Manifest Characterization and Testing for Certain Latent Properties [1]

Abbreviated Title: Manifest characterization

A. Yuan and B. Clarke

Department of Statistics, University of British Columbia and National Genome Center at Howard University

Abstract

Work due to Junker (1993) and more recently due to Junker and Ellis (1997) characterized desired latent properties of an educational testing procedure in terms of a collection of other manifest properties. This is important because one can only propose tests for manifest quantities, not latent ones. Here, we complete the conversion of a pair of latent properties to equivalent conditions in terms of four manifest quantities and identify a general method for producing tests for manifest properties.

[1] AMS 1991 subject classifications: Primary 62P15; secondary 62F03.
KEY WORDS: manifest, latent, educational testing.
Partial funding for this project was provided by NSERC Operating Grant OGP-0138122.

§1 Introduction

Item response theory, IRT, is the statistical theory of standardized tests which are commonly used in educational testing applications. The goal is to combine the data generated by many examinees answering a collection of test items so as to estimate the value of a parameter, or trait, say Θ, for each of the examinees. The parameter is intended to quantify a latent trait such as 'mathematics ability'. The traits are latent in the sense that they cannot be measured directly.

More specifically, desired properties of such testing procedures are called latent when their statement depends explicitly on the the latent trait being estimated. Most efforts at modeling involve the introduction of latent parameters: One specifies parameters to reflect aspects of the physical problem, such as the individual achievement of each examinee, the difficulty of each test item, specialized knowledge that certain examinees might possess, etc. This is in contrast to the manifest properties of a testing procedure which can be phrased in a way that does not explicitly depend on the latent trait Θ.

The limitation of a latent model is that it contains parameters for the latent trait: The experimenter wants to estimate each examinee's parameter value and so cannot perform hypothesis tests on a conjectured model in the usual way. The solution to this problem is to convert latent models, or latent statements about them, to physically meaningful manifest properties which are provably mathematically equivalent to the latent properties actually wanted. Manifest statements are amenable to conventional hypothesis testing and the problem then becomes the identification of optimal hypothesis tests for the manifest quantities. In this way one can test hypotheses that are independent of the latent structure, at least in principle.

The goal of characterizing latent properties in terms of manifest properties to be tested begins with Stout (1987). Stout (1987) considers the requirement that the smallest dimension of Θ for which the manifest model is a mixture of models conditional on θ is one; this is denoted $d_L = 1$. Clearly, this is a latent property. Stout (1987) defines a manifest analog called essential unidimensionality, denoted $d_E = 1$, and gives a hypothesis test for it.

Junker (1993) established a variety of results giving cases in which collections of latent properties were implied, or were implied by, collections of manifest properties. Our first result here is a version of Theorem 5.2b in Junker (1993) which gives $d_L = 1$ and another latent property – local asymptotic discrimination or LAD – as a consequence of four manifest properties, one of which is $d_E = 1$. We weaken one of the other three manifest properties to an asymptotic form, re-prove Junker's result and then establish the converse. We then propose hypothesis tests for the three manifest properties which currently do not have associated tests. Taken together this is the major contribution of our paper: The results here complete the manifest characterization of unidimensionality of Θ and LAD and provide a way to test for whether both of them are satisfied.

A more recent contribution due to Junker and Ellis (1997) characterizes monotone unidimensional models in IRT contexts in more generality, see also Ellis and Junker

(1997). The main result in this work characterizes monotone unidimensional models in terms of two properties: conditional association and vanishing conditional dependence. The first of these appears in our theorem mentioned above. So, it remains to deal with the second. We identify a hypothesis test for it in Section 5. However, as with the other tests here, extensive development will be necessary before it could actually be used.

The structure of the paper is as follows. In Section 2, we provide the key definitions, notation, and background to make the main characterization theorem intelligible. We state and prove this theorem in Section 3. In Section 4, we give tests for the manifest conditions identified in the theorem. In Section 5, we give one further test to indicate how the characterization result in Junker and Ellis (1997) can be used. In Section 6, we discuss briefly how testing might be done in practice.

§2. Notation and prelimilaries

The setting in which our results apply is the following. Let $\mathbf{x}_J = (x_1, x_2, \ldots, x_J)$ be an outcome of $\mathbf{X}_J = (X_1, ..., X_J)$, a binary response variable in which each of the X_i's takes values zero (wrong) or one (right). One can impose restrictions on the marginal distribution for the test items of an examinee $P(\mathbf{X}_J = \mathbf{x}_J)$ by making assumptions on the conditional distribution $P(\mathbf{X}_J = \mathbf{x}_J | \mathbf{\Theta} = \theta)$ for the vector of examinee responses $\mathbf{X}_J$ given the latent trait Θ. This follows from writing

$$P(\mathbf{X}_J = \mathbf{x}_J) = \int P(\mathbf{X}_J = \mathbf{x}_J | \mathbf{\Theta} = \theta) dF(\theta) \tag{2.1}$$

in which the sampling distribution of the latent variable $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_d)$ is $F(\theta)$. Often one requires conditional independence given Θ i.e., that $P(\mathbf{X}_J = \mathbf{x}_J | \mathbf{\Theta} = \theta) = \Pi_{j=1}^{J} P(\mathbf{X}_j = \mathbf{x}_j | \mathbf{\Theta} = \theta)$. The marginal distribution $P(\mathbf{X}_J = \mathbf{x}_J)$, in which Θ does not appear explictly, defines the manifest structure. By contrast, the latent variable appears explicitly in the marginal distribution $F(\theta)$ and conditional distributions $P(\mathbf{X}_J = \mathbf{x}_J | \mathbf{\Theta} = \theta)$ which define the latent structure of the sequence of response variables, see Cressie and Holland (1983).

In this context, there are three latent assumptions which are typically made in IRT, see Birnbaum (1968), Holland and Rosenbaum (1986), Rosenbaum (1987), Holland (1990) and Junker (1993). The first is called local independence (LI): The conditional probability for $\mathbf{X}_J$ given Θ in the integral of (2.1) factors as noted into a product of univariate probabilities. LI is just the usual factoring of the densities definition of statistical independence; 'local' here just means the property holds for a range of θ's. The second is called monotonicity (M): For each j the probability $P(X_j = 1 | \Theta = \theta)$ is increasing in θ. This has the interpretation that, roughly, the higher a value of the latent trait an examinee has, the more likely the examinee is to get question j right. The third latent assumption typically made is that the dimensionality d of Θ is much smaller than the test length J. In particular we want $d = 1$. See also Stout (1990) and Junker (1991) for related work on unidimensionality and essential independence.

To continue, many definitions are necessary. We group them into 4 classes. The first class has four members and pertains to dimensionality. Often we write $d_L = 1$ to mean more than $d = 1$. Following Stout (1990) we use the following.

Definition 2.1: The statement $d_L = 1$ means that one is the least dimension for which (2.1) holds and P_θ satisfies LI and M.

This concept of dimensionality will be translated into essential unidimensionality below. Essential unidimensionality requires the properties local asymptotic discrimination, LAD, and essential independence, EI. To define LAD, let $A_j = A_j(X_j)$ for $j = 1, ..., J$ be a sequence of random variables satisfying $\sup_j |A_j(\cdot)| \leq M < \infty$ for some positive M. The functions A_j are called uniformly bounded item scores. They are ordered if $A_j(0) \leq A_j(1)$. Moreover, ordered uniformly bounded item scores are said to be asymptotically discriminating if $(1/J)\sum_{j=1}^{J}(A_j(1) - A_j(0))$ is positive and bounded away from 0, as $J \to \infty$. Denote the mean of the item scores by $\bar{A}_J = (1/J)\sum_{j=1}^{J} A_j(X_j)$, and, with a slight abuse of notation, write $\bar{A}_J(\theta) = E(\bar{A}_J|\theta)$. When $d_L = 1$, $\overline{A}_J(\theta)$ may be inverted to produce estimates of θ directly. In particular, we use $\overline{A}_J^{-1}(\cdot)$ to denote the inverse function for $\overline{A}_J(\theta)$.

Now, from Junker (1993), LAD is formally defined as follows.

Definition 2.2: We say that $\mathbf{X}_J$ is locally asymptotically discriminating, LAD, if for every set of asymptotically discriminating item scores, to every θ there corresponds an interval N_θ containing θ and an $\epsilon_\theta > 0$ such that for any $t \in N_\theta$ with $t \neq \theta$ we have

$$\text{liminf}_{J\to\infty} \frac{\overline{A}_J(t) - \overline{A}_J(\theta)}{t - \theta} \geq \epsilon_\theta.$$

Next, consider the following analog to LI, taken from Junker (1993), modified from Stout (1990):

Definition 2.3: We say that $\mathbf{X}_J$ is essentially independent (EI) with respect to Θ if

$$\lim_{J\to\infty} Var(\overline{A}_J|\Theta = \theta) = 0$$

for every set of uniformly bounded item scores $\{A_j(\cdot) : j = 1, 2, \ldots\}$.

Using Definitions 2.2 and 2.3, Junker (1993) has the following.

Definition 2.4: We say that $\mathbf{X}_J$ is essentially unidimensional and write $d_E = 1$ if there exists Θ such that $\mathbf{X}_J$ is EI and LAD with respect to Θ.

This is not identical to the usage in Stout (1987) or Stout (1990), but is close and more appropriate here. Observe that that $d_E = 1$ is not, strictly, a manifest condition: LAD is latent and $d_E = 1$ depends on LAD. However, Stout (1987) has given a hypothesis test for $d_E = 1$ and since our goal is to test $d_L = 1$ and LAD, the latent nature of $d_L = 1$ remaining in $d_E = 1$ after replacing LI by EI is not important in practice. Henceforth, we implicitly assume unidimensionality although our statements hold, possibly with minor modifications, for the multidimensional case too.

The second class of definitions has five members and pertains to the conditional covariance between items. The first was introduced by Junker (1993).

Definition 2.5: We say that the covariances given the sum are nonpositive, CSN, if and only if for any $i < j \leq J$ the covariance between items i and j given the mean is negative, that is

$$\text{Cov}(X_i, X_j|\overline{X}_J) \leq 0.$$

We weaken Junker's definition to an asymptotic criterion on $\text{Cov}(X_i, X_j|\overline{X}_J)$, so it remains manifest.

Definition 2.6: The sequence $\mathbf{X}_J$ satisfies asymptotic CSN, ACSN, if and only if, for all $1 \leq i < j \leq J$, and all θ, we have that

$$P_\theta(\text{Cov}(X_i, X_j|\overline{X}_J) \geq \epsilon) \to 0,$$

for any $\epsilon > 0$, as $J \to \infty$.

It would be equivalent to require $\text{Cov}(X_i, X_j|\overline{X}_J)$ to be asymptotically nonpositive in the marginal probability P from (2.1).

The latent version of Definition 2.5 or 2.6 used in Junker (1993) is the following.

Definition 2.7: The sequence $\mathbf{X}_J$ satisfies the property that, locally, the covariances given the sum are non-positive, LCSN, if and only if

$$\text{Cov}(X_i, X_j|\overline{X}_J, \theta) \leq 0, \quad \forall i \neq j.$$

The third class of definitions is also based on covariances, but they are between functions of subvectors of $\mathbf{X}_J$. There are three members in this class. The first is from Junker (1993).

Definition 2.8: We say that $\mathbf{X}_J$ is locally associated, LA, if and only if for all θ, and all coordinatewise nondecreasing functions f and g, and all finite response vectors $\mathbf{Y}$ taken from $\mathbf{X}_J$ we have that

$$\text{Cov}(f(\mathbf{Y}), g(\mathbf{Y})|\Theta = \theta) \geq 0.$$

Definitions 2.7 and 2.8 are ony used in the proof of Theorem 3.1.

The second is from Holland and Rosenbaum (1986). It is the following.

Definition 2.9 We say that $\mathbf{X}_J$ is conditionally associated, CA, if and only if for every pair of disjoint, finite response vectors $\mathbf{Y}$ and $\mathbf{Z}$ in $\mathbf{X}$, and for every pair of coordinatewise nondecreasing functions $f(\mathbf{Y})$ and $g(\mathbf{Y})$, and for every function $h(\mathbf{Z})$, and for every $c \in \text{range}(h)$ we have that

$$\text{Cov}(f(\mathbf{Y}), g(\mathbf{Y})|h(\mathbf{Z}) = c) \geq 0,$$

for any c in the range of h.

The third is from Junker and Ellis (1997). It is a hybrid of CA and ACSN.

Definition 2.10 We say that $\mathbf{X}_J$ has vanishing conditional dependence, VCD, if and only if, for any partition $(\mathbf{Y}, \mathbf{Z})$ of the response vector $\overline{X}_J$, and any measureable functions f and g we have that

$$\lim_{m\to\infty} \text{Cov}(f(\mathbf{Y}), g(\mathbf{Z})|X_{J+1}, ..., X_{J+m}) = 0$$

almost sure.

Note that in this definition, the asymptotics are in test length rather than number of examinees.

The fourth class of definitions pertains to monotonicity, M. It has one member: the manifest analog of monotonicity, from Junker (1993).

Definition 2.11: Let $\overline{X}_{(j)} = \overline{X}_J - X_j/J$. We say manifest monotonicity, MM, holds if

$$E(X_i|\overline{X}_{(j)}) \text{ is nondecreasing as a function of } \overline{X}_{(j)}$$

for all $j \leq J$ and all J.

With these definitions in hand, we can informally state our main theorem: Taken together, the two latent conditions $d_L = 1,$ and LAD are equivalent to the four conditions $d_E = 1, ACSN, MM,$ and CA taken together. Three of these four conditions are manifest and we can identify hypothesis tests for them. The first condition, essential unidimensionality, already has a hypothesis test, see Stout (1987). We remark also that Junker and Ellis (1997) used CA and VCD to characterize monotone unidimensional representations of models, i.e., LI, M, and $d_l = 1$.

The fourth class of definitions are the regularity conditions we require for our formal results. The first of these comes from Junker (1993). We assume that $\mathbf{X_J}$ has been embedded in a sequence of binary response variables $\mathbf{X}$ and that for any finite response vector $\mathbf{Y}$ in $\mathbf{X_J}$

$$E(f(\mathbf{Y})|\Theta = \theta) \qquad \text{is continuous in } \theta \tag{2.1}$$

for any function $f(\mathbf{Y})$. We require the differentiability of conditional expectations, namely that for each J and each $j \leq J$

$$\sup_{j,J,u} |\frac{\partial}{\partial u} E(X_j|\overline{X}_J = u)| \leq M < \infty. \tag{2.2}$$

To make use of LCSN, we require an analogous to (2.2): For each J and $j \leq J$ we have that

$$\sup_{j,J,u} |\frac{\partial}{\partial u} E(X_j|\overline{X}_J = u, \Theta = \theta)| \leq M_\theta < \infty. \tag{2.3}$$

We also require the regularity conditions that permit application of the corollary to Theorem 4.1 in Clarke and Ghosh (1995). First, we assume the characteristic functions $f_j(t, \theta)$ of the response variables X_j, conditional on θ, are jointly continuous in (t, θ) uniformly in j and we denote the conditional density of X_j by $p(x_j|\theta) = p_\theta(x_j)$, with respect to counting measure for instance, when we need it. Next, we define $\overline{\mu_J}(\theta) = E_\theta \overline{X}$ where $\overline{X}$ is the sample mean of the first J X_j's and for $j = 1, ..., J$ we set $\Sigma_j(\theta) = \text{Var}_\theta(X_j)$ with mean $\overline{\Sigma}_J(\theta)$. Given this, we make a general definition.

Definition 2.12: A sequence of functions $< f_n(\theta) > |_{j=1}^{\infty}$ is locally invertible at θ_0 if and only if there is a neighborhood N_{θ_0} of θ_0 so that for all j $f_j|_{N_{\theta_0}} : N_{\theta_0} \longmapsto f_j(N_{\theta_0})$ is invertible, for $\theta \in N_{\theta_0}^c$ we have that $f_j(\theta) \in f_j(N_{\theta_0})^c$ and we have that the set $\cap_{j=1}^{\infty} f_j(N_{\theta_0})$ contains an open set around $\lim_{j \to \infty} f_j(\theta_0)$, assumed to exist.

Now, we require that $\bar{\mu}_j(\theta)$, $\bar{\Sigma}_J^{-1}(\theta)$, and $\nabla\bar{\mu}_J(\theta)^t\bar{\Sigma}_J^{-1}(\theta)\nabla\bar{\mu}_J(\theta)$ have Taylor expansions that are uniformly good in j, and that $\bar{\mu}_J(\theta)$ be locally invertible in the sense of Definition 2.11.

§3. The Main Result

To see the necessity of weakening CSN to ACSN we restate a result of Junker (1993).

Theorem : Suppose $\mathbf{X}$ is a sequence of binary responses and θ is unidimensional. If (2.1), (2.2) and (2.3) hold then:
a) CA, $d_E = 1$, LCSN, MM $\Leftrightarrow$ $d_L = 1$, LAD
b) CA, $d_E = 1$, CSN, MM $\Rightarrow$ $d_L = 1$, LAD
Proof : See Theorem 5.2 in Junker (1993).

It is seen that LCSN is latent and permits the biconditional in (a), whereas CSN which is manifest is so strong that a converse is unobtainable for (b). Relaxing CSN to ACSN will permit us to retain (b) and obtain the converse.

Theorem 3.1 *Forward Direction:* Assume (2.1) and (2.2). Then

$$CA, d_E = 1, ACSN, \text{and} \;\; MM, \tag{3.1}$$

taken together, imply the two latent conditions

$$d_L = 1, \text{and} \;\; LAD. \tag{3.2}$$

Backward Direction: Assume the logarithm of any density $p_\theta(x)$ is concave in x and that the regularity conditions at the end of Section 2 are satisfied. Then, the conditions in (3.2) taken together imply the conditions in (3.1), taken together.

Remark : The Forward proof is, mostly, a modification of techniques used in Junker (1993). The assumption of logconcavity is used for the backward direction so that the Corollary to Theorem 4.1 in Clarke and Ghosh (1995) can be applied. The proof of that result uses Theorem 2.8 in Joag-dev and Proschan (1983).
Proof : We start with the Forward Direction because, although harder, it is more important in practice.

By Definition , $d_E = 1$ implies that both LAD and EI are satisfied. So, it is enough to get $d_L = 1$. By definition, $d_L = 1$ is equivalent to LI and M taken together. We get M from Proposition 4.1 in Junker (1993). It states that EI, LAD and MM taken together imply M.

To obtain LI, we use Proposition 3.2 from Junker (1993) which shows that LI is equivalent to LA and LCSN taken together. The first of these, LA, follows by use of Proposition 3.1 in Junker (1993) which gives that CA, $d_E = 1$ and (2.1) taken together implies LA. We show how the second of these, LCSN, follows from the assumptions.

We begin by observing that

$$\text{Cov}(X_i, X_j|\theta) = \lim_{J\to\infty} \text{Cov}(X_i, X_j|\alpha_J(\theta) \leq \overline{X}_J \leq \beta_J(\theta)), \tag{3.3}$$

where $\alpha_J(\theta)$ and $\beta_J(\theta)$ are two functions satisfying $\beta_J(\theta) - \alpha_J(\theta) \searrow 0$ as $J \to \infty$. This follows from Lemma 3.1 in Junker (1993). Now, we can use ACSN to get LCSN. It is enough, by (3.3), to show that

$$\lim_{J\to\infty} \text{Cov}(X_i, X_j|\alpha_J \leq \overline{X}_J \leq \beta_J) \leq 0, \tag{3.4}$$

for all $\beta_J - \alpha_J \searrow 0$. To obtain (3.4), we follow Junker (1993). Note the standard identity

$$\text{Cov}(X_i, X_j|\alpha_J \leq \overline{X}_J \leq \beta_J) = E(\text{Cov}(X_i, X_j|\overline{X}_J)|\alpha_J \leq \overline{X}_J \leq \beta_J)$$
$$+\text{Cov}(E(X_i|\overline{X}_J), E(X_j|\overline{X}_J)|\alpha_J \leq \overline{X}_J \leq \beta_J). \tag{3.5}$$

It is enough to show both terms on the right hand side of (3.5) go to zero.

First, recall $\mathbf{X}$ is a sequence of binary responses so that $\text{Cov}(X_i, X_j|\overline{X}_J) \leq 1$ and write $\chi(A)$ to denote the indicator function for a set A. Now, given $\epsilon > 0$ the first term to control is:

$$E(\text{Cov}(X_i, X_j|\overline{X}_J)|\alpha_J \leq \overline{X}_J \leq \beta_J)$$
$$\leq \epsilon + E(\chi(\text{Cov}(X_i, X_j|\overline{X}_J) \geq \epsilon)|\alpha_J \leq \overline{X}_J \leq \beta_J). \tag{3.6}$$

Since $0 \leq \chi(\text{Cov}(X_i, X_j|\overline{X}_J) \geq \epsilon) \leq 1$, we have

$$0 \leq \overline{\lim}_{J\to\infty} E(\chi(\text{Cov}(X_i, X_j|\overline{X}_J) \geq \epsilon)|\alpha_J \leq \overline{X}_J \leq \beta_J) \leq 1. \tag{3.7}$$

So, the expectation of the middle quantity in (3.7) equals

$$\overline{\lim}_{J\to\infty} E(E(\chi(\text{Cov}(X_i, X_j|\overline{X}_J) \geq \epsilon)|\alpha_J \leq \overline{X}_J \leq \beta_J))$$
$$= \overline{\lim}_{J\to\infty} E(\chi(\text{Cov}(X_i, X_j|\overline{X}_J) \geq \epsilon)$$
$$= \int \lim_{J\to\infty} P_\theta(\text{Cov}(X_i, X_j|\overline{X}_J) \geq \epsilon) dF(\theta). \tag{3.8}$$

Now, by ACSN, the limit in (3.8) is zero, so the first term on the right in (3.5) is zero.

The second term on the right in (3.5) goes to zero by use of (2.2) and the same argument as is used to prove Lemma 5.1 in Junker (1993). Thus, LCSN follows.

Backward Direction: By Theorem 5.2 (a) of Junker (1993) we see that $d_L = 1$ and LAD imply CA, $d_E = 1$, and MM, three of the four conditions we must establish. Thus, we only have to prove ACSN. Since $d_L = 1$ implies LI, we can use the log concavity and the regularity conditions at the end of Section 2 to restate the Corollary to Theorem 4.1 in Clarke and Ghosh (1995) as LI implies ACSN. This completes the proof. □

In principle, the foregoing can be extended beyond settings in which the X_i's assume finitely many values. See Clarke and Yuan (2000) for a brief discussion.

§4. Testing for CSN, MM, and CA

In the three subsections here we give hypothesis tests for the three manifest conditions in Theorem 3.1. We recall that it is not necessary to give a test for $d_E = 1$ because Stout (1987) has already done so.

The tests we identify follow a common pattern: Identify a statistic which is a function of UMVU estimators, show this function is consistent for the quantity of interest, establish an appropriate form of asymptotic normality for the statistics, use existing results for the normal case to obtain the hypothesis testing optimality of the limiting procedure.

Suppose that each of m examinees writes a J-item dichotomous test. Usually, there are more examinees than test items, i.e., $m >> J$, so our asymptotics will be as $m \to \infty$ for fixed J. Let the scores of the i^{th} examinee be denoted by $\mathbf{X}_i = (X_{i,1}, ..., X_{i,J})$, for $i = 1, ..., m$. This means that $x_{i,j}$ is the i^{th} examinee's score on the j^{th} item.

Without knowing the value of θ for a given examinee, all we can do is assign the mixture density P from (2.1) to a given vector $\mathbf{X}_i$. This means that if we do not have access to a quantity such as θ on which to condition, then we are assuming that the $\mathbf{X}_i$'s are IID with respect to P. Within a given $\mathbf{X}_i$, the $X_{i,j}$'s are not independent (unless we condition on θ), but between different $\mathbf{X}_i$'s they would be independent. We comment that in this and later sections, omitted details of proofs can be found in Clarke and Yuan (2000).

4.1 Testing for CSN

Let $\bar{x}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J} x_{i,j}$ denote the average over item scores for the i^{th} examinee and let $\bar{x}_{\cdot j} = \frac{1}{m}\sum_{i=1}^{m} x_{i,j}$ be the average over the examinee's scores on the j^{th} item. To construct a test of CSN, we denote the generic score of an examinee on items p and q by X_p and X_q and write $\bar{X} = (1/J)\sum_{j=1}^{J} X_j$ for the generic test score of this examinee. (X_p and X_q are summands in $\bar{X}$.) If CSN holds then we expect that

$$r_{p,q}(\bar{x}) = \mathrm{Cov}(X_p, X_q | \bar{X} = \bar{x}) \leq 0, \tag{4.1}$$

for $p, q = 1, ..., J$, with $p \neq q$, and $\bar{x} = 0, 1/J, ..., (J-1)/J, 1$ since the item scores are binary.

Expression (4.1) means that there are $J+1$ values of $\bar{x}$ that we have to consider. So, for a given collection of m examinees, partition the space of all data vectors from the J items into disjoint sets based on the value of $\bar{x}$. We define

$$A_k = \{\mathbf{x}_i : \ \ \overline{x}_i = \frac{k}{J} \ \}, \quad 0 \leq k \leq J. \tag{4.2}$$

Writing $l_k = |A_k|$ for the cardinality of A_k, a natural estimator for $r_{p,q}(\frac{k}{J})$ from (4.1) is given by

$$\hat{r}_{p,q}(\frac{k}{J}) = \frac{1}{l_k}\sum_{i \in A_k} (x_{i,p} - \overline{x}_{\cdot p}(\frac{k}{J}))(x_{i,q} - \overline{x}_{\cdot q}(\frac{k}{J})), \tag{4.3}$$

where $\overline{x}_{\cdot p}(\frac{k}{J}) = \frac{1}{l_k}\sum_{i\in A_k} x_{i,p}$. Note that functions of the sets A_k, such as l_k and $\overline{x}_{\cdot,p}(\cdot)$ are dependent random variables. Indeed, there is a negative correlation between l_k and $l_{k'}$. (In a multinomial, the correlation between cell counts goes to a negative constant as the sum over all the cells increases.) Also, note there are $J+1$ values of k, and there are $J(J-1)/2$ pairs of items p, q which give potentially distinct values of $\hat{r}_{p,q}(\frac{k}{J})$. Our first result in this section gives the asymptotic behavior of $\hat{r}_{p,q}(\frac{k}{J})$ for each fixed value of p, q, k. Let

$$\mathbf{V}_m = \frac{1}{m}\sum_{i=1}^{m}\Big(X_{i,p}\chi(\overline{X}_i = \frac{k}{J}), X_{i,q}\chi(\overline{X}_i = \frac{k}{J}), X_{i,p}X_{i,q}\chi(\overline{X}_i = \frac{k}{J}), \chi(\overline{X}_i = \frac{k}{J})\Big)^T$$

$$:= \frac{1}{m}\sum_{i=1}^{m} V_{m,i}^T,$$

and note the $V_{m,i}$'s are IID random vectors. The estimated variance matrix $\mathrm{Cov}(V_1) = \Sigma_{p,q,k}$ of the $V_{m,i}$'s is

$$\hat{\Sigma}_{p,q,k} = \frac{1}{m}\sum_{i=1}^{m}(V_{m,i} - \mathbf{V}_m)(V_{m,i} - \mathbf{V}_m)^T.$$

Let $g(a,b,c,d) = c/d - ab/d^2$, $(\nabla g)(a,b,c,d) = (-b/d^2, -a/d^2, 1/d, (2ab - dc)/d^3)$, and $\boldsymbol{\mu}_{p,q,k} = E(\mathbf{V}_1)$. Then we can write

$$\sigma^2_{p,q,k} = \nabla g(\boldsymbol{\mu}_{p,q,k})\Sigma_{p,q,k}(\nabla g(\boldsymbol{\mu}_{p,q,k}))^T,$$

and

$$\hat{\sigma}^2_{p,q,k} = \nabla g(\mathbf{V}_m)\hat{\Sigma}_{p,q,k}(\nabla g(\mathbf{V}_m))^T.$$

Let $\overset{D}{\to}$ and $\overset{a.s.}{\to}$ denote convergence in distribution, and convergence almost sure, respectively. Holding the testlength J fixed, we have the following asymptotics in m.

Proposition 4.1 Assume $m \to \infty$. Then we have
(i) Consistency:

$$\hat{r}_{p,q}(\frac{k}{J}) \overset{a.s.}{\to} r_{p,q}(\frac{k}{J}),$$

and (ii) Asymptotic Normality:

$$\sqrt{m}(\hat{r}_{p,q}(\frac{k}{J}) - r_{p,q}(\frac{k}{J}))/\hat{\sigma}_{p,q,k} \overset{D}{\to} N(0,1).$$

Proof: (i) Write out $\hat{r}_{p,q}(k/J)$ as a collection of sums of m indicator functions weighted by the possible outcomes. Then apply the strong law of large numbers to each summation.

(ii) Theorem A in Serfling (1980, p. 122) gives

$$\sqrt{m}(g(\mathbf{V}_m) - g(\boldsymbol{\mu}_{p,q,k})) \xrightarrow{D} N(0, \sigma^2_{p,q,k}),$$

and $\hat{\sigma}^2_{p,q,k}$ is consistent for $\sigma^2_{p,q,k}$. □

Proposition 4.1 gives the asymptotics for fixed p, q, and k. Thus, we can estimate $\Sigma_{p,q,k}$ and use the asymptotic normality to test the hypothesis $r_{p,q}(\frac{k}{J}) \leq 0$ for any triple, (p, q, k). However, the hypothesis $H : CSN$ is that $r_{p,q}(\frac{k}{J})$ is non positive for all $J(J-1)(J+1)/2$ triples (p, q, k). Thus, by Proposition 4.1 we could test H using $J(J-1)(J+1)/2$ normal tests. However, we want to avoid performing so many tests.

Consider the condition $CSN(p, q)$ which is that the covariance between X_p and X_q given the mean $\bar{X}$ is nonpositive for all values of $\bar{x}$. The first part of the following theorem gives an asymptotically UMP level α test for any null hypothesis of the form $H_{p,q} = H : CSN(p, q)$ for given p and q with $p \neq q$. This test is based on the statistic $\hat{T}_{p,q} = \max_k(\hat{r}_{p,q}(\frac{k}{J}))$, permiting us to examine the covariance between test items. Clearly, it is equivalent to write $H : CSN$ as $H : CSN(p, q)$ for all distinct pairs p, q. The second part of our theorem below extends the test of $H : CSN(p, q)$ to $H : CSN$ by taking the maximum over all pairs (p, q). Thus, $H : CSN$ is equivalent to $H : \max_{p,q,k} r_{p,q}(\frac{k}{J}) \leq 0$ and we can base an asymptotically UMP level α test on the asymptotics of $\hat{T} = \max_{p,q,k}(\hat{r}_{p,q}(\frac{k}{J}))$. Our definition of asymptotic UMP level α is the following:

Definition 4.1: A testing procedure ϕ_m based on a sequence of test statistics $\{T_m\}$ is *asymptotic uniformly most powerful* (AUMP) level α for $H : \Omega_H$ versus $K : \Omega_K$ if and only if i)

$$\sup_{\theta \in \Omega_H} \lim_m E_\theta \phi_m \leq \alpha$$

and ii) for all $\theta \in \Omega_K$

$$\lim_m E_\theta \phi_m = \arg \left[\max_{\{\phi'_m\}} \lim_m E_\theta \phi'_m \right],$$

where the maximum is taken over all sequences $\{\phi'_m\}$ of asymptotically level α tests based on the same sequence $\{T_m\}$ for H versus K.

Note that clause (ii) permits there to be another sequence of statistics say S_m for which there might be another, different, AUMP level α test for H versus K. In the present context $T_m = \hat{T}_{p,q}$ is asymptotically normal. Even if S_m also has a limiting normal distribution we have not ruled out the possibility that a test based on S_m is AUMP level α and better than the test based on $\hat{T}_{p,q}$. Development of a testing procedure that would be asymptotically UMP level α over large classes of sequences of statistics is difficult, especially if the limiting distribution of S_m differs from the limiting distribution of $\hat{T}_{p,q}$.

To state our theorem, we require definitions suitable for fixed values p, q, and then analogous definitions when we take suprema over p and q. In the following, we let $r'_{p,q}(\frac{k}{J}) = r_{p,q}(\frac{k}{J})/\sigma_{p,q,k}$ be the rescaled version of $r_{p,q}(\frac{k}{J})$, $\hat{r}'_{p,q}(\frac{k}{J}) = \hat{r}_{p,q}(\frac{k}{J})/\hat{\sigma}_{p,q,k}$, $T_{p,q} = \max_k r'_{p,q}(\frac{k}{J})$ and $\hat{T}_{p,q} = \max_k \hat{r}'_{p,q}(\frac{k}{J})$. Since the $\sigma_{p,q,k}$'s are positive, $H_{p,q}$ can be reformulated as $H_{p,q}: T_{p,q} < 0$. Likewise for H: Let $T = \max_{p,q,k} r'_{p,q}(\frac{k}{J})$ and $\hat{T} = \max_{p,q,k} \hat{r}'_{p,q}(\frac{k}{J})$ so that H is $H : T < 0$. For $H_{p,q}$ with fixed p and q, let the range of $r'_{p,q}(\frac{k}{J})$ be $\{r_1 < r_2 < \cdots < r_d\}$ with $1 \leq d \leq J$, and denote the inverse of $r'_{p,q}(\frac{\cdot}{J})$ at r_k, for $1 \leq k \leq d$ by $A_k(p,q) = r'^{-1}_{p,q}(r_k)$. Obviously, $A_k(p,q) \cap A_j(p,q) = \emptyset$ for $k \neq j$. It is seen that for $l \in A_d(p,q)$, $r'_{p,q}(\frac{l}{J}) = \max_k r'_{p,q}(\frac{k}{J})$. Denote the multiplicity of $A_d(p,q)$ by $|A_d(p,q)|$. Now, $r_{p,q}$ is one-to-one if and only if $|A_k(p,q)| = 1$ for all $k = 1, ..., d$ and $d = J$. Let $\mathbf{Z}(A_d(p,q)) = (\hat{r}'_{p,q}(\frac{k}{J}) : k \in A_{d(p,q)})$. Note $\mathbf{Z}(A_d(p,q))$ is a vector of length $|A_d(p,q)|$, and we may denote it by $(Z_1, ..., Z_{|A_{d(p,q)}|})$. Also note that Z_k and $Z_{k'}$ are independent when $k \neq k'$.

Consider the one dimensional distributions $F_{A_d(p,q)}(x) = \Phi(x)^{|A_d(p,q)|}$, the distribution of the maximum of the $|A_d(p,q)|$ entries in the vector $\mathbf{Z}(A_d(p,q))$. That is, $F_{A_d(p,q)}(x)$ is the distribution function of $\max\{Y_1, ..., Y_{|A_d(p,q)|}\}$, where $(Y_1, ..., Y_{|A_d(p,q)|})$ are IID $N(0,1)$ because, for any x, $P(\max(Y_1, ..., Y_{|A_d(p,q)|}) \leq x) = \Pi_{i=1}^{A_d(p.q)} P(Y_i \leq x)$. Denote the $(1-\alpha)$-th percentile of $F_{A_d(p,q)}(x)$ by $F^{-1}_{A_d(p,q)}(1-\alpha)$.

For testing $H : CSN$, we use definitions similar to those used for $H_{p,q}$. Analogously, we record the following definitions: For fixed J, the maximum value of $r'_{p,q}(\frac{k}{J})$ over p, q and k is r'_d, and $A_d = r'^{-1}(r'_d)$, is the collection of triples (p,q,k) at which $r'_{\cdot,\cdot}(\frac{\cdot}{J})$ achieves the same maximal value r'_d. Let $\mathbf{z}(A_d) = (\hat{r}'_{p,q}(\frac{k}{J}) : (p,q,k) \in A_d)$ be the vector of conditional covariances with entries in A_d. Write $F_{A_d}(x) = \Phi(x)^{A_d}$, denote its $1-\alpha$ percentile by $F^{-1}_{A_d}(1-\alpha)$. We have the following.

Theorem 4.1 (i) An asymptotic level α test of $H_{p,q}$ is given by the rejection rule

$$\sqrt{m}\hat{T}_{p,q} \geq F^{-1}_{A_d(p,q)}(1-\alpha).$$

(ii) When $|A_d| = 1$, an AUMP level α test of $H_{p,q}$ is given by the rejection rule

$$\sqrt{m}\hat{T}_{p,q} \geq \Phi^{-1}(1-\alpha).$$

(iii) An asymptotic level α test for CSN is given by the rejection rule

$$\sqrt{m}\hat{T} \geq F^{-1}_{A_d}(1-\alpha).$$

(iv) When $|A_d| = 1$, an AUMP level α test for CSN is given by the rejection rule

$$\sqrt{m}\hat{T} \geq \Phi^{-1}(1-\alpha).$$

Proof: We only consider (i) and (ii); the proofs for (iii) and (iv) are similar.

(i) For $\epsilon > 0$, let $B_k(\epsilon) = \{|\hat{r}'_{p,q}(\frac{k}{J}) - r'_{p,q}(\frac{k}{J})| \leq \epsilon\}$ and let $B(\epsilon) = \cap_k B_k(\epsilon)$. We suppose ϵ is so small that the intervals $[r_k - \epsilon, r_k + \epsilon]$ for $k = 1, ..., d$ are disjoint. Now, consider the expression

$$\sqrt{m}(\hat{T}_{p,q} - T_{p,q}) = \chi_{B(\epsilon)}\sqrt{m}(\max_k \hat{r}'_{p,q}(\frac{k}{J}) - \max_k r'_{p,q}(\frac{k}{J}))$$

$$+\chi_{B^c(\epsilon)}\sqrt{m}(\max_k \hat{r}'_{p,q}(\frac{k}{J}) - \max_k r'_{p,q}(\frac{k}{J})). \tag{4.4}$$

The first term on the RHS of (4.4) is

$$\chi_{B(\epsilon)} \max_{k\in A_d(p,q)} \sqrt{m}(\hat{r}'_{p,q}(\frac{k}{J}) - r_d). \tag{4.5}$$

Using Prop. 4.1(ii), and (i) we can use (4.5) to show that the first term on the RHS of (4.4) converges to $F_{A_d(p,q)}$. The second term on the RHS of (4.4) goes to zero because χ_{B^c} goes to 0.

(ii) One can see that $H_0 : CSN(p,q)$ is asymptotically equivalent to $H_0 : T_{p,q} \leq 0$. Moreover, in the limit $\psi = T_{p,q}$ can be treated as a parameter, so it is as if we are testing $H_0 : \psi \leq 0$ versus $H_A : \psi > 0$. Since $p_\psi(\hat{T}_{p,q})$, the density for $\hat{T}_{p,q}$, has a monotone likelihood ratio in its normal limit, Theorem 2 in Lehmann (1986, p. 78) implies that the critical function given in (i) is UMP level α. A technical argument verifies that Definition 4.1 is satisfied. □

Note that we have not actually identified the values at which the maxima occur and that $|A_{d(p,q)}|$ or $|A_d|$ are unknown. Thus, to use Theorem 4.1 in practice, we might construct the 95% confidence intervals for the $r'_{p,q}(\frac{k}{J})$'s from the $\hat{r}'_{p,q}(\frac{k}{J})$'s. If the interval from the largest $\hat{r}'_{p,q}(\frac{k}{J})$ does not overlap with the other intervals, it suggests that $|A_{d(p,q)}| = 1$ so that (ii) or (iv) may be used for the testing. On the other hand, if several such intervals corresponding to the largest $\hat{r}'_{p,q}(\frac{k}{J})$'s overlap, it suggests $|A_{d(p,q)}| > 1$ and the number of overlapping intervals might be $|A_{d(p,q)}|$, similarly for A_d.

One can develop a parallel to Junker and Ellis (1997) for certain continuous cases also. Suppose the $X_{i,p}$'s have compact supports covered by a common compact set S. Let $A = \{t : r_{p,q}(t) = \sup_s r_{p,q}(s)\}$. When A has finitely many elements, we get results similar to those in the discrete case. When A has countably infinitely many elements, a more technical approach gives results similar in spirit to the discrete case. However, the details and the reasoning are quite different. See Clarke and Yuan (2000).

4.2 Testing for MM

It will be seen that the central ideas for testing MM are similar to those for testing CSN, and the results are parallel.

Consider the average score of the i^{th} examinee over the J items, but subtract the term for the j^{th} item. Denote this by $\bar{x}_{i,(j)} = \frac{1}{J}\sum_{r=1}^{J} x_{ir} - \frac{x_{ij}}{J}$. As a generic random variable this is $\overline{X}_{(j)} = \frac{1}{J}\sum_{i=1}^{J} X_i - X_j/J$, in which j indexes the item. Now, the quantity we test to see if MM is satisfied will be obtained from $\Delta_{k,(j)} = E(X_j|\overline{X}_{(j)} = \frac{k+1}{J}) - E(X_j|\overline{X}_{(j)} = \frac{k}{J})$, where $k = 0, ..., J-1$ and $j = 1, ..., J$. Indeed, MM can now be expressed as $\Delta_{k,(j)} \geq 0$, for all j and k. So, $H_0 : MM$ is equivalent to $H_0 : \min_{0\leq k\leq J-1, 1\leq j\leq J} \Delta_{k,(j)} \geq 0$.

To develop a natural estimator of $\Delta_{k,(j)} \geq 0$ and so a test statistic for $H : MM$, we partition the collection of examinees' binary response vectors based on the values of $\bar{x}_{i,(j)}$: Let $B_{k,(j)} = \{\mathbf{x}_i : \ \bar{x}_{i,(j)} = \frac{k}{J}\ \}$, where $k = 0, 1, ..., J-1$ and $j = 1, ...J$. Write the cardinality of $B_{k,(j)}$ as $l_{k,j} = |B_{k,(j)}|$ and drop the subscript j. Now, a natural estimate of $\Delta_{k,(j)}$ is

$$\hat{\Delta}_{k,(j)} = \frac{1}{l_{k+1}} \sum_{i \in B_{k+1}} x_{i,j} - \frac{1}{l_k} \sum_{i \in B_k} x_{i,j}. \tag{4.6}$$

Let

$$\mathbf{V}_m = \frac{1}{m} \sum_{i=1}^{m} \Big(X_{i,j}\chi(\overline{X}_{i,(j)} = \frac{k}{J}), X_{i,j}\chi(\overline{X}_{i,(j)} = \frac{k+1}{J}), \chi(\overline{X}_{i,(j)} = \frac{k}{J}), \chi(\overline{X}_{i,(j)} = \frac{k+1}{J}) \Big)^T$$

$$:= \frac{1}{m} \sum_{i=1}^{m} V_{m,i}^T,$$

with mean

$$\boldsymbol{\mu}_{k,j} = E(\mathbf{V}_1) = \Big(E(X_j\chi(\overline{X}_j = \frac{k}{J})), E(X_j\chi(\overline{X}_j = \frac{k+1}{J})), P(\overline{X}_j = \frac{k}{J}), P(\overline{X}_j = \frac{k+1}{J}) \Big)^T,$$

and write its covariance as $\Sigma_{k,j} = \mathrm{Cov}(\mathbf{V}_1)$. This time let $g(a,b,c,d) = (a/c) - (b/d)$ with $(\nabla g)(a,b,c,d) = (1/c, -1/d, -a/c^2, b/d^2)$. It is easy to see that $g(\mathbf{V}_m) = \hat{\Delta}_{k,(j)}$, and $g(\boldsymbol{\mu}_{k,j}) = \Delta_{k,(j)}$. Let $\sigma^2_{k,j} = \nabla g(\boldsymbol{\mu}_{k,j})\Sigma_{k,j}(\nabla g(\boldsymbol{\mu}_{p,q,k}))^T$ and let $\hat{\sigma}^2_{k,j} = \nabla g(\mathbf{V}_m)\hat{\Sigma}_{k,j}(\nabla g(\mathbf{V}_m))^T$, where

$$\hat{\Sigma}_{p,q,k} = \frac{1}{m} \sum_{i=1}^{m} (V_{m,i} - \mathbf{V}_m)(V_{m,i} - \mathbf{V}_m)^T.$$

Parallel to Proposition 4.1 we have the following.

Proposition 4.2 Assume $m \to \infty$, then we have in the mixture distribution P from (2.1)
(i) Consistency:

$$\hat{\Delta}_{k,(j)} \overset{a.s.}{\to} \Delta_{k,(j)},$$

and (ii) Asymptotic Normality:

$$\sqrt{m}(\hat{\Delta}_{k,(j)} - \Delta_{k,(j)})/\hat{\sigma}_{k,j} \overset{D}{\to} N(0,1),$$

where $\hat{\sigma}_{k,j}$ will be specified in the proof.

Proof: This parallels the proof of Prop. 4.1. For (i) write $\hat{\Delta}_{k,(j)}$ as a collection of sums of m indicator functions weighted by the possible outcomes and apply the strong law of large numbers to each summation. For (ii) use Theorem A in Serfling (1980, p. 122). □

To use Prop. 4.2 to test MM, let $\Delta = \max_{k,j} \Delta_{k,(j)}$, and $\hat{\Delta} = \max_{k,j} \hat{\Delta}_{k,(j)}$. Then testing MM is equivalent to testing $H_M : \Delta \leq 0$ vs $K_M : \Delta > 0$. In the same spirit as Theorem 4.1, let the range of $\Delta_{k,(j)}$ be $\{\delta_1 < \delta_2 < \cdots < \delta_d\}$ with the maximum of $\Delta_{k,(j)}$ over pairs (i,j) being δ_d. (We assume $-\infty < \inf_{k,j} \Delta_{k,(j)} \leq \sup_{k,j} \Delta_{k,(j)} < \infty$.) Set $A_d = \Delta^{-1}(\delta_d)$ to be the indices $\{(k,j)\}$ where $\Delta_{k,j}$ is maximal. Now we write $\mathbf{z}(A_d) = (\hat{\Delta}_{k,j} : (k,j) \in A_d)$, with asymptotic variance matrix $\Sigma(A_d)$, consistently estimated by $\hat{\Sigma}(A_d)$. Finally, let $\Phi_\Sigma(x_1, ..., x_k)$ be the distribution function of the k-dimensional normal distribution with mean zero (vector) and covariance matrix Σ. let $F_{A_d}(x) = \Phi_{A_d}(x, ..., x)$, $\hat{F}_{A_d}(x) = \Phi_{\hat{\Sigma}(A_d)}(x, ..., x)$, with $1-\alpha$ percentile $\hat{F}^{-1}_{A_d}(1-\alpha)$. Parallel to Theorem 4.1 we get AUMP level α tests for MM.

Theorem 4.2 (i) An asymptotic level α test of H_M is given by the rejection rule

$$\sqrt{m}\hat{\Delta} \geq \hat{F}^{-1}_{A_d}(1-\alpha).$$

(ii) When $|A_d| = 1$, an AUMP level α test of H_M is given by the rejection rule

$$\sqrt{m}\hat{\Delta} \geq \Phi^{-1}_{\Sigma(\hat{A}_d)}(1-\alpha).\Box$$

Proof: Let $B_{k,j}(\epsilon) = \{|\hat{\Delta}_{k,j} - \Delta_{k,j}| \leq \epsilon\}$, and $B(\epsilon) = \cap_{k,j} B_{k,j}(\epsilon)$ where $\epsilon > 0$ is so small that the intervals $(\delta_i \pm \epsilon)$ are disjoint for $i = 1, ..., d$. As in Theorem 4.1 it is enough to consider

$$\begin{aligned}\sqrt{m}(\hat{\Delta} - \Delta) &= \chi_{B(\epsilon)}\sqrt{m}(\max_{k,j} \hat{\Delta}_{k,(j)} - \max_{k,j} \Delta_{k,(j)}) \\ &+ \chi_{B^c(\epsilon)}\sqrt{m}(\max_{k,j} \hat{\Delta}_{k,(j)} - \max_{k,j} \Delta_{k,(j)}).\Box \qquad (4.7)\end{aligned}$$

4.3 Testing for CA

Here we develop an AUMP level α test for CA. The ideas are similar to those for CSN and MM. To see this, however, we must consider different orderings on response vectors so as to give a condition equivalent to CA but more amenable to testing. We begin by developing the mechanics to represent the functions that appear in the definition of CA.

First note that in the definition of CA, we can assume without loss of generality that the coordinatewise nondecreasing functions are non-negative. This follows because all indices are finite and the covariance will be unchanged if we subtract the infimum from each function.

Let S_J be the set of all vectors of length J with all entries 0 or 1. That is S_J is the set of vectors $s_1 = (0, ..., 0)$, $s_2 = (1, 0, ..., 0)$, $s_3 = (0, 1, 0..., 0)$, ..., $s_{J+1} = (0, ..., 0, 1)$, $s_{J+2} = (1, 1, 0..., 0)$, ..., $s_K = (1, 1, ...1)$, where $K = 2^J$. These s_i's represent the outcomes of the data vector $(X_1, ..., X_J)$. Next, define the partial ordering $\prec_0$ on the set S_J: For s_i and s_k in S_J we write $s_i \prec_0 s_k$ if and only if each coordinate of s_i is less than or equal to the corresponding coordinate of s_k. In the above sequence of vectors defining S_J we have $s_1 \prec_0 s_2$, and $s_1 \prec_0 s_3, ...,$ but there is no order specified between s_2 and s_3, s_2 and s_4.... In particular, no order is specified within a collection of s_i's having the same number of nonzero entries in different locations.

Nevertheless, we can extend the partial order $\prec_0$ by specifying some of the remaining size relationships. There are many ways to do this. We say that $\prec$ is a refinement of $\prec_0$ if $\prec$ retains all the orderings of $\prec_0$ while adding, in a consistent fashion, at least one new order between a pair of members that were not ordered under $\prec_0$. Let Λ be the collection of all maximal refinements $\prec$ of $\prec_0$; the refinements are maximal in the sense that there is no nontrivial refinement of them. Heuristically, $\prec$ extends $\prec_0$ if and only if there is a coordinatewise nondecreasing function f on S_J so that the ordering $\prec$ on S_J is the same as that induced by ordering on the values f assumes. Thus, $\prec = \prec_f$ for some f. Now, Λ is the collection of orders that reduce to $\prec_0$ and can be derived from f's that are strictly monotonic.

Next, we represent selections from the coordinates of the vector of length J by S_J. That is, let S_j be the set of all ordered subvectors of length j from $(1, 2, 3, ..., J)$. For instance, since the ordering is retained we have that for J=10 and j=5, $(2, 3, 6, 8, 9) \in S_j$ but $(2, 6, 3, 8, 9) \notin S_j$. Let $\Omega = \Omega_J$ be the disjoint union of all S_j's for $j \leq J$. Now, any element ω of Ω can be written as $\omega = \omega_j$ for some $\omega_j \in S_j$.

Now let $\prec \in \Lambda$ be an ordering on S_J and let $\omega \in \Omega$ be an ordered subvector of $(1, 2, ..., J)$ of length j, so $\omega = \omega_j \in S_j$. Next, define $S(\omega) = S(\omega_j)$ to be the set of all vectors of length j with entries zero or one, where the entries in the vectors are indexed by the j (ordered) entries of $\omega(j)$. For instance, if $J = 10, j = 3$ and $\omega_3 = (2, 5, 8)$, then $S(\omega(3)) = \{(x_2, x_5, x_8) : x_j = 0, 1; j = 2, 5, 8.\}$. Clearly, the cardinality of $S(\omega_j)$ is $|S(\omega_j)| = 2^j$. Observe next that ω defines a restriction of any ordering $\prec \in \Lambda$ on S_J to $\prec_{\omega_j}$ on $S(\omega)$; we write this as $\prec_\omega$.

For each fixed selection of coordinates ω of length j, let $\Lambda(\omega)$ be the set of all complete orderings on the set of possible outcomes for those coordiantes $S(\omega)$. For each fixed $\prec' (\omega) \in \Lambda(\omega)$, (each of them is a refinement of some $\prec_\omega$) all the elements of $S(\omega)$ can be listed by the ordering $\prec' (\omega)$ as $s_1, s_2, ..., s_{2^j}$, with $s_i \prec' (\omega) s_k$ when $i < k$. For this fixed $\prec' (\omega)$, let A_i be the ordered complement of the first i j-vectors s_1,...,s_j in $S(\omega)$. That is, $A_i = \{s_k : k \geq i\}$ for $i = 1, ..., 2^j$. Let $R(\prec' (\omega)) = \{A_1, ..., A_{2^j}\}$ be the sequence of sets of ordered vectors. Clearly, $A_l \supset A_{l+1}$. For different $\prec' (\omega)$'s, there are different classes $R(\prec' (\omega))$'s of A_i's. However, these classes will often have some common A_i's, since there are some common natural ordering relationships for different $\prec' (\omega)$'s. Let $\mathcal{S}(\prec_\omega)$ be the collection of all the different A_i's, that is

$$\mathcal{S}(\prec_\omega) = \cup_{\prec'(\omega) \in \Lambda(\omega)} R(\prec' (\omega)). \tag{4.8}$$

Next, we use these sets to define the set ordering relation and the set we actually want. For fixed j and $\omega(j)$, let $X(\omega(j))$ be the subvector of X from the index vector $\omega(j)$. Now, define $\mathcal{F}(\omega(j))$ to be the collection of all the non-negative, coordinatewise nondecreasing functions of $Y_j = X(\omega(j))$. With the above definitions we have the following characterization of $\mathcal{F}(\omega(j))$ because all members of $\mathcal{F}(\omega)$ are restrictions of functions on all of X.

Proposition 4.3 A non-negative coordinate-wise nondecreasing function f satisfies $f \in \mathcal{F}(\omega(j))$ if and only if there exists an ordering $\prec$ in $\Lambda(\omega(j))$, and a collection of real numbers $a_i \geq 0$, and a sequence of sets A_i in $\mathcal{S}(\prec_{\omega(j)}), i = 1, ..., 2^j$, so that

$$f(\cdot) = \sum_{i=1}^{2^j} a_i \chi_{A_i}(\cdot). \tag{4.9}$$

Proof: First, we show that f can be written as a telescoping sum for an increasing sequence of sets. That is, we show f can be represented using constants a_i on sets A_i where $i < i'$ implies that $A_i \subset A_{i'}$. Thus, re-order the vectors in $S(\omega(j))$ as $s_{(1)}, s_{(2)}, ..., s_{(2^j)}$ such that

$$f(s_{(1)}) \leq f(s_{(2)}) \leq ... \leq f(s_{(2^j)}).$$

Letting $a_1 = f(s_{(1)})$, and $a_i = f(s_{(i)}) - f(s_{(i-1)})$ for $i = 2, ..., 2^j$ gives the 'only if' part.

For the 'if' part, we must show that for $s_l, s_k \in S(\omega(j))$, with s_k coordinatewise greater than s_l, we have $f(s_l) \leq f(s_k)$. In fact, we have $s_l \prec s_k$, thus $l < k$. Thus, $f(s_k)$ is $f(s_l)$ plus at least one more non-negative term to represent outcomes $s_{l+1}, ..., s_k$. $\square$

We use Prop. 4.3 to get a condition equivalent to CA. For this we need another sub-vector $Z_{j'}$ of X_J with length j'. We assume that $Z_{j'}$ has no intersection with Y_j and denote its domain by $S_{j'}$. Write Y_j and $Z_{j'}$ as $X(\omega(j))$ and $X(\omega'(j'))$ respectively, where $\omega(j)$ and $\omega'(j')$ are non-overlapping sub-vectors of $(1, 2, ..., J)$ with length j and j' respectively. Now we have the following.

Proposition 4.4 The criterion CA is equivalent to

$$\min_{(j,j',\omega,\omega',\prec,\prec',A,B,D)} \text{Cov}(\chi_A(X(\omega(j))), \chi_B(X(\omega(j)))|X(\omega'(j')) \in D) \geq 0, \tag{4.10}$$

where the operation $\min_{(j,j',\omega,\omega',\prec,\prec',A,B,D)}$ denotes

$$\min_{j+j' \leq J;\, \omega(j),\omega'(j') \in \Omega, \omega(j) \cap \omega'(j') = \phi;} \; \min_{\prec \in \Lambda(\omega(j)) \prec' \in \Lambda(\omega'(j'));} \; \min_{A,B \in \mathcal{S}(\prec_\omega), D \subset \mathcal{S}(\prec'_{\omega'})} .$$

Proof: First write CA as a minimum over partitions of X_J. Then using Prop. 4.3, write each of the three functions in the definition of CA as a sum of the form (4.9). Taking the minimum over such functions identifies the other three minimizations. $\square$

Now that CA has been converted into a condition which is an explicit minimum it is amenable to the same sort of procedure as we used for CSN and MM. Indeed, Prop. 4.4 identified (4.10) as the central quantity for testing CA.

To fix notation, write $\psi = (j, j', \omega, \omega', \prec, \prec', A, B, D)$. The parameter ψ varies over $\Psi = \{\psi : 1 \leq j, j'; j + j' \leq J; \omega(j), \omega'(j') \in \Omega; \omega(j) \cap \omega'(j') = \phi; \prec \in \Lambda(\omega(j)), \prec' \in$

$\Lambda(\omega'(j')); A, B \in \mathcal{S}(\prec_\omega), \mathcal{D} \subset \mathcal{S}(\prec_{\omega'})\}$. The set Ψ is finite, though usually of enormous cardinality. Now, for $\psi \in \Psi$ we let

$$r_\psi = -\mathrm{Cov}(\chi_A(X(\omega(j))), \chi_B(X(\omega(j)))|X(\omega'(j') \in D). \qquad (4.11)$$

So, CA is equivalent to $\max_{\psi\in\Psi} r_\psi \leq 0$.

We develop an estimator for r_ψ as follows. Suppose we have m examinees with scores denoted $\mathbf{x}_i = (x_{i,1}, ..., x_{i,J})$, for $i = 1, ..., m$. We will be conditioning on $D \in \mathcal{S}(\prec'_{\omega'(j')})$ so such D's will define the subset of examinees over which we will average. For fixed D, let $G = G_D = \{i : \mathbf{x}_i(\omega'(j')) \in D\}$, and set $l = |G|$. Now, for $A, B \in \mathcal{S}(\prec_{\omega(j)})$ the averages of examinees' scores over G are $\overline{\chi}_A(D) = (1/l)\sum_{i\in G} \chi_A(\mathbf{x}_i(\omega(j)))$ and $\overline{\chi}_B(D) = (1/l)\sum_{i\in G} \chi_B(\mathbf{x}_i(\omega(j)))$. So,

$$\hat{r}_\psi = -\frac{1}{l}\sum_{i\in G}(\chi_A(\mathbf{x}_i(\omega(j))) - \overline{\chi}_A(D))(\chi_B(\mathbf{x}_i(\omega(j))) - \overline{\chi}_B(D)) \qquad (4.12)$$

is an estimator for r_ψ in (4.11).

Let $\mathbf{y}_i = \mathbf{x}_i(\omega(j))$ and $\mathbf{z}_i = \mathbf{x}_i(\omega'(j'))$;

$$\mathbf{V}_m = \frac{1}{m}\sum_{i=1}^m \Big(\chi_A(\mathbf{y}_i)\chi_D(\mathbf{z}_i), \chi_B(\mathbf{y}_i)\chi_D(\mathbf{z}_i), \chi_{A\cap B}(\mathbf{y}_i)\chi_D(\mathbf{z}_i), \chi_D(\mathbf{z}_i)\Big)^T$$

$$:= \frac{1}{m}\sum_{i=1}^m V_{m,i}^T,$$

and

$$\boldsymbol{\mu}_\psi = E(\mathbf{V}_1), \qquad \mathrm{Cov}(\mathbf{V}_1) = \Sigma_\psi.$$

It is seen that

$$\boldsymbol{\mu}_\psi = \Big(P(\mathbf{y}_i \in A, \mathbf{z}_i \in D), P(\mathbf{y}_i \in B, \mathbf{z}_i \in D), P(\mathbf{y}_i \in A\cap B, \mathbf{z}_i \in D), P(\mathbf{z}_i) \in D)\Big)^T.$$

Denote its covariance by $\Sigma_{k,j} = \mathrm{Cov}(\mathbf{V}_1)$. This time, let $g(a, b, c, d) = ab/d^2 - c/d$ with $(\nabla g)(a, b, c, d) = (b/d^2, a/d^2, -1/d, (dc-2ab)/d^3)$. It is easy to see that $g(\mathbf{V}_m) = \hat{r}_\psi$, and $g(\boldsymbol{\mu}_\psi) = r_\psi$. Let $\sigma_\psi^2 = \nabla g(\boldsymbol{\mu}_\psi)\Sigma_\psi(\nabla g(\boldsymbol{\mu}_\psi))^T$ and denote its moment estimate by $\hat{\sigma}_\psi^2 = \nabla g(\mathbf{V}_m)\hat{\Sigma}_\psi(\nabla g(\mathbf{V}_m))^T$, where

$$\hat{\Sigma}_\psi = \frac{1}{m}\sum_{i=1}^m (V_{m,i} - \mathbf{V}_m)(V_{m,i} - \mathbf{V}_m)^T.$$

Now, analogously to Proposition 4.1 and 4.2, we have the following.

Proposition 4.5 Fix j and j' with $j + j' \leq J$ and $\omega(j)$ and $\omega'(j') \in \Omega$ non-overlapping, i.e., $\omega(j) \cap \omega'(j') = \phi$. Next, choose ordering relations $\prec \in \Lambda(\omega(j))$ and $\prec' \in \Lambda(\omega'(j'))$.

Let $A, B \in \mathcal{S}(\prec_{\omega(j)})$ and let $D \in \mathcal{S}(\prec'_{\omega'(j')})$. Then, as $m \to \infty$, we have the following limits. (i) Consistency: Expression (4.12) converges a.s. to expression (4.11). That is,

$$\hat{r}_\psi \stackrel{a.s.}{\to} r_\psi.$$

(ii) Asymptotic normality: For $\hat{\sigma}_\psi$ as above,

$$\sqrt{m}(\hat{r}_\psi - r_\psi)/\hat{\sigma}_\psi \stackrel{D}{\to} N(0,1).\Box$$

Finally, as a parallel to Theorems 4.1 and 4.2, we state a result giving an AUMP level α test for CA. Let $R = \max_{\psi \in \Psi} r_\psi$, $\hat{R} = \max_{\psi \in \Psi} \hat{r}_\psi$. Now, testing CA is equivalent to testing $H_{CA} : R \leq 0$, vs $KCA : R > 0$.

As before, let the domain of r_ψ be $\{r_1 < r_2 < \cdots < r_d\}$, $A_d = (r)^{-1}(r_d)$ be the collection of indices ψ at which r_ψ achieves the same maximum value r_d. Let $\mathbf{z}(A_d) = (\hat{r}_\psi : \psi \in A_d)$ with asymptotic variance $\Sigma(A_d)$, and consistent estimator $\hat{\Sigma}(A_d)$. Set $F_{A_d}(x) = \Phi_{\Sigma(A_d)}(x, ..., x)$, with $1-\alpha$ percentile denoted $F^{-1}_{A_d}(1-\alpha)$. These are approximated by $\hat{F}_{A_d}(x) = \Phi_{\hat{\Sigma}(A_d)}(x, ..., x)$, and $\hat{F}^{-1}_{A_d}(1-\alpha)$.

Theorem 4.3 An asymptotic level α test of H_{CA} is given by the rejection rule

$$\sqrt{m}\hat{R} \geq \hat{F}^{-1}_{A_d}(1-\alpha).$$

(ii) When $|A_d| = 1$, an AUMP level α test of H_{CA} is given by the rejection rule

$$\sqrt{m}\hat{R} \geq \Phi^{-1}_{\hat{\Sigma}(A_d)}(1-\alpha).$$

Proof: The proof of this result is similar to the proof of Theorem 4.1 and 4.2 $\Box$

§5. Testing VCD

Recall that Junker and Ellis (1997) characterized monotone unidimensional models in terms of VCD and CA. Since Section 4 gave a test for CA, it remains to give a test for VCD. Note that the asymptotics in VCD are in test length m rather than number of examinees. Here, we identify an AUMP test of VCD for fixed m. This test is similar to that for CA but the construction is simpler because the functions $f(\cdot)$ and $g(\cdot)$ in the covariance are arbitrary. However, this suggests that a larger sample size will be necessary for the asymptotics to be effective.

Let $X_{J,m} = (X_{J+1}, ..., X_{J+m})$, and let $\mathcal{S}_{J,m}$ be the domain of $X_{J,m}$. Let $\mathcal{S}_j$'s, ω and Ω be as in section 4.3. Also write $\mathcal{F}(\omega(j))$ for the collection of all measureable functions of $X(\omega(j))$. Parallel to Prop. 4.3 we have that for any $j = 1, ..., J$, and any $f(\cdot) \in \mathcal{F}(\omega(j))$ there is a sequence $a_i \in I\!R$ for $i = 1, ..., 2^j$ so that

$$f(X(\omega(j))) = \sum_{i=1}^{2^j} a_i \chi_{s_i}(X(\omega(j))),$$

where the s_i's are elements of $\mathcal{S}(\omega(j))$. Letting $\omega^c(j)$ be the complement of $\omega(j)$, we have the following parallel to Prop. 4.4.

Proposition 5.1 VCD is equivalent to the condition that for each m there is an $\epsilon = \epsilon(m)$, with $\epsilon(m)$ going to zero, so that

$$\max_{j,\omega(j),a,b,D} |Cov(\chi_A(X(\omega(j)), \chi_B(X(\omega^c(j))|X_{J,m} \in D)| \leq \epsilon, \tag{5.1}$$

in which the operation $\max_{j,\omega(j),A,B,D}$ denotes the maximum over

$$1 \leq j < J; \quad \omega(j) \in \Omega; \quad A \in \mathcal{S}(\omega(j)); \quad B \in\in \mathcal{S}(\omega^c(j); \quad \text{and} \quad D \in \mathcal{S}_{J,m}.$$

Proof: If VCD holds, take $Y = X(\omega(j)$, $Z = X(\omega^c(j)$, $f(Y) = \chi_A(Y)$ and $g(Z) = \chi_B(Z)$. Then, (5.1) follows from

$$\lim_{m\to\infty} Cov(\chi_A(X(\omega(j)), \chi_B(X(\omega^c(j))|X_{J,m} \in D) \to 0.$$

If (5.1) holds then, for $1 \leq j < J$, $\omega(j), \omega^c(j) \in \Omega$, $f(\cdot) \in \mathcal{F}(\omega(j))$, and $g(\cdot) \in \mathcal{F}_{\omega^c(j)}$, we have

$$|Cov(f(X(\omega(j)), g(X(\omega^c(j))|X_{J,m} \in D)|$$

$$\leq \sum_{r=1}^{2^j} \sum_{t=1}^{2^{j'}} |a_r b_t Cov(\chi_{A_r}(X(\omega(j)), \chi_{B_t} X(\omega^c(j)))|X_{J,m} \in D)| \leq 2^J \overline{a}\overline{b}\epsilon \to 0,$$

as $m \to \infty$, where $j' = J - j$, $\overline{a} = \max\{|a_1|, ..., |a_{2^j}|\}$, and $\overline{b} = \max\{|b_1|, ..., |b_{2^{j'}}|\}$. □

Now, let $\phi = (j, \omega(j), A, B, D)$ and $\Phi = \{\phi : 1 \leq j < J; \quad \omega(j) \in \Omega; \quad A \in \omega(j); \quad B \in \omega^c(j); \quad D \in \mathcal{S}_{J,m}\}$. For $\phi \in \Phi$ consider

$$q_\phi = Cov(\chi_A(X(\omega(j)), \chi_B(X(\omega^c(j))|X_{J,m} \in D). \tag{5.2}$$

For fixed m and ϕ, denote $\mathbf{x}_i = (x_{i,1}, ..., x_{i,J}, x_{i,J+1}, ..., x_{i,J+m})$ where $i = 1, ..., n$. Let $G = G_D = \{i : \mathbf{x}_{J,m} = D\}$, and set $l = |G|$. The averages of examinees' scores over G are $\overline{\chi}_A(D) = (1/l)\sum_{i\in G} \chi_A(\mathbf{x}_i(\omega(j)))$ and $\overline{\chi}_B(D) = (1/l)\sum_{i\in G} \chi_B(\mathbf{x}_i(\omega^c(j)))$. So,

$$\hat{q}_\phi = \frac{1}{l} \sum_{i\in G} (\chi_A(\mathbf{x}_i(\omega(j))) - \overline{\chi}_A(D))(\chi_B(\mathbf{x}_i(\omega^c(j))) - \overline{\chi}_B(D)) \tag{5.3}$$

is an estimator of q_ϕ. Let $\hat{\sigma}_\phi$ be the asymptotic variance of $\hat{q}_\phi$ constructed as in Section 4.3. Parallel to Prop. 4.5 we have the following.

Proposition 5.1. For fixed $\phi \in \Phi$, we have, as $n \to \infty$, (i) Consistency: Expression (5.3) converges to expression (5.2). That is,

$$\hat{q}_\phi \stackrel{a.s.}{\to} q_\phi.$$

(ii) Asymptotic normality: For $\hat{\sigma}_\phi$ as above,

$$\sqrt{n}(\hat{q}_\phi - q_\phi)/\hat{\sigma}_\phi \xrightarrow{D} N(0,1),$$

where σ_ϕ is the asymptotic variance matrix of $\hat{q}_\phi$ specified as that for $\hat{r}_\psi$ in Section 4.3.

Let $H_{VCD(m)}$ be the hypothesis that VCD is true for fixed m. Let $\overline{Q} = \max_{\phi\in\Phi} q_\phi$, $\underline{Q} = \min_{\phi\in\Psi} q_\phi$, $\hat{\overline{Q}} = \max_{\phi\in\Phi} \hat{q}_\phi$ and $\hat{\underline{Q}} \min_{\phi\in\Phi} \hat{q}_\phi$. Now, $H_{VCD(m)}$ is equivalent to: $-\epsilon(m) \leq \underline{Q}$ and $\overline{Q} \leq \epsilon(m)$. Let A_d be as in Theorem 4.3, but with r_ψ replaced by q_ϕ in the definition and let $\mathbf{z}(A_d)$, $\Sigma(A_d)$, $\Phi_{\Sigma(A_d)}$ and F_{A_d} be as in Theorem 4.3. Let B_d be the counterpart of A_d with max replaced by min. Similar to Theorem 4.3, we have the following.

Theorem 5.1 (i) An asymptotic level α test of $H_{VCD(m)}$ is given by the rejection rule

$$\sqrt{n}(\hat{\underline{Q}} + \epsilon(m)) \leq -\hat{F}^{-1}_{B_d}(1-\alpha/2)$$

or

$$\sqrt{n}(\hat{\overline{Q}} - \epsilon(m)) \geq \hat{F}^{-1}_{A_d}(1-\alpha/2).$$

(ii) When $|A_d| = |B_d| = 1$, an AUMP level α test of H is given by the rejection rule

$$\sqrt{n}(\hat{\underline{Q}} + \epsilon(m)) \leq -\Phi^{-1}_{\hat{\Sigma}(B_d)}(1-\alpha/2)$$

or

$$\sqrt{n}(\hat{\overline{Q}} - \epsilon(m)) \geq \Phi^{-1}_{\hat{\Sigma}(A_d)}(1-\alpha/2).\square$$

The test identified in Theorem 5.2 is different from the approach in Bartolucci and Fortina (2000). They defined a desireable property MTP_2 of $\mathbf{X}_J$ (see Definition 1 in Bartolucci and Fortina 2000) and observed that Rosenbaum (1986) showed CA implies MTP_2. Thus, if their test of MTP_2 rejects, a fortiori one can reject CA. Their test is based on an ML approach (likelihood ratio) and converges asymptotically to a mixture of chi-squared distributions. By contrast, ours is based on the asymptotic normality of unbiased estimators.

§6. Discussion

The main contribution of this paper is the conversion of two latent properties into a set of equivalent manifest properties and the provision of a way to get routine hypothesis tests for manifest properties. We gave hypothesis tests for three of the four conditions in our characterization Theorem 3.1 – ACSN, MM, and CA. We also gave a test for VCD, a manifest condition that arises in a different characterization result due to Junker and Ellis (1997). These tests demonstrate the general feasibility of testing manifest conditions by use of best unbiased estimators. Moreover, we have demonstrated a weak optimality for this procedure.

The major limitation of the approach here is that we only have theoretical feasibility for one sequence of tests. In particular, there remains the question of how large a sample size is necessary for the normal approximation to be effective. Rough calculations suggest the sample sizes necessary for the weak optimality shown here to hold approximately are essentially never available in practice. As is suggested in Yuan and Clarke (2000), the enormous sample sizes seem to arise because one is asking for many disjoint occurrences of asymptotic normality. Obviously, asking only for asymptotic normality on the midrange of the test statistic on which one is conditioning will reduce the sample sizes somewhat.

To see the necessity for getting smaller sample size consider the simple minded approach of using Bonferroni and normality for each member in the set over which minimization is done in, for instance, CSN. Studies (see, eg. Port, 1994, p.685) indicate that, in the one dimensional case, for "reasonable" distributions (such as the binomial for instance) the normal approximation of the standardized sum ($\sqrt{m}\frac{\sum_{i=1}^{m} X_i}{m}$) is quite accurate for $m \geq m_0 = 25$. The minimal sample size m_0 and the accuracy δ of the normal approximation are related by $m_0 \propto 1/\delta^2$. For the same level of accuracy in the d-dimensional case, Bonferroni's inequality gives accuracy δ/d for each marginal dimension. This means the minimal sample size m_0 is $m_0 \propto d^2/\delta^2$ in the d-dimensional case, which gets large, fast.

One alternative for seeing if reasonable sample sizes exist is to use a Berry-Esseen bound in place of the asymptotic normality. It is unclear whether this will be better because the Berry-Esseen Theorem uses a stronger mode of convergence. A second issue is that there might be better statistics on which to base hypothesis tests than the ones we used. One of these is a modification of Fisher's exact test. Both of these are discussed in Yuan and Clarke (2000).

Acknowledgements: The authors would like to express their gratitude to two anonymous referees whose comments greatly improved this paper.

References

[1] Bartolucci, F. and Fortina, A. (2000). A likelihood ratio test for MTP_2 with binary variables. *Ann. Statist.* Vol. 28, No. 4, in press.

[2] Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability (Part 5). In F. Lord and M. Novick (Eds.), *Statistical theorems of mental test scores*, (pp. 397-479). Reading, MA: Addison-wesley.

[3] Clarke, B. S. and Ghosh, J. K. (1995). Posterior convergence given the mean. *Ann. Statist.* **23** No. 6: 2116–2144.

[4] Cressie, N. and Holland, P. W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika.* **48**: 129-141.

[5] Ellis, J. and Junker, B. W. (1997). Tail measurability in monotone latent variable models. *Psychometrika* , Vol. 62, No. 4, 495-523.

[6] Holland, P. W. and Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent trait models. *Ann. Statist.* **14**: 1523-1543.

[7] Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika.* **55**: 577-601.

[8] Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **10**: 286-295.

[9] Junker, B. W. and Ellis, J. (1997). A characterization of monotone unidimensional latent variable models. *Ann. Statist.* Vol. 25, No. 3, 1327-1343.

[10] Junker, B. W. (1993). Conditional association, essential independence, and monotone unidimensional item response models. *Ann. Statist.* **21** No. 3: 1359-1378.

[11] Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika.* **56**: 255-278.

[12] Lehmann, E. L. (1986). *Testing Statistical Hypothesis.* Wiley.

[13] Port, S. C. (1994). *Theoretical Probability for Applications.* Wiley.

[14] Rosenbaum, P. R. (1987). Comparing item characteristic curves. *Psychometrika.* **52** No. 2: 217-233.

[15] Serfling, R. J. (1980) *Approximation Theorems of mathematical Statistics*, Wiley

[16] Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika.* **52**: 589-617.

[17] Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika.* **55**: 293-325.

[18] Yuan, A. and Clarke, B. (2000). Manifest characterization and testing for certain latent traits. *Technical Report # 192*, Department of Statistics, University of British Columbia.