DIMENSIONALITY STRUCTURE ANALYSIS IN LATENT TRAITS ESTIMATION

HAE RIM KIM

1. Item Response Theory

Let θ be the trait (ability, skill, etc.) to be measured. For dichotomous item (scored 0 or 1), the item response function is the probability P or $P(\theta)$ of a correct response to the item. It is assumed that $P(\theta)$ increases as θ increases. A commonly assumed model for this probability can be the (three-parameter) logistic function

$$P \equiv P(\theta) = c + \frac{1-c}{1+exp\{-1.7a(\theta-b)\}}$$

where a is called the item discrimination parameter, b the difficulty parameter, and c the guessing parameter, respectively.

It is also assumed that probability of success on an item depends only on three item parameters and ability θ, referring to the assumption of *local independence (LI)*. Let $X_i = 0$ or 1 be the score on item i, then LI can be written as

$$P(X_1 = 1, ..., X_n = 1 \mid \theta) = \prod_{i=1}^{n} P(X_i \mid \theta)$$

where n is the length of the test (see Lord (1980) for details). Strictly test dimensionality is defined to be the dimensionality of the vector θ holding LI. In practice, however, the notion of essential dimensionality of Stout (1990) is acceptable instead.

2. A Dimensionality Structure Analysis

Definition of *DETECT*

Suppose $\{X_i; 1 \leq i \leq n\}$ is a test of n items. Suppose $A_1, A_2, \ldots, A_r$ are all non-empty subsets of the test $\{X_i\}$, $A_i \bigcap A_j = \emptyset$ for all $1 \leq i < j \leq r$, and $\bigcup_{i=1}^{r} A_i = \{X_i\}$. Then, $\mathcal{P} = \{A_1, A_2, \ldots, A_r\}$ is called a *r-subset (r-cluster) partition* of the test.

For an item pair (X_i, X_j), define a weighted sum of conditional covariance estimates of X_i and X_j as

$$\widehat{Cov}_{ij} = \frac{1}{J} \sum_{k=0}^{n-2} J_k \widehat{Cov}(X_i, X_j | S_{ij} = k). \tag{1}$$

Here S_{ij} is the observed correct score on the $(n-2)$ remaining items except for items i and j, J_k is the number of examinees with score $S_{ij} = k$, and J is the total number of examinees.

Let Ω be the set of all pairs of item indices, i.e.,

$$\Omega = \{(i, j), 1 \leq i < j \leq n\}.$$

Note that Ω has $n(n-1)/2$ elements.

The index *DETECT* is defined as

$$DETECT(\mathcal{P}) = \frac{2}{n(n-1)} \sum_{(i,j) \in \Omega} \delta_{ij} (\widehat{Cov}_{ij} - \overline{Cov}), \tag{2}$$

where $\mathcal{P}$ is any specified r-subset (r-cluster) partition of the test, $\overline{Cov}$ is the average of $\widehat{Cov}_{ij}$ over all $n(n-1)/2$ item pairs, and

$$\delta_{ij} = \begin{cases} 1 & \text{if items } X_i \text{ and } X_j \text{ are in the same cluster} \\ -1 & \text{otherwise.} \end{cases}$$

The index δ_{ij} manipulates the $(\widehat{Cov}_{ij} - \overline{Cov})$ term in (2), to be added or subtracted according as items X_i and X_j belong to the same cluster or not; when both items belong to the same cluster the *centered* (it is centered at $\overline{Cov}$) conditional covariance estimate $(\widehat{Cov}_{ij} - \overline{Cov})$ is added, while it is subtracted otherwise.

As a function of $\mathcal{P}$, *DETECT* can take on value ranging from negative to positive. One object here is to find one particular partition of the test *maximizing DETECT*. Maximum *DETECT* is expected when test is classified into dimensionally "correct" subsets.

Conditional Covariance

The total score (or the remaining total score) can be thought as (best) measuring a linear composite θ of the two traits having weights determined primarily by the influence of each trait; this linear composite is called the *test composite*. Thus $\widehat{Cov}_{ij}$ is interpreted as an estimator of $Cov(X_i, X_j \mid \theta)$ for an appropriate level of θ corresponding to the observed score S_{ij} value.

When a test is strictly (or essentially) unidimensional, $\widehat{Cov}_{ij} = 0$ (or approximately zero) because local independence holds for unidimensional θ approximately estimated by S_{ij}, except for statistical error caused by score unreliability and by estimation noise. Thus, maximum *DETECT* value will also be expected to be zero (or approximately zero) except for statistical error.

When a test is multidimensional and has *simple structure*, positive covariances should be produced between items on the same cluster measuring one trait, while negative covariances between items from different clusters except for statistical error.

Recall from (2) that δ makes the negative conditional covariances positive in calculating *DETECT* when they are from between-cluster items. Therefore, except for statistical error, *DETECT* can be expected to be maximized at the "correct" cluster formation.

DETECT$_{max}$

Denote $DETECT_{max}$ to be the maximum *DETECT* value calculated over all possible partitions of a test. According to the theoretical results of Kim (1994) and Zhang (1996), it becomes clear that the main objective is to find a partition that maximizes, or approximately maximizes, *DETECT*, because then one suspects that this partition, except for statistical error, correctly indicates the underlying multidimensional structure. The number of sizable clusters (that contain at least a certain number of items, say 4) in this partition that maximizes *DETECT* is judged to be the number of dimensions present in the test, and the average direction of the cluster that an item is located in corresponds to the dominant dimension the item is (best) measuring. The minimal cluster size restriction helps prevent the identification of dimensions having only a minor influence as well as helping reduce the possibility of statistical noise being opportunistically yet incorrectly judged by *DETECT* as contributing a (minor) dimension.

Since each estimated conditional covariance $\widehat{Cov}_{ij}$ contributes to a measure of the lack of unidimensionality resulting from violation of local independence (LI), the size of $DETECT_{max}$ can be viewed as an indicator that quantifies the *amount* of departure from unidimensionality. This amount of departure from unidimensionality is interpreted as the magnitude of departure from the *unidimensional composite direction* determined by a weighted average of all the underlying latent dimensions, these dimensions represented by item clusters in the approximate simple structure case. This composite direction can be thought of intuitively as the single dimension best measured by the test, somewhat like the psychologist's g on an intelligent test. $DETECT_{max}$ is expected to be close to zero for unidimensional data, while it reaches a substantially larger value for heavily multidimensional data.

Unfortunately, for a finite-length unidimensional test, there exists statistical bias in the index *DETECT* due to the lack of reliability of the conditioning scores S_{ij}, as recognized by Rosenbaum (1984), Holland & Rosenbaum (1986), Douglas, Kim & Stout (1994) and Kim et al. (in press) among others. That is it can be proposed under appropriate assumptions that

$$Cov(X_i, X_j|S_{ij}) > 0, \qquad \text{for all } 1 \le i < j \le n.$$

Therefore,

$$E[Cov(X_i, X_j|S_{ij})] > 0, \qquad \text{for all } 1 \le i < j \le n.$$

Notice that

$$E[Cov(X_i, X_j|S_{ij})] = \sum_{k=0}^{n-2} \text{Prob}\,(S_{ij} = k)\, Cov(X_i, X_j|S_{ij} = k). \tag{3}$$

By comparing (1) and (3), we see that $\widehat{Cov}_{ij}$ is a reasonable estimate of $E[Cov(X_i, X_j|S_{ij})]$. Hence, the claimed statistical bias of the $\widehat{Cov}_{ij}$ and hence of *DETECT* will occur. In order to correct this bias, the average $\overline{Cov}$ is subtracted from each $\widehat{Cov}_{ij}$ before it is combined into *DETECT*. This is why the $(\widehat{Cov}_{ij} - \overline{Cov})$ term is used in (2) rather than $\widehat{Cov}_{ij}$. Indeed, after this bias correction, as will be seen later, $DETECT_{max}$ remains small for unidimensional data as desired. Simulation studies show that this correction, designed for the correction of the positive bias in the unidimensional case, has no visible deleterious impact in the multidimensional case. That is, as desired, $DETECT_{max}$ remains large for strongly multidimensional data while staying near zero for the unidimensional data.

Table 1 below roughly categorizes a suggested quantitative interpretation of the amount of departure from unidimensionality, or the amount of multidimensionality, which is indicated by the maximum *DETECT* value. It should be stressed that the "amount" of multidimensionality is distinct from the number of dimensions; a two-dimensional data set could display a large amount of multidimensionality if the two dimensions are each well measured and are weakly correlated while an eight-dimensional data set could display very weak multidimensionality if there is only one dominant dimension and/or the multiple dimensions are highly correlated. In Table 1, the *DETECT* value has been multiplied by 100 for convenience.

Table 1
A Categorization of $DETECT_{max}$ as an Index
of Amount of Multidimensionality

$DETECT_{max}$	Multidimensionality
0.0 – 0.19	unidimensional
0.2 – 0.39	weak
0.4 – 0.79	moderate
0.8 –	strong

Theoretical justification of *DETECT* is developed by Zhang (1996) and it supports well the use of *DETECT*. Also it is essential to search for the meaningful cluster formation which maximizes the *DETECT*, in fact requiring enormous computation. Recently the Genetic Algorithm is adapted by Zhang (1996).

3. Real and Simulated Data Analysis

Real Data Analysis

Two Analytical Reasoning sections of an administration of the GRE have eight passages with 38 items. We chose four passages for which the data was complete and which seemed likely to be dimensionally distinct. These four passages have a total of 19 items with numbers of items/passage being 5, 4, 4, and 6. The number of examinees we used was 2477. *DETECT* was maximized at four clusters with the maximum *DETECT* value being 8.34×10^{-3}. Strikingly, the four clusters found by *DETECT* corresponded **exactly** to the items associated with each of the four passages.

Simulated Data Analysis

The three parameter logistic (3PL) model is used in the generation of dichotomously scored data. A 40 item test is split into several dimensionally distinct clusters. Each cluster is *unidimensional* to form a simple structure test. That is, all the items within a cluster load on *one* ability trait and the unidimensional ability varies over separate clusters. From unidimensional (1D) to four dimensional (4D) cases are simulated. Table 2 gives the number of items in each dimension. Item parameters are generated independently of items and of respective parameters within an item from the normal distribution.

Table 2
Number of Items in the Test and
in Each Dimensionally Distinct Cluster

	Number of Items in the Test	Number of Items in Each Dimensionally Distinct Cluster
1D	40	40
2D	40	20/20
3D	40	13/13/14
4D	40	10/10/10/10

The correlation coefficient between ability traits is one of the important factors to determine the extent of multidimensionality. In this simulation study, six different values, 0.3, 0.5, 0.7, 0.8, 0.85, and 0.9, are employed as the correlation coefficients among ability traits generated from the multivariate normal distribution. In each simulation model all possible pairs of ability traits have identical correlation coefficients. 6000 response vectors are generated per model, and then the data are cross validated with 3000 examinee responses used for constructing item clusters and the other 3000 examinee responses used for calculating *DETECT*. Note that all the values of *DETECT* presented in this paper are multiplied by 100 for ease of presentation.

The values of *DETECT* are displayed in Table 3 for the unidimensional simulated data with the increasing number of clusters up to 5. As expected, all 5 *DETECT* values remain fairly small.

Table 3

DETECT in the Unidimensional Case

	Number of Clusters				
	1	2	3	4	5
1D	0.0000	0.0340	0.0421	0.0524	0.0342

Table 4 shows $DETECT_{max}$ values for the two, three, and four dimensional cases at the different correlation coefficients. Notice that *DETECT* is maximized at the correct dimensionally-based cluster partitions in all these cases. It is interesting to observe that the size of $DETECT_{max}$ is a function of correlation coefficient. For example, the smaller the correlation coefficient, the larger $DETECT_{max}$, implying larger amounts of lack of unidimensionality. In all cases when the traits are highly correlated, there exists less multidimensionality revealing smaller $DETECT_{max}$. Also it is noteworthy that the size of $DETECT_{max}$ in Table 4 roughly explains the strength of multidimensionality of the data.

Table 4

$DETECT_{max}$ in the Two, Three, Four Dimensional Cases

	Correlation Coefficient					
	0.3	0.5	0.7	0.8	0.85	0.9
2D	2.8446	1.7669	0.9105	0.8155	0.5224	0.4401
3D	2.0013	1.5984	1.0471	0.6417	0.4472	0.3706
4D	1.5148	1.1825	0.7834	0.4424	0.3077	0.2748

4. Closing

The estimated conditional covariance based index *DETECT* for assessing the dimensionality structure of educational/psychological test data is defined and investigated extensively in order to discern its properties. Through analyses of simulated data, *DETECT* has been shown to display effective performance in identifying the number of dimensions present in test data as well as in identifying the items contributing to each dimension in the case of approximate simple structure and both the mixed and approximate simple structure cases for two dimensional data. *DETECT* has been shown to function effectively on identifying the paragraph-based items if a verbal test as producing separate dimensions. Also it quantifies the lack of unidimensionality of the data.

Recently, a theoretical justification for *DETECT* is made by defining its theoretical analogue, called theoretical *DETECT* (see Zhang and Stout, 1995). We can see that under certain reasonable conditions, the theoretical *DETECT* will be maximized at the correct simple structure cluster partition of the test items with the number of clusters in this partition corresponding to the number of dimensions of the test, for example, the clusters corresponding to items associated with the distinct paragraphs of a reading comprehension test. The properties of this theoretical *DETECT* are under further investigation. More investigation on the asymptotic behavior regarding *DETECT* is also planned for a future study as well as additional simulations to study the performance of *DETECT* when the dimensionality is at least three and approximate simple structure does not hold.

References

Douglas, J., Kim, H. R., & Stout, W. F. (1994). Exploring and explaining the lack of local independence through conditional covariance functions. Paper presented at the 1994 annual meeting of the American Educational Research Association, New Orleans, LA.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics, 14*, 1523-1543.

Kim, H. R., et al. (in press). Conditional covariance based nonparametric multidimensionality assessment. *Applied Psychological Measurement.* Invited Paper.

Kim, H. R. (1994). New techniques for the dimensionality assessment of standardized test data. Unpublished doctoral dissertation, Department of Statistics, University of Illinois at Urbana-Champaign.

Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problems.* Lawrence Erlbuam Associates.

Rosenbaum, P. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49*, 425-435.

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensional assessment and ability estimation. *Psychometrika, 55*, 293-326.

Zhang, J. (1996). *Some Fundamental Issues in Item Response Theory with Applications.* Unpublished doctoral dissertation, Department of Statistics, University of Illinois at Urbana-Champaign.

Zhang, J., & Stout, W. F. (1995). Theoretical results concerning *DETECT.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April 1995.

DEPARTMENT OF APPLIED STATISTICS SANGJI UNIVERSITY