教育心理学研究, 2001, 49, 491—499

多次元項目反応理論での相関のある特性値の線形結合に関するテスト情報関数

星 野 崇 宏[1]

教育評価・心理測定では測定対象である特性に複数の下位特性がある場合がしばしばである。また，特に入学試験などの何らかの決定を求められる場合は，複数の下位特性をそのまま多次元として評価するのではなく，測定者が設定した重みによる下位特性の線形結合という1次元の評価がなされることがある。本研究では相関のある複数の特性の線形結合についての評価を容易にするために，項目反応理論を用いたテスト編集において重要な意味をもつテスト情報関数，及び項目情報関数の提案を新たに行った。本研究で提案された情報関数は局外母数の積分消去により，多次元の特性があってもテストを行う前に1次元で評価でき，特性間の相関を考慮できるという点で複数の下位特性を有する特性の評価のためのテスト編集を可能にする。また本研究では，異なる重みや相関係数の下での情報関数の比較のための方法が提案された。数値例によって重み付け，相関係数を様々に変化させても導出できることが確認され，本方法による多次元項目反応理論での項目選択法の有用性が示唆された。また，アメリカ国立教育統計センターの作成した幾何学と統計学についての項目プールから選択した項目群に対して本方法が適用された。

キーワード：項目プール，適応型テスト，テスト編集，多次元項目反応モデル

問題と目的

多次元項目反応理論におけるテスト情報関数

教育評価や心理測定の現場で利用されている，能力測定を目的としたテストや，性格特性を測定する尺度の大部分が下位テストや下位尺度を含む。これは，測定の対象である構成概念が，多くの場合広がりを有する為である。例えば，英語の能力を測定する際には，「読解」「文法」「会話」「リスニング」などの，英語の能力を構成すると思われる，いくつかの互いに相関を持った構成概念を測定する下位テストが作成される。ここで古典的テスト理論に基づいてテストを利用する際は，各下位テスト（下位尺度）に対する得点が算出され，テスト使用者が事前に設定した重み付けを用い，テストの総得点が算出される。英語検定や数学検定，入学試験における合否判定や，能力・性格特性の判定などの決定は，しばしばこの重み付けによる総得点によって行われることがある。

しかし，古典的テスト理論では，被験者集団の性質とテストの項目の性質を分離することが困難である。これに対して近年様々な場面で利用されてきている項目反応理論（池田, 1994；Lord, 1980；Lord & Novick, 1968；芝, 1991；繁桝, 1998）は，個人の特性(能力値)とテスト項目についての特性を分離してモデリングが出来るため，適応型テスト（Adaptive testing）など様々な新しいテスト形態を提供するための基礎理論を与えることができる。

では，項目反応理論の観点から，下位テストや下位尺度が存在する状況における能力・性格特性の判定はどのように行うべきであろうか？

テスト項目群が複数の特性を測定している場合の項目反応理論は，多次元項目反応理論（Multidimensional item response theory）と呼ばれる。多次元項目反応理論モデルは，1次元を測定している項目反応理論モデルの自然な拡張であり，これらは1因子の因子分析モデルと多因子の因子分析モデルの関係と同じである。1つの特性を測定する1次元の項目反応理論を用いた教育評価や心理測定では，識別力や困難度などの項目パラメータが既知のテスト項目の集合である項目プールから，テスト使用者がテストの目的に従って項目選択を行うことがしばしばある。これはテストの編集と呼ばれるが，この過程で有用な情報を与えるのがテスト情報関数である。これは特性値の関数であり，特性値を所与とした場合の特性値の推定量の漸近分布の分散

[1] 東京大学総合文化研究科／日本学術振興会
bayesian@jasmine.ocn.ne.jp
本研究は，日本学術振興会科学研究費補助金の補助を受けて行われた。

の逆数であるため，テスト使用者にとっては，これを用いると特性のどの値（またはその近傍）でそのテストにどれくらいの精度があるかがわかる。そこで，テスト使用者はその目的に従って目標となるテスト情報関数の大きさや形を設定し，それに近づくように項目プールからテスト項目を選択することができる点でテスト情報関数の利用価値は大きい。また，テスト情報関数はデータの関数ではないので，テストを行う前に利用することができる。

さて，1次元の項目反応理論モデルでは有用性が非常に高いとされるテスト情報関数であるが，多次元項目反応理論モデルにおいては測定する特性の次元の正方行列となる。実際の評価測定においては既に論じたように，多次元の特性が測定される状況が多いにもかかわらず，多次元項目反応理論モデルの利用があまりされない。その一因として，このテスト情報関数が行列形式になるために実際の評価・測定場面での利用が困難であることがあると思われる。

望ましいテスト情報関数：本研究の目的

ここで特性が多次元にわたる場合の望ましいテスト情報関数は

1　関心のある特性の線形結合の値のみ決定すれば情報関数の大きさが決まる。

1次元の場合のテスト情報関数の有用性は，特性値 θ のみの関数となることである。この場合は，テスト情報関数はテスト実施前に利用できる。多次元の特性については，実際のテスト場面で利用される時は，関心の対象がある特定の特性のみか，多次元の特性の線形結合になる場合が多い。従って，テスト情報関数が関心のある特性・線形結合の値のみの関数になっていれば，従来のテスト情報関数と同じ利用法が可能である。そのためには関心のある母数以外の局外母数に関しては情報関数から消去されている必要がある。

2　特性間の相関に関する情報を利用できる。

事前知識として特性間の相関がわかっている場合に，その相関構造を利用せずに各特性が直交しているという仮定の下に作られたテスト情報関数は歪んだ結果を与えるであろう。また，テストを行った集団から得られた相関の推定値を用いることもできれば望ましい。

この2つの条件を満たすものであると思われる。以上の条件が満たされることは，項目選択のために有用である項目情報関数においても望ましい条件である。

多次元項目反応理論におけるテスト情報関数（または項目情報関数）に関する先行研究は，主に適応型テスト（Adaptive testing）における項目選択法の提案を中心としていくつか行われている。しかし，上記に挙げた望ましい性質いずれをも満たすテスト情報関数の導出はいまだなされていない。

Segall (1996) や Luecht (1996) は多次元の適応型テストにおける項目選択のための方法を提案している。これらの研究では，適応型テストにおいて，すでに実施された項目群からテスト情報関数を算出し，さらに能力母数の推定の情報量を最大化するために，次に提示するのに最適な項目を選択することが目的である。テスト情報関数そのものは行列であるので，これらの研究ではテスト情報関数の行列式を最大化することを項目選択の基準としている。特に Segall の研究では，最尤法だけではなくベイズ的な方法もあわせて提案されていて，後者の方法は特性間の相関の情報も考慮することができる。しかし，後者の方法は「対数事後分布の二回微分の，尤度による期待値」というベイズ統計学の規範から逸脱した基準を利用している。また，これらの方法では線形結合の情報関数を導出できない。

van der Linden (1999) は，多次元項目反応理論モデルは多くの場合，テストが実は多次元の特性のベクトル θ そのものではなく，その重み付け和（線形結合）$\boldsymbol{\theta}^* = \mathbf{a}^t\boldsymbol{\theta}$ を測定することを目的として利用されることが多いことに注目し，$\boldsymbol{\theta}^*$ の推定量の漸近分散，つまりテスト情報関数 $I(\boldsymbol{\theta}^*)$ の逆数を

$$V(\hat{\boldsymbol{\theta}}^*|\boldsymbol{\theta}) = I(\boldsymbol{\theta}^*)^{-1} = \mathbf{a}^t I(\boldsymbol{\theta})^{-1}\mathbf{a} \tag{1}$$

とした。（ここで t は行列やベクトルの転置を表す。）

しかしこの方法では θ の各要素の相関についての情報をくみいれることはできない。特に英語のテストの例などのように，各下位テストに高い相関がある場合，相関を無視した線形結合の情報関数は適当ではないと思われる。また，θ の各要素すべての値を特定しないとテスト情報関数の大きさがわからないので，関心のある θ^* の大きさとテスト情報関数の大きさが一対一対応しない。従って，適応型テストでは，各項目を被験者に実施するごとに，その能力推定値 $\hat{\theta}$ が得られるためにこの方法を用いることができるが，事前に項目プールから適切な項目を選ぶテストの編集に用いるのは難しい。Reckase & McKinley (1991) は，方向微分係数の考え方を用いてすでに同様な方法を提案しているが，これも van der Linden と同様の欠点を有する。そこで，本研究では，特性間の相関を考慮し，かつ局外母数の積分消去の行われた，テスト使用者にとって関心のある能力の線形結合についてのみの関数となる

テスト情報関数を提案し，シミュレーションと具体的な適用例によってその妥当性を検証する。

多次元項目反応モデルとそのテスト情報関数

ここでは項目反応理論モデルを，２値データの２パラメータロジスティックモデルに限定して表現する。(ここでは説明の簡約化のために２値データのモデルを記すが，本研究において提案される方法は名義応答モデル(Nominal response model)や部分採点モデル(Partial credit model)などの多値データに対するモデルにも何の変更なしに適用できる。) K次元の特性を測定する多次元項目反応モデルにおいて，J個の項目に対する反応が，特性値を所与とすると互いに独立の時，特性値を所与とした条件付き尤度は

$$L(u_1,u_2,\cdots,u_J|\theta=(\theta_1,\cdots\theta_K)^t)=\prod_j^J P_j(\theta)^{u_j}Q_j(\theta)^{(1-u_j)} \tag{2}$$

である。

但し，u_j は，項目 j に対して正答するか ($u_j=0$) 誤答するか ($u_j=1$) を表す確率変数であり，θ_k は特性 k についての特性値である。

ここで，特性値 θ の分布を $p(\theta)$ とおくと，周辺尤度 $L(u_1,u_2,\cdots,u_J)$ は，

$$L(u_1,u_2,\cdots,u_J)=\int L(u_1,u_2,\cdots,u_J|\theta)p(\theta)d\theta \tag{3}$$

となる。

またここで，$P_j(\theta)$ は θ を所与とした場合の正答確率である。これを２パラメータのロジスティック回帰モデルで表現するなら

$$P_j(\theta)=\frac{\exp D(w_{j1}\theta_1+w_{j2}\theta_2+,\cdots,+w_{jK}\theta_K+b_j)}{1+\exp D(w_{j1}\theta_1+w_{j2}\theta_2+,\cdots,+w_{jK}\theta_K+b_j)} \tag{4}$$

である (Reckase, 1996)。但し，w_{jk} は第 j 項目の第 k 特性の識別力パラメータであり，b_j は第 j 項目の困難度パラメータ，そしてDはロジスティック特性曲線の係数である。

教育評価などでは各教科はそれぞれ１つの能力・特性のみ測定している，という仮定がしばしば置かれる。この場合は特性 k のみ測定するとされる項目において，θ_k 以外の特性への識別力パラメータを０に固定したモデルとなり，これは明らかに多次元項目反応理論モデルの下位モデルとなる。つまり，条件付き尤度は，

$$L(u_1^1,\cdots,u_{J1}^1,\cdots,u_1^K,\cdots,u_{JK}^K|\theta=(\theta_1,\cdots,\theta_K)^t)$$
$$=\prod_k^K\prod_j^{J_k}P_j(\theta_k)^{u_j^k}Q_j(\theta_k)^{(1-u_j^k)} \tag{5}$$

となることがわかる。但し，u_j^k は特性 k を測定する j 番目の項目についての反応についての確率変数，J_k は特性 k を測定する項目の数であり，$J=\sum_{k=1}^K J_k$ である。後述の適用例など，本方法が用いられる場合の多くにおいてはこのモデルの適用が適切であると思われる (本方法の適用範囲はこの下位モデルに限定されない)。

また，多次元項目反応モデルにおけるテスト情報関数の行列の第 st 成分は，フィッシャー情報行列の θ に関する部分行列の第 st 成分であることを考えると，

$$I_{st}(\theta)=\sum_j^J\frac{\partial P_j(\theta)/\partial\theta_s\times\partial P_j(\theta)/\partial\theta_t}{P_j(\theta)Q_j(\theta)} \tag{6}$$

となる (Segall, 1996)。

特に２パラメータのロジスティックモデルの場合，

$$I_{st}(\theta)=\sum_j^J D^2 w_{js}w_{jt}P_j(\theta)Q_j(\theta) \tag{7}$$

となる

さて，テスト使用者の関心が，複数の特性の線形結合 θ^*，つまり

$$\boldsymbol{\theta}^*=\mathbf{a}^t\boldsymbol{\theta}$$

にあるとする。

ここで，係数 $\mathbf{a}$ は，テスト使用者によって事前に決定されている係数であり，θ^* における各下位テストに対する重要度の「重み」である。ここで，以降の式変形のために $\mathbf{a}^t\mathbf{a}=1$ とするが，この制約は実際の利用上はなんら制約とはならない。θ^* のテスト情報関数 $I(\theta^*)$ を θ のテスト情報関数 $I(\theta)$ から直接導出することはできない。なぜなら，θ はK次元であるのに対して，θ^* は１次元であるため，パラメータ変換によるフィッシャー情報行列の導出の条件が満たされないからである (Stuart & Ord, 1987)。従って，まずは θ^* を第１成分とする，θ と同じ次元Kを有するパラメータ $\theta^*_{\mathrm{all}}=A\theta$ への変換を行い，パラメータ変換後のフィッシャー情報行列 $I(\theta^*_{\mathrm{all}})$ を導出する。

ここで，Aはその第１列が $\mathbf{a}^t$ であるような$K\times K$の直交行列とする (これは後述の式変形のための設定である)。a が定まっても，Aの２行以降が一意に決まらない。しかしここではこの不定性を考慮しなくても，以降２行以降が結果として消去されるので，a に直交するように適当に決定しても問題はない(後述)。また，後の記法の簡約化の為に，

$$A=\begin{pmatrix}\mathbf{a}^t\\ A\_\end{pmatrix} \tag{8}$$

とする。

従って，

$$\theta^*_{\mathrm{all}}=A\theta=\begin{pmatrix}\theta^*\\ \theta^*_{\mathrm{nu}}\end{pmatrix} \tag{9}$$

となる。ここで θ^*_{nu} は $\theta^*_{\mathrm{nu}}=A\_\theta$ となる$K-1$次元のベクトルであり，後に積分消去される局外母数である。ここで，$\mathbf{a}^t\mathbf{a}=1$ である必要がある。また，任意に θ の

要素（各特性・科目）に対する重みを付けた場合は，これを正規化して a とする必要がある。

このとき，容易に

$$I(\theta^*_{\mathrm{all}})=AI(\theta)A^t \tag{10}$$

であることがわかる。

ここで $I(\theta^*_{\mathrm{all}})^{-1}$ の11成分は van der Linden (1999) の提案した $\mathbf{a}^t I(\theta)^{-1}\mathbf{a}$ と同じになるが，これには局外母数 θ^*_{nu} が含まれる。そこで，θ^*_{nu} を積分消去することを考える。ここで，$\hat{\theta}^*_{\mathrm{all}}$ は θ^*_{all} の，$\hat{\theta}^*$は θ^*のそれぞれ推定量とすると，漸近的に

$$V(\hat{\theta}^*_{\mathrm{all}}|\theta^*_{\mathrm{all}})=I(\theta^*_{\mathrm{all}})^{-1} \tag{11}$$

が成立することから，

$$V(\hat{\theta}^*_{\mathrm{all}}|\theta^*)=\int V(\hat{\theta}^*_{\mathrm{all}}|\theta^*_{\mathrm{all}})p(\theta^*_{\mathrm{nu}}|\theta^*)d\theta^*_{\mathrm{nu}}=\int I(\theta^*_{\mathrm{all}})^{-1}p(\theta^*_{\mathrm{nu}}|\theta^*)d\theta^*_{\mathrm{nu}} \tag{12}$$

となる。よって，

$$V(\hat{\theta}^*|\theta^*)=\int I(\theta^*_{\mathrm{all}})^{-1}_{11}p(\theta^*_{\mathrm{nu}}|\theta^*)d\theta^*_{\mathrm{nu}} \tag{13}$$

従って，

$$I(\theta^*)^{-1}=\int I(\theta^*_{\mathrm{all}})^{-1}_{11}p(\theta^*_{\mathrm{nu}}|\theta^*)d\theta^*_{\mathrm{nu}}=\int \mathbf{a}^t I(\theta=A^t\theta^*_{\mathrm{all}})^{-1}\mathbf{a}p(\theta^*_{\mathrm{nu}}|\theta^*)d\theta^*_{\mathrm{nu}} \tag{14}$$

ここで，$I(\ )^{-1}_{11}$ は $I(\ )$ の逆行列の11成分である。

ここで θ の分布が，データから，またはテスト作成・標準化において平均 μ，共分散Σの正規分布に従うと特定されているとする。

すると，

$$\theta^*_{\mathrm{all}}=A\theta\sim N\left(\begin{pmatrix}\mathbf{a}^t\mu\\ A\_\mu\end{pmatrix},\begin{pmatrix}\mathbf{a}^t\Sigma\mathbf{a} & \mathbf{a}^t\Sigma A^t\_\\ A\_\Sigma\mathbf{a} & A\_\Sigma A^t\_\end{pmatrix}\right) \tag{15}$$

となり，

$$\mu_0=A\_\mu+A\_\Sigma\mathbf{a}\{\mathbf{a}^t\Sigma\mathbf{a}\}^{-1}(\theta^*-\mathbf{a}^t\mu),$$
$$\Sigma_0=A\_\Sigma A^t\_-A\_\Sigma\mathbf{a}\{\mathbf{a}^t\Sigma\mathbf{a}\}^{-1}\mathbf{a}^t\Sigma A^t\_ \tag{16}$$

とおくと，$\theta^*_{\mathrm{nu}}|\theta^*$ の分布は平均 μ_0，共分散行列 Σ_0 の $K-1$ 次元多変量行列に従う。ここで，θ^*_{nu} についての積分は，$\theta^*_{\mathrm{nu}}|\theta^*$ の分布から乱数を発生させることによるモンテカルロ平均，または区分求積法(Mislevy, 1984)などによって十分近似が可能である。

例えば区分求積法では，$\Sigma_0=BB^t$ とし，$\theta'=B^{-1}(\theta^*_{\mathrm{nu}}-\mu_0)$ とすると，θ' の分布は平均 0，共分散行列 I の正規分布となり，

$$I(\theta^*)^{-1}=\int \mathbf{a}^t I\left(\theta=A^t\begin{pmatrix}\theta^*\\ B\theta'+\mu_0\end{pmatrix}\right)^{-1}\mathbf{a}N(\theta'|0,I)d\theta'$$
$$\approx\sum_k^q\ \mathbf{a}^t I\left(\theta=A^t\begin{pmatrix}\theta^*\\ BR_k+\mu_0\end{pmatrix}\right)^{-1}\mathbf{a}\,W(R_k) \tag{17}$$

と近似できる。ここで Q_k は区分求積点のベクトル，$W(R_k)$ はそれに対応する重み，q は求積点の数である。

ここで $A^t\_A\_=I-\mathbf{aa}^t$ の関係が成立することから，

$$\theta=a\theta^*+A^t\_BR_k+(I-\mathbf{aa}^t)(\mu+\Sigma\mathbf{a}(\mathbf{a}^t\Sigma\mathbf{a})^{-1}(\theta^*-\mathbf{a}^t\mu)) \tag{18}$$

となる。従って，$A_$が関係するのは右辺第2項だけであるが，ここで $B=A\_B^*$（ここで B^* は $A_$の関数でない）となり，第2項の A^t_B の部分は $(I-\mathbf{aa}^t)B^*$ となり $A_$は関係しない。よって，前述したように，$A_$の不定性があっても，$I(\theta^*)$ は同一の値をとる。

ここで，区分求積法を用いると，求積点が少ない場合は，$I(\theta^*)$ の曲線が θ^* の関数として明示的に表せる。しかし，区分求積法は次元が大きいと近似の精度は悪くなる（Meng & Stephen, 1996）。さらに，θ の次元が大きいと，I の逆行列の要素を I の要素によって明示的に表せなくなる。従って特性の次元が大きいときはモンテカルロ平均を用いるとよい。（参考のため，付録1に2次元の時の θ^* のテスト情報関数 $I(\theta^*)$ を明示的に表した。）

異なる重み付けや相関の大きさのときの情報関数の比較

本研究で提案された，特性の線形結合 θ^* の情報関数は，あくまで所与の重みと相関 $\boldsymbol{a}, \Sigma$の下でのものであり，異なる重みと相関 $\boldsymbol{a}, \Sigma$においては比較は無意味である。なぜなら，情報関数の定義より，漸近的に，

$$V(\hat{\theta}^*|\theta^*)=\frac{1}{I(\theta^*)} \tag{19}$$

であり，このことから θ^* の分散が同じ場合は情報量の大小がテストの信頼性の大小に直結する（Sijtsma & Molenaar, 1987 ; Nicewander, 1993）。しかし分散が異なる場合，直接 θ^* の情報関数の大小から $\hat{\theta}^*$ の推定精度を比較しても意味が無い。また，θ^* の分散は $Var(\theta^*)=\mathbf{a}^t\Sigma\mathbf{a}$ より，a を変化させると一般に θ^* の分散は変化する。従って，重み付けや相関の大きさが異なった場合の情報関数間の比較をするためには，θ^* の分散を1にするように，$\theta^*_N=\theta^*/(\sqrt{\mathbf{a}^t\Sigma\mathbf{a}})$ と置き，θ^*_N の情報関数

$$I(\theta^*_N)=I(\theta^*)\times\mathbf{a}^t\Sigma\mathbf{a} \tag{20}$$

を比較すればよい（この方法はテスト情報関数の「標準化」と呼ぶことができる）。

基準集団準拠検査への適用

また，本方法によって，θ^* についてのテスト情報関数だけでなく，各項目の θ^* についての項目情報関数も同様に作成できる。従って，基準集団準拠検査において，受験者集団全体の上位 α%とそれ以外を識別し

たい場合は，項目プールから $p(\theta^*)$ の上位 α%点に対応する θ^*_α の近傍で項目情報関数の値の大きい項目を選ぶことによって，最適なテストの構成が可能である。また，１つの特性のみ測定していると仮定される(つまり，測定している特性以外の識別力パラメータがゼロ）場合や，２パラメータロジスティックモデルを採用した場合，項目情報関数の行列の対角要素のいくつかがゼロになったり，行列式がゼロになり，逆行列の計算ができなくなるため，本方法を用いて線形結合の項目情報関数を計算できない(付録２参照)。従ってその場合は，既に選択された項目群からなる，特性の線形結合のテスト情報関数と，その項目を入れた後の特性の線形結合のテスト情報関数の差が最大になるような項目を選ぶという方法をとればよい。

ある特性が他の特性のテスト情報関数に与える影響

$\boldsymbol{\theta}^*=\mathbf{a}^t\boldsymbol{\theta}$ の $\mathbf{a}$ を，ある成分のみ１で他が０であるようなベクトルにすることにより（つまり A を permutation matrix にすることにより)，本方法から特性値の線形結合だけではなく，各特性の周辺的なテスト情報関数を構成できる。従って，特性 j のテスト情報関数における特性 k の影響なども論じることができる。これは，特性間に相関がある場合は重要な情報を与えてくれる。また，θ^* の次元は１次元である必要はないので，K 次元の特性を測定するテストにおける別々の２つや３つの特性の関係を知ることもできる。

数 値 例 1

ここでは，２次元の特性 θ_A, θ_B をそれぞれ独立に測定する２パラメータロジスティックモデルを仮定する項目群A，B（各８項目）が存在し，特性が二変量正規分布（どちらも分散１）に従うとする。項目群A，Bの項目パラメータは，全て識別力が２（測定していない特性に対しては０）であり，困難度は両群とも12，６，４，０，０，－４，－６，－12である（普通の１次元 IRT の表記では −6, −3, −2, 0, 0, 2, 3, 6)。２つの特性の相関をいくつかに変化させたときの θ^* のテスト特性関数をプロットし，重みが４：１のときを FIGURE 1 に，２：１のときを FIGURE 2 に，１：１のときを FIGURE 3 に示した。それぞれ横軸は θ^* の大きさ，縦軸はテスト情報関数の値である。同様に，特性の相関が0.3の時，重み付けを４：１，３：１，２：１，１：１と変化させ，それぞれの θ^* の情報関数をプロットし，FIGURE 4 に示した。ここで，FIGURE 1 から４までテスト情報関数が多峰になっているのは項目の困難度の散らばりが大きいからである。ここであえて多峰になるように項目を設定した理由は，相関や重み付けの変更に従って情報関数がどのように変わるかがモードの位置から容易にわかるからである。困難度がある値の付近に集中するように

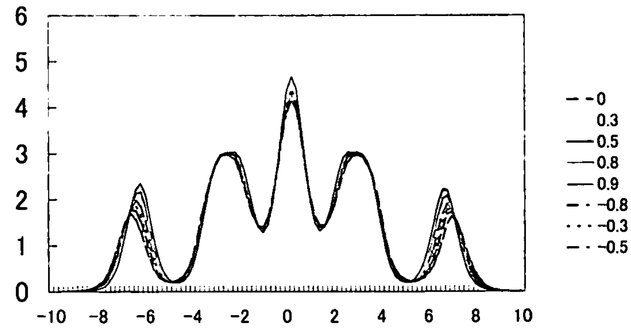

FIGURE 1　重みが４：１の時のテスト情報関数

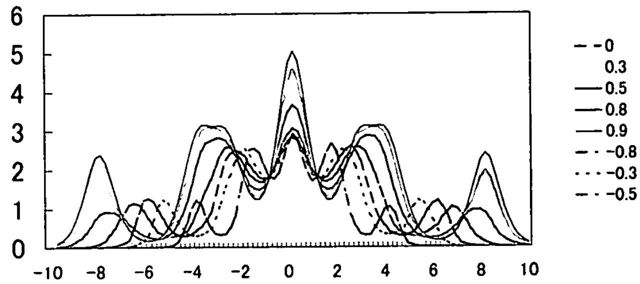

FIGURE 2　重みが２：１の時のテスト情報関数

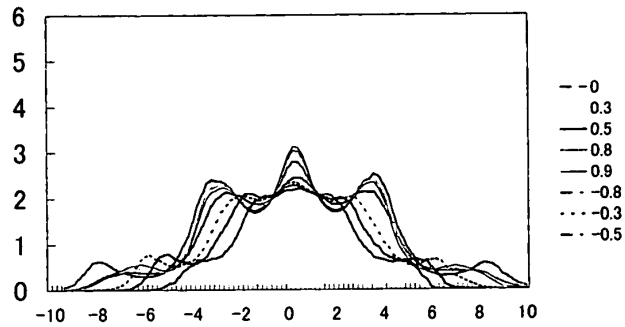

FIGURE 3　重みが１：１の時のテスト情報関数

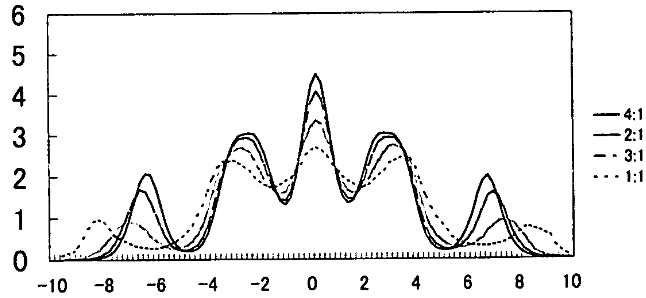

FIGURE 4　重みを変化させた時の情報関数（相関 0.3）

項目を選択すると，テスト情報関数は単峰になる。FIGURE 1から3までにより，この項目群を用いたテストでは，相関が小さくなると，テスト情報関数の形は縮小する。この結果は，今回使用した項目のパラメータが特性の原点を中心として対称に配置されていることに起因するものであり，安易な一般化は危険である。しかし，相関が小さくなるほど θ^* の分布の分散が小さくなる（$Var(\theta^*)=\mathbf{a}'\Sigma\mathbf{a}$ より）ことから，了解可能な結果であると思われる。FIGURE 1から3より，重み付けは1：1に近づく方がより相関の変化の影響を受けることも同時にわかるが，これも θ^* に2つの特性が均等に影響を与える方が2つの特性の相関の影響が大きくなるはずであるので，当然の結果であると思われる。また，前述したように，ここで各情報関数の大小によってテストの信頼性の大小を論じることはできない。そこで，相関が0.8のときの線形結合 θ^*_N の標準化されたテスト情報関数をFIGURE 5に示した。ここでは，各情報関数の大小から，特定の θ^*_N の値の下でどの重み付けが最大の信頼性を与えるかがわかる。FIGURE 5から，重み付けが一方に偏るほど，情報関数の多峰性が増すことがわかる。また，TABLE 1には，相関と重み付けを変えたときの，θ^* の分布における上位30%点での標準化されたテスト情報関数の値を掲載した。TABLE 1から，相関が高い時は重み付けが1：1の時の方が，相関が低い時は一方に重みが偏っている場合の方がテスト情報関数は大きくなることがわかる。また，同じ重み付けの時には，相関の高いほうがテスト情報関数は大きな値をとることがわかる。

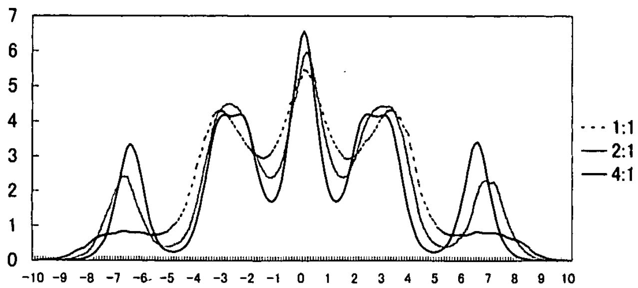

FIGURE 5　重みを変化させた時の標準化されたテスト情報関数（相関 0.8）

TABLE 1　相関と重み付けを変えた時の上位30%でのテスト情報関数の値(標準化後)

相関	4：1	2：1	1：1
-0.8	1.6864	0.8470	0.4118
-0.5	2.0033	1.3860	1.0463
-0.3	2.2205	1.7652	1.4757
0	2.5991	2.3596	2.1506
0.3	3.0187	2.9754	2.9002
0.5	3.3277	3.4185	3.4391
0.8	3.8657	4.2502	4.3526

数値例 2

すでに適用例1で使われた16個の項目に加え，新たに項目パラメータ既知の12項目から項目を追加することを考える。ここでTABLE 2は特性の相関が0.3であるとき，それらの項目のパラメータの値と，重み付け $\boldsymbol{a}$ を変化させた場合の，各項目を適用例1の項目群に加え作成された17項目からなるテストの θ^* での上位30%点におけるテスト情報関数の大きさを示した。また，それぞれの重み付けでのテスト情報関数の大きさの順位を記した。TABLE 2から，$\boldsymbol{a}$（重み付け）が変わると，θ^* のある特定の値での標準誤差も変化するため，テスト編集時に採択するべき項目は変わることがわかる。例えば項目1から4までは特性1に対する識別力が高い。よって，特性2より特性1を測定する度合が高いと言えるが，測定目的である θ^* が特性1に対する重みを増すほど同じ重みの時での情報量の順位が高くなることがわかる。このことから，本研究で提案された方法が，θ^* に関しての各項目の情報の大きさという，既存の多次元項目反応理論の諸手法では得られない情報を与えてくれることがわかる。また，数値例1では，相関が0.3の時，一方の特性に対する重みの比率が大きくなるほど，テスト情報関数の値は大きくなっていた（TABLE 1より）が，TABLE 2からはそのような単調な関係が得られていない。これは，数値例1では1つの項目は特性を1つしか測定していなかったからであると考えられる。

TABLE 2　追加された項目の入ったテストについての上位30%点でのテスト情報関数の値（相関 0.3）

項目番号	特性1の識別性	特性2の識別性	困難度	4：1		2：1		1：1	
1	3	1.5	-1	6.866	2	7.132	3	4.815	7
2	3	1.5	-2	7.937	1	10.840	1	5.377	5
3	3	1.5	-3	5.143	3	6.711	4	4.148	8
4	3	1.5	-4	3.369	9	3.484	10	3.186	11
5	2.5	2	-1	5.063	4	6.425	5	6.351	2
6	2.5	2	-2	4.739	5	7.600	2	8.758	1
7	2.5	2	-3	3.770	6	4.783	6	5.310	6
8	2.5	2	-4	3.155	11	3.077	11	3.058	12
9	1.5	2	-1	3.548	8	3.759	8	3.440	10
10	1.5	2	-2	3.563	7	4.264	7	5.682	4
11	1.5	2	-3	3.289	10	3.735	9	5.691	3
12	1.5	2	-4	3.026	12	3.064	12	3.507	9

適用例

この節では，実際の項目プールからいくつか項目を選び，特性値の線形結合のテスト情報関数を求める。

ここでは，アメリカ国立教育統計センター（National Center for Education Statistics）による，1996年度の全米での小学4年生に実施された数学テスト（被験者64364人）に項目反応理論を適用して得られた項目プールを利用する。ここでは数学は5つの下位特性に分類されており，それぞれの特性を測定する項目が複数作成されて，実施されている。本研究では，Geometry & Spatial Sense（以後 Geometry），Data Analysis, Statistics & Probability（以後 Data Analysis）の2つの下位特性を測定する項目を20項目ずつ選び出した。各項目の識別性，困難度パラメータの推定値を TABLE 3 に示す。ここで，項目プール作成時に推定された Geometry に関する特性と Data Analysis に関する特性の相関の推定値は0.79である。FIGURE 6 に，項目1から15までを利用した場合（=set①）と，項目6から20を利用した場合（=set②）の *Geometry* : *Data Analysis* = 2 : 1

TABLE 3 Geometry と Data Analysis の各項目のパラメータの推定値

項目	識別性	困難度	項目	識別性	困難度
Geometry 1	1.298	2.119	Data Analysis 1	1.048	0.720
Geometry 2	0.511	2.573	Data Analysis 2	2.468	1.435
Geometry 3	1.050	1.889	Data Analysis 3	1.158	0.602
Geometry 4	1.142	2.035	Data Analysis 4	2.143	1.437
Geometry 5	0.848	1.669	Data Analysis 5	1.009	0.811
Geometry 6	0.369	-0.675	Data Analysis 6	0.694	-0.154
Geometry 7	0.712	1.545	Data Analysis 7	1.077	0.258
Geometry 8	0.617	-0.766	Data Analysis 8	0.621	0.361
Geometry 9	0.520	1.171	Data Analysis 9	0.973	-0.111
Geometry10	2.322	0.258	Data Analysis10	0.648	-0.941
Geometry11	1.983	0.449	Data Analysis11	1.281	0.585
Geometry12	0.632	-0.425	Data Analysis12	1.258	-0.292
Geometry13	1.469	-0.215	Data Analysis13	0.943	0.216
Geometry14	1.836	-0.545	Data Analysis14	0.633	-0.165
Geometry15	0.839	1.091	Data Analysis15	0.899	0.089
Geometry16	0.635	-1.779	Data Analysis16	0.432	-2.676
Geometry17	0.510	-1.602	Data Analysis17	0.453	-0.432
Geometry18	1.273	-1.586	Data Analysis18	1.064	-0.929
Geometry19	1.323	-1.276	Data Analysis19	1.092	-1.805
Geometry20	0.624	-1.467	Data Analysis20	0.674	-1.471

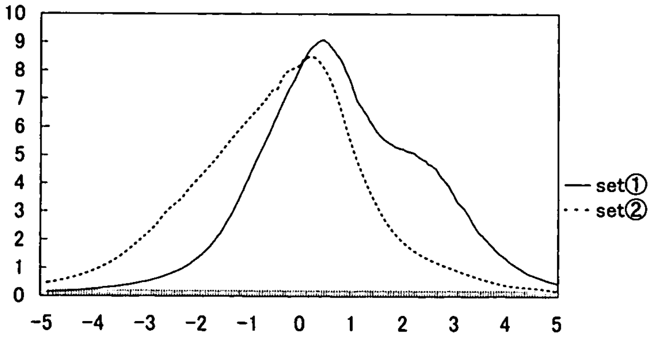

FIGURE 6 アメリカ国立教育統計センターのテストのテスト情報関数（相関 0.79）

の時の特性値の線形結合 θ^*_N の情報関数を示す。set ①での情報関数に対して，set ②の情報関数の方が低い能力値の被験者の測定に適していることがわかる。本研究の利点は，特性値の線形結合を測定する際に，このような形でテストの目的に応じてテストフォームの選択や項目選択が可能である点にある。

結　語

本研究では，複数の特性が存在し，かつそれらに相関が存在するような状況下で，関心がそれらの特性の線形結合にあるとき，評価・測定を項目反応理論を用いて行うための基礎となる「テスト情報関数」の計算方法を開発した。教育評価・心理測定の実際場面では，ある特性の測定を行う際に複数の下位特性の測定を行いながら，最終的には1次元の評価が行われることが多いので，本方法で提案された手法は実用的であると考えられる。本方法は，テストの編集時に利用できるよう，テストデータを必要としないでテスト情報関数を算出する方法であるが，特性値の分布が事前情報として与えられるときだけでなく，テストの被験者集団から得られたデータを用いて分布（またはそのパラメータ）を推定し，それを利用することも可能である。また，本手法は項目情報関数にも適用できるため，適応型テストにも利用可能である。本方法によって様々な重み付けや相関関係のもとでも情報関数の導出が可能であることは数値例を通じて確認された。また，本研究では重み付けや相関がどのように変化しても，線形結合 θ^* の分散が一定になるというような制約を置かなかったが，代りに異なる重み付けや相関を有する θ^* のテスト情報関数間の比較を行う方法も提案され，数値例では解釈可能な結果が導かれた。

引用文献

Allen, N.L., Jenkins, F., Kulick, E., & Zelenak, C.A. 1997 Technical report of the NAEP 1996 State Assessment Program in Mathematics. Washington, DC : National Center for Education Statistics.

池田　央　1994　現代テスト理論　朝倉書店

Lord, F.M. 1980 Applications of item response theory to practical testing problems. Hillsdale: Lawrence Erlbaum Associates.

Lord, F.M., & Novick, M.R. 1968 Statistical theories of mental test scores. Reading, Mass. : Addison-Wesley.

Luecht, R.O. 1996 Multidimensional computerized adaptive testing in a certification or licensure context. *Applied psychological measurement*, **20**, 389—404.

Meng, X-L., & Stephen, S. 1996 Fitting full-information item factor models and an empirical investigation of Bridge sampling. *Journal of the American Statistical Association*, **91**, 1254—1267.

Mislevy, R.J. 1984 Estimating latent distributions. *Psychometrika*, **49**, 359—381.

Nicewander, W.A. 1993 Some relationship between the information function of irt and the signal/noise ratio and reliability coefficient of classical test theory. *Psychometrika*, **58**, 139—141.

Reckase, M.D. 1996 A linear logistic multidimensional model. In W.J. van der Linden, & R.K. Hambleton (Eds.), *Handbook of modern item response theory*. New York : Springer-Verlag. Pp.271—286.

Reckase, M.D., & McKinley, R.L. 1991 The discriminating power of items that measure more than one dimension. *Applied psychological measurement*, **15**, 361—373.

Segall, D.O. 1996 Multidimensional adaptive testing. *Psychometrika*, **61**, 331—354.

芝 祐順 編 1991 項目反応理論 東京大学出版会

繁桝算男 編 1998 新版 心理測定法 放送大学教育振興会

Sijtsma, K., & Molenaar, I.W. 1987 Reliability of test scores in nonparametric item response theory. *Psychometrika*, **52**, 79—97.

Stuart, A., & Ord, J.K. 1987 Kendall's advanced theory of statistics. London : C. Griffin.

van der Linden, W.J. 1998 Bayesian item selection criteria for adaptive testing. *Psychometrika*, **63**, 201—216.

van der Linden, W.J. 1999 Multidimensional adaptive testing with a minimum error-variance criterion. *Journal of educational and behavioral statistics*, **63**, 201—216.

謝　辞

繁桝算男先生（東京大学），南風原朝和先生（東京大学）には貴重なお時間を割いて頂き，本研究の理論的な側面や論文構成などに関して大変有用なコメントを頂きました。この場をお借りして，御礼申し上げます。

付　記

本研究による方法はSAS/IMLでプログラムを行いました。利用をご希望の方は著者までご連絡ください。

(2001.3.9 受稿, 7.31 受理)

付録 1　特性が 2 次元のときの θ^* に関するテスト情報関数

重みベクトル $\boldsymbol{a}=(a_1,\sqrt{1-a_1^2})^t$ とおくと

$$
\mathrm{I}(\theta)^{-1}\approx\sum_k^q\left((1-a_1^2)\sum_j^J\frac{(\partial P_j(\theta)/\partial\theta_1)^2}{P_j(\theta)Q_j(\theta)}+2a_1\sqrt{1-a_1^2}\sum_j^J\frac{\partial P_j(\theta)/\partial\theta_1\times\partial P_j(\theta)/\partial\theta_2}{P_j(\theta)Q_j(\theta)}+a_1^2\sum_j^J\frac{(\partial P_j(\theta)/\partial\theta_2)^2}{P_j(\theta)Q_j(\theta)}\right)
$$
$$
/\left(\left\{\sum_j^J\frac{(\partial P_j(\theta)/\partial\theta_1)^2}{P_j(\theta)Q_j(\theta)}\right\}\times\left\{\sum_j^J\frac{(\partial P_j(\theta)/\partial\theta_2)^2}{P_j(\theta)Q_j(\theta)}\right\}-\left\{\sum_j^J\frac{\partial P_j(\theta)/\partial\theta_1\times\partial P_j(\theta)/\partial\theta_2}{P_j(\theta)Q_j(\theta)}\right\}^2\right)W(R_k)
$$
$$
=\sum_k^q\frac{\dfrac{\sum_j^J\left(\sqrt{1-a_1}\partial P_j(\theta)/\partial\theta_1+a_1\partial P_j(\theta)/\partial\theta_2\right)^2}{P_j(\theta)Q_j(\theta)}}{\left\{\sum_j^J\dfrac{(\partial P_j(\theta)/\partial\theta_1)^2}{P_j(\theta)Q_j(\theta)}\right\}\times\left\{\sum_j^J\dfrac{(\partial P_j(\theta)/\partial\theta_2)^2}{P_j(\theta)Q_j(\theta)}\right\}-\left\{\sum_j^J\dfrac{\partial P_j(\theta)/\partial\theta_1\times\partial P_j(\theta)/\partial\theta_2}{P_j(\theta)Q_j(\theta)}\right\}^2}W(R_k) \tag{21}
$$

さらに，2パラメータのロジスティック項目反応モデルを仮定すると，

$$
I(\theta)^{-1}\approx\sum_k^q\left(\sum_j^J\frac{\left(\sqrt{1-a_1^2}Dw_{j1}P_j(\theta)+a_1Dw_{j2}P_j(\theta)\right)^2}{P_j(\theta)Q_j(\theta)}\right)/(\{\sum_j^JD^2w_{j1}^2P_j(\theta)Q_j(\theta)\}\times\{\sum_j^JD^2w_{j2}^2P_j(\theta)Q_j(\theta)\}-
$$
$$
\{\sum_j^JD^2w_{j1}w_{j2}P_j(\theta)Q_j(\theta)\}^2)\times W(R_k) \tag{22}
$$

以上の式において θ は式(19)で表される。

付録 2　項目情報関数の逆行列が求められないことについて

1　項目が 1 つの特性のみ測定しているとする場合

ここで関心下の項目を j とし，それが特性 k のみを測定しているとする。すると本文の式(5)より，

$$\partial P_j(\theta_k)/\partial\theta_s=0 \quad (for \ \ k \neq s)$$

また，式(6)より，$I(\theta)$ の非対角成分と k 以外の対角成分が 0 となるので，$I(\theta)$ は常に非正則になる。

2　2 パラメータロジスティックモデルに従う場合

特性が 2 つの場合を例に挙げる。項目 j の情報関数 $I(\theta)_j$ の第 $(s,\ t)$ 要素は

$$I_{st}(\theta)_j=D^2 w_{js}w_{jt}P_j(\theta)Q_j(\theta) \tag{23}$$

となる。従って，行列式は $D^4({w^2}_{j1}{w^2}_{j2}-(w_{j1}w_{j2})^2)P_j(\theta)^2Q_j(\theta)^2=0$ となるので項目情報関数の行列は常に非正則になる。

A Test Information Function for Linear Combinations of Traits Without Nuisance Traits in Multidimensional Item Response Theory

TAKAHIRO HOSHINO (GRADUATE SCHOOL OF ARTS AND SCIENCES, UNIVERSITY OF TOKYO) JAPANESE JOURNAL OF EDUCATIONAL PSYCHOLOGY, 2001, 49, 491—499

In educational evaluation and psychological measurement, we often measure traits that contain several sub-traits. When tests, such as entrance examinations, are used for decision-making, often test users do not evaluate the sub-traits directly, but rather evaluate the linear combination of sub-traits in which weights were pre-specified by the test users. In the present research, we propose a new test information function and item information that enable us to evaluate linear combinations of sub-traits in which the traits are correlated. Nuisance dimensions of linear combinations are integrated out in the proposed information function. This will enable test editing in which linear combinations of sub-traits can be evaluated prior to executing the test. We also propose a method for comparing values of information functions that have different weights and/or different correlations. A numerical simulation study showed the validity of our methods. We also applied our methods to the item pool of Geometry and Statistics of the 1996 NAEP test.

Key Words : item pool, adaptive testing, test editing, multidimensional item response theory