A GEOMETRICAL APPROACH TO ITEM RESPONSE THEORY

J.O. Ramsay*

How critical is the concept of the latent trait to modern test theory? The appeal to some unobservable characteristic modulating response probability can lead to some confusion and misunderstanding among users of psychometric technology. This paper looks at a geometric formulation of item response theory that avoids the need to appeal to unobservables. It draws on concepts in differential geometry to represent the trait being measured as a differentiable manifold within the space of possible joint item response probabilities given conditional independence. The result is a manifest and in principle observable representation of the trait that is invariant under one-to-one transformations of trait scores. These concepts are illustrated by analyses of an actual test.

1. Introduction

The top panel in Figure 1 displays two three-parameter logistic (3PL) item response functions,

$$P(\theta)=c+(1-c)/(1+\exp[-1.7a(\theta-b)]), \ a=1, \ b=-1, \ 1, \ c=0.2. \qquad (1)$$

The variable θ determines or modulates the probability P of a correct response on a test item, and is often communicated to users of psychometric technology as a representation of "ability", "proficiency", or whatever one understands as determining test performance.

But values of the argument θ are both unobservable and undefined, whereas values of the function are, at least in principle. For example, we can transform θ to an alternative representation of ability that may seem more natural to a test user. Suppose that one did not find appealing the fact that θ in (1) has no lower or upper limit; the lower limit of someone's understanding of calculus could reasonably be imagined to be zero, for example, and actualized in practical terms by the performance of small children on calculus exams. Such a perspective might suggest the transformation

$$\gamma(\theta)=\exp(\theta/2) \qquad (2)$$

as a more meaningful representation, and, since probability of success has to remain unchanged, one might substitute $\theta=2\ln(\gamma)$ in (1) to obtain

$$P^*(\gamma)=c+(1-c)/(1+\exp[-1.7a(2\ln(\gamma)-b)]). \qquad (3)$$

Theta

exp(Theta/2)
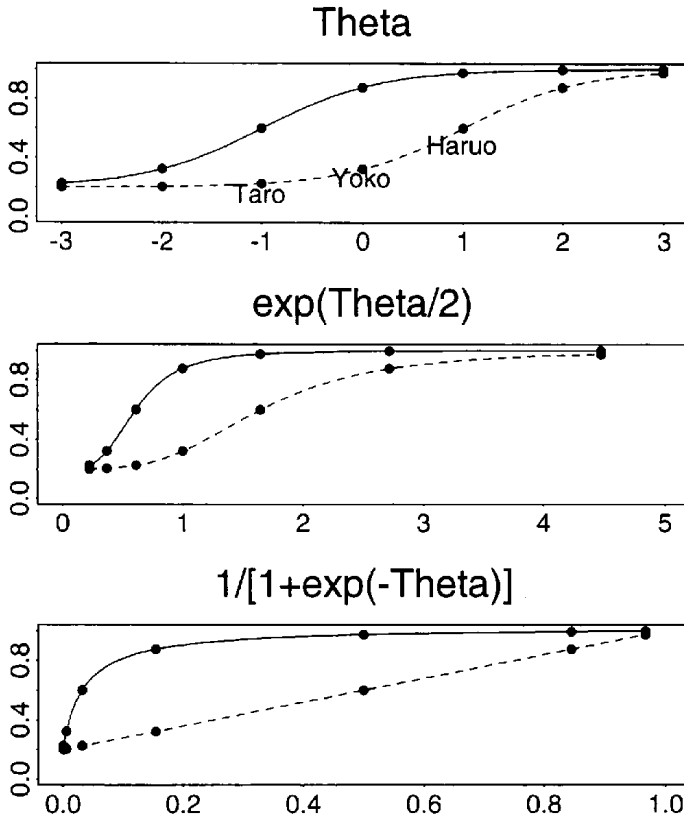
1/[1+exp(-Theta)]

Fig. 1   Two item response functions plotted in terms of latent variable θ and, after the appropriate re-expressions, as functions of two transformations of θ. Seven values of θ and their corresponding transformed values are indicated by the dots.

These functions appear in the second panel of Figure 1, and obviously have somewhat different characteristics.   The dots, for example, which correspond to equally spaced values of θ, are no longer equally spaced with respect to γ, even though their ordinate or probability values remain the same.

On the other hand, traits corresponding to performances which can be perfect, such as calculating the derivatives of polynomials, might be better represented on the unit interval $[0, 1]$, and this can be achieved by the transformation

$$\phi(\theta)=1/[1+\exp{(-\theta)}] \tag{4}$$

and the consequence of substituting $\theta=\ln{[\phi(1-\phi)]}$ into (1) is displayed in the third panel.

There does not seem to be any reason in most applications other than these "natural interpretation" considerations for choosing among these three representations; all make identical statements about probabilities of success for specific individuals on specific items.   Of course, we might superimpose on the problem some properties that would result in some pleasant mathematical properties; users

of the Rasch model sometimes appeal to the principle of specific objectivity, expressible in terms a functional equation relating item and examinee characteristics (Fischer, 1995) as a justification the θ-measure. But, even supposing that the very restrictive Rasch model offered an adequate account of item performance, what aside from mathematical nicety would recommend the constraint of specific objectivity to users? Nature, someone is likely to quickly point out, is not always mathematically convenient.

Consequently, P is a very curious mathematical object; a sort of quasi-function for which the range is clearly defined and anchored in observables, but with a domain that can be monotonically transformed at will. Put algebraically, let g be any strictly increasing transformation of θ. Then its inverse g^{-1} exists, and if $\gamma = g(\theta)$, then the new item response function P^* defined by

$$P^*(\gamma) = P[g^{-1}(\gamma)]$$

and the original function P are, for all practical purposes, identical. We shall use the notation P_j^* to refer to an item response function of an arbitrary argument, and reserve P_j for the situation where a specific argument continuum is intended. In this sense, P_j^* stands for a class of functions rather than a specific example.

Unfortunately, I believe, the latent trait concept is widely misunderstood and abused in the community of statistically less-informed consumers of modern test analysis technology. My experience, and a careful reading of most of texts in the area, has convinced me that there are too many people who imagine that modern test theory permits the estimation of ability or proficiency on a numerical scale where computing differences, sums, and other quantities is meaningful.

Let, for example, examinees Taro, Yoko, and Haruo have values of θ of -1, 0, and 1, respectively. Can we say that Yoko is as more proficient than Taro as Haruo is than Yoko? Certainly not in the scales of γ and ϕ. Such a statement is only supportable if we have some considerations able to withstand legal scrutiny that would determine representation θ, or any linear transformation of it, over all other possible candidates.

The aim of this paper is to replace the inherently ambiguous concept of the latent variable θ by concepts drawn from differential geometry, with a view to expressing the unidimensional modern test theory model entirely ·in terms of observables. A consequence of this will be a measure of proficiency that is completely invariant under admissible transformations of the latent trait, and for which algebraic operations like differencing have a specific meaning in terms of experimentally measurable outcomes.

2. The response probability manifold $\mathcal{M}$

Suppose now that we plot the values of the two item response functions in Figure 1 against each other. We are allowed to do this provided we believe that

the two probabilities of success behave independently of each other given a value of θ, or that the joint probability of success conditional on θ is $P_1(\theta)P_2(\theta)$. This is the principle of conditional independence that underlies most of modern test theory.

Figure 2 is the result, and this plot will be unchanged no matter which representation θ, γ, or ψ is used since, after all, these argument values are simply not used in the plotting process. We now see that performance on this two-item test is represented by a curve within the square of possible pairs of success probabilities. Individual performances, such as those of Taro, Yoko, and Haruo, now become points along this curve, and we can easily read off their two probabilites of success by noting their two coordinate values.

This graphical device can be extended to any number of items : A three-item test is depicted as a curve within a cube, and an n-item test as a curve within an n-dimensional hypercube. The key point is that any position along this curve corresponds to an n-tuple of probabilities of success, and thus is estimable from data. If Yoko, for example, could be induced under conditions of no memory to take the test a few hundred times, her position would be defined to within a quite reasonable level of accuracy using elementary statistical techniques.

The technical definition of a structure like the curve in Figure 2 is a *manifold*,
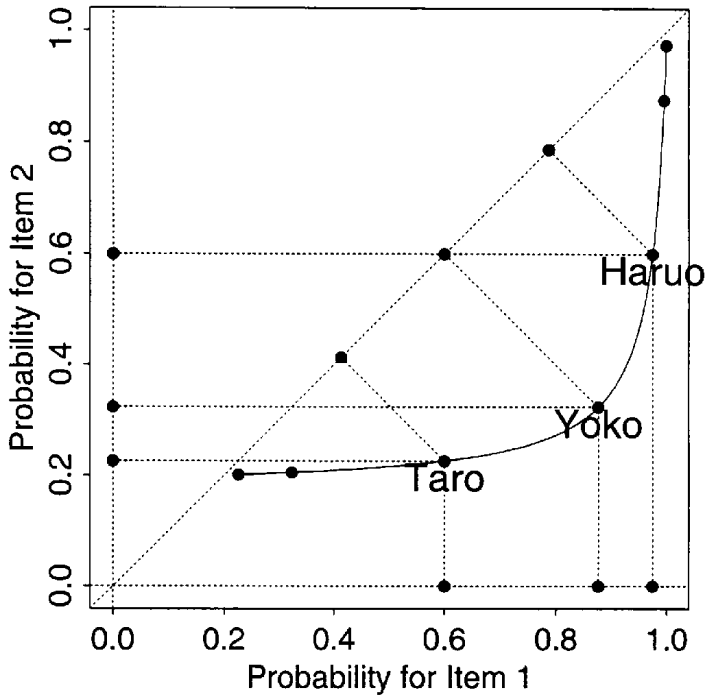

Fig. 2   The probability values for the two item response functions in Figure 1 are plotted against each other, with the dots corresponding to those in that Figure. Three possible coordinate axes for defining positions along the response probability manifold, assuming that item response functions are monotone, are suggested by the horizontal, vertical and diagonal dashed lines.

and if we, perhaps not unreasonably, assume that each P_j is smooth in the sense of being an infinitely differentiable function of θ, and consequently that only strictly monotonic and smooth transformations g of θ are possible, the space curve is called a *differentiable manifold*. The Appendix can be consulted for a more careful definition.

One might use the term *response probability space* to refer to the n-cube, and *response probability manifold* to refer to the space curve itself, which we shall designate as $\mathscr{M}$. Curve $\mathscr{M}$ is essentially the trajectory within joint probability space that individuals must follow to move from extreme ignorance to extreme knowledge.

2.1 Extrinsic coordinate maps

How are we to measure positions along this space curve $\mathscr{M}$? There are many possibilities, among which are, to be sure, our latent trait continua θ, γ and ψ. A *coordinate map* is a differentiable one-to-one mapping $\phi : \mathscr{M} \to X$ from the response manifold to an interval on the real line, and its inverse ϕ^{-1} determines positions on $\mathscr{M}$ in terms of values within this interval X. We turn now to the task of defining a coordinate map.

2.1.1 Success probability for a fixed item

Let it be supposed at this point that each item response function is strictly monotone, something that would be true if one restricted oneself to the three-parameter logistic family (1) with positive coefficients a. Monotonicity is often assumed in discussions of probabilistic questions connected with item response theory (Ellis and Junker, 1995; Junker, 1993; Holland and Rosenbaum, 1986). This assumption permits the following two possibilites for coordinate maps, displayed in Figure 2.

We can begin with a simple illustration of a coordinate map by projecting points on the curve $\mathscr{M}$ on to the first axis (P_1 values), or indeed any other axis. Suppose, for illustration, we use the probability of success on the first item to define this mapping. Then $X=[0, 1]$, and coordinate function ϕ is the inverse of this projection; that is, the process of locating the point on $\mathscr{M}$ that corresponds to a given probability value. For example, for Yoko, whose θ value is 0, the success probabilities for all items can be computed as $P_j[P_1^{-1}\{P_1(0)\}]$. From (1) we have that

$$P_1^{-1}(p) = 0.588 \ln [(p-0.2)/(1-p)] - 1,$$

so that the j-th coordinate a point on $\mathscr{M}$ corresponding to a first-item success probability of p is

$$P_j^*(p) = c_j + (1-c_j)/(1+\exp [-1.7a_j(\ln [(p-0.2)/(1-p)]) - 1 - b_j)]).$$

Since there is a one-to-one correspondence between positions in $\mathscr{M}$ and $[0, 1]$ defined by this relation, the coordinate map is then essentially the relation between $\mathscr{M}$ and

the whole real line defined by $P_1^{-1}(p)$.

2.1.2 Expected score τ

An interesting choice of coordinate map from a practical perspective would be the projection on the diagonal line in the hypercube, also depicted on in Figure 2. Points on this line correspond to values of the function

$$\tau(\theta) = P_1(\theta) + P_2(\theta) + \cdots + P_n(\theta), \tag{5}$$

or simply the value of the expected total score. This is guaranteed to be a differentiable monotone function of curve position by monotonicity, and has the distinct advantage of being directly estimable from data, of being symmetric in the indices j, and perhaps most importantly of being easily understood by the user community. Indeed, Ellis and Junker (1995) have proven that, for all practical purposes, total score becomes equivalent to any latent variable θ as the number of monotone items increases without limit. Total score was also used Ramsay (1991) to order examinees as a part of a smoothing procedure for nonparametric estimates of item response functions, and technique that has been shown by Douglas (1995) to provide a consistent estimate of item response functions.

The interval X for τ is $[0, n]$, and the expression for the j-th coordinate in terms of τ of a point in $\mathscr{M}$ in terms of score value x is then defined by the n functions

$$P_j^*(x) = P_j[\tau^{-1}(x)]. \tag{6}$$

Although the expected total score is expressible analytically if 3PL curves are used, computation of its inverse requires numerical techniques such as linear interpolation, but this should not present any difficult computational problems provided the slope of the expected total score function is not too close to zero.

The use of τ as a measure of ability has one useful advantage; since expected and observed scores are on the same metric, one can plot the proportion of examinees having a fixed observed score who pass an item on the same plot as the item response function, and gives a visual impression of the goodness-of-fit of the model. In fact, this sort of display has been used as an alternative to item response function estimation for test administrations involving very large numbers of examinees, and is called the item-test regression function. Ramsay (1991) exploited this approach in comparing goodness-of-fits of nonparametric estimates to parametric fits.

Figure 3 displays 25 three-parameter logistic item response functions estimated by maximum marginal likelihood estimation (Bock and Aitkin, 1980) from data from 2735 examinees for the quantitative subscale of the General Management Aptitude Test (GMAT). The left panel of Figure 4 indicates the expected score function $\tau = \Sigma_j P_j(\theta)$, and the right panel contains the 25 item response functions plotted against expected total score. We see that the τ measure has eliminated regions over which almost all curves are nearly flat.

Note, however, that use of the expected total score coordinate map should not be seen as justifying the use of observed total score as a means of estimating an
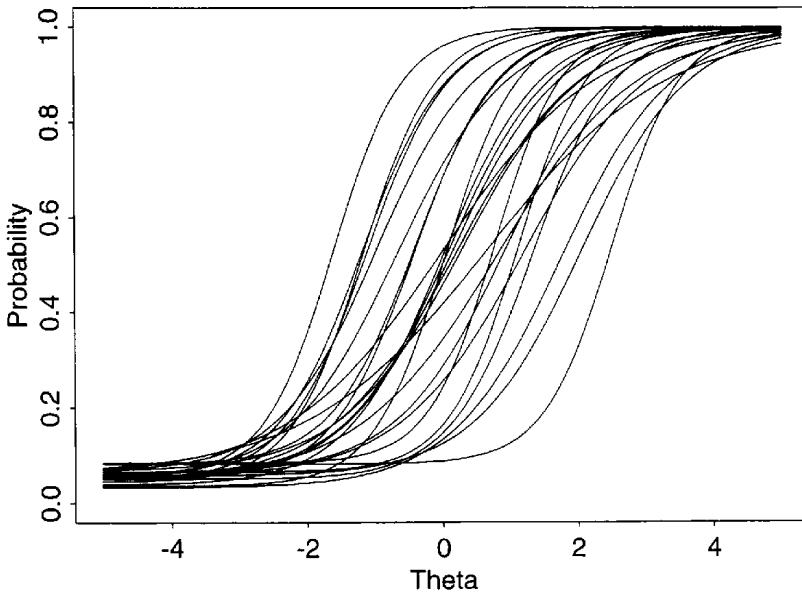
Fig. 3   Three-parameter logistic item response curves estimated for 25 items in the General
Management Aptitude Test estimated for an administrations to 2735 examinees.

examinee's position on $\mathcal{M}$; modern test theory has developed more efficient tech-
niques such as maximum likelihood and Bayesian estimation.

Each of the coordinate maps θ, γ, ψ, τ, or fixed item success probability P_j,
involve constructing a relationship between a continuum X known to be
monotonically related to position along response probability manifold $\mathcal{M}$.   In the
case of the latter two, the continuum is a function of the set of probabilities, P_1, ...,
P_n, and therefore firmly anchored in observables in the sense that we can collect
enough data in principle to independently define to an arbitrary precision location
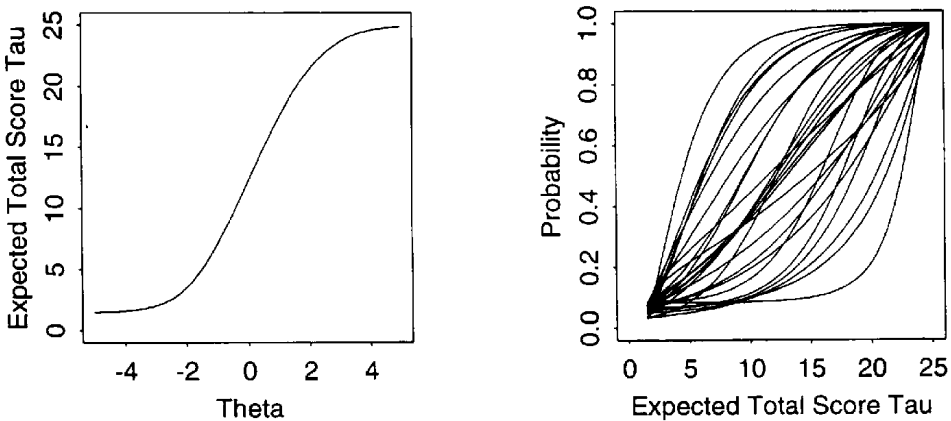on the independent variable dimension.   Levine (1984) has explored in a very

Fig. 4   The left panel shows the expected total score τ for the GMAT Quantitative scale, and
the right panel displays the item response functions as functions of τ.

systematic way the construction of functions of this nature in his account of what he calls *formula score theory*.

We prefer, therefore, these last two possibilities, τ and P_j, because they are not latent and because the merits and demerits of choosing one of them (or some other function of response probabilities) are likely to be clearer to users. All five, however, are *extrinsic* to the manifold in the sense that the relationship of position on the domain of the coordinate map is only indirectly related to position within the manifold itself.

2.2 The intrinsic coordinate map : Arc length

We may also measure ability or proficiency in terms of distance s along the manifold from its beginning, termed *arc length*. This measure is therefore intrinsic to the manifold itself, and is in certain senses a "natural" measure. Arc length has the following expression in terms any extrinsic map with argument x

$$s(x) = \int_{x_0}^{x} \sqrt{\sum_{j}^{n} [DP_j(u)]^2}\, du, \tag{7}$$

where the notation $DP_j(u)$ refers to the value of the first derivative of P_j with respect to independent variable u, and where the lower limit x_0 is the lowest possible value of x, including possibly $-\infty$. This may also be put more compactly as

$$s = D^{-1} \| D\mathbf{P} \|,$$

where $\mathbf{P}$ is the vector of n item response functions, and D^{-1} denotes the operation of partial integration. Since s is a differentiable and monotonic transformation of X, its inverse $\phi = s^{-1}$ defines the coordinate map directly. Moreover, the measure has the advantage over expected score of not requiring monotonicity of item
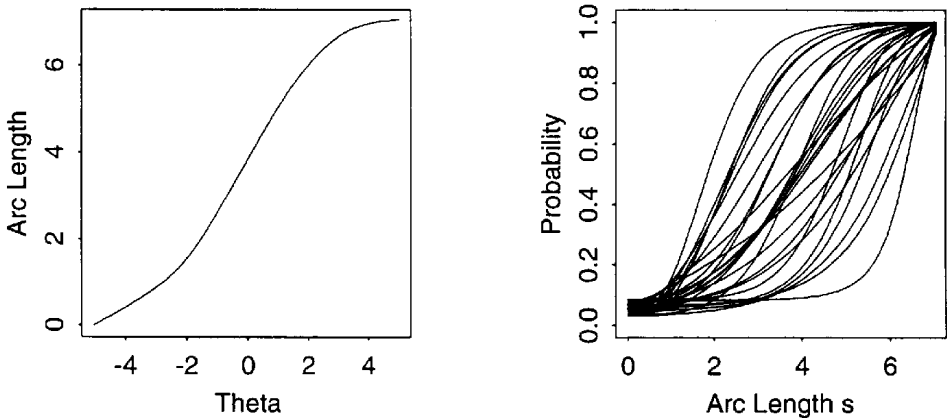

Fig. 5   The left panel shows the relationship between the θ measure of ability and arc length s, and the right panel plots the item responce functions with respect to arc length.

response functions, since the integrand in (7) will be nonnegative, and hence the integral nondecreasing, even when some $DP_j(\theta)$ values are negative.

For example, for the 3PL family, this amounts to

$$s(\theta) = \int_{-\infty}^{\theta} \sqrt{\sum_{j}^{n} a_j^2(1-c_j)^2[\exp 1.7a_j(u-b_j)]^2/[1+\exp 1.7a_j(u-b_j)]^4}\, du. \qquad (8)$$

Figure 5 displays in the left panel the relationship between arc length s and θ for the 25-item GMAT test, where the integration in (8) was carried out numerically. The item response functions of s are given in the right panel, where we see that probabilities of success do not change much for values of s between 0 and 1. An item is clearly easy if its probability approaches one for low to medium values of s, and difficult if there is substantial change only for very high s values. Figure 6 shows the probability density functions for ability for the GMAT test in terms of both θ and arc length. These probability density functions are a result of applying a kernel density smoothing procedure to the values of θ estimated by maximum likelihood for each of the 2735 examinees.

Another reason to prefer arc length as an ability measure is that certain important quantites are considerably simplified when expressed in terms of s. The length of the tangent vector $DP^*(s)$ is automatically unity; $\parallel DP^*(s) \parallel = 1$ for all s. This is because, by the chain rule and (7)

$$DP_j(x) = DP_j^*(s)D_s(x) = DP_j^*(s) \parallel D\mathbf{P}(x) \parallel$$

so that

$$DP_j^*(s) = DP_j(x)/\parallel D\mathbf{P}(x) \parallel.$$

From a practical perspective $DP_j(s)$ is *item discriminability*, the change in success proability for item j that will result from a small increase in s, and because
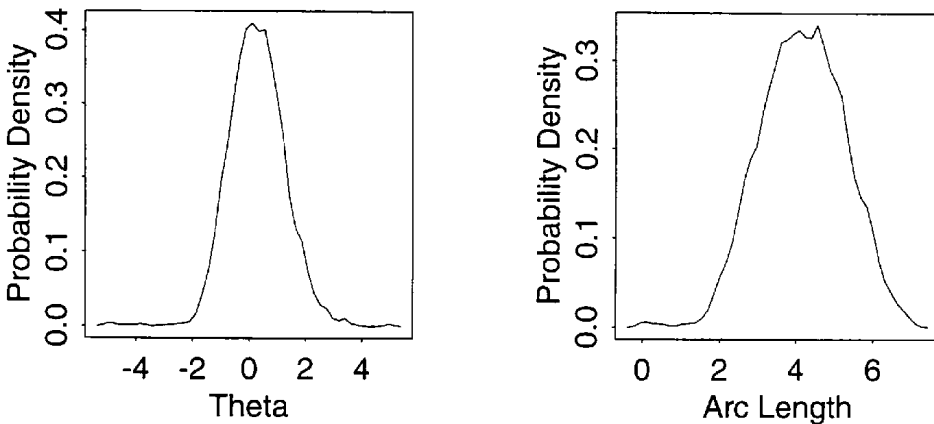

Fig. 6   The left panel shows a kernel density estimate of the probability density function for the maximum likelihood estimates of the θ measure of ability for the GMAT test, and the right panel plots the probability density with respect to arc length.
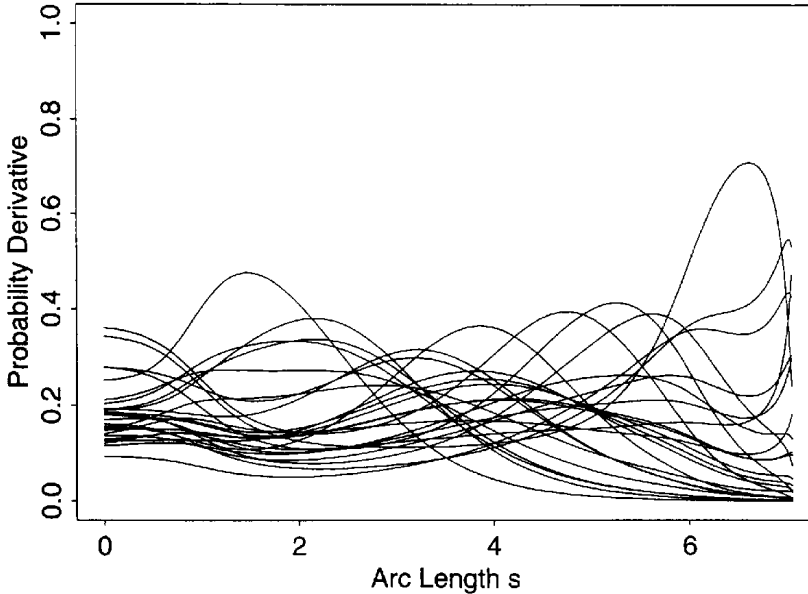
Fig. 7   The functions $DP_j(s)$, the impacts of small increases in arc lengh on item success probabilities. These have unit sums of squares.

this is bounded below by 0 because of monotonicity and above by 1, $DP_j(s)$ can be interpreted as a scale-free "importance" measure of improving one's performance by Δs in the same manner as for a correlation coefficient. Figure 7 displays the $DP_j(s)$ curves for the GMAT test.

The practical problems that confront test developers can be dealt with in terms of arc length measure without the need to assume a fixed metric for θ. It is easy to show that there is a one-to-one correspondence between arc length along a manifold $\mathscr{M}_0$ within a subspace $\mathscr{P}_0$ of a response probability space and arc length along the manifold $\mathscr{M}$ within the over-space $\mathscr{P}$ provided that all item characteristic functions are strictly monotonic. This implies that one can construct an individual's position in $\mathscr{M}$ knowing his position within $\mathscr{M}_0$ defined by the subset of items administered, for example, in a computerized adaptive testing situation.

Similarly, differential item functioning (DIF) results in manifolds for subgroups that do not belong to the same trajectory, and may as easily be assessed in terms of s as in terms of any other measure. But it is interesting to speculate whether two non-overlapping manifolds for males and females, for example, which are nevertheless of the same length and terminate in the same place should be regarded as really problematic. This issue is considered further at the end of the paper.

2.3   Item and test information functions

The test information function is essential to test appraisal since it is for modern test theory the counterpart of reliability in classical test theory. Defined relative to an arbitrary ability measure θ it is
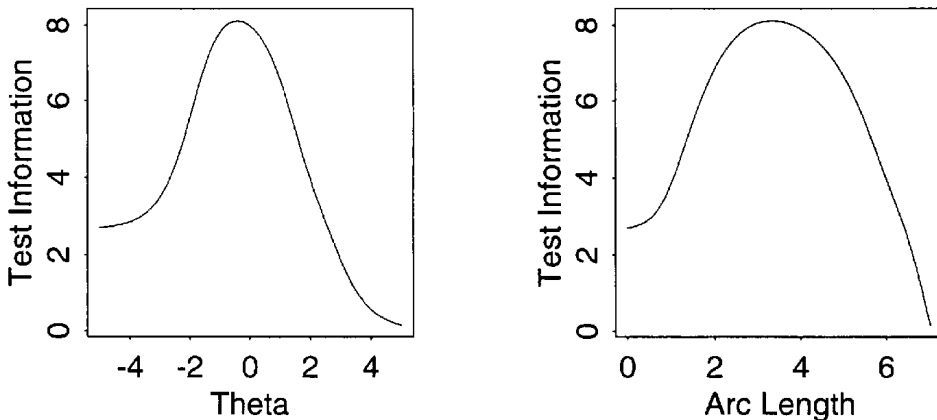
Fig. 8   The left panel displays the test information function for the GMAT test with respect to θ measure, and the right panel with respect to arc length measure.

$$I(\theta)=\sum_j [DP_j(\theta)]^2/[P_j(\theta)(1-P_j(\theta))]. \qquad (9)$$

The error standard deviation function

$$\sigma(\theta)=1/\sqrt{I(\theta)}$$

is the lowest achievable standard error of estimate of an individual's ability given that his true ability is θ, and therefore assesses directly the stability of an ability estimate. However, since $I(\theta)$ depends upon the first derivative of the item response functions with respect to θ, it is not invariant under monotone transformations. It's re-expression with respect to transformed value $\gamma=g(\theta)$ is

$$I^*(\gamma)=I(\theta)/[D\gamma(\theta)]^2, \qquad (10)$$

so that for arc length measure it becomes simply

$$I^*(s)=I(\theta)/\sum_j DP_j^2(\theta)=I(\theta)/\parallel D\mathbf{P}(\theta)\parallel^2. \qquad (11)$$

Figure 8 displays the test information functions for the GMAT test relative to the θ measure and arc length.

3.   Extensions and discussion

The discussion of $\mathscr{A}$ so far has been predicated on items being scored dichotomously. How might the hyper-cubic probability response space concept be extended to items involving more than one category of response? Although this paper confines itself to considering two-state items, we can get some intuitions by considering a test consisting of a three-category item and a two-category item. The three response probabilities for the former must, of course, sum to one, and consequently they define a point lying within a two-dimensional equilateral triangle having unit sides. Crossing this structure with the unit interval representing

success on the second item, the result is a prism with triangular ends and a square base. Similarly, two three-option items define a four-dimensional structure having three-dimensional faces that are prisms. For further discussion of these concepts, one can consult Ramsay (1995), who uses these notions to define a non-dimensional item analysis technique. But however complex and unimaginable the resulting geometry of the response probability space may be, the response probability manifold $\mathcal{M}$ remains a space curve, or a smooth one-dimensional structure within this space.

When items are polytomous, the concept of monotonicity no longer plays a particularly important role either theoretically or practically. At this point arc length possibly becomes the most obvious manifest measure of ability. Its expression in terms of a set of item response functions $P_{jm}(x)$ is now

$$s(x) = \int_{x0}^{x} \sqrt{\sum_{j}^{n} \sum_{m}^{M_j} [DP_{jm}(u)]^2}\, du. \tag{12}$$

The concept of $\mathcal{M}$ as a space curve obviously has many interesting extensions. It could be postulated to be a manifold of dimension K, meaning that there exists a family of coordinate maps that are diffeomorphic to R^K. Or, it may be that the concept of a manifold of fixed dimensionality should be replaced by something broader, such as manifolds for which dimensionality changes smoothly from one value to another as a function of position within $\mathcal{M}$. It might be argued, for example, that examinees occupying positions in the response probability space far from some vertex representing ideal behavior exhibit higher dimensional variation than those in its immediate neighborhood. Ramsay (1995) has described a technique for estimating the response probability space position of individual examinees using similarity-based smoothing that seems promising in terms of identifying more complex structures of this kind.

The main goal of this paper has been to offer a mathematically unambiguous account of item response theory. This will appeal, it is hoped, to those who share the author's uneasy feeling about exactly what a latent trait "means", especially when interacting with end-users of psychometric technology who lapse very easily into imagining that item response theory has provided a metric for measuring ability where none existed before. The latent trait concept is inessential to item response theory ; it is the geometrical notion of a manifold within probability response space that is critical. A particular realization of a set of item response functions $P^*(x)$ is no more than one among an arbitrarily large number of coordinate maps. The existence of a coordinate map is critical to the concept of a manifold, but the characteristics of a specific map are incidental.

It is the goal of the theory of differentiable manifolds to describe the characteristics of a particular manifold in ways that are invariant with respect to particular choices of coordinate map. Unidimensional item response theory presents the simplest of cases : a one-dimensional space curve. The natural and, it can be

shown, only invariant characterization is in terms of arc length, s. Measured in terms of arc length, ability does indeed have metric properties: sums and differences of arc lengths have clearly defined meanings in terms of navigating within the probability response space along curve $\mathscr{A}$. Moreover, there is nothing whatever that is latent about arc length; given appropriate data from replicated responses, one can estimate arc length empirically arbitrarily well. The methods of Ramsay (1995) serve this end rather well even when responses are not replicated.

Nevertheless, one can still ask whether the metric implicit in the arc length measure would be actually interesting to a non-psychometrician. Perhaps not. After all, what one really wants is some measure of the effort or cost of moving from position s_1 to position s_2 along $\mathscr{A}$ rather than merely the distance between the two points. How much time does my child have to spend in school to reach the level of proficiency represented by a specific value s? How much money must be fed into a city's educational system to raise the average proficiency of a specific target group from s_1 to s_2? These, it would seem, are the really interesting metric questions, and they cannot be resolved within the context of test data alone.

Appendix

The Appendix offers a technical definition of a differentiable manifold, following Boothby (1975).

An *m-manifold* $\mathscr{A}$ is a space that is locally Euclidean, meaning that each point has a neighborhood that is topologically equivalent or homeomorphic to an open subset of R^m. It must also possess the more primitive properties of being separable in the sense that any two distinct points have non-intersecting neighborhoods, and of having a countable basis of open sets. A line or curve segment is a one-manifold that is globally Euclidean, but a circle is an example of a one-manifold that is not. But note that there is nothing in the definition of $\mathscr{A}$ that requires it to be a subset of or to be embedded in a Euclidean space of higher dimension.

Let U be an open set of $\mathscr{A}$ and let $\phi: U \to R^n$ be a homeomorphism from U to an open subset of R^m, termed a *coordinate map*. Each pair (U, ϕ) is called a *coordinate neighborhood*. Let (V, ψ) be a second coordinate neighborhood such that the intersection $U \cap V$ is nonempty. Then for any point $p \in U \cap V$ there are two coordinate map values, $\phi(p)$ and $\psi(p)$. Since both coordinate maps are invertible because they are homeomorphisms, one can move from one map value to the other by the two transformations $\psi \circ \phi^{-1}$ and $\phi \circ \psi^{-1}$. These two transformations are themselves homeomorphisms from one open subset of R^n to another, and in effect define two changes of coordinate system for $U \cap V$.

A *diffeomorphism* is an infinitely differentiable homeomorphism between two subsets. Two coordinate neighborhoods (U, ϕ) and (V, ψ) are C^∞-*compatible* if the transformations $\psi \circ \phi^{-1}$ and $\phi \circ \psi^{-1}$ are diffeomorphisms in addition to being homeomorphisms. That is, one can change from one coordinate sysem to the other

infinitely smoothly.

Finally, a *differentiable manifold* $\mathcal{M}$ possesses a family of coordinate neighborhoods (U_a, ϕ_a) such that

1. the U_a cover $\mathcal{M}$,
2. any two intersecting coordinate neighborhoods in the family are C^∞-compatible, and
3. any coordinate neighborhood (V, ψ) compatible with every (U_a, ϕ_a) with which it intersects is also in the family.

Acknowledgments

REFERENCES

Bock, R.D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an *EM* algorithm, *Psychometrika*, **46**, 443-459.

Boothby, W.M. (1975). *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press.

Douglas, J. (1995). Joint consistency of nonparmetric item characteristic curve and ability estimation, *Psychometrika*, in press.

Ellis, J.L. and Junker, B.W. (1995). Tail-measurability in monotone latent variable models. Carnegie-Mellon University, unpublished manuscript.

Fischer, G.H. (1995). Derivations of the Rasch model. In G.H. Fischer and I.W. Molenaar (Eds.) *Rasch Models: Foundations, Recent Developments, and Applications*. New York: Springer-Verlag.

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

Junker, B.W. (1993). Conditional association, essential independence and monotone unidimensional item response models, *Annals of Statistics*, **21**, 1259-1378.

Levine, M.V. (1984). *An Introduction to Multilinear Formula Score Theory* (Report 84-4). Champaign, IL: University of Illionois, Department of Educational Psychology, Model-Based Measurement Laboratory.

Ramsay, J.O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, **56**, 611-630.

Ramsay, J.O. (1995). A similarity-based smoothing approach to nondimensional item analysis. *Psychometrika*, **60**, 323-339.