THE VALIDATION OF MODELS BASED ON LATENT STRUCTURES

C. Mitchell Dayton and George B. Macready, University of Maryland

A. INTRODUCTION

Many research areas share a common methodological concern with fitting pre-established dependency structures to data gathered from human subjects under conditions which introduce error into the measurement processes. Data are summarized, often, as a series of "successes" or "failures", while a theoretical model postulates some kind of sequential dependency among the tasks. The purpose of this paper is to summarize a class of probabilistic models which is useful for analyzing data purported to reflect hierarchic structures. The historical antecedents for the models discussed in this paper stem from the work of Lazarfeld and Henry (1968). Recent advances in estimation and hypothesis testing are due to Proctor (1970), Murray (1971), Goodman (1974, 1975, 1976), Dayton and Macready (1976), and Macready and Dayton (1977a). For a more complete overview of the theory underlying these models and for applications to real data sets, the above references, as well as Macready and Dayton (1977b), may be consulted.

B. THE GENERAL MODEL AND SOME SPECIAL CASES

It is assumed that all respondents (subjects) can be, in theory, identified with a set of "latent classes" which represent the levels of an a priori hierarchic structure. Furthermore, this presentation is limited to dichotomous response data; that is, we assume K distinct tasks, each of which can be scored 0,1 for a sample of n respondents (such 0,1 scoring may result from a true point variable, or from artificial dichotomization of a continuous variable). For convenience, let $\underline{u}_s$ be an observed response vector with elements 0,1 and let $\underline{v}_j$ be one of q theoretical vectors corresponding to the latent classes in the hierarchic structure. The basic concept of the general probabilistic model is that the observed vectors arise from the theoretical vectors due to response errors which obey a law of local independence. Using the notation $P(\cdot)$ for probabilities, the model is:

$$(1)\quad P(\underline{u}_s) = \sum_{j=1}^{q} P(\underline{u}_s \mid \underline{v}_j)\cdot\theta_j$$

$$(2)\quad P(\underline{u}_s \mid \underline{v}_j) = \prod_{i=1}^{K} \alpha_i^{a_{ijs}}(1-\alpha_i)^{b_{ijs}}\beta_i^{c_{ijs}}(1-\beta_i)^{d_{ijs}}$$

where the parameters are:

- θ_j - the true proportion of respondents which falls in the j^{th} latent class
- α_i - the probability of an intrusion error on task i
- β_i - the probability of an omission error on task i
- $a_{ijs}, b_{ijs}, c_{ijs}, d_{ijs}$ - numerical coefficients which are 0,1 and which relate the observed vector to the theoretical vector (methods for determining these coefficients are shown in connection with special cases of the model which are described below)

Note that $P(\underline{u}_s \mid \underline{v}_j)$ is the conditional probability that the observed vector, $\underline{u}_s$, arises from the j^{th} latent class through the occurrence of appropriate intrusion and/or omission errors. Such conditional probabilities are, in turn, generated by a product of probabilities associated with the individual tasks. The association of such (unconditional) probabilities with the tasks is equivalent to assuming a condition of local independence in the sense that a respondent's behavior is independent (without memory) across tasks.

Although the general model as presented in equations (1) and (2) can be fitted to data under certain circumstances, most of the applications which have been pursued to date have centered about simplified forms of the model. In Section C., we present some special cases so that the form of the models and the notation utilized are made clear. Two classes of models are distinguished: Extreme Groups Models and Hierarchic Models. In the cases of Extreme Groups, there are only two theoretical vectors - one corresponding to complete "failure", $\underline{v}_1 = (0\ 0\ 0\ \ldots\ 0)$, and one corresponding to complete "success", $\underline{v}_2 = (1\ 1\ 1\ \ldots\ 1)$. All other observed vectors must arise by intrusion or omission errors. Two special cases of the Extreme Groups Model are:

Case 1 - each task has an unique intrusion and omission error component (α_i and β_i);

Case 2 - all intrusion occurs at a constant rate (α) and all omission at a constant rate (β).

Within the class of Hierarchic Models, we include all linear and non-linear (e.g., convergent or divergent) hierarchies of arbitrary complexity. Four special cases of the Hierarchic Model are distinguished:

Case 1 - α_i and β_i as for the Extreme Groups Model, above;

Case 2 - separate error rates (α_i) per task, but intrusion and omission occur at this same rate for a given task (that is, Case 1 with $\alpha_i = \beta_i$);

Case 3 - intrusion (α) and omission (β) constant across tasks as in Case 2 of the Extreme Groups Model, above;

Case 4 - a single error rate for intrusion and omission across all tasks (i.e., Case 3 with $\alpha = \beta$).

It is evident that many other special cases can be defined by appropriate restrictions (or generalizations) with respect to the operation of errors. However, the cases referenced above have been studied both theoretically and practically. Ordinarily, the hierarchic structure (set of latent classes) can be specified in detail on the basis of a priori considerations (e.g., a Guttman scale implies a linear hierarchy), but the way in which error probabilities enter into the model is more open to speculation. Thus, the various cases distinguished for Hierarchic Models permit some flexibility in fitting real data sets.

C. PARAMETRIZATION OF THE MODELS

(1) Case 1 of the Extreme Groups Model - For this special case, the only *a priori* vectors are $\underline{v}_1' = (0\ 0\ 0\ \ldots\ 0)$ and $\underline{v}_2' = (1\ 1\ 1\ \ldots\ 1)$ and the probabilistic model in equations (1) and (2) simplifies considerably. Thus,

$$P(\underline{u}_s) = P(\underline{u}_s \mid \underline{v}_1)\cdot\theta_1 + P(\underline{u}_s \mid \underline{v}_2)\cdot\theta_2$$

Further, let the elements in $\underline{u}_s$ be denoted x_{is}, so that $\underline{u}_s' = (x_{1s}, x_{2s}, \ldots, x_{Ks})$; then,

$$P(\underline{u}_s \mid \underline{v}_1) = \prod_{i=1}^{K} \alpha_i^{x_{is}}(1-\alpha_i)^{1-x_{is}}$$

$$P(\underline{u}_s \mid \underline{v}_2) = \prod_{i=1}^{K} \beta_i^{1-x_{is}}(1-\beta_i)^{x_{is}}$$

The topic of estimating parametric values from real data sets is discussed in Section D., below; since, in general, the various patterns of observed score vectors may have different probabilities of occurring under the models, it is apparent that estimation must be based on the total set of 2^K observed score vectors. With the exception of Case 2 of the current model, this requirement to have data summarized for each of the 2^K possible observed score vectors holds for all of the special cases considered in this paper.

(2) Case 2 of the Extreme Groups Model - If we restrict the intrusion errors, α_i, to a constant value (α) across the K tasks and the omission errors, β_i, to a constant value (β) across the K tasks, Case 2 is obtained. For this case, the probabilistic model takes on an especially simple form since the number of errors necessary to account for the observed score vectors is a function of the total "score" $X = \sum_{i=1}^{K} x_i$ associated with such a vector. Thus,

$$P(X \mid \underline{v}_1) = ({}_K C_X)\alpha^X(1-\alpha)^{K-X}$$

$$P(X \mid \underline{v}_2) = ({}_K C_X)\beta^{K-X}(1-\beta)^X$$

where ${}_K C_X$ is the combinations operator. Note that each of these conditional probabilities is of the form of a binomial and the model becomes, in effect, the *mixture* of two binomial processes with binomial parameters α and β, and with mixture, θ_1. Since all observed score vectors which yield the same score, X, have the same probability of occurring under this model, data can be analyzed from scores alone. That is, unlike Case 1 where the 2^K patterns are needed, the data can be summarized as K+1 score frequencies. Of course, this simplification makes the model less flexible with respect to representing real data sets.

(3) Case 1 of the Hierarchic Model - This model is summarized in equations (1) and (2) without simplification. Unfortunately, for arbitrary hierarchies (including linear forms), it does not seem to be possible to obtain estimates for all of the parameters simultaneously from real data sets by conventional estimation procedures (maximum likelihood). For this reason, the model as embodied in Case 1 is non-identifiable and we must turn our attention to the restricted cases in order to arrive at practical solutions.

(4) Case 2 of the Hierarchic Model - For this case, the intrusion and omission error rates are restricted to be equal for a given task, but each task has an unique error parameter (α_i). Thus, the model in equation (2) becomes:

$$P(\underline{u}_s \mid \underline{v}_j) = \prod_{i=1}^{K} \alpha_i^{a_{ijs}}(1-\alpha_i)^{1-a_{ijs}}$$

where a_{ijs} is 0,1 and determined as follows: let the i^{th} element in $\underline{u}_s$ be x_{is} and the i^{th} element in $\underline{v}_j$ be t_{ij}. Then,

$$a_{ijs} = \begin{cases} 0 & \text{if } x_{is} - t_{ij} = 0 \\ 1 & \text{otherwise} \end{cases}$$

In effect, whenever corresponding elements in the observed and theoretical vectors fail to match, a_{ijs} is given the value 1 and this introduces the error parameter into the model for this task. Otherwise, the value $1-\alpha_i$ enters and this is the probability of *not* making an error for the ith task.

(5) Case 3 of the Hierarchic Model - In dealing with hierarchic structures, we have had the most experience in applying this case since the number of parameters which must be estimated remains reasonably small even for fairly large numbers of tasks. The notion of separate intrusion and omission error rates is retained, but these rates are assumed to be constant (or homogeneous) across tasks. For notational purposes, let α be this constant intrusion error rate and β be the constant omission error rate. Then, the model in equation (2) becomes:

$$P(\underline{u}_s \mid \underline{v}_j) = \alpha^{a_{js}}(1-\alpha)^{b_{js}}\beta^{c_{js}}(1-\beta)^{d_{js}}$$

where the coefficients a_{js}, b_{js}, c_{js}, and d_{js} are determined from the following rules based on the elements x_{is} of $\underline{u}_s$ and the elements t_{ij} of $\underline{v}_j$:

a_{js} is the number of times $t_{ij} = 0$ when $x_{is} = 1$ (number of intrusions)

b_{js} is the number of times $t_{ij} = 0$ when $x_{is} = 0$ (number of non-intrusions)

c_{js} is the number of times $t_{ij} = 1$ when $x_{is} = 0$ (number of omissions)

d_{js} is the number of times $t_{ij} = 1$ when $x_{is} = 1$ (number of non-omissions)

D. ESTIMATION OF PARAMETERS

Within certain broad limits of identifiability, parameter estimates can be obtained by means of computer-based algorithms for both types of Extreme Groups Model and for cases 2, 3, and 4 of the Hierarchic Model. The method of estimation which is employed ordinarily is that of maximum likelihood. Unfortunately, there are no simple, algebraic formulae which can be derived for these models since they are non-linear in the parameters. Nevertheless, computerized procedures can locate the maximum likelihood estimate if initial guessed values for all parameters are used and, then, iteratively improved until they converge on the appropriate values. Programs developed by us have been based on Fisher's method of "scoring" and, in general, the final solution does not depend upon good choices for initial guessed values (i.e., the algorithm is relatively insensitive to starting values). However, "boundary problems" arise with some regularity, and the programs have options which will force the final solution to take on acceptable values (i.e., all the θ_j, α_i, and β_i are restricted to the interval 0,1).[1]

For a total of n respondents, the likelihood for the sample is:

$$(3)\quad L = \prod_{s=1}^{n} P(\underline{u}_s) = \prod_{s=1}^{n} \sum_{j=1}^{q} P(\underline{u}_s / \underline{v}_j)\cdot\theta_j$$

and the general method of solution involves solving the system of partial derivatives:

$$\partial \mathrm{Log}_e L/\partial\theta_j = 0 \quad , \quad j = 1,\ldots,q-1$$

$$\partial \mathrm{Log}_e L/\partial\alpha_i = 0 \quad , \quad i = 1,\ldots, K$$

$$\partial \mathrm{Log}_e L/\partial\beta_i = 0 \quad , \quad i = 1,\ldots, K$$

with suitable restrictions placed on the α_i and β_i to provide non-singularity (identifiability) for the system. Computation of the derivatives is greatly simplified if Fisher's method of scoring (Rao, 1965) is used and solution of the system can be pursued iteratively by the method of Newton-Raphson. An important by-product of this approach is that the matrix of partial second derivatives provides a basis for estimating large-sample sampling variances of the parameter estimates (i.e., the inverse of this matrix, with signs changed, contains asymptotic variances and covariances for the estimates when maximum likelihood estimates are substituted in the second partial derivatives). Further discussion of the conditions for identifiability and problems concerning boundaries for the estimates is presented in Dayton and Macready (1976).

E. ASSESSING GOODNESS OF FIT

Given the computerized estimation procedures which are available, it is possible to derive maximum likelihood estimates of the parameters if there are sufficient degrees of freedom once all parameters are specified and if the system of equations based on (3) is identifiable. However, the estimates do not necessarily provide good fit to the observed data. An assessment of goodness of fit can be made in several ways, but the simplest procedure is to utilize the maximum likelihood estimates of $P(\underline{u}_s)$ for each of the 2^K types of observed score vectors and, then, to apply an ordinary (Pearson) chi-square goodness-of-fit test based on observed and expected frequencies for these 2^K types. An alternative method which yields generally comparable values for the test statistic is the likelihood ratio chi-square test which, in effect, compares the expected frequencies (generated as for the Pearson case) with those arising under an unrestricted multinomial model. Degrees of freedom for both types of test are computed as $2^K - m - 1$, where m is the number of independent parameters estimated under the probabilistic model (i.e., for Case 1 of the Extreme Groups Model $m = 2K + 1$, while for Case 2, $m = 3$; for Hierarchic Models, under Case 2, $m = K + q - 1$, under Case 3, $m = q + 1$, and under Case 4, $m = q$).

In addition to assessing how well a given model fits an observed set of data, it is possible to compare the differential fit of alternate models under certain circumstances. A general rule is that the model with fewer parameters must be derivable, in theory, by a process of parameter restriction from the model which has the greater number of parameters. For example, Cases 1 and 2 of the Extreme Groups Model can be compared for differential fit since Case 2 can be derived from Case 1 by the restrictions $\alpha_i = \alpha$, $\beta_i = \beta$ for $i = 1,\ldots,K$. However, Cases 2 and 3 of the Hierarchic Model cannot be compared since neither case can be obtained from the other by a single set of restrictions; note, nevertheless, that Case 4 can be derived from either Case 2 or Case 3 and can be compared with either of these. An appropriate test statistic for comparing the relative fits of two models which meet the preceding conditions can be based on the difference between their respective goodness-of-fit chi-square values (based on either the Pearson or likelihood ratio approach) with degrees of freedom equal to the difference in degrees of freedom from these same two tests.

FOOTNOTE

[1]The authors will make available at no cost a Users Manual and single-copy listings of FORTRAN programs for all cases discussed above, with the exception of Case 1 of the Hierarchic Model. Written requests should be sent to the Department of Measurement & Statistics, College of Education, University of Maryland, College Park, Md. 20742

REFERENCES

Dayton, C. M. & Macready, G. B. A probabilistic model for validation of behavioral hierarchies. Psychometrika, 1976, 41, 189 - 204.

Goodman, L. A. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, 1974, 61, 215-231.

______ The analysis of systems of qualitative variables when some of the variables are unobservable. Part I - a modified latent structure approach. American Journal of Sociology, 1975, 79, 1179-1259.

______ A new model for scaling response patterns: an application of the quasi-independence concept. Journal of the American Statistical Association, 1976, 70, 755-768,

Lazarfeld, P. F. & Henry, H. W. Latent Structure Analysis. Boston: Houghton-Mifflin, 1968.

Macready, G. B. & Dayton, C. M. The use of probabilistic models in the assessment of mastery. Journal of Educational Statistics, 1977a, 2, 99-120.

______ Statistical comparisons among hierarchies based on latent structure models. Paper read the the Annual Meeting of the American Educational Research Association, New York, 1977b

Murray, J. R. Statistical models for qualitative data with classification errors. Ph.D. Thesis, University of Chicago, 1971.

Proctor, C. H. A probabilistic formulation and statistical analysis for Guttman scaling. Psychometrika, 1970, 35, 73-78.

Rao, C. R. Linear Statistical Inference and its Applications. New York: Wiley, 1965.