Reports from the Research Laboratories

of the

Department of Psychiatry

University of Minnesota

Detecting Latent Clinical Taxa, VII:

Maximum Likelihood Solution and Empirical and Artificial Data Trials

of the Multi-indicator Multi-taxonomic Class Normal Theory[1]

Robert R. Golden, Shirley H. Tyan

and

Paul E. Meehl

Report No. PR-74-5 October 1974

[1] This Research was supported in part by grants from the Psychiatry Research Fund and the National Institute of Mental Health, Grant Number MH 24224

pdf by ljy October 2006

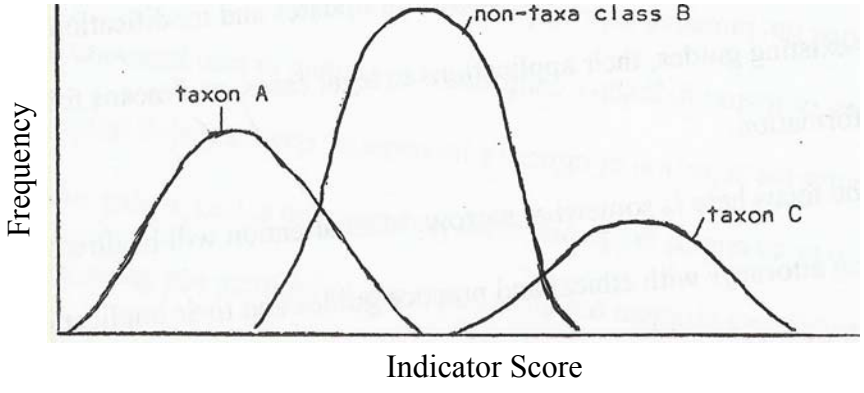
TABLE OF CONTENTS

I. Introduction

In a previous research report in the present series, the minimum chi-square single indicator normal theory's parameter estimation accuracy and taxon detection power in an empirical trial were both sufficient to encourage further study of the theory (Golden and Meehl, 1973a). It was also shown by analytical development in that report that the indicator latent distributions within the taxon and the nontaxon class on keys are quasi-normal, which is the sole assumption upon which most of the normal taxometric theory rests, if they are sums of dichotomous items which are pairwise slightly correlated within the taxon and the nontaxon class. The only shortcoming of the minimum chi-square calculation method (see Meehl et al., 1969) is that it is very time consuming and expensive even on a high-speed computer. The maximum likelihood solution by Hasselblad (1966), besides providing for much quicker calculation, is not restricted to one taxon plus the nontaxon class, is amenable to a multi-indicator generalization, as shown in this report, and uses estimators that are known to have optimal properties. In the present report it is shown that results of several artificial and empirical data trials of the generalized maximum likelihood method are sufficiently encouraging to warrant further study of the method.

The present taxometric theory, as others in previous reports is intended for the detection of 'real' empirical classes. When all the members of such an empirical class are considered to have the same etiology, such as the presence of a mutated gene, a germ

or a neural defect, whereas all non-members are considered not to have this etiology, then the class will be called a ‘taxon’ and the complement of the class will be called the ‘nontaxon class’ or the ‘extra-taxon class.’ Previously developed theories in this series have allowed for only a single taxon and a single nontaxon class, the major reason being that work on the present methodological problem resulted from interest in testing a substantive theory concerning schizotypes and nonschizotypes (Meehl, 1962; 1965; 1973). Generally however, a taxonomy will consist of more than one taxon and, possibly, more than one nontaxa class. The present theory allows for this possibility under the assumption that the indicator distribution is quasi-normal within each taxon and within each nontaxa class as in the example in the figure below. It should be noted that by ‘nontaxon A class’ we would be referring to the union of classes B and C but by ‘nontaxa class’ we refer to B. The point is that in the present report it is continually necessary to refer to two kinds of empirical classes, in general, the taxon and the nontaxa class, and for this purpose we will use the term ‘taxonomic class.’ In the example below each of A, B and C is a taxonomic class, whereas the union of any of them pairwise or higher order is not (by definition).



The theory allows for several different taxonomic class orderings. For example, consider the possibilities when there are three taxonomic classes. If one were using the amount of blood sugar as an indicator then a diabetes taxon *might* be detected on the high end and a hypoglycemia taxon on the low end, the middle taxonomic class being normals. In psychopathology this taxon, nontaxa class, taxon pattern would potentially be likely when measures of bi-polar personality traits are used, such as extroversion-introversion, where each direction of extreme deviation is indicative of psychopathology.

The case of two contiguous taxa might turn out to be illustrated by the normal-neurotic-psychotic example where a number of different psychopathology measures possibly could be used.

While it it intended that the term 'taxonomic class' is to refer to a real rather than fictional entity, the fact that it has proved difficult to demonstrate the existence of such taxonomic classes in psychopathology is illustrated by our need to resort to hypothesized examples above. Because of this difficult state of affairs, it is useful to remember that taxometric theory can only be required to detect the existence of a taxonomy and not necessarily to establish its essential nature. In the first stage of detecting existence, the theory can produce erroneous results of two kinds. First the theory may not be powerful enough to detect a truly existent underlying taxonomy. Second, and probably more important, the theory can produce spurious detections. The likelihood of this kind of error can be decreased substantially by using well-tested consistency tests (Meehl, 1965; 1973; Golden and Meehl, 1973b) which check the adequacy of the degree-of-fit of the theory. Several consistency tests are developed in section X.

II. The Maximum Likelihood Solution

An outline of the calculations of the method is give below; the interested reader is referred to the original article by Hasselblad (1966), as further analytical development is provided there.

Suppose that there are n taxonomic classes denoted by the subscript j with distributions on an indicator x and the taxonomic class means, variances, and base-rates denoted by μ_j, σ_j and p_j ($\sum_{j=1}^{n} p_j = 1$). Let x be divided into N intervals denoted by the subscript i such that the interval width is small compared to σ_j. Then let q_{ij} be the density of the jth taxon in the ith interval and let this be approximated by

$$q_{ij} = \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\{-(x_i - \mu_j)^2/2\sigma_j^2\};$$

also, let Q_i be the compound density for the ith interval or

$$Q_i = \sum_{j=1}^{n} q_{ij}\, p_j .$$

The only values that are known are the compound sample distribution interval frequencies f_i, $i = 1, 2, 3, \ldots, N$. Hasselblad shows that the maximum likelihood estimates of the unknown latent parameters can be found by the steepest descent iterative procedure which results in the following equations:

$$\mu_j = \sum_{i=1}^{N} (f_i/Q_i)\, q_{ij}\, x_i \,/\, \left(\sum_{i=1}^{N} (f_i/Q_i)\, q_{ij}\right),$$

$$\sigma_j^2 = \sum_{i=1}^{N} (f_i/Q_i)\, (x - \mu_j)^2 \,/\, \left(\sum_{i=1}^{N} (f_i/Q_i)\, q_{ij}\right), \text{ and}$$

$$p_j = \sum_{i=1}^{N} (f_i \, q_{ij} \, P_j / Q_i) / (\sum_{i=1}^{N} f_i) \, .$$

The iterative procedure begins with initial guesses of μ_j, σ_j and p_j. It is not shown how accurate the initial guesses must be or if, indeed, convergence to the true values will necessarily obtain. Scarborough (1962) shows that the method converges whenever the following condition is met: Let the true value of a root x satisfy the equation $x = \phi(x, y, z, \ldots)$; then if $\left|\frac{\partial \phi}{\partial x}\right| < 1$ in a neighborhood of x which contains the successive approximations of x, then the approximations converge to the true value of x.

III. Artificial and Empirical Data Trials

The calculation scheme was applied to a variety of artificial and real data samples. First, three MMPI keys developed to discriminate between the sexes were analyzed by the method. The results are given in Table 1. The method gave very accurate parameter estimates on the second and third keys after just 100 iterations, at which point the estimates demonstrated strong convergence. However, for the first key, the base-rate estimate is markedly in error, especially after the process had been continued until convergence was apparent. It is difficult to explain this singular error, as other methods have estimated the base-rate equally well for the three keys. The exact significance of the larger number of iterations required for apparent convergence is not known, but the result is illustrative of a general finding that if several hundred iterations are required for convergence then the results should not be trusted and will fail further consistency tests. The chi-square goodness-of-fit values (by comparing the estimated and the observed

Table 1. Examples of empirical data trials of the maximum-likelihood single-indicator method using MMPI keys to identify the sexes.

	P_1	μ_1	σ_1	P_2	μ_2	σ_2	χ^2 *
First Key (N = 1105)							
Initial guess	.500	6.00	3.00	.500	13.00	3.00	
No. of interations							
50	.417	9.45	2.50	.582	14.40	2.16	14.89
100	.449	9.66	2.58	.551	14.51	2.12	14.58
200	.483	9.88	2.67	.516	14.64	2.07	14.37
300	.499	9.98	2.71	.501	14.69	2.05	14.33
True sample value	.389	9.57	2.55	.611	14.10	2.56	
Error	.110	.41	.16	–.110	.59	–.51	
Second Key (N = 1105)							
Initial guess	.432	7.36	2.37	.568	11.97	2.26	
No. of iterations							
10	.428	7.33	2.36	.572	11.95	2.27	13.34
50	.411	7.24	2.32	.589	11.89	2.29	13.28
100	.368	6.98	2.23	.632	11.72	2.35	13.18
True sample value	.389	7.31	2.42	.611	11.68	2.47	
Error	–.021	–.33	–.19	.021	.04	–.12	
Third Key (N = 1105)							
Initial guess	.500	6.00	3.00	.500	13.00	3.00	
No. of interations							
10	.418	8.64	2.54	.582	13.02	2.42	15.25
50	.418	8.59	2.48	.582	13.06	2.38	14.92
100	.416	8.58	2.48	.584	13.06	2.38	14.92
200	.409	8.54	2.46	.591	13.03	2.40	14.91
True sample value	.389	8.60	2.57	.611	12.84	2.53	
Error	.020	–.06	–.11	–.020	.19	–.13	

* $\chi^2_{.05} = 26.30$

compound sample frequency distributions) are similar in magnitude and do not approach significance for the three keys. It is seen then that non-significant χ^2 values do *not* guarantee accurate parameter estimates. The initial guesses were obtained by a method which makes use of probability paper and which will be discussed below (Harding, 1949).

Artificial data were generated by means of a normal random number generator (see p. 3ff of Golden and Meehl, 1973b for the method). Intervals were formed simply by rounding each generated number to its greatest integer part; parameter values and sample sizes were such that this method produced an acceptable interval coarseness. A perusal of Table 2 shows that the method produced accurate parameter estimates with 100 iterations when the mean separations are as small as one within taxonomic class sigma-unit and base-rates as disproportionate as .1 and .9. While none of the chi-square values approach significance, the goodness-of-fit of the theory is evidence[d] mainly by the acceptable accuracy of the parameter estimates.

The results of the method when the initial estimates are poor is illustrated by an example given in Table 3. It can be seen that initial guesses can be off quite a bit without apparently affecting the accuracy after a large enough number of iterations. However, several hundred iterations were required for convergence to be apparent even though accuracy was sufficient after 100 iterations. Several extrapolation methods have been tried to increase the speed of convergence but without much success. At present 100 to 200 iterations appear to be sufficient for adequate estimation accuracy when initial guesses are not grossly in error, even though a much larger number is required for convergence to be easily discernible.

It is clear that intervals must be such that the frequencies for each are large enough to avoid excessive sampling error. One procedure is to combine pairs of contiguous intervals until (a) the interval frequencies are apparently large enough to avoid excessive sampling error and (b) the compound frequency curve is always monotonically increasing or decreasing on each side of the local maxima and minima, the number of each being less than or equal to the postulated number of taxonomic classes. However, it has been found that for compound samples of a few hundred in size, taxonomic class mean separations of two within-taxonomic class sigma units and equal base-rates, which are about optimal conditions in psychopathology measurement, the base-rate estimate can be in error by as much as .10, for example. The required interval frequency size and a safe method of interval construction should be determined more carefully by Monte Carlo study.

The maximum likelihood method has several advantages over the minimum chi-square method, mainly generalizability and much fewer, although still numerous, calculations. Both methods apparently can produce multiple solutions but the existence of several solutions is more easily discovered with the minimum chi-square method. Whether this is an advantage or not depends on whether or not the different solutions correspond to real but different taxonomies. Suffice it to say that in the present work there is no evidence to indicate that possible existence of several taxonomies underlying a properly selected set of indictors is a matter of real concern although, of course, this may not always be the case. In this connection, it should be noted that when the solution is unique then it has

Table 2. Examples of artificial data trials of the maximum-likelihood single-indicator method.

	P_1	μ_1	σ_1	P_2	μ_2	σ_2	χ^2 *
Example 1							
Initial guess	.500	7.43	2.64	.500	13.50	1.91	
Iteration estimate**	.402	7.76	2.02	.598	14.02	2.39	16.90
True value	.429	8.05	2.06	.571	14.22	2.18	
Error	–.027	–.29	–.04	.027	–.20	.21	
Example 2 (N = 1000)							
Initial guess	.500	7.42	2.56	.500	15.00	2.31	
Iteration estimate**	.344	8.70	1.92	.656	12.73	2.16	9.64
True value	.402	8.85	2.06	.598	13.09	2.08	
Error	–.058	–.15	–.14	.058	–.36	.08	
Example 3 (N = 1000)							
Initial guess	.350	8.00	2.23	.650	13.00	2.26	
Iteration estimate**	.327	8.56	2.29	.673	14.11	2.29	18.68
True value	.310	8.05	2.06	.690	14.22	2.18	
Error	.017	.51	.23	–.017	–.11	.11	
Example 4 (N = 1000)							
Initial guess	.200	7.00	1.73	.800	14.00	2.00	
Iteration estimate**	.256	7.76	2.06	.744	11.78	1.85	18.80
True value	.300	8.00	2.00	.700	12.00	1.73	
Error	–.044	–.33	.06	.044	–.22	.12	
Example 5 (N = 1000)							
Initial guess	.500	8.01	2.66	.500	13.42	1.73	
Iteration estimate**	.219	8.23	2.38	.781	14.34	2.13	15.74
True value	.208	8.05	2.06	.792	14.22	2.18	
Error	.011	.18	.32	–.011	.12	.05	
Example 6 (N = 1000)							
Initial guess	.106	10.11	1.89	.894	11.15	2.16	
Iteration estimate**	.110	10.44	2.01	.889	11.11	2.16	9.20
True value	.100	9.48	1.86	.900	11.13	2.02	
Error	.010	.96	.15	–.011	–.02	.14	
Example 7 (N = 1000)							
Initial guess	.500	7.42	2.65	.500	13.50	1.91	
Iteration estimate**	.128	8.54	2.12	.872	14.31	2.17	12.22
True value	.117	8.05	2.06	.883	14.22	2.18	
Error	.011	.49	.06	–.011	.09	.01	

* $\chi^2_{.05} = 26.30$ ** All estimates are after 100 iterations

Table 3. Example of artificial data trial of the maximum-likelihood single-indicator method when guesses were poor (N=1000).

	P_1	μ_1	σ_1	P_2	μ_2	σ_2	χ^2 *
Initial guess	.450	4.575	1.17	.550	9.350	1.32	
Number of iterations							
50	.104	9.618	1.72	.896	11.205	2.14	8.93
100	.106	10.113	1.89	.894	11.150	2.16	9.08
150	.110	10.444	2.06	.890	11.114	2.16	9.21
200	.115	10.871	2.12	.885	11.063	2.16	9.29
250	.115	10.964	2.15	.885	11.051	2.16	9.29
300	.116	11.007	2.16	.884	11.045	2.16	9.29
350	.116	11.027	2.16	.884	11.045	2.16	9.29
400	.116	11.027	2.16	.884	11.045	2.16	9.29
450	.116	11.038	2.16	.884	11.041	2.16	9.29
500	.116	11.039	2.16	.884	11.041	2.16	9.29
550	.116	11.041	2.16	.884	11.041	2.16	9.29
600	.116	11.041	2.16	.884	11.041	2.16	9.29
650	.116	11.041	2.16	.884	11.041	2.16	9.29
True sample value	.100	9.020	1.90	.900	11.253	2.04	
Error	.016	2.021	.26	–.016	–.212	.12	

* $\chi^2_{.05} = 26.30$

been analytically shown that both maximum likelihood and minimum chi-square methods produce the same result for large enough samples (Cramé, 1946).

IV. Use of Probability Paper for Initial Guesses

The initial guesses of the parameters can be obtained by a procedure described by Harding (1949). The method makes use of probability paper and the assumption that the taxonomic class frequency distributions are normal. Generally, when the compound frequency distribution is polymodal and the sample size very large, the method is sufficiently accurate by itself. Since the procedure is based on the same assumption of normality within taxonomic classes it lends itself well to the problem of making initial guesses for the maximum likelihood method.

Two dozen cumulative density functions resulting from composites of various numbers of artificial normal frequency distributions, for which the compound sample sizes were 1000, were analyzed without knowledge of the true component parameters by using the probability paper method. The following general conclusions were formed:

a) when the separation between component means is about two intra-component sigma units then the parameters can be estimated with at least 90% accuracy,
b) two components are distinguishable from one component when the separation in the means is above one intra-component sigma unit,
c) when the estimates are such that the estimated mean separation is less than two intra-component sigma units the results are not to be trusted, and

d) apparently, the estimates are always accurate enough to serve as initial guesses for the maximum likelihood method.

These conclusions are known to be true only when the within component population distributions are perfectly normal. In practice, distributions are not perfectly normal, of course; therefore, it remains to be determined if the probability paper method is sufficient for the unknown robustness of the maximum likelihood method with respect to the normality assumption. Pending such a study it would appear safe to use the method with the following stipulations:

a) the method is not more accurate than is indicated above for ideal conditions and

b) the method might be mainly useful in suggesting alternative latent situations which can be subjected to testing by the normal theory.

V. Multi-indicator Generalizations

When several indicators are analyzed singly they can produce discrepant estimates of the base-rates (some or all being erroneous) such as the three male-female keys did. Possibly the only way to rectify such a situation when singly analyzing indicators is to increase the sample size. Since more than one indicator is usually available, it follows that one is behooved to consider a simultaneous multi-indicator approach as presumably this lessens the sample size requirement. The maximum likelihood iterative scheme can easily be adapted to the multi-indicator situation by simultaneously estimating the parameters of

each indicator just as is done if analyzed singly except changing the estimate of the common base-rate parameter to be the average of the estimates produced by the previous iteration.

Analytical proof that such a method should correctly converge is not provided but an empirical trial of the method is encouraging in this respect. The three male-female keys used previously were analyzed and the results are given in Table 4. The results of single indicator analyses were used as initial guesses and convergence was apparent after 50 iterations. The results are remarkable in that all the parameter estimates are extremely accurate. Further encouragement is gained from the fact that the intra-taxa correlations between the indicators were each between .3 and .5; thus the method might not require a strong within taxonomic class independence condition to be met.

Also, the method was tried with an artificial data sample in which the three indicators were distributed multivariate normally within taxonomic class, all within-taxonomic class correlations being equal to .5. The resulting parameter estimates, given in Table 5, were also quite accurate.

It is intuitively clear that a multi-indicator method should be the more accurate of the two approaches, especially if all indicators are of about equal validity and are weakly correlated within taxonomic class. Also, it is true that some (most?) taxonomies require the use of more than one indicator for a complete specification. For such a taxonomy, using fewer than a complete set of indicators would lead to an incorrect result. One the other hand, it also is true that sometimes only a single candidate indicator is available.

Table 4. An empirical trial of the first multi-indicator generalization of the maximum-likelihood solutions using the three MMPI keys to identify the sexes ($N = 1105$).

	P_1	μ_1	σ_1	P_2	μ_2	σ_2	χ^2 *
First Key							
Initial guess	.499	9.98	2.71	.501	14.68	2.05	
Estimate after 50 iterations	.431	9.55	2.54	.569	14.45	2.15	14.70
True sample value	.389	9.57	2.55	.611	14.10	2.56	
Error	.042	–.02	–.01	–.042	.35	–.41	
Second Key							
Initial guess	.368	6.98	2.23	.632	11.72	2.35	
Estimate after 50 iterations	.431	7.34	2.36	.569	11.96	2.26	13.40
True sample value	.389	7.31	2.42	.611	11.68	2.47	
Error	.042	.03	–.07	–.042	.28	–.21	
Third Key							
Initial guess	.410	8.54	2.46	.590	13.03	2.39	
Estimate after 50 iterations	.431	8.66	2.51	.569	13.11	2.37	14.90
True sample value	.389	8.60	2.57	.611	12.84	2.53	
Error	.042	.06	.06	–.042	.27	–.16	

* $\chi^2_{.05} = 26.30$

Table 5. An artificial trial of the first multi-indicator generalization of the maximum likelihood solution using the three MMPI keys to identify the sexes ($N = 1000$).

	P_1	μ_1	σ_1	P_2	μ_2	σ_2	χ^2 *
First Key							
Initial guess	.31	9.90	2.15	.69	15.00	2.07	
Estimate after 200 iterations	.28	9.68	1.83	.72	14.89	2.03	6.30
True sample value	.30	10.04	2.12	.70	14.93	2.05	
Error	–.02	–.36	–.29	.02	–.04	–.02	
Second Key							
Initial guess	.31	12.12	3.27	.69	17.97	3.08	
Estimate after 200 iterations	.28	15.09	3.98	.72	17.30	3.39	17.90
True sample value	.30	13.77	3.27	.70	17.94	3.08	
Error	–.02	1.32	.71	.02	–.64	.31	
Third Key							
Initial guess	.31	11.55	1.52	.69	20.53	1.54	
Estimate after 200 iterations	.28	12.54	1.60	.72	16.86	1.61	6.20
True sample value	.30	12.60	1.53	.70	16.94	1.52	
Error	–.02	–.06	.07	.02	–.08	.09	

* $\chi^2_{.01} = 20.80$; $\chi^2_{.05} = 11.60$

Also, it is possible that some taxonomies are optimally detected when a certain single indicator is used alone, that indicator being the only one with the required taxon-specific variance. In this situation, the use of further indicators, even though each being adequately discriminative, may merely cause what Loevinger (1957) called "psychometric drift." One way this could happen is by violation of within taxa independence assumptions which are used in the solution below. In summary then, the comparison of the single and the multiple indicator methods will not just be a mathematical matter but also an empirical one.

The multi-indicator generalization of the maximum-likelihood development by Hasselblad can easily proceed under the assumption of intra-taxonomic class indicator independence. Let x_{ik} be the score of the ith individual on the kth indicator and $\tilde{x}_i$ be the vector of the ith individual scores on the p indicators ($x_{i1}, x_{i2}, \ldots, x_{ip}$). Also, let $\phi(\tilde{x}_i)$ denote the compound density and $\phi_j(\tilde{x}_i)$ denote the jth taxonomic class density where $j = 1, 2, 3, \ldots, m$ at the point $\tilde{x}_i$. Thus

$$\phi(\tilde{x}_i) = \sum_{j=1}^{m} p_j \, \phi_j(\bar{x}_i)$$

where p_j is the base-rate of the jth taxonomic class. The likelihood function is given by

$$L = \prod_{j=1}^{N} \phi(\tilde{x}_i) ,$$

and the maximum likelihood estimates of the latent parameters p_j, μ_{jk} and σ_{jk}, where k denotes the kth indicator, are obtained by solving the following set of $2mp + m - 1$ simultaneous equations:

$$\frac{\partial \log \mathrm{L}}{\partial p_j} = 0 \qquad j = 1, 2, 3, \ldots, m-1$$

$$\frac{\partial \log \mathrm{L}}{\partial \mu_{jk}} = 0 \qquad j = 1, 2, 3, \ldots, m \tag{1}$$

$$\frac{\partial \log \mathrm{L}}{\partial \sigma_{jk}} = 0 \qquad k = 1, 2, 3, \ldots, p$$

The following two assumptions make it possible to easily obtain solutions of (1):

(a) all indicators are independent within each taxonomic class; i.e.,

$$\phi_{\mathrm{j}}(\tilde{x}_i) = \prod_{k=1}^{p} \phi_{jk}(x_{ik}) \quad \text{and}$$

(b) the interval density is equal to the normal curve ordinate at the interval midpoint; i.e.,

$$\phi_{jk}(x_{ik}) = \frac{1}{\sqrt{2\pi}\,\sigma_{jk}} \exp\left\{-\frac{1}{2}\frac{\left(x_{ik}-\mu_{jk}\right)^2}{\sigma_{jk}^2}\right\} .$$

We now have

$$\log \mathrm{L} = \sum_{i=1}^{N} \log\{\phi(\tilde{x}_i)\}$$

$$= \sum_{i=1}^{N} \log\left\{\sum_{j=1}^{m} p_j\, \phi_j(\tilde{x}_i)\right\} ,$$

and it follows that

$$\frac{\partial \log \mathrm{L}}{\partial p_j} = \sum_{i=1}^{N}\left\{\frac{\phi_j(\tilde{x}_i) - \phi_m(\tilde{x}_i)}{\sum_{j=1}^{m} p_j \phi_j(\tilde{x}_i)}\right\} = 0, \qquad j = 1, 2, 3, \ldots, m-1 \tag{2}$$

$$\frac{\partial \log \mathrm{L}}{\partial \mu_{jk}} = \sum_{i=1}^{N}\left\{\frac{p_j \phi_j(\tilde{x}_i)\left[\dfrac{x_{ik}-\mu_{jk}}{\sigma_{jk}^2}\right]}{\sum_{j=1}^{m} p_j \phi_j(\tilde{x}_i)}\right\} = 0 , \text{ and} \tag{3}$$

$$\frac{\partial \log \mathrm{L}}{\partial \sigma_{jk}} = \sum_{i=1}^{\int \mathrm{N}} \left\{ \frac{p_j \phi_j(\tilde{x}_i) \left[\frac{(x_{ik} - \mu_{jk})^2 - \mu_{jk}}{\sigma_{jk}^2} \right]}{\sum_{j=1}^{m} p_j \phi_j(\tilde{x}_i)} \right\} = 0 . \tag{4}$$

Letting $Q_i = \sum_{j=1}^{m} p_j \phi_j(\tilde{x}_i)$ and $q_{ij} = \phi_j(\tilde{x}_i)$ then (3) and (4) can be written as

$$\mu_{jk} = \frac{\sum_{i=1}^{N} p_j q_{ij} x_{ik} \Big/ Q_i}{\sum_{i=1}^{N} p_j q_{ij} \Big/ Q_i} \quad \text{and} \tag{5}$$

$$\sigma_{jk} = \frac{\sum_{i=1}^{N} p_j q_{ij} (x_{ik} - \mu_{jk})^2 \Big/ Q_i}{\sum_{i=1}^{N} p_j q_{ij} \Big/ Q_i} . \tag{6}$$

Writing (2) as

$$\sum_{i=1}^{N} \left\{ \frac{q_{ij}}{Q_i} \right\} = \sum_{i=1}^{N} \left\{ \frac{q_{im}}{Q_i} \right\}$$

and multiplying by p_j, we have

$$p_j \sum_{i=1}^{N} \left\{ \frac{q_{ij}}{Q_i} \right\} = p_j \sum_{i=1}^{N} \left\{ \frac{q_{im}}{Q_i} \right\} \tag{2a}$$

Summing (2a) over j, we get

$$\sum_{j=1}^{m} \sum_{i=1}^{N} \left\{ \frac{p_j q_{ij}}{Q_i} \right\} = \sum_{i=1}^{N} \left\{ \frac{q_{im}}{Q_i} \right\} \tag{2b}$$

and since the left side of (2b) is N, it follows from (2a) that

$$p_j \sum_{i=1}^{N} \left\{ \frac{q_{ij}}{Q_i} \right\} = p_j N$$

or

$$p_j = \frac{\sum_{i=1}^{N} \left\{ \frac{p_j q_{ij}}{Q_i} \right\}}{N} \quad . \tag{7}$$

Thus equations (5), (6) and (7) can then be solved in an iterative fashion just as was done in the single indicator method. It has been found that the estimate of (5) must be substituted into (6) *within the iteration* and likewise that of (6) (and of (5)) into (7) for the iterative process to converge correctly.

This completes the derivation of the multi-indicator generalization of the maximum likelihood solution. Unfortunately a computer program for the method has not been completed and there are no trial results to report at this time.

VI. The Independence Assumption

The most important comparison of two multi-indicator generalizations would appear to be that of the relative robustness with respect to indicator independence within taxonomic classes. Even though the first method is not derived from explicitly stated assumptions as the second method is, and it is not known if the required assumptions for the first method do, in fact, include an independence assumption, a few Monte Carlo results indicate that the parameter estimates lose accuracy as within taxonic class dependency is greatly increased. Indicator independence within taxonomic class is an idealization that is not always easily approximated in practice. In the area of psychopathology measurement, not only do the most powerful taxometric indicators frequently correlate within taxonomic class because of shared valid variance but also because of shared error variance due to common sources of error of measurement.

While the (unknown) formal independence requirements of the first generalization could be either stronger or weaker than those of the second, most importantly, it is not known *which method is more robust* with respect to either (a) indicator independence within taxonomic class or (b) within taxonomic class normality. That is, it is not so much the strength of the assumptions but more the robustness with respect to the assumptions that is the matter of ultimate concern. Apparently, only a Monte Carlo investigation will allow for such comparison of the methods.

VII. The Normality Assumption

The assumption of intra-taxonomic class normality has several consequences which are important in such matters as construction of indicators from items and the identification of a taxonomy. Since the present taxonomic work uses unweighted sums of MMPI items, the present analysis is in terms of indicators which are sums of Bernoulli variables. The reasonableness or unreasonableness of the normality assumption can be more easily determined with aid from the result that the sum of (many) independent Bernoulli random variables is distributed asymptotically normal, the approximation being better the larger the number of such variables (see p. 24 of Golden and Meehl, 1973a). It should be pointed out that it is *not* necessary that the Bernoulli variable parameters all be the same; the demonstration referred to above shows that the variables can have any set of parameter values. Monte Carlo study should show that in practice, for short enough keys, the approximation is dependent on the dispersion of the item means, being better when the dispersion is small. It should likewise be shown that for reasonably long keys, say about 25 items or more, the distributions are approximately normal if the correlations between

each pair of items are close to zero. While pairwise item independence is a sufficient condition it has not been shown that it is a necessary one. It would appear that for a long enough key it is only necessary that the average of the absolute values of the correlations be near zero. These matters are terribly complex to handle analytically but a Monte Carlo study will be simple and straightforward and will provide a sufficiently precise and general result.

Since the method requires that indicators be constructed from items that are approximately independent within taxonomic class, the indicators have nearly zero homogeneity within taxonomic class. Thus, whatever degree of the homogeneity of the key does exist for the compound group is caused mainly, if not solely, by the compound group being a mixture of the taxonomic classes. Conversely, when a compound frequency distribution is determined to be the resultant of two or more overlapping normal frequency distributions then it is implied that the items of the key should be on the average independent within the corresponding taxonomic classes. This last result leads naturally to consistency tests which are discussed below. Most importantly here though, it follows that keys that are considered to be even moderately homogeneous measures of a single dimension within the taxonomic class are most likely not suited for use with the present method; the reason being, of course, that homogeneity implies non-normality of the intra-taxonomic class distributions.

For the first multi-indicator method, in which, in view of some Monte Carlo trials, strict independence between indicators is not required, it is interesting to note that while items within the key must be on the average independent within taxonomic class, items from the different keys presumably can be rather highly dependent within taxonomic

class. Thus, in constructing keys from a large set of discriminative items what is necessary is to somehow sort the items into groups such that the average absolute correlation within each group is near zero within taxonomic class; the keys are then formed from these groups paying no attention to cross-key correlation. Apparently two items from two different groups can be correlated moderately high on the average; possibly, the keys could even have the properties of parallel tests. Curiously, the item independence requirement for the maximum covariance method is diametrically opposite of this. The major assumption of the maximum covariance method is roughly that the indicators be independent within taxonomic class. The covariance of two keys is zero if all the items of one key are independent of all those of the other key. This result follows from the fact that the covariance of sums of items is a linear function of all the item cross-key covariances (McNemar, 1963, p.206). Although pairs of items from different keys must be weakly correlated on the average, items within a key can be highly correlated; that is, keys can be homogeneous within taxonomic class.

Thus, for example, if scales 7 and 8 of the MMPI are two of a set of candidate indicators of a purported taxonomy, then the normal theory is most likely more appropriate than the maximum covariance theory since the keys are regarded to be very heterogeneous and yet are highly correlated for most populations. On the other hand, if keys are known to have high internal consistency for compound samples, being developed by such methods as factor analysis, then the maximum covariance method is more appropriate. While such *a priori* considerations are sometimes useful they are not sufficient for determining if one theory or another is adequate in a particular instance. It is the ultimate responsibility of the consistency tests, which are discussed below, to

determine this. Suffice it to say here that if the assumptions of a theory are such that the real situation deviates more from the ideal condition than can be tolerated for accurate enough estimation of the parameters (the method is not robust enough), then a good consistency test will detect this disparity and indicate that the theory should be rejected.

The normality assumption implies that the compound distribution must have certain properties. The resultant of two sufficiently distinct but overlapping normal distributions differs from the normal one in being either skewed or platykurtic or bimodal. Thus the compound frequency distribution should be initially checked for the presence of one or more of these characteristics as the disjunction serves as a necessary requirement for the normal theory. Sampling irregularities make this step sometimes appear difficult; suffice it to say that when this happens it probably indicates that the sample size is too small for the normal method although accurate Monte Carlo guidance is yet to be obtained.

In view of the result mentioned above, the compound frequency distribution is non-normal because the compound item correlations or covariances are not zero. Analysis of the compound covariances leads to another taxonomic theory closely related to the normal theory. Letting the subscripts 'left' and 'right' denote the left and right taxonomic classes respectively, then the compound covariance between two items x, y is given by

$$\sigma_{xy} = P\sigma_{xy\,\text{right}} + Q\sigma_{xy\,\text{left}} + PQ\Delta p_x \Delta p_y$$

where

$\sigma_{xy\,\text{left}}$ is the covariance within the left taxonomic class,

$\sigma_{xy\,\text{right}}$ is the covariance within the right taxonomic class,

P is the base-rate of the right taxonomic class and

Δp_x and Δp_y are the differences in the item plus-rates for the two taxonomic classes for items x and y (see p. 51 of Meehl, 1965).

Assume as before, $\sigma_{xy\,\text{left}} = \sigma_{xy\,\text{right}} = 0$. Thus for n items and two taxa, there are $2n + 1$ unknowns (the taxonomic class item plus-rates and the base-rate) and there are $\binom{10}{2} = \frac{n(n-1)}{2}$ covariance mixture equations such as above. Thus if the system of equations

$$\left.\begin{aligned}\sigma_{ij} = PQ\left(p_{i\,\text{right}} - p_{i\,\text{left}}\right)\left(p_{j\,\text{right}} - p_{j\,\text{left}}\right)\\ p_i = p\left(p_{i\,\text{right}} - p_{i\,\text{left}}\right)\end{aligned}\right\} \quad \text{for } i, j = 1, 2, 3, \ldots, n;\ (i \neq j) \tag{8}$$

has an unique solution (it is overdetermined when $n \geq 3$) then it can be concluded that the items satisfy the latent conditions required by the normal theory (it is interesting that this is so even though the method does not make any reference to normal distributions) and analysis of the key compound frequency distribution by the maximum likelihood method should yield consistent results. Unfortunately, methods for solving (8), such as one by Brown (1967), have been found to be highly sensitive to sampling error and severely lack robustness with respect to assumption of zero within taxonomic class covariance.

VIII. Development of Indicators

When indicators are sums of items then consideration must be given to selecting items that are independent within taxonomic classes; however, there are no established methods for doing this. One possibility is provided by consistency hurdles theory, an independent taxonomic detection theory discussed in another report in this series (Golden, et al., 1974). Suffice it to say here that the method selects items which should be

sufficiently independent within taxonomic class as the method is based on the same requirement. If used in this manner the two methods would serve as checks on each other.

A method that has been found not to work well is to construct keys by factor analysis of the compound sample, forming a key from items that load highly on the same factor. A few such studies with the MMPI items indicate that factors tend to result from items that are not only discriminative of the taxonomic classes but are also correlated within the taxonomic classes. This result would evidently be obtained if the factor structures within the taxonomic classes were similar to that for the compound group.

One item selection theory that follows from the same assumption of independence proceeds as follows. Suppose there exists a large pool of items, some taxometrically discriminative, some not, and we do not know which is which, and we wish to choose a subset of items, each discriminative and each pair satisfying the independence requirement. Let i, j, k denote three items. Then if the three items are mutually independent within taxonomic classes it is easily shown that for the compound group

$$\mathrm{F}_i =_{\mathrm{def}} \frac{\sigma_{ij}\sigma_{ik}}{\sigma_{jk}} = \frac{\left(PQ\Delta p_i \Delta p_j\right)\left(PQ\Delta p_i \Delta p_k\right)}{\left(PQ\Delta p_j \Delta p_k\right)} = \left(PQ\Delta p_i^2\right)$$

where Δp_i is the difference in taxonomic class item means; or a constant for all j and k. For each i, all (j,k) pairs that give approximately the same estimate of F_i are determined. Again for each i, the largest subset, call it S_i, of those (j,k) pairs is determined such that every pair of items in S_i gives approximately the same value for F_i. Such a subset will be formed for every item i, and if it is large enough it can be further considered as a set of items for a key. If there are sets of mutually quasi-independent-within-taxonomic class items in the original pool then they should turn up as S_i subsets. It is clear however, that an S_i subset

does not necessarily consist of items which are quasi-independent within taxonomic classes.

The sample value of F_i, denoted by $\hat{F}_i$, will contain error due to sampling error in the three covariances; hence, it will not be exactly equal to $PQ\Delta p_i^2$. A second reason that F_i will not exactly equal $PQ\Delta p_i^2$ is that the three items will not be perfectly pairwise independent within taxonomic classes. Thus these two sources of error when taken into account give a range for acceptable values of F_i. In order to obtain an estimate of this range, consider each triplet of $\hat{\sigma}$'s as likewise containing these two kinds of error; namely, sampling error, α_{ij} and dependence error, β_{ij}. Since

$$\hat{F}_{i\,|\,jk} = \frac{\hat{\sigma}_{ij}\hat{\sigma}_{ik}}{\hat{\sigma}_{jk}}$$

and letting Δ denote a small deviation of $\hat{F}$ from $PQ\Delta p_i^2$ we have

$$\Delta\hat{F}_{i|jk} \doteq \frac{\hat{\sigma}_{ik}}{\hat{\sigma}_{jk}}\left(\alpha_{ij}+\beta_{ij}\right)+\frac{\hat{\sigma}_{ij}}{\hat{\sigma}_{jk}}\left(\alpha_{ik}+\beta_{ik}\right)-\frac{\hat{\sigma}_{ij}\hat{\sigma}_{ik}}{\hat{\sigma}_{jk}^2}\left(\alpha_{jk}+\beta_{jk}\right)$$

If we restrict the within taxonomic class sample covariance to be less than δ, then each β is equal to $P\delta + Q\delta = \delta$. In order to assess the α's we must know the sampling variance for the covariance of two Bernoulli variables.

The derivation for sampling covariance proceeds as follows. Let i and j be two subscripts which refer to two Bernoulli random variables. The sample covariance between the two variables is given by

$$s_{ij} = \frac{\sum_k \left(x_{ik}-\bar{x}_i\right)\left(x_{jk}-\bar{x}_j\right)}{n-1}$$

$$= \frac{\sum_{k=1}^{n} x_{ik}x_{jk}}{n-1} - \frac{n}{n-1}\bar{x}_i\bar{x}_j$$

the variance of s_{ij} is given by

$$\mathrm{E}\{s_{ij} - \mathrm{E}(s_{ij})\}^2 = \mathrm{E}(s_{ij}^2) - \{\mathrm{E}(s_{ij})\}^2 .$$

The first term can be developed as follows:

$$\mathrm{E}\left(s_{ij}^2\right) = \mathrm{E}\left[\frac{1}{n-1}\sum_k x_{ik}x_{jk} - \frac{n}{n-1}\overline{x}_i\overline{x}_j\right]^2$$

$$= \mathrm{E}\left[\frac{1}{(n-1)^2}\left(\sum_k x_{ik}x_{jk}\right)^2 - \frac{2n}{(n-1)^2}\overline{x}_i\overline{x}_j\sum_k x_{ik}x_{jk} + \left(\frac{n}{n-1}\right)^2\left(\overline{x}_i^2\overline{x}_j^2\right)\right]$$

$$= \frac{1}{(n-1)^2}\mathrm{E}\left(\sum_k x_{ik}x_{jk}\right)^2 - \frac{2n}{(n-1)^2}\mathrm{E}\left(\overline{x}_i\overline{x}_j\sum_k x_{ik}x_{jk}\right) + \left(\frac{n}{n-1}\right)^2\mathrm{E}\left(\overline{x}_i^2\overline{x}_j^2\right)$$

If we let $\mathrm{E}(x_i) = p_i$, $\mathrm{E}(x_j) = p_j$ and $\mathrm{E}(x_ix_j) = p_{ij}$, we have

$$\mathrm{E}\left(s_{ij}^2\right) = \frac{1}{(n-1)^2}\left[np_{ij} + n(n-1)p_{ij}\right]$$

$$-\frac{2n}{(n-1)}\left[\frac{1}{n}p_{ij} + \frac{n-1}{n}p_{ij}^2 + \frac{n-1}{n}p_{ij}p_i + \frac{n-1}{n}p_{ij}p_j + \frac{(n-1)(n-2)}{n}p_ip_jp_{ij}\right]$$

$$+\left[\frac{n}{(n-1)}\right]^2\left[\frac{n(n-1)(n-2)(n-3)}{n^4}p_i^2p_j^2 + \frac{4n(n-1)(n-2)}{n^4}p_ip_jp_{ij}\right.$$

$$+\frac{n(n-1)(n-2)}{n^4}p_i^2p_j + \frac{2n(n-1)}{n^4}p_{ij}^2 + \frac{2n(n-1)}{n^4}p_ip_{ij}$$

$$\left.+\frac{n(n-1)(n-2)}{n^4}p_ip_j^2 + \frac{2n(n-1)}{n^4}p_{ij}p_j + \frac{n(n-1)}{n^4}p_ip_j + \frac{1}{n^3}p_{ij}\right]$$

$$= p_{ij}\left[\frac{n}{(n-1)^2} - \frac{2}{(n-1)^2} + \frac{1}{n(n-1)^2}\right] + p_{ij}^2\left[\frac{n}{n-1} - \frac{2}{n-1} + \frac{2}{n(n-1)}\right]$$

$$+p_{ij}p_i\left[-\frac{2}{n-1} + \frac{2}{n(n-1)}\right] + p_{ij}p_j\left[-\frac{2}{n-1} + \frac{2}{n(n-1)}\right]$$

$$+p_ip_jp_{ij}\left[-\frac{2(n-2)}{n-1} + \frac{4(n-2)}{n(n-1)}\right] + \frac{(n-2)(n-3)}{n(n-1)}p_i^2p_j^2$$

$$+p_i^2p_j\left[\frac{n-2}{n(n-1)}\right] + p_ip_j^2\left[\frac{n-2}{n(n-1)}\right] - \frac{1}{n(n-1)}p_ip_j \ .$$

The second term is written in terms of the same parameters as follows:

$$\left\{\mathrm{E}\left(s_{ij}\right)\right\}^2 = \left(p_{ij} - p_i p_j\right)^2 = p_{ij}^2 - 2p_i p_j p_{ij} + p_i^2 p_j^2 \ .$$

After further simplification we have

$$\mathrm{E}\left\{s_{ij} - \mathrm{E}\left(s_{ij}\right)\right\}^2 = \frac{1}{n} p_{ij} - \frac{n-2}{n(n-1)} p_{ij}^2 - \frac{2}{n} p_i p_{ij}$$

$$+ p_i p_j p_{ij} \left[\frac{2(3n-4)}{n(n-1)}\right] + \frac{n-2}{n(n-1)} p_i p_j^2$$

$$+ \frac{n-2}{n(n-1)} p_i^2 p_j + \frac{1}{n(n-1)} p_i p_j - \frac{2(2n-3)}{n(n-1)} p_i^2 p_j^2 \ .$$

Letting $\sigma_{ij} = p_{ij} - p_i p_j$ or $p_{ij} = \sigma_{ij} + p_i p_j$, we have for the sampling variance of the covariance σ_{ij} between two Bernoulli variables, with parameters p_i and p_j, for samples of size n, the expression

$$\frac{1}{n}\left(\sigma_{ij} + p_i p_j\right) - \frac{n-2}{n(n-1)}\left(\sigma_{ij} + p_i p_j\right) - \frac{2}{n}\left(\sigma_{ij} + p_i p_j\right)$$

$$+ p_i p_j \left(\sigma_{ij} + p_i p_j\right)\left[\frac{2(3n-4))}{n(n-1)}\right] + p_i p_j^2 \left[\frac{n-2}{n(n-1)}\right] + p_i^2 p_j \left[\frac{n-2}{n(n-1)}\right]$$

$$+ p_i p_j \left[\frac{1}{n(n-1)}\right] - p_i^2 p_j^2 \left[\frac{2(2n-3))}{n(n-1)}\right] \ .$$

The Monte Carlo method was used to check the formula at a few values of the parameters and the results agreed very well with those of the formula. In Table 6 the standard deviation of s_{ij} is given in terms of various values of n, p_i, p_j and σ_{ij} .

Returning to the expression for $\Delta\hat{\mathrm{F}}_i$ we are now able to estimate upper and lower bounds. To obtain an upper bound we assume that $\beta_{ij} = \beta_{ik} = \delta > 0$, and since the probability that α_{ij} and α_{ik} deviate positively and α_{jk} deviates negatively by a sigma unit or more is less than .05 and sigmas of the α's are less than .01 for $n \geq 500$, we have

$$\Delta\mathrm{F}_{i|ik} < \frac{\hat{\sigma}_{ik}\hat{\sigma}_{ik} + \hat{\sigma}_{ij}\hat{\sigma}_{jk} + \hat{\sigma}_{ij}\hat{\sigma}_{ik}}{\hat{\sigma}_{jk}^2}\left(.01 + \delta\right) = \mathrm{B}_{i|jk}$$

nearly all the time.

In similar fashion, a lower bound can be shown to be the negative of the upper bound. Thus for each *i*, pairs of items are selected so that the pairs of $F_{i|jk} \pm B_{i|jk}$ intervals intersect.

The method was tried in an example using 10 items, 5 of which discriminated strongly between the sexes ($\Delta p_i > .30$) and 5 which did not ($-.10 < \Delta p_i < .10$). Two of the discriminative items were strongly correlated within the sexes while all other eight pairs of discriminative items were not. The method successfully picked the four most discriminative and weakly correlated items. After a computer program is completed the method will be studied by more extensive empirical and artificial data trials and in conjunction with the maximum likelihood taxometric method. Although the method will not necessarily produce independent item sets (that is, if there are no such sets, the method may produce a spurious set), it should detect such sets of items (for large enough samples) if, in fact, they do exist.

Table 6. Values of the sampling standard deviation of the covariance of two Bernoulli variables as a function of sample size (N), covariance (σ_{ij}), and the two variable parameters, p_i and p_j.

			p_i			
N	σ_{ij}	p_j	.2	.4	.6	.8
100	.00	.2	.0161			
		.4	.0197	.0241		
		.6	.0197	.0241	.0241	
		.8	.0161	.0197	.0197	.0161
	.05	.2	.0204			
		.4	.0206	.0240		
		.6	.0174	.0232	.0240	
		.8	—*	.0174	.0206	.0204
	.10	.2	.0228			
		.4	.0202	.0229		
		.6	—	.0211	.0229	
		.8	—	—	.0202	.0228
	.15	.2	.0240			
		.4	—	.0205		
		.6	—	.0170	.0205	
		.8	—	—	—	.0240
500	.00	.2	.0072			
		.4	.0088	.0107		
		.6	.0088	.0107	.0107	
		.8	.0072	.0088	.0088	.0072
	.05	.2	.0091			
		.4	.0092	.0107		
		.6	.0078	.0103	.0107	
		.8	—	.0078	.0092	.0091
	.10	.2	.0102			
		.4	.0090	.0102		
		.6	—	.0094	.0102	
		.8	—	—	.0090	.0102
	.15	.2	.0107			
		.4	—	.0091		
		.6	—	—	.0091	
		.8	—	—	—	.0107
1000	.00	.2	.0051			
		.4	.0062	.0076		
		.6	.0062	.0076	.0076	
		.8	.0051	.0062	.0062	.0051
	.05	.2	.0064			
		.4	.0065	.0076		
		.6	.0055	.0073	.0076	
		.8	—	.0055	.0065	.0064
	.10	.2	.0072			
		.4	.0064	.0072		
		.6	—	.0066	.0072	
		.8	—	—	.0064	.0072
	.15	.2	.0076			
		.4	—	.0064		
		.6	—	.0054	.0064	
		.8	—	—	—	.0076

*Note: The formula is evaluated only when the following conditions are met: (i) $p_{ij} \geq 0$, (ii) $p_{ij} - p_i \leq p_i(1 - p_i)\, p_j(1 - p_j)$, (iii) $p_i \geq p_{ij}$, $p_j \geq p$, (iv) $1 - p_i \geq p_j - p_{ij}$ and (v) $1 - p_j \geq p_i - p_{ij}$.

As the above method is cumbersome, although encouraging, it is of interest to describe a simple method that failed, as it leads to another important fact. In selecting items for keys it is necessary that an item meet two conditions. First, items must discriminate sufficiently between the taxonomic classes and second, an item must not be correlated with the other selected items within the taxonomic class. Thus, the total manifest covariance between items x and y for the compound sample must be such that while the first two terms of the covariance mixture equation are relatively small, the third is relatively large.

$$\sigma_{xy} = P\sigma_{xys} + Q\sigma_{xyn} + PQ\Delta\bar{x}\,\Delta\bar{y} \ .$$

If we sum the covariances of an item with each of the other items then we would expect this sum to be large to the extent the item is discriminative if we assume that the within taxonic class covariances to be about the same size no matter the degree of discrimination. Unfortunately, a trial of this procedure with the items of the male-female example was not very encouraging. Some insight into why this happened is gained by plugging typical numerical values into the covariance mixture equation. The within sex covariances are observed to be between .05 and .10 typically. While the contribution to the total covariance from the within sex covariance is .05 to .10, the contribution from the third term is typically not more than (.5)(.5)(.3)(.3) = .025, which is not only considerably less but only twice the sampling standard deviation of the covariance terms for large sample sizes. It is not difficult to imagine that the rank order of the item covariance sums is mainly influenced by the sampling fluctuation of the within taxonomic class terms.

The important fact, illustrated by this example, is that even for a pair of very discriminative items, a major portion of the total covariation comes from within the

taxonomic classes and a minor portion from the separation of taxonomic classes. That is, all pairs of items in the male-female keys are of this nature, and there is no reason to believe MMPI items for other taxonomies should behave differently in this regard.

IX. Classification of Individuals

The primary purpose of the present taxometric theories is for taxonomy detection and description and not necessarily classification of individuals. Normally, however, it will be of interest to compare the taxonomic classes on a number of other variables, not used as defining indicators, for purposes such as taxonomic class identification, substantive theory testing and the like and, in order to do this, classification of individuals may be necessary. Since the normality assumption has already been used and so justified by passage of the consistency tests, yet to be discussed, it is natural to use the estimated multinormal indicator density functions for each taxonomic class for the classification procedure also. It is noted that at this point it is necessary to make the additional assumption that the intra-taxonomic class indicator correlations are near zero. Using the estimated base-rates, Bayes' Rule can be used to obtain the posterior probabilities that an individual, with a given set of scores, belongs to each of the taxonomic classes. Presumably, he would be assigned to the most likely taxonomic class as this scheme can be shown to minimize the number of misclassifications when the prior probabilities are equal.

Classification of individuals also gives rise to an iterative bootstrapsing generalization of the method. Letting the two taxonomic classes be denoted by subscripts l [for left] and r [for right], the proportion of those correctly classified in taxon r, P_r, and those for taxon l, P_l, can be determined by a method given by Golden and Meehl (p. 31,

1973a). Writing the covariance mixture equation for each of the two classification groups we have

$$\sigma_{xy_{\mathrm{cl}}} = P_{\mathrm{l}}\sigma_{xy_{\mathrm{l}}} + Q_{\mathrm{l}}\sigma_{xy_{\mathrm{r}}} + P_{\mathrm{l}}Q_{\mathrm{l}}\Delta\bar{x}\Delta\bar{y} \text{ , and}$$
$$\sigma_{xy_{\mathrm{cr}}} = P_{\mathrm{r}}\sigma_{xy_{\mathrm{l}}} + Q_{\mathrm{r}}\sigma_{xy_{\mathrm{r}}} + P_{\mathrm{r}}Q_{\mathrm{r}}\Delta\bar{x}\Delta\bar{y} \tag{9}$$

(cl: classified as left; cr: classified as right)

where

$\sigma_{xy_{\mathrm{l}}}$ and $\sigma_{xy_{\mathrm{r}}}$ are the manifest covariances for individuals classified in groups l and r and

$\Delta\bar{x}$ and $\Delta\bar{y}$ are the estimated differences in the taxonomic class means.

Thus (9) can be solved simultaneously for the two unknowns $\sigma_{xy_{\mathrm{l}}}$ and $\sigma_{xy_{\mathrm{r}}}$. Now, with estimates of the intra-taxonomic class covariances, previously assumed to be zero, we are presumably in a position to improve the estimates of the multi-indicator density functions and, hence, improve the classification accuracy. The procedure can be applied repeatedly in an iterative fashion, possibly producing estimates that converge to approximately the true values. Again, Monte Carlo study is indicated.

It should be noted that the method does produce underestimates of the intra-taxonomic class covariances. This is because the taxonomic classes are not the latent groups considered in (9); the latent groups which we referred to are the taxonomic class members which are classified correctly.

Another method of estimation of the covariances results from the fact that the conditional covariance between two output indicators *x* and *y* (conditioned on the input variable w) is not a function of the input variable value when the indicators are multivariate normal within taxonomic classes. Anderson (1958) showed that

$$\sigma_{xy|w} = \sigma_{xy} - \frac{\sigma_{xw}\sigma_{yw}}{\sigma_w} \ .$$

Permutation of the subscripts gives

$$\sigma_{xw|y} = \sigma_{xw} - \frac{\sigma_{xy}\sigma_{wy}}{\sigma_y} \quad \text{and}$$

$$\sigma_{yw|x} = \sigma_{yw} - \frac{\sigma_{xy}\sigma_{xw}}{\sigma_x} \ .$$

Thus, we have three equations in three unknowns (σ_{xy}, σ_{wy}, σ_{xw}). The conditional covariances can be estimated by choosing a range of values on the input variable that has nearly all members from the same taxonomic class. Possibly the system of equations can be solved numerically by the method developed by Brown (1967) as existence of an explicit solution is not evident.

X. Consistency Tests

It is very important to have adequate consistency tests for any taxometric or psychometric theory. For a discussion of the consistency tests and the development of certain examples see Meehl (1973) and Golden and Meehl (1973b). Briefly, the purpose of the consistency test is to ascertain if the assumptions of the theory are adequate approximations of the truth in that the taxon detection is not spurious and the parameter estimates are accurate enough for a given substantive problem.

A very significant result occurred in the context of the preliminary study in the detection of schizotypes which illustrates the crucial role of consistency tests. Thirty items were selected which discriminated by .30 or more in the difference in item plus-rates between individuals already determined to be very likely schizotypes versus those

determined to be very likely nonschizotypes. However, when these items were made into a single key the normal method estimates of the base-rate and the other latent parameters was totally off the mark from numerous other consistent results. The χ^2 goodness-of-fit value was near the expected value and the general appearance of the fit of the theoretical to the observed compound frequency distribution was very good. The average correlation within the taxonomic classes turned out to be .5 and subsequent Monte Carlo trials showed that this is far too high for the method to work adequately. Thus, what is needed among the final set of consistency tests is one which determines if the average within taxonomic class item intercorrelation is sufficiently small.

The most obvious consistency test might be thought to consist of comparing the estimated compound distribution with the observed one. While the chi-square quantity measures such a discrepancy we do not wish to determine if this value exceeds a critical value used for statistical significance but rather if it exceeds a value which is required for accurate parameter estimation and non-spurious detection. Clearly it is not sufficient for the present purposes to determine if the chi-square value is significant or not at, say, the .50 level. The failure of the method with the first male-female key illustrates the need for consistency tests more sensitive than this. On the other hand, other artificial and real data trials have shown that a highly significant value can result even though parameter estimates are quite accurate. More importantly, though, a low value, say near the expected value, can easily be obtained even though parameter estimates are totally off the mark as was the case in the two cited examples above. Experimentation with the method of calculation, such as variation of the interval frequency size, has not led to any improvement in the usefulness of the chi-square value as a consistency test. A chi-square

goodness-of-fit value for the compound distribution near the expected value or below can *only* be considered as faintly encouraging, but certainly not by any means sufficient for confirming the existence of a taxonomy; on the other hand, while a highly significant value behooves one to look for further discorroboration of the theory, it does not *necessarily* refute either the existence of a taxonomy or the accuracy of the estimates. In short, the chi-square value appears to help very little in determining if the assumptions are close enough to the truth.

A number of consistency tests were developed by Meehl (1965, 1968, 1973), a few of which have been studied and modified slightly by the Monte Carlo method and thereby shown to work surprisingly well for the maximum covariance theory (Golden and Meehl, 1973b). Some further consistency tests are developed below.

A. *The case of a single indicator*

When only a single indicator is available the possibilities for consistency testing are limited but there are a few simply derived tests which immediately suggest themselves. First, the base-rate weighted taxonomic class means μ_j should be approximately equal to the compound mean μ where

$$\mu = \sum_j p_j \mu_j \ .$$

When μ is determined by the above equation then estimation errors in p_j and μ_j, denoted by Δp_j and $\Delta \mu_j$, cause an error $\Delta \mu$. The exact differential of μ is given by

$$\mathrm{d}\mu = \sum_j (\mu_j \, \mathrm{d}p_j + p_j \, \mathrm{d}\mu_j)$$

from which it follows that

$$\Delta \mu \doteq \sum_j (\hat{\mu}_j \, \Delta p_j + \hat{p}_j \, \Delta \mu_j) \ .$$

Maximum allowable values for estimation errors Δp_j and $\Delta \mu_j$ along with the estimates $\hat{\mu}_j$ and $\hat{p}_j$ can then be used to estimate the maximum allowable discrepancy between the grand mean when calculated as a function of estimates of the latent taxa means and base-rates and when calculated directly as a sample mean. Preliminary trials of this consistency test indicate that when it is failed the estimates are grossly in error; however, when the test is passed there is little assurance that the estimates are accurate enough. For example, taking the first male-female key (Table 1) it would be reasonable to require that

$$| \Delta p_{\mathrm{m}} | \leq .1 \ ,$$

$$| \Delta p_{\mathrm{f}} | \leq .1 \ ,$$

$$| \Delta \mu_{\mathrm{m}} | = s_{\mathrm{m}} \doteq \hat{s}_{\mathrm{m}} = 2.71 \text{ and}$$

$$| \Delta \mu_{\mathrm{f}} | = s_{\mathrm{f}} \doteq \hat{s}_{\mathrm{f}} = 2.05 \ .$$

Thus the maximum value of $\Delta \mu$ must be less than (.10)(9.98) + (.10)(14.69) + (.499)(2.71) + (.501)(2.05) = 4.84. The grand mean when calculated from the latent parameter estimates ($p_{\mathrm{f}} \mu_{\mathrm{f}} + p_{\mathrm{m}} \mu_{\mathrm{m}}$) is (.501)(14.69) + (.499)(9.98) = 12.34 and this differs by only .01 from the compound sample mean of 12.33; thus the consistency test in this instance is easily passed even though the estimate of p_{f} is in error by .110 which according to the chosen error limits is not quite acceptable. The test is passed in all the other examples as it should be. It appears that further Monte Carlo study will confirm that the test should be used only as a preliminary one for early detection of gross errors.

An analogous test to the above one results from expressing the compound variance (rather than the mean) in terms of the taxonomic class means, base-rates, and variances. The derivation of the formula for any number of taxonomic classes consists simply of repeated application of the variance mixture equation for two taxonomic classes.

There is always a limit on the degree to which taxonomic class distributions can overlap and still allow for accurate taxonomic detection. It appears that for the present method the taxonomic class means must be separated by at least one to one and one-half intra-taxonomic class sigma units (Cohen [1966] suggests that a two sigma separation is necessary) and that the base-rates for a dichotomous taxonomy must not be more disproportionate than .1 and .9. Parameter estimates not meeting these conditions are quite possible very erroneous and cannot be trusted without further corroboration.

If the indicator is a key consisting of the sum of dichotomous items then the possibility for consistency testing is greatly increased. First, the condition derived in the item selection procedure above in which triplets of covariances are considered should be satisfied by all the items of the key or such a selected subset of the original items should produce a result consistent with the original key. Second, each item can be used in the output role and the many relations between output and input variables developed by Meehl (1965, 1968) can be checked to see if they are approximately satisfied. Third, the method given above to estimate the intra-taxonomic class covariance by repeatedly solving a set of two simultaneous covariance mixture equations can be applied to each item pair. Monte Carlo study will show that the average and the variance of these covariance estimates cannot exceed certain limits without causing intolerable error; possibly such limits can be chosen so that the converse is also true.

B. *The case of multiple indicators*

When there are several indicators, the possibilities for consistency testing are greatly improved by the addition of several tests to those above. A number of such tests are developed by Meehl (1968) and a few more are given below.

For some taxonomies it should be true that when the method is applied to the indicators taken singly, the separate base-rate estimates should be in adequate agreement with each other and with the single estimate of the multi-indicator method. Such a consistency test does assume that the taxonomies detected for each indicator taken singly are the same as one detected when they are taken jointly. Presumably, this assumption is sometimes known to be true and sometimes known not to be at all true; of course, in the latter case the test cannot be used.

A preliminary trial in the detection of the schizotype taxon with MMPI keys indicates that it will be possible to develop three heterogeneous keys for use with the normal method, each having taxonomic class mean differences of nearly two intra-taxonomic class sigma units. For such a typical situation there exists a large portion of the three dimensional indicator hyperspace where nearly all of the individuals belong to the taxon. To demonstrate this let x, y, and z be the three indicators and assume that they are distributed trivariate normally within the taxon and within the nontaxon class just as the normal theory requires. The taxon density function, for example, is given by

$$\phi_s(x,y,z) = \frac{1}{(2\pi)^{3/2}|\mathrm{D}|^{1/3}} \exp\left(-\frac{x^2}{2}\right)$$

where

$$x^2 = (x,y,z)\begin{bmatrix} 1 & r_{xy} & r_{xz} \\ r_{yx} & 1 & r_{yz} \\ r_{zx} & r_{zy} & 1 \end{bmatrix}^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

and D is the determinant of the correlation matrix. Consider the proportion of the members of the taxon with scores such that $x > \mathrm{d}$, $y > \mathrm{d}$ and $z > \mathrm{d}$ for some cut d. This octant

proportion R can be calculated from the following equation if the correlation matrix is known and d is specified in intra-taxonomic class sigma units from the taxon mean:

$$\mathrm{R} = \int_{\mathrm{d}}^{\infty} \int_{\mathrm{d}}^{\infty} \int_{\mathrm{d}}^{\infty} \phi(x, y, z)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\,.$$

The triple integral can be evaluated numerically such as by the Gaussian method. If d is zero (each indicator is cut at the taxon mean) then R varies from .125 when the within taxon correlations are zero to .266 when these correlations are each .90; generally R is roughly a monotonic function of the average of the three correlations. Likewise it can be shown that the same quantity for the non-taxon class is less than .006. Thus, for the positive octant of the hyperspace formed by cuts at the taxon means it is expected that a proportion of the compound sample not be less than $.125P + .006(1 - P)$ and not more than $.266P + .006(1 - P)$ where P is the taxon base-rate which has been estimated by the normal method. By comparing these limits of the octant proportion with the observed value we have a consistency test. In the male-female three key example the limits are .125(.61) + .006(.39) = .077 + .002 = .079 and .266(.61) + .006(.39) = .162 + 002 = .164, and these surround the observed value of .141. Suffice it to say that numerous other closely related consistency tests using different hyperspace regions are suggested by such an approach.

The three indicator system for a dichotomous taxonomy can be checked by another procedure which begins by dichotomizing each of the indicators at, say, the estimated hitmax cuts. Using the score 1 for scores above hitmax cut and 0 for below or equal, we let $p_{x\mathrm{r}}$, $p_{y\mathrm{r}}$ and $p_{z\mathrm{r}}$ denote the right taxonomic class means and $p_{x\mathrm{l}}$, $p_{y\mathrm{l}}$ and $p_{z\mathrm{l}}$ the left taxonomic class means. Then under the condition of independence within the taxonomic class, we have the system of seven equations

$$p_x = Pp_{x\mathrm{r}} + Qp_{x\mathrm{l}}$$

$$p_y = Pp_{y\mathrm{r}} + Qp_{y\mathrm{l}}$$

$$p_z = Pp_{z\mathrm{r}} + Qp_{z\mathrm{l}}$$

$$p_{xy} = Pp_{x\mathrm{r}}\, p_{y\mathrm{r}} + Qp_{x\mathrm{l}}\, p_{y\mathrm{l}}$$

$$p_{xz} = Pp_{x\mathrm{r}}\, p_{z\mathrm{r}} + Qp_{x\mathrm{l}}\, p_{z\mathrm{l}}$$

$$p_{yz} = \mathrm{P}p_{y\mathrm{r}}\, p_{z\mathrm{r}} + Qp_{y\mathrm{l}}\, p_{z\mathrm{l}}$$

$$p_{xyz} = Pp_{x\mathrm{r}}\, p_{y\mathrm{r}}\, Pp_{z\mathrm{r}} + Qp_{x\mathrm{l}}\, p_{y\mathrm{l}}\, p_{z\mathrm{l}}$$

where the left-side parameters are directly observable and there are seven unknowns on the right side. Evidently the equations are not easily solved for an explicit solution, but a numerical iterative method by Brown (1967) can be used. The method requires initial estimates for the unknowns and these can be obtained from the normal method parameter estimates. The consistency tests consist of comparing these initial estimates with the final estimates. Initial guesses further off than .10 from the true value appear always to give incorrect results but such results appear to be detectable since they have always been impossible values (negative and greater than unity probabilities) in trials so far. It remains to be determined if the method is robust enough with respect to the independence assumption and sampling errors in the compound proportions but preliminary trials are not encouraging. Passage of the tests would appear to provide strong corroboration of the normal theory. However as the tests were not passed for the three male-female keys, the method for solving the equations appears to lack sufficient robustness.

Another consistency test results from the fact that the sum of two or more normally distributed variables is also normally distributed. First the base-rate estimates should be the same. If there are three indicators, then there are seven different sets that can be

analyzed singly. Second the mean and variance of the sum within a taxonomic class are given by

$$\mu = \Sigma\mu_j$$

$$\sigma^2 = \Sigma\ \sigma_i^2 + 2\ \sum_{i \neq j}\ \sigma_{ij}$$

where the μ_i and the σ_i are the within taxonomic class means and sigmas of the individual keys and the σ_{ij} are the within taxonomic covariances for pairs of individual keys. When there are only two indicators summed then the single within taxonomic class covariance between the two keys can be calculated form the second equation. In view of this interesting result, it is noted that the idea of summing keys can be used either primarily to estimate the within taxonomic class covariance matrices or as a set of consistency tests, possibly both.

Another consistency test evolves from the fact that in the typical taxonomic situation the separation between the latent means is one to two within taxonomic class sigma units. With three such indicators the probability of taxonomic class membership determined, say, by Bayes' Rule is most often very high or very low. If the probability of taxon membership is regarded as a random variable then the frequency distribution of this variable is strongly U-shaped. Preliminary Monte Carlo results indicate that this function is more nearly flat when the taxonomic detection is a spurious one.

When items are randomly assigned to, say, three keys then we can expect that the estimates of the latent means and variances should be fairly close from one key to the next. It is difficult to derive analytically useful results concerning just how close the parameter estimates should be. It is noted that items can be assigned randomly to keys any number of times.

When constructing keys by randomly assigning taxonomically discriminative items to the keys, there results an interesting relationship between the average correlation between items of a single key x within a taxonomic class, $\overline{r}_{wx}$, the average correlation between items of key x with those of key y within the same taxonomic class, $\overline{r}_{bxy}$, and the correlation between two keys within the same taxonomic class, $\overline{r}_{xy}$. McNemar (1963) showed that the three average correlations are always related by the formula

$$r_{xy} = \frac{nm\overline{r}_{bxy}}{\sqrt{n + n(n-1)\overline{r}_{wx}}\sqrt{m + m(n-1)\overline{r}_{wy}}}$$

where n and m are the number of items in keys x and y respectively. If items are randomly assigned to the two keys it is reasonable to assume that $\overline{r}_{bxy} \doteq \overline{r}_{wx} \doteq \overline{r}_{wy} \doteq \overline{r}$. When $n = m$, the formula then simplifies to

$$r_{xy} = \frac{n\overline{r}}{1 + (n-1)\overline{r}}.$$

This relationship between r_{xy} and $\overline{r}$ shows what the maximum value of $\overline{r}$ may be in order that r_{xy} is sufficiently small with respect to the (unknown) robustness or the method with respect to dependence between x and y within taxonomic classes. In Figure 1, r_{xy} is plotted as a function of the key length n for different values of $\overline{r}$ and perusal of the graphs reveals several interesting results. Since r_{xy} cannot be greater than .5 for the maximum covariance theory (Golden and Meehl, 1973b) it is seen that it is necessary that $\overline{r} < .1$ always and it is necessary that $n \leq 20$ for $\overline{r} = .05$. Also, a few artificial data trials have shown that the first multi-indicator normal method will not allow $\overline{r}$ to be as high as .2, possibly not much over .1 for $n = 20$, an optimal length for discrimination with MMPI

keys. Finally it would appear keys should be about as short as possible, which is 15 to 20 items usually, as ~~shorter~~ keys [shorter than that] begin to strongly violate the normality assumption.

Monte Carlo study is required to obtain precise information on the acceptable areas of (n, $\bar{r}$) values. With such results, a nice consistency test will consist of estimating r_{xy} by one of the procedures given above, then determining $\bar{r}$ from

$$\frac{r_{xy}}{n-(n-1)r_{xy}} ,$$

and then checking to see if this value is low enough for the taxometric method being used. However, as indicated above, such consistency tests should be difficult to pass when using MMPI items as such a method of key construction will be likely to yield $\bar{r}$ and $\bar{r}_{xy}$ values are far too large.

The final consistency test to be given here consists of comparing two ways of estimating the hitmax cut between two taxonomic classes. Unlike the numerous other approximative hitmax-cut estimation methods developed in previous reports (Meehl, 1965, 1968), the present method requires the indicator within taxonomic class distributions to be quasi-normally distributed. Just as in the former methods, it is assumed that the output indicator y and input indicator x are independent within taxonomic class. There is no restriction on the within taxonomic class distributions of the output indicator, so an item or a key not meeting the normality assumption could be used for example. If one considers the manifest output mean as a function of the input interval, denoted by $\bar{y}(x)$, then it will be shown that $\bar{y}(x)$ usually has a point of inflection very near or at the hitmax-cut.

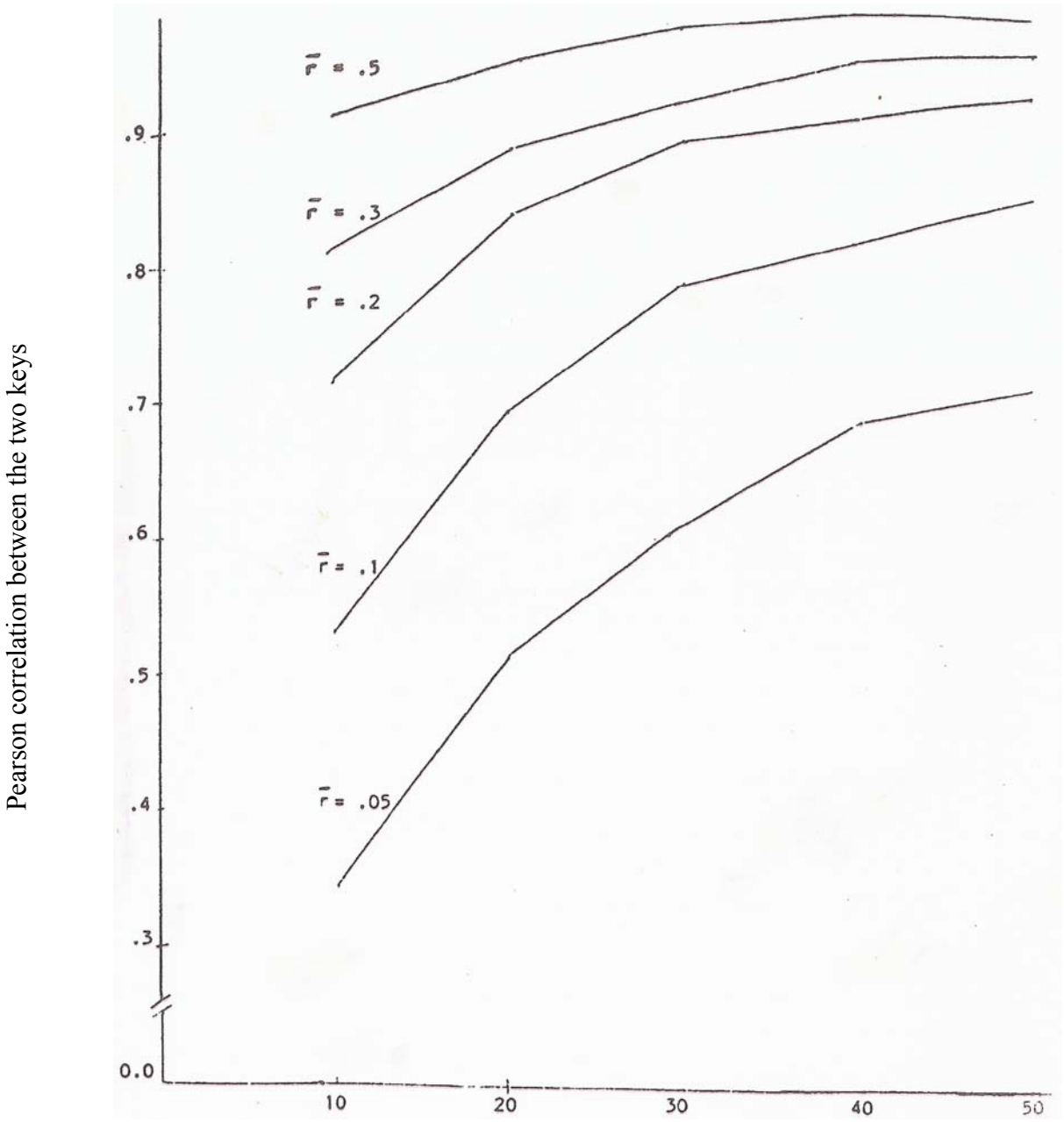


Figure 1. The Pearson correlation between two keys as a function of the common keylength for different values of the common average inter-item correlation within key, $\bar{r}$.

The number of right-taxonomic class members in interval x is approximately

$$f_{\mathrm{r}}(x) = \frac{N_{\mathrm{r}}}{\sqrt{2\pi}\sigma_{\mathrm{r}}} e^{-(x-\mu_{\mathrm{r}})^2/2\sigma_{\mathrm{r}}^2} \;;$$

likewise

$$f_{\mathrm{l}}(x) = \frac{N_{\mathrm{l}}}{\sqrt{2\pi}\sigma_{\mathrm{l}}} e^{-(x-\mu_{\mathrm{l}})^2/2\sigma_{\mathrm{l}}^2} .$$

At the hitmax-cut h, $f_{\mathrm{l}}(\mathrm{h}) = f_{\mathrm{r}}(\mathrm{h})$ or

$$\frac{N_{\mathrm{l}}}{\sigma_{\mathrm{l}}} e^{-(\mathrm{h}-\mu_{\mathrm{l}})^2/2\sigma_{\mathrm{l}}^2} = \frac{N_{\mathrm{r}}}{\sigma_{\mathrm{r}}} e^{-(\mathrm{h}-\mu_{\mathrm{r}})^2/2\sigma_{\mathrm{r}}^2}$$

Taking the natural logarithm of each side and rearranging terms, we have

$$\frac{(\mathrm{h}-\mu_{\mathrm{r}})^2}{2\sigma_{\mathrm{r}}^2} - \frac{(\mathrm{h}-\mu_{\mathrm{l}})^2}{2\sigma_{\mathrm{l}}^2} = \ln \frac{N_{\mathrm{r}}\sigma_{\mathrm{l}}}{N_{\mathrm{l}}\sigma_{\mathrm{r}}},$$

or

$$(\sigma_{\mathrm{l}}^2 - \sigma_{\mathrm{r}}^2)\,\mathrm{h}^2 + (2\mu_{\mathrm{l}}\,\sigma_{\mathrm{r}}^2 - 2\mu_{\mathrm{r}}\,\sigma_{\mathrm{l}}^2)\,\mathrm{h} + \mu_{\mathrm{r}}^2\,\sigma_{\mathrm{l}}^2 - \mu_{\mathrm{l}}^2\,\sigma_{\mathrm{r}}^2 - 2\,\sigma_{\mathrm{r}}\,\sigma_{\mathrm{l}}^2 \ln \frac{N_{\mathrm{r}}\sigma_{\mathrm{l}}}{N_{\mathrm{l}}\sigma_{\mathrm{r}}} = 0$$

Since the above equation is a quadratic function in h, it can be solved by the quadratic formula to give h explicitly in terms of the six other latent parameters.

The derivation of the second method for determining h consists of writing $\left.\frac{\mathrm{d}^2\bar{y}(x)}{\mathrm{dx}^2}\right|_{x=\mathrm{h}}$ in terms of the same latent parameters. Let the manifest output mean for input interval x be written as

$$\bar{y}_x = p_x\,\bar{y}_{\mathrm{r}} + (1 - p_x)\,\bar{y}_{\mathrm{l}} \quad \text{or} \quad \bar{y}_x = (\bar{y}_{\mathrm{r}} - \bar{y}_{\mathrm{l}})\,p_x - \bar{y}_{\mathrm{l}}$$

where p_x is the proportion of the individuals in interval x who are members of the right taxonomic class, and $\bar{y}_{\mathrm{r}}$ and $\bar{y}_{\mathrm{l}}$ are the two taxonomic class means. It follows that

$$\frac{\mathrm{d}\overline{y}_x}{\mathrm{d}x} = \left(\overline{y}_\mathrm{r} - \overline{y}_\mathrm{l}\right)\frac{\mathrm{d}p_x}{\mathrm{d}x} \quad \text{and} \quad \frac{\mathrm{d}^2\overline{y}_x}{\mathrm{d}x^2} = \left(\overline{y}_\mathrm{r} - \overline{y}_\mathrm{l}\right)\frac{\mathrm{d}^2 p_x}{\mathrm{d}x^2} \ .$$

Since $p_x = \dfrac{f_{\mathrm{r}x}}{f_{\mathrm{l}x} + f_{\mathrm{r}x}}$, we have

$$\frac{\mathrm{d}p_x}{\mathrm{d}x} = \frac{f_{\mathrm{l}x}\dfrac{\mathrm{d}f_{\mathrm{r}x}}{\mathrm{d}x} - f_{\mathrm{r}x}\dfrac{\mathrm{d}f_{\mathrm{l}x}}{\mathrm{d}x}}{\left(f_{\mathrm{l}x} + f_{\mathrm{r}x}\right)^2}$$

and

$$\frac{\mathrm{d}^2 p_x}{\mathrm{d}x^2} = \frac{\left(f_{\mathrm{l}x}\dfrac{\mathrm{d}^2 f_{\mathrm{r}x}}{\mathrm{d}x^2} - f_{\mathrm{r}x}\dfrac{\mathrm{d}^2 f_{\mathrm{l}x}}{\mathrm{d}x^2}\right)\left(f_{\mathrm{l}x} + f_{\mathrm{r}x}\right) + 2\left(\dfrac{\mathrm{d}f_{\mathrm{l}x}}{\mathrm{d}x} + \dfrac{\mathrm{d}f_{\mathrm{r}x}}{\mathrm{d}x}\right)\left(f_{\mathrm{l}x}\dfrac{\mathrm{d}f_{\mathrm{l}x}}{\mathrm{d}x} - f_{\mathrm{l}x}\dfrac{\mathrm{d}f_{\mathrm{r}x}}{\mathrm{d}x}\right)}{\left(f_{\mathrm{l}x} + f_{\mathrm{r}x}\right)^3}$$

In order to simplify the above expression, we write

$$\frac{\mathrm{d}f_x}{\mathrm{d}x} = \frac{-(x-\mu)}{\sigma^2} f_x = \mathrm{A}_x f_x \quad \text{and} \quad \frac{\mathrm{d}^2 f_x}{\mathrm{d}x^2} = \frac{(x-\mu)^2 - \sigma^2}{\sigma^4} f_x = \mathrm{B}_x f_x \ .$$

The numerator for $\dfrac{\mathrm{d}^2 f_x}{\mathrm{d}x^2}$ can now be written as

$$(\mathrm{B_r} f_\mathrm{l} f_\mathrm{r} - \mathrm{B_l} f_\mathrm{l} f_\mathrm{r})(f_\mathrm{l} + f_\mathrm{r}) + 2(\mathrm{A_l} f_\mathrm{l} + \mathrm{A_r} f_\mathrm{r})(\mathrm{A_l} f_\mathrm{l} f_\mathrm{r} - \mathrm{A_r} f_\mathrm{r} f_\mathrm{l})$$

which simplifies to

$$\{f_\mathrm{l} f_\mathrm{r}\,(\mathrm{B_r} - \mathrm{B_l})(f_\mathrm{l} + f_\mathrm{r}) + 2(\mathrm{A_l} f_\mathrm{l} + \mathrm{A_r} f_\mathrm{r})(\mathrm{A_l} - \mathrm{A_r})\}$$

which can be rewritten as

$$f_\mathrm{l} f_\mathrm{r}\,(\mathrm{C_l} f_\mathrm{l} + \mathrm{C_r} f_\mathrm{r})$$

where

$$\mathrm{C_l} = (\mathrm{B_r} - \mathrm{B_l}) + 2\mathrm{A_l}\,(\mathrm{A_l} - \mathrm{A_r}) \quad \text{and} \quad \mathrm{C_r} = (\mathrm{B_r} - \mathrm{B_l}) + 2\mathrm{A_r}\,(\mathrm{A_l} - \mathrm{A_r}) \ .$$

At hitmax h, $f_\mathrm{l} = f_\mathrm{r} = f_\mathrm{h}$, we have

$$\left.\frac{\mathrm{d}^2\overline{y}_x}{\mathrm{d}x^2}\right|_{x=\mathrm{h}} = \left(\overline{y}_\mathrm{l} - \overline{y}_\mathrm{r}\right)\frac{f_\mathrm{h}^3\left(\mathrm{C_l} + \mathrm{C_r}\right)}{8 f_\mathrm{h}^3} = \frac{\left(\overline{y}_\mathrm{l} - \overline{y}_\mathrm{r}\right)\left(\mathrm{C_l} + \mathrm{C_r}\right)}{8} \ .$$

Since $C_l + C_r = 2(B_r - B_l) + 2A_l^2 - A_r^2$ which simplifies to $\frac{2}{\sigma_l^2} - \frac{2}{\sigma_r^2}$ we have

$$\left.\frac{d^2\bar{y}}{dx^2}\right|_{x=h} = \frac{(\bar{y}_l - \bar{y}_r)}{4}\frac{\sigma_r^2 - \sigma_l^2}{\sigma_r^2\sigma_l^2}$$

which is precisely zero only when $\sigma_r = \sigma_l$. Generally, with estimates of $\hat{\sigma}_r$ and $\hat{\sigma}_l$ available one can use a numerical differentiation method to determine where $\frac{d^2\bar{y}}{dx^2}$ is closest to $\frac{(\bar{y}_l - \bar{y}_r)}{4}\frac{\sigma_r^2 - \sigma_l^2}{\sigma_r^2\sigma_l^2}$ and this should be adequately close to the previous hitmax-cut estimate.

The same method can be applied using just the first derivative but it would appear that the second derivative can be more easily evaluated near hitmax by a numerical method as it is close to zero at hitmax when variances are nearly equal. In either method the accuracy of the hitmax estimate will be limited by that of the numerical differentiation method when applied to sample data and this is only determinable by Monte Carlo study.

REFERENCES

Anderson, T.W. (1958) *An introduction to multivariate statistical analysis*. New York: Wiley.

Brown, K.M. (1967) Algorithm 316, Solutions of simultaneous non-linear equations. *Communications of the ACM*, *10*, No. 22, 728-729.

Cohen, A.C. (1966) Estimation of parameters for a mixture of normal distributions by Victor Hasselblad. *Technometrics*, *8*, 445-446.

Cramér, H. (1946) *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.

Golden, 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* (Report No. PR–73–2). Minneapolis, MN: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., & Meehl, P.E. (1973b) *Detecting latent clinical taxa, V: A Monte Carlo study of the maximum covariance method and associated consistency tests* (Report No. PR–73–3). Minneapolis, MN: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., Tyan, S.H., & Meehl, P.E. (1974) *Detecting latent clinical taxa, VI: Analytical development and empirical trials of the consistency hurdles theory* (Report No. PR–74–4). Minneapolis, MN: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Harding, J.P. (1949) The use of probability paper for the graphical analysis of polymodal frequency distributions. *Journal of the Marine Biological Association*, *28*(1), 141-153.

Hasselblad, V. (1966) Estimation of parameters for a mixture of normal distributions. *Technometrics*, *8*, 431-444.

Loevinger, J. (1957) Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635–694.

McNemar, Q. (1963) *Psychological Statistics*. New York: Wiley.

Meehl, P.E. (1962) Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, *17*, 827–838. Reprinted in Meehl, *Psychodiagnosis: Selected papers* (pp. 135–155). Minneapolis, MN: University of Minnesota Press, 1973.

Meehl, P.E. (1964) *Manual for use with checklist of schizotypic signs* (Report No. PR–73–5). Minneapolis, MN: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Meehl, P.E. (1965) *Detecting latent clinical taxa by fallible quantitative indicators lacking an accepted criterion* (Report No. PR–65–2). Minneapolis, MN: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Meehl, P.E. (1968) *Detecting latent clinical taxa, II: A simplified procedure, some additional hitmax cut locators, a single-indicator method, and miscellaneous theorems* (Report No. PR–68–4). Minneapolis, MN: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Meehl, P.E. (1973) MAXCOV-HITMAX: A taxonomic search method for loose genetic syndromes. In *Psychodiagnosis: Selected papers* (pp. 200–224). Minneapolis, MN: University of Minnesota Press.

Meehl, P.E., Lykken, D.T., Burdick, M.R., & Schoener, G.R. (1969) *Identifying latent clinical taxa, III. An empirical trial of the normal single-indicator method, using MMPI Scale 5 to identify the sexes* (Report No. PR–69–1). Minneapolis, MN: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Scarborough, J.B. (1962) *Numerical mathematical analysis*. Baltimore, MD: Johns Hopkins Press.