PSYCHOMETRIKA—VOL. 59, NO. 3, 381–389
SEPTEMBER 1994

ON THE ESTIMATION OF POLYCHORIC CORRELATIONS AND THEIR ASYMPTOTIC COVARIANCE MATRIX

KARL G. JÖRESKOG

UPPSALA UNIVERSITY

A general theory for parametric inference in contingency tables is outlined. Estimation of polychoric correlations is seen as a special case of this theory. The asymptotic covariance matrix of the estimated polychoric correlations is derived for the case when the thresholds are estimated from the univariate marginals and the polychoric correlations are estimated from the bivariate marginals for given thresholds. Computational aspects are also discussed.

Key words: ordinal variables, polychoric correlations, maximum likelihood, asymptotic covariance matrix.

Theory and applications of structural equation models when some or all of the observed variables are ordinal have been considered by several authors, for example, Muthén (1984), Lee and Poon (1987), Poon and Lee (1987), Lee, Poon, and Bentler (1990), Jöreskog (1990), and Aish and Jöreskog (1990). Typically the estimation of the model is done in two steps. The first step involves estimating polychoric, polyserial and other correlations for the observed variables or rather for the underlying response variables. The second step estimates the parameters of the model by weighted least squares using a weight matrix which must be a consistent estimate of the asymptotic covariance matrix of the correlations estimated in the first step. This paper considers only the first step and the case when all variables are ordinal.

Lee and Poon (1987) and Poon and Lee (1987) assume underlying multivariate normality and estimate the polychoric correlations and the thresholds jointly by minimizing a single fit function (ML or GLS) based on *all* the sample proportions in the k-way contingency table, where k is the number of observed ordinal variables. This requires heavy computations involving the numerical evaluation of the multinormal distribution function. Muthén (1984), Lee, Poon, and Bentler (1990), and Jöreskog (1990) take a simpler approach and estimate the thresholds from the univariate marginals and the polychoric correlations from the bivariate marginals. Thus, different parameters are obtained from *different* fit functions. This is the estimator to be investigated here.

Different formulas for the weight matrix have been given by Muthén (1984, Equation 22) and by Lee, Poon, and Bentler (1990, Equation 11). These involve matrices of the order of total number of parameters and may therefore be computationally cumbersome particularly if the total number of categories is large. This paper gives a procedure for estimating the asymptotic covariance matrix of polychoric correlations which effectively eliminates the thresholds and therefore requires only a matrix of the order of number of estimated correlations. Since it is not necessary to store the raw data in memory, this procedure has the additional advantage of working well in large

The research was supported by the Swedish Council for Research in the Humanities and Social Sciences (HSFR) under the program *Multivariate Statistical Analysis*. The author thanks a reviewer for pointing out an error in the original version of the paper.

Requests for reprints should be sent to Karl G. Jöreskog, Department of Statistics, Uppsala University, PO Box 513, S-75120 Uppsala, SWEDEN.

0033-3123/94/0900-92010$00.75/0

samples and being feasible and relatively straightforward even on computers with limited memory.

A general theory for parametric inference in contingency tables is given in section 2. The estimation of the polychoric correlations is considered in section 3. The asymptotic covariance matrix of the polychoric correlations is derived in section 4.

General Theory

Consider m mutually exclusive events occurring with probabilities $\pi_1, \pi_2, \ldots, \pi_m$, where each $\pi_a = \pi_a(\boldsymbol{\theta})$ is a function of some parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_t)$, $t < m$, in some parameter space. It is assumed that the $\pi_a(\boldsymbol{\theta})$ are positive, continuous, and differentiable functions of $\boldsymbol{\theta}$ and that $\sum \pi_a(\boldsymbol{\theta}) = 1$ at every point $\boldsymbol{\theta}$ of the parameter space.

In a random sample of size N let n_a be the observed frequency (count) of event a and let $p_a = n_a/N$ be the corresponding sample proportion. Furthermore, let $\mathbf{p}' = (p_1, p_2, \ldots, p_m)$, $\boldsymbol{\pi}'(\boldsymbol{\theta}) = [\pi_1(\boldsymbol{\theta}), \pi_2(\boldsymbol{\theta}), \ldots, \pi_m(\boldsymbol{\theta})]$, and $\mathbf{A}_\theta$ be the matrix of order $m \times t$ with elements $(\partial \pi_i / \partial \theta_j)$. Consider the fit function

$$F[\mathbf{p}, \boldsymbol{\pi}(\boldsymbol{\theta})] = \sum_{a=1}^{m} p_a[(\ln p_a - \ln \pi_a(\boldsymbol{\theta})], \tag{1}$$

to be minimized with respect to $\boldsymbol{\theta}$. If $n_a = 0$, set $p_a \ln p_a = 0$ in (1). Minimizing $F(\boldsymbol{\theta})$ is equivalent to maximizing the log likelihood function

$$\ln L = \sum_{a=1}^{m} n_a \ln \pi_a(\boldsymbol{\theta}) = N \sum_{a=1}^{m} p_a \ln \pi_a(\boldsymbol{\theta}).$$

The gradient vector and information matrix are

$$\frac{\partial F}{\partial \boldsymbol{\theta}} = -\sum_a \left(\frac{p_a}{\pi_a}\right)\left(\frac{\partial \pi_a}{\partial \boldsymbol{\theta}}\right) = -\mathbf{A}'_\theta \mathbf{D}_\pi^{-1} \mathbf{p}, \tag{2}$$

$$\mathbf{E}_\theta = \sum_a \left(\frac{1}{\pi_a}\right)\left(\frac{\partial \pi_a}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \pi_a}{\partial \boldsymbol{\theta}}\right)' = \mathbf{A}'_\theta \mathbf{D}_\pi^{-1} \mathbf{A}_\theta, \tag{3}$$

where $\mathbf{D}_\pi = \text{diag}\,[\pi_1(\boldsymbol{\theta}), \pi_2(\boldsymbol{\theta}), \ldots, \pi_m(\boldsymbol{\theta})]$.

Assume that $\mathbf{p}$ converges in probability to $\boldsymbol{\pi}_0$ as $N \to \infty$. Let $\boldsymbol{\theta}_0$ be the value of $\boldsymbol{\theta}$ that minimizes $F[\boldsymbol{\pi}_0, \boldsymbol{\pi}(\boldsymbol{\theta})]$ and assume that $\mathbf{A}_0 = \mathbf{A}_{\theta=\theta_0}$ is of rank t. We say that the model holds if $\boldsymbol{\pi}_0 = \boldsymbol{\pi}(\boldsymbol{\theta}_0)$, which is the case considered here. The noncentral case when $\boldsymbol{\pi}_0 \neq \boldsymbol{\pi}(\boldsymbol{\theta}_0)$ will be considered in another paper. Furthermore, let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator of $\boldsymbol{\theta}$, that is, the value of $\boldsymbol{\theta}$ that minimizes $F[\mathbf{p}, \boldsymbol{\pi}(\boldsymbol{\theta})]$. Then:

1. $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$,
2. $N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically distributed as

$$N^{1/2}(\mathbf{A}'_0 \mathbf{D}_0^{-1} \mathbf{A}_0)^{-1} \mathbf{A}'_0 \mathbf{D}_0^{-1} [\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\theta}_0)], \tag{4}$$

where $\mathbf{D}_0 = \text{diag}\,[\pi_1(\boldsymbol{\theta}_0), \pi_2(\boldsymbol{\theta}_0), \ldots, \pi_m(\boldsymbol{\theta}_0)]$,

3. If the model holds, then $N^{1/2}(\hat{\boldsymbol\theta} - \boldsymbol\theta_0)$ is asymptotically normally distributed with mean vector zero and covariance matrix $\mathbf{E}_0^{-1} = \mathbf{E}_{\theta=\theta_0}^{-1} = (\mathbf{A}_0'\mathbf{D}_0^{-1}\mathbf{A}_0)^{-1}$.

Furthermore, let $\boldsymbol\theta$ be partitioned as $\boldsymbol\theta = (\boldsymbol\theta_1, \boldsymbol\theta_2)$, where $\boldsymbol\theta_1$ is of order t_1 and $\boldsymbol\theta_2$ is of order t_2, $t_1 + t_2 = t$, and let $\boldsymbol\theta_0 = (\boldsymbol\theta_{10}, \boldsymbol\theta_{20})$. Assume that $\tilde{\boldsymbol\theta}_2$ is a consistent estimator of $\boldsymbol\theta_{20}$ independent of $\boldsymbol\theta_1$ and let $\hat{\boldsymbol\theta}_1 = \hat{\boldsymbol\theta}_1(\tilde{\boldsymbol\theta}_2)$ be the value of $\boldsymbol\theta_1$ that minimizes $F[\mathbf{p}, \boldsymbol\pi(\boldsymbol\theta_1, \tilde{\boldsymbol\theta}_2)]$, that is, $\hat{\boldsymbol\theta}_1$ is a pseudo maximum likelihood estimator (Parke, 1986) of $\boldsymbol\theta_1$ for given $\boldsymbol\theta_2 = \tilde{\boldsymbol\theta}_2$. Let $\mathbf{A}_0 = [\mathbf{A}_{10}, \mathbf{A}_{20}]$, where $\mathbf{A}_{10} = \partial\boldsymbol\pi/\partial\boldsymbol\theta_1'$ and $\mathbf{A}_{20} = \partial\boldsymbol\pi/\partial\boldsymbol\theta_2'$ evaluated at $\boldsymbol\theta_0$. If $\mathbf{A}_{10}$ has rank t_1 then:

4. $N^{1/2}(\hat{\boldsymbol\theta}_1 - \boldsymbol\theta_{10})$ is asymptotically distributed as

$$N^{1/2}\{(\mathbf{A}_{10}'\mathbf{D}_0^{-1}\mathbf{A}_{10})^{-1}\mathbf{A}_{10}'\mathbf{D}_0^{-1}[\mathbf{p} - \boldsymbol\pi(\boldsymbol\theta_0)] - (\mathbf{A}_{10}'\mathbf{D}_0^{-1}\mathbf{A}_{10})^{-1}\mathbf{A}_{10}'\mathbf{D}_0^{-1}\mathbf{A}_{20}(\tilde{\boldsymbol\theta}_2 - \boldsymbol\theta_{20})\} \tag{5}$$

and the two terms in (5) are asymptotically independent.

Proofs of these propositions are sketched in the Appendix.

Estimation of Polychoric Correlations

Observations on an ordinal variable are assumed to represent responses to a set of ordered categories, such as a five-category Likert scale. It is only assumed that a person who responds in one category has more of a characteristic than a person who responds in a lower category.

Ordinal variables are not continuous variables and should not be treated as if they are. Ordinal variables do not have origins or units of measurements. Means, variances, and covariances of ordinal variables have no meaning.

It is common practice to treat numbers like 1, 2, 3, 4, representing the ordered categories of an ordinal variable as numbers on an interval scale and use a covariance matrix computed in the usual way to estimate a structural equation model. What is so bad with this is not so much that the distribution is non-normal; more importantly the distribution is not continuous: there are only four distinct values in the distribution. The use of ordinal variables in structural equation models requires other techniques than those which are used for continuous variables.

For an ordinal variable z, it is assumed that there is an *underlying* continuous variable z^* which is normally distributed with mean μ_z and variance σ_z^2. Muthén (1984) call these underlying variables latent response variables. They are not the same as the latent variables in the structural equation model. Assuming that there are m categories on z, we write $z = i$ to mean that z belongs to category i. The actual score values in the data may be arbitrary and are irrelevant as long as the ordinal information is retained. That is, low scores correspond to low-order categories of z that are associated with smaller values of z^* and high scores correspond to high-order categories that are associated with larger values of z^*.

The connection between z and z^* is

$$z = i \Leftrightarrow \tau_{i-1} < z^* \le \tau_i,\ i = 1, 2, \ldots, m,$$

where

$$\tau_0 = -\infty,\ \tau_1 < \tau_2 < \cdots < \tau_{m-1},\ \tau_m = +\infty,$$

are parameters called threshold values. If there are m categories, there are $m - 1$ threshold parameters.

Since only ordinal information is available about z^*, the mean μ_z and variance σ_z^2 are usually not identified and are therefore set to zero and one, respectively. However, when the same ordinal variable is measured one or more times, as in longitudinal or panel studies and in multigroup studies, it is possible to estimate the means and variances of these variables (relative to a fixed origin and scale) by specifying the thresholds to be the same for the same variable over time and/or groups. In the following we assume that $\mu_z = 0$ and that $\sigma_z^2 = 1$. Otherwise, replace $\tau_i^{(z)}$ by $(\tau_i^{(z)} - \mu_z)/\sigma_z$ in what follows.

Consider k ordinal variables $z_1, z_2, \ldots, z_k$ with $m_1, m_2, \ldots, m_k$ categories, respectively. Altogether there are $\sum_{i=1}^{k} m_i - k + k(k - 1)/2$ parameters to be estimated, namely the thresholds $(\tau_1^{(g)}, \tau_2^{(g)}, \ldots, \tau_{m_g-1}^{(g)})$, g = 1, 2, ..., k, and the polychoric correlations ρ_{gh}, $h < g$. The parameters are usually estimated from the univariate and bivariate marginal likelihoods, that is, the thresholds are estimated from the univariate marginal distribution and the polychoric correlations from the bivariate marginal distributions for given thresholds; see Olsson (1979), Muthén (1984), and Jöreskog and Sörbom (1988). The univariate and bivariate marginal likelihoods all have the general form given in the previous section.

Olsson (1979) considered the case $k = 2$ and studied two methods for estimating the parameters:

i. Estimate the thresholds and the polychoric correlation jointly from the bivariate marginal distribution.
ii. Estimate the thresholds from the univariate marginal distribution and then the polychoric correlation from the bivariate marginal distribution for given thresholds.

In both methods, an iterative procedure must be used to estimate the parameters. Practical experience suggests that the two methods give almost identical estimates. Method ii. is computationally simple and is used most often in practice. This is the method considered here. Method i. would have the disadvantage that different estimates of thresholds for one variable may be obtained from different pairs of variables where this variable is included.

The model for the univariate marginal of variable g is

$$\pi_a^{(g)}(\boldsymbol{\theta}) = \int_{\tau_{a-1}^{(g)}}^{\tau_a^{(g)}} \phi(u)du, \tag{6}$$

where $\phi(u)$ is the standard normal density function. The parameter vector is

$$\boldsymbol{\theta} = \boldsymbol{\tau}_g = (\tau_1^{(g)}, \tau_2^{(g)}, \ldots, \tau_{m_g-1}^{(g)}).$$

Application of the general theory gives the maximum likelihood estimator $\tilde{\boldsymbol{\tau}}_g$ of $\boldsymbol{\tau}_g$ with asymptotic covariance matrix $\boldsymbol{\Psi}_{gg}/N$, say. The maximum likelihood estimator is given explicitly as

$$\tilde{\tau}_a^{(g)} = \Phi^{-1}(p_1^{(g)} + p_2^{(g)} + \cdots + p_a^{(g)}), \; a = 1, \ldots, m_g - 1,$$

where Φ^{-1} is the inverse of the standard normal distribution function, and where $p_i^{(g)}$, $i = 1, 2, \ldots, m_g$, are the sample proportions in the univariate marginal distribution of variable g. Obviously, since $p_i^{(g)}$ converges in probability to the corresponding population proportion $\pi_i^{(g)}$, $\bar{\tau}_a^{(g)}$ is consistent.

According to Proposition (b), $\bar{\boldsymbol{\tau}}_g$ is asymptotically linear in the proportions $\mathbf{p}_g$ of the univariate marginal distribution of variable g. Let $\mathbf{A}_g$ be the matrix of order $m_g \times m_g - 1$ (here illustrated with $m_g = 5$):

$$\mathbf{A}_g = \begin{pmatrix} \phi_1 & 0 & 0 & 0 \\ -\phi_1 & \phi_2 & 0 & 0 \\ 0 & -\phi_2 & \phi_3 & 0 \\ 0 & 0 & -\phi_3 & \phi_4 \\ 0 & 0 & 0 & -\phi_4 \end{pmatrix},$$

where $\phi_i = \phi(\tau_i^{(g)})$, and let $\mathbf{D}_\pi$ be the diagonal matrix $\mathbf{D}_\pi = \text{diag}\,(\pi_1^{(g)}, \pi_2^{(g)}, \ldots, \pi_m^{(g)})$, both matrices evaluated at $\boldsymbol{\theta}_0$. Then $\bar{\boldsymbol{\tau}}_g \sim (\mathbf{A}_g'\mathbf{D}_\pi^{-1}\mathbf{A}_g)^{-1}\mathbf{A}_g'\mathbf{D}_\pi^{-1}\mathbf{p}_g = \mathbf{B}_g'\mathbf{p}_g$, say.

Let $\boldsymbol{\Psi}_{gh}$ be N times the asymptotic covariance matrix of $\bar{\boldsymbol{\tau}}_g$ and $\bar{\boldsymbol{\tau}}_h$. Then:

$$\boldsymbol{\Psi}_{gh} = \mathbf{B}_g'\boldsymbol{\pi}_{gh}\mathbf{B}_h, \tag{7}$$

since Cov $(\mathbf{p}_g, \mathbf{p}_h) = \boldsymbol{\pi}_{gh} - \boldsymbol{\pi}_g\boldsymbol{\pi}_h'$, and $\mathbf{B}_g'\boldsymbol{\pi}_g = \mathbf{0}$, where $\boldsymbol{\pi}_{gh}$ is a matrix of population probabilities of the bivariate marginal distribution of variables g and h and $\boldsymbol{\pi}_g$ and $\boldsymbol{\pi}_h$ are vectors of population probabilities of the univariate marginal distributions of variables g and h. Equation (7) holds for $g \neq h$. It holds for $g = h$ as well, if $\boldsymbol{\pi}_{gh}$ in (7) is interpreted as $\mathbf{D}_\pi$.

The model for the bivariate marginal of variables g and h is

$$\pi_{ab}^{(gh)}(\boldsymbol{\theta}) = \int_{\tau_{a-1}^{(g)}}^{\tau_a^{(g)}} \int_{\tau_{b-1}^{(h)}}^{\tau_b^{(h)}} \phi_2(u, v; \rho_{gh})du\,dv, \tag{8}$$

where $\phi_2(u, v; \rho)$ is the density function of the standardized bivariate normal distribution with correlation ρ. The parameter vector is

$$\boldsymbol{\theta} = \boldsymbol{\theta}_{gh} = (\tau_1^{(g)}, \tau_2^{(g)}, \ldots, \tau_{m_g-1}^{(g)}, \tau_1^{(h)}, \tau_2^{(h)}, \ldots, \tau_{m_h-1}^{(h)}, \rho_{gh}) = (\boldsymbol{\tau}_g, \boldsymbol{\tau}_h, \rho_{gh}), \tag{9}$$

consisting of the thresholds for the two variables and the polychoric correlation ρ_{gh}.

To maximize the bivariate likelihood, the fit function

$$F(\rho, \bar{\boldsymbol{\tau}}_g, \bar{\boldsymbol{\tau}}_h) = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} p_{ab}^{(gh)}(\ln p_{ab}^{(gh)} - \ln \pi_{ab}^{(gh)}), \tag{10}$$

is minimized with respect to ρ for given $\bar{\boldsymbol{\tau}}_g$ and $\bar{\boldsymbol{\tau}}_h$. Here $p_{ab}^{(gh)}$ are the sample proportions in the bivariate marginal distribution for variables g and h. The value of ρ that minimizes F is the estimate $\hat{\rho}_{gh}$ of the polychoric correlation $\rho_{gh}^{(0)}$. This estimate satisfies the equation

$$g(\hat{\rho}, \bar{\boldsymbol{\tau}}_g, \bar{\boldsymbol{\tau}}_h) = 0, \tag{11}$$

where $g = \partial F/\partial\rho$.

Application of (5) shows that

$$(\hat{\rho}_{gh} - \rho_{gh}^{(0)}) \sim \text{tr}\,[\boldsymbol{\alpha}_{gh}'(\mathbf{p}_{gh} - \boldsymbol{\pi}_{gh})] + \boldsymbol{\beta}_g'^{(gh)}(\bar{\boldsymbol{\tau}}_g - \boldsymbol{\tau}_g^{(0)}) + \boldsymbol{\beta}_h'^{(gh)}(\bar{\boldsymbol{\tau}}_h - \boldsymbol{\tau}_h^{(0)}), \tag{12}$$

where $\boldsymbol{\alpha}_{gh}$ is a matrix of order $m_g \times m_h$, and $\boldsymbol{\beta}_g^{(gh)}$ and $\boldsymbol{\beta}_h^{(gh)}$ are vectors of order $m_g - 1$ and $m_h - 1$, respectively, and $\mathbf{p}_{gh}$ and $\boldsymbol{\pi}_{gh}$ are matrices of sample proportions and probabilities of the bivariate distribution of variables g and h. Since $\mathbf{p}_{gh}$, $\bar{\boldsymbol{\tau}}_g$, and $\bar{\boldsymbol{\tau}}_h$ are asymptotically normal, it follows that $\hat{\rho}_{gh}$ is asymptotically normal. Proposition (d) states that the first term is asymptotically independent of the other terms.

A typical element of $\boldsymbol{\alpha}_{gh}$ is

$$\alpha_{ab}^{(gh)} = D^{-1} \frac{1}{\pi_{ab}^{(gh)}} \frac{\partial \pi_{ab}^{(gh)}}{\partial \rho_{gh}}, \tag{13}$$

and typical elements of $\boldsymbol{\beta}_g^{(gh)}$ and $\boldsymbol{\beta}_h^{(gh)}$ are

$$\beta_{gi}^{(gh)} = D^{-1} \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \frac{1}{\pi_{ab}^{(gh)}} \frac{\partial \pi_{ab}^{(gh)}}{\partial \rho_{gh}} \frac{\partial \pi_{ab}^{(gh)}}{\partial \tau_i^{(g)}}, \tag{14}$$

$$\beta_{hj}^{(gh)} = D^{-1} \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \frac{1}{\pi_{ab}^{(gh)}} \frac{\partial \pi_{ab}^{(gh)}}{\partial \rho_{gh}} \frac{\partial \pi_{ab}^{(gh)}}{\partial \tau_j^{(h)}}, \tag{15}$$

where

$$D = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \frac{1}{\pi_{ab}^{(gh)}} \left(\frac{\partial \pi_{ab}^{(gh)}}{\partial \rho_{gh}} \right)^2.$$

These quantities are to be evaluated at $\boldsymbol{\theta}_0$. The required derivatives are given by Olsson (1979).

Let $u_{gh} = \mathrm{tr}\,(\boldsymbol{\alpha}'_{gh}\mathbf{p}_{gh})$. This is asymptotically uncorrelated with $\bar{\boldsymbol{\tau}}_g$ and $\bar{\boldsymbol{\tau}}_h$, so that (12) represents the regression of $\hat{\rho}_{gh}$ on $\bar{\boldsymbol{\tau}}_g$ and $\bar{\boldsymbol{\tau}}_h$ with residual u_{gh}. It does not follow, however, that u_{gh} is asymptotically uncorrelated with $\bar{\boldsymbol{\tau}}_i$ for $i \neq g$ and $i \neq h$, despite the fact that $\hat{\rho}_{gh}$ does not depend on $\bar{\boldsymbol{\tau}}_i$. Therefore, although (12) can be used to derive the asymptotic variance of $\hat{\rho}_{gh}$, it can not be used directly to obtain the asymptotic covariance between different polychoric correlations.

Equation (12) can be developed further, however. Since $\bar{\boldsymbol{\tau}}_g$ and $\bar{\boldsymbol{\tau}}_h$ are asymptotically linear in $\mathbf{p}_g$ and $\mathbf{p}_h$, respectively, which in turn are linear in $\mathbf{p}_{gh}$, it follows that $\hat{\rho}_{gh}$ is asymptotically linear in $\mathbf{p}_{gh}$.

Let

$$\boldsymbol{\Gamma}_{gh} = \boldsymbol{\alpha}_{gh} + \mathbf{B}_g \boldsymbol{\beta}_g^{(gh)} \mathbf{1}'_h + \mathbf{1}_g \boldsymbol{\beta}'^{(gh)}_h \mathbf{B}'_h, \tag{16}$$

where $\mathbf{1}_g$ is a column vector of order m_g with all elements equal to 1. Then

$$\hat{\rho}_{gh} \sim \mathrm{tr}\,(\boldsymbol{\Gamma}'_{gh}\mathbf{p}_{gh}). \tag{17}$$

Gunsjö (1994) gives a more direct derivation of an expression for $\hat{\rho}_{gh}$ which can be shown to be equivalent to (17).

Asymptotic Covariance Matrix

The estimated thresholds and polychoric correlations are all asymptotically linear in the sample proportions of the univariate and bivariate marginal distributions as shown in section 3. Since these proportions are linear in all the sample proportions of the k-way contingency table, it follows that the joint distribution of all the estimated parameters is asymptotically normal.

Let

$$\hat{\boldsymbol{\tau}} = (\hat{\boldsymbol{\tau}}_1, \hat{\boldsymbol{\tau}}_2, \ldots, \hat{\boldsymbol{\tau}}_k),$$

be the vector of all estimated thresholds, and let

$$\hat{\boldsymbol{\rho}} = (\hat{\rho}_{21}, \hat{\rho}_{31}, \hat{\rho}_{32}, \hat{\rho}_{41}, \hat{\rho}_{42}, \hat{\rho}_{43}, \ldots, \hat{\rho}_{k,k-1}),$$

be the vector of estimated polychoric correlations. Our objective is to find the asymptotic covariance matrix of $\hat{\boldsymbol{\rho}}$. Using (17), this is obtained as

$$NA\text{Cov}\,(\hat{\rho}_{gh}, \hat{\rho}_{ij}) = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \sum_{c=1}^{m_i} \sum_{d=1}^{m_j} \gamma_{ab}^{(gh)} N \text{ Cov}\,(p_{ab}^{(gh)}, p_{cd}^{(ij)}) \gamma_{cd}^{(ij)}. \tag{18}$$

To evaluate (18), note that

$$N \text{ Cov}\,(p_{ab}^{(gh)}, p_{cd}^{(ij)}) = \pi_{abcd}^{(ghij)} - \pi_{ab}^{(gh)} \pi_{cd}^{(ij)}.$$

where $\pi_{abcd}^{(ghij)}$ are the probabilities of the four-way contingency table for variables g, h, i, j, which can be estimated consistently by the corresponding sample proportions $p_{abcd}^{(ghij)}$. However, as shown in the next section the four-way contingency tables need not be obtained. Let

$$\omega_{gh} = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \gamma_{ab}^{(gh)} \pi_{ab}^{(gh)}. \tag{19}$$

Then

$$NA\text{Cov}\,(\hat{\rho}_{gh}, \hat{\rho}_{ij}) = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \sum_{c=1}^{m_i} \sum_{d=1}^{m_j} \gamma_{ab}^{(gh)} \pi_{abcd}^{(ghij)} \gamma_{cd}^{(ij)} - \omega_{gh} \omega_{ij}, \tag{20}$$

which may be estimated as

$$Est[NA\text{Cov}\,(\hat{\rho}_{gh}, \hat{\rho}_{ij})] = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \sum_{c=1}^{m_i} \sum_{d=1}^{m_j} \hat{\gamma}_{ab}^{(gh)} p_{abcd}^{(ghij)} \hat{\gamma}_{cd}^{(ij)} - \hat{\omega}_{gh} \hat{\omega}_{ij}, \tag{21}$$

where $\hat{\gamma}_{ab}^{(gh)}$ and $\hat{\omega}_{gh}$ are (16) and (19) evaluated at $\hat{\boldsymbol{\theta}}_{gh} = (\tilde{\boldsymbol{\tau}}_g, \tilde{\boldsymbol{\tau}}_h, \hat{\rho}_{gh})$. Equations (20) and (21) hold for every pair of variables gh and ij with $g \neq h$ and $i \neq j$.

Computational Aspects

The asymptotic covariance matrix can only be estimated from a large sample. However, all the computations can be done without storing the raw data in memory. The following procedure works with samples of unlimited size.

Instead of storing the raw data in memory, one can store all the univariate and bivariate contingency tables, which can be obtained by reading the raw data once. All parameter estimates can be obtained quickly from these contingency tables. After each bivariate likelihood has been maximized, the estimates of $\boldsymbol{\alpha}_{gh}$, $\boldsymbol{\beta}_g^{(gh)}$, and $\boldsymbol{\beta}_h^{(gh)}$ are computed. Since this is done for each bivariate contingency table separately, these quantities can be stored in the same space for all contingency tables. Then $\hat{\boldsymbol{\Gamma}}_{gh}$ is computed and saved in the same space as the bivariate contingency table $\mathbf{p}_{gh}$. Finally,

$\hat{\omega}_{gh}$ is computed. This is a scalar for each contingency table. The asymptotic covariance matrix, which is of order $k(k-1)/2 \times k(k-1)/2$, is then computed as follows.

Let $\kappa_{vghij} = 1/N$, if $z_{vg} = a$, $z_{vh} = b$, $z_{vi} = c$, $z_{vj} = d$ and $\kappa_{vghij} = 0$, otherwise, where z_{vg} is the score fore case v on variable g in the raw data. Then, the first term of (21) is

$$\sum_{v=1}^{N} \kappa_{vghij} \hat{\gamma}_{ab}^{(gh)} \hat{\gamma}_{cd}^{(ij)}.$$

Thus, this is obtained by reading the raw data a second time and for each case multiplying $\hat{\gamma}_{ab}^{(gh)}$ and $\hat{\gamma}_{cd}^{(ij)}$ for all combinations of g, h, i, j and cumulating over all cases in the data. The second term in (21) is simply $-\hat{\omega}_{gh}\hat{\omega}_{ij}$.

Appendix

Sketches of proofs of Propositions 1 through 4 follows.

1. Consider $F[\mathbf{p}, \boldsymbol{\pi}(\boldsymbol{\theta})]$ and $F[\boldsymbol{\pi}_0, \boldsymbol{\pi}(\boldsymbol{\theta})]$ as functions of $\boldsymbol{\theta}$. Since $F[\mathbf{p}, \boldsymbol{\pi}(\boldsymbol{\theta})]$ converges uniformly in probability to $F[\boldsymbol{\pi}_0, \boldsymbol{\pi}(\boldsymbol{\theta})]$ and $F[\boldsymbol{\pi}_0, \boldsymbol{\pi}(\boldsymbol{\theta})]$ has a unique minimum at $\boldsymbol{\theta}_0$, $\hat{\boldsymbol{\theta}}$ must converge to $\boldsymbol{\theta}_0$.
2. Expanding $\partial F/\partial \boldsymbol{\theta}$ at $\hat{\boldsymbol{\theta}}$ around $\boldsymbol{\theta}_0$ to linear terms, equating this to $\mathbf{0}$, and solving for $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ gives the required result.
3. If the model holds, $\boldsymbol{\pi}(\boldsymbol{\theta}_0) = \boldsymbol{\pi}_0$. Result (2) then shows that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ is asymptotically linear in $\mathbf{p} - \boldsymbol{\pi}_0$. Since $\mathbf{p} - \boldsymbol{\pi}_0$ is asymptotically normal with mean vector zero and covariance matrix $\mathbf{D}_0 - \boldsymbol{\pi}_0\boldsymbol{\pi}_0'$, the result follows by noting that $\mathbf{A}_0'\mathbf{D}_0^{-1}\boldsymbol{\pi}_0 = \mathbf{0}$.
4. Writing $\mathbf{A}_\theta = [\mathbf{A}_{1\theta}, \mathbf{A}_{2\theta}]$, where $\mathbf{A}_{1\theta} = \partial\boldsymbol{\pi}/\partial\boldsymbol{\theta}_1'$ and $\mathbf{A}_{2\theta} = \partial\boldsymbol{\pi}/\partial\boldsymbol{\theta}_2'$, the first derivative of $f(\boldsymbol{\theta}_1) = F[\mathbf{p}, \boldsymbol{\pi}(\boldsymbol{\theta}_1, \bar{\boldsymbol{\theta}}_2)]$ is

$$\frac{\partial f}{\partial \boldsymbol{\theta}_1} = -\mathbf{A}_{1\theta}'\mathbf{D}_\pi^{-1}\mathbf{p} = -\mathbf{A}_{1\theta}'\mathbf{D}_\pi^{-1}[\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\theta}_1, \bar{\boldsymbol{\theta}}_2)].$$

$\hat{\boldsymbol{\theta}}_1$ must therefore satisfy

$$\hat{\mathbf{A}}_{1\theta}'\hat{\mathbf{D}}_\pi^{-1}[\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2)] = \mathbf{0}, \tag{22}$$

where $\hat{\mathbf{A}}_{1\theta}$ and $\hat{\mathbf{D}}_\pi$ are $\mathbf{A}_{1\theta}$ and $\mathbf{D}_\pi$ evaluated at $[\hat{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2]$. Using the identity

$$\mathbf{p} - \boldsymbol{\pi}(\hat{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2) = [\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\theta}_{10}, \bar{\boldsymbol{\theta}}_2)] - [\boldsymbol{\pi}(\hat{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2) - \boldsymbol{\pi}(\boldsymbol{\theta}_{10}, \bar{\boldsymbol{\theta}}_2)],$$

it is seen that (22) is equivalent to

$$\hat{\mathbf{A}}_{1\theta}'\hat{\mathbf{D}}_\pi^{-1}[\boldsymbol{\pi}(\hat{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2) - \boldsymbol{\pi}(\boldsymbol{\theta}_{10}, \bar{\boldsymbol{\theta}}_2)] = \hat{\mathbf{A}}_{1\theta}'\hat{\mathbf{D}}_\pi^{-1}[\mathbf{p} - \boldsymbol{\pi}(\boldsymbol{\theta}_{10}, \bar{\boldsymbol{\theta}}_2)]. \tag{23}$$

By Taylor expansion,

$$\boldsymbol{\pi}(\hat{\boldsymbol{\theta}}_1, \bar{\boldsymbol{\theta}}_2) - \boldsymbol{\pi}(\boldsymbol{\theta}_{10}, \bar{\boldsymbol{\theta}}_2) = \mathbf{A}_{1*}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}), \tag{24}$$

$$\boldsymbol{\pi}(\boldsymbol{\theta}_{10}, \bar{\boldsymbol{\theta}}_2) - \boldsymbol{\pi}(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20}) = \mathbf{A}_{2*}(\bar{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20}), \tag{25}$$

where $\mathbf{A}_{1*}$ is $\mathbf{A}_{1\theta}$ evaluated at $[\boldsymbol{\theta}_1^*, \bar{\boldsymbol{\theta}}_2]$ with $\boldsymbol{\theta}_1^*$ between $\hat{\boldsymbol{\theta}}_1$ and $\boldsymbol{\theta}_{10}$ and where $\mathbf{A}_{2*}$ is $\mathbf{A}_{2\theta}$ evaluated at $[\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_2^*]$ with $\boldsymbol{\theta}_2^*$ between $\bar{\boldsymbol{\theta}}_2$ and $\boldsymbol{\theta}_{20}$. Substituting (24) and (25) into (23), gives

$$\hat{\mathbf{A}}_{1\theta}'\hat{\mathbf{D}}_\pi^{-1}\mathbf{A}_{1*}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) = \hat{\mathbf{A}}_{1\theta}'\hat{\mathbf{D}}_\pi^{-1}(\mathbf{p} - \boldsymbol{\pi}_0) - \hat{\mathbf{A}}_{1\theta}'\hat{\mathbf{D}}_\pi^{-1}\mathbf{A}_{2*}(\bar{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20}), \tag{26}$$

Equation (5) follows from (26) by noting that $plim(\hat{\mathbf{A}}'_{1\theta}\hat{\mathbf{D}}_\pi^{-1}\mathbf{A}_{1*}) = \mathbf{A}'_{10}\mathbf{D}_0^{-1}\mathbf{A}_{10}$, $plim(\hat{\mathbf{A}}'_{1\theta}\hat{\mathbf{D}}_\pi^{-1}) = \mathbf{A}'_{10}\mathbf{D}_0^{-1}$, and $plim(\hat{\mathbf{A}}'_{1\theta}\hat{\mathbf{D}}_\pi^{-1}\mathbf{A}_{2*}) = \mathbf{A}'_{10}\mathbf{D}_0^{-1}\mathbf{A}_{20}$ and that $\mathbf{A}'_{10}\mathbf{D}_0^{-1}\mathbf{A}_{10}$ is nonsingular. $\mathbf{A}_{1*}$ and $\mathbf{A}_{2*}$ can be estimated consistently by $\hat{\mathbf{A}}_{1\theta}$ and $\hat{\mathbf{A}}_{2\theta}$, respectively, so that a linearized version of the pseudo maximum likelihood estimator is

$$\hat{\boldsymbol{\theta}}_1(\tilde{\boldsymbol{\theta}}_2) = (\hat{\mathbf{A}}'_{1\theta}\hat{\mathbf{D}}_\pi^{-1}\hat{\mathbf{A}}_{1\theta})^{-1}\hat{\mathbf{A}}'_{1\theta}\hat{\mathbf{D}}_\pi^{-1}\mathbf{p} - (\hat{\mathbf{A}}'_{1\theta}\hat{\mathbf{D}}_\pi^{-1}\hat{\mathbf{A}}_{1\theta})^{-1}\hat{\mathbf{A}}'_{1\theta}\hat{\mathbf{D}}_\pi^{-1}\hat{\mathbf{A}}_{2\theta}\tilde{\boldsymbol{\theta}}_2. \quad (27)$$

In a more general context, Pierce (1982, section 4) shows that the two terms in (27) are asymptotically independent, because $\tilde{\boldsymbol{\theta}}_2$ is consistent and the first term in (27) is the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_1(\boldsymbol{\theta}_{20})$ of $\boldsymbol{\theta}_{10}$ that would be obtained if $\boldsymbol{\theta}_2$ was known and equal to its population value.

References

Aish, A. M., & Jöreskog, K. G. (1990). A panel model for political efficacy and responsiveness: An application of LISREL7 with weighted least squares. *Quality and Quantity, 24*, 405–426.

Gunsjö, A. (1994). *Faktoranalys av ordinala variabler.* Doctoral dissertation. Uppsala University, Department of Statistics. (In Swedish)

Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity, 24*, 387–404.

Jöreskog, K. G., & Sörbom, D. (1988). *PRELIS—A program for multivariate data screening and data summarization. A preprocessor for LISREL* (2nd ed.). Chicago, IL: Scientific Software.

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

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.

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

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443–460.

Parke, W. P. (1986). Pseudo maximum likelihood estimation: The asymptotic distribution. *The Annals of Statistics, 14*, 355–357.

Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statistics, 10*, 475–478.

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

Manuscript received 3/2/92
Final version received 9/15/93