

## NEGATIVE BINOMIAL FACTOR ANALYSIS

Haruhiko Ogasawara\*

A latent variable model for observed variables representing frequencies is proposed. The data type for the model is a subjects by variables two-way frequency table. The model has two groups of latent variables. The first group of latent variables represents the characteristics of subjects and corresponds to common factors in factor analysis. On the other hand, each of latent variables in the second group is related to one of the manifest variables and corresponds to a specific factor in factor analysis. The manifest variables in the model, when given the values of common latent variables, follow the negative binomial distributions. The latent variables in the first and second groups are integrated out of the model. The parameters in the model are estimated by the marginal maximum likelihood method, using a kind of the EM algorithm. The communality, specificity, and reliability for an observed variable are defined.

If a group of subjects take several tests whose scores are represented by counts such as the numbers of errors, the data may be summarized by a subjects by tests frequency table. The typical models applied to such data are the log-linear model (see e.g., Bishop, Fienberg & Holland, 1975) and the log-bilinear model (see Goodman, 1985, 1986, 1991). When the number of subjects is moderate, the models may be easily fitted. However, if we have a large number of subjects, the number of parameters increases rapidly. Thus, in such situations we may not obtain a concise description of the data by fitting these models. Most of these models deal with the two modes, subjects and tests, symmetrically. However, if the relationships among tests are of primary interest, then the assessment of individual subjects is not the focus and the parameters which are concerned with subjects become nuisance parameters (see Clogg, 1986).

This situation corresponds to that of the ordinary factor analysis for continuous variables, where we have several difficulties by dealing with factor scores for subjects as unknown fixed parameters. In the random model, the factor scores are treated as random variables which do not appear in the likelihood to be maximized. For the frequency data mentioned earlier, the Poisson factor analysis model (Ogasawara, 1996a, 1998b) has been proposed, where the parameters corresponding to subjects in the log-bilinear model are taken as random variables. In this model, the latent variables are numerically integrated out of the model and it is possible to treat multiple latent variables with factor rotation if necessary.

The name of factor analysis in the Poisson model comes from the fact that the relationships among the manifest variables are explained by assuming the latent

---

*Key Words and Phrases:* negative binomial distribution, Poisson distribution, specific factor, communality, marginal maximum likelihood, frequency table

\* 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

variables (factors) which are common to the manifest variables. However, the Poisson factor analysis model is different from the ordinary factor analysis model in that the former does not assume specific factors. This point is also applied to the usual item response models (see e.g., Lord & Novick, 1968). That is, for ordinary item response models, given the values of latent variables (abilities), the probability of an event such as responding to an item correctly is the same for all subjects. However, it is quite natural to consider the variation of subjects which is specific to each manifest variable as assumed by the ordinary factor analysis model. The Poisson factor analysis resembles the principal component analysis without specific factors. The negative binomial factor analysis model proposed in this paper contains specific factors and is an extension of the Poisson factor analysis model. Tsutakawa (1988) proposed a model which is similar to our model. Though the specific variation corresponding to each manifest variable is considered in his model, the common variation comes from only one common latent variable following the negative gamma distribution. Our model will deal with several latent variables as is usual in ordinary exploratory factor analysis.

## Model

The usual factor analysis model for continuous variables is a kind of regression model in which the explanatory variables are the unobserved latent variables. The factor analysis model proposed in this paper is based on the negative binomial regression model (Lawless, 1987; see also Gardner, Mulvey & Shaw, 1995). Let  $X_{ij}$  be a variable representing a count (e.g., the number of errors) for the  $i$ -th ( $i=1, \dots, N$ ) individual and the  $j$ -th ( $j=1, \dots, p$ ) test. In the proposed model, the probability of  $X_{ij}=x_{ij}$  is described by the probability function for the Poisson distribution with the parameter  $\eta_{ij}\exp(\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \mu_j)$ :

$$P(X_{ij}=x_{ij} \mid \boldsymbol{\theta}_i, \eta_{ij}, \boldsymbol{\delta}_j, \mu_j) = (\eta_{ij} \exp(\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \mu_j))^{x_{ij}} \times \exp(-\eta_{ij} \exp(\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \mu_j)) / x_{ij}!, \quad (1)$$

$$(i=1, \dots, N; j=1, \dots, p),$$

where  $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{iq})'$  is the vector of  $q$  latent variables for the  $i$ -th subject. The latent variables will be called "common factors" considering their similarity to those in the ordinary factor analysis model for continuous variables. The vector  $\boldsymbol{\delta}_j = (\delta_{j1}, \dots, \delta_{jq})'$  contains the regression coefficients of the  $j$ -th manifest variable on the  $q$  latent common factors and correspond to factor loadings in the ordinary factor analysis model or discrimination parameters in the IRT model. Here, the regression coefficients will be called "factor loadings". The parameter  $\mu_j$  is an intercept parameter which determines the probability of  $X_{ij}=x_{ij}$  for a typical subject with  $\boldsymbol{\theta}_i = \mathbf{0}$ . The latent variable  $\eta_{ij}$  takes positive values and correspond to the score of the  $i$ -th subject for the "specific factor" for the  $j$ -th variable. Transforming the form of the Poisson parameter in (1),

$$\eta_{ij} \exp(\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \mu_j) = \exp(\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \ln \eta_{ij} + \mu_j), \tag{2}$$

$$(i=1, \dots, N; j=1, \dots, p),$$

is obtained. Hence,  $\ln \eta_{ij}$  corresponds to the term of the product of the factor loading of the  $j$ -th specific factor and its factor score for the  $i$ -th subject in the ordinary factor analysis model.

We can assume various distributions for  $\boldsymbol{\theta}_i$ . One of them is the  $q$ -dimensional standardized normal distribution. For  $\eta_{ij}$ , the gamma distribution with mean one and variance  $a_j$  is assumed, that is,

$$f(\eta_{ij} | a_j) = \frac{1}{\Gamma(1/a_j) a_j^{1/a_j}} \eta_{ij}^{\frac{1}{a_j}-1} \exp(-\eta_{ij}/a_j), \tag{3}$$

$$(i=1, \dots, N; j=1, \dots, p).$$

The reason for assuming the gamma distribution for  $\eta_{ij}$  is a computational one:  $\eta_{ij}$  can be integrated out. Though the term  $\ln \eta_{ij}$  takes the values in  $(-\infty, +\infty)$ , its distribution is not symmetric like the normal distribution for  $\boldsymbol{\theta}_i$ . In principle, it is possible to assume the log-normal distribution for  $\eta_{ij}$ . However, the integration of  $\eta_{ij}$  may have to be numeric/may not be possible symbolically. It is also possible to assume the exponential gamma distribution for each element of  $\boldsymbol{\theta}_i$ , which is the same distribution as  $\ln \eta_{ij}$ . However, such a model may present a problem in factor rotation.

Let  $h_{ij} = \exp(\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \mu_j)$ . Then, with the assumption of (3),  $\eta_{ij}$  is integrated out of (1) as

$$P(X_{ij} = x_{ij} | \boldsymbol{\theta}_i, \boldsymbol{\delta}_j, \mu_j, a_j) = \int_0^{+\infty} \frac{h_{ij}^{x_{ij}} \eta_{ij}^{x_{ij} + \frac{1}{a_j} - 1} \exp\left\{-\left(h_{ij} + \frac{1}{a_j}\right)\eta_{ij}\right\}}{\Gamma(1/a_j) a_j^{1/a_j} x_{ij}!} d\eta_{ij}$$

$$= \frac{h_{ij}^{x_{ij}} \Gamma\left(x_{ij} + \frac{1}{a_j}\right)}{\Gamma(1/a_j) \left(h_{ij} + \frac{1}{a_j}\right)^{x_{ij} + \frac{1}{a_j}} a_j^{1/a_j} x_{ij}!} \tag{4}$$

$$= \frac{h_{ij}^{x_{ij}} \prod_{k=0}^{x_{ij}-1} (1 + ka_j)}{(1 + h_{ij} a_j)^{x_{ij} + \frac{1}{a_j}} x_{ij}!},$$

$$(i=1, \dots, N; j=1, \dots, p),$$

where  $\prod_{k=0}^{-1} (\cdot) = 1$ . Equation (4) represents the probability function of the negative binomial distribution with shape parameter  $1/a_j$  and scale parameter  $a_j h_{ij}$ . The mean and variance of  $X_{ij}$  given  $\boldsymbol{\theta}_i$  with the probability function (4) are

$$E(X_{ij} | \boldsymbol{\theta}_i, \cdot) = h_{ij} \text{ and } Var(X_{ij} | \boldsymbol{\theta}_i, \cdot) = h_{ij} + a_j h_{ij}^2, \tag{5}$$

(see e.g., Lawless, 1987), where the center dot ‘ $\cdot$ ’ denotes the parameters concerned. The Poisson factor analysis corresponds to the case of  $a_j = 0$  and  $E(X_{ij} | \boldsymbol{\theta}_i, \cdot) = Var(X_{ij} | \boldsymbol{\theta}_i, \cdot)$ . The term  $a_j h_{ij}^2$  in (5) is the extra-Poisson or overdispersion variation (see e.g., Engel, 1984; Van Duijn, 1993) contributed by the “specific factor”

$\eta_{ij}$  when given  $\theta_i$ .

In the extended models of Rasch's multiplicative Poisson model (Rasch, 1960/1980) for subjects by tests frequency tables (e.g., Jansen & Van Duijn, 1992; Van Duijn, 1993; Ogasawara, 1996b), the Poisson parameter (see (2)) takes the form such as

$$a_i\beta_j = \exp(\ln a_i + \ln \beta_j) = \exp(a'_i + \beta'_j). \tag{6}$$

This shows that the models with (6) are equivalent to the log-linear model without interaction terms. The model of (1) with (4) is different from (6) in that the interaction terms in the log-linear model take the bilinear forms like the factor analysis model.

**Estimation**

We assume the local independence of  $X_{ij}$ , ( $j=1, \dots, p$ ) given  $\theta_i$  as is usual for the IRT models. With this assumption the first-stage marginal likelihood  $L^*$  with  $\eta_{ij}$  integrated out by (4) is formally described as

$$\begin{aligned} L^* &= \prod_{i=1}^N \prod_{j=1}^p P(X_{ij} = x_{ij} \mid \theta_i, \delta_j, \mu_j, a_j) \\ &= \prod_{i=1}^N L_i^*(\boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{a} \mid \theta_i, \mathbf{x}_i), \end{aligned} \tag{7}$$

where  $L_i^*(\cdot \mid \cdot)$  denotes the first-stage marginal likelihood of the parameters with respect to the  $i$ -th subject;  $\boldsymbol{\delta} = (\delta'_1, \dots, \delta'_p)'$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$ ,  $\boldsymbol{a} = (a_1, \dots, a_p)'$  and  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ . Since the likelihood  $L^*$  contains unknown latent variable vectors  $\theta_i$ , ( $i=1, \dots, N$ ) the second-stage marginal likelihood  $L^*$  is obtained by assuming prior distributions for  $\theta_i$ , ( $i=1, \dots, N$ ) in the following way:

$$L^* = \prod_{i=1}^N \int_{R(\theta_i)} L_i^*(\boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{a} \mid \theta_i, \mathbf{x}_i) t(\theta_i) d\theta_i, \tag{8}$$

where  $t(\theta_i)$  is a probability density function for  $\theta_i$  and is

$$t(\theta_i) = \frac{1}{(2\pi)^{q/2}} \exp(-\boldsymbol{\theta}'_i \boldsymbol{\theta}_i / 2), \quad (i=1, \dots, N) \tag{9}$$

in the case of the independent  $q$ -dimensional standardized normal distributions. In order to estimate the unknown parameters,  $L^*$  in (8) should be maximized with respect to the parameters. However, since the algebraic integration in (8) is difficult, we use a numerical approximation to the integration. For maximizing  $L^*$ , a kind of the EM algorithm (Bock & Aitkin, 1981) is employed.

The marginal likelihood  $L^*$  is approximated by  $L$  as follows:

$$\begin{aligned} L^* &\cong L = \prod_{i=1}^N \sum_{m_1=1}^r \dots \sum_{m_q=1}^r \left\{ \prod_{j=1}^p P(X_{ij} = x_{ij} \mid \mathbf{y}, \boldsymbol{\delta}_j, \mu_j, a_j) \right\} \times A(y_{m_1}) \times \dots \times A(y_{m_q}) \\ &= \prod_{i=1}^N g(\mathbf{x}_i \mid \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{a}), \end{aligned} \tag{10}$$

where  $\mathbf{y}=(y_{m_1}, \dots, y_{m_q})'$  is one of  $r^q$  lattice points in the distribution of  $\theta_i$  with the weight  $A(y_{m_1}) \times \dots \times A(y_{m_q})$  being proportional to  $l(\mathbf{y})$ . Let

$l = \ln L$  and

$$L_i = \prod_{j=1}^p P(X_{ij} = x_{ij} | \mathbf{y}, \boldsymbol{\delta}_j, \mu_j, a_j), \tag{11}$$

then from (4) the gradient vector for the parameters follows:

$$\begin{aligned} \frac{\partial l}{\partial(\boldsymbol{\delta}'_j, \mu_j)'} &= \sum_{i=1}^N \sum_{m_1=1}^r \dots \sum_{m_q=1}^r \frac{\partial \ln L_i}{\partial(\boldsymbol{\delta}'_j, \mu_j)'} \times \frac{L_i A(y_{m_1}) \times \dots \times A(y_{m_q})}{g(\mathbf{x}_i | \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a})} \\ &= \sum_{i=1}^N \sum_{m_1=1}^r \dots \sum_{m_q=1}^r \frac{\partial \ln L_i}{\partial(\boldsymbol{\delta}'_j, \mu_j)'} \phi(\mathbf{y} | \mathbf{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a}) \\ &= \sum_{i=1}^N \sum_{m_1=1}^r \dots \sum_{m_q=1}^r \left\{ x_{ij} - \left( x_{ij} + \frac{1}{a_j} \right) \frac{h_j a_j}{1 + h_j a_j} \right\} \phi(\mathbf{y} | \mathbf{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a}) (\mathbf{y}', 1)' \tag{12} \\ &= \sum_{i=1}^N \sum_{m_1=1}^r \dots \sum_{m_q=1}^r \frac{x_{ij} - h_j}{1 + h_j a_j} \phi(\mathbf{y} | \mathbf{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a}) (\mathbf{y}', 1)', \\ &(j=1, \dots, p), \end{aligned}$$

where  $h_j = \exp(\mathbf{y}' \boldsymbol{\delta}_j + \mu_j)$  and

$$\begin{aligned} \phi(\mathbf{y} | \mathbf{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a}) &= \frac{L_i A(y_{m_1}) \times \dots \times A(y_{m_q})}{g(\mathbf{x}_i | \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a})} \\ &(i=1, \dots, N) \end{aligned} \tag{13}$$

is the posterior probability of the lattice point  $\mathbf{y}$  given the observation  $\mathbf{x}_i$  and the values of the parameters  $\boldsymbol{\delta}, \boldsymbol{\mu}$  and  $\mathbf{a}$ . Similarly,

$$\begin{aligned} \frac{\partial l}{\partial a_j} &= \sum_{i=1}^N \sum_{m_1=1}^r \dots \sum_{m_q=1}^r \left\{ \sum_{k=0}^{x_{ij}-1} \frac{k}{1 + ka_j} + \frac{1}{a_j^2} \ln(1 + h_j a_j) - \left( x_{ij} + \frac{1}{a_j} \right) \frac{h_j}{1 + h_j a_j} \right\} \\ &\times \phi(\mathbf{y} | \mathbf{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a}), \quad (j=1, \dots, p) \end{aligned} \tag{14}$$

is obtained.

Equations (13) and (14) correspond to the gradient vector in the negative binomial regression model (Lawless, 1987, (2.3), (2.4)) with the additional weights  $\phi(\cdot | \cdot)$  for each observation. Actually, the weight  $\phi(\cdot | \cdot)$  includes unknown parameters and is not a given one. However, substituting initial appropriate values for  $\boldsymbol{\delta}, \boldsymbol{\mu}$  and  $\mathbf{a}$  and treating  $\phi(\cdot | \cdot)$  as if they were fixed known weights, the Hessian matrix of  $-l$  including expected values is obtained as

$$\begin{aligned}
 E\left(-\frac{\partial^2 l}{\partial(\boldsymbol{\delta}'_j, \boldsymbol{\mu}'_j) \partial(\boldsymbol{\delta}'_j, \boldsymbol{\mu}'_j)}\right) &= \sum_{i=1}^N \sum_{m_1=1}^r \cdots \sum_{m_q=1}^r \frac{h_j}{1+h_j a_j} \phi(\mathbf{y}|\mathbf{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a}) \\
 &\quad \times \begin{bmatrix} \mathbf{y}\mathbf{y}' & \mathbf{y} \\ \mathbf{y}' & 1 \end{bmatrix}, \\
 E\left(-\frac{\partial^2 l}{\partial(\boldsymbol{\delta}'_j, \boldsymbol{\mu}'_j) \partial a_j}\right) &= \mathbf{0}, \\
 -\frac{\partial^2 l}{\partial a_j^2} &= \sum_{i=1}^N \sum_{m_1=1}^r \cdots \sum_{m_q=1}^r \left\{ \sum_{k=0}^{x_{ij}-1} \frac{k^2}{(1+ka_j)^2} + \frac{2}{a_j^3} \ln(1+h_j a_j) - \frac{2h_j}{a_j^2(1+h_j a_j)} \right. \\
 &\quad \left. - \left(x_{ij} + \frac{1}{a_j}\right) \frac{h_j^2}{(1+h_j a_j)^2} \right\} \times \phi(\mathbf{y}|\mathbf{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a}), \\
 &\quad (j=1, \dots, p).
 \end{aligned} \tag{15}$$

Note that the second partial derivatives with respect to the parameters for the  $j$ -th variable and those for the  $j^*$ -th ( $j \neq j^*$ ) variable are zero when assuming  $\phi(\cdot | \cdot)$  as fixed values. The iteration proceeds separately for each manifest variable as:

$$\begin{pmatrix} \boldsymbol{\delta}_j \\ \boldsymbol{\mu}_j \\ \alpha_{j/(i-1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\delta}_j \\ \boldsymbol{\mu}_j \\ \alpha_{j/(i)} \end{pmatrix} + \left[ E\left(\frac{-\partial^2 l}{\partial(\boldsymbol{\delta}'_j, \boldsymbol{\mu}'_j, \alpha'_j) \partial(\boldsymbol{\delta}'_j, \boldsymbol{\mu}'_j, \alpha'_j)}\right) \right]^{-1} \left[ \frac{\partial l}{\partial(\boldsymbol{\delta}'_j, \boldsymbol{\mu}'_j, \alpha'_j)} \right]_{(i)}, \tag{16}$$

where the subscript  $(i)$  indicates the value in the  $i$ -th iteration and  $E(\cdot)$  denotes (15) as a whole. The iteration in (16) is performed only once when  $\phi(\cdot | \cdot)$  is given (one-step scoring method; Sammel et al., 1997). With the revised values of the parameters,  $\phi(\cdot | \cdot)$  is re-estimated. The next iteration in (16) proceeds with the new  $\phi(\cdot | \cdot)$ . The cycle continues until convergence is obtained. The vector of the converged values of the parameters  $\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\mu}}$  and  $\hat{\mathbf{a}}$  is the marginal maximum likelihood estimator of the parameter vector.

For evaluating the variation of the estimates when  $\boldsymbol{\theta}_i, (i=1, \dots, N)$  are given, the information matrix given by (16) with the parameters replaced by their estimates become the estimated asymptotic variance-covariance matrix of the parameter estimates. However, in our case (15) is not the exact Hessian or Fisher information matrix, since  $\phi(\cdot | \cdot)$  is regarded as if it is a fixed term in (15). The estimated matrix tends to be larger than the exact one. For the exact second partial derivatives of  $-l$ , the derivatives of  $\phi(\cdot | \cdot)$  should be evaluated or the following approximation should be used (Louis, 1982; Mislevy & Sheehan, 1989),

$$\begin{aligned}
 I &= \sum_{i=1}^N \left\{ \sum_{m_1=1}^r \cdots \sum_{m_q=1}^r \frac{\partial \ln L_i}{\partial(\boldsymbol{\delta}'_j, \boldsymbol{\mu}'_j, \mathbf{a}'_j)} \phi(\mathbf{y}|\mathbf{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a}) \right\} \\
 &\quad \times \left\{ \sum_{m_1=1}^r \cdots \sum_{m_q=1}^r \frac{\partial \ln L_i}{\partial(\boldsymbol{\delta}'_j, \boldsymbol{\mu}'_j, \mathbf{a}'_j)} \phi(\mathbf{y}|\mathbf{x}_i, \boldsymbol{\delta}, \boldsymbol{\mu}, \mathbf{a}) \right\}.
 \end{aligned} \tag{17}$$

We use (17) for the approximation of the exact information matrix.

When the number of common factors is more than one and all elements of  $\boldsymbol{\delta}_j, (j=1, \dots, p)$  are free parameters, we have rotational indeterminacy as is the case for ordinary factor analysis model. To remove the indeterminacy some of the elements of  $\boldsymbol{\delta}_j, (j=1, \dots, p)$  can be set to fixed values and the confirmatory factor

analysis model is obtained. In the case of exploratory factor analysis, appropriate  $(q^2 - q)/2$  elements of the loading matrix can be set to zero and the remaining free parameters estimated. The unrotated loadings may be rotated as  $\delta_j T$  where  $T(q \times q)$  is a transformation matrix for loadings. Though (17) can be used for the confirmatory factor analysis model, it cannot be used for the exploratory factor analysis model because in the latter case the estimated loadings are the (marginal) maximum likelihood estimates with restrictions. For such cases the augmented information matrix (see Silvey, 1975) should be substituted for (17). For orthogonal rotation, the restrictions for rotated loadings are as follows (Archer & Jennrich, 1973):

$$r_{ij} = \left( \frac{\partial c(\Delta)}{\partial \Delta'} \Delta - \Delta' \frac{\partial c(\Delta)}{\partial \Delta} \right)_{ij} = 0, \quad (q \geq i > j \geq 1), \tag{18}$$

where  $(\cdot)_{ij}$  denotes the  $(i, j)$ th element of the parenthesized matrix and  $c(\Delta)$  is the function to be optimized by the  $p \times q$  rotated loading matrix  $\Delta$ . For instance, in the case of the normalized varimax rotation,

$$c(\Delta) = \frac{1}{4} \sum_{j=1}^q \left\{ \sum_{i=1}^p \frac{\delta_{ij}^4}{\left( \sum_{m=1}^q \delta_{im}^2 \right)^2} - \frac{1}{p} \left( \sum_{i=1}^p \frac{\delta_{ij}^2}{\sum_{m=1}^q \delta_{im}^2} \right)^2 \right\}. \tag{19}$$

Equation (18) can be expressed as a vector  $\mathbf{r} = (r_{21}, r_{31}, r_{32}, \dots, r_{q,(q-1)})' = \mathbf{0}$ . The asymptotic variance-covariance matrix for the estimators with the restriction is obtained as the submatrix  $I^*$  in the inverse of the augmented information matrix  $I_A$ :

$$I_A^{-1} = \begin{bmatrix} I & \partial \mathbf{r}' / \partial (\boldsymbol{\delta}', \boldsymbol{\mu}', \boldsymbol{\alpha}')' \\ \partial \mathbf{r} / \partial (\boldsymbol{\delta}', \boldsymbol{\mu}', \boldsymbol{\alpha}') & O \end{bmatrix}^{-1} = \begin{bmatrix} I^* & \# \\ \# & \# \end{bmatrix}, \tag{20}$$

where  $\partial \mathbf{r} / \partial (\boldsymbol{\mu}', \boldsymbol{\alpha}') = O$  and  $\#$ 's denote submatrices which are not used here.

**Reliability, Community and Specificity**

Equation (5) shows that the extra-Poisson variation, which comes from a specific factor when  $\boldsymbol{\theta}_i$  is given, is  $a_j h_{ij}^2$ . The marginal variation becomes larger than that since  $h_{ij}$  includes the random vector  $\boldsymbol{\theta}_i$ . Meredith (1971) showed that the marginal variance of the Poisson distributed variable  $X_{ij}$ ,  $Var(V_{ij})$ , is described as the sum of the variance of the Poisson parameter  $h_{ij}$  in its prior distribution,  $Var(h_{ij}|\cdot)$ , and the marginal expectation of  $X_{ij}$ ,  $E(X_{ij})$ :

$$Var(X_{ij}) = Var(h_{ij}|\cdot) + E(X_{ij}). \tag{21}$$

Further, he defined the reliability of the Poisson distributed variable  $X_{ij}$  with a prior distribution as

$$\rho(X_{ij}) = 1 - \frac{E(X_{ij})}{Var(X_{ij})} \tag{22}$$

based on the result of classical test theory for speeded tests. Reliability is generally defined in psychological measurement as the ratio of the variation without measurement error to the total variation in a variable. Communality in factor analysis is the variation of common factors or its proportion to the total variance in a manifest variable. In the Poisson factor analysis, where there are no specific factors, the communality becomes equivalent to the reliability. However, when specific factors exist, the reliability (22) becomes greater than the communality.

In this section, the reliability, communality, and specificity (the proportion of the variation of a specific factor to the total variance of an observed variable; see Harman, 1976, p. 19) are defined and discussed for the negative binomial factor analysis model. First, we derive the marginal mean and variance of  $X_{ij}$ . Let  $h_{ij}^*$  be  $\eta_{ij}\exp(\lambda_{ij})$ , where  $\lambda_{ij} = \boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \mu_j$ . (Note that  $h_{ij}^* = \eta_{ij} h_{ij}$ .) Then,

$$\begin{aligned}
 E(X_{ij}) &= \sum_{x_{ij}=0}^{\infty} \int_{R(\eta_{ij})} \int_{R(\boldsymbol{\theta}_i)} x_{ij} h_{ij}^{*x_{ij}} \frac{\exp(-h_{ij}^*)}{x_{ij}!} t(\boldsymbol{\theta}_i) f(\eta_{ij} | \cdot) d\boldsymbol{\theta}_i d\eta_{ij} \\
 &= \int_{R(\eta_{ij})} \int_{R(\boldsymbol{\theta}_i)} h_{ij}^* \sum_{(x_{ij}-1)=0}^{\infty} h_{ij}^{*(x_{ij}-1)} \frac{\exp(-h_{ij}^*)}{(x_{ij}-1)!} t(\boldsymbol{\theta}_i) f(\eta_{ij} | \cdot) d\boldsymbol{\theta}_i d\eta_{ij} \\
 &= \int_{R(\eta_{ij})} \int_{R(\boldsymbol{\theta}_i)} \eta_{ij} \exp(\lambda_{ij}) t(\boldsymbol{\theta}_i) f(\eta_{ij} | \cdot) d\boldsymbol{\theta}_i d\eta_{ij} \\
 &= \int_{R(\boldsymbol{\theta}_i)} \exp(\lambda_{ij}) t(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i.
 \end{aligned} \tag{23}$$

Similarly,

$$\begin{aligned}
 E(X_{ij}(X_{ij}-1)) &= \sum_{x_{ij}=0}^{\infty} \int_{R(\eta_{ij})} \int_{R(\boldsymbol{\theta}_i)} h_{ij}^{*2} t(\boldsymbol{\theta}_i) f(\eta_{ij} | \cdot) d\boldsymbol{\theta}_i d\eta_{ij} \\
 &= \int_{R(\boldsymbol{\theta}_i)} \exp(2\lambda_{ij}) t(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i \times \int_{R(\eta_{ij})} \frac{\eta_{ij}^{\frac{1}{a_j}+2-1} \exp\left(-\frac{\eta_{ij}}{a_j}\right)}{\Gamma(1/a_j) a_j^{1/a_j}} d\eta_{ij} \\
 &= (a_j + 1) \int_{R(\boldsymbol{\theta}_i)} \exp(2\lambda_{ij}) t(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i
 \end{aligned} \tag{24}$$

is obtained. When  $t(\boldsymbol{\theta}_i) = \frac{1}{(2\pi)^{q/2}} \exp(-\boldsymbol{\theta}_i' \boldsymbol{\theta}_i / 2)$ , (23) and (24) become

$$\begin{aligned}
 E(X_{ij}) &= \int_{R(\boldsymbol{\theta}_i)} \frac{1}{(2\pi)^{q/2}} \exp\left(\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + \mu_j - \frac{\boldsymbol{\theta}_i' \boldsymbol{\theta}_i}{2}\right) d\boldsymbol{\theta}_i \\
 &= \exp\left(\mu_j + \frac{\boldsymbol{\delta}_j' \boldsymbol{\delta}_j}{2}\right),
 \end{aligned} \tag{25}$$

$$\begin{aligned}
 E(X_{ij}(X_{ij}-1)) &= (a_j + 1) \int_{R(\boldsymbol{\theta}_i)} \frac{1}{(2\pi)^{q/2}} \exp\left(2\boldsymbol{\theta}_i' \boldsymbol{\delta}_j + 2\mu_j - \frac{\boldsymbol{\theta}_i' \boldsymbol{\theta}_i}{2}\right) d\boldsymbol{\theta}_i \\
 &= (a_j + 1) \exp(2(\mu_j + \boldsymbol{\delta}_j' \boldsymbol{\delta}_j)), \\
 &(i=1, \dots, N; j=1, \dots, p).
 \end{aligned} \tag{26}$$

Hence,

$$\begin{aligned} \text{Var}(X_{ij}) &= (a_j + 1) \exp(2(\mu_j + \delta'_j \delta_j)) + \exp\left(\mu_j + \frac{\delta'_j \delta_j}{2}\right) \\ &\quad - \exp(2\mu_j + \delta'_j \delta_j), \end{aligned} \tag{27}$$

( $i=1, \dots, N$ ;  $j=1, \dots, p$ ),

follows. Using the relationships

$$\text{Var}(X_{ij}) = \text{Var}_{(\theta_i, \eta_{ij})}(E(X_{ij} | \theta_i, \eta_{ij})) + E_{(\theta_i, \eta_{ij})}(\text{Var}(X_{ij} | \theta_i, \eta_{ij})) \tag{28}$$

(see e.g., Rao, 1973, p. 97, (2b.3.6); Ogasawara, 1998a), the reliability is

$$\begin{aligned} \frac{\text{Var}_{(\theta_i, \eta_{ij})}(E(X_{ij} | \theta_i, \eta_{ij}))}{\text{Var}(X_{ij})} &= 1 - \frac{E_{(\theta_i, \eta_{ij})}(\text{Var}(X_{ij} | \theta_i, \eta_{ij}))}{\text{Var}(X_{ij})} \\ &= 1 - \frac{E(X_{ij})}{\text{Var}(X_{ij})} \\ &= \frac{(a_j + 1)\exp(2(\mu_j + \delta'_j \delta_j)) - \exp(2\mu_j + \delta'_j \delta_j)}{(a_j + 1)\exp(2(\mu_j + \delta'_j \delta_j)) + \exp\left(\mu_j + \frac{\delta'_j \delta_j}{2}\right) - \exp(2\mu_j + \delta'_j \delta_j)}. \end{aligned} \tag{29}$$

On the other hand, the communality is defined as

$$\begin{aligned} \frac{\text{Var}_{\theta_i}\{E_{\eta_{ij}}(E(X_{ij} | \theta_i, \eta_{ij}))\}}{\text{Var}(X_{ij})} &= \frac{\text{Var}_{\theta_i}\{E_{\eta_{ij}}(\eta_{ij} \exp(\lambda_{ij}))\}}{\text{Var}(X_{ij})} \\ &= \frac{\text{Var}_{\theta_i}(\exp(\lambda_{ij}))}{\text{Var}(X_{ij})}, \end{aligned} \tag{30}$$

where using (23) to (26) the numerator in the right-hand side of (30) is

$$\begin{aligned} \text{Var}_{\theta_i}(\exp(\lambda_{ij})) &= \int_{K(\theta_i)} \exp(2\lambda_{ij}) t(\theta_i) d\theta_i - \left[ \int_{K(\theta_i)} \exp(\lambda_{ij}) t(\theta_i) d\theta_i \right]^2 \\ &= \exp(2(\mu_j + \delta'_j \delta_j)) - \exp(2\mu_j + \delta'_j \delta_j). \end{aligned} \tag{31}$$

Thus, the communality defined by (30) is

$$\frac{\exp(2(\mu_j + \delta'_j \delta_j)) - \exp(2\mu_j + \delta'_j \delta_j)}{(a_j + 1)\exp(2(\mu_j + \delta'_j \delta_j)) + \exp\left(\mu_j + \frac{\delta'_j \delta_j}{2}\right) - \exp(2\mu_j + \delta'_j \delta_j)} \tag{32}$$

Further, the specificity is defined as

$$\begin{aligned} \frac{\text{Var}_{\eta_{ij}}\{E_{\theta_i}(E(X_{ij} | \theta_i, \eta_{ij}))\}}{\text{Var}(X_{ij})} &= \frac{\text{Var}_{\eta_{ij}}\{E_{\theta_i}(\eta_{ij} \exp(\lambda_{ij}))\}}{\text{Var}(X_{ij})} \\ &= \frac{\text{Var}_{\eta_{ij}}\left\{\eta_{ij} \exp\left(\mu_j + \frac{\delta'_j \delta_j}{2}\right)\right\}}{\text{Var}(X_{ij})} \\ &= \frac{a_j \exp(2\mu_j + \delta'_j \delta_j)}{(a_j + 1)\exp(2(\mu_j + \delta'_j \delta_j)) + \exp\left(\mu_j + \frac{\delta'_j \delta_j}{2}\right) - \exp(2\mu_j + \delta'_j \delta_j)}. \end{aligned} \tag{33}$$

From the above definitions of communality and specificity, we still have remaining variation in the reliability. The proportion of the remaining variation is from (29), (32) and (33)

$$\frac{a_j\{\exp(2(\mu_j + \delta'_j\delta_j)) - \exp(2\mu_j + \delta'_j\delta_j)\}}{(a_j + 1)\exp(2(\mu_j + \delta'_j\delta_j)) + \exp\left(\mu_j + \frac{\delta'_j\delta_j}{2}\right) - \exp(2\mu_j + \delta'_j\delta_j)} \geq 0. \tag{34}$$

The value of (34) is zero only when  $a_j=0$  or  $\delta'_j\delta_j=0$ .

**Numerical Examples**

Numerical examples are based on four sets of artificial data (Data A, B, C and D). They have been generated such that they follow the distributions similar to actual distributions of the numbers of errors in psychological tests (see e.g., Ogasawara, 1996b). In each data set,  $N=1,000$  and  $p=6$ . Table 1 shows the elementary statistics for the data with their population values. The population values of the parameters are shown in Tables 2, 3, 4 and 5. The four sets of data were constructed by generating 1,000 random observations for each data set which follow the distributions determined by the parameters. Data A and B are based on

Table 1  
Population and Observed Moments

Variable Data	Means		Variances		Observed Product-Moment Correlation Coefficients						
	Pop.	Ob.	Pop.	Ob.							
A	1	1.08	1.17	1.84	2.03	1.000					
	2	1.79	1.76	3.84	3.56	.115	1.000	N = 1,000			
	3	1.08	1.05	1.98	2.06	.110	.103	1.000			
	4	1.79	1.76	4.21	4.00	.169	.110	.131	1.000		
	5	1.08	1.07	2.11	2.00	.141	.066	.038	.130	1.000	
	6	1.79	1.82	4.59	4.98	.125	.098	.027	.102	.076	1.000
B	1	1.20	1.23	2.64	2.89	1.000					
	2	1.97	2.05	5.90	6.09	.211	1.000	N = 1,000			
	3	1.20	1.20	2.85	2.68	.183	.267	1.000			
	4	1.97	2.10	6.45	7.92	.200	.274	.223	1.000		
	5	1.20	1.20	3.05	3.07	.330	.286	.205	.284	1.000	
	6	1.97	2.08	7.01	7.54	.232	.254	.100	.282	.181	1.000
C	1	1.16	1.27	2.32	2.60	1.000					
	2	1.91	1.92	5.07	4.53	.190	1.000	N = 1,000			
	3	1.16	1.19	2.50	2.44	.180	.179	1.000			
	4	1.91	1.97	5.56	5.85	.201	.198	.168	1.000		
	5	1.16	1.20	2.68	2.41	.129	.141	.160	.207	1.000	
	6	1.91	1.97	6.04	6.48	.174	.133	.101	.229	.272	1.000
D	1	1.28	1.30	3.44	3.90	1.000					
	2	2.12	2.04	7.98	7.41	.317	1.000	N = 1,000			
	3	1.28	1.21	3.71	2.83	.316	.291	1.000			
	4	2.12	2.25	8.72	10.02	.074	.118	.081	1.000		
	5	1.28	1.31	3.98	2.93	.126	.097	.037	.245	1.000	
	6	2.12	2.13	9.46	9.47	.090	.123	.047	.319	.207	1.000

Note : Pop. = Population values, Ob. = Observed values.

one-factor models: Data A is generated from the population with relatively low correlations ( $\delta=(.4,.4,.4,.4,.4,.4)$ ) and Data B from the population with relatively high correlations ( $\delta=(.6,.6,.6,.6,.6,.6)$ ). Data C and D are based on two-factor models: the population loading matrix for Data C is

$$A' = \begin{pmatrix} .5 & .5 & .2 & .2 & .2 \\ .2 & .2 & .2 & .5 & .5 \end{pmatrix},$$

which shows a weak simple structure and that for Data D is

$$A' = \begin{pmatrix} .7 & .7 & .7 & .1 & .1 & .1 \\ .1 & .1 & .1 & .7 & .7 & .7 \end{pmatrix},$$

which represents a clear simple structure. These two population loading matrices obviously optimize the raw/normalized varimax criteria. In addition,  $\mu=(.0,.5,.0,.5,.0,.5)$  and  $\alpha=(.4,.4,.5,.5,.6,.6)$  were assumed for each population. In the sample product-moment correlation matrices in Table 1, the largest value is about .3, which may seem low. However, in data representing the number of errors (incorrect responses), it is not unusual to have large frequencies of  $X_{ij}=0$ , giving low product-moment correlations. The simulated data are comparable with those observed in such typical situations.

Tables 2 and 3 show the parameter estimates and their standard errors for the one-factor models. The tables contain the population and sample values of the communality (Com.), specificity (Spe.), residual in reliability (Com. $\times$ Spe.) and reliability (Rel.; defined by the last term of (29), not 1-observed ( $E(\cdot)/Var(\cdot)$ )). In

Table 2  
Results for Data A

Variable No.	$\delta_j$ Pop.	PS Est. (SE)	NB Est. (SE)	$\mu_j$ Pop.	PS Est. (SE)	NB Est. (SE)
1	.4	.471 (.031)	.515 (.068)	.0	.048 (.033)	.025 (.048)
2	.4	.409 (.025)	.344 (.059)	.5	.485 (.025)	.508 (.037)
3	.4	.356 (.027)	.337 (.065)	.0	-.011 (.030)	-.006 (.046)
4	.4	.555 (.026)	.474 (.067)	.5	.418 (.030)	.458 (.042)
5	.4	.395 (.031)	.345 (.071)	.0	-.012 (.030)	.005 (.045)
6	.4	.638 (.024)	.300 (.059)	.5	.400 (.033)	.552 (.040)

  

Variable No.	$\alpha_j$ Pop.	NB Est. (SE)	Com.		Spe.		Com. $\times$ Spe.		Rel.	
			Pop.	Est.	Pop.	Est.	Pop.	Est.	Pop.	Est.
1	.4	.265 (.081)	.11	.20	.26	.18	.04	.05	.41	.43
2	.4	.415 (.061)	.14	.11	.33	.36	.06	.05	.54	.51
3	.5	.612 (.094)	.10	.07	.30	.35	.05	.04	.45	.46
4	.5	.428 (.075)	.13	.19	.38	.32	.07	.08	.58	.58
5	.6	.675 (.100)	.10	.07	.33	.37	.06	.05	.49	.49
6	.6	.762 (.076)	.12	.06	.42	.52	.07	.05	.61	.63

Note: Pop.=Population values, Est.=Estimates, SE=Standard errors, PS=Poisson model, NB=Negative binomial model, Com.=Communalities, Spe.=Specificities, Rel.=Reliabilities.

Table 3  
Results for Data B

Variable No.	$\delta_j$ Pop.	PS Est. (SE)	NB Est. (SE)	$\mu_j$ Pop.	PS Est. (SE)	NB Est. (SE)
1	.6	.626 (.026)	.562 (.050)	.0	.098 (.033)	.060 (.047)
2	.6	.706 (.021)	.615 (.046)	.5	.567 (.029)	.540 (.042)
3	.6	.608 (.029)	.551 (.056)	.0	.075 (.032)	.038 (.047)
4	.6	.805 (.019)	.660 (.044)	.5	.539 (.031)	.537 (.045)
5	.6	.705 (.027)	.653 (.054)	.0	.032 (.034)	-.023 (.052)
6	.6	.757 (.019)	.556 (.047)	.5	.558 (.029)	.591 (.043)

  

Variable No.	$\alpha_j$ Pop.	NB Est. (SE)	Com.		Spe.		Com. × Spe.		Rel.	
			Pop.	Est.	Pop.	Est.	Pop.	Est.	Pop.	Est.
1	.4	.422 (.073)	.24	.21	.22	.24	.09	.09	.55	.54
2	.4	.406 (.061)	.29	.30	.26	.26	.12	.12	.67	.69
3	.5	.539 (.082)	.22	.19	.25	.28	.11	.10	.58	.57
4	.5	.432 (.062)	.26	.32	.30	.26	.13	.14	.69	.72
5	.6	.492 (.090)	.20	.25	.28	.23	.12	.12	.61	.61
6	.6	.629 (.074)	.24	.21	.33	.37	.14	.13	.72	.72

Note: Pop.=Population values, Est.=Estimates, SE=Standard errors, PS=Poisson model, NB=Negative binomial model, Com.=Communalities, Spe.=Specificities, Rel.=Reliabilities.

the tables the results for the Poisson factor analysis model ( $\alpha_j=0$ ) are also included. For the numerical integration for  $\theta_i$ , five points with their weights in the standardized normal distribution were used. The estimates are similar to their population values. The estimated values of the above indices (Com. etc.) are also similar to their population values. Overall, the results of the estimates for the Poisson factor analysis model are not very different from those for the negative binomial factor analysis. However, there are large differences in the standard errors, which has been previously noted for the results of the Poisson and negative binomial regressions (Lawless, 1987, p. 221). From Tables 2 and 3 we see that it is not appropriate to use the standard errors of the estimates in the Poisson factor analysis model when there is considerable variation for a specific factor.

Tables 4 and 5 show the results for two-factor models. For the integration of  $\theta_i$ ,  $5^2$  lattice points in the bivariate standardized normal distribution with corresponding weights were used. The initial loading matrix has been estimated with  $\delta_{12}=(\delta'_i)_{12}=0$ . After convergence, the estimated loadings were rotated by the normalized varimax method. Though some of the parameter estimates are off their population values, overall the estimates recover the population values. The standard errors for the Poisson factor analysis model are much smaller than those for the negative binomial factor analysis model. In Data A-D, the residual reliabilities are less than 20% and smaller than the communalities or specificities.

Table 4  
Results for Data C

Variable No.	$\delta_j$		PS		NB	
	Pop.		Est. (SE)	Est. (SE)	Est. (SE)	Est. (SE)
1	.5	.2	.461 (.029)	.270 (.036)	.411 (.092)	.255 (.104)
2	.5	.2	.633 (.027)	.074 (.024)	.495 (.091)	.167 (.060)
3	.5	.2	.570 (.032)	.130 (.026)	.437 (.096)	.203 (.089)
4	.2	.5	.574 (.029)	.420 (.034)	.367 (.099)	.370 (.103)
5	.2	.5	.336 (.038)	.480 (.034)	.298 (.115)	.484 (.123)
6	.2	.5	.089 (.026)	.880 (.028)	.199 (.052)	.612 (.123)

  

Variable No.	$\mu_j$ Pop.	PS	NB	$\alpha_j$ Pop.	NB
		Est. (SE)	Est. (SE)		Est. (SE)
1	.0	.114 (.032)	.118 (.046)	.4	.462 (.085)
2	.5	.473 (.032)	.519 (.049)	.4	.342 (.089)
3	.0	.023 (.036)	.552 (.050)	.5	.537 (.099)
4	.5	.455 (.034)	.537 (.043)	.5	.528 (.068)
5	.0	.037 (.035)	.249 (.054)	.6	.400 (.101)
6	.5	.318 (.044)	.469 (.075)	.6	.418 (.152)

  

Variable No.	Com.		Spe.		Com. × Spe.		Rel.	
	Pop.	Est.	Pop.	Est.	Pop.	Est.	Pop.	Est.
1	.19	.16	.23	.28	.08	.08	.50	.52
2	.24	.24	.29	.27	.10	.08	.62	.60
3	.18	.15	.27	.30	.09	.08	.54	.53
4	.22	.21	.33	.35	.11	.11	.66	.66
5	.17	.22	.30	.23	.10	.09	.57	.53
6	.20	.31	.36	.25	.12	.13	.68	.69

Note: Pop.=Population values, Est.=Estimates, SE=Standard errors, PS=Poisson model, NB=Negative binomial model, Com.=Communalities, Spe.=Specificities, Rel.=Reliabilities.

**Discussion**

The existence of specific factors is often assumed in factor analysis model for continuous variables. However, in the usual exploratory factor analysis model, a specific factor cannot be separated from a measurement error. The two components are usually combined together as a unique factor in factor analysis model. To estimate the specific-factor loadings, restrictions such as administering parallel tests or imposing test-retest situations are required (see e.g., Jöreskog & Sörbom, 1976 ; Jöreskog, 1978). On the other hand, in the case of negative binomial factor analysis, it is possible to separate communality and specificity if several manifest variables are available for the same subjects. Further, even if only one manifest variable is available, the reliability for the variable can be easily estimated by using (22) or (29). This cannot be done in other models for continuous variables without data reconstructing methods such as split halving.

In some cases, part of the manifest variables may not have specific factors.

Table 5  
Results for Data D

Variable No.	$\delta_j'$		PS		NB	
	Pop.		Est. (SE)	Est. (SE)	Est. (SE)	Est. (SE)
1	.7	.1	.820 (.028)	.298 (.028)	.838 (.069)	.145 (.049)
2	.7	.1	.847 (.022)	.169 (.025)	.621 (.053)	.221 (.050)
3	.7	.1	.730 (.031)	.020 (.024)	.673 (.064)	.040 (.046)
4	.1	.7	.224 (.024)	.856 (.021)	.131 (.048)	.699 (.064)
5	.1	.7	.165 (.025)	.552 (.027)	.129 (.053)	.514 (.065)
6	.1	.7	.010 (.022)	.991 (.024)	.083 (.047)	.801 (.074)

  

Variable No.	$\mu_j$ Pop.	PS		NB		$\alpha_j$ Pop.	NB	
		Est. (SE)		Est. (SE)			Est. (SE)	
1	.0	-.063 (.042)		-.088 (.063)		.4	.570 (.087)	
2	.5	.392 (.038)		.498 (.047)		.4	.672 (.055)	
3	.0	-.036 (.040)		-.028 (.057)		.5	.597 (.078)	
4	.5	.461 (.038)		.559 (.055)		.5	.731 (.063)	
5	.0	.125 (.032)		.136 (.050)		.6	.790 (.061)	
6	.5	.305 (.044)		.441 (.062)		.6	.731 (.075)	

  

Variable No.	Com.		Spe.		Com. $\times$ Spe.		Rel.	
	Pop.	Est.	Pop.	Est.	Pop.	Est.	Pop.	Est.
1	.31	.43	.19	.13	.12	.14	.63	.70
2	.36	.32	.23	.26	.15	.14	.74	.72
3	.29	.29	.22	.18	.14	.11	.65	.58
4	.33	.33	.26	.27	.17	.18	.76	.78
5	.27	.17	.25	.33	.16	.11	.68	.60
6	.31	.38	.28	.22	.18	.20	.78	.81

Note: Pop.=Population values, Est.=Estimates, SE=Standard errors, PS=Poisson model, NB=Negative binomial model, Com.=Communalities, Spe.=Specificities, Rel.=Reliabilities.

For such cases,  $a_j$  for these variables may be set equal to zero. That is, the Poisson distributed variables and other types of discrete and continuous variables can coexist with the negative binomial variables (see Moustaki, 1996; Sammel et al., 1997; Ogasawara, 1998a). In order to confirm  $a_j > 0$ , the chi-square statistic using  $-2 \times \log$  likelihood ratio may be used. The chi-squares for comparison of the Poisson and negative binomial models for Data A, B, C and D are 911, 1141, 623 and 950, respectively with common  $d.f.=6$ , which support the hypothesis that some of the  $a_j$  are not zero for each data set. A convenient method is to compute  $\hat{a}_j / \widehat{SE}(\hat{a}_j)$ . Under  $H_0: a_j = 0$ , the ratio is asymptotically distributed with  $N(0, 1)$ . Though we have to pay attention to the problem of multiple comparison, all  $\hat{a}_j / \widehat{SE}(\hat{a}_j)$  in Tables 2-5 are large enough to reject  $H_0$ .

Similar to the Poisson factor analysis model with correlated common factors (Ogasawara, 1998b), it is possible to include unknown factor correlations and to estimate them in the negative binomial factor analysis model (see Mislevy, 1984).

## REFERENCES

- Archer, C.O. & Jennrich, R.I. (1973). Standard errors for rotated factor loadings. *Psychometrika*, **38**, 581-592.
- 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.
- 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.
- Engel, J. (1984). Models for response data showing extra-Poisson variation. *Statistica Neerlandica*, **38**, 159-167.
- Gardner, W., Mulvey, E.P. & Shaw, E.C. (1995). Regression analysis of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, **118**, 392-404.
- 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.
- Harman, H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.
- Jansen, M.G.H. & Van Duijn, M.A.J. (1992). Extensions of Rasch's multiplicative Poisson model. *Psychometrika*, **57**, 405-414.
- Jöreskog, K.G. (1977). Structural analysis of covariance and correlation matrices. *Psychometrika*, **43**, 443-447.
- Jöreskog, K.G. & Sörbom, D. (1976). Statistical models and methods for test-retest situations. In D.N.M. De Gruijter and L.J.Th. van der Kamp (Eds.), *Advances in psychological and educational measurement* (pp. 135-157). Wiley: New York.
- Lawless, J.F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, **15**, 209-225.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.
- Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, B*, **44**, 226-233.
- Meredith, W. (1971). Poisson distributions of error in mental test theory. *British Journal of Mathematical and Statistical Psychology*, **24**, 49-82.
- Mislevy, R.J. (1984). Estimating latent distributions. *Psychometrika*, **49**, 359-381.
- Mislevy, R.J. & Sheehan, K.M. (1989). Information matrices in latent-variable models. *Journal of Educational Statistics*, **14**, 335-359.
- Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology*, **49**, 313-334.
- Ogasawara, H. (1996a). A proposal of Poisson factor analysis. *Abstracts of the 26th International Congress of Psychology* (p. 421).
- Ogasawara, H. (1996b). Rasch's multiplicative Poisson model with covariates. *Psychometrika*, **61**, 73-92.
- Ogasawara, H. (1998a). A factor analysis model for a mixture of various types of variables. *Behaviormetrika*, **25**, 1-12.
- Ogasawara, H. (1998b). A log-bilinear model with latent variables. *Behaviormetrika*, **25**, 95-110.
- 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).
- Sammel, M.D., Louise, M.R. & Legler, J.M. (1997). Latent variable models for mixed discrete and

- continuous outcomes. *Journal of Royal Statistical Society, B*, **59**, 667-678.
- Silvey, S.D. (1975). *Statistical inference*. London: Chapman & Hall.
- Tsutakawa, R.K. (1988). Mixed model for analyzing geographic variability in mortality rates. *Journal of the American Statistical Association*, **83**, 37-42.
- Van Duijn, M.A.J. (1993). *Mixed models for repeated count data*. Leiden: DSWO Press.

(Received December 1998, Revised July 1999)