $\boldsymbol{\delta}$ based on N independent observations, $L^{\#}$, is

$$L^{\#}=\prod_{i=1}^{N} L_i^*(\boldsymbol{\delta} \mid \boldsymbol{x}_i, \boldsymbol{\theta}_i). \tag{2}$$

To obtain estimates of $\boldsymbol{\delta}$, $L^{\#}$ should be maximized. However, since the unknown (nuisance) parameters $\boldsymbol{\theta}_i$, $(i=1, \cdots, N)$ are involved in (2), this can not be obtained directly. So, we maximize the following marginal maximum likelihood which is derived by integrating out $\boldsymbol{\theta}_i$ which are treated as random variables:

$$L^*=\prod_{i=1}^{N} \int_{R(\theta_i)} L_i^*(\boldsymbol{\delta} \mid \boldsymbol{x}_i, \boldsymbol{\theta}_i) h(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \tag{3}$$

where $h(\boldsymbol{\theta}_i)=\prod_{k=1}^{q} e(\theta_{ik})$ and no unknown parameter is in $e(\theta_{ik})$ which is a common distribution to the q distributions of factors. For instance, the uniform distribution of the range $[0, 1]$ and the normal distribution with mean zero and unit variance may be realistic ones. However, note that the type of the distribution of a latent variable in a multifactor model might change by factor rotation without the assumption of normality or sphericity on $\boldsymbol{\theta}_i$.

For $f_j(\cdot)$, the normal distribution has been most frequently used. However, the gamma or log-normal distribution may be used for measures of response times (Ogasawara, 1995). While the distribution functions of the normal and logistic distributions have been used for dichotomous or ordered categorical variables, the multivariate logit (Bock, 1972) has been used for nominal variables. In addition, Poisson distribution is an appropriate one for the distribution of frequencies (e.g., the number of errors; Resch, 1980; Ogasawara, 1996a).

3. Estimation of parameters by the EM algorithm

It is often difficult to obtain the result of integration in (3). For such cases, the numerical integration can be used as an approximation. In IRT, Bock and Aitkin (1981) showed that the marginal maximum likelihood estimation became simple by using a kind of the EM (Expectation and Maximization) algorithm (Dempster, Laird & Rubin, 1977). After this work, the method has been applied to various data and various types of manifest variables, which has shown usefulness of the algorithm (Bock, Gibbons & Muraki, 1988; Muraki, 1990, 1992; Ogasawara, 1996b).

Estimation of parameters in (3) by the EM algorithm is described in the following. The marginal likelihood is

$$\begin{aligned} L^* \cong L &= \prod_{i=1}^{N} g(\boldsymbol{x}_i \mid \boldsymbol{\delta}) \\ &= \prod_{i=1}^{N} \sum_{m_1=1}^{r} \cdots \sum_{m_q=1}^{r} L_i(\boldsymbol{\delta}_i \mid \boldsymbol{x}_i, \boldsymbol{y}) B(\boldsymbol{y}), \end{aligned} \tag{4}$$

where $\boldsymbol{y}=(y_{m_1}, \cdots, y_{m_q})'$ is the vector which represents a lattice point in the q-dimensional factor space whose density function is $h(\boldsymbol{\theta}_i)$. The scalar $B(\boldsymbol{y})=\prod_{k=1}^{q} A(y_{m_k})$ is the weight proportional to $h(\boldsymbol{y})$.

The partial derivative of ln L in (4) with respect to $\boldsymbol{\delta}_j$ is

$$\frac{\partial \ln L}{\partial \boldsymbol{\delta}_j} = \sum_{i=1}^{N} \sum_{m_1=1}^{r} \cdots \sum_{m_q=1}^{r} \frac{\partial \ln L_i(\boldsymbol{\delta} \mid \boldsymbol{x}_i, \boldsymbol{y})}{\partial \boldsymbol{\delta}_j} \frac{L_i(\boldsymbol{\delta} \mid \boldsymbol{x}_i, \boldsymbol{y}) B(\boldsymbol{y})}{g(\boldsymbol{x}_i \mid \boldsymbol{\delta})}, \quad (j=1, \cdots, p_1).$$

If we denote $L_i(\boldsymbol{\delta} \mid \boldsymbol{x}_i, \boldsymbol{y}) B(\boldsymbol{y}) / g(\boldsymbol{x}_i \mid \boldsymbol{\delta})$ by $\phi(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta})$, $\phi(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta})$ is the posterior probability of $\boldsymbol{y}$ given the values of $\boldsymbol{x}_i$ and $\boldsymbol{\delta}$. Using this, above equation becomes

$$= \sum_{i=1}^{N} \sum_{m_1=1}^{r} \cdots \sum_{m_q=1}^{r} \frac{\partial \ln f_j(x_{ij} \mid \boldsymbol{y}, \boldsymbol{\delta}_j)}{\partial \boldsymbol{\delta}_j} \phi(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta}). \tag{5}$$

Supposing that $\phi(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta})$ are given, the information matrix is described as:

$$E\left(\frac{-\partial^2 \ln L}{\partial \boldsymbol{\delta}_j \partial \boldsymbol{\delta}_j'}\right) = \sum_{i=1}^{N} \sum_{m_1=1}^{r} \cdots \sum_{m_q=1}^{r} E\left(\frac{-\partial^2 \ln f_j(x_{ij} \mid \boldsymbol{y}, \boldsymbol{\delta}_j)}{\partial \boldsymbol{\delta}_j \partial \boldsymbol{\delta}_j'}\right) \phi(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta}), \quad (j=1, \cdots, p_1). \tag{6}$$

When $j=p_1+1, \cdots, p$, $f_j(\cdot)$ in (5) and (6) should be replaced by $P_j(\cdot)$.

From (5) and (6), the estimate of $\boldsymbol{\delta}_j$ given $\phi(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta})$ is iteratively obtained as follows:

$$\boldsymbol{\delta}_{j(i+1)} = \boldsymbol{\delta}_{j(i)} + \left\{ E\left(\frac{-\partial^2 \ln L}{\partial \boldsymbol{\delta}_{j(i)} \partial \boldsymbol{\delta}_{j(i)}'}\right) \right\}^{-1} \frac{\partial \ln L}{\partial \boldsymbol{\delta}_{j(i)}}, \tag{7}$$

where the subscript (i) indicates that the value is at the i-th iteration. Though the value of $\phi(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta})$ in (6) is regarded as known, actually it contains unknown parameters. So, starting with some initial values of $\boldsymbol{\delta}$, the values of $\boldsymbol{\delta}_j$ are improved by (7). Using the improved values, $\phi(\boldsymbol{y} \mid \boldsymbol{x}_i, \boldsymbol{\delta})$ is renewed in the iteration in (7). These procedures should be repeated until estimated values are stable.

The inverse of the information matrix in (6) is the estimator of the asymptotic covariance matrix of the parameters when $\boldsymbol{y}$ are known. However, since in the iterative procedure $\boldsymbol{y}$ were actually unknown, the following estimate should be used as an approximation to the exact information matrix (Louis, 1982):

$$I=\sum_{i=1}^{N}\left[\left\{\sum_{m_1=1}^{r}\cdots\sum_{m_q=1}^{r}\frac{\partial \ln L_i(\boldsymbol{\delta}\mid\boldsymbol{x}_i,\boldsymbol{y})}{\partial\boldsymbol{\delta}}\phi(\boldsymbol{y}\mid\boldsymbol{x}_i,\boldsymbol{\delta})\right\}\right.$$
$$\left.\times\left\{\sum_{m_1=1}^{r}\cdots\sum_{m_q=1}^{r}\frac{\partial \ln L_i(\boldsymbol{\delta}\mid\boldsymbol{x}_i,\boldsymbol{y})}{\partial\boldsymbol{\delta}'}\phi(\boldsymbol{y}\mid\boldsymbol{x}_i,\boldsymbol{\delta})\right\}\right]. \tag{8}$$

4. Cases of the normal, binomial and Poisson distributions

In the former sections only general results were given. This was because in principle $f_j(\cdot)$ and $P_j(\cdot)$ can be any types of probability (density) functions. In this section, results of a mixture of p_1 normal distributions, p_2 binomial distributions and p_3 Poisson distributions are derived.

(1) normal distribution

Suppose that given $\boldsymbol{\theta}_i$, x_{ij} is normally distributed with mean $\boldsymbol{\theta}_i'\boldsymbol{\alpha}_j+\mu_j$ and variance ψ_j^2, where $\boldsymbol{\alpha}_j$ is the vector of factor loadings of the j-th manifest variable on q common factors, μ_j is the marginal mean of x_{ij} and ψ_j^2 corresponds to the variance of the j-th unique factor. If each element of $\boldsymbol{\theta}_i$ is independently distributed with mean zero and unit variance, the proportion of the communality of the j-th variable is given by $\boldsymbol{\alpha}_j'\boldsymbol{\alpha}_j/(\boldsymbol{\alpha}_j'\boldsymbol{\alpha}_j+\psi_j^2)$.

The function $f_j(\cdot)$ is described as:

$$f_j(x_{ij}\mid\boldsymbol{\theta}_i,\boldsymbol{\delta}_j)=\frac{1}{\sqrt{2\pi}\,\psi_j}\exp\left\{-\frac{(x_{ij}-\boldsymbol{\theta}_i'\boldsymbol{\alpha}_j-\mu_j)^2}{2\psi_j^2}\right\},\quad (j=1,\cdots,p_1), \tag{9}$$

where $\boldsymbol{\delta}_j=(\boldsymbol{\alpha}_j',\mu_j,\psi_j)'$. From (9), $\partial\ln f_j(\cdot)/\partial\boldsymbol{\delta}_j$ in (5) is

$$\frac{\partial\ln f_j(\cdot)}{\partial(\boldsymbol{\alpha}_j',\mu_j)'}=\frac{x_{ij}-\boldsymbol{y}'\boldsymbol{\alpha}_j-\mu_j}{\psi_j^2}\begin{bmatrix}\boldsymbol{y}\\1\end{bmatrix},$$
$$\frac{\partial\ln f_j(\cdot)}{\partial\psi_j}=-\frac{1}{\psi_j}+\frac{(x_{ij}-\boldsymbol{y}'\boldsymbol{\alpha}_j-\mu_j)^2}{\psi_j^3}, \tag{10}$$

and $E(\cdot)$ in the right-hand side of (6) becomes

$$E\left(\frac{-\partial^2\ln f_j(\cdot)}{\partial\boldsymbol{\delta}_j\partial\boldsymbol{\delta}_j'}\right)=\begin{bmatrix}\boldsymbol{y}\boldsymbol{y}' & \boldsymbol{y} & 0\\ \boldsymbol{y}' & 1 & 0\\ 0 & 0 & 2\end{bmatrix}\frac{1}{\psi_j^2}. \tag{11}$$

The results of the case when all of manifest variables are normally distributed (Rubin & Thayer, 1982) apply here. For this case, the iterative process of (7) is unnecessary. Inserting the results of (10) into (5), $\boldsymbol{\delta}_j$ are obtained algebraically. The solution corresponds to the case of multiple regression analysis, where $\boldsymbol{y}$ are observed independent variables with the weight $\phi(\boldsymbol{y}\mid\boldsymbol{x}_i,\boldsymbol{\delta})$ for the i-th observation. Thus,

$$(\boldsymbol{\alpha}_j', \mu_j)' = \left[\sum_{i=1}^{N}\sum_{m_1=1}^{r}\cdots\sum_{m_q=1}^{r}\begin{pmatrix}\boldsymbol{y}\boldsymbol{y}' & \boldsymbol{y}\\ \boldsymbol{y}' & 1\end{pmatrix}\phi(\boldsymbol{y}\,|\,\boldsymbol{x}_i, \boldsymbol{\delta})\right]^{-1}$$
$$\times\sum_{i=1}^{N}\sum_{m_1=1}^{r}\cdots\sum_{m_q=1}^{r}\begin{pmatrix}\boldsymbol{y}\\ 1\end{pmatrix}x_{ij}\phi(\boldsymbol{y}\,|\,\boldsymbol{x}_i, \boldsymbol{\delta}),$$
$$\psi_j^2 = \frac{\sum_{i=1}^{N}\sum_{m_1=1}^{r}\cdots\sum_{m_q=1}^{r}(x_{ij}-\boldsymbol{y}'\boldsymbol{\alpha}_j-\mu_j)^2\phi(\boldsymbol{y}\,|\,\boldsymbol{x}_i, \boldsymbol{\delta})}{\sum_{i=1}^{N}\sum_{m_1=1}^{r}\cdots\sum_{m_q=1}^{r}\phi(\boldsymbol{y}\,|\,\boldsymbol{x}_i, \boldsymbol{\delta})} \tag{12}$$
$$=\frac{1}{N}\sum_{i=1}^{N}\sum_{m_1=1}^{r}\cdots\sum_{m_q=1}^{r}(x_{ij}-\boldsymbol{y}'\boldsymbol{\alpha}_j-\mu_j)^2\phi(\boldsymbol{y}\,|\,\boldsymbol{x}_i, \boldsymbol{\delta}).$$

(2) binomial distribution with logistic function

Let x_{ij} be the value (one or zero) of a dichotomous manifest variable which is distributed with logistic function. Then, the probability function is described as:

$$P_j(x_{ij}\,|\,\boldsymbol{\theta}_i, \boldsymbol{\delta}_j) = \frac{\exp\{(-\boldsymbol{\theta}_i'\boldsymbol{\alpha}_j-\mu_j)(1-x_{ij})\}}{1+\exp(-\boldsymbol{\theta}_i'\boldsymbol{\alpha}_j-\mu_j)}, \quad (j=p_1+1, \cdots, p_1+p_2), \tag{13}$$

where $\boldsymbol{\alpha}_j$ is the vector of the parameters of so-called discrimination and corresponds to the factor loadings of the j-th variable, μ_j is an intercept which determines the probability of $x_{ij}=1$ when $\boldsymbol{\theta}_i=\mathbf{0}$, and $\boldsymbol{\delta}_j=(\boldsymbol{\alpha}_j', \mu_j)'$.

Using $1/(1+\exp(-1.7x)\cong\int_{-\infty}^{x}(1/\sqrt{2\pi})\exp(-x^2/2)dx$, the proportion of the variance of common factors is approximated by $\dfrac{\boldsymbol{\alpha}_j'\boldsymbol{\alpha}_j/1.7^2}{(\boldsymbol{\alpha}_j'\boldsymbol{\alpha}_j/1.7^2)+1}$.

From (13), $\partial \ln P_j(\cdot)/\partial\boldsymbol{\delta}_j$ in (5) is

$$\frac{\partial \ln P_j(\cdot)}{\partial(\boldsymbol{\alpha}_j', \mu_j)'} = \left\{x_{ij}-\frac{1}{1+\exp(-\boldsymbol{y}'\boldsymbol{\alpha}_j-\mu_j)}\right\}\begin{bmatrix}\boldsymbol{y}\\ 1\end{bmatrix}, \tag{14}$$

and $E(\cdot)$ in the right-hand side of (6) is

$$E\left(\frac{-\partial^2 \ln P_j(\cdot)}{\partial\boldsymbol{\delta}_j\partial\boldsymbol{\delta}_j'}\right) = \begin{bmatrix}\boldsymbol{y}\boldsymbol{y}' & \boldsymbol{y}\\ \boldsymbol{y}' & 1\end{bmatrix}\frac{\exp(-\boldsymbol{y}'\boldsymbol{\alpha}_j-\mu_j)}{\{1+\exp(-\boldsymbol{y}'\boldsymbol{\alpha}_j-\mu_j)\}^2}. \tag{15}$$

(3) Poisson distribution

Suppose that x_{ij} is the value of a Poisson distributed manifest variable, which takes values of nonnegative integers. This is the model of Ogasawara (1996b). Poisson distribution is a distribution for relatively low frequencies, which is used as a distribution of the number of errors for example in the field of psychological measurement. Thus, considering these applications, x_{ij} is assumed to be inversely related to $\boldsymbol{\theta}_i$. That is, $-\boldsymbol{\theta}_i$ is defined as inability which is positively correlated with x_{ij}. Since the expectation of a Poisson variable must be positive, the logarithm of the expectation is described as $(-\boldsymbol{\theta}_i)'\boldsymbol{\alpha}_j+\mu_j$. The probability function becomes

$$P_j(x_{ij}\,|\,\boldsymbol{\theta}_i, \boldsymbol{\delta}_j) = \exp\{x_{ij}(-\boldsymbol{\theta}_i'\boldsymbol{\alpha}_j+\mu_j)\}\times\exp\{-\exp(-\boldsymbol{\theta}_i'\boldsymbol{\alpha}_j+\mu_j)\}/x_{ij}!, \tag{16}$$
$$(j=p_1+p_2+1, \cdots, P),$$

where $\boldsymbol{\alpha}_j$ is a kind of factor loadings ; μ_j is the intercept which determines the value of the expectation of x_{ij} when $\boldsymbol{\theta}_i$ is zero, and $\boldsymbol{\delta}_j=(\boldsymbol{\alpha}_j', \mu_j)'$. From (16), $\partial \ln P_j(\cdot)/\partial\boldsymbol{\delta}_j$ in (5) is

$$\frac{\partial \ln P_j(\cdot)}{\partial(\boldsymbol{\alpha}_j', \mu_j)'}=\{x_{ij}-\exp(-\boldsymbol{y}'\boldsymbol{\alpha}_j+\mu_j)\}\begin{bmatrix}-\boldsymbol{y}\\ 1\end{bmatrix}, \tag{17}$$

and $E(\cdot)$ in the right-hand side of (6) is

$$E\left(\frac{-\partial^2 \ln P_j(\cdot)}{\partial\boldsymbol{\delta}_j\partial\boldsymbol{\delta}_j'}\right)=\begin{bmatrix}\boldsymbol{y}\boldsymbol{y}' & -\boldsymbol{y}\\ -\boldsymbol{y}' & 1\end{bmatrix}\exp(-\boldsymbol{y}'\boldsymbol{\alpha}_j+\mu_j). \tag{18}$$

Let $\lambda_{ij}=\exp(-\boldsymbol{\theta}_i'\boldsymbol{\alpha}_j+\mu_j)$ and let $E(X_{ij})$ and $Var(X_{ij})$ be the marginal expectation and variance, respectively, of the variable X_{ij} which takes the value x_{ij} in (17). Then

$$Var(X_{ij})=E(X_{ij})+Var(\lambda_{ij}) \tag{19}$$

(Meredith, 1971). When $\boldsymbol{\theta}_i$ are fixed, that is, $Var(\lambda_{ij})=0$, well known relationship $E(X_{ij})=Var(X_{ij})$ in Poisson distribution is obtained. Since the variation $(E(X_{ij}))$ can be interpreted as a specific one to the j-th variable, the communality is defined as $Var(\lambda_{ij})/(E(X_{ij})+Var(\lambda_{ij}))$. Incidentally, Meredith (1971) defined this as reliability. From $\lambda_{ij}\cong 1+(-\boldsymbol{\theta}_i'\boldsymbol{\alpha}_j+\mu_j)$, $Var(\lambda_{ij})\cong\boldsymbol{\alpha}_j'\boldsymbol{\alpha}_j$. Using this approximation, the proportion of the communality of the j-th variable is estimated by $\boldsymbol{\alpha}_j'\boldsymbol{\alpha}_j/(E(X_{ij})+\boldsymbol{\alpha}_j'\boldsymbol{\alpha}_j)$, where $E(X_{ij})$ can be replaced by the sample mean.

An anonymous referee pointed out that the communalities for various manifest variables can be defined as follows. Generally, the following formula (see e.g., Rao (1973), p. 97, (2b. 3.6)) holds

$$Var(X_{ij})=E_{\theta_i}(Var(X_{ij}\mid\boldsymbol{\theta}_i))+Var_{\theta_i}(E(X_{ij}\mid\boldsymbol{\theta}_i)), \tag{20}$$

where the subscript $\boldsymbol{\theta}_i$ indicates that the expectation and variance are taken with respect to it. The communality is defined as $Var_{\theta_i}(E(X_{ij}\mid\boldsymbol{\theta}_i))/Var(X_{ij})$. This is a general and comprehensive definition of the communality.

5. Numerical examples

Two data sets are used which have been constructed from actual data. Each data set consists of the scores of subtests of the adult ability, tests, Tests A/C ($N=$ 1,495) and Test B ($N=1{,}493$). Contents of the subtests are shown in Table 1. Among six subtests in each data set, two subtests are assigned to the manifest variables of normal, binomial or Poisson distribution : that is, the case of $p_1=p_2=p_3=2$ and $p=6$ in Section 4. In the normal model, the number-correct score in the subtest represents the value of the manifest variable. For logistic model, the number-correct score was divided into low or high categories by a threshold around sample mean, which constituted a dichotomous variable. The numbers of errors in subtests were assigned to the Poisson distributed variables.

Table 1
Contents of tests

Data set	Subtest	Task	Type of variable
Tests A/C $N=1{,}495$	A1	Letter search	Normal
	A2	Finding figures	Logistic
	A3	Symmetric figure I	Poisson
	A4	Rearranging figure	Normal
	C1	Digit-symbol	Logistic
	C2	Symmetric figure II	Poisson
Test B $N=1{,}493$	B1	Figure comparison	Normal
	B2	Figure matching	Logistic
	B3	Figure classification	Poisson
	B4	Part and whole	Normal
	B5	Construction of square	Logistic
	B6	Surface development	Poisson

Results of model fitting are shown in Tables 2 and 3. The number of common factors was set to one or two. In the case of two-factor model, an initial factor loadings were estimated by fixing the loading of the first subtest on the second factor at zero considering indeterminacy of rotation. These factor loadings were rotated to simple structure by the normalized varimax method. Integration of $\boldsymbol{\theta}_i$ was approximated by five points in each dimension of common factors which were assumed to have standardized normal distributions. The standard errors of rotated factor loadings in Tables 2 and 3 were obtained by the results for the ML method with restrictions (Ogasawara, 1996c).

First, the results of one-factor model are examined. Though the estimated values of factor loading can not directly be compared between different types of manifest variables, the stability of each loading may be assessed by comparing it to the value of corresponding standard error. All of the values of factor loadings are large enough when compared to their standard errors, which concludes that each subtest has common variation in spite of the difference in the types of measures. The estimated communality ranges from 7% to 33%.

Next, the results of two-factor model are examined. It is also difficult to compare directly the values of factor loadings between different types of measures. However, by comparing the factor loadings across common factors within each manifest variable, it is possible to assign each manifest variable to one of the common factors. By this method it is seen that in both of Tests A/C and Test B, the second factor consists of the Poisson variables and the first factor remaining variables. This may probably come from the fact that the variables of logistic function have been constructed from number-correct scores. For this reason the variables of logistic function may have formed the same group as the normal variables. The patterns of rotated factor loadings in Tables 2 and 3 seem to represent clear simple structures. The stability of the patterns is confirmed by comparing them with their estimated standard errors.

Comparing the results of the two-factor models with those of the one-factor

Table 2
Results of Tests A/C

Model Subtest (Type)	One-factor $-2\log L=30957.76$ I (SE)		μ_j (SE)	ψ_j (SE)	Communality (%)
A1 (Normal)	1.678 (.118)		16.292 (.109)	3.750 (.074)	16.7
A2 (Logistic)	1.051 (.098)		.277 (.066)	···	27.7
A3 (Poisson)	.669 (.020)		.648 (.025)	···	15.9
A4 (Normal)	1.873 (.112)		10.809 (.099)	3.290 (.065)	24.5
C1 (Logistic)	.911 (.089)		.012 (.063)	···	22.3
C2 (Poisson)	.822 (.026)		.006 (.035)	···	32.9
Model Subtest (type)	Two-factor (normalized varimax) $-2\log L=30570.83$ I (SE)	II (SE)	μ_j (SE)	ψ_j (SE)	Communality (%)
A1 (Normal)	2.464 (.135)	.335 (.102)	16.305 (.107)	3.271 (.085)	36.6
A2 (Logistic)	1.752 (.174)	.469 (.089)	.358 (.083)	···	53.2
A3 (Poisson)	.183 (.020)	.667 (.023)	.616 (.027)	···	16.9
A4 (Normal)	1.870 (.120)	.992 (.121)	10.828 (.099)	3.128 (.066)	31.4
C1 (Logistic)	1.558 (.155)	.307 (.076)	.019 (.076)	···	46.6
C2 (Poisson)	.162 (.024)	.855 (.030)	−.056 (.039)	···	35.5

Note: I: The first factor, II: The second factor. SE: Standard error.

Table 3
Results of Test B

Model Subtest (Type)	One-factor $-2\log L=33546.11$ I (SE)		μ_j (SE)	ψ_j (SE)	Communality (%)
B1 (Normal)	1.714 (.204)		35.557 (.182)	6.102 (.107)	7.3
B2 (Logistic)	.848 (.089)		−.141 (.067)	···	19.9
B3 (Poisson)	.590 (.023)		−.134 (.030)	···	25.3
B4 (Normal)	2.309 (.139)		15.461 (.125)	3.743 (.078)	27.6
B5 (Logistic)	.980 (.091)		−.137 (.067)	···	24.9
B6 (Poisson)	.752 (.017)		.977 (.026)	···	14.0
Model Subtest (type)	Two-factor (normalized varimax) $-2\log L=32566.83$ I (SE)	II (SE)	μ_j (SE)	ψ_j (SE)	Communality (%)
B1 (Normal)	3.743 (.172)	−.100 (.164)	35.582 (.164)	5.123 (.112)	34.8
B2 (Logistic)	2.229 (.211)	.241 (.088)	−.184 (.092)	···	63.5
B3 (Poisson)	.122 (.020)	.609 (.024)	−.157 (.032)	···	27.3
B4 (Normal)	3.226 (.114)	1.091 (.112)	15.488 (.114)	2.798 (.096)	59.7
B5 (Logistic)	1.779 (.148)	.511 (.087)	−.154 (.083)	···	54.3
B6 (Poisson)	.081 (.018)	.805 (.019)	.929 (.028)	···	15.8

Note: I: The first factor, II: The second factor, SE: Standard error.

models, the followings are observed. The values of μ_j are fairly stable. The values of ψ_j in the two-factor models are reduced from those in the one-factor model (slight reduction in Tests A/C). The values of communalities in the two-factor models have increased from the one-factor models, though the increases in Poisson variables are slight.

From the values of $-2\times\log$ likelihood in Tables 2 and 3, χ^2 values are obtained as 386.93 for Tests A/C and 979.28 for Test B with the degree of freedom, 5 for each data set considering the rotational freedom. These values are used as testing the

Table 4
Results of Simulation ($N=1{,}500$)

Subtest (Type)	True values I	II	μ_j	ψ_j	Communality (%)
S1 (Normal)	1	.3	0	1	52.2
S2 (Logistic)	1	.3	0	---	27.4
S3 (Poisson)	1	.3	−.5	---	51.7[a]
S4 (Normal)	.3	1	0	1	52.2
S5 (Logistic)	.3	1	0	---	27.4
S6 (Poisson)	.3	1	−.5	---	50.3[a]
Subtest (type)	Estimated values (normalized varimax) $-2\log L=22052.84$ I (SE)	II (SE)	μ_j (SE)	ψ_j (SE)	Communality (%)
S1 (Normal)	1.009 (.045)	.158 (.036)	−.006 (.037)	1.011 (.031)	50.5
S2 (Logistic)	.962 (.094)	.266 (.065)	−.074 (.062)	---	25.6
S3 (Poisson)	1.154 (.036)	.334 (.033)	−.579 (.056)	---	58.6
S4 (Normal)	.285 (.040)	1.109 (.043)	−.052 (.039)	.986 (.038)	57.4
S5 (Logistic)	.158 (.061)	.989 (.096)	−.018 (.063)	---	25.8
S6 (Poisson)	.306 (.034)	.999 (.036)	−.429 (.049)	---	50.3

Note: I : The first factor, II: The second factor, SE: Standard error.
a. The difference of the two population communalities comes from the difference of the estimated population means which have been obtained by sample means.

added goodness-of-fit by the two-factor model to the one-factor model. They are highly significant ($p<.001$), which favor the two-factor models over one-factor models in these data.

In the results of the numerical examples, the same type(s) of manifest variables represented each common factor. Since it was suspected that the difference of the types of manifest variables might have affected the estimated factor patterns, the following simulation with artificial data was carried out. The artificial data consist of the six manifest variables (S1-S6) of the same pattern of variable types as the above numerical examples.

The true values of the parameters are shown in Table 4, where each common factor consists of three manifest variables of distinct types of variables. The factor pattern represents a simple structure. Based on the population values, 1,500 random observations were generated with the assumption of the independent standardized normal distribution for each of the latent variables. In Table 4, the estimated values by the normalized varimax are provided. They are close to the true values except that the two estimated loadings (S1 for Factor II and S5 for Factor I) have negative biases. As a whole, these results suggest the appropriateness of our method.

6. Discussion

It is well known that the factor analysis model is a special case of covariance structure analysis (CSA) models. So, some remarks concerning CSA are given in this section. CSA is basically a method for continuous variables, though the ordered categorical variables are treated in the ways depending on models and methods. Most of the discrete variables are connected to underlying continuous measurement variables by assuming thresholds. That is, even in the case of discrete variables, a latent covariance matrix is assumed in the model. In this paper, the proposed model does not depend on the covariance matrix. Instead, the method of marginal maximum likelihood in IRT was applied to both of continuous and discrete variables. Using this method, different types of distributions are treated simultaneously.

In this paper the factor analysis model among CSA models was considered. Though the factor loadings in the normally distributed variables give 'covariance' structures, the parameters in the Poisson distributed variables cannot easily be described as part of covariances. Instead, they should be called the parameters in association structures, since they are the parameters in discrete variables.

References

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scores in two or more nominal categories. *Psychometrika,* **37**, 29-51.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: 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.

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

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.

Jöreskog, K.G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika,* **59**, 381-389.

Lawley, D.N., & Maxwell, A.E. (1971). *Factor analysis as a statistical method* (2nd ed.). London: Butterworths.

Lee, S.Y. & Poon, W.Y. (1986). Maximum likelihood estimation of polyserial correlations. *Psychometrika,* **51**, 113-121.

Lee, S.Y. & Poon, W.Y. (1987). Two-step estimation of multivariate polychoric correlations. *Communications in Statistics : Theory and Methods,* **16**, 307-320.

Lee, S.Y., Poon, W.Y., & Bentler, P.M. (1990a). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics and Probability Letters,* **9**, 91-97.

Lee, S.Y., Poon, W.Y., & Bentler, P.M. (1990b). 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.

Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

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.

Meredith, W. (1971). Poisson distribution of error in mental test theory. *British Journal of Mathematical and Statistical Psychology,* **24**, 49-82.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement,* **14**, 59-71.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement,* **16**, 159-176.

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

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics,* **22**, 43-65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika,* **49**, 115-132.

Muthén, B.O., & Sattora, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika,* **60**, 489-503.

Ogasawara, H. (1995). The gamma-gamma regression for the distribution model of event times. *Japanese Psychological Research,* **37**, 70-79.

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

Ogasawara, H. (1996b). Poisson factor analysis: A latent variable model for item response data which are represented by frequencies. *The 4th Item Response Seminar*, University of Tokyo, Tokyo, Japan. (in Japanese)

Ogasawara, H. (1996c). Standard errors for rotated factor loadings by normalized orthomax method. *The Japanese Journal of Behaviormetrics,* **23**, 122-129. (in Japanese)

Poon, W.Y., & Lee, S.Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika,* **52**, 409-430.

Rao, C.R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: Wiley.

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

Rubin, D.B., & Thayer, D.T. (1982). EM algorithms for ML factor analysis. *Psychometrika,* **47**, 69-76.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement,* No. 17.

(Received December, 1996, Revised October, 1997)