A TEST OF LOCAL INDEPENDENCE IN LATENT STRUCTURE ANALYSIS

H. I. Patel, University of Wisconsin-Milwaukee

1. Introduction

In many areas, especially in the social and behavioral sciences, we are interested in assessing some trait or characteristic of individuals which cannot be observed directly. Suppose each individual belongs, unknown to us, to one of several classes. Then since these classes are unobservable, they are called latent classes. In this situation we usually observe the response items which provide the information pertinent to the latent classes. The latent classes are characterized by a set of latent parameters which can be estimated by solving a series of accounting equations linking the latent parameters and the empirically observed response frequencies. The main assumption to be made for solving this system of accounting equations is the axiom of local independence which establishes the independence of the responses of different items within a given latent class. In other words the dependence among the response-items is explained by the latent variable.

The original work of latent structure analysis is due to Lazarsfeld [1950, 1954]. Various techniques for estimating the latent parameters have been proposed by Lazarsfeld and Dudman [1951], Green [1951], Gibson [1955, 1959], McHugh [1954, 1956], Anderson [1954, 1959], Madansky [1958, 1960] and others. The relation of latent structure analysis to factor analysis has been studied by Bargmann [1957], Gibson [1951] and Green [1952]. An excellent treatment of latent structure analysis can be found in Lazarsfeld and Henry [1968].

Under the assumption of local independence the maximum likelihood estimates of the latent parameters are obtained using an iterative procedure and a test of goodness fit is proposed.

2. Formulation of the Model

We define a set of dichotomous variables $X_1, X_2, \ldots, X_p$ to be observed on a randomly selected individual by $P(X_i=1) = p_i$ and $P(X_i=0) = 1-p_i$, $i = 1,2,\ldots,p$. Then the vector $\underset{\sim}{X}$ has a point multivariate binomial distribution whose f.m.g.f. is given by (Krishnamoorthy [1947])

$$(1)\quad \varphi(t_1,\ldots,t_n) = (1 + \Sigma\, p_i t_i + \sum_{ij} p_{ij} t_i t_j + \ldots + p_{12\ldots p} t_1 t_2 \ldots t_p),$$

where

$$p_{ij} = P(X_i=1, X_j=1)$$
$$\ldots$$
$$p_{12\ldots p} = P(X_1=1, X_2=1, \ldots, X_p=1).$$

We now consider that each individual in this population can be classified into one of the m+1 latent classes B^α $(\alpha = 0,1,\ldots,m)$, so that $P(B^\alpha) = \pi_\alpha$ and $\pi_0+\pi_1+\ldots+\pi_m=1$. Let H be the hypothesis of local independence which establishes independence of $X_1, X_2, \ldots, X_p$ within a latent class B^α (see Lazarsfeld and Henry [1968]). The hypothesis H can be stated as

$$(2)\quad P(X_{i_1}=1, X_{i_2}=1, \ldots, X_{i_k}=1 \mid B^\alpha) = p^\alpha_{i_1} p^\alpha_{i_2} \ldots p^\alpha_{i_k},\ 1\le i_1<i_2<\ldots<i_k\le p,\ k=2,\ldots,p,$$

where $P(A\mid B^\alpha)$ denotes the conditional probability of the event A given that an individual belongs to class B^α and $p^\alpha_j = P(X_j=1\mid B^\alpha)$. Hence the characteristic function of $\underset{\sim}{X} = (X_1,\ldots,X_p)$, under H, is

$$(3)\quad \phi(t_1,\ldots,t_p) = \sum_{\alpha=0}^{m} \pi_\alpha \prod_{j=1}^{p} (1-p^\alpha_j + p^\alpha_j e^{it_j}),$$

and by the inversion theorem it can be shown that the probability function (p.f.) of $\underset{\sim}{X}$ is

$$(4)\quad f(x_1,\ldots,x_p) = \sum_{\alpha=0}^{m} \pi_\alpha \prod_{i=1}^{p} (p^\alpha_i)^{x_i} \times (1-p^\alpha_i)^{1-x_i},\quad x_i=0,1\ (i=1,2,\ldots,p).$$

The p.f. in (4) may be considered as a mixture of $f^\alpha(x_1,\ldots,x_p)$ for $\alpha=0,1,\ldots,m$. where

$$f^\alpha(x_1,\ldots,x_p) = \prod_{i=1}^{p} (p^\alpha_i)^{x_i} (1-p^\alpha_i)^{1-x_i}.$$

The alternative hypothesis H_a is that the dependence among $X_1, X_2, \ldots, X_p$ cannot be explained by the latent variable, i.e., the p.f. of $\underset{\sim}{X}$ cannot be written as in (4).

3. Maximum Likelihood Estimates under H

If $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_N$ is a random sample from the population with p.f. (4), the log of the likelihood function, under H, is given by

$$(5)\quad \log L = \sum_{t=1}^{N} \log \sum_{\alpha=0}^{m} \pi_\alpha W_{\alpha t},$$

where

$$W_{\alpha t} = \prod_{i=1}^{p} (p^\alpha_i)^{x_{it}} (1-p^\alpha_i)^{1-x_{it}},$$
$$\alpha = 0,1,\ldots,m;\ t=1,2,\ldots,N.$$

Setting the first partial derivatives of log L with respect to the parameters equal to zero, we get

(6) $\sum_{t=1}^{N} W_{\alpha t}/g_t = \sum_{t=1}^{N} W_{ot}/g_t, \ \alpha=1,2,\ldots,m,$

(7) $p_i^\alpha = \left(\sum_{t=1}^{N} W_{\alpha t}x_{it}/g_t\right)/\sum_{t=1}^{N}(W_{\alpha t}/g_t),$

where $g(t)=\pi_o W_{ot}+\pi_1 W_{1t}+\ldots+\pi_m W_{mt}$. Multiplying both sides of equation (6) by π_α, and summing over α, we get

$$\sum_{\alpha=0}^{m}\sum_{t=1}^{N}\pi_\alpha W_{\alpha t}/g_t = \sum_{\alpha=0}^{m}\sum_{t=1}^{N}\pi_\alpha W_{0t}/g_t$$

or

(8) $\pi_\alpha = (\pi_\alpha/N)\sum_{t=1}^{N} W_{\alpha t}/g_t, \ \alpha=1,2,\ldots,m.$

If we have a set of initial estimates of $\pi_\alpha(\alpha=1,2,\ldots,m)$ and p_i^α $(\alpha=0,1,\ldots,m;i=1,2,\ldots,p)$, the equations (7) and (8) can be used as the basis for iteration scheme. A method for obtaining initial estimates is given in the next section.

It should be observed that the distribution given by the p.f. W_α is a member of the exponential family and hence the population in (4) is a mixture of m+1 sub-populations each being a member of the exponential family. Hasselblad [1969] has given an algorithm for obtaining the maximum likelihood estimates in finite mixtures of distributions. Based on over three hundred trials with several mixtures of distributions, he concludes that (i) the likelihood function increases monotonically, ignoring the oscillation, (ii) the procedure attains a relative maximum of the likelihood function, and (iii) in general, different sets of initial estimates give the same final estimates to a certain accuracy. The same convergence properties will be possessed by our algorithm.

4. Initial Estimates

Various methods are available for estimating the latent parameters. Such estimates can be used as initial estimates in ite--tive procedure of maximum likelihood -nod. In this section we propose a method based on a factor analysis solution. The estimates obtained by this method are not unique and perhaps are less efficient than the present methods available. However, they can be used as initial estimates in the iterative procedure given in the previous section.

We have

(9)
$$\sigma_i^2 = \mathrm{Var}\,(X_i) = p_i(1-p_i), i=1,2,\ldots,p$$
$$\sigma_{ij} = \mathrm{Cov}\,(X_i,X_j) = p_{ij} - p_i p_j,$$
$$i\neq j=1,2,\ldots,p.$$

Further, $p_i=\pi_o p_i^o+\pi_1 p_i^1+\ldots+\pi_m p_i^m$.

The axiom of local independence implies the pairwise independence. Writing

$$p_{ij}^\alpha = P(X_i=1,X_j=1 \mid B^\alpha),$$

we get

(10) $p_{ij}^\alpha = p_i^\alpha p_j^\alpha, \ \alpha=0,1,\ldots,m,$

$$i\neq j=1,2,\ldots,p.$$

Hence $p_{ij} = \sum_{\alpha=0}^{m} \pi_\alpha p_i^\alpha p_j^\alpha$.

With these notations, under the assumption of local independence, we get the covariance matrix of $\underset{\sim}{X}$ as

(11) $\Sigma = D_u+YVY'$,

where D_u is a diagonal matrix with the ith diagonal element equal to

$$\sum_{\alpha=0}^{m}\pi_\alpha p_i^\alpha(1-p_i^\alpha),$$

$$Y = \begin{bmatrix} p_1^o-p_1^1 & p_1^o-p_1^2 & \cdots & p_1^o-p_1^m \\ p_2^o-p_2^1 & p_2^o-p_2^2 & \cdots & p_2^o-p_2^m \\ \cdot & \cdot & \cdots & \cdot \\ p_p^o-p_p^1 & p_p^o-p_p^2 & \cdots & p_p^o-p_p^m \end{bmatrix}$$

and the elements of matrix V are

$$v_{ii} = \pi_i(1-\pi_i), \ i=1,2,\ldots,m$$
$$v_{ij} = -\pi_i\pi_j, \quad i\neq j=1,2,\ldots,m.$$

Since V is a symmetric and positive definite matrix, it can be written as TT', where T is a triangular matrix with real elements. Thus the correlation matrix, P, of $\underset{\sim}{X}$ can be written as

(12) $P = D_w + \Phi\Phi'$,

where $D_w = D_\sigma^{-1}D_u D_\sigma^{-1}$, $\Phi = D_\sigma^{-1}YT$ and D_σ is a diagonal matrix with elements $\sigma_1,\sigma_2,\ldots,\sigma_p$.

Thus the correlation matrix of $\underset{\sim}{X}$ has the structure of factor analysis which we arrived at through the assumption of local independence. The matrix Σ should be estimated by the sample covariance matrix of $\underset{\sim}{X}$. Let F be any admissible factor analysis solution for Φ, such that $-1 \leq f_{i\alpha} < 1$ for all elements of F and $0 \leq \sum_{\alpha=1}^{m} f_{i\alpha}^2 < 1$ for all i.

Then $F^* = FA$ satisfying the above admissible constraints is also a solution, where A is an orthogonal matrix. If some knowledge about the parameters π_α's is available from the past experience, we can estimate Y as $\hat{Y}=D_\sigma F^* T^{-1}$, such that each element of $\hat{Y}$ is less than 1 in absolute value. Thus $p_i^0 - p_i^\alpha$ can be estimated for all i and $\alpha = 1,2,\ldots,m$. Using the relation $\sum_{\alpha=0}^{m} \pi_\alpha p_i^\alpha = p_i$, we can estimate all p_i^α. p_i are to be estimated from the sample. It should be noted that all estimates p_i^α must lie on the open interval (0,1).

It is clear that when m=1 (i.e., dichotomous latent variable), the matrix Φ can be estimated uniquely, if π_0 is known and hence all p_i^α can be estimated uniquely. When $\pi_0,\pi_1,\ldots,\pi_m$ are not known, it is recommended to take equal values for all π_α's. A drawback of this method of obtaining initial estimates is that they are based on a derived solution of the factor analysis. Further, it may be noted that any factor analysis method can be used to obtain an estimate of Φ.

5. Test of Goodness of Fit

Let N be the number of individuals interviewed and $N_{j_1 j_2 \ldots j_p}$ the number of individuals with response vector $(j_1,j_2,\ldots,j_p)$, where $j_i = 0$ or 1. If $p_{j_1 j_2 \ldots j_p} = P(X_1=j_1,X_2=j_2,\ldots,X_p=j_p)$, the observed probability of the response vector $(j_1,j_2,\ldots,j_p)$ is equal to $N_{j_1 j_2 \ldots j_p}/N$. The maximum likelihood estimate of $p_{j_1 j_2 \ldots j_p}$ is given by

$$\hat{p}_{j_1 j_2 \ldots j_p} = \sum_{\alpha=0}^{m} \hat{\pi}_\alpha \prod_{i=1}^{p} (\hat{p}_i^\alpha)^{j_i}(1-\hat{p}_i^\alpha)^{1-j_i},$$

where $\hat{\pi}_\alpha$ and $\hat{p}_i^\alpha$ are the maximum likelihood estimates of π_α and p_i^α. In sample data one may not observe all of the 2^p distinct vectors $(j_1,j_2,\ldots,j_p)$. Suppose k is the number of distinct observed vectors. If $k < 2^p$, the probabilities of the observed vectors should be adjusted so as to make the total expected frequency N. Finally

$$\chi^2 = \sum_{j_1,j_2,\ldots,j_p} \left(N_{j_1 j_2 \ldots j_p} - N\hat{p}_{j_1 j_2 \ldots j_p}\right)^2 / N\hat{p}_{j_1 j_2 \ldots j_p}$$

with k-1-[(m+1)p+m] d.f. can be used as a test statistic.

Acknowledgement

This research was supported, in part, by the office of Naval Research, Contract Number N00014-69-A-0423, Department of Statistics, University of Georgia.

REFERENCES

[1] Anderson, T.W., "On estimation of parameters in latent structure analysis", Psychometrika, 19, (1954) 1-10.

[2] ____________, "Some scaling methods and estimation procedures in the latent class model", Probability and Statistics, Ulf Grenander, ed., New York: Wiley, 1959, 9-38.

[3] Bargmann, R. E., "A study of independence and dependence in multivariate normal analysis", Inst. of Stat., University of North Carolina, 1957, No. 186.

[4] Gibson, W. A., "Applications of Mathematics of multiple factor analysis to problems of latent structure analysis", Unpublished Ph.D . dissertation, University of Chicago, 1951.

[5] ____________, "An extension of Anderson's solution for the latent structure equations", Psychometrika, 20, (1955), 69-73.

[6] ____________, "Three multivariate methods: factor analysis, latent structure analysis and latent profile analysis", Psychometrika, 24, (1959), 229-52.

[7] Green, Bert F., "A general solution for the latent class model of latent structure analysis", Psychometrika, 16, (1951), 151-66.

[8] ____________, "Latent structure analysis and its relation to factor analysis", J. A. S. A., 47, (1952) 71-76.

[9] Hasselblad, Victor, "Estimation of finite mixture of distributions from the exponential family", J.A.S.A., 64, (1969), 1459-71.

[10] Krishnamoorthy, A. S., "Multivariate binomial and Poisson distributions", Sankhye, 11, (1947), 117-24.

[11] Lazarsfeld, Paul F., "The logical and mathematical foundation of latent structure analysis", Measurement and Prediction, S.A. Stouffer, et al., Princeton; Princeton University Press, 1950.

[12] ________________, "A conceptual introduction to latent structure analysis", Mathematical Thinking in Social Sciences, Paul F. Lazarsfeld ed., New York: The Free Press, 1954.

[13] ________ and Dudman, Jack, "The general solution of the latent class case", The Use of Mathematical Models in the Measurement of Attributes, Paul F. Lazarsfeld, ed. Rand Research Memorandum, 1951, No. 455.

[14] ________ and Henry N. W., Latent Structure Analysis, Houghton Mifflin Company, 1968.

[15] Madansky, A., "Identification and estimation in latent class analysis", Unpublished Ph.D. dissertation, University of Chicago, 1958.

[16] _____, "Determinantal Methods in latent class analysis" Psychometrika, 25, (1960), 183-98.

[17] McHugh, R. B., "On the scaling of psychological data by latent structure analysis", Unpublished doctoral dissertation, University of Minnesota, 1954.

[18] McHugh, R. B., "Efficient estimation and local identification in latent class analysis", Psychometrika, 21, (1956), 331-47; correction, 23, (1958), 273-74.