(2)

FORMULA SCORING

BASIC THEORY AND APPLICATIONS

Michael V. Levine

Model Based Measurement Laboratory
University of Illinois
210 Education Building
Champaign, IL 61820

December 1989

Prepared under Contract No. N00014-83K-0397, NR 150-518
and No N00014-86K-0482, NR 4421546.

Sponsored by the Cognitive Science Program
Office of Naval Research.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE — Form Approved OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release: distribution unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) Measurement Series 89-1	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Michael V. Levine Model Based Measurement Lab. 6b OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Cognitive Science Program Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) University of Illinois 210 Education Building, 1310 S. Sixth St. Champaign, IL 61820	7b ADDRESS (City, State, and ZIP Code) Code 1142PT 800 North Quincy St. Arlington, VA 22217-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-83K-0397 N00014-86K-0482
8c. ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO.
61153N	RR04204	RR04204-01	NR 150-518 NR 4421546

11 TITLE (Include Security Classification)
Formula Scoring, Basic Theory and Applications

12 PERSONAL AUTHOR(S)
Levine, Michael V.

13a TYPE OF REPORT	13b TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final Report	FROM ______ TO ______	1989, December	

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Latent trait theory, item response theory, formula score, Rasch model, equating, foundations, quasidensities, densities, non-parametric density estimation, (continued).
05	09		

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Formula scoring is a systematic study of measurement statistics expressed as sums of products of item scores. The theory is currently being used to compute non-parametric estimates of ability distributions, item response functions, and option response functions. The theory has been used to design algorithms for estimating item response functions from adaptive test data (on-line calibration), monitoring and correcting drift in observed score distributions for adaptive tests (on-line equating), computing optimal tests for cheating, and combining appropriateness measurement information from several subtests. In this paper a portion of the theory is developed from a few principles. Applications are considered to the problems of deciding whether ability has the same distribution in two demographic groups, to finding latent class models that are equivalent to item response models, and to controlling drift in adaptive testing programs. Keywords:

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Charles Davis	22b TELEPHONE (Include Area Code) 202-696-4046 22c OFFICE SYMBOL ONR 1142CS

DD Form 1473, JUN 86 — Previous editions are obsolete — SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

UNCLASSIFIED

18. continued ...
ability distributions, identifiability. (S!)

FORMULA SCORING

BASIC THEORY AND APPLICATIONS

Michael V. Levine

Model Based Measurement Laboratory
University of Illinois
210 Education Building
Champaign, IL 61820

December 1989

Prepared under Contract No. N00014-83K-0397, NR 150-518
and No. N00014-86K-0482, NR 4421546.

Sponsored by the Cognitive Science Program
Office of Naval Research.

ABSTRACT

Formula scoring is the systematic study of measurement statistics expressed as linear combinations of products of item scores. The theory is currently being used to compute non-parametric estimates of ability distributions, item response functions, and option response functions. The theory has been used to design algorithms for estimating item response functions from adaptive test data (on-line calibration), monitoring and correcting drift in observed score distributions for adaptive tests (on-line equating), computing optimal tests for cheating, and combining appropriateness measurement information from several subtests. In this paper a portion of the theory is developed from a few principles. Applications are considered to the problems of deciding whether ability has the same distribution in two demographic groups, to finding latent class models that are equivalent to item response models, and to controlling drift in adaptive testing programs.

FORMULA SCORING
BASIC THEORY AND APPLICATIONS

Preface

For several years, Bruce Williams and I have been presenting applications of a new approach to measurement, which we call *formula scoring*. Our presentations to the annual ONR Contractor's Conferences have been punctuated with the phrase, "It can be shown ... ." This technical report begins a series of papers providing proofs of these claims. An attempt will be made to derive formula score theory from a few basic principles.

This version of the report is being used to introduce graduate students to the work in our laboratory. Very explicit, computational proofs are provided for some basic results. A shorter version is being prepared for publication.

Thanks to Bruce Williams and Fritz Drasgow there are many data-based applications[1] of formula scoring, which are now starting to appear in print[2]. The data-based applications are not suitable for motivating this paper because Bruce's programs use concepts that are developed in later papers. Therefore an alternative way to motivate the report had to be found.

Three examples of results that can be obtained with the theory have been selected to motivate the theory. I don't think the results would have been discovered without the theory. Each seems surprising - at least to me - and somewhat contrary to conventional psychometric wisdom. Each result can be easily proven with the theory. And each result seems hard to prove without reproducing the reasoning in the theory.

Some Examples to Motivate the Theory

Formula score theory can be used to derive some unexpected, hopefully useful, consequences of the assumptions of item response theory. Three examples follow.

The examples are valid for parametric and non-parametric item response models. Except where noted, the results hold for all "continuous, one-dimensional, probabilistic item response models for bounded abilities." Thus, item response functions are permitted to have any shape, provided they are continuous functions of one variable with values strictly between zero and one. The cumulative distribution of ability also is permitted to have any shape, provided there is some - possibly very large - interval such that the distribution is zero or one outside this interval.

Example One: Checking for ability distribution differences

A quick way to recognize ability distribution differences is to check average tests scores. Thus, if girls on the average have higher test scores than boys on an unbiased test it is safe to conclude that ability is distributed differently among girls and boys. The converse obviously is not true because very different distributions may have same mean.

Using observed scores to check for group ability differences is believed to be uniquely uncomplicated for the Rasch model. Since the number right score is a sufficient statistic for estimating ability it might be expected that it is possible to determine the presence or absence of group ability differences by comparing distributions of number right score. This (incorrect) assertion can also be expressed as follows:

> There is a set of statistics $X_0, X_1, \ldots X_n$ such that the group ability distributions are different if and only if at least one of the statistics has different expected values among girls and boys.

Here n is the number of items on the test, and X_j is the statistic which is one if exactly j items were answered correctly and zero otherwise.

The theory shows that the Rasch model is *not* unique in having a small number of diagnostic statistics. The theory also shows what can *and cannot* be concluded when corresponding pairs of expectations are equal.

For *any* item response model, Rasch model or other, there is a set of statistics $X_1, X_2, \ldots X_J$ such that if at least one pair of corresponding expected values differ, then the group ability distributions are different. But if corresponding expected values are equal, then the distributions still may be different. However, it can be shown that no statistical test (using only the answers to the n items for data) exists that can demonstrate the difference! In particular, for a test satisfying the Rasch model if boys and girls have equal expected X_j's , then ability may be distributed differently in the two populations, but no analysis of test data can be used to demonstrate the difference. Details follow the proof of Theorem One.

Recall that for the Rasch model each item response function P_i

$$P_i(t) = \text{Prob}(\text{correct answer for item } i \mid \text{ability} = t)$$

can be written in the form $P_i(t) = [1+e^{-(t-b_i)}]^{-1}$ for some constant b_i . To avoid mathematical digressions irrelevant to the main points of this paper, it will generally be assumed that for $i \neq j$, $b_i \neq b_j$. Thus no two Rasch model items have exactly the same item response function.

As an example of another model having a small set of diagnostic statistics, consider the generalization of the Rasch model having item response functions given by the following equation

$$P_i(t) = c_i + (1-c_i)[\ 1 + e^{-a(t-b_i)}]^{-1} .$$

As with the Rasch model, it will generally be assumed that different items have different difficulties. Thus if $i \neq j$, $b_i \neq b_j$. For this model J is less than or equal to the number of items, and X_j can be taken to be the score that is one if item j is answered correctly and zero otherwise. (If for some $i \neq j$, $b_i = b_j$, then a somewhat more complicated set of X_j must be used, but J is still small.)

Incidentally, these results are related to the identifiability of ability distributions. Since different distributions can give the same vector of expected X_j's, the ability distribution is not identifiable, even when the item response functions are completely specified.

Example Two: How to turn an item response model for an ability continuum into an isomorphic latent class model with finitely many classes

Suppose we are given an item response model with continuous item response functions $\neq 0,1$ and a continuous ability density f. Using the theoretical results in this paper it can be shown that it is possible to select abilities $t_0 < t_1 < \ldots t_J$ and numbers $p(t_0), p(t_1), \ldots p(t_J)$ such that for each item response pattern u^*, the "manifest probability"

Prob(Sampling an examinee with item response patttern u^*) ,

which is ordinarily computed by i tegrating the likelihood function,

$$\int_c^d \text{lik}(u^* \mid \text{ability} = t)\, f(t)\, dt$$

can be computed by evaluating the sum

$$\sum_{j=0}^{J} \text{lik}(u^* \mid \text{ability} = t_j)\, p(t_j) \ .$$

For the item response functions given by the formulas in Example One, J can by set equal to the number of items.

Since the manifest probabilities sum to one, $\Sigma\, p(t_j) = 1$. Thus if $p(t_j) \geq 0$ for $j \leq J$, we have a latent class model with $J+1$ classes that is isomorphic to the continuous latent trait model.

I haven't found a simple proof based only on the results in this paper of the existence t_j with $p(t_j) \geq 0$. However the result also is true and is proven in next paper in this series. In any event, even when some of the $p(t_j)$ are negative the result seems able to greatly reduce computation times in some applications noted below.

Example Three: On-line equating or Simulation results without simulation

Consider two subtests, say, word knowledge (WK) and arithmetic reasoning (AR), of a computer administered adaptive test such as the adaptive version of the Armed Services Vocational Aptitude Battery (ASVAB). Suppose the item pool for WK has just been changed by introducing some new items that haven't been ministered often enough to highly motivated examinees to have well estimated item response functions. To analyze and control the effect of the new items on the distribution of an observed score $\hat{\theta}_{WK}$ we wish to calculate three functions, usually computed by simulation:

$$F_1(t) = \text{Expectation } \{ \hat{\theta}_{WK} \mid \theta_{WK} = t \}$$
$$F_2(t) = \text{Variance } \{ \hat{\theta}_{WK} \mid \theta_{WK} = t \}$$
$$P(x|t) = \text{Prob } \{ \hat{\theta}_{WK} \leq x \mid \theta_{WK} = t \} .$$

F_1 and F_2 show how the first two conditional moments of the observed score are affected by the new items and can be used to make corrections. For example, if $F_2(-1)$ is observed to increase very much when the new items replace easy old items then countermeasures such as adding more easy items can be tried. $P(x|t)$ provides the remaining moments. It can be used to predict how the marginal distribution of $\hat{\theta}_{WK}$ will be affected by future changes in the ability distribution.

Since the item response functions for the new items are not known, simulation is not possible. (When the score $\hat{\theta}_{WK}$ is a Bayes mode or maximum likelihood ability estimate, then item parameter estimates derived from small samples of not highly motivated examinees may be used to compute the score, but such estimates are not suitable for including in a simulation.) Thus, the following result is of interest.

It is generally possible to use the item response functions for the *old* WK items to compute functions $c_0(t), c_1(t), \ldots c_K(t)$ and to sort examinees into groups using only an AR score $\hat{\theta}_{AR}$. According to the theory, the conditional expectation of $\hat{\theta}_{WK}$ (computed from item scores for both old and new items) can be calculated with the formula

$$\text{Expectation} (\hat{\theta}_{WK} \mid \theta_{WK}=t) =$$

$$\sum_{k=0}^{K} c_k(t) \text{ Expectation } (\hat{\theta}_{WK} \mid \hat{\theta}_{AR} \text{ is in the } k\textit{th} \text{ score group}) .$$

In words, we use $\hat{\theta}_{AR}$ to group examinees and then compute the conditional expected WK score as a linear combination $\hat{\theta}_{WK}$ group averages. The $\hat{\theta}_{WK}$ score is computed using item scores for both old WK items and new WK items. However, only the well estimated old WK item response functions are used to compute the coefficients of the linear combination. In this way the effect of introducing new items on an observed score *at each ability level* can be calculated from actual data. Since the method does not use item parameter estimates for the new items, it is not adversely affected by item parameter estimation error on the new items.

A similar formula gives the conditional variance since for the same c_j and groups

$$\text{Expectation} (\hat{\theta}^2_{WK} \mid \theta_{WK}=t)$$

$$= \sum_{k=0}^{K} c_k(t) \text{ Expectation}(\hat{\theta}^2_{WK} \mid \hat{\theta}_{AR} \text{ is score group } k) .$$

Finally, for the random variable defined by

$$X = \begin{cases} 1 \text{ if } \hat{\theta}_{WK} \le x \\ 0 \text{ otherwise} \end{cases}$$

the conditional distribution of $\hat{\theta}_{WK}$ is given by

$$\text{Prob}\{ \hat{\theta}_{WK} \le x \mid \theta_{WK} = t\} = \text{Expectation}\{ X \mid \theta_{WK} = t\}$$

$$= \sum_{k=0}^{K} c_k(t) \text{ Expectation}\{ X \mid \hat{\theta}_{AR} \text{ is in group } k \} \quad .$$

The calculation of these three conditional expected values illustrates a more general result described in the discussion of "quasidensities" (Section Two, below).

NOTES

1. Formula score theory currently is being used to compute non-parametric maximum likelihood estimates of ability distributions, item response functions, and option response functions. The theory has been used to design algorithms for estimating item response functions from adaptive test data without interrupting testing (on-line calibration), to compute optimal tests for cheating, and to combine appropriateness measurement information from several subtests. The theory yields measures of item bias and test dimensionality. The theory seems to lead to a tractible, non-parametric, multidimensional item response theory, which is currently being developed. The theory is also being applied to what might be called "online equating," i.e., monitoring and correcting changes in the distribution of observed scores for an adaptive test as the test's item pool is replenished.

2. Drasgow, F., Levine, M.V., Williams, B., McLaughlin, M.E., and Candell, G.L. Modelling incorrect responses with multilinear formula score theory. Applied Psychological Measurement, In press, 1989; Drasgow, F., Levine, M.V., and McLaughlin, M.E. Multitest extensions of appropriateness indices. Applied Psychological Measurement, accepted for publication, 1989.

Section One

Formula Score Theory and Equivalent Distributions

Formula score theory systematically studies measurement statistics expressed as linear combinations of products of item scores. The theory begins with an equivalence relation on ability distributions.

We consider a fixed test of n items. A pair of distributions F and G are defined to be *equivalent* relative to the test if every statistic computed from the test's item scores has the same distribution under the hypothesis

$$H_0\text{: Ability has cumulative distribution } F$$

as under the alternative hypothesis

$$H_1\text{: Ability has cumulative distribution } G\ .$$

Notice that there is no way whatsoever to use item responses on the test being analyzed to distinguish between a pair of equivalent distributions. For if F is equivalent to G and if the statistic X is used for hypothesis testing, then decisions based on X will be no more valid than decisions based on the flip of a coin or other irrelevant random process.

Notice also that equivalence is defined relative to a fixed test of specified items. Thus a pair of distributions may be equivalent relative to the test, but distinguishable if one more item is added to the test. In fact, if one of the items is replaced by a slightly different item, the equivalence relation may be changed. This is a significant limitation of the present algebraic version of the theory. Later papers on applications use metric concepts to get around this problem.

The main result of this section is a characterization of equivalent distributions in terms of the expected values of finitely many statistics. Comments on implications and applications of this result are at the end of this section.

Item Response Theory and Formula Score Theory

To make the paper more nearly self-contained and to make explicit just what assumptions of item response theory are used to prove the new results, we begin with some definitions from item response theory.

An item response model provides a probability measure for set $\{a\}$, which is interpreted as a set of possible or actual examinees. There are two types of random variables in item response theory: observed *item scores* $u_1(a)$, $u_2(a)$, ... $u_n(a)$ and unobserved *abilities* $\theta(a)$. Item scores are either one or zero. "$u_i(a)=1$" is interpreted as "examinee a successfully answered item i ."

In this paper, the abilities $\theta(a)$ are numbers. However, after some routine changes, all of the results in this paper and their proofs generalize to multidimensional abilities, i.e., vector-valued $\theta(a)$'s .

Item response theory relates item scores to abilities with functions P_i called *item response functions*

$$P_i(t) = \text{Prob}(\ u_i=1|\theta=t\ ) \ .$$

$P_i(t)$ is interpreted as the probability of observing $u_i(a) = 1$, when examinee a is sampled from all those with ability t .

In this paper, details about the item response functions are generally left unspecified. Only continuity and a weak condition, $0 < P_i(t) < 1$, are assumed. These conditions are also implied by the parametric formulas of most item response models.

Formula scoring differs from much of item response theory on the domain of definition of the item response functions. In item response models $P_i(t)$ is usually defined for all numbers t, despite the fact that the models predict essentially the same behavior from examinees with ability 20 and 20,000 and despite the fact that applications of the parametric models usually proceed as if abilities were bounded.

In this section the domain of definition of the item response functions can be bounded or unbounded. However, in the following sections $P_i(t)$ is defined only for t in an interval of finite length. Some discussion of this point is at the end of this section.

The main assumption of item response theory is *local independence*. It asserts that item responses are conditionally independent, i.e., for any sequence of zeros and ones

$$u_1^*, u_2^*, \ldots u_n^*$$

and any ability t

$$\text{Prob}\{ u_1=u_1^* \ \& \ u_2=u_2^* \ \ldots \ u_n=u_n^* \mid \theta = t\} = \Pi_i \ \text{Prob}\{ u_i=u_i^* \mid \theta=t\} \ .$$

In item response theory analyses of data, the item responses are recorded and inferences are made about θ. Only the item responses are observed. Thus if the word "statistic" is to be reserved for random variables that are functions of the observables, only functions of the u_i are statistics. Since the range of each u_i is finite, every function of the u_i is a random variable. *Thus* X *is a statistic if and only if* X *is a function of item scores.*

The set of all statistics for a test is obviously a vector space since a linear combination of functions of item scores is a function of item scores. Since the u_i take on only finitely many values, every statistic can be written as a polynomial in the item scores. In fact, since $u_i^2 = u_i$

every statistic is a linear combination of the following statistics, which are called *elementary formula scores*,

$$1$$

$$u_1,\ u_2,\ \ldots\ u_n$$

$$u_1u_2,\ u_1u_3,\ \ldots\ u_{n-1}u_n$$

$$\cdot\ \cdot\ \cdot$$

$$\prod_{i=1}^{n} u_i\ .$$

Thus the elementary formula scores, or some subset of the these scores, form a basis for the vector space of all statistics. Since there are finitely many (2^n) elementary formula scores, *the set of all statistics is a finite dimensional vector space.*

The *regression function* $R_X(\cdot)$ or conditional expectation function of a statistic X

$$R_X(t) = E(X|\theta=t)$$

expresses the conditional expected value of the statistic as a function of ability. Since every statistic is a linear combination of the elementary formula scores, local independence implies that each regression function can be written in at least one way as a linear combination of the following functions

$$1$$

$$P_1(t),\ \ldots\ P_n(t)$$

$$P_1(t)P_2(t),\ P_1(t)P_3(t),\ \ldots\ P_{n-1}(t)P_n(t)$$

$$\cdot\ \cdot\ \cdot$$

$$\prod_{i=1}^{n} P_i(t)\ .$$

The central concept of formula score theory is the canonical space. The *canonical space* (CS) of a test is the vector space of regression functions of statistics. Obviously it is the vector space spanned by the *square-free monomials*, i.e. the products of item response functions without repeated factors, listed above. Thus, *the canonical space is a finite dimensional vector space of continuous, real-valued functions.*

An Alternative Characterization of Equivalent Distributions

Using the canonical space it is possible to derive a simpler test for equivalent distributions. The definition would have us check the *distribution* of *every* statistic. It will be shown that only *finitely* many statistics need to be considered and that all that needs to be known about each statistic is its expected value. First, some notation.

F will be used in all sections of this paper to denote the (generally unknown) ability distribution. For any statistic X and number x, the distribution function of X evaluated at x can be written

$$\mathrm{Prob}(X \leq x) = \int \mathrm{Prob}(X \leq x \mid \theta = t)\, dF(t) \ .$$

If G is F or any other distribution, then the distribution of X relative to G evaluated at x will be denoted by $P(x;X,G)$. Thus

$$P(x;X,G) = \int P(X \leq x \mid \theta = t)\, dG(t) \ .$$

Similarly, the expected value of X and the expected value of X relative to distribution G are denoted by

$$E(X) = \int E(X \mid \theta = t)\, dF(t)$$

$$E(X;G) = \int E(X \mid \theta = t)\, dG(t) \ .$$

Using this notation the definition of equivalent distributions given earlier can be succinctly expressed: *Two distributions* F_1 *and* F_2 *are* *equivalent if for all statistics* X *and real* x

$$P(x;X,F_1) = P(x;X,F_2) \; .$$

Theorem One is an alternative characterization of equivalent distributions.

<u>Theorem One</u>: <u>Let</u> J+1 <u>be the dimension of the canonical space</u>. <u>Then there are</u> J <u>statistics</u> $X_1, X_2, \ldots X_J$ <u>such that</u> F_1 <u>is equivalent to</u> F_2 <u>if and only if</u>

$$E(X_j;F_1) = E(X_j;F_2) \quad \underline{\text{for}} \;\; j=1, \ldots , J \; .$$

<u>Furthermore</u>, <u>if</u> $Y_0, Y_1, \ldots Y_J$ <u>are any statistics with linearly independent regression functions</u>, <u>then</u> F_1 <u>is equivalent to</u> F_2 <u>if and only if</u> $E(Y_j;F_1) = E(Y_j;F_2)$ <u>for</u> $j=0, 1, \ldots J$.

Proof: Let $h_0, \ldots h_J$ be a basis for the canonical space. Since the constant function is in the CS, h_0 can be taken to be the constant function, $h_0(t) = 1$. Since the h_j are in the CS, there are statistics X_j such that $h_j(t) = E(X_j|\theta=t)$ for $0 \le j \le J$. For any statistic X and real x , the regression function of the indicator random variable, χ

$$\chi = \begin{cases} 1, \text{ if } X(u_1, \ldots , u_n) \le x \\ 0, \text{ if } X(u_1, \ldots , u_n) > x \end{cases}$$

is in the canonical space and consequently can be written

$$E(\chi|\theta=t) = \sum_{j=0}^{J} \alpha_j h_j(t) \; .$$

Therefore for $i=1,2$

$$\begin{aligned} P(x;X,F_i) &= \int \Sigma_j \alpha_j h_j(t) dF_i(t) \\ &= \Sigma_j \alpha_j E(X_j;F_i) \; . \end{aligned}$$

Since $E(X_0;F_1) = \int 1 \; dF_1(t) = 1 = E(X_0;F_2)$,

$$E(X_j;F_1) = E(X_j;F_2) \quad \text{for} \quad j=1, \ldots, J$$

implies that F_1 and F_2 are equivalent. Conversely, each X_j can be written as a sum of products of the binary item scores,

$$X_j = \sum_{\nu=1}^{2^n} a_\nu v_\nu$$

where $v_1, v_2, \ldots v_\nu, \ldots v_{2^n}$ is an enumeration of the 2^n elementary formula scores. Since v_ν is either zero or one, for $i=1$ or 2

$$E(v_\nu;F_i) = 1-P(0;v_\nu,F_i) .$$

Therefore "F_1 is equivalent to F_2" implies

$$\begin{aligned} E(X_j;F_1) &= \sum_\nu a_\nu E(x_\nu;F_1) \\ &= \sum_\nu a_\nu [1-P(0;v_\nu,F_1)] \\ &= E(X_j;F_2) . \end{aligned}$$

Finally, if $J+1$ statistics Y_j have linearly independent regression functions g_j then for some non-singular $(J+1)\times(J+1)$ matrix $A=(a_{ij})$, $g_j(\cdot)=\sum_k a_{jk}h_k(\cdot)$. The remainder of the proof follows routinely from $E(Y_j;F_i) = \sum_k a_{jk}E(X_k;F_i)$ for $j=0, 1, \ldots J$ and $i=1,2$.

Implications and Applications

The theorem has negative implications for distribution estimation. We have observed that when J is small, two distributions with clearly different shapes can be equivalent. As noted in Example Two a discrete distribution on a few points may turn out to be indistinguishable from a distribution with a continuous density. Thus, even when item response functions are known, it is not possible to consistently estimate the ability distribution without additional assumptions.

Note that for some applications it is valuable to know that ability distributions are equivalent. Returning to Example One of the Prefa -, if the ability distributions for boys and girls are equivalent relative to the test, then any selection procedure based on test results is as likely to select a boy as a girl.

The theorem shows, as was asserted in Example One, that by checking finitely many pairs of expected values, a difference between the ability distributions can be demonstrated. In Section 3 it is shown that J can be small. For the Rasch model and its generalization, J can be taken equal to the number of test items and X_j can be taken to be the *jth* item score. Thus a necessary and sufficient condition for there to be a *demonstrable* difference between distributions is that there be at least one item on which the proportion of boys passing the item is different from the proportion of girls.

For other models J can be large and the X_j may be complicated. Models with large J are discussed in Section 4. The task of computing J and X_j is also discussed in Section 4.

Example Two illustrates a second situation in which distribution equivalence may have practical importance. In Example Two we considered replacing an ability distribution having a continuous density with a step function having finitely many steps. The goal in doing so was to reduce integrals to sums. (In Section 3 a procedure for calculating the location and size of the steps is described.) In optimal appropriateness measurement[1] it is necessary to integrate over ability to obtain a uniformly most powerful test for cheating and other forms of aberrance. Even for unidimensional tests a great deal of computing is required to compute the theoretical manifest probabilities in Example Two. For multidimensional tests and "multi-unidimensional" test batteries such as ASVAB considerably more computation is required.

So far we have successfully avoided computing multiple integrals in our analyses of test batteries in which each subtest measures a different ability[2] by using approximations. The results in this section indicate an alternative, more general way[3] to calculate probabilities. Since an integral must be evaluated for each of thousands of examinees and since multivariate quadrature requires a lot of computation, replacing a continuous multivariate with an equivalent discrete distribution on a small number of points is very desireable.

This section is concluded with comments on the issue of bounded and unbounded ability continua, which is raised by Theorem One.

Why Bounded Abilities

Sometimes whatever is being measured by a test is intrinsically bounded. Adding extremely hard items to a test generally changes what is being measured and may cause a test to fail to be unidimensional. Thus a calculus item is not a very hard arithmetic item but an item measuring an ability or achievement other than what is being measured by a grade school subtraction test. At the other extreme, a child totally ignorant of subtraction occupies a lower end point on the measurement scale.

Theorem One raises questions about the domain of definition of the P_i and also motivates considering bounded continua. Suppose that on a particular test no examinee has an ability outside the interval $[-5,5]$. Then there can be a pair of *inequivalent* distributions F_1 and F_2 such that $F_1(t) = F_2(t)$ for $|t| \leq 5$, even though no empirical study can distinguish between F_1 and F_2 . This awkward situation can be kept from occuring by defining the item response functions as functions of abilities in $[-5,5]$. If the P_i are defined only for $|t| \leq 5$, then the CS becomes a set of functions defined on an interval. Distributions that agree on the interval will then be equivalent in the sense of Theorem One as well as in the intuitive sense. Thus by treating the P_i as functions of a *bounded* variable the intuitive and technical meanings of "equivalent" can be brought closer together. Alternatively, attention can be limited to ability distributions that are zero or one outside this interval. Both options are developed in the next section.

The assumption of boundedness turns out to be very weak. In any practical measurement situation, it can be trivially satisfied by considering a very large interval, an interval so large that the probability of sampling an examinee outside the interval for all practical purposes is zero. For theoretical work, boundedness can be imposed on a test model by transforming abilities without affecting the only assumptions being made about item response functions: continuity and $0 < P_i(t) < 1$.

NOTES

1. Levine, M.V. and Drasgow, F., Optimal Appropriateness Measurement. Psychometrika, 1989.

2. Drasgow, F., Levine, M.V., and McLaughlin, M.E. Multitest extensions of appropriateness indices. Applied Psychological Measurement, accepted for publication, 1989.

3. The method can be thought of as a quadrature technique developed for evaluating the integrals that occur in psychometric applications. The selection of the quadrature points and weights is discussed in Section 3. Each quadrature formula is exact for some set of integrands. The new method is exact for integrating functions in the CS.

Section Two

An Inner Product and Quasidensities

When abilities are bounded, the CS has an inner product with a simple statistical interpretation. And each distribution function can be treated as if it had a continuous derivative. This "derivative," the quasidensity, is the subject of this section.

In the remainder of this paper it will be assumed that there are numbers $c<d$ such that Prob $(c \leq \theta \leq d) = 1$. Item response functions will be treated as functions defined on $[c,d]$, and the canonical space will be a set of functions defined on $[c,d]$. After these changes are made the function $<\cdot,\cdot>$ defined on pairs of functions f,g in the CS by

$$<f,g> = \int_c^d f(t)g(t)dt$$

becomes an inner product.

Note that when the ability distribution has a density and this density is in the CS, then the inner product has a statistical interpretation. For if $R(t) = E(X|\theta=t)$ is the regression function of a statistic X and if the ability distribution has a density f also in the CS, then $<R,f>$ is the expectation of X . The major result of this section is to generalize this property to situations in which the ability density is not in the CS and to situations in which the ability distribution is not differentiable. It will be shown that there is a *unique* continuous function g in the CS such that for all statistics X

$$\begin{aligned} E(X) &= \int_c^d E(X|\theta=t)\ dF(t) \\ &= \int_c^d E(X|\theta=t)\ g(t)dt \\ &= <R_X,g> \ . \end{aligned}$$

<u>Theorem Two</u>: <u>If</u> $P(c \leq \theta \leq d) = 1$, <u>then there is a unique continuous function</u>

g <u>in the CS such that for every statistic</u> X

$$E(X) = \int_c^d E(X|\theta=t)\ g(t)dt\ .$$

<u>Proof</u>: Let $h_0, h_1, \ldots h_J$ be an orthonormal basis for the CS relative to its inner product $<\cdot,\cdot>$. Thus $<h_i,h_j> = 1$ or zero according to whether $i =, \neq j$. For each $j \leq J$ a statistic X_j can be found such that $E(X_j|\theta=t) = h_j(t)$ because every function in the CS is the regression function of at least one statistic. Let X be any statistic and R_X its regression function. Since the h_j form a basis for the CS, R_X can be written

$$R_X(\cdot) = \Sigma_j\ b_j h_j(\cdot)$$

for some constants b_j . Since the h_j are orthonormal $<R_X,h_j> = b_j$ and

$$R_X(\cdot) = \Sigma_j\ <R_X,h_j>h_j(\cdot)\ .$$

Consequently

$$\begin{aligned}
E(X) &= \int_c^d R_X(t)\ dF(t) \\
&= \int_c^d \Sigma_j\ <R_X,h_j>h_j(t)\ dF(t) \\
&= \Sigma_j\ <R_X,h_j> \int_c^d h_j(t)dF(t) \\
&= \Sigma_j\ <R_X,h_j>E(X_j) \\
&= \Sigma_j \int_c^d R_X(t)h_j(t)dt\ E(X_j) \\
&= \int_c^d R_X(t)\ \Sigma_j\ E(X_j)h_j(t)dt \\
&= \int_c^d E(X|\theta=t)g(t)dt
\end{aligned}$$

for $g = \Sigma\ E(X_j)h_j(\cdot)$ in the CS.

To prove uniqueness, suppose that for some h in the CS

$$E(X) = \int_c^d R_X(t)h(t)dt$$

for all statistics X . Since the h_j form a basis, $h(\cdot) = \Sigma\ \alpha_j h_j(\cdot)$ for some constants α_j . Since the h_j are orthonormal, for $X = X_j$

$$\begin{aligned} E(X_j) &= \int_c^d R_{X_j}(t)h(t)dt \\ &= \int_c^d h_j(t) \sum_k \alpha_k h_k(t)dt \\ &= \sum_k \alpha_k \langle h_j, h_k \rangle \\ &= \alpha_j \ . \end{aligned}$$

Thus $h=g$, as was to be proven.

If $G=F$ or any other distribution function, then G will be called a *distribution on [c,d]* if for $t<c$, $G(d) - G(t) = 1$. If G is F or any other distribution on $[c,d]$ then a function g in the canonical space is called the *quasidensity*[1] *for G* if for all statistics X

$$E(X;G) = \int_c^d E(X|\theta=t)\ g(t)dt \ .$$

Note that Theorem Two implies that every distribution on $[c,d]$ has a unique quasidensity. Furthermore the proof shows that the quasidensity for G can be written as

$$g(\cdot) = \sum_{j=0}^{J} E(X_j;G)h_j(\cdot)$$

where $\{h_j\}_{j=0}^J$ is any orthonormal basis for the CS and each X_j satisfies $R_{X_j} = h_j$. Since the quasidensity is unique, the choice of the orthonormal basis and statistics X_j used in the formula is inconsequential.

At the end of this section some facts about quasidensity densities are listed and proven. The quasidensity for the unit step at -1 is shown to have the simple form $g(t) = \sum_{j\leq J} h_j(-1)h_j(t)$ where $\{h_j\}_{j=0}^J$ is any orthonormal basis for the CS. This formula was used to compute an approximation to the quasidensity for the unit step at -1 . The first 19

h_j's for 100 three parameter logistic items by the methods in Section 4. Figure One shows the graph of $q(t) = \sum_{j \le 18} h_j(-1)h_j(t)$. If $q(t)$ is multiplied times any of the 100 logistic functions and integrated, the result should be very close to $P_i(-1)$. $|P_i(-1) - \int_c^d P_i(t)q(t)\,dt|$ was found to be generally small, as shown in Table One.

For shorter tests, the quasidensity of the unit step function can be computed without approximation. The graph shown in Figure One is typical.

The precision of the approximation shown in Table One serves to illustrate a point developed in Section Four: For some purposes, high dimensional canonical spaces can be approximated by much lower dimensional spaces.

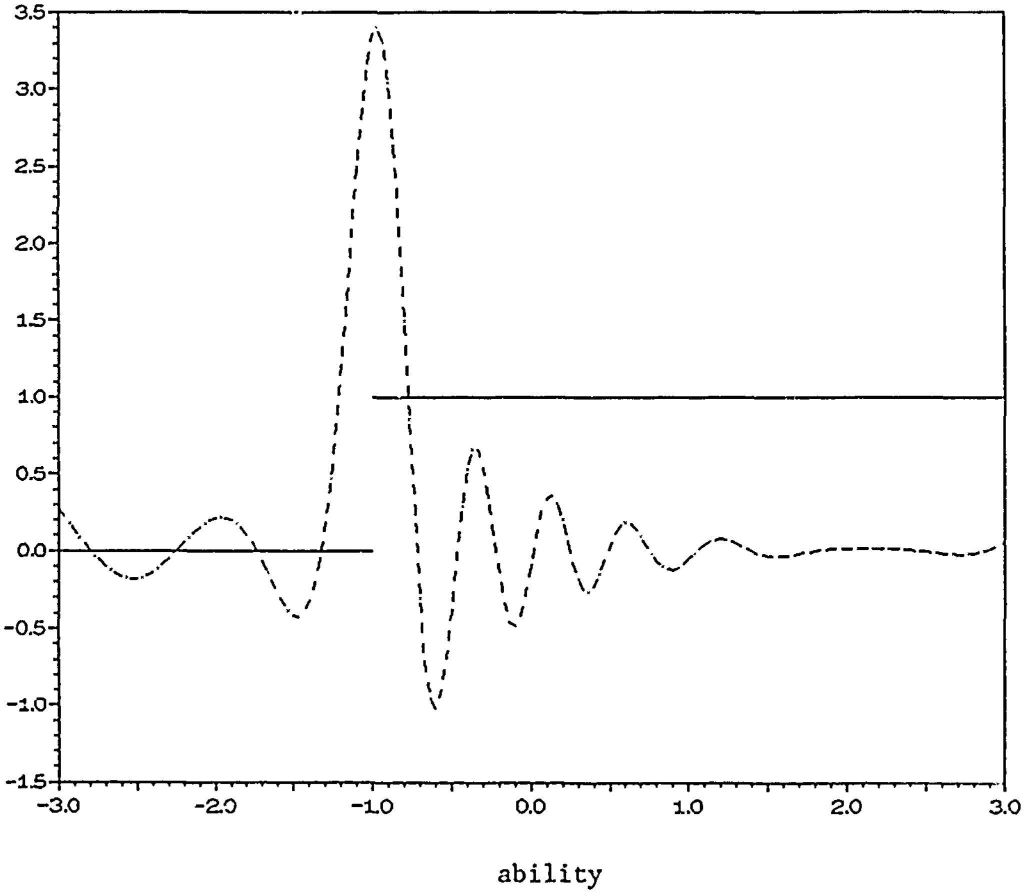

Figure One: Cumulative distribution function for the unit step function at $\theta = -1$ and its quasidensity

Table One: $P_i(-1)$ and an Approximation

item	$P_i(-1)$	$\int P_i q$	diff	item	$P_i(-1)$	$\int P_i q$	diff
1.	.1223	.1223	.0000	2.	.0601	.0601	.0000
3.	.0852	.0852	.0000	4.	.1639	.1639	.0000
5.	.1449	.1449	.0000	6.	.1878	.1878	.0000
7.	.2958	.2958	.0000	8.	.2058	.2058	.0000
9.	.2601	.2601	.0000	10.	.3345	.3345	.0000
11.	.2380	.2380	.0000	12.	.2093	.2093	.0000
13.	.3024	.3023	.0001	14.	.2965	.2965	.0000
15.	.3385	.3385	.0000	16.	.4869	.4869	.0000
17.	.2798	.2795	.0003	18.	.7576	.7575	.0001
19.	.4482	.4483	.0000	20.	.8665	.8665	.0000
21.	.7634	.7634	.0000	22.	.9014	.9012	.0002
23.	.7804	.7804	.0000	24.	.9054	.9054	.0000
25.	.8695	.8696	.0000	26.	.1391	.1391	.0000
27.	.2832	.2832	.0000	28.	.2334	.2334	.0000
29.	.1463	.1463	.0000	30.	.1504	.1504	.0000
31.	.1396	.1396	.0000	32.	.1374	.1374	.0000
33.	.2578	.2578	.0000	34.	.2314	.2313	.0001
35.	.2262	.2262	.0000	36.	.1881	.1880	.0000
37.	.2521	.2521	.0000	38.	.2788	.2788	.0001
39.	.3256	.3256	.0000	40.	.2676	.2673	.0003
41.	.3734	.3734	.0000	42.	.5322	.5322	.0000
43.	.6150	.6149	.0001	44.	.6617	.6614	.0003
45.	.7948	.7948	.0001	46.	.7852	.7851	.0001
47.	.7835	.7835	.0000	48.	.8159	.8159	.0000
49.	.8228	.8227	.0001	50.	.9064	.9062	.0001
51.	.1133	.1133	.0000	52.	.0662	.0662	.0000
53.	.0605	.0605	.0000	54.	.2013	.2013	.0000
55.	.2024	.2024	.0000	56.	.2697	.2697	.0000
57.	.3809	.3809	.0000	58.	.1809	.1809	.0000
59.	.3495	.3495	.0000	60.	.3370	.3370	.0000
61.	.1521	.1521	.0000	62.	.2812	.2812	.0000
63.	.2931	.2931	.0000	64.	.2673	.2673	.0000
65.	.2569	.2569	.0000	66.	.3876	.3876	.0000
67.	.4459	.4459	.0000	68.	.6903	.6903	.0000
69.	.6179	.6179	.0000	70.	.8457	.8454	.0003
71.	.7718	.7718	.0000	72.	.7427	.7427	.0000
73.	.8167	.8167	.0000	74.	.8800	.8800	.0000
75.	.8775	.8774	.0000	76.	.1406	.1406	.0000
77.	.2074	.2074	.0000	78.	.2022	.2022	.0000
79.	.0660	.0660	.0000	80.	.2454	.2454	.0000
81.	.2858	.2858	.0000	82.	.0996	.0996	.0000
83.	.1365	.1365	.0000	84.	.1368	.1368	.0001
85.	.2095	.2095	.0000	86.	.1741	.1740	.0000
87.	.2888	.2888	.0000	88.	.2685	.2684	.0001
89.	.3565	.3565	.0000	90.	.4457	.4457	.0000
91.	.3742	.3742	.0000	92.	.3632	.3632	.0000
93.	.7894	.7894	.0000	94.	.4970	.4970	.0000
95.	.7856	.7856	.0000	96.	.7681	.7681	.0000
97.	.8536	.8532	.0004	98.	.7984	.7984	.0000
99.	.8159	.8159	.0000	100.	.9671	.9674	.0003

Averages: .414162 .414137 .000025

An Application of Quasidensities

As an illustrative application[2], we return to Example Three of the Preface. Let X be a statistic such as $\hat{\theta}_{WK}$ for which we desire $E(X|\theta=t)$. Let $M_1, M_2, \ldots M_K$ be binary random variables indicating group membership. For example in Example Three, K numbers x_k in the range of $\hat{\theta}_{AR}$ can be used to define variables of the form

$$M_k = 1 \text{ if } |\hat{\theta}_{AR} - x_k| \le .5 \text{ , else zero}$$

dividing examinees into K not necessarily disjoint groups. Let $q_1, \ldots q_K$ be the quasidensities for the (conditional) distributions

$$F_k(t) = \text{Prob } (\theta \le t \mid M_k = 1) \ .$$

Suppose K is large enough and the F_k different enough so that some subset of the q_k forms a basis for the CS. Let $q(\cdot;s)$ be the quasidensity of the unit step at s in $[c,d]$. Then there must be numbers $c_k=c_k(s)$ such that

$$q(t;s) = \sum_{k\le K} c_k(s)q_k(t) \ , \ c\le t\le d \ .$$

From the definition of $q(\cdot;s)$ we have

$$E(X|\theta=s) = \int_c^d E(X|\theta=t)q(s;s) \ dt \ .$$

Thus

$$\begin{aligned} E(X|\theta=s) &= \int_c^d E(X|\theta=t) \sum_{k\le K} c_k(s)q_k(t) \ dt \\ &= \sum_{k\le K} c_k(s) \int_c^d E(X|\theta=t)q_k(t) \ dt \\ &= \sum_{k\le K} c_k(s) \ E(X|M_k=1) \ . \end{aligned}$$

Thus the regression function on the left - expressing a conditioning on an *unobserved* ability - equals a linear combination expected values of observed scores for the *objectively defined* groups.

To apply this result K is taken to be large, $q(\cdot;s)$ is computed with the identity (derived at the end of this section)

$$q(\cdot;s) = \sum_j h_j(s)h_j(t) \ .$$

The q_k are estimated by maximum likelihood. The $c_k(\cdot)$ are computed for each s by minimizing a quadratic objective function such as

$$Q(c_1, \ldots c_K) = \int_c^d [q_s(t) - \Sigma\, c_k(s)q_k(t)]^2 dt \ .$$

In this way a conditional expected value of a statistic given ability can be computed when simulation is not possible or practical.

In addition to the three examples in Example Three, there is the interesting special case of $X=u_{n+1}$, the item score for a new item, and

$$E(X|\theta=t) = P_{n+1}(t) \ ,$$

its item response function. Thus the formula at the bottom of page 24 expresses an unknown item response function as a linear combination of the expected values of statistics.

Summary of Properties of Quasidensities

Throughout this summary, let $\{h_j\}_{j=0}^{J}$ be an orthonormal basis for the CS and $\{X_j\}_{j=0}^{J}$ be statistics satisfying $E(X_j|\theta=t) = h_j(t)$ for $c \le t \le d$.

Properties One, Two, and Three are useful for guessing the shape of the quasidensity when F has a density in the CS or is closely approximated by a distribution on [c,d] with a density in the CS. Property Four can be used even if no close approximation of F has a density in the CS. Property Five underscores the identifiability of the quasidensity by exhibiting a strongly consistent (albeit, inefficient) estimate for the quasidensity.

Defining Property of Quasidensities: A function g in the CS is the quasidensity for G if for all statistics X

$$\int_c^d E(X|\theta=t)\, dG(t) = \int_c^d E(X|\theta=t) g(t)\, dt$$

Formula for Quasidensities: $g(t) = \Sigma_j E[X_j;G] h_j(t)$

Quasidensity for Step Functions: Let G_s be the unit step at s and $q(\cdot;s)$ its quasidensity. Then

$$q(t;s) = \Sigma_j h_j(s) h_j(t)$$

Proof: $E[X_j;G_s] = \int_c^d h_j(t)\, dG_s(t) = h_j(s)$

Property One: If G has a continuous density G′ and G′ is in the canonical space then G′ is the quasidensity of G .

Proof: $\langle R_X, G' \rangle = E(X;G)$ for all statistics X .

Property Two: If G has a (not necessarily continuous) density G′ then the quasidensity of G is the projection of G′ into the canonical space in the sense that the quasidensity g is the unique minimizer in

<u>the CS of</u>

$$\int_c^d [G'(t) - g(t)]^2 dt .$$

<u>Proof</u>: The general function in the CS can be written $h(t,d) = \Sigma_j [E(X_j;G) - d_j] h_j(t)$ for some vector of constants d. Since $E(X_j;G) = \int_c^d h_j(t) G'(t) dt$ and since the h_j are linearly independent it suffices to show that $h(t,0)$ is a minimizer. This follows from the identity

$$\int_c^d [G'(t) - h(t,d)]^2 dt = \int_c^d G'^2 - \Sigma E(X_j;G)^2 + \Sigma d_j^2 .$$

<u>Property Three</u>: <u>If distributions are close, then their quasidensities are close in the following sense</u>:

<u>If</u> F_1 <u>and</u> F_2 <u>be distributions on</u> $[c,d]$ <u>with quasidensities</u> q_1 <u>and</u> q_2 <u>and</u> $\int_c^d [F_1(t) - F_2(t)]^2 dt \le \epsilon$, <u>then</u> $\int_c^d [q_1(t) - q_2(t)]^2 dt \le \epsilon$

<u>Proof</u>: For $i=1,2$ F_i can be written $F_i = q_i + (F_i - q_i) = q_i + r_i$. For any orthonormal basis $\{h_j\}$, $\langle r_i, h_j \rangle = 0$ for each j . Thus for any h in the CS, $\langle r_i, h \rangle = 0$. Consequently

$$\begin{aligned} \int_c^d [F_1(t) - F_2(t)]^2 dt &= \int_c^d [q_1(t) - q_2(t)]^2 dt \\ &\quad + 0 \\ &\quad + \int_c^d [r_1(t) - r_2(t)]^2 \, dt \\ &\ge \int_c^d [q_1(t) - q_2(t)]^2 dt . \end{aligned}$$

<u>Property Four</u>: <u>The quasidensity of the limit of a convergent sequence of distributions on</u> $[c,d]$ <u>is the limit of the corresponding sequence of quasidensities</u>. <u>More precisely</u>,

<u>If</u> $\{G_n\}$ <u>is a sequence of distribution functions on</u> $[c,d]$ <u>weakly convergent to a distribution</u> G <u>on</u> $[c,d]$, <u>then the sequence of</u>

<u>quasidensities of the</u> G_n <u>converges uniformly to the quasidensity of</u> G .

<u>Proof</u>: Let X be any statistic. Since the regression function for X is continuous, by Helly's second theorem $\lim E(X,G_n) = \lim \int_a^b E(X|\theta=t)\, dG_n(t) = E(X;G)$. Uniformity follows from the continuity of quasidensities.

The ability distribution clearly isn't determined by item response data. This is obvious from Theorem One. When J is small, markedly different distributions can be equivalent. The quasidensity, on the other hand, can be recovered from item response data. The formula for the quasidensity shows that all one needs to estimate the quasidensity from data is the expected values of finitely many statistics.

<u>Property Five</u>: <u>The quasidensity is determined by item response data in the sense that there is a strongly consistent quasidensity estimation procedure</u>.

<u>Proof</u>: The variance of each X_j must be finite because there are only finitely many possible values for X_j , one for each of 2^n possible response patterns. Consequently $X_{j,N}$, the sample average for N randomly sampled examinees, tends to $E(X_j)$ with probability one as sample size is increased. In fact, the multivariate strong law of large numbers implies that the vector of sample means $<X_{0,N}, \ldots X_{J,N}>$ almost surely converges to the vector of expected values $<E(X_0), \ldots, E(X_J)>$. Since the quasidensity g for the ability distribution F satisfies

$$g(t) = \sum_{j=0}^{J} E(X_j)h_j(t)$$

the random function defined by

$$g_N(t) = \sum_{j=0}^{J} X_{j,N}\, h_j(t) \ , \quad c \le t \le d$$

almost surely converges to the quasidensity. Furthermore, the convergence must be uniform in t because the h_j are continuous on $[c,d]$.

NOTES

1. The term seems apt because the prefix "quasi" means "to some degree, in some manner." Although $g(t)$ may be negative,
$$\int_c^d h(t)\, dG(t) = \int_c^d h(t)g(t)\, dt$$
at least for every function h in the CS.

2. There is a technical problem beyond the scope of this paper that arises in applications of this type. When the CS has been computed from only a subset of the test items then $R_X(t) = E[X|\theta=t]$ may not be in the CS. In this case the analysis yields an estimate of the projection of R_X into a subspace of the CS computed from all the test items. We have observed that when only a small number of items have not been included the projection and $R_X(t)$ agree to several decimals, provided the not included items are not extremely easy, extremely hard or otherwise atypical.

Section Three

The Canonical Space

Logistic Models and the Examples

This section contains proofs and additional details for assertions made earlier about the examples. We begin the study of computing the dimensionality of the CS and selecting basis functions h_j and statistics X_j for some simple models.

The Rasch Model and its Generalization

In Examples One and Two it was asserted that the generalization of the Rasch Model has J less than or equal to the number of items and that the item response functions or some subset of them form a basis.

If $P_i(t) = c_i + (1-c_i)[\ 1 + e^{-a(t-b_i)}]^{-1}$ then we can solve for e^{at} and obtain

$$e^{at} = e^{ab_i} \frac{P_i(t)-c_i}{1-P_i(t)}$$

Thus for $i \neq j$

$$e^{ab_i} [\ 1-P_j(t)\ ][\ P_i(t)-c_i\ ] = e^{ab_j} [\ 1-P_i(t)\ ][\ P_j(t)-c_j\ ]\ .$$

If $b_i \neq b_j$, then this equation can be simplified to obtain an expression of form

$$P_i(t)P_j(t) = a + bP_i(t) + cP_j(t)$$

where a, b, and c are independent of t . Thus any product of two item response functions can be rewritten as a linear combination of the item response functions plus a constant. Using this fact it's easy to prove the assertions concerning these models in Example One.

<u>If item response functions satisfy the formula for the Rasch model or its generalization with</u> $b_i \neq b_j$ <u>for</u> $i \neq j$, <u>then</u>

1. <u>The dimensionality of the canonical space is less than or equal to one plus the number of items</u>
2. <u>The constant function and the item response functions or some subset of these functions form a basis for the CS</u>
3. <u>The item scores satisfy the condition on the</u> X_i <u>in Theorem One and Example One</u>.

<u>Proof</u>: Since the square-free monomials span the canonical space, it is sufficient to show that every square-free monomial can be expressed as a linear combination of the P_i plus the constant function $h_0(t)=1$. Any square-free monomial containing two or more of the item response functions can be written in form RP_iP_j for $i \neq j$ for R equal to a square-free monomial not divisible by P_i or P_j . Thus $RP_iP_j = aR + bRP_i + cRP_j$ can be rewritten as the linear combination of three square-free monomials, each of which has fewer factors than the original monomial. By iterating this process one eventually obtains a linear combination of square-free monomials depending on one of the P_i or none of the P_i (i.e. h_0). Thus h_0 and the P_i span the CS, which proves 1. and 2. The remaining assertion follows from $E(u_i|\theta=t) = P_i(t)$.

Selecting Points for Example Two

In Example Two we considered changing integrals to sums. It was asserted that there were numbers $t_0, t_1, \ldots t_J$ and $p(t_0), p(t_1), \ldots p(t_J)$ such that for any vector of zeros and ones u^*, the manifest or pattern probability

$$\int \text{lik}(u^*|t)\ dF(t)$$

could be written

$$\Sigma_k\ \text{lik}(u^*|t_k)p(t_k)\ .$$

This is an example of a more general result, proven in this subsection: For any statistic X (including the statistic that is one if the observed item response pattern equals u^* and zero otherwise)

$$\int E[X|\theta=t]\ dF(t) = \Sigma_k\ E[X|\theta=t_k]p(t_k)\ .$$

The choice of the t_k and computation of the $p(t_k)$ is also discussed. We use the notation $q(\cdot;t_k)$ for the quasidensity of the unit step function at t_k and the fact that $q(\cdot;t_k) = \Sigma_j\ h_j(t_k)h_j(\cdot)$ for any orthonormal basis for the CS.

The result need only be proven for bounded ability continua since any item response model with continuous $P_i \neq 0,1$ can be transformed by an invertible transformation to a bounded model. The proof is split into two parts: The existence of a basis consisting of quasidensities and interpretation of the $p(t_k)$.

The results indicate the following procedure for selecting points and computing p's for a model with CS having basis $\{h_j\}_{j=0}^{J}$:

1. Choose $t_0, t_1, \ldots t_J$ such that the matrix

$$\begin{bmatrix} h_0(t_0) & h_1(t_0) & \cdots & h_J(t_0) \\ h_0(t_1) & h_1(t_1) & \cdots & h_J(t_1) \\ \vdots & \vdots & & \vdots \\ h_0(t_J) & h_1(t_J) & \cdots & h_J(t_J) \end{bmatrix}$$

is nonsingular

2. Compute $p(t_0), p(t_1), \ldots p(t_J)$ by solving the linear equations $g(t_j) = \Sigma_k\, p(t_k) q(t_j;t_k)$ for $j=0,1, \ldots J$ where g is the quasidensity of F.

For the generalization of the Rasch model, the procedure can be simplified; the $p(t_k)$ can be found by solving the system of linear equations

$$E(u_i) = \Sigma_k\, p(t_k) P_i(t_k) \qquad i=1, \ldots n$$
$$1 = \Sigma_k\, p(t_k)$$

Generalizations and proofs follow.

<u>If a test has continuous item response functions</u> $\neq 0,1$ <u>defined on an interval</u> $[c,d]$ <u>then the CS has a basis consisting of quasidensities of unit step distributions</u>.

<u>Proof</u>: Let $\{h_j\}_{j=0}^{J}$ be an orthonormal basis for the CS and let $h(t)$ denote the column vector

$$h(t) = \langle h_0(t), h_1(t), \ldots h_J(t)\rangle^T .$$

Since the h_j are linearly independent there must be $J+1$ values of t such that the vectors $h(t_0), h(t_1), \ldots h(t_J)$ are linearly independent. It follows that the partitioned matrix $[h(t_0), h(t_1), \ldots h(t_J)]$ has an inverse, say $A=(a_{ij})$. Consequently, using Kronecker's delta notation each h_j can be written as a linear combination of the quasidensities $q(\cdot;t_i)$

$$\begin{aligned} h_j(t) &= \Sigma_k\, h_k(t)\, \delta_{kj} \\ &= \Sigma_k\, h_k(t) \left(\Sigma_m h_k(t_m) a_{mj}\right) \\ &= \Sigma_m\, a_{mj}\, \Sigma_k h_k(t) h_k(t_m) \end{aligned}$$

$$= \Sigma_m a_{mj} q(t;t_m) .$$

Thus the quasidensities form a basis for the CS.

As a corollary, we have

<u>The quasidensities of unit step distributions at</u> $t_0, t_1, \ldots t_J$ <u>span the CS if and only if</u> $[h(t_0), h(t_1), \ldots h(t_J)]$ <u>is non-singular</u>.

In practice on this type of problem we compute the t_k recursively. After having chosen $t_0, t_1, \ldots t_k$ we choose t_{k+1} such that $h(t_{k+1})$ makes a relatively large angle with its projection into the linear space spanned by $h(t_0), h(t_1), \ldots h(t_k)$.

After the t_k are selected the calculation of the $p(t_k)$ is straight forward. Since the quasidensities for the t_k form a basis for the CS, the ability distribution's quasidensity is a linear combination of the $q(\cdot;t_k)$ and the coefficients of the combination are unique. The $p(t_k)$ are simply the coefficients of the linear combination.

<u>Let</u> $\{q(\cdot;t_k)\}_{k=0}^{J}$ <u>be a basis for the CS and the quasidensity for the ability distribution be</u> $\Sigma_k p(t_k)q(\cdot;t_k)$. <u>Then for any statistic</u> X , $E(X) = \Sigma_k E(X|\theta=t_k)p(t_k)$. <u>In particular for any vector of zeros and ones</u> u^* , $\mathrm{Prob}(u=u^*) = \Sigma_k \mathrm{Prob}(u=u^*|\theta=t_k)p(t_k)$.

<u>Proof</u>: Let X be any statistic. Then from the defining property of quasidensities

$$E(X) = \int_c^d E(X|\theta=t)\, \Sigma_k p(t_k)q(t;t_k)\, dt$$

$$= \Sigma_k p(t_k) \int_c^d E(X|\theta=t)q(t;t_k)\, dt$$

$$= \Sigma_k p(t_k)\, E(X|\theta=t_k)$$

In particular for any vector of zeros and ones u^* if X is the random variable that is one if $u=u^*$ and zero otherwise, $\mathrm{Prob}(u=u^*) = E(X)$

$= \Sigma_k \text{ Prob}(u=u^* | \theta=t_k)p(t_k)$.

Models with Very Large J

If J is small, as is the case with the Rasch model and its generalization, then standard techniques can be used for computing an orthonormal basis. ,wever, if the dimensionality of the CS is as large as the number of square-free monomials (2^n) then computing an orthonormal basis is problematical. To conclude this section it is shown that for the most commonly used item response models, the three parameter logistic models, J+1 typically is equal to its upper bound.

Item response functions are three parameter logistic (3PL) if

$$P_i(t) = c_i + (1-c_i)[1 + e^{-a_i(t-b_i)}]^{-1}$$

for some item parameters $a_i>0$, b_i, and c_i in $[0,1)$. It is natural to consider the item parameters random variables because in most applications they are estimated from data. Suppose the sampling distribution of the estimated parameters has a continuous density. Then the following result is of interest.

<u>If the joint distribution of the</u> n <u>item parameter vectors</u> $\langle a_i,b_i,c_i\rangle$ <u>has a continuous density</u>, <u>then with probability one the CS of the 3PL item response model defined with sampled item parameters will have dimension</u> 2^n

Thus, for example, if one begins with the any published set of estimated item parameters for an application of the 3PL model and adds an independent normally distributed "error" with zero mean and very small variance, say 10^{-10}, to each of the 3n parameters, then with probability one either one of the a's or c's will be moved outside its allowed range or a 3PL model with J as large as it possible can be will be obtained.

<u>Proof</u>: With probability one, the functions

$$e^{a_1 t}, e^{a_2 t}, \ldots e^{a_n t}$$

will be algebraicly independent over the reals, i.e. will not satisfy any nontrivial polynomial with real coefficients. But if $J+1<2^n$ then one of the square-free monomials can be expressed as a linear combination of the remaining monomials. On multiplying both sides of the equation giving one monomial as a linear combination of the others by positive $\Pi_i[e^{a_i t} + e^{a_i b_i}]$ one obtains a polynomial in the $e^{a_i t}$ and a contradiction to the hypothesis $J+1<2^n$.

Section Four

Large Canonical Spaces

Consider Example Three for a test with large CS for an application currently in progress. In a large scale simulation we are attempting to monitor and control the changes in a Bayes modal ability estimate as new items are introduced into a 100 item adaptive test item pool. The item response function estimates for the new items are not expected to be very accurate because of motivation, test format, and ability distribution differences between the item response function estimation sample and the examinees in the application. The methods to be reviewed in this section permit us to compute as many as we need of the roughly 2^{100} orthornormal h_j for the test consisting of old items.

The trick is to compute the h_j one-at-a-time in such a way that the h_j needed to complete the application are computed first. Thus the CS is treated as the union of nested vector spaces CS_K

$$CS_K = \mathrm{Span}(h_0, h_1, \ldots h_K)$$

where functions in only a dozen or so spaces can be and need be accurately computed. Some details follow.

We wish to approximate $E(\hat{\theta}_{WK}|\theta=t) = R(t)$, where $\hat{\theta}_{WK}$ is the Bayes mode adaptive test score. It turns out that although J is very large, the projection $\hat{R}$ of R into the twelfth space

$$\hat{R}(t) = \Sigma_{j\leq 12} <R,h_j>h_j(t)$$

is very close to $R(t)$. Now if $\hat{q}(\cdot;s) = \Sigma_{j\leq 12} h_j(s)h_j(t)$ is the projection $q(\cdot;s)$ into the twelfth space then $\int_c^d E(\hat{\theta}_{WK}|\theta=t)\hat{q}(t;s)\,dt = \hat{R}(t)$. Thus if we can write $\hat{q}(\cdot;s)$ as

$$\hat{q}(\cdot;s) = \Sigma_{k\leq K} c_k(s)q_k(\cdot)$$

a linear combination of quasidensities for the K AR score groups $q_k(\cdot)$, then

$$\hat{R}(t) = \Sigma_{k\leq K} c_k(s) E[\hat{\theta}_{WK}|\hat{\theta}_{AR} \text{ is in } M_k] .$$

The point is that if an application can be completed using $h_0, h_1, \ldots h_K$ only then it may be possible to proceed *as if* $J=12$.

This section describes a general technique used by our laboratory for calculating the h_j one-at-a-time in such a way that functions that are likely to be needed for an application are well approximated by a function in CS_K for small K .

The General Method

The first step of our approach to large spaces is to select a set of functions $\{f_\nu\}_{\nu=1}^{N}$ that span the CS *and* are such that the function of two variables $\Sigma_\nu f_\nu(s)f_\nu(t)$ can be easily evaluated. For example if $f_1, f_2, \ldots f_\nu, \ldots f_{2^n}$ is any enumeration of the square-free monomials then the f_ν span the CS. Furthermore for any s and t

$$\sum_{\nu=1}^{2^n} f_\nu(s)f_\nu(t) = \prod_{1}^{n} [1 + P_i(s)P_i(t)]$$

can be evaluated with 2n-1 multiplications and n additions. (This identity can be verified by induction on test length n .) Other examples of tractible spanning sets and additional criteria for spanning sets are discussed below.

There are two important points to be emphasized here. Although there are generally billions of f_ν to enter into the sum $H(s,t) = \Sigma_\nu f_\nu(s)f_\nu(t)$, the multiplicative formula for $H(s,t)$ requires only n additions and 2n-1 multiplications. Second, the ordering of the f_ν is inconsequential. Whereas the outcome of a Gram-Schmidt orthogonalization applied to the square-free monomials or any other large set of functions f_ν would be very order dependent, the calculation of H is not.

The next step in computing the h_j can be carried out with commercial software or can be converted to a eigenvalue/eigenvector problem: Compute positive numbers λ and functions h not identically equal to zero such that each h is in the CS and satisfies

$$\lambda h(\cdot) = \int_c^d H(\cdot,t)h(t)\ dt$$

where $H(s,t) = \Sigma_\nu\ f_\nu(s)f_\nu(t)$. There will be only finitely many different values of λ such that there is some $h\neq 0$ in the CS satisfying the equation. Since the h's are in the CS there can be only finitely many linearly independent solutions h for any λ . Thus any maximal set of linearly independent solutions can be subscripted and arranged in order of their subscript so that $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_K > 0$ for some $K \leq J$ and $\lambda_j h_j(\cdot) = \int_c^d H(\cdot,t)h_j(t)\ dt$.

Without loss of generality we can set $\langle h_j,h_j\rangle = 1$ since $h_j(t)$ is a solution for λ_j if and only if $h_j(\cdot)/\langle h_j,h_j\rangle$ is. Since the set of all h's corresponding to any λ form a vector space, they can be selected to be orthonormal. Since it can be shown h's with different λ's are orthogonal, the h_j will form an orthonormal set of vectors In fact it is easy to show that when the f_ν span the CS, K=J and the h_j computed in this way form an orthonormal basis for the CS. If an application suggests a set of f_ν that don't span the CS, then K<J and the h_j will be a basis for whatever subspace the f_ν span.

Note that except in the unusual case that more than one h correpsonds to one λ , the h's are fully ordered by their λ's . Even if for some j , $\lambda_j = \lambda_{j+1}$, the h's corresponding to different λ's will be ordered and we can still speak of h_j occuring early or late in the sequence of h's .

The ordering is important because for various reasons (cumulative numerical errors and the fact that λ_j is very close to zero for large j) the h_j that occur early in the sequence are relatively easy to compute (although the remaining h_j can be very hard to compute).

There are two related advantages in arranging the computation of basis functions as described above. The h_j with large λ_j , which are easy to compute, can be computed without computing the h_j with small λ , which can be very hard to compute. This is important because λ_j generally measures the relative importance h_j in representing functions in several senses. For example, if f_ν is approximated by its projection into $CS_K = \text{span}(h_0, \ldots h_K)$, which turns out to be $\hat{f}_\nu(\cdot) = \sum_{j\le K} \langle f_\nu, h_j \rangle h_j$ for $K<J$, then the total error

$$\sum_\nu \int [f_\nu(t) - \hat{f}_\nu(t)]^2 \, dt$$

is simply $\sum_{j>K} \lambda_j$. (This sum can be evaluated even if J is very large because

$$\int_c^d H(t,t)\, dt - \sum_{j\le K} \lambda_j = \sum_{j>K} \lambda_j \text{ .)}$$

As a bonus, the method also delivers a set of statistics X_j needed for Example One and Theorem One (i.e., statistics such that $h_j(t) = E[X_j | \theta = t]$ for all t in $[c,d]$). Details are given in the final subsection.

Some Examples of Spanning Sets

In addition to the square-free monomials we use the 2^n likelihood functions for short tests. Here

$$f_\nu(t) = \prod_{i=1}^{n} P_i(t)^{u^*_{i,\nu}} [1-P_i(t)]^{(1-u^*_{i,\nu})}$$

where $u^*_1, \ldots u^*_\nu, \ldots u^*_{2^n}$ is any enumeration of the 2^n item response patterns. For these functions

$$H(s,t) = \Sigma_\nu \; f_\nu(s)f_\nu(t)$$
$$= \prod_1^n \{P_i(s)P_i(t)+[1-P_i(s)][1-P_i(t)]\} \; ,$$

which can be easily evaluated. (This also can be proven by induction on test length n after noting that each likelihood function can be written as $f_\nu(t) = \prod_{i=1}^{n} \{u^*_{i,\nu}P_i(t) + (1-u^*_{i,\nu})[1-P_i(t)]\}$.) These functions certainly span the CS because any square-free monomial can be written as a linear combination of likelihood functions. (To prove this, simply write the general monomial $\prod_{j\le r} P_{i_j}$ as the sum of the likelihoods for patterns u^* with $u^*_{i_1}=u^*_{i_2}= \ldots u^*_{i_r}=1$.)

For adaptive tests and long tests satisfying (exactly or approximately) an algebraic property described below, we use likelihood functions for selected subtests. For example to study a fixed length adaptive test of 15 items with a 100 item pool it is natural to consider the $\binom{100}{15} \ll 2^{100}$ likelihood functions with fifteen factors since every statistic computed from an examinee's score depends on only 15 item scores.

The discussion of the Rasch model introduces a second rationale for forming the f_ν from the likelihood functions for short subtests. Recall that for the Rasch model every polynomial in the CS could be rewritten as a "polynomial" in the CS, no monomial of which contained 2 or more factors. This property is remarkably general. For the 3PL model (and most of its generalizations) every polynomial in the CS can be rewritten as a linear combination of monomials with five or fewer factors, at least to a surprisingly high degree of approximation[1].

When every function in the CS can be expressed as a linear combination of square-free monomials with five or fewer factors, then the CS is spanned by the likelihood functions from subtests with five factors. There are still an enormous number of likelihood functions f_ν that can be formed from from all five item subtests. Nonetheless $H(s,t) = \Sigma\ f_\nu(s)f_\nu(t)$ can be computed efficiently for these functions as follows:

Let $F_i(s,t)$ abbreviate $P_i(s)P_i(t) + [1-P_i(s)][1-P_i(t)]$.
Let $H_i^m(s,t)$ denote the sum of the likelihood functions for all i item subtests formed from the first m items.

To initialize set

$$H_1^1(s,t) = F_1(s,t)$$

$$H_i^1(s,t) = 0 \quad \text{for} \quad i=2,3,\ldots 5 \ .$$

To update, compute

$$H_i^{m+1} = F_{m+1}H_{i-1}^m \quad \text{for} \quad i=2,\ldots 5$$

$$H_1^{m+1} = F_{m+1} + H_1^m$$

If in the update step H_5^{m+1} is computed first, followed by H_4^{m+1} , etc., then H_j^{m+1} can be written over H_j^m and the amount of storage required by the algorithm can be kept small.

Most of our current applications to one dimensional ability tests use this algorithm. Although some of the CS may be left out, the algorithm in practice works very well. It is the only algorithm that has consistently produced useful results with long tests.

Reduction of Proofs to Matrix Algebra

A number of assertions were made without proof concerning the solutions for the functional equation

$$\psi(h) = \lambda h$$

where $\psi(h)(\cdot) = \int_c^d H(\cdot,t)h(t)\,dt$ for $H(s,t) = \Sigma_\nu f_\nu(s)f_\nu(t)$. By taking advantage of the finite dimensionality of the CS these proofs can be obtained with matrix algebra. In this section the reduction to matrix algebra is indicated after a few of the assertions are proven directly.

First ψ is a transformation of the CS to itself because the f_ν are in the CS and $\psi(h) = \Sigma_\nu \langle f_\nu,h\rangle f_\nu$ is a linear combination of the f_ν . ψ is thus a linear mapping of a finite dimensional vector space into itself.

To show that the eigenfunctions of ψ span the CS it is neccesary to show that ψ maps the CS onto the CS. Equivalently, since the CS is finite dimensional, one may show $\psi(h)=0$ implies $h=0$. To show this one can write $f_\nu(\cdot) = \sum_{j=0}^{J} a_{\nu j}g_j(\cdot)$ for some orthonormal basis $\{g_j\}_{j=0}^{J}$. The matrix $A=(a_{\nu j})$ must have rank $J+1$ since the f_ν span the CS. If $\psi(h)=0$, then $0 = \langle g_j,\psi(h)\rangle = e_j^T A^T A\langle g,h\rangle$, $j=0,\dots J$ where e_j is the jth unit vector and $\langle g,h\rangle$ is the column vector of $\langle g_j,h\rangle$'s . Thus $A^T A\langle g,h\rangle=0$. Since $A^T A$ has rank $J+1$, $\langle g,h\rangle=0$, i.e., h is orthogonal to each g_j . Thus $h=0$.

The existence of eigenfunctions in the CS and the fact that the eigenfunctions span the CS can be shown with matrix algebra. To introduce matrix notation, for each t in $[c,d]$ let $f(t)$ be the column vector with νth coordinate $f_\nu(t)$. Then $H(s,t)$ is the scalar product of $f(s)$ and $f(t)$. Let Q denote the matrix of definite integrals

$Q = \int_c^d f(t)f^T(t)\,dt$, i.e. Q is the matrix with typical entry $q_{\nu\nu'} = \langle f_\nu, f_{\nu'}\rangle$.

Q must be positive definite or positive semidefinite since for any vector a , $a^TQa = \int_c^d [a\cdot f(t)]^2\,dt \geq 0$. Therefore for some K , Q can be written $Q = [a^0,a^1, \ldots a^K]^T D[a^0,a^1, \ldots a^K]$ for $K+1$ orthonormal vectors a^j and a diagonal matrix D having positive diagonal entries $d_j>0$. For $0\leq j\leq K$ let h_j be defined by

$$h_j(t) = d_j^{-1/2}\, a^j\cdot f(t) .$$

Since each h_j is a linear combination of functions in the CS, each must be in the CS. The h_j are orthonormal since

$$\begin{aligned}\langle h_j,h_k\rangle &= d_j^{-1/2} d_k^{-1/2}\, a^{jT} \int_c^d f(t)f^T(t)\,dt\, a^k \\ &= d_j^{-1/2} d_k^{-1/2}\, a^{jT}Qa^k . \\ &= \begin{cases} 0 , & \text{if } j\neq k \\ 1 , & \text{if } j=k \end{cases}\end{aligned}$$

In fact the h_j must be eigenfunctions of ψ because

$$\begin{aligned}\psi(h_j) &= \int_c^d f^T(t)f(\cdot)d_j^{-1/2}\, a^j\cdot f(t)\,dt \\ &= d_j^{-1/2}\, a^{jT} \int_c^d f(t)f^T(t)\,dt\, f(\cdot) \\ &= d_j^{-1/2}\, a^{jT}Qf(\cdot) \\ &= d_j^{1/2}\, a^{jT}f(\cdot) \\ &= d_j h_j .\end{aligned}$$

K must equal J because otherwise ψ would not map the CS onto the CS. Thus the eigenfunctions form an orthonormal basis for the CS.

The Statistics X_j

In Example One and Theorem One statistics with regression functions equal to h_j were needed. Of course such statistics exist because every function in the CS, by definition, is the regression function of at least one statistic. Finding a statistic matching a function fortunately turns out to be easy for bases formed from eigenfunctions.

When the h_j are obtained as eigenfunctions, these statistics are calculated in two steps. First, the examinee's data is transformed into a continuous function $X(t)$. Then a statistic is obtained by computing $\langle X,h_j\rangle/\lambda_j$.

For concreteness consider the second example of the general method in which each f_ν is a likelihood function. The general technique applied to this example gives $X(t)$ equal to the familiar likelihood function as the random function

$$X(t) = \prod_{i=1}^{n} [u_i P_i(t) + (1-u_i)Q_i(t)]$$

and $X_j = \int_c^d X(t)h_j(t)\, dt/\lambda_j$.

To verify that the regression function for this statistic is h_j , we compute as follows. The regression function for X_j evaluated at $\theta = s$ is

$$\begin{aligned} E[X_j|\theta=s] &= \lambda_j^{-1} E[\int X(t)h_j(t)\, dt|\theta=s] \\ &= \lambda_j^{-1} \int \prod_{i=1}^{n}{}' [P_i(s)P_i(t) + Q_i(s)Q_i(t)]h_j(t)\, dt \\ &= \lambda_j^{-1} \int H(s,t)h_j(t)\, dt \\ &= h_j(s) . \end{aligned}$$

The general rule for obtaining a random function $X(t)$ for arbitrary f_ν is to make the replacements

$$P_1(s) \rightarrow u_1$$

$$P_2(s) \to u_2$$

$$\vdots$$

$$P_n(s) \to u_n$$

in $f_\nu(s)$ to obtain a random variable $Y_\nu(u)$ from $f_\nu(s)$. A random function X is defined by

$$X(t) = \Sigma_\nu \, Y_\nu(u) f_\nu(t) \; .$$

Finally, a random variable having regression function equal to the jth basis function is obtained as $\int_c^d X(t)h_j(t)\,dt/\lambda_j$. To summarize

<u>Let</u> $H(s,t) = \Sigma\, f_\nu(s)f_\nu(t)$ <u>for functions in the CS</u> f_ν <u>not necessarily spanning the CS</u>. <u>Let</u> h <u>satisfy</u> $\int_c^d H(\cdot,t)h(t)\,dt = \lambda h(\cdot)$ <u>for positive</u> λ. <u>For each</u> t <u>in</u> $[c,d]$ <u>let</u> $X(t)$ <u>be the random variable obtained by replacing each</u> $P_i(s)$ <u>by</u> u_i <u>in the formula defining</u> $H(s,t)$. <u>If</u> $X_j = \langle X,h_j\rangle/\lambda_j$, <u>then</u> $E[X_j|\theta=t] = h_j(t)$ <u>for</u> $c \le t \le d$.

Note, the transformation $f_\nu(s) \to Y_\nu$ generally cannot be defined on the CS because if two items have the same item response function, then we can have $f_\nu(\cdot) = f_{\nu'}(\cdot)$ as functions in the CS but $Y_\nu \neq Y_{\nu'}$. The problem can be avoided by regarding $f_\nu(s)$ as a polynomial with real coefficients in algebraicly independent variables $P_1(s), P_2(s), \ldots P_n(s)$.

<u>Proof</u>: $E[X(t)|\theta=s] = H(s,t)$.

NOTES

1. Levine, M. and Williams, B. <u>Latent trait theory as fundamental measurement</u>, Paper presented at Society for Mathematical Psychology Annual Conference, Irvine, California, 1989.

Acknowledgements

This research was supported by the Cognitive Science Program of the Office of Naval Research contract N00014-83K-0397, NR 150-518 and N00014-86K-0482, NR 4421546. I am indebted to Bruce Williams, Tim Davey, Charles Davis, Fritz Drasgow, Brian Junker, and Gary Thomasson for detailed comments on earlier versions of this work. Conversations with J.O. Ramsay were also useful.

1983/08/04
University of Illinois/Levine

Dr. Terry Ackerman
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Robert Ahlers
Code N711
Human Factors Laboratory
Naval Training Systems Center
Orlando, FL 32813

Dr. James Algina
1403 Norman Hall
University of Florida
Gainesville, FL 32605

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Mail Stop: 10-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N712
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Bruce Bloxom
Defense Manpower Data Center
99 Pacific St.
Suite 155A
Monterey, CA 93943-3231

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans F. Crombag
Faculty of Law
University of Limburg
P.O. Box 616
Maastricht
The NETHERLANDS 6200 MD

Dr. Timothy Davey
American College Testing Program
P.O. Box 168
Iowa City, IA 52243

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Bldg., Rm. 4112
University of Maryland
College Park, MD 20742

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1K226
Piscataway, NJ 08854

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. R. Darrell Bock
University of Chicago
NORC
6030 South Ellis
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code 7B
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd., North
Chapel Hill, NC 27514

Dr. Robert M. Carroll
Chief of Naval Operations
OP-01B2
Washington, DC 20350

Dr. Raymond E. Christal
UES LAMP Science Advisor
AFHRL/MOEL
Brooks AFB, TX 78235

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
Los Angeles, CA 90089-1061

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
224B Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. George Englehard, Jr.
Division of Educational Studies
Emory University
210 Fishburne Bldg.
Atlanta, GA 30322

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. P-A. Federico
Code 51
NPRDC
San Diego, CA 92152-6800

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
P.O. Box 168
Iowa City, IA 52243

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
U.S. Army Headquarters
DAPE-MRR
The Pentagon
Washington, DC 20310-0300

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Alfred R. Fregly
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor Street
Chicago, IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Mr. Dick Hoshaw
OP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
3-104 Educ. N.
University of Alberta
Edmonton, Alberta
CANADA T6G 2G5

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Elec. and Computer Eng. Dept.
University of South Carolina
Columbia, SC 29208

Dr. Douglas H. Jones
Thatcher Jones Associates
P.O. Box 6640
10 Trafalgar Court
Lawrenceville, NJ 08648

Dr. Brian Junker
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Milton S. Katz
European Science Coordination
Office
U.S. Army Research Institute
Box 65
FPO New York 09510-1500

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

DORNIER GMBH
P.O. Box 1420
D-7990 Friedrichshafen 1
WEST GERMANY

Prof. Edward Haertel
School of Education
Stanford University
Stanford, CA 94305

Dr. Ronald K. Hambleton
University of Massachusetts
Laboratory of Psychometric
and Evaluative Research
Hills South, Room 152
Amherst, MA 01003

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Grant Henning
Senior Research Scientist
Division of Measurement
Research and Services
Educational Testing Service
Princeton, NJ 08541

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 63
San Diego, CA 92152-6800

Dr. Paul W. Holland
Educational Testing Service, 21-T
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 92010

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
Box 7246, Meas. and Eval. Ctr.
University of Texas-Austin
Austin, TX 78703

Dr. Leonard Kroeker
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Jerry Lehnus
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541-0001

Dr. Robert L. Linn
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. George B. Macready
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. James R. McBride
The Psychological Corporation
1250 Sixth Avenue
San Diego, CA 92101

Dr. Clarence C. McCormick
HQ, USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

Mr. Christopher McCusker
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. Robert McKinley
Law School Admission Services
Box 40
Newtown, PA 18940

Dr. James McMichael
Technical Director
Navy Personnel R&D Center
San Diego, CA 92152-6800

Mr. Alan Mead
c/o Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Headquarters Marine Corps
Code MPI-20
Washington, DC 20380

Dr. Ratna Nandakumar
Dept. of Educational Studies
Willard Hall, Room 213
University of Delaware
Newark, DE 19716

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Norman, OK 73071

Deputy Technical Director
NPRDC Code 01A
San Diego, CA 92152-6800

Director, Training Laboratory,
NPRDC (Code 05)
San Diego, CA 92152-6800

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152-6800

Director, Human Factors
& Organizational Systems Lab,
NPRDC (Code 07)
San Diego, CA 92152-6800

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Dr. James B. Olsen
WICAT Systems
1875 South State Street
Orem, UT 84058

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Office of Naval Research,
Code 125
800 N. Quincy Street
Arlington, VA 22217-5000

Assistant for MPT Research,
Development and Studies
OP 0187
Washington, DC 20370

Dr. Judith Orasanu
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Peter J. Pashley
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Department of Operations Research,
Naval Postgraduate School
Monterey, CA 93940

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MOA
Brooks AFB, TX 78235

Mr. Steve Reiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
905 Orchid Way
Carlsbad, CA 92009

Dr. Dan Segall
Navy Personnel R&D Center
San Diego, CA 92152

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr. William Sims
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street, Suite 120
Alexandria, VA 22314-1713

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305

Dr. Richard C. Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Judy Spray
ACT
P.O. Box 168
Iowa City, IA 52243

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
Code-131
San Diego, CA 92152-6800

Dr. John Tangney
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
103 S. Mathews Avenue
Urbana, IL 61801

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Thomas J. Thomas
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 440
St. Paul, MN 55114

Dr. Frank L. Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Thomas A. Warm
FAA Academy AAC934D
P.O. Box 25082
Oklahoma City, OK 73125

Dr. Brian Waters
HumRRO
12908 Argyle Circle
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Ronald A. Weitzman
Box 146
Carmel, CA 93921

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 51
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90089-1061

German Military Representative
ATTN: Wolfgang Wildgrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
NRC MH-176
2101 Constitution Ave.
Washington, DC 20418

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
Code 51
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
03-T
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550

Mr. Anthony R. Zara
National Council of State
Boards of Nursing, Inc.
625 North Michigan Avenue
Suite 1544
Chicago, IL 60611