

Reports from the Research Laboratories
of the Department of Psychiatry
University of Minnesota

**Detecting Latent Clinical Taxa, VI:
Analytical Development and Empirical Trials
of the Consistency Hurdles Theory.***

by

**Robert R. Golden, Shirley H. Tyan and
Paul E. Meehl**

October 1974

Report Number PR-74-4

Contents

I. Introduction 2

II. Development of the Consistency Hurdles Theory 2

III. Empirical Trials of the Theory..... 14

IV. Discussion 18

V. A method for the Selection of Discriminative indicators 23

VI. Conclusions 26

Notes for this digital copy

Subscripts ‘t’ for taxon and ‘c’ for complement groups were used in later publications. This research report uses ‘s’ (schizotype), ‘n’ (nonschizotype), and ‘c’ (compound/mixed taxon-nontaxon subsets).

Pagination differs from the original. Some mis-typings were corrected and tables and figures better aligned with text.

*This research was supported in part by grants from the Psychiatry Research Fund and the National Institute of Mental Health, Grant Number MH 24224.

I. Introduction

In previous reports in the present research report series it has been assumed that there exist a few, say less than a dozen, candidate taxometric indicators. Given one or more discriminative indicators the various theories developed can be used to estimate the parameters of the underlying latent situation (Meehl, 1965, 1968; Golden & Meehl, 1973a). In this report it is shown that a theory is frequently needed for the initial selection of the discriminative indicators from a set of candidate indicators which can be quite large in size.

The present theory provides for both indicator selection and parameter estimation and does this in an integrated fashion. An indicator is rejected if either (a) it is not discriminative enough or (b) it is not in adequate agreement with the assumptions required for the estimation of the latent parameters. Another theory with weaker assumptions than the consistency hurdles theory is also developed just for the selection of discriminative indicators.

The central idea underlying the consistency hurdles theory is that indicators which are both discriminative and in agreement with the assumptions of the theory will behave consistently with each other in various ways. A series of consistency tests (see Golden & Meehl, 1973b for a general discussion) are used in a consecutive hurdles fashion; when an indicator fails a hurdle it is rejected at that point and the data contributed by that indicator are not used in the remaining calculations. The hurdles are repeatedly applied to each indicator until none of the remaining indicators fail any of them.

II. Development of the Consistency Hurdles Theory

A random number generator was used to simulate responses (with values 0 and 1) of one thousand individuals, five hundred from a taxon and 500 from the non-taxon class, to each of fifteen dichotomous indicators such that for each pair of indicators the population within-taxon and within-non-taxon class phi-coefficients were zero. Analysis of these data will serve as an example of the method. The indicator parameters are given in Table 1. Note that nine items are extremely weak and six have marginal to substantial validities.

Table 1. Artificial data population item parameters

Item	Taxon mean	Non-taxon mean	Difference
1	.364	.301	.063
2	.400	.134	-.266*
3	.696	.758	-.052
4	.582	.251	.331*
5	.887	.754	.133
6	.899	.708	.191*
7	.637	.258	.379*
8	.515	.471	.041
9	.625	.371	.244*
10	.844	.481	.363*
11	.268	.201	.067
12	.667	.572	.095
13	.737	.809	-.072
14	.625	.557	.072
15	.902	.797	.105

*Difference between taxon and non-taxon class means is large enough for item to have at least marginal validity.

The method proceeds as follows:

A) Let y_i denote the Bernoulli item, $i = 1, 2, \dots, n$ and x_i the key formed by summing the $n - 1$ other items; i.e.,

$$x_i = \sum_{\substack{j=1 \\ (j \neq i)}}^n y_j$$

For each item, using the key x_i as the “input” variable and an item y_i as the “output” variable, we can create the function

$$\bar{y}_d(x) = \bar{y}_a(x) - \bar{y}_b(x)$$

where \bar{y}_a is the mean on y of those cases lying above a cut x on the input key and \bar{y}_b is the corresponding quantity for below the x -cut. It will be shown that we can require that

$$\max\{\bar{y}_d(x)\} > .10 = \delta_1. \quad (\mathbf{c}_1)$$

Assume that \bar{y}_{sx} and \bar{y}_{nx} , the within taxon and non-taxon class regression functions are each constant for all x . it is easily shown that

$$\bar{y}_d(x) = \frac{H_s(x)}{N_a(x)} - \frac{M_s(x)}{N_b(x)} \bar{y}_s = \frac{H_n(x)}{N_b(x)} - \frac{M_n(x)}{N_a(x)} \bar{y}_n$$

where $H_s(x)$ and $M_s(x)$ are the numbers of hits and misses among the taxon members, $H_n(x)$ and $M_n(x)$ are the corresponding quantities for the non-taxon class, and $N_a(x)$ and $N_b(x)$ are the total numbers above and below a cut score x (Meehl, 1968). It would be reasonable to assume that for non-truncated taxon and non-taxon class distributions intersecting in only one point that as $x \rightarrow x_{\max}$,

$$\frac{H_s}{N_a} \rightarrow 1, \quad \frac{M_s}{N_b} \rightarrow P, \quad \frac{H_n}{N_b} \rightarrow 1 - P, \quad \text{and} \quad \frac{M_n}{N_a} \rightarrow 0,$$

where P is the base-rate of the taxon, likewise as $x \rightarrow x_{\min}$,

$$\frac{H_n}{N_b} \rightarrow 1, \quad \frac{M_n}{N_a} \rightarrow 1 - P, \quad \frac{H_s}{N_a} \rightarrow 1, \quad \text{and} \quad \frac{M_s}{N_b} \rightarrow 0.$$

It follows that

$$\begin{aligned} \max \{ \bar{y}_d(x) \} &> \max \left\{ \lim_{x \rightarrow x_{\max}} \bar{y}_d(x), \lim_{x \rightarrow x_{\min}} \bar{y}_d(x) \right\} \\ &= \max \left\{ (1-P)(\bar{y}_s - \bar{y}_n), P(\bar{y}_s - \bar{y}_n) \right\} \geq \frac{1}{2}(\bar{y}_s - \bar{y}_n) . \end{aligned}$$

Experience has shown that it is generally true that $\bar{y}_d(x)$ is a convex downward parabolic shaped curve with a local maximum between x_{\min} and x_{\max} near the hitmax cut¹ when the discrimination by the key and the output item is strong enough. Analytical demonstration of this fact has unfortunately proved elusive so far. However, a Monte Carlo study of $\bar{y}_d(x)$ which will be given in a forthcoming report in this series, does indicate that discriminative items which are weakly correlated with the input key within the taxon and non-taxon class have $\bar{y}_d(x)$ curves which are concave downward. Also, it appears to be true and will be demonstrated below that items for which $\bar{y}_s - \bar{y}_n < .20$ are too weak for the method to give accurate parameter estimates. Detection of these weak items can be partially accomplished by the requirement $\max \{ \bar{y}_d(x) \} > .10$. The results obtained for the example are given in the table below. It is necessary to require the frequencies N_a and N_b to be large enough to avoid selecting maxima not associated with the hitmax cut but resulting from sampling error. So far it has been found adequate to restrict N_a and N_b to be fifty or more.

Items 3, 5, 8, 12, 13 and 15 are deleted by consistency test c_1 . If the $\bar{y}_d(x)$ curves are smoothed by a method such as moving averages then items 1, 2 and 11 are also deleted.

¹ The relationships between $\max \{ \bar{y}_d(x) \}$ and the hitmax cut is developed in PR-68-4, p. 14ff [section “[Quasi]-Proof”]. (The hitmax cut is defined to be the cut on the input key which maximizes the total number of hits; in other words, it is the abscissa of the point of intersection of the taxon and non-taxon (unrelativized) frequency curves.)

The local maximum of the mean above less the mean below for each of fifteen artificial data items

Item	$\max \{ \bar{y}_d(x) \}$
1	.12
2	.19
3	.02
4	.22
5	.08
6	.19
7	.18
8	.08
9	.24
10	.22
11	.11
12	.07
13	.05
14	.11
15	.08

B) After deletion of k items by \mathbf{c}_1 then the x_i keys are recalculated using the $n - k - 1$ items for each. When condition \mathbf{c}_1 is applied again no items are deleted in the example. The interval of x_i associated with $\max \{ \bar{y}_d(x) \}$ for each output item y_i is found; denote this estimated hitmax interval value by h_i and the mean of the h_i values by \bar{h} . The next hurdle then is to require that

$$|h_i - \bar{h}| < \delta_2. \quad (\mathbf{c}_2)$$

It appears difficult to show mathematically that δ_2 can be safely fixed at any adequately low value. It is easily shown that if items are perfectly independent within the taxon and non-taxon class, then the h_i values will not deviate by more than one interval because of the fact that the various input keys are not composed of exactly the same items and thus

have different distributions. The effects of other sources of deviation of the h_i values are more difficult to evaluate. However, a crudely derived value for δ_2 , 15% of the range of the input key, has worked well on both artificial and real data. This parameter has to be studied further but it appears that normally it can be set at one or two interval widths after weak items have been deleted. In the example, the hitmax interval is estimated to be 2 for all 9 items remaining except #4, where it is estimated to be 3. By the “15% of the range” rule no items are deleted.

C) The manifest tail means can be used as estimates of \bar{y}_{si} and \bar{y}_{ni} (again using key x_i as the input variable). The idea here is to take enough of the tails to avoid excessive sampling error but not so much as to obtain excessive bias due to the contaminating taxon or non-taxon class. So far, the crude rule of making each tail range be about 25% of the total range has worked quite well. There can be considerable distribution overlap in the tails; certainly, the tails need not be “exposed flanks.” The use of non-exposed flanks causes \hat{y}_{si} to be too small and the \hat{y}_{ni} to be too large if y_i satisfies the independence assumption. It usually happens that the output item y_i is somewhat positively correlated with x_i within the taxon and the non-taxon class and this correlation tends to counterbalance the non-exposed flank biasing effect. In order that an item be discriminative enough it can be required that

$$\hat{y}_{si} - \hat{y}_{ni} > \delta_3 \quad (\mathbf{c}_3)$$

where \hat{y}_{si} and \hat{y}_{ni} are tail estimates. Since $\hat{y}_{si} - \hat{y}_{ni}$ can be as low [as] .20 and the item still be considered discriminative enough (as will be shown below) and since \hat{y}_{si} will tend to be too small and \hat{y}_{ni} will tend to be too large but not so that $\hat{y}_{si} - \hat{y}_{ni} < .10$ (according to \mathbf{c}_1) it would seem appropriate to set $\delta_3 = .15$. As shown in the table below, consistency test \mathbf{c}_3 eliminates items 1, 11 and 14, each weakly discriminative, thereby leaving items 2, 4, 6, 7, 9 and 10. Comparison of these estimates with true values given in Table 1 shows that generally these estimates are within .05 of the true values.

Tail estimates of the taxon and non-taxon
class parameters for the remaining nine items

	Item	\hat{y}_{si}	\hat{y}_{ni}	$\hat{y}_{si} - \hat{y}_{ni}$
1.	1	.378	.246	.132
2.	11	.276	.202	.072
3.	14	.620	.536	.084
4.	2	.402	.169	.233
5.	4	.559	.218	.341
6.	6	.877	.717	.160
7.	7	.601	.259	.342
8.	9	.557	.406	.151
9.	10	.818	.532	.286

D) In the hitmax interval cases are an approximately 50% mixture of the taxon and the non-taxon class so it is true that

$$\frac{1}{2}\hat{y}_{si} + \frac{1}{2}\hat{y}_{ni} \doteq \hat{y}_{hi}$$

where h denotes the hitmax interval. Hence, we can require that

$$\left| \frac{1}{2}\hat{y}_{si} + \frac{1}{2}\hat{y}_{ni} - \hat{y}_{hi} \right| < \delta_4 \quad \text{(C4)}$$

where the parameter δ_4 will not be zero because (a) \bar{y}_h is in error due to sampling, (b) the hitmax interval has not been correctly determined, (c) the estimates \hat{y}_{si} and \hat{y}_{ni} are in error, (d) the taxa proportions in the interval are not exactly .50 due to coarseness of intervals and (e) the assumption of the theory (\hat{y}_{si} and \hat{y}_{ni} are constant for all i) is not true. Let ε denote the difference between the theoretical value in terms of the true latent parameter values and the true population value hitmax compound mean. Dropping the subscript i , we have

$$\varepsilon = P_h \bar{y}_{sh} + (1 - P_h) \bar{y}_{nh} - \bar{y}_h$$

where h denotes the hitmax interval and all parameters are the true values. Then

$$d\varepsilon = \frac{\partial \varepsilon}{\partial P_h} dP_h + \frac{\partial \varepsilon}{\partial \bar{y}_{sh}} d\bar{y}_{sh} + \frac{\partial \varepsilon}{\partial \bar{y}_{nh}} d\bar{y}_{nh} + \frac{\partial \varepsilon}{\partial \bar{y}_h} d\bar{y}_h$$

and

$$\Delta \varepsilon \doteq (\bar{y}_{sh} - \bar{y}_{nh}) \Delta P_h + P_h \Delta \bar{y}_{sh} + (1 - P_h) \Delta \bar{y}_{nh} - \bar{y}_h$$

where Δ denotes a small error in the parameter. Under the assumption given in (e) above, and with fine enough intervals we have

$$\Delta \varepsilon \doteq (\bar{y}_s - \bar{y}_n) \Delta P_h + \frac{1}{2} (\Delta \bar{y}_{sh} + \Delta \bar{y}_{nh}) - \Delta \bar{y}_h.$$

Usually we are able confidently to assume that $\bar{y}_s - \bar{y}_n < .6$ and $\Delta P_h < .25$. In the second term, $\Delta \bar{y}_{sh}$ and $\Delta \bar{y}_{nh}$ will tend to have opposite signs when the tail estimation method is used; each of $|\Delta \bar{y}_{sh}|$ and $|\Delta \bar{y}_{nh}|$ should be less than .10 and their algebraic sum less than, say, .06. The third term represents sampling error; assuming that the number in the hitmax interval is at least 100 (keys are quite short at this stage) it follows that one probable error unit in the estimate of \bar{y}_h is less than .05. Thus we have

$$\Delta \varepsilon \leq .15 + .03 - (\pm .10);$$

hence, it would seem reasonable to set $\delta_4 = .30$. In the artificial data example, no items are deleted by \mathbf{c}_4 .

E) Since items were eliminated by \mathbf{c}_3 we now recalculate the keys x_i and apply tests \mathbf{c}_2 , \mathbf{c}_3 , and \mathbf{c}_4 again. In the example all items pass these tests this time except item 6 on \mathbf{c}_2 (this is the weakest of the remaining items). Recalculation of the x_i and application of tests \mathbf{c}_2 , \mathbf{c}_3 , and \mathbf{c}_4 eliminates item 4 (a strong item). Finally, the next cycle does not eliminate any of the remaining four Items (2, 4, 9 and 10).

F) The final tail estimates for each of the four remaining items are given in Table 2 and can be used to solve for the base-rate P by

$$P \hat{\bar{y}}_{si} + (1 - P) \hat{\bar{y}}_{ni} = \bar{y}_i \quad \text{or} \quad P = \frac{\hat{\bar{y}}_i - \hat{\bar{y}}_{ni}}{\hat{\bar{y}}_{si} - \hat{\bar{y}}_{ni}}.$$

Item #	non-taxon			taxon			estimate of base-rate
	estimate	true value	error	estimate	true value	error	
2	.128	.134	-.006	.364	.400	-.046	.680
4	.292	.251	.041	.591	.582	.009	.626
9	.374	.371	.003	.614	.625	-.011	.603
10	.519	.481	.038	.830	.844	-.014	.557
						Average	.617
						Corrected for bias	.574
						True value	.582

Table 2. Item parameter and base-rate estimates for artificial data example (N = 1000)

G) The error in P_i , $\Delta P_i = \hat{P}_i - P$, resulting from using the above equation is a function of $\Delta \bar{y}_i$ (sampling error), $\Delta \bar{y}_{ni} = \hat{\bar{y}}_{ni} - \bar{y}_{ni}$ and $\Delta \bar{y}_{si} = \hat{\bar{y}}_{si} - \bar{y}_{si}$:

$$\Delta P_i \doteq \frac{1}{\bar{y}_{si} - \bar{y}_{ni}} \Delta \bar{y}_i - \frac{\bar{y}_i - \bar{y}_{ni}}{(\bar{y}_{si} - \bar{y}_{ni})^2} \Delta \bar{y}_{si} + \frac{\bar{y}_i - \bar{y}_{si}}{(\bar{y}_{si} - \bar{y}_{ni})^2} \Delta \bar{y}_{ni} .$$

If it is assumed that $\Delta \bar{y}_{ci} = \Delta \bar{y}_{ni} = -\Delta \bar{y}_{si}$ we have

$$\Delta P_i = \frac{\Delta \bar{y}_i}{\bar{y}_{si} - \bar{y}_{ni}} + \frac{2P\bar{y}_{si} + 2(1-P)\bar{y}_{ni} - \bar{y}_{si} - \bar{y}_{ni}}{(\bar{y}_{si} - \bar{y}_{ni})^2} \Delta \bar{y}_{ci}$$

which simplifies to

$$\Delta P_i = \frac{\Delta \bar{y}_i}{\bar{y}_{si} - \bar{y}_{ni}} + \frac{(2P-1)}{(\bar{y}_{si} - \bar{y}_{ni})} \Delta \bar{y}_{ci} .$$

It is reasonable to assume that $\Delta \bar{y}_{ci} < .05$ and that $|\Delta \bar{y}_i| < 2 \frac{(.5)}{(1000)^{1/2}} \doteq .03$. To get some

idea of the size of variation in P_i values for which to generally allow, let $P = .60$, for example and it follows that

$$\frac{-.04}{\bar{y}_{si} - \bar{y}_{ni}} < \Delta P_i < \frac{+.04}{\bar{y}_{si} - \bar{y}_{ni}}$$

which results in limits of $\pm .2$ for $\bar{y}_{si} - \bar{y}_{ni} = .2$ and limits of $\pm .1$ for $\bar{y}_{si} - \bar{y}_{ni} = .4$. If we

have n items, each with $\bar{y}_{si} - \bar{y}_{ni} = \bar{y}_s - \bar{y}_n$ and $\Delta y_{ci} = \Delta y_c$ and estimate P by taking the

average of the n individual estimates, then

$$\begin{aligned}\Delta P &= \hat{P} - P = \frac{\sum \hat{P}_i}{n} - P = \frac{\sum (\hat{P}_i - P)}{n} = \frac{\sum \Delta P_i}{n} \\ &= \frac{\sum \Delta \bar{y}_i}{n(\bar{y}_s - \bar{y}_n)} + \frac{\sum (2P - 1) \Delta \bar{y}_c}{\bar{y}_s - \bar{y}_n}.\end{aligned}$$

The expected value of ΔP is given by

$$E(\Delta P) = \frac{(2P - 1) \Delta \bar{y}_c}{\bar{y}_s - \bar{y}_n}.$$

which shows that there is a bias in the estimation of P by this procedure if $\Delta \bar{y}_c \neq 0$.

The above formula for $E(\Delta P)$ was used to obtain the following table which gives the bias in the estimation of P for various values of $\bar{y}_s - \bar{y}_n$ and P when $\Delta \bar{y}_c = .05$.

The expected error in the base-rate estimate when $\Delta \bar{y}_c = .05$
for various values of P and $\bar{y}_s - \bar{y}_n$

the true value of P	the true value of the difference in the item parameters $\bar{y}_s - \bar{y}_n$			
	.2	.3	.4	.5
.5	0.00	0.00	0.00	0.00
.6	.05	.03	.02	.02
.7	.10	.07	.05	.04
.8	.15	.10	.07	.06
.9	.20	.15	.10	.08

Thus, the bias can be very large and one is clearly behooved to make use of a correction term. The general form of the bias term for n items is

$$\frac{2P - 1}{n} \sum_i \frac{\Delta \bar{y}_{ci}}{\bar{y}_{si} - \bar{y}_{ni}}$$

or

$$\frac{(2\hat{P} - 1) \Delta \hat{y}_c}{n} \sum_i \frac{1}{\hat{y}_{si} - \hat{y}_{ni}}.$$

Taking $\Delta\bar{y}_c = .05$ and using the parameter estimates of the example, this expression has the value of .043; thus the approximate unbiased estimate is .574. The sampling variance of ΔP which only takes into account the sampling error of \hat{y}_i is given by

$$\sum \frac{\text{var } \Delta\bar{y}_i}{n^2 (\bar{y}_{si} - \bar{y}_{ni})^2} = \sum \frac{\text{var } \hat{y}_i}{n^2 (\bar{y}_{si} - \bar{y}_{ni})^2} \leq \frac{.25}{Nn} \sum \frac{1}{(\bar{y}_{si} - \bar{y}_{ni})^2} .$$

If it is assumed for simplicity that $\bar{y}_s - \bar{y}_n$ is the same for all items, then a conservative 95% confidence interval for the resulting sampling standard deviation is

$$\pm \frac{1}{N^{1/2} n^{1/2} (\bar{y}_s - \bar{y}_n)}$$

and from this expression we get the very important result that it is clearly more important to have a few strong items rather than many weak items. For example, 5 items with $\bar{y}_s - \bar{y}_n = .4$ gives a better estimation of P than 16 items with $\bar{y}_s - \bar{y}_n = .2$.

The sampling standard deviation of ΔP which takes into account the sampling error in \hat{y}_i , \hat{y}_{si} and \hat{y}_{ni} can be derived as follows. From above we have

$$\Delta P = \frac{\sum \Delta P_i}{n} \quad \text{and} \quad P_i = \frac{\hat{y}_i - \hat{y}_{ni}}{\hat{y}_{si} - \hat{y}_{ni}} .$$

Thus

$$\text{var}(\Delta P) = \frac{\sum \text{var}(\Delta P_i)}{n^2}$$

$$\text{var}(\Delta P) = \frac{\text{var}\left(\frac{a_i}{b_i}\right)}{n^2}$$

where $a_i = \hat{y}_i - \hat{y}_{ni}$ and $b_i = \hat{y}_{si} - \hat{y}_{ni}$. It will be necessary to assume that \hat{y}_i , \hat{y}_{si} and \hat{y}_{ni} are mutually independent and that $\hat{y}_i - \hat{y}_{ni}$ and $\hat{y}_{si} - \hat{y}_{ni}$ are independent even though this is not strictly true. The error resulting from this assumption would not seem to be of significant size but an analytical demonstration is not provided. From the

assumption it follows that the expected value of the quotient $\frac{a_i}{b_i}$ is the quotient of the

expected values of a_i and b_i , or

$$E\left(\frac{a_i}{b_i}\right) = \frac{E(a_i)}{E(b_i)} = \frac{\bar{a}_i}{\bar{b}_i} .$$

Now we have

$$\begin{aligned} \text{var}\left(\frac{a_i}{b_i}\right) &= E\left[\frac{a_i}{b_i} - E\left(\frac{a_i}{b_i}\right)\right]^2 \\ &= E\frac{a_i^2}{b_i^2} - \frac{\bar{a}_i^2}{\bar{b}_i^2} \\ &= \frac{\text{var}(a_i) + \bar{a}_i^2}{\text{var}(b_i) + \bar{b}_i^2} - \frac{\bar{a}_i^2}{\bar{b}_i^2} . \end{aligned}$$

Since

$$\begin{aligned} \text{var}(a_i) &\doteq \frac{\hat{y}_i(1-\hat{y}_i)}{N} + \frac{\hat{y}_{ni}(1-\hat{y}_{ni})}{\frac{1}{4}N} , \\ \text{var}(b_i) &\doteq \frac{\hat{y}_{si}(1-\hat{y}_{si})}{\frac{1}{4}N} + \frac{\hat{y}_{ni}(1-\hat{y}_{ni})}{\frac{1}{4}N} , \\ \bar{a}_i &\doteq \hat{y}_i - \hat{y}_{ni} , \quad \text{and} \quad \bar{b}_i \doteq \hat{y}_{si} - \hat{y}_{ni} \end{aligned}$$

the resulting expression for $\text{var}(\Delta P)$ is

$$\frac{1}{n^2} \sum_i \left\{ \frac{\hat{y}_i(1-\hat{y}_i) + 4(\hat{y}_{ni}(1-\hat{y}_{ni})) + N(\hat{y}_i - \hat{y}_{ni})^2}{4\hat{y}_{si}(1-\hat{y}_{si}) + 4(\hat{y}_{ni}(1-\hat{y}_{ni})) + N(\hat{y}_{si} - \hat{y}_{ni})^2} - \frac{(\hat{y}_i - \hat{y}_{ni})^2}{(\hat{y}_{si} - \hat{y}_{ni})^2} \right\}$$

Suppose $P = .5$, $N = 1000$, $\hat{y}_i = .5$ and $\hat{y}_{si} - \hat{y}_{ni} = .2$ or that all items are marginally valid;

then

$$\begin{aligned} \text{var}(\Delta P) &= \frac{1}{n^2} \cdot n \cdot \frac{.25 + .96 + 10}{.96 + .96 + 40} - \frac{10}{40} \\ &= \frac{.017}{n} \quad \text{or} \\ \text{SD}(\Delta P) &= \frac{.130}{\sqrt{n}} \end{aligned}$$

which is .065 when $n = 4$, or .043 when $n = 9$ and .032 when $n = 16$. We see that for marginally discriminative items and a sample size of 1000 it is necessary for the twenty items to pass all the hurdles for the base-rate estimate to be accurate enough for most purposes. If items discriminate each by as much as .40 then $\text{var}(\Delta P) = \frac{.004}{n}$ or

$$SD(\Delta P) = \frac{.06}{\sqrt{n}} \text{ which is only .03 when } n = 4.$$

III. Empirical Trials of the Theory

Several empirical trials in which MMPI items are used to identify the sexes give an idea of how well the theory works with real data in terms of sample size, discrimination power of the items, inter-item dependency within taxa and the proportion of the items in the initial item pool which are discriminative. Two samples of 430 men and 720 women were used to obtain 15 items that discriminate by .30 to .40 (difference in the item mean plus-rates), 12 by .20 to .30, 15 by .10 to .20 and 33 by $-.10$ to $+.10$ for a total of 75 items. Twenty-seven of the items can be taken as discriminating adequately enough to serve as useful taxonomic indicators (.20 or more difference).

It is of interest first to mention the results of applying other well-known taxometric theories to the 75 items. The hierarchical nearest-neighbor method developed by Ward (1963) was used with the 75 items and the method failed to come close in correct identification of the sexes at any stage of the iterative procedure. Factor analysis, using both the principle components and varimax rotation methods, of the compound distribution failed to segregate the discriminative items in any way. Also, inverted factor analysis developed by Stephenson (1938) failed to produce factors which were loaded with mostly members of just one sex. It may be thought that since only 27 of the 75 items were adequately discriminative the data were concocted to be too tough for any taxometric method. But we would contend in reply that this would usually be a rather high concentration of valid candidate indicators at least in preliminary stages of a taxonomic investigation in personality measurement.

When the consistency hurdles method was applied it was found that 28 of the items produced \bar{y}_d curves that had local maxima which were sufficient in size and were well-

defined. Further analysis of the 28 items by the method eliminated all but 7 for which the male and female item plus-rates were estimated fairly accurately (usually within .10) and the base-rate was estimated to be .670 which compares well enough with the true value of .630 (see Table 3). While there is some room for improvement in the method in that it retained less than one-fourth of the actually valid items it did provide estimates of the base-rate and item plus-rates which yielded nearly 90% correct classification of individuals according to biological sex by application of Bayes' Rule.

Item	females			males			estimate of base-rate
	estimate	true value	error	estimate	true value	error	
1	.48	.69	-.21	.08	.06	.02	.74
2	.67	.74	-.07	.24	.31	-.07	.68
3	.44	.55	-.11	.14	.15	-.01	.71
4	.53	.54	-.01	.04	.11	-.07	.58
5	.63	.77	-.14	.48	.48	.00	.98
6	.33	.34	-.01	.08	.27	-.19	.90
7	.74	.69	.05	.36	.55	-.19	.69
Average (corrected for bias)							.67
True sample value							.63

Table 3. Item parameter and base-rate estimates for the seven items which remained in the seventy-five item male-female trial (N = 1150)

If only the 15 most discriminative items are used then each of the above methods do result in nearly 90% correct identification of the sexes. It should be noted that the consistency hurdles method does provide more information in that the item plus-rates are estimated quite accurately (usually within .05) for each sex even when a compound sample size as small as 200 is used (see Table 4).

While it was found that the consistency hurdles method worked very well for samples as small as 100 when the fifteen most discriminative items were used, it was also found that the sample size had to be at least 500 for the 75 item trial.

Item	females			males			estimate of base-rate
	estimate	true value	error	estimate	true value	error	
1	.66	.64	.02	.05	.08	-.03	.53
2	.37	.30	.07	.09	.10	-.01	.57
3	.87	.90	-.03	.41	.50	-.09	.55
4	.84	.74	.10	.42	.44	-.02	.46
5	.91	.84	.07	.46	.40	.06	.48
6	.87	.78	.09	.25	.36	-.11	.54
7	.54	.52	.02	.05	.14	-.09	.56
8	.47	.38	.09	.08	.08	.00	.55
9	.80	.84	-.04	.30	.36	-.06	.57
Average (corrected for bias)							.527
True sample value							.500

Table 4. Item parameter and base-rate estimates for nine items that remained in the small sample, fifteen strong discriminator items (mean difference in item parameters is greater than .30) male-female trial (N = 200)

Item	females			males			estimate of base-rate
	estimate	true value	error	estimate	true value	error	
1	.58	.60	-.02	.35	.36	-.01	.68
2	.79	.66	.13	.17	.38	-.21	.63
3	.79	.66	.13	.31	.46	-.15	.59
4	.73	.61	.12	.17	.39	-.22	.63
5	.83	.73	.10	.29	.48	-.19	.65
6	.81	.77	.04	.41	.48	-.07	.63
7	.72	.61	.11	.17	.37	-.20	.64
Average (corrected for bias)							.606
True sample value							.630

Table 5. Item parameter and base-rate estimates for the seven items that remained in the twelve moderately (mean difference in item parameters between .20 and .30) discriminative item male-female trial (N = 1150)

When the set of twelve items that discriminate between .20 and .30 was used for the total group of 1150 the consistency hurdles estimates were marginally accurate (usually within .10 to .20) while the base-rate estimate of .606 was close to the true value of .630 (see Table 5). The results of using the items that discriminate by .10 to .20 were

marginally acceptable at best. The item parameters sometimes were substantially in error and the base-rate estimate of .558 is a bit off the mark, the true value being .630 (see Table 6).

Item	females			males			estimate of base-rate
	estimate	true value	error	estimate	true value	error	
1	.80	.61	.19	.22	.48	-.26	.59
2	.91	.87	.04	.69	.69	.00	.53
3	.55	.50	.05	.23	.32	-.09	.62
4	.44	.38	.06	.19	.22	-.03	.54
5	.87	.71	.16	.39	.61	-.22	.60
6	.76	.55	.21	.13	.46	-.33	.62
7	.90	.75	.15	.44	.60	-.16	.57
Average (corrected for bias)							.558
True sample value							.630

Table 6. Item parameter and base-rate estimates for the seven items that remained in the fifteen weakly discriminative (mean difference in item parameters between .10 and .20) item male-female trial (N = 1150).

Another tough test of the method gives some idea about the limit of the method's detection power in terms of the strength of the indicators. Two diagnostic groups of male inpatients diagnosed as schizophrenic and male inpatients with other psychiatric diagnoses were used. Eight items were found that discriminated between the two groups by .10 to .20 and one by .26. All of the items were eliminated except two for which the item plus-rates were estimated to within .10 points. From these two items the base-rate of the schizophrenics was estimated to be .35. The true sample value was .28 but it should be noted that some of those diagnosed as not schizophrenic are (presumably) schizophrenic, whereas the other diagnosis error, where a truly non-schizophrenic is diagnosed as schizophrenic probably does not occur nearly as often. Thus the .28 value is probably too low for the proper taxonomic base-rate and the estimated base-rate is larger than the schizophrenic base-rate as it should be.

IV. Discussion

From the empirical trials it appears that for parameters to be estimated within .05 on the average it is required that there be a small number (say, five to ten) items that are quasi-independent within the taxon and the non-taxon class and that discriminate by .20 or more. Also, the sample size must be five hundred or larger depending on the portion of the initial item pool which is discriminative.

Five quasi-independent and discriminative items allow for accurate enough classification of individuals. It can be shown numerically that five independent items that discriminate by .20 will produce a misclassification rate via Bayes' Rule which is only .20 for equal base-rates. For five .30 difference items the misclassification rate is only .10 and for five .40 difference items it is just .05. A few powerful quasi-independent items then can [work] better than a larger batch of weak ones. This should perhaps not be surprising, despite the opposite leaning in traditional psychometrics, when one considers how powerful diagnostic syndromes have been employed in medicine, including psychiatry. Physicians have for centuries found it more useful to identify a few "strong signs" of each nosological entity than to "summate scores" of numerous feeble ones. With these results it is not alarming that the method retained seven or fewer items for the estimation of the parameters and the classification of the individuals in each of the reported empirical trials.

An artificial data trial shows that even if items are correlated within the taxon and the non-taxon class as high as .25, which is probably quite high for any taxon detectable with MMPI items, then parameters are still estimated accurately. The data were generated by the method given in Golden et al. (1974b) such that the 10 items discriminated by .50 and the base-rate was .500, the sample size being 1000. The item parameters were estimated to within .05 on the average and the base-rate was estimated perfectly to the third digit, .5003 (see Table 7). Of course, this trial illustrates that there is robustness with respect to the independence assumption only for strongly discriminative items and more extensive Monte Carlo study is required. But it is interesting to note that a preliminary trial in the attempted detection of the schizotype taxon (Meehl, 1962) indicates that there are a dozen or more items that discriminate by nearly .50 or more. Pending further Monte Carlo study of the method's robustness with respect to the assumptions, mainly that of independence,

it would appear that there probably is adequate robustness for the usual situation in psychopathology taxometrics.

Item	taxon			non-taxon			estimate of base-rate
	estimate	true value	error	estimate	true value	error	
1	.75	.75	.00	.21	.25	-.04	.49
2	.78	.75	.03	.17	.25	-.08	.51
3	.78	.75	.03	.25	.25	.00	.49
4	.75	.75	.00	.16	.25	-.09	.54
5	.78	.75	.03	.21	.25	-.04	.48
6	.81	.75	.06	.26	.25	.01	.49
7	.74	.75	-.01	.21	.25	-.04	.53
8	.97	.75	.22	.24	.25	-.01	.49
Average (corrected for bias)							.5003
True sample value							.500

Table 7. Item parameter and base-rate estimates for the eight items that remained in the artificial data trial where items were highly correlated ($r = .25$) within the taxon and the non-taxon class ($N = 1000$)

It might be mentioned here that the reason an item is removed from the input key when in the role of an output indicator is due to the requirement of independence between the output and input indicators. If an item were included in the input key when there are as few as, say, five items remaining this correlation could become substantial. Of course, the increase in programming complexity and the amount of calculation is enormous.

It is hoped that when there is not adequate robustness for accurate enough estimation that all of the items are deleted. If the Monte Carlo study shows this not to be true in certain situations then the method will require modification probably in the form of adding more consistency tests.

There are two key ideas to the present form of the method: (1) the \bar{y}_d curve has a maximum near the hitmax cut when there is a low enough correlation and (2) the input key, while possibly consisting of highly dependent items still discriminates adequately between the taxon and the non-taxon class and can be used to begin a bootstrapping process. Although it has not been rigorously shown that (1) is true it has been

investigated somewhat by Monte Carlo study and the result appears to always hold when correlations are low enough. When the correlations become too high (about .50) the method dramatically fails in producing a \bar{y}_d curve that has a local maximum in that the curve is concave upward. The complete results of the study will be given in a forthcoming report in this series. It could be that the consistency hurdles method is otherwise robust to the strong assumption of independence and that the appearance of a well-defined maximum of the curve is too tough a requirement which eliminates many good items. Fortunately, the more discriminative items will have better defined maxima than the less discriminative ones for the same amount of dependence; thus it is more likely that an item rejection mistake by the \bar{y}_d maximum requirement be made on the weaker items.

When the initial item pool contains only 10 to 30% items that discriminate by .20 or more then the latent distributions on the input key are not very well separated and the \bar{y}_d maximum is not very well-defined as when the distribution separation is, say, two or more intra-taxon sigma units between the latent means. Such a faintly defined maximum is only reliably detectable with a large enough sample. Thus the lower the portion of discriminative items in the initial pool, the larger the sample required. The male-female trial described above showed that when 27 of 75 Items were discriminative by .20 or more then a sample size of at least 1000 was required. Under such considerations it would certainly seem advantageous to smooth the \bar{y}_d curve. Scarborough (1962) gives the rationale for smoothing curves so that each observation is weighted by the inverse of the probable error of measurement. That is, it is necessary to have an estimate of the sampling error of $\bar{y}_d(x)$ at each value of x . Since $\bar{y}_d(x) = \bar{y}_a(x) - \bar{y}_b(x)$,

$$\begin{aligned} \text{var } \bar{y}_d(x) &= \text{var } \bar{y}_a(x) + \text{var } \bar{y}_b(x) \\ &= \frac{\bar{y}_a(1-\bar{y}_a)}{N_a} + \frac{\bar{y}_b(1-\bar{y}_b)}{N_b} \end{aligned}$$

and we have

$$SD(\bar{y}_d(x)) = \sqrt{\frac{\bar{y}_a(1-\bar{y}_a)}{N_a} + \frac{\bar{y}_b(1-\bar{y}_b)}{N_b}} .$$

From this formula, it is seen that the sampling variance is larger in the extremes of the distribution. Thus, it is best to have the hitmax near the middle of the distribution and, fortunately, this should nearly always be the case. Examples of smoothed \bar{y}_d curves are given in Figure 1. The method of moving averages (Kendall, 1951), which makes use of interlaced least-squares fitted polynomials (degree three was used in the examples), was used with a weighting function $1/SD(\bar{y}_d(x))$ according to a method developed by Forsythe (1957).

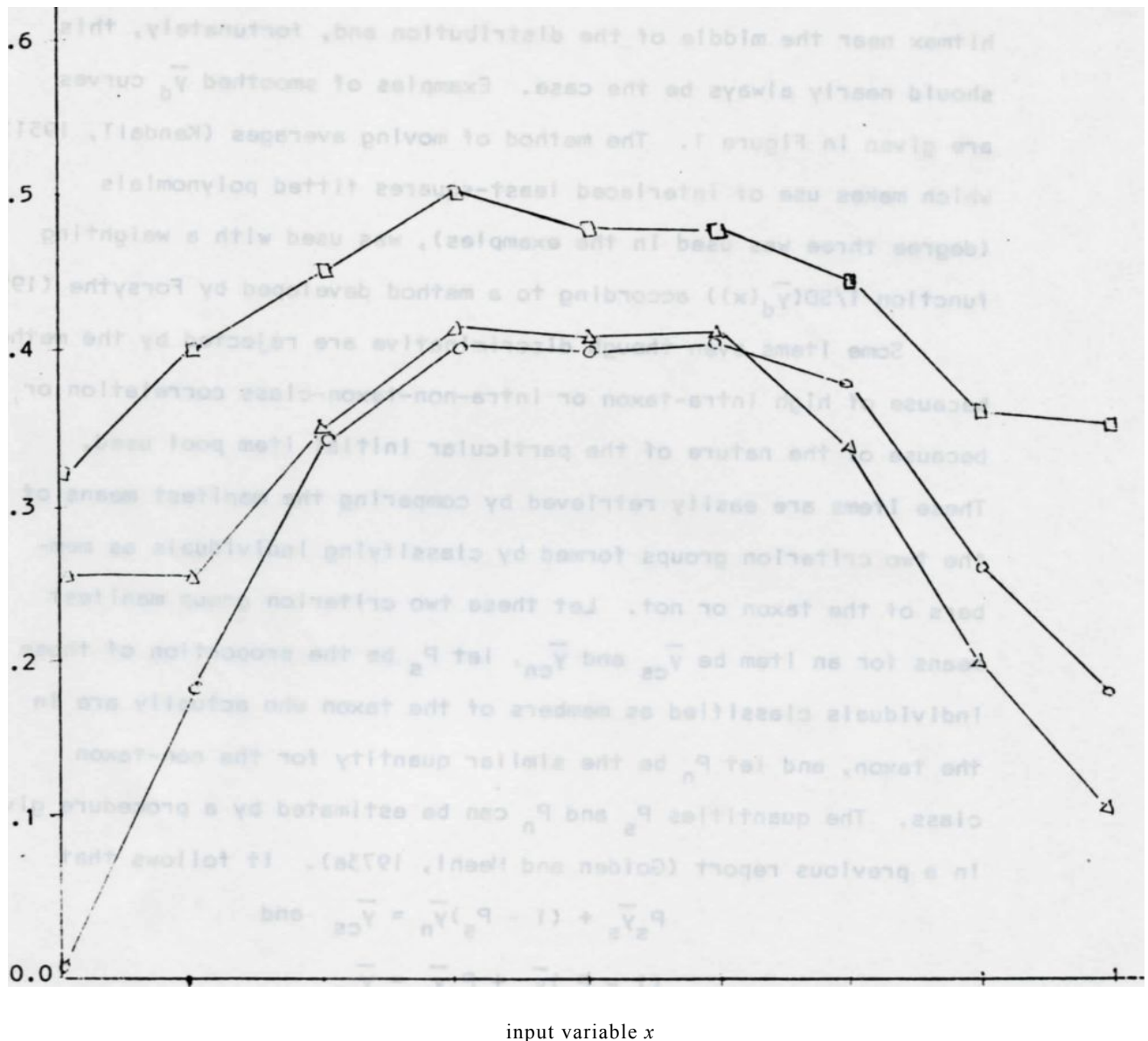


Figure 1. Typical $\bar{y}_d(x)$ curves for three moderately discriminative items in the 75 item male-female example ($N = 1150$). The curves were smoothed by the moving averages method. Y-axis is mean above less the mean below on output variable ($\bar{y}_d(x)$)

Some items even though discriminative are rejected by the method because of high intra-taxon or intra-non-taxon-class correlation or because of the nature of the particular initial item pool used. These items are easily retrieved by comparing the manifest means of the two criterion groups formed by classifying individuals as members of the taxon or not. Let these two criterion group manifest means for an item be \bar{y}_{cs} and \bar{y}_{cn} , let P_s be the proportion of those individuals classified as members of the taxon who actually are in the taxon, and let P_n be the similar quantity for the non-taxon class. The quantities P_s and P_n can be estimated by a procedure given in a previous report (Golden and Meehl, 1973a). It follows that

$$P_s \bar{y}_s + (1 - P_s) \bar{y}_n = \bar{y}_{cs} \quad \text{and}$$

$$(1 - P_n) \bar{y}_s + P_n \bar{y}_n = \bar{y}_{cn}$$

where \bar{y}_s and \bar{y}_n are latent taxon and non-taxon class means and are the two unknowns in this pair of simultaneous equations. The above pair of equations can be applied to all the items in the initial pool and the validity as measured by, say $\bar{y}_s - \bar{y}_n$, can be determined. The recaptured discriminative items and the originally retained items can be used in the consistency hurdles method again for possibly improving parameter estimates, development of discriminative keys, further confirmation of the solution obtained with the original item “seed” and in identifying and studying the nature of the taxon.

Since the method normally ends up with ten or less items, the recapture procedure allows for the possibility of selecting enough items to develop keys. These keys can be used with the maximum-covariance theory (Golden & Meehl, 1973a) and the normal theory (Golden et al., 1974) for further corroboration tests of the detected taxon as these other taxometric methods are based on different assumptions.

The manifest joint proportions for the sample should be approximately equal to those calculated from the estimates of the base-rate and the latent item means. For example, for three items x , y and z it can be shown that

$$PP_{sx}P_{sy}P_{sz} + (1 - P)P_{nx}P_{ny}P_{nz} = P_{xyz}$$

$$PP_{sx}P_{sy} + (1 - P)P_{nx}P_{ny} = P_{xy}$$

$$PP_{sx}P_{sz} + (1 - P)P_{nx}P_{nz} = P_{xz}$$

$$PP_{sy}P_{sz} + (1 - P)P_{ny}P_{nz} = P_{yz}$$

under the condition of independence within taxon and non-taxon class.

he allowable differences between the observed and the theoretical values of the joint proportions can be derived by expressing the exact differential of the difference in terms of the partial derivatives of the parameters as was done above to obtain $\Delta\varepsilon$. Thus we have a final consistency test for those items retained by the method.

V. A Method for the Selection of Discriminative indicators

Probably the method can be most improved by developing a method of initial item selection so that the concentration of discriminative items is at least as high as, say, 50%. One such method results from the fact that for an item to be discriminative it is necessary that the mean of the item when used as an output variable be a monotonically increasing function of the input variable value. In the 75 male-female item set, most of the discriminative items can be distinguished from the rest by simply observing the presence or absence of the monotonicity by eye (see Figure 2). However generally a more sophisticated analytical method is required; such a method is developed below.

Consider a set of candidate discriminative indicators, $x_i, i = 1, 2, 3, \dots, n$, each dichotomous. As before, let one of the indicators be considered as an output variable y while the sum of the remaining $n - 1$ indicators is the input variable $x = (\sum x_i) - y$. Assume that the mean of the taxon members with score x is given by

$$\bar{y}_s(x) = m_s x + b_s \quad (A_1)$$

and that for the non-taxon class members by

$$\bar{y}_n(x) = m_n x + b_n \quad (A_1)$$

Thus $\bar{y}_s(x) - \bar{y}_n(x) = (m_s - m_n)x + b_s - b_n > 0$ for all x if an item is discriminative. The problem is, then, given an observable function of the compound distribution such as the compound mean $\bar{y}(x)$, to estimate the values of the latent parameters m_s, m_n, b_s , and b_n under assumption A_1 . It is a simple matter to observe that

$$\lim_{x \rightarrow x_{\max}} \bar{y}(x) = m_s n + b_s$$

$$\lim_{x \rightarrow x_{\min}} \bar{y}(x) = b_n$$

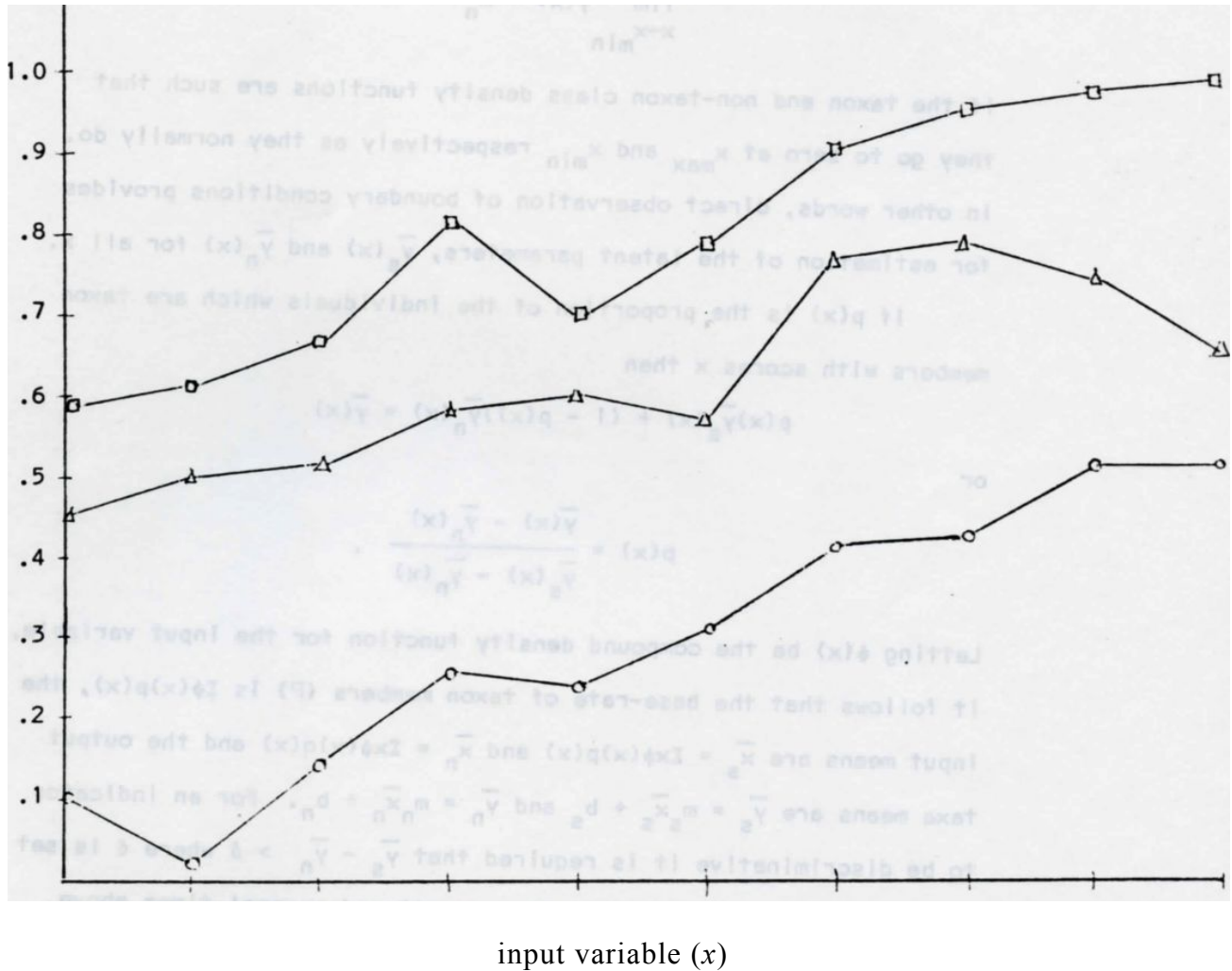


Figure 2. Typical monotonically increasing $\bar{y}(x)$ curves for three discriminative items in the 75 item male-female example ($N = 1150$). The curves were not smoothed. Y-axis is the compound manifest mean of output variable ($\bar{y}(x)$)

and that

$$\lim_{x \rightarrow x_{\max}} \bar{y}(x) = m_s \quad \text{and}$$

$$\lim_{x \rightarrow x_{\min}} \bar{y}(x) = m_n$$

if the taxon and non-taxon class density functions are such that they go to zero at x_{\max} and x_{\min} respectively as they normally do. In other words, direct observation of boundary conditions provides for estimation of the latent parameters, \bar{y}_{sr} and \bar{y}_{nr} for all x .

If $p(x)$ is the proportion of the individuals which are taxon members with scores x then

$$p(x)\bar{y}_s(x) + (1 - p(x))\bar{y}_n(x) = \bar{y}(x)$$

or

$$p(x) = \frac{\bar{y}(x) - \bar{y}_n(x)}{\bar{y}_s(x) - \bar{y}_n(x)} .$$

Letting $\phi(x)$ be the compound density function for the input variable, it follows that the base-rate of taxon members (P) is $\Sigma\phi(x)p(x)$, the input means are $\bar{x}_s = \Sigma x\phi(x)p(x)$ and $\bar{x}_n = \Sigma x\phi(x)q(x)$ and the output taxa means are $\bar{y}_s = m_s\bar{x}_s + b_s$ and $\bar{y}_n = m_n\bar{x}_n + b_n$. For an indicator to be discriminative it is required that $\bar{y}_s - \bar{y}_n > \delta$ where δ is set to be adequately large, say, .20 as mentioned several times above.

It should turn out that $P\bar{x}_s + Q\bar{x}_n \doteq \bar{x}$ where \bar{x} is the observable compound mean on x . Presumably, errors in the estimation of b_s , b_n , m_s and m_n would cause errors in the $p(x)$ which would cause an error in P and/or \bar{x}_s and \bar{x}_n thereby causing the above condition not to obtain. Letting $\Delta\bar{x}_s$, $\Delta\bar{x}_n$ and ΔP denote errors in the parameter estimates we see that since

$$d\bar{x} = \frac{\partial\bar{x}}{\partial\bar{x}_s}d\bar{x}_s + \frac{\partial\bar{x}}{\partial\bar{x}_n}d\bar{x}_n + \frac{\partial\bar{x}}{\partial p}dP$$

we have

$$\Delta\bar{x} \doteq P\Delta\bar{x}_s + Q\Delta\bar{x}_n + (\bar{x}_s - \bar{x}_n)\Delta P .$$

It remains to be seen what reasonable values for $\Delta\bar{x}_s$, $\Delta\bar{x}_n$ and ΔP are. It is hoped that the method produces a discrepant estimate of the manifest compound input mean for those items which are incorrectly estimated to have values of $\bar{y}_s - \bar{y}_n$ larger than δ .

The item selection method thus relaxes the independence assumption in the consistency hurdles theory to one of linear regression of the output item on the input scale within the taxon and the non-taxon class. For the method to work well it is necessary that the parameters of the linear regression functions be estimated from the tails of the compound distribution on the input scale. As when the tails are used in the

consistency hurdles method there are the problems of impurity in the tail and sample size in the tail. Unfortunately these two problems permit solution only at the expense of the other. It is hoped that Monte Carlo study and curve-smoothing techniques will provide for adequate simultaneous solution. Suffice it to say for now that when graphs of discriminative items are analyzed by eye with the aid of a ruler, the parameter estimates are not appreciably less accurate than those of the consistency hurdles method in the few trials run so far.

VI. Conclusions

The consistency hurdles method gave sufficiently encouraging results in several empirical trials to indicate that further study is justifiable.

The method is designed to meet two major requirements. First, to eliminate as early in the calculations as possible those indicators that are not behaving in a way that is consistent with both the assumptions of the theory and relations satisfied only if there is adequate discrimination. This requirement was seen as important since other taxometric methods generally fail to work when there are many non-discriminative indicators in the pool. Evidently, invalid indicators do not just get ignored by these methods but they cause the methods to malfunction. In psychopathology measurement, it is true that there are typically at least fifty per cent invalid indicators since apparently there are not adequate indicator selection methods. Second, [it is important] to choose indicators that behave consistently with assumptions so that they can be used to accurately estimate the latent parameters.

It always is possible that we end up with valid and consistent indicators but for the wrong taxonomy. The method itself can never tell us when this happens. For example, the items of the social introversion scale (S_i) of the MMPI were intended to be used as an empirical trial of the method in that there is no good reason to postulate an underlying taxon. While the desired result was that all (or nearly all) of the items would be rejected, this was not the case; eleven items were retained. However, it was fairly clear the taxon detected was schizoidia, not introversion, as the base-rate and item parameter estimates and individual classification were very similar to those obtained in another study which was explicitly designed to detect the schizoid taxon (Golden et al., 1974a).

It is hoped that Monte Carlo study will show that there are only 0, 1 or 2 indicators left out of an initial pool if and only if (a) there is not a dichotomous taxonomy or (b) there are not more than two of the indicators which are both adequately discriminative and in sufficient agreement with the assumptions of the method. In any event such a Monte Carlo study is required and it may indicate revision of the method.

In the Monte Carlo study it will be interesting to study the effects of

- (a) the within-taxon and non-taxon class correlations,
- (b) differences between the taxon and the non-taxon correlation matrices,
- (c) the taxon base-rate,
- (d) the proportion of the original pool of candidate items which are discriminative, and
- (e) the number of taxa.

The main desired result is that whenever items are found to behave consistently and are used to estimate parameters in the final step, the resulting estimates are accurate enough. That is, estimates are never produced which are too erroneous. In the event erroneous estimates are produced for any sort of artificial data it will be indicated that other consistency tests are required. Thus, the first desired result is that taxometric detection is never spurious and is accurate. The second desired result is that the method proves to be powerful enough in the detection of taxonomies.

The present empirical trials do serve to indicate that the detection power of the method is its most attractive demonstrated feature.

Finally, empirical trials should be used to test the power of the method for the avoidance of spurious taxonomic detection. Two possible examples of empirical non-taxonomic variables which could be used for such trials are social introversion among normals and general intelligence among normals.

REFERENCES

- Forsythe, G. D. (1957). Generation and use of orthogonal polynomials for data fitting with a digital computer. *Journal of the Society of Industrial Applied Mathematics*, 5, 74-88.
- Golden, R. R. & Meehl, P. E. (1973a). *Detecting latent clinical taxa, IV: An empirical study of the maximum covariance method and the normal minimum chi-square method using three MMPI keys to identify the sexes*. Reports from the Research Laboratories of the Department of Psychiatry, University

- of Minnesota, Report No. PR-73-2.
- Golden, R. R. & Meehl, P. E. (1973b). *Detecting latent clinical taxa, V: A Monte Carlo study of the maximum covariance method and associated consistency tests*. Reports from the Research Laboratories of the Department of Psychiatry, University of Minnesota, Report No. PR-73-3.
- Golden, R. R. & Meehl, P. E. (1974a ~~in press~~). *Detecting latent clinical taxa, VIII: A preliminary study in the detection of the schizoid taxon using MMPI items as indicators*. Reports from the Research Laboratories of the Department of Psychiatry, University of Minnesota, [Report No. PR-74-6].
- Golden, R. R., Meehl, P. E. & Tyan, S. H. (1974b ~~in press~~). *Detecting latent clinical taxa, IX: A Monte Carlo method for testing taxometric theories*. Reports from the Research Laboratories of the Department of Psychiatry, University of Minnesota, [Report No. PR-74-7].
- Kendall, M. G. (1951). *The advanced theory of statistics*. Vol II. (3rd ed.) London: Butler and Tanner Limited.
- Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, 17, 827-838.
- Meehl, P. E. (1965). *Detecting latent clinical taxa by fallible quantitative indicators lacking an accepted criterion*. Reports from the Research Laboratories of the Department of Psychiatry, University of Minnesota, Report No. PR-65-2.
- Meehl, P. E. (1968). *Detecting latent clinical taxa, II: A simplified procedure, some additional hitmax cut locators, a single-indicator method, and miscellaneous theorems*. Reports from the Research Laboratories of the Department of Psychiatry, University of Minnesota, Report No. PR-68-4.
- Scarborough, J. B. (1962). *Numerical mathematical analysis*. (5th ed.) Baltimore: John Hopkins Press.
- Stephenson, W. (1938). The inverted factor technique. *British Journal of Psychology*, 26, 344-361.
- Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.