Behavior Research Methods & Instrumentation
1978, Vol. 10 (1), 47-48

PROGRAM ABSTRACTS/ALGORITHMS

A test for validating learning hierarchies based on maximum likelihood estimates

GORDON RAE
College of Education, Aberdeen, Scotland AB9 1FA

During the past 15 years, a number of indices have been used to validate postulated connections between pairs of skills in learning hierarchies. These have included Gagné and Paradise's (1961) "proportion positive transfer," the five variants in it proposed by Walbesser and Eisenberg (1972), Guttman's coefficient of reproducibility, and the phi correlation. White (1974) has pointed out that all of these indices are unsatisfactory for one or more of the following reasons: Sometimes the index can have a value that indicates a hierarchical connection even when the skills are really independent; errors of measurement are ignored; the index lacks a sampling distribution; and subjective judgments are necessary on how many questions of a set must be answered correctly for a "pass" in a skill. White and Clark (1973) have produced a test that avoids all these difficulties. The test can only be applied when two or more questions are used for each skill. When only one question is used, no estimate of the size of errors of measurement is possible.

Suppose that two questions are used for each of the two skills, I and II, and these are distributed by a sample of N subjects, whose responses are distributed as shown in Table 1. In this table f_i denotes an observed cell frequency and π_i, the probability that a member of the sample will be classified in that particular cell ($i = 1, \ldots, 9$; $\Sigma f_i = N$). Using White and Clark's (1973) notation, let the proportion of the population with neither skill be P_0; with skill I only, P_I; with skill II only, P_{II}; and with both skills, P_B. The hypothesis that those with skill II are totally included among those with skill I (i.e., the connection is valid) can be stated as H_0: $P_{II} = 0$, with the general alternative H_a: $P_{II} > 0$.

Table 1
Sample Distribution of Subjects Who Answered Two Questions for Each of Two Skills

	II		
I	0	1	2
2	f_1 π_1	f_2 π_2	f_3 π_3
1	f_4 π_4	f_5 π_5	f_6 π_6
0	f_7 π_7	f_8 π_8	f_9 π_9

Note—I = number of questions correct, lower skill; II = number of questions correct, higher skill.

Let the probability of someone with skill I answering correctly any question for skill I and θ_a and the probability of someone without skill I answering correctly any question for skill I be θ_b. Let θ_c and θ_d be the corresponding probabilities for skill II. Each of the πs shown in Table 1 is a function of the P and θ parameters. The test focuses on π_9, the probability that a member of the sample will be classified in the I-0, II-2 cell, which is given by:

$$\pi_9 = P_0(1 - \theta_b)^2\theta_d^2 + P_I(1 - \theta_a)^2\theta_d^2$$
$$+ P_{II}(1 - \theta_b)^2\theta_c^2 + P_B(1 - \theta_a)^2\theta_c^2$$

(White & Clark, 1973, Table 1).

White and Clark solve this question for π_9 by assuming that $\theta_b = 0$, and $\theta_c = 1$, and obtaining estimates for θ_a, θ_d, P_0, and P_I using marginal totals from the 3×3 contingency table. A value of $\hat{P}_{II}$ is also chosen: either zero under H_0, or a particular value under H_a. The remaining parameter, P_B, is estimated using $\hat{P}_B = 1 - \hat{P}_0 - \hat{P}_I - \hat{P}_{II}$. Once a value for $\hat{\pi}_9$ is found, it is easy to estimate the probability that the observed frequency in the I-0, II-2 cell will be $0, 1, 2, \ldots,$ since f_9 is a binomial random variable with parameters (N, π_9) under either hypothesis.

In the present test the values of the parameters θ_a, θ_d, P_0, and P_I are estimated from cell frequencies using the method of maximum likelihood. Since this method uses more information than White and Clark's (1973) marginal total method, it should provide more accurate estimates of these parameters and, hence, of π_9. The method of maximum likelihood is one of selecting an estimate $\hat{\lambda}_i$ for the parameter λ_i which will maximize the likelihood function L given, in this instance, by:

$$L = \pi_1^{f_1}\pi_2^{f_2} \ldots \pi_k^{f_k}.$$

Since log L is a monotonic function and attains its maximum when L is a maximum, the problem reduces to the solution of four simultaneous nonlinear equations:

$$\frac{\partial \log L}{\partial \lambda_i} = 0, \qquad (1)$$

where $\underline{\lambda} = (\theta_a, \theta_d, P_0, P_I)$.

To solve the likelihood Equations 1, some first approximation $\underline{\lambda}_0$ is assumed to be available for each of the parameters to be estimated. If the second approximation is $\underline{\lambda}$ and $\underline{\delta} = \underline{\lambda} - \underline{\lambda}_0$, then it can be shown that:

$$\underline{\delta} = \underline{B}^{-1}\underline{\phi}/N, \qquad (2)$$

where, for large samples, the elements of Matrices $\underline{B}$ and $\underline{\phi}$ are given by:

$$b_{ut} = \Sigma \frac{1}{\pi_i} \frac{\partial \pi_i}{\partial \lambda_u} \frac{\partial \pi_i}{\partial \lambda_t},$$

and

$$\phi_t = \Sigma \frac{f_i}{\pi_i} \frac{\partial \pi_i}{\partial \lambda_t}.$$

(A detailed mathematical proof will be provided by the author on request).

I have produced a program for solving the maximum likelihood Equations 1 and obtaining estimates for the parameters $\lambda = \theta_a, \theta_d, P_0, P_1$. This involves selecting trial values for the four parameters (using White and Clark's estimates based on marginal totals) and then solving Equation 2 to obtain the additive corrections $\underline{\delta}$. This operation is repeated with corrected values each time until stable values of $\underline{\lambda}$ are obtained (in the program this is arbitrarily defined as $|\underline{\delta}| < 10^{-5}$). Expressions for π_i ($i = 1, \ldots, 9$) in terms of the P and θ parameters are those provided by White and Clark (1973, Table 1). It is important to note that after some stage, Matrix B need not be recalculated for each approximation. Only Matrix ϕ has to be calculated at each stage and used in conjunction with the same inverse matrix $C = B^{-1}$ (kept constant from some stage) to obtain closer approximations (Rao, 1965, pp. 305-309). In the program, the same inverse Matrix C is used after the fifth iteration. When convergence is achieved, or the number of iterations exceeds 20, Matrices C and ϕ are calculated for the last approximation values, and the final estimates and their variances are obtained. The variance of the final estimate $\hat{\lambda}_i$ is obtained from C_{ii}, the ith diagonal element in the inverse of B, that is:

$$V(\hat{P}_0) = \frac{C_{11}}{N},\ V(\hat{P}_1) = \frac{C_{22}}{N}, \text{etc.}$$

Since $\hat{P}_B = 1 - \hat{P}_0 - \hat{P}_1 - \hat{P}_{11}$, where $\hat{P}_{11} = 0$ under H_0 or a particular value under H_a,

$$V(\hat{P}_B) = V(\hat{P}_0) + V(\hat{P}_1) + 2\,\mathrm{Cov}(\hat{P}_0; \hat{P}_1)$$

$$= \frac{C_{11} + C_{22} + 2C_{12}}{N}.$$

In most cases the estimates obtained by White and Clark's (1973) marginal total method and the maximum likelihood method described above are in close agreement. Consider, however, the following two distributions of f_i ($i = 1, \ldots, 9$): (i) 25, 0, 20, 2, 13, 0, 38, 0, 2; (ii) 20, 5, 20, 15, 0, 0, 30, 8, 2. In both cases the marginal totals are the same and White and Clark's method yields a value for $\hat{\pi}_9 = .00763$, despite the fact that two frequency distributions are different. Using the maximum likelihood method, the values for $\hat{\pi}_9$ are .00729 and .00926, respectively.

Input. The input data are the number of postulated connections to be validated and, for each connection, the values for f_i ($i = 1, \ldots, 9$) and the value for P_{11} (either zero under H_0, or a particular value under H_a).

Output. For each connection the following are given as printed output: estimates for all parameters and their variances, $\hat{\pi}_i$, and the probabilities of obtaining frequencies 0, 1, . . . 5 in the I-0, II-2 cell.

Computer and Language. This program is written in FORTRAN IV and was prepared and tested on the Aberdeen University ICL 470 computer. Double precision is used for increased accuracy.

Restrictions. The present program is limited to $f_i < 10^4$ and the number of connections being less than 10^4. This should be more than sufficient for most problems but, if necessary, these limitations can be changed by modifying two FORTRAN statements.

Availability. A source listing, which includes data and printed output for a sample analysis, may be obtained without charge from Gordon Rae, Department of Educational Psychology, Aberdeen College of Education, Hilton Place, Aberdeen, Scotland AB9 1FA.

REFERENCES

Gagné, R. M., & Paradise, N. E. Abilities and learning sets in knowledge acquisition. *Psychological Monographs*, 1961, **75**(Whole No. 518).

Rao, C. R. *Linear statistical inference and its applications*. New York: Wiley & Sons, 1965.

Walbesser, H. H., & Eisenberg, T. A. A review of research on behavioural objectives and learning hierarchies. *Mathematics Education Reports*. Columbus, Ohio: ERIC Information Analysis Center for Science, Mathematics and Environmental Education, 1972.

White, R. T. Indexes used in testing the validity of learning hierarchies. *Journal of Research in Science Teaching*, 1974, **11**, 61-66.

White, R. T., & Clark, R. M. A test of inclusion which allows for errors of measurement. *Psychometrika*, 1973, **38**, 77-86.

(Accepted for publication September 8, 1977.)