Japanese Psychological Research
1976, Vol. 18, No. 2, 82–90

A STATISTICAL PROCEDURE FOR THE LATENT PROFILE MODEL

YOSHIO TAKANE[1]

Department of Psychology, University of Tokyo

A maximum likelihood estimation procedure is presented for the latent profile model when the conditional distributions of manifest variates given specific latent classes are assumed to be normal, together with the associated tests of the goodness of fit. By incorporating means of constraining estimated parameters in various ways, an important class of statistical hypotheses, constant and equality, about the structures of latent classes can be tested. Illustrative examples are included with the results suggesting strongly the usefulness of the current approach.

Gibson (1959), proposing the latent profile model as an extention of the latent class model of the latent structure analysis, discussed possible failures of the usual (linear) factor analysis. One such situation is when test items with wide range of difficulties are factor-analyzed simultaneously. Due to the nonlinear relations between test scores, some superfluous factors tend to appear (McDonald, 1965). This is called a difficulty factor problem.

The latent profile model (LPM) presents one approach to nonlinearity problems (Anderson, 1959; McDonald, 1962); it enforces linearity of the relations between manifest variates and latent variates by constraining the latter to be dichotomous. Since $0^n=0$, $1^n=1$ and $0^n1^m=0$ for n and m being any natural numbers, and by the orthogonality conditions of latent variates (no observational units belong to more than one latent class simultaneously), all nonlinear terms of polynomials in latent variates necessarily vanish. Thus, a polynomial model always reduces to a linear one (Takane, 1972).

In addition to the ability of coping with the nonlinearity which exists among data elements LPM has a conceptual advantage over the usual factor analysis model with its rotational indeterminacy. LPM solutions, on the other hand, are unique (or at least locally identifiable in the sense of McHugh, 1956) under some general regularity conditions (Neyman, 1949). Moreover, the natural interpretation of factors (latent variates) is already embedded in the structural features of the model; i.e., they represent latent classes from which each observational unit is recruited.

In this paper we give a maximum likelihood estimation (MLE) procedure for the latent profile model when within-class distributions of manifest variates are assumed to be normal. The major consequences of the procedure are that the likelihood ratio statistic is now available for the test of the goodness of fit of the model and for the test of the number of latent classes, and that specific hypotheses about the structures of latent classes, which can be expressed by linear equality constraints, can be tested. This latter case permits confirmatory type analyses (Jöreskog, 1969) with the latent profile model. Besides MLE generally yields estimates of parameters with statistically better pro-

[1] I would like to thank Drs. Tanaka, Shiba and Yanai for their helpful comments on earlier draft of this paper. A request for the reprint should be directed to Dept. of Psychology, Univ. of Tokyo, 7–3–1, Hongo, Bunkyo-ku, Tokyo.

perties; it gives BAN (Best Asymptotically Normal) estimates, though not necessarily unbiased, while the existing procedures which are applicable to the latent profile model provides only CAN (Consistent Asymptotically Normal) estimates (Anderson, 1954).

Estimation Procedure

We state the latent profile model as

$$f(\mathbf{x}_k;\ \boldsymbol{\theta}) = \sum_{a=1}^{t} f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a) g_a \tag{1}$$

where $f(\mathbf{x}_k;\ \boldsymbol{\theta})$ is the marginal density of $\mathbf{x}_k$ which is an n-component vector of k'th observation on n manifest variates (N is the sample size), $f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)$ is the conditional density of $\mathbf{x}_k$ given latent class a, g_a is the relative class size of latent class a (marginal probability of latent class a), t is the number of latent classes, and $\boldsymbol{\theta}$ is a vector of unknown parameters of the form

$$\boldsymbol{\theta} = (g_a, \boldsymbol{\theta}_a;\ a=1, \ldots, t)$$

in which $\boldsymbol{\theta}_a$ is the vector of latent class parameters of class a. We assume that

$$f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a) = \prod_{i=1}^{n} f_{ia}(x_{ki};\ \boldsymbol{\theta}_{ia}) \tag{2}$$

(Local Independence) where $\boldsymbol{\theta}_a = (\boldsymbol{\theta}_{1a}, \ldots, \boldsymbol{\theta}_{na})$, and further that $f_{ia}(x_{ki};\ \boldsymbol{\theta}_{ia})$ is normal with mean m_{ia} and variance s_{ia}^2 so that $\boldsymbol{\theta}_{ia} = (m_{ia},\ s_{ia}^2)$. Given N sets of observations $\mathbf{x}_k$ the likelihood function L can be stated as

$$L = \prod_{k=1}^{N} f(\mathbf{x}_k;\ \boldsymbol{\theta}) \tag{3}$$

where

$$f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a) = \exp\left\{-\sum_{i=1}^{n} (x_{ki} - m_{ia})^2 / 2s_{ia}^2\right\} \Big/ \left\{\prod_{i=1}^{n} (\sqrt{2\pi} s_{ia})\right\}. \tag{2$'$}$$

Since we should have

$$\sum_{a=1}^{t} g_a = 1, \tag{4}$$

we are to maximize the log of L under the restriction (4) with respect to latent parameters. Consider the quantity

$$Q = \ln L - \lambda\left(\sum_{a=1}^{t} g_a - 1\right) \tag{5}$$

where

$$\ln L = \sum_{k=1}^{N} \ln f(\mathbf{x}_k;\ \boldsymbol{\theta}) \tag{6}$$

and λ is a Lagrangean multiplier. We are to solve the zeros of the first derivatives of Q with respect to unknown parameters. We have

$$\partial Q / \partial g_a = \sum_{k=1}^{N} \{f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a) / f(\mathbf{x}_k;\ \boldsymbol{\theta})\} - \lambda. \tag{7}$$

Hence setting $\partial Q / \partial g_a = 0$ gives

$$\lambda = \sum_{k=1}^{N} \{f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a) / f(\mathbf{x}_k;\ \boldsymbol{\theta}\}. \tag{8}$$

(Here and elsewhere we omit all $\wedge$ symbols for the estimates in order to avoid awkward notations.)

Multiplying both sides of (8) by g_a we obtain

$$\begin{aligned} g_a \lambda &= \sum_{k=1}^{N} \{f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a) g_a / f(\mathbf{x}_k;\ \boldsymbol{\theta})\} \\ &= \sum_{k=1}^{N} \{f_a(\mathbf{x}_k, a;\ \boldsymbol{\theta}_a) / f(\mathbf{x}_k;\ \boldsymbol{\theta})\}, \\ & a = 1, \ldots, t, \end{aligned} \tag{9}$$

where $f_a(\mathbf{x}_k, a;\ \boldsymbol{\theta}_a)$ is the joint probability density of manifest response pattern $\mathbf{x}_k$ and latent class a. If we sum (9) over the latent classes ($a=1, \ldots, t$) we have

$$\lambda = \lambda \sum_{a=1}^{t} g_a = \sum_{k=1}^{N} \left\{\sum_{a=1}^{t} [f_a(\mathbf{x}_k, a;\ \boldsymbol{\theta}_a) / f(\mathbf{x}_k;\ \boldsymbol{\theta})]\right\} = N \tag{10}$$

by noting that

$$\sum_{a=1}^{t} f_a(\mathbf{x}_k, a;\ \boldsymbol{\theta}_a) = f(\mathbf{x}_k;\ \boldsymbol{\theta}).$$

Hence by (9) and (10) we obtain

$$g_a = \sum_{k=1}^{N} \{f_a(\mathbf{x}_k, a;\ \boldsymbol{\theta}_a)/f(\mathbf{x}_k;\ \boldsymbol{\theta})\}/N. \quad (11)$$

We next differentiate Q with respect to m_{ia} and set the derivative to zero.

$$\partial Q/\partial m_{ia} = g_a \sum_{k=1}^{N} (1/f(\mathbf{x}_k, \boldsymbol{\theta}))\cdot (\partial f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/\partial m_{ia}) = 0. \quad (12)$$

Assuming $g_a \neq 0$ we have

$$\sum_{k=1}^{N} (1/f(\mathbf{x}_k;\ \boldsymbol{\theta}))\cdot (\partial f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/\partial m_{ia}) = 0 \quad (12)'$$

where

$$\partial f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/\partial m_{ia} = f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)\cdot (x_{ki} - m_{ia})/s_{ia}^2. \quad (13)$$

Thus, again assuming that s_{ia}^2 is finite we have

$$\sum_{k=1}^{N} \{f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/f(\mathbf{x}_k;\ \boldsymbol{\theta})\}\cdot (x_{ki} - m_{ia}) = 0. \quad (14)$$

Solving (14) for m_{ia} we obtain

$$m_{ia} = \sum_{k=1}^{N} \{f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/f(\mathbf{x}_k;\ \boldsymbol{\theta})\}\ x_{ki}. \quad (15)$$

Differentiating Q with respect to s_{ia}^2 we have

$$\partial Q/\partial s_{ia}^2 = g_a \sum_{k=1}^{N} (1/f(\mathbf{x}_k;\ \boldsymbol{\theta}))\cdot (\partial f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/\partial s_{ia}^2) = 0 \quad (16)$$

where

$$\partial f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/\partial s_{ia}^2 = f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)\cdot \{(x_{ki} - m_{ia})^2/s_{ia}^2 - 1\}/2s_{ia}^2. \quad (17)$$

Substituting (17) into (16) we obtain

$$\sum_{k=1}^{N} \{f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/f(\mathbf{x}_k;\ \boldsymbol{\theta})\}\cdot \{(x_{ki} - m_{ia})^2/s_{ia}^2 - 1\} = 0. \quad (18)$$

Solving this equation for s_{ia}^2, we have

$$s_{ia}^2 = \sum_{k=1}^{N} \{f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/f(\mathbf{x}_k;\ \boldsymbol{\theta})\}\cdot (x_{ki} - m_{ia})^2. \quad (19)$$

Since the conditional probability of latent class given manifest response pattern $\mathbf{x}_k$, $f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/f(\mathbf{x}_k;\ \boldsymbol{\theta}) = f_a(a|\ \mathbf{x}_k;\ \boldsymbol{\theta})$, is a function of latent parameters, (11), (15) or (19) do not give completely explicit solutions of g_a, m_{ia} or s_{ia}^2. However, they can be used as updating equations for the estimates of latent parameters in the iterative scheme, together with Eqs. (1) and (2)′.

Similar MLE procedures have been developed for the latent class model with dichotomous manifest variates by Goodman (1974) and Henry (1975), and a minimum chi-square method by de Leeuw (1973) which gives BAN estimates as does a maximum likelihood method.

Although the convergence characteristic of the iterative scheme presented (which, in numerical analysis term, may be appropriately called nonlinear block Gauss-Seidel method), has not been established theoretically, in all cases we have tested it consistently maximizes the likelihood function (3) (i.e., the value of the likelihood function never goes down), provided a good initial start is available at the time of initiation of the iterations. Unfortunately sensible initial estimates of parameters are important for our iterative procedure to work reliably. We use Mooijaart's (1973) procedure as an initialization procedure to our algorithm. His procedure is an extention of Lazarsfeld-Anderson (Lazarsfeld & Henry, 1968; Anderson, 1954) type procedures which, unlike Green's (1951) procedure, avoids ad hoc estimations of unobservable quantities by splitting manifest variates into non-overlapping subsets, and by considering only the moments of variables belonging to different subsets. The difficulty with these procedures, however, is that they use only partial information pertaining to the data, and consequently at an unnegligible rate, result in unfeasible solutions. It is reported that Mooijaart's procedure, by the use of Moore-Penrose inverse, substantially reduces the probability of improper solutions. Nonetheless, Mooijaart estimators are only CAN, giving a justification to develop an MLE procedure in the present study.

Statistical Hypothesis Testing

A Goodness of Fit Test

A reasonable test statistic for the goodness of fit of the model can be derived based on Neyman-Pearson's general method of a likelihood ratio. Define

$$\lambda = L(\omega_1)/L(\omega_2)$$

where $L(\omega)$ is the likelihood of the model ω whose parameters are estimated by an MLE method and where the model ω_1 is subsumed under the model ω_2. It has been shown (Wilks, 1962) that

$$\chi^2 = -2 \ln \lambda = -2\{\ln L(\omega_1) - \ln L(\omega_2)\} \tag{20}$$

is distributed asymptotically chi-square with degrees of freedom being the difference in degrees of freedom between the two models (The degrees of freedom of a model is the number of observations ($N \times n$) minus the number of independently estimated parameters), and provides a consistent and the most powerful test of the goodness of fit among consistent tests of the same hypothesis. We have

$$\lim_{q\to\infty} \Pr\,(\chi^2 \geqslant \chi_q{}^2(\alpha) | X \epsilon\, \omega_1) = \alpha\,.$$

Two most important applications of the above general result are a test of the number of latent classes and tests of specific hypotheses on parameters to which we now turn.

A Test of the Number of Latent Classes

For the determination of the number of significant latent classes we can rely on essentially the same idea as discussed above. In this case we compare the likelihood of a model with t latent classes against that of the model with $t+1$ classes. Again the negative two times the log likelihood ratio is distributed asymptotically chi-square, this time with $2n+1$ *df*. If it is significant the added class has significantly improved the goodness of fit and the number of latent classes is at least $t+1$. We continue the comparison until no additional class contributes to the goodness of fit significantly.

Tests of Specific Hypotheses on Parameters

At the final stages of a research we may have more or less specific hypotheses about the parameters being estimated. Note that statistical hypotheses are restrictions on a model. In the next section we discuss ways of incorporating such constraints into the iterative scheme presented earlier. A statistical test of specific hypotheses consists, again, of the comparison of the likelihoods of two models, one with the restrictions and the other without the restrictions. Then the χ^2 defined in (20) is asymptotically chi-square with *df* equal to the difference between the degrees of freedom of the two models, which is the reduction in the number of free parameters as a consequence of imposed constraints.

Other Ingredients of the MLE Procedure

Constraints

We discuss ways to incorporate two types of constraints currently availabe in the program, MAXLPM, which has been written along the line presented in this paper. One is fixed value constraints and the other equality constraints.

The fixed value constraints prescribe certain parameters to fixed values. It is easy enough to implement this type of restrictions in the present iterative scheme, except perhaps the restrictions on class sizes which are already constrained to add up to unity so that the additional constraints may interact with the previous one. However, it can be shown that the fixed constraints on class sizes can be resolved in the same way as those on conditional expectations and variances. In any case we just skip the updating phase of the parameters whose values are fixed to constants throughout iterations.

Equality constraints are those which specify subsets of parameters whose estimates supposedly take equal but non-prespecified values. Currently equality constraints can be imposed within the same types of latent parameters. That is, equality constraints such as $m_{ia}=s_{ia}^2=g_{a'}$ are not permissible.

For illustrative purposes, assume $m_{ia}=m_{i'a'}$. Then we have the derivative of Q' (Q defined in (5) with $m_{i'a'}$ replaced by m_{ia}) with respect to m_{ia}:

$$\partial Q'/\partial m_{ia} = \sum_{k=1}^{N}(1/f(\mathbf{x}_k;\ \boldsymbol{\theta}_a))\cdot [g_a(\partial f_a(\mathbf{x}_k;\ \boldsymbol{\theta}_a)/\partial m_{ia}) + g_{a'}(\partial f_{a'}(\mathbf{x}_k;\ \boldsymbol{\theta}_{a'})/\partial m_{ia})]\,.$$

Hence,

$$m_{ia}^{(c)} = \left[(g_a/s_{ia}^2)\sum_{k=1}^{N}x_{ki}f_a(\mathbf{x}_k;\ \boldsymbol{\theta})/f(\mathbf{x}_k;\ \boldsymbol{\theta})+(g_{a'}/s_{i'a'}^2)\cdot\sum_{k=1}^{N}x_{ki'}f_{a'}(\mathbf{x}_k;\ \boldsymbol{\theta}_{a'})/f(\mathbf{x}_k;\ \boldsymbol{\theta})\right]/N(g_a/s_{ia}^2+g_{a'}/s_{i'a'}^2) = (m_{ia}^{(u)}g_a/s_{ia}^2+m_{i'a'}^{(u)}g_{a'}/s_{i'a'}^2)/(g_a/s_{ia}^2+g_{a'}/s_{i'a'}^2) \tag{21}$$

where $m_{ia}^{(c)}$ and $m_{ia}^{(u)}$ are constrained and unconstrained estimates of m_{ia}. Similarly we have, for constraints, $s_{ia}^2=s_{i'a'}^2$ and $g_a=g_{a'}$,

$$s_{ia}^{(c)2} = (g_a s_{ia}^{(u)2}+g_{a'}s_{i'a'}^{(u)2})/(g_a+g_{a'}) \tag{22}$$

where $s_{ia}^{(u)2}$ and $s_{i'a'}^{(u)2}$ are constrained and unconstrained estimates of s_{ia}^2, and

$$g_a^{(c)} = (g_a^{(u)}+g_{a'}^{(u)})/2\,. \tag{23}$$

You will note that in any case constrained estimates are weighted or unweighted averages of corresponding unconstrained estimates.

Among other types of constraints which are interesting in terms of empirical relevance, but yet unavailable in the current version of MAXLPM is

$$s_{ia} = c_{ia}m_{ia}$$

for some constant c_{ia}, where we may further constrain

$$c_{ia} = c \text{ for all } i \text{ and } a,$$
$$c_{ia} = c_i \text{ for all } a,$$
$$c_{ia} = c_a \text{ for all } i.$$

This is essentially equivalent to Weber's law in psychophysics. It is our ultimate hope, however, that the general linear equality and inequality constraints of the form $\mathbf{K}\boldsymbol{\theta}\geqslant\mathbf{c}$ be incorporated in the estimation procedure.

Some Acceleration Techniques

The MLE procedure described in this paper may sometimes be slow in convergence, particularly when the numbers of observation units and of manifest variates are very large. Here we suggest possible acceleration techniques for more rapid convergence.

A particularly relevant technique for the type of fixed point iteration method as the one employed in MAXLPM, the introduction of a relaxation parameter may significantly cut down the number of iterations required of convergence (Ortega & Rheinbolt, 1970; Ramsay, 1975). For some relaxation parameter ω the updating equation can be written

$$\boldsymbol{\theta}^{(k+1)} = w^{(k)}\boldsymbol{\theta}^{(k)}+(1-w^{(k)})G(\boldsymbol{\theta}^{(k)})$$

where $G(\boldsymbol{\theta}^{(k)})$ is the updating equation of $\boldsymbol{\theta}$ before the relaxation parameter is introduced, and the parenthesized subscripts on $\boldsymbol{\theta}$ and w indicate iteration numbers. The w must be strictly less than 1. If G is a linear function, an optimal relaxation parameter can be analytically solvable. However in nonlinear case only approximate updating scheme for adjustable $w^{(k)}$ has been given. The effectiveness of the relaxation parameter, however, has been confirmed elsewhere by the present author in the alternating least squares iterations for estimating an additive constant in metric multidimensional scaling.

Another possibility is to use the higher order information. Newton-Raphson method at least reduces the number of iterations needed for convergence, though it takes more time per iteration since extra computation is required to evaluate the second order derivatives. However, a fewer iterations may well make up for the additional time required per iteration and still leave some surplus, since the caluculation of $f_a(a|\mathbf{x}_k;\ \boldsymbol{\theta})$ and taking various means with $f_a(a|\mathbf{x}_k;\ \boldsymbol{\theta})$ as weights are the main computational load. It is suggested that Newton-Raphson method be applied separately for the estimations of class sizes, conditional expectations and variances in order to avoid the large order Hessian matrix which is to be inverted. An important by-product of using the second order method is that the negative of the inverse of the Hessian evaluated at the convergence point is proportional to the variance-covariance estimates of the estimated parameters.

Although MAXLPM does not use the Newton-Raphson method, we give the expressions of the second order derivatives which can be evaluated at the optimal point, then inverted and finally multiplied by χ^2/df to obtain the variance-covariance estimates.

We give only the general expressions.

$$\partial \ln L/\partial \theta_s = \sum_{k=1}^{N} (\partial \ln f(\mathbf{x}_k;\ \boldsymbol{\theta}/\partial \theta_s) \quad (24)$$

$$\partial^2 \ln L/\partial \theta_s \partial \theta_t = \sum_{k=1}^{N} (\partial^2 \ln f(\mathbf{x}_k;\ \boldsymbol{\theta})/\partial \theta_s \partial \theta_t) \quad (25)$$

$$\partial \ln f(\mathbf{x}_k;\ \boldsymbol{\theta})/\partial \theta_s = (1/f(\mathbf{x}_k;\ \boldsymbol{\theta}))\cdot (\partial f(\mathbf{x}_k;\ \boldsymbol{\theta})/\partial \theta_s) \quad (26)$$

$$\partial^2 \ln f(\mathbf{x}_k;\ \boldsymbol{\theta})/\partial \theta_s \partial \theta_u = (1/f(\mathbf{x}_k;\ \boldsymbol{\theta}))\cdot (\partial^2 f(\mathbf{x}_k;\ \boldsymbol{\theta})/\partial \theta_s \partial \theta_u) - (1/f(\mathbf{x}_k;\ \boldsymbol{\theta}))^2 \cdot (\partial f(\mathbf{x}_k;\ \boldsymbol{\theta})/\partial \theta_s)(\partial f(\mathbf{x}_k;\ \boldsymbol{\theta})/\partial \theta_u) \quad (27)$$

For the specific results we refer to Takane (1976).

Results and Discussion

In this section some empirical evidence as to the feasibility of the procedure will be presented. The first example is artificial in which an arbitrary set of 'true' latent parameters are hypothesized and observations are generated according to the structural model of the hypothesized latent classes. Assumed latent parameters are given in Column (A) of Table 1. This case is error free in terms of the structural model (i.e., no error was added when the data were generated); there is, of course, some sampling error, since the sample size is finite ($N=100$). With the artificial data we have prior knowledge as to which latent class each observation unit is recruited from, so that we can calculate estimates of latent parameters based on the class assignment of the observations (which, of course, is unknown in practice). The estimates are given in Column (B) of Table 1 along with the standard errors of the estimates. The corresponding MAXLPM results are listed in the third column of Table 1 in the same format as in the second column. The convergence is obtained in 13 iterations for the stopping criterion being that the relative improvement in the log likelihood be less than 1.0×10^{-6} starting from the initial estimates (Column (D)) obtained by Mooijaart's procedure.

The next result is with the same data, but the solutions are obtained under several external constraints; m_{11} and m_{22} are fixed to .9, m_{41} and m_{42} are set equal, and within-class dispersions are also set equal across latent classes. The estimates under these restrictions are given in Column (E) of Table 1. The difference in the fit between the constrained and unconstrained cases is chi-square 44.782 with 8 degrees of freedom, which is significant. Thus, the constrained latent structure is not

TABLE 1

Hypothesized latent parameters and their estimates in the first example

	Size									
Class I	(A) .600		(B) .590 (.049)		(C) .649 (.085)		(D) .535		(E) .658	
	mean	vari.	*mean*	vari.	*mean*	vari.	*mean*	vari.	*mean*	vari.
Manifest variates										
1	.900	.090	.896	.082	.878	.075	.910	.112	.900#	.108
			(.038)	(.015)	(.063)	(.014)				
2	.200	.160	.151	.173	.212	.208	.440	.265	.220	.163
			(.055)	(.032)	(.061)	(.039)				
3	.800	.160	.807	.135	.845	.131	.626	.261	.825	.171
			(.049)	(.025)	(.048)	(.025)				
4	.400	.240	.442	.186	.437	.176	.370	.207	.592#	.207
			(.057)	(.035)	(.054)	(.032)				
5	.900	.090	.936	.085	.938	.061	1.070	.021	.952	.089
			(.039)	(.016)	(.033)	(.012)				
Class II	.400		.410 (.049)		.351 (.085)		.465		.342	
Manifest variates										
1	.700	.210	.743	.139	.749	.169	.744	.112	.737	.108
			(.059)	(.031)	(.073)	(.042)				
2	.900	.090	.877	.087	.888	.073	.459	.265	.900#	.163
			(.039)	(.019)	(.050)	(.020)				
3	.400	.240	.302	.207	.168	.206	.586	.261	.189	.171
			(.087)	(.067)	(.082)	(.052)				
4	.800	.160	.808	.158	.879	.138	.847	.207	.592#	.207
			(.063)	(.035)	(.069)	(.035)				
5	.500	.250	.390	.188	.294	.185	.316	.021	.250	.089
			(.068)	(.042)	(.080)	(.050)				

(A) Hypothesized latent parameters
(B) Estimated latent class parameters based on the prior knowledge of class assignment
(C) Unconstrained estimates of latent parameters from MAXLPM
(D) Initial estimates of latent parameters by Mooijaart's procedure
(E) Constrained estimates of latent parameters from MAXLPM
Standard errors of estimates in parantheses. The # indicates that the parameters are either fixed or equated (All within-class variances are equated across classes, but not indicated in the table).

acceptable against the unconstrained structure, and the specific hypothesis implied in the constraints should be rejected on the statistical ground. The convergence is obtained in six iterations starting from the unconstrained estimates.

The next example is also hypothetical and has been adapted from Example III from Gibson (1959). He illustrates how a ridiculous interpretation may be led from the factor analysis result of these data. Hypothesized structure is given in Fig. 1. For illustrative purposes let us suppose that the manifest variates in this example measure some kind of ability of subjects. Suppose further that they all measure the same ability and that higher scores mean more able. We have three latent classes, one with low ability (Class I), another with middle ability (Class II) and a third with high ability (Class III). Manifest variates 1 and 2 discriminate

TABLE 2

Estimates of latent parameters in the second example: Three-class solution

Manifest variates	Latent class 1		2		3	
	mean	vari.	*mean*	vari.	*mean*	vari.
1	−3.163	1.029	1.027	1.003	0.997	0.923
2	−2.946	0.962	0.983	1.008	1.169	0.973
3	−2.871	1.171	−0.066	1.085	3.016	1.017
4	−2.472	1.206	−0.030	1.068	2.467	0.930
5	−1.928	0.892	−0.075	1.000	1.876	0.942
6	−1.471	0.956	0.200	1.010	1.389	1.150
7	−0.961	1.058	0.017	0.097	0.715	1.070
8	−0.791	0.958	−0.907	0.946	2.917	0.913
9	−0.904	1.000	−0.926	0.958	2.837	1.023
Class size	.250		.530		.220	

Class I from the other two classes, but does not differentiate between Class II and Class III, whereas manifest variates 8 and 9 discriminate Class III from other classes but does not differentiate between Class I and Class II. Other variates have equal discriminating power over the entire range of population, there are distinct differences in the over all discriminating power among variates (high in 3 and low in 7). Within-class variances are assumed to be unity for all variates. Note that variates 1, 2, 8 and 9 are nonlinearly related to latent classes in reference to other variates (there is no necessity for the classes to be equally spaced, but even if we can space them off so that these variates are linearly related to latent classes, then other variates will no longer be related linearly to the classes).

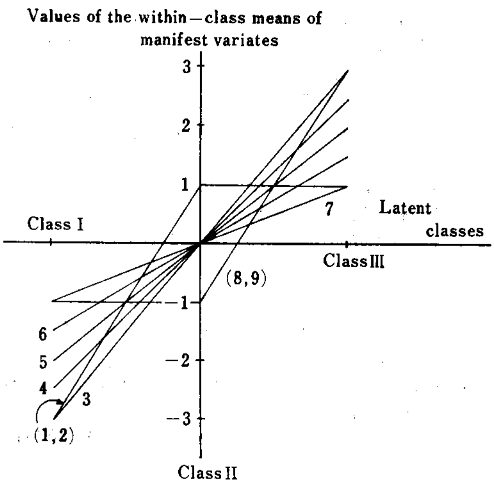

FIG. 1. Hypothesized 'true' structure in the second example.

Class sizes are .25, .50 and .25, respectively.

A typical factor analysis of these data obtains two relatively independent factors which represent high and low portions of the same ability, which intuitively violates the assumption that the manifest variates measure one ability. The enforcement of linearity leads to a conclusion that high and low portions of an ability are two different abilities.

A set of observations are generated according to the hypothesized structure and then submitted to MAXLPM. When the data are generated a small random error (about 1/10 of the within-class standard deviation in magnitude) is added to each observation. The three class solution from MAXLPM is given in Table 2. The result clearly indicates the adequacy of the latent profile model for the type of hidden underlying structure hypothesized as well as the feasiblity of the estimation procedure in the presence of random error perturbations. The same set of data has been analyzed under the different hypothesis about the number of latent classes. The two-class solution indicates that the difference in the chi-square statistic from the three-class solution is 297.250 with 19 *df* which is significant at 1% level. Thus, the addition of the third class has improved the goodness of fit significantly and the number of latent classes is at least three. An attempt has been made to make the same type of comparison between three- and four-class solutions. However, Mooijaart's procedure has failed to obtain any feasible initial estimates. Apparently four latent classes are an overestimation.

In this paper it has been demonstrated, through these admittedly small Monte Carlo experiments, that our estimation procedure works reasonably well even in the presence of structural errors of certain degrees as well as in the presence of sampling errors. Associated goodness of fit under the various restrictions provide useful information pertaining to the covariation structures among manifest variates, particularly when the covariation is nonlinear.

References

ANDERSON, T. W. 1954 On estimation of parameters in latent structure analysis. *Psychometrika*, **19**, 1–10.

ANDERSON, T. W. 1959 Some scaling models and estimation procedures in the latent class model. In U. Grenander (Ed.), *Probability and statistics*. New York: Wiley.

GIBSON, W. A. 1959 Three multivariate models: Factor analysis, latent structure analysis and latent profile analysis. *Psychometrika*, **24**, 229–252.

GOODMAN, L. A. 1974 Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, **61**, 215–231.

GREEN, B. F. 1951 A general solution for the latent class model of latent structure analysis. *Psychometrika*, **16**, 151–160.

HENRY, N. W. 1975 Maximum likelihood estimation of parameters of latent class models. *Gary Income Maintenance Experiment*, Purdue University, Lafayette, Indiana.

JÖRESKOG, K. G. 1969 A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, **34**, 183–202.

LAZARSFELD, P. F., & HENRY, N. W. 1968 *Latent structure analysis*. Boston: Houghton Mifflin.

DE LEEUW, J. 1973 Estimation in latent class analysis. *Research Bulletin*. RB 001–73, Dept. of Data Theory, University of Leiden, Leiden, The Netherlands.

MCDONALD, R. P. 1962 A note on the derivation of the general latent class model. *Psychometrika*, **27**, 203–206.

MCDONALD, R. P. 1965 Difficulty factors and nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology*, **18**, 11–24.

MCHUGH, R. B. 1956 Efficient estimation and local identification in latent class analysis. *Psychometrika*, **21**, 331–347.

MOOIJAART, A. 1973 A contribution to latent class analysis. *Research Bulletin*. RB-MT-001–73, Dept. of Psychology, University of Leiden, Leiden, The Netherlands.

NEYMAN, J. 1949 Contribution to the theory of χ^2-test. *Proceedings of Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.

ORTEGA, J. M., & RHEINBOLT, W. C. 1970 *Iterative solution of nonlinear equations in several variables*. New York: Academic Press.

RAMSAY, J. O. 1975 Solving implicit equations in psychometric data analysis. *Psychometrika*, **40**, 337–360.

TAKANE, Y. 1972 Estimation of the recruitment latent class by least squares methods. *Japanese Psychological Research*, **14**, 87–102.

TAKANE, Y. 1976 A maximum likelihood procedure for the latent profile model: Some second order derivatives. Unpublished manuscript, University of Tokyo.

WILKS, S. S. 1962 *Mathematical statistics*. New York: Wiley.

(Received Feb. 25, 1976)