

UDC 371. 279. 6

独立に推定されたテスト項目パラメータの共通尺度上での等化法 ——共通第1次学力試験の選択科目の差異の分析への適用——

前川 眞一*

鈴木 規夫**

目 次

- | | |
|--|----------------------------|
| 1. はじめに
2. 項目反応理論による等化
3. 方法
4. 方法の比較
5. 共通尺度上でのテスト特性の比較 | 6. 考察
参考文献
付. 最適化の方法 |
|--|----------------------------|

1. は じ め に

同じ能力を測っていると想定されるテストが複数個存在し、異なる受験者集団により受験されている場合、それぞれに含まれる項目が同一でなければ、そのテスト得点を互いにそのままの形で直接比較することには無理があり、何らかの形で各テストの得点を分析する必要がある。例えば、昭和60年度以降の共通第1次学力試験の数学のテストは、大問Ⅰ～Ⅲは共通で、大問Ⅳ～Ⅵの中から受験生が任意の2問を選択回答するというものであり、この場合、その選択のパターンにより3つの数学のテストが存在することになる。この問題に関して、同じく選択制である理科と社会について清水（1983）にも述べられているように、各テストの難易度と、そのテストを受験した集団の構成員の能力とは互いに関連しており、あるテストが易しかったのか、

それともそのテストを受験した集団の構成員の能力が高かったのかを明らかにするためには、何らかの分析が必要となる。例えば項目反応理論（Item Response Theory, IRT, または潜在特性理論とも呼ばれる。詳細は例えば、Hambleton & Swaminathan (1985) 等を参照。）に基づき、テスト項目の難易度の分析を行うとすれば、項目パラメータの値が与えられていることが大前提となるが、各テストが、異なった受験者集団により受験されている場合には、通常項目パラメータの等化という作業が必要となる。本論文では、そのための幾つかの方法を提起する。

2. 項目反応理論による等化

n 個の項目がある同一の能力を測っていると想定される場合、IRTは能力値 θ を持つ個人がその中のある項目 i に正答する確率が項目パラメータ、 $\mu_i = (a_i, b_i, c_i)'$ を用いて、

* 大学入試センター研究開発部 進学適性研究部門

** 大学入試センター研究開発部 情報処理研究部門

大学入試センター研究紀要 No. 17, 1988, p. 249～271（昭和63年1月20日受付）

©1988 THE NATIONAL CENTER FOR UNIVERSITY ENTRANCE EXAMINATION

$$P(\theta | \mu_i) = c_i + \frac{1 - c_i}{1 + \text{EXP}\{-1.7a_i(\theta - b_i)\}} \quad (2.1)$$

と書き表されることを仮定する。ここに a_i , b_i , c_i はそれぞれ項目 i の識別力, 困難度及びチャンスレベルと呼ばれる項目パラメータであり, θ 尺度の原点と単位は任意である。これを 3 パラメータロジスティック項目特性曲線 (ICC) という。今, n 個の項目のパラメータの値が与えられている場合, その中から任意の n_j 個の項目を集めて作られたテスト j , $j = 1, 2, \dots, t$, をそれぞれ異なる N_j 人の個人が受けたとしても, (2.1) に基づいて推定される各個人の能力の値は, その精度の違いを除けば比較可能である。しかし, 項目パラメータが未知の場合, その推定を各テストごとに各個人の能力の推定と同時に行うとすると, (2.1) に含まれる θ の原点と単位の不定性により, 各テストごとに独自の尺度が構成され, その比較のためには各尺度の原点と単位を調整して同一の共通尺度に変換する必要が生じる。即ち, 各テストごとに定義される能力の尺度を θ_j , また, 項目 i がテスト j に含まれる場合, その尺度上での項目パラメータの値を (a_{ij}, b_{ij}, c_{ij}) , $i \in j, j = 1, 2, \dots, t$, とすると, 任意の 2 つの尺度 k と j との間には,

$$\theta_k = A_{j(k)}\theta_j + B_{j(k)} \quad (2.2)$$

$$a_{ik} = a_{ij}/A_{j(k)} \quad (2.3)$$

$$b_{ik} = A_{j(k)}b_{ij} + B_{j(k)} \quad (2.4)$$

$$c_{ik} = c_{ij} \quad (2.5)$$

ただし, $A_{j(k)}, B_{j(k)}$ はテスト j を k に等化するための変換の係数

という線形の関係が存在し, この線形変換の係数及び共通尺度上での項目パラメータの値を求めることが, IRT における等化の問題である。この問題はこれまでにいくつかの方法で解かれてきているが, [例えば Marco (1977), 芝 (1978), Haebara (1980, 1981),

Linn, Levine, Hastings, & Wardrop (1981), Stocking & Lord (1983) 等, また Hambleton & Swaminathan (1985) を参照] それらは, 主に共通の項目を共有する 2 つのテストの等化を取り扱い, (2.3) 又は (2.4) を基にしてテスト対間の変換係数を推定することを目的とするものである。そして, 共通尺度としては, どちらか一方のテストの尺度, 例えばテスト k があるテストの旧版, テスト j が新版である時には, 旧版の尺度が用いられ, 等化により新たに共通尺度を作るということは積極的には行われていない。したがって, 推定値の間で (2.3) ~ (2.5) は正確には成立しないため, 項目 i がテスト j と k の両方に含まれる場合には, テスト k を用いて推定されたものと, テスト j からの等化によって得られたものの 2 通りの項目パラメータの推定値,

$$(\hat{a}_{ik}, \hat{b}_{ik}, \hat{c}_{ik}) \text{ 及び}$$

$$(\hat{a}_{ij}/\hat{A}_{j(k)}, \hat{A}_{j(k)}\hat{b}_{ij} + \hat{B}_{j(k)}, \hat{c}_{ij})$$

が存在することになり, どちらか一方の尺度に強い意味のある場合 (例えば, 旧版の推定値を変更したくない場合) を除けば, 変換係数を推定した後, 改めて共通項目のパラメータの値の調整をする必要が生じる。

また, t 個のテストが存在し, それらがテスト j とテスト $j-1$ とが共通項目を持つように並べられているとするならば, 上記の何れかの方法によりテスト対ごとの変換係数 $A_{j(j-1)}, B_{j(j-1)}$, $j = 2, 3, \dots, t$, が与えられた時に, 全てのテスト独自の尺度は逐次的に,

$$\theta_1 = A_{j(1)}\theta_j + B_{j(1)} \quad (2.6)$$

ただし, $j = 3, 4, \dots, t$, に関して,

$$A_{j(1)} = \prod_{l=2}^j A_{j(l-1)}$$

$$B_{j(1)} = \prod_{l=2}^j [B_{j(l-1)} \prod_{m=2}^{l-1} A_{m(m-1)}]$$

という形で θ_1 尺度へ変換することが可能である。しかしこの方法に関しては, テストの数

が多くなった時に誤差が累積していく可能性、及び最終結果がテストの並べ方に依存することの2つの問題点が指摘されている。

(Haebara, 1981)

以上の共通項目を含むテストを、異なる受験者に課した場合の等化の方法であるが、同一受験者が複数のテストを受験する場合には、各個人の能力の推定値 $\theta_{ij,l}=1,2,\dots,N_j$, $j=1,2,\dots,t$, の間の、

$$\theta_{lk}=A_{j(k)}\theta_{lj}+B_{j(k)}+\text{誤差} \quad (2.7)$$

という関係を用いてテストを等化することが可能である。この点に関しては、野口(1983, 1986)を参照されたい。

本論文では、共通項目を共有するテストが複数存在し各項目パラメータが異なる受験者集団のデータを用いて推定されている場合には、全ての項目パラメータを同時に共通の尺度に位置づけるための方法を考察する。

3. 方 法

各テストごとにそのテストの受験者集団の能力の平均が0、分散が1となるように定義された θ_j 尺度上での項目パラメータの推定値を、それがデータであるという意味で、

$$\begin{aligned} \mathbf{x}_{ij} &= (a_{ij}, b_{ij}, c_{ij})', \\ i &\in j, j=1,2,\dots,t \end{aligned} \quad (3.1)$$

また、 n 個の項目パラメータの共通尺度 θ 上での値を、

$$\boldsymbol{\mu}_i = (\alpha_i, \beta_i, \gamma_i)', i=1,2,\dots,n \quad (3.2)$$

とする。各テストには、高々 $n_j \leq n$ 個の項目しか含まれていないため、 $\mathbf{x}_{ij}$ は項目 i がテスト j に含まれる場合に限って観測されている。また、 $\mathbf{x}_{ij}$ はそれぞれ N_j 人の異なる受験者集団をもちいて推定されているものとする。そして、共通尺度から各尺度への変換、

$$\theta_j = u_j + v_j \theta, j=1,2,\dots,t,$$

を考え、それに対応する項目パラメータの変

換を

$$\boldsymbol{\mu}_{ij} = (\alpha_{ij}, \beta_{ij}, \gamma_{ij})', \quad (3.4)$$

ただし、

$$\alpha_{ij} = \alpha_i / v_j \quad (3.5)$$

$$\beta_{ij} = u_j + v_j \beta_i \quad (3.6)$$

$$\gamma_{ij} = \gamma_i \quad (3.7)$$

とする。そして、 $\mathbf{x}_{ij}$ を $\boldsymbol{\mu}_{ij}$ の誤差を含んだ観測値であるとみなし、 $\mathbf{x}_{ij}$ のセットが与えられた時、それを最も良く説明する、(3.5)～(3.7)という構造を持った $\boldsymbol{\mu}_{ij}$ を推定することを考える。このように、IRTにおける等化の問題を、共通尺度から各尺度への変換の係数と共通尺度上での項目パラメータの値との同時推定として定義することにより、テスト対ごとの変換係数を推定していく方法が持つテスト対の選び方への解の依存性、累積誤差の可能性、及び共通尺度上での項目パラメータの値の不定性の問題を無くすることができる。なお共通尺度の性格としては、例えば全受験者集団の能力の平均がゼロ、分散が1となるようなものとしておく。

今、モデルとデータとのズレの指標として、

$$Q_k = \sum_{j=1}^t w_j \sum_{i \in j} D_k(\mathbf{x}_{ij}, \boldsymbol{\mu}_{ij}), \quad (3.8)$$

ただし、 w_j はテスト j に固有な重み、また $\sum_{i \in j}$ はテスト j に含まれる全ての項目に関する和を表す、

という形を考える。そして、 $\mathbf{x}_{ij}$ と $\boldsymbol{\mu}_{ij}$ のズレを表す測度である $D_k(\mathbf{x}_{ij}, \boldsymbol{\mu}_{ij})$ としては、そのメトリックの取り方により様々なものが考えられるが、ここでは以下の4種類を提案する。まず、一般的なものとして、項目パラメータの差の2次形式で定義される、

$$\begin{aligned} D_1(\mathbf{x}_{ij}, \boldsymbol{\mu}_{ij}) &= \\ (\mathbf{x}_{ij} - \boldsymbol{\mu}_{ij})' W_{ij}^{-1} (\mathbf{x}_{ij} - \boldsymbol{\mu}_{ij}), \end{aligned} \quad (3.9)$$

ただし、 W_{ij} は任意の正定値重み行列、及び、ICCの差の2乗和で定義される、

$$D_2(\mathbf{x}_{ij}, \mu_{ij}) = \int_{-\infty}^{+\infty} w_{ij}(\theta) [P(\theta | \mathbf{x}_{ij}) - P(\theta | \mu_{ij})]^2 f_j(\theta) d\theta \quad (3.10)$$

ただし、 $f_j(\theta)$ はテスト j の受験者集団の θ_j 尺度上での能力分布の密度関数を、また、 $w_{ij}(\theta)$ は θ の関数として定義される重みを表す、

の2つの基準を考える。

D_1 はMarco (1977)、芝 (1978)、及びLinn et al. (1981)の流れを、また D_2 はcharacteristic curve methodと呼ばれるHaebara (1980, 1981)、及びStocking & Lord (1983)の流れを汲むものであり、両者の違いは、回帰分析における回帰係数の比較に例えるならば、 D_1 は回帰係数の差異を直接に比較したものであり、 D_2 は、まず予測値を計算し、その差異を比較するものであると言えよう。

本論文の主題は、この2つの方法の比較にあるが、参考として、次の2つの測度をも取り上げることにする。

すなわち、ICCのログオッズの差の2乗和で定義される、

$$D_3(\mathbf{x}_{ij}, \mu_{ij}) = \int_{-\infty}^{+\infty} \{L[P(\theta | \mathbf{x}_{ij}), c_{ij}] - L[P(\theta | \mu_{ij}), \gamma_i]\}^2 f_j(\theta) d\theta \quad (3.11)$$

ただし、

$$L(x, y) = \ln\{(x-y)/(1-x)\}, \quad 0 < y < x < 1$$

したがって、

$$D_3(\mathbf{x}_{ij}, \mu_{ij}) = \int_{-\infty}^{+\infty} [a_{ij}(\theta - b_{ij}) - \alpha_{ij}(\theta - \beta_{ij})]^2 f_j(\theta) d\theta \\ = (a_{ij} - \alpha_{ij})^2 + (a_{ij}b_{ij} - \alpha_{ij}\beta_{ij})^2 \quad (3.12)$$

および、パラメータの差の2乗和で定義される、

$$D_4(\mathbf{x}_{ij}, \mu_{ij}) = (1/a_{ij} - 1/\alpha_{ij})^2 + (b_{ij} - \beta_{ij})^2 + (c_{ij} - \gamma_i)^2 \quad (3.13)$$

の2つである。なお、 D_3 の(3.12)への展開は、

各尺度ごとに θ の平均が0、分散が1となっていることを用いた。また、 D_4 は α_{ij} の逆数を取ることににより、 v_j に関して線形化することをねらったものである。

上記の4つの基準はそれぞれ最小2乗基準であり、2乗の項の左側にはデータが、右側にはモデルの値のみが含まれていることに注意すると、各 Q_k の最小値を与える μ_i 及び (u_j, v_j) の推定値は、 Q_k のそれぞれに対する偏微分をゼロと置いた時に得られる連立方程式の解として与えられることになり、これは、例えばガウス・ニュートン法等を用いることにより数値的に解くことが可能である。推定すべき変数の数は、3パラメータロジスティックモデルの場合は $2t + 3n - 2$ 、2パラメータロジスティックモデルの場合は $2t + 2n - 2$ である。ここに現れる -2 は、共通尺度の原点と単位の不定性に関する自由度を示している。

Q_2 に関しては、その積分を解析的に行うことが難しいため、それを、

$$D_2'(\mathbf{x}_{ij}, \mu_{ij}) = \sum_{l=1}^m w_{ij}(\theta_l) [P(\theta_l | \mathbf{x}_{ij}) - P(\theta_l | \mu_{ij})]^2 f_j(\theta_l) \quad (3.14)$$

で置き換えることにし、テスト j の受験者の能力の分布が標準正規分布に近いと考えられる場合には、 θ を例えば -4 から $+4$ まで m 個の等間隔に区切ったものを θ_l とし、 $f_j(\theta_l)$ としては、その各点における標準正規分布の密度関数をとることにする。

実際の Q_k の最小化は、全ての μ_i が与えられた場合、変換係数 (u_j, v_j) の推定が各テストごとに独立に行えること、また、 Q_k が項目 i を含むテスト j に関して、

$$Q_k = \sum_{i=1}^n \sum_{j \ni i} w_j D_k(\mathbf{x}_{ij}, \mu_{ij}) \quad (3.15)$$

ただし、 $\sum_{j=i}^t$ は項目*i*を含む全てのテストに関する和を示す、

と書けることから、全ての変換係数が与えられた場合、各項目パラメータの推定が独立に行えることを利用して、条件付きの最小化を繰り返す交互最小2乗法(ALS)〔例えば高根(1980)を参照〕を用いることにする。また Q_2 と Q_3 に関しては、(3.3)の逆変換、

$$\begin{aligned}\theta &= q_j + r_j \theta_j, \\ \text{ただし、} q_j &= -u_j/v_j, r_j = 1/v_j\end{aligned}\quad (3.16)$$

を定義し、

$$P(\theta | \mu_{ij}) = P(q_j + r_j \theta | \mu_i) \quad (3.17)$$

に注意すると、これらに関する偏微分を取った方が計算が簡単となる。同様に Q_4 に関しては、 $1/\alpha_i$ を推定すると線形となり都合がよい。

また、共通尺度 θ 上での全受験者の能力の平均と分散はそれぞれ、

$$\bar{\theta} = \sum_{j=1}^t N_j q_j / \sum_{j=1}^t N_j \quad (3.18)$$

$$S^2 = \sum_{j=1}^t N_j (q_j^2 + r_j^2) / \sum_{j=1}^t N_j - \bar{\theta}^2$$

となるため、繰り返し計算のステップ中でこの値を、0と1とするような基準化を行うことにより、共通尺度を意味付けることができる。

推定の実アルゴリズムをまとめると、以下のステップの繰り返しとなる。

- (1) (u_j, v_j) 又は (q_j, r_j) , $j=1, 2, \dots, t$ の初期値、及び重みの値を計算する。
- (2) 各項目ごとに、 (u_j, v_j) 又は (q_j, r_j) が全て与えられているものとして Q_k を最小とする $(\alpha_i, \beta_i, \gamma_i)$ を求める。
- (3) 各テストごとに、 $(\alpha_i, \beta_i, \gamma_i)$ が全て与えられているものとして Q_k を最小とする (u_j, v_j) 又は (q_j, r_j) を求める。

- (4) 推定された値を、(3.18)の $\bar{\theta}$ と S^2 を用いて、

$$\left. \begin{aligned}\alpha_i &\leftarrow \alpha_i \times S \\ \beta_i &\leftarrow (\beta_i - \bar{\theta})/S\end{aligned} \right\} i=1, 2, \dots, n \quad (3.19)$$

$$\left. \begin{aligned}u_j &\leftarrow u_j + v_j \bar{\theta} \\ v_j &\leftarrow v_j \times S \\ q_j &\leftarrow (q_j - \bar{\theta})/S \\ r_j &\leftarrow r_j/S\end{aligned} \right\} j=1, 2, \dots, t \quad (3.20)$$

と基準化する。

- (5) Q_k の減り具合を見て、収束を判定する。
- (6) 収束してなければ(2)へ戻る 各繰り返しごとにステップ2と3において、 Q_k の値は必ず小さくなるか前回の値と同じ値をとり、 Q_k は極小値を持つ関数であるため、このALSのアルゴリズムは、必ず Q_k の極小値に収束することが言える。なお全ての方法に関して、 W 、又は w で表されている重みの値は既知であることに注意されたい。 Q_k の各変数に関する偏微分、及びアルゴリズムの詳細に関しては、付録を参照されたい。

4. 方法の比較

本節では、第3節に述べた4つの方法により、いくつかのデータを解析し、その結果の比較を行う。用いたデータセットは昭和60年、61年及び62年度の共通第1次学力試験の数学I及びIIに含まれる項目及び昭和59年度の理科と社会の各教科に含まれる項目の5つである。含まれる項目総数、 n 、はそれぞれ37、35、40、140、及び282である。

数学の場合は、各年度とも大問IからIIIに含まれる項目が共通項目となり、大問IV～VIからの選択のパターンにより全部で3つの数学のテスト、即ち(I—III, IV, V), (I—III, IV, VI)及び(I—III, V, VI)が存在することになる。各年度ごとに項目の包含関係を、図-1～3に示した。

テスト 項目	1	2	3
I - III 1 18			
IV 19 24			
V 25 33			
VI 34 37			

図-1 昭和60年度数学における項目の分配

テスト 項目	1	2	3
I - III 1 21			
IV 22 26			
V 27 29			
VI 30 35			

図-2 昭和61年度数学における項目の分配

テスト 項目	1	2	3
I - III 1 17			
IV 18 25			
V 26 36			
VI 37 40			

図-3 昭和62年度数学における項目の分配
 (各行は大問を、その中の数字は項目(枝問)の通し番号を示す。また斜線部分は項目がテストに含まれていることを示す。)

テスト 項目	1	2	3	4	5	6
物理 1 23						
化学 24 56						
生物 57 102						
地学 103 140						

図-4 昭和59年度理科における項目の分配

テスト 項目	1	2	3	4	5	6	7	8	9	10	11	12	13
倫・社 1 39													
政・経 40 87													
世界史 88 132													
日本史 133 184													
地理A 185 232													
地理B 233 282													

図-5 昭和59年度社会における項目の分配

(各行は科目を、その中の数字は項目(枝間)の通し番号を示す。また斜線部分は項目がテストに含まれていることを示す。)

理科の場合は、物理、化学、生物、及び地学の4科目からの任意の2科目の選択制であり、全部で6種類の理科のテストが存在する。また、社会の場合は倫理・社会、政治・経済、世界史、日本史、地理A、及び地理Bの6科目からの2科目選択制であるが、倫理・社会と政治・経済、及び地理Aと地理Bの組み合わせは禁止されているため、全部で13のテストが存在する。理科及び社会の項目の包含関係を、図-4と図-5に示す。

まず、各テストの受験者集団から無作為に5,000人を選び、その受験者集団における能力の平均が0、分散が1となるような尺度上での項目パラメータの値を推定した。なお、4つの方法を比較するために、推定においては2パラメータロジスティックモデルを用い、計算はGETAB [Baillie (1980)] というプログラムを用いて行った。その推定値のうち、昭和60年度の数学に関するものを表1に示す。

以上のようにして作成された5つのデータセットを、それぞれ第3節に述べる4つの方法を用いて解析したが、その際重みとしては、 w_j は全て1、 D_1 の W_{ij} は 2×2 の単位行列、 D_2' の $w_{ij}(\theta_i)$ は、それを全て1とした。また D_2' の $f_j(\theta)$ としては、一様分布と標準正規分布の密度関数の2通りを用いた。(D_2' の θ_i 区間数は $m > 20$ になると解は殆ど同等であることが判明したため、以下では $m = 50$ の例を挙げる。)したがって、実際には、各データセットごとに、5つの等化結果が得られることになる。以下の記述においてはそれらを、 Q_1 、 Q_2 、 u 、 Q_{2N} 、 Q_3 、及び Q_4 と略す。初期値としては全て $u_j = 0$ 、 $v_j = 1$ を用い、それぞれの Q の値を、観測された全ての x_{ij} の数で割ったものの平方根が各繰返しで、0.00001以上が小さくならないことをもって収束の基準とした。5つのデータセットのうち、表1に対応する昭和60年度の数学の等化結果を、 α 、 β 、 u 、 v 、

表1 昭和60年度数学データセット(テストごとに推定された項目パラメータの値)

1	1	1	0.8739	-1.2348
2	2	1	0.7766	-1.5025
3	3	1	0.7730	-0.7170
4	4	1	0.7683	-1.0395
5	5	1	0.7154	-0.1911
6	6	1	1.1075	-1.5988
7	7	1	1.2119	-1.6347
8	8	1	0.9870	-2.3094
9	9	1	1.1738	-0.6040
10	10	1	0.9137	-0.2432
11	11	1	0.9124	0.3163
12	12	1	0.7542	-0.5844
13	13	1	0.6999	-0.6428
14	14	1	0.8051	-0.4149
15	15	1	1.1303	-0.5344
16	16	1	1.0655	-0.5840
17	17	1	0.3512	1.4849
18	18	1	0.5990	0.5163
19	19	1	1.4023	-0.5455
20	20	1	1.4336	-0.5777
21	21	1	1.3753	-0.1706
22	22	1	1.3955	-0.1699
23	23	1	1.1851	0.3385
24	24	1	1.1635	0.5933
25	25	1	2.2536	-0.2502
26	26	1	2.1027	-0.4490
27	27	1	2.4199	-0.3619
28	28	1	2.0707	-0.2630
29	29	1	2.2085	-0.2641
30	30	1	2.0803	-0.0910
31	31	1	1.9939	0.3116
32	32	1	2.1080	0.4956
33	33	1	1.3691	1.7661
34	1	2	1.2188	-1.1014
35	2	2	1.1104	-1.2776
36	3	2	1.1024	-0.6321
37	4	2	1.0380	-0.9747
38	5	2	0.9155	-0.2914
39	6	2	1.5440	-1.1748
40	7	2	1.5530	-1.2840
41	8	2	1.3297	-1.7687
42	9	2	1.4992	-0.4216
43	10	2	1.1767	-0.1639
44	11	2	1.1448	0.2865
45	12	2	0.9466	-0.5332
46	13	2	0.9439	-0.5724
47	14	2	1.0617	-0.4136
48	15	2	1.5802	-0.3463
49	16	2	1.5319	-0.3949
50	17	2	0.4557	1.2237
51	18	2	0.7536	0.4262
52	19	2	4.0285	-0.3749
53	20	2	4.6253	-0.3985
54	21	2	5.7300	-0.1331
55	22	2	5.9834	-0.1311
56	23	2	4.9536	0.1042
57	24	2	4.3432	0.2420
58	34	2	0.1358	-1.7052
59	35	2	0.2043	-7.5645
60	36	2	0.1832	0.1727
61	37	2	0.1296	9.0594
62	1	3	1.3887	-0.9478
63	2	3	1.3076	-1.1068
64	3	3	1.0665	-0.5516
65	4	3	1.0490	-0.8937
66	5	3	0.9362	-0.1537
67	6	3	3.3920	-0.7752
68	7	3	3.9332	-0.8077
69	8	3	2.3040	-1.1628
70	9	3	2.0881	-0.1226
71	10	3	1.5875	0.1625
72	11	3	1.5131	0.5825
73	12	3	0.8180	-0.3494
74	13	3	0.8220	-0.3918
75	14	3	0.9109	-0.2084
76	15	3	1.2051	-0.0321
77	16	3	1.1890	-0.0760
78	17	3	0.3293	2.1347
79	18	3	0.5776	1.0807
80	25	3	0.5437	-1.1254
81	26	3	0.5149	-1.1611
82	27	3	0.5613	-1.5712
83	28	3	0.4991	-1.2661
84	29	3	0.5763	-1.1570
85	30	3	0.5008	-0.7408
86	31	3	0.7613	0.4258
87	32	3	0.7978	0.7893
88	33	3	0.7088	3.0947
89	34	3	0.0545	-3.5857
90	35	3	0.1882	-7.7217
91	36	3	0.0801	1.3384
92	37	3	0.2554	5.5412

表2 昭和60年度数学の等化結果

a. α_i パラメータ

項目\方法	Q1	Q2U	Q2N	Q3	Q4
1	1.1776	1.1603	1.1658	1.1630	1.2165
2	1.0849	1.0536	1.0670	1.0703	1.1085
3	0.9887	0.9887	0.9885	0.9831	1.0300
4	0.9567	0.9645	0.9595	0.9491	1.0069
5	0.8561	0.8685	0.8618	0.8484	0.9110
6	2.1182	1.5424	1.7183	2.0366	1.8333
7	2.3515	1.5907	1.7822	2.2477	1.9775
8	1.5946	1.1802	1.3600	1.5476	1.5213
9	1.6154	1.5505	1.5594	1.5812	1.6546
10	1.2465	1.2146	1.2128	1.2218	1.2817
11	1.2057	1.1867	1.1691	1.1825	1.2527
12	0.8308	0.8560	0.8531	0.8283	0.8886
13	0.8197	0.8342	0.8334	0.8178	0.8659
14	0.9205	0.9409	0.9388	0.9187	0.9762
15	1.2978	1.3014	1.3017	1.3020	1.3545
16	1.2596	1.2575	1.2599	1.2633	1.3094
17	0.3727	0.3847	0.3866	0.3742	0.3916
18	0.6328	0.6511	0.6562	0.6338	0.6718
19	3.2418	2.0468	2.1395	3.0554	1.8684
20	3.6547	2.1622	2.2733	3.4539	1.9579
21	4.3746	2.2722	2.3603	4.1558	1.9686
22	4.5526	2.3211	2.4101	4.3271	2.0069
23	3.7804	1.9540	2.1014	3.5915	1.6972
24	3.3620	1.8303	2.0179	3.1871	1.6339
25	1.1957	0.8954	1.0821	1.3378	0.9542
26	1.1207	0.8568	1.0384	1.2538	0.9015
27	1.2690	0.8532	1.0901	1.4201	0.9913
28	1.0984	0.8255	1.0135	1.2289	0.8761
29	1.2007	0.9207	1.1001	1.3427	0.9983
30	1.1031	0.8814	1.0370	1.2341	0.8793
31	1.2442	1.2517	1.3297	1.3879	1.2210
32	1.3107	1.3246	1.4225	1.4622	1.2822
33	0.9780	1.0278	1.1951	1.0886	1.0493
34	0.0833	0.0972	0.0964	0.0858	0.0789
35	0.1726	0.1910	0.1908	0.1706	0.2143
36	0.1154	0.1341	0.1349	0.1183	0.1138
37	0.1699	0.1644	0.1782	0.1613	0.2034

b. β_i パラメータ

項目\方法	Q1	Q2U	Q2N	Q3	Q4
1	-0.9626	-1.0728	-1.0677	-1.0713	-1.0275
2	-1.1412	-1.2617	-1.2554	-1.2559	-1.2123
3	-0.5690	-0.6314	-0.6249	-0.6237	-0.6046
4	-0.8461	-0.9575	-0.9557	-0.9651	-0.9165
5	-0.2005	-0.2260	-0.2172	-0.2151	-0.2106
6	-1.0838	-1.1136	-1.0792	-1.0171	-1.0907
7	-1.1292	-1.1695	-1.1346	-1.0606	-1.1410
8	-1.5787	-1.6536	-1.6076	-1.5745	-1.5996
9	-0.3844	-0.3744	-0.3602	-0.3367	-0.3561
10	-0.1236	-0.0866	-0.0704	-0.0429	-0.0752
11	0.2872	0.3667	0.3878	0.4223	0.3646
12	-0.4510	-0.4891	-0.4842	-0.4955	-0.4636
13	-0.4918	-0.5339	-0.5277	-0.5348	-0.5070
14	-0.3273	-0.3501	-0.3444	-0.3494	-0.3293
15	-0.3207	-0.3053	-0.3005	-0.3046	-0.2821
16	-0.3605	-0.3503	-0.3444	-0.3476	-0.3258
17	1.2918	1.5491	1.5348	1.5876	1.5371
18	0.5053	0.6440	0.6360	0.6599	0.6479
19	-0.4121	-0.3681	-0.3922	-0.3893	-0.4496
20	-0.4342	-0.3947	-0.4181	-0.4037	-0.4798
21	-0.1645	-0.0544	-0.0829	-0.1370	-0.1148
22	-0.1636	-0.0526	-0.0810	-0.1350	-0.1134
23	0.1467	0.3024	0.2656	0.1107	0.2961
24	0.3087	0.4969	0.4527	0.2565	0.5121
25	-0.4561	-0.5632	-0.5088	-0.4376	-0.6676
26	-0.4749	-0.5854	-0.5283	-0.4622	-0.6918
27	-0.6415	-0.8029	-0.7173	-0.6339	-0.9308
28	-0.5103	-0.6293	-0.5633	-0.4937	-0.7473
29	-0.4812	-0.6004	-0.5489	-0.4805	-0.6956
30	-0.2836	-0.3439	-0.3072	-0.2420	-0.4307
31	0.2300	0.2801	0.2740	0.3256	0.2803
32	0.4186	0.5151	0.4978	0.5576	0.5243
33	1.6652	2.0686	1.8972	2.1030	2.1109
34	-2.9633	-2.2116	-2.2163	-2.4562	-2.6256
35	-8.6526	-7.8338	-7.8565	-8.7581	-6.9951
36	0.8841	0.5059	0.5129	0.6187	0.7406
37	8.3486	7.5601	7.1925	8.0568	6.2367

表2 昭和60年度数学の等化結果 (続)

項目\方法	c. u_j				
	Q1	Q2U	Q2N	Q3	Q4
1	0.1017	-0.0556	-0.0098	0.1082	-0.0469
2	-0.0134	-0.1187	-0.1021	-0.0239	-0.0443
3	-0.0435	0.2149	0.1415	-0.0329	0.1258

項目\方法	d. v_j				
	Q1	Q2U	Q2N	Q3	Q4
1	1.5639	1.0479	1.0862	1.6898	1.0217
2	0.8888	0.8855	0.8695	0.7965	0.8412
3	0.8709	1.1485	1.1182	0.9684	1.2761

項目\方法	e. q_j				
	Q1	Q2U	Q2N	Q3	Q4
1	-0.0651	0.0531	0.0090	-0.064	0.0459
2	0.0151	0.1340	0.1175	0.030	0.0527
3	0.0499	-0.1871	-0.1265	0.034	-0.0986

項目\方法	f. r_j				
	Q1	Q2U	Q2N	Q3	Q4
1	0.6394	0.9543	0.9206	0.5918	0.9787
2	1.1251	1.1294	1.1501	1.2559	1.1888
3	1.1482	0.8707	0.8943	1.0327	0.7837

q , 及び r のそれぞれにつき, 表2の a ~ f に示す。

今回の分析に用いたデータセットは, 人工的に作られたものではなく, 実際の共通第1次学力試験のデータであり, 各パラメータの‘真の値’は不明である。したがって, シミュレーション研究とは異なり, 例えば表2に示される各推定値をその真の値と比較して, 各方法の優劣を論じることができない。したがって, ここでは5つの方法間の類似性というものを手掛りとして各方法を比較してみることにした。即ち,

$$\alpha_{i(L)}, \beta_{i(L)}, i=1, 2, \dots, n,$$

$$u_{j(L)}, v_{j(L)}, q_{j(L)}, r_{j(L)},$$

$$j=1, 2, \dots, t \quad L=1, 2, \dots, 5$$

を, 方法 L を用いて推定されたパラメータの値とする時, 方法 L と方法 m との非類似度を, 例えば $\alpha_{i(L)}$ を用いた場合には,

$$DA_{Lm} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\alpha_{i(L)} - \alpha_{i(m)})^2} \quad (4.1)$$

で定義する。同様に,

$$DB_{Lm} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\beta_{i(L)} - \beta_{i(m)})^2} \quad (4.2)$$

$$DU_{Lm} = \sqrt{\frac{1}{t} \sum_{j=1}^t (u_{j(L)} - u_{j(m)})^2} \quad (4.3)$$

$$DV_{Lm} = \sqrt{\frac{1}{t} \sum_{j=1}^t (v_{j(L)} - v_{j(m)})^2} \quad (4.4)$$

$$DQ_{Lm} = \sqrt{\frac{1}{t} \sum_{j=1}^t (q_{j(L)} - q_{j(m)})^2} \quad (4.5)$$

$$DR_{Lm} = \sqrt{\frac{1}{t} \sum_{j=1}^t (r_{j(L)} - r_{j(m)})^2} \quad (4.6)$$

を定義し, それを基にして, 各方法の他の4つの方法との平均的非類似度を,

$$DA_L = \frac{1}{4} \sum_{m=1}^4 DA_{Lm} \quad (4.7)$$

$$DB_L = \sum_{m=1}^4 DB_{Lm} / 4 \quad (4.8)$$

$$DU_L = \sum_{m=1}^4 DU_{Lm} / 4 \quad (4.9)$$

$$DV_L = \sum_{m=1}^4 DV_{Lm} / 4 \quad (4.10)$$

$$DQ_L = \sum_{m=1}^4 DQ_{Lm} / 4 \quad (4.11)$$

$$DR_L = \sum_{m=1}^4 DR_{Lm} / 4 \quad (4.12)$$

と定義する。この値を各データセットごとに

表3に示した。

これを見て分かるように、どのパラメータで比較してみても、方法 Q_{2U} と Q_{2N} は他の方法との類似度が高く、 Q_4 は他の方法と比べて異なる結果を出すことが多い。また Q_1 と Q_3 は、データセットにより、その類似度が異なっている。特に注目に値するのは昭和59年度社会のデータセットにおける Q_4 の解の異質性で、これは殆ど全ての項目の α_i の値が他の方法の推定値よりも掛け離れて大きな値をとったことに由来している。また、(4.1)～(4.6)

表3 平均非類似度

a. 昭和60年度数学

パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
DA	0.587724*	0.418249○	0.390882○	0.533907	0.472312
DB	0.289601	0.200962*	0.203451	0.250290	0.353185*
DU	0.115105	0.120756	0.089009○	0.111940	0.091954
DV	0.286786	0.208936*	0.200449○	0.320416*	0.249121
DQ	0.103679	0.109464	0.082303	0.095742	0.081247○
DR	0.213958*	0.142988	0.134037○	0.207075	0.174061

b. 昭和61年度数学

パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
DA	0.988148	0.797594	0.920465	0.776078○	1.51814*
DB	0.462347	0.374744○	0.471878	0.399495	0.50085*
DU	0.159133	0.111701○	0.122688	0.140807	0.16277*
DV	0.364603	0.290328○	0.324508	0.321596	0.60730*
DQ	0.073456	0.047151	0.046482○	0.076941*	0.06568
DR	0.174921	0.142001○	0.166311	0.150042	0.22728*

c. 昭和62年度数学

パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
DA	0.125786	0.087510○	0.093851	0.114290	0.126836*
DB	0.140075*	0.080303○	0.082358	0.124264	0.110489
DU	0.148217	0.091959	0.090677○	0.169609*	0.134153
DV	0.200538	0.146951	0.141850○	0.274993*	0.231345
DQ	0.158493*	0.088808	0.087741○	0.156421	0.130327
DR	0.191839	0.136000	0.131823○	0.247315*	0.231754

d. 昭和59年度理科

パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
DA	0.276511	0.256994○	0.260233	0.334220	0.519939*
DB	0.318280	0.280467○	0.295402	0.343450	0.428216*
DU	0.281115	0.204787	0.204207○	0.247556	0.382347*
DV	0.374364	0.276687	0.270345○	0.322448	0.580722*
DQ	0.168644	0.123026○	0.125445	0.143591	0.183543*
DR	0.243576	0.171356	0.170042○	0.207281	0.338518*

e. 昭和59年度社会

パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
DA	0.741238	0.743544	0.731509	0.717073○	2.51429*
DB	0.534004	0.411343	0.378229○	0.379271	0.98741*
DU	0.237749	0.220333	0.217220○	0.219118	0.62577*
DV	0.775789	0.761963	0.754347	0.753483○	2.66231*
DQ	0.149849	0.132148	0.131638○	0.138881	0.37843*
DR	0.272097	0.259138	0.254434○	0.265576	0.80288*

○は各行の最大類似度を示す。

*は各行の最小類似度を示す。

表4 平均相関係数

a. 昭和60年度数学					
パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
RA	0.93656	0.960132	0.964497 ○	0.940279	0.916641 *
RB	0.99467	0.997147	0.997496 ○	0.996816	0.993345 *
RU	-0.13160 *	0.263380	0.343968 ○	-0.038730	0.168478
RV	0.33599 *	0.590199	0.591859 ○	0.482523	0.453437
RQ	-0.16055 *	0.274865	0.358742 ○	0.014797	0.172447
RR	0.36292 *	0.657695	0.741343 ○	0.621597	0.558338
b. 昭和61年度数学					
パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
RA	0.986554	0.990593 ○	0.982703	0.988727	0.979535 *
RB	0.922766	0.942959	0.905074	0.945820 ○	0.897318 *
RU	0.985587 *	0.995700	0.994847	0.996693	0.996903 ○
RV	0.968256 ○	0.956878	0.968063	0.911169 *	0.964162
RQ	0.992636	0.996864	0.997335	0.995861	0.990948 *
RR	0.966026	0.952669	0.957240	0.950396 *	0.979822 ○
c. 昭和62年度数学					
パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
RA	0.98230 *	0.990907 ○	0.988402	0.99057	0.982588
RB	0.99811 *	0.999167 ○	0.999167 ○	0.99851	0.998272
RU	0.74505	0.769697	0.760652 ○	0.280835 *	0.558449
RV	-0.68672 *	0.200042 ○	0.197354	-0.11976	-0.029302
RQ	0.69516	0.715511	0.730711 ○	0.23224 *	0.494139
RR	-0.70675 *	0.212502 ○	0.208820	-0.03639	-0.070141
d. 昭和59年度理科					
パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
RA	0.967963	0.969503 ○	0.967297	0.955854	0.911057 *
RB	0.978401	0.984543 ○	0.978110	0.976797	0.971313 *
RU	0.698750	0.720857 ○	0.711667	0.653545	0.369450 *
RV	0.517437	0.570636 ○	0.554511	0.348622	0.079547 *
RQ	0.821827	0.835767 ○	0.826834	0.798183	0.717292 *
RR	0.561750	0.563291 ○	0.548003	0.370185	0.044001 *
e. 昭和59年度社会					
パラメータ\方法	Q1	Q2U	Q2N	Q3	Q4
RA	0.976557 *	0.981137	0.984100 ○	0.983931	0.977389
RB	0.961638	0.967761	0.973286 ○	0.971301	0.938159 *
RU	0.810560	0.791140	0.800843	0.818362 ○	0.436333 *
RV	0.736793 ○	0.664440	0.723091	0.704722	0.316395 *
RQ	0.832838	0.844339 ○	0.841276	0.837698	0.454017 *
RR	0.774899	0.790366	0.816247 ○	0.736189	0.491180 *

○は各行の最大相関係数を示す。

*は各行の最小相関係数を示す。

に対応する推定値間の相関係数の平均をRA, RB, RU, RV, RQ, 及びRRとして表4に示すが, ここにも上述の結果とほぼ同じ傾向が現れている。ただし, RU, RV, RQ, 及びRRに関しては, 相関を計算するのに用いたデータの数 n_j が, 各データセットごとに3, 3, 3, 6, 及び14と極めて少ないことに注意されたい。

以上の結果から, どの方法が最も優れているかの判断を下すことは早計ではあるが, 限られた解析例ではあるが, それらが他の方法と異なる解を出す可能性が多いという意味で

Q₃及びQ₄の使用はあまり好ましいことではないと言えるであろう。Q₁とQ₂に関しては, その重みの取り方により, 上記の結果からでは推測しきれない特徴が現れることがあると考えられるが, Q₂の解は安定した, 平均的なものであると言えるであろう。

5. 共通尺度上でのテスト特性の比較

第3節に示すいずれかの方法を用いて, 全受験者集団の能力の平均と分散を(0, 1)とした共通尺度上での項目パラメータの値,

及びそこへの各尺度の変換係数が推定されたならば、その値を用いることにより各テストの特性を共通尺度上で表すことが可能となる。例えば、テスト j の得点が、ここに含まれる n_j 個の項目の平均正答数として与えられる場合、テスト特性曲線 (TCC),

$$T_j(\theta) = \sum_{i \in j} P(\theta | \mu_i) / n_j \quad (5.1)$$

が、その得点の θ への回帰曲線を表すが、これは、共通尺度上の能力 θ の者がテスト j を受けた時に得る得点の期待値である。したがって、 t 個のTCCがある θ の値で比べることにより、テストの難易度を比較することが可能となる。ただ各々のTCCは、 θ の非線形関数であるため、比較のために用いる θ の値により、その大小関係は変わることになり、各テストの難易度を一意に定めることはできない。(能力の低いものにとってはテスト j がテスト j' より易しく、高いものにとってはその逆であるということも起こりうる。) 平均的なテストの難易度としては、例えば全受験者集団の平均能力値 $\theta = 0$ で、TCCを評価したもの、 $T_j(0) = 1, 2, \dots, t$, を用いることも考えられる。

また、各テストを受けた受験者集団の能力の平均値は、(3.16)により共通尺度上では、

$$\bar{\theta}_j = q_j \quad (5.2)$$

また、分散は、

$$S_j^2 = r_j^2 \quad (5.3)$$

となるが、この値を比較することにより各受験者集団の平均的能力を比べることができる。

以下に上記の方法の共通第1次学力試験の理科と社会の各科目の難易度の比較への適用例を示す。ただ、ここで興味があるのは、2科目の組み合わせで定義されるテストの難易度ではなく、各科目の難易度であるため、(5.1)に示すTCCを直接に使うことはできない。そこで、次の科目特性曲線とでも呼ぶ

べきものを定義し、

$$S_m(\theta) = \sum_{i \in m} P(\theta | \mu_i) / m_m \quad (5.4)$$

ただし、 m は各科目を示し、 m_m は科目 m に含まれる項目の数を、また $i \in m$ は科目 m に含まれる項目に関する和を示す、この値を上述の議論のように、 $\theta = 0$ で評価したもの、 $S_m(0)$ 、をもって科目 m の難易度とする。 $S_m(0)$ は平均的能力の受験者が科目 m に含まれる m_m 個の項目のうちの何パーセントに正答するか期待値であるため、その値が大きい程、その科目は易しいということになる。このように定義される科目難易度を、昭和59年の理科と社会の各科目ごとに算出したものを表5-a、及び表6-aの第1行に示した。これは、いずれも Q_{2N} を最小とする解の値を用いて計算したものである。ここでは、この値、及び(5.2)で定義される各テストごとの受験者集団の学力の平均値を、清水(1983)の方法による結果と比較してみる。

清水(1983)の方法は、基本的には、

$$y_{mjk}, m=1, 2, \dots, M, \text{ただし } M \text{ は科目総数}$$

$$j=1, 2, \dots, t,$$

$$k=1, 2, \dots, N_j,$$

を個人 k が、 j というテストを受けた時に取る科目 m の得点とする時、その期待値に、

$$E(y_{mjk}) = \text{mean} + A_m + B_j \quad (5.5)$$

という線形モデルを仮定するものであり、この A_m で科目 m の難易度を、 B_j でテスト j の受験者集団の学力を定義する。

ただ、 y_{mjk} は全ての m と j との組み合わせについて観測されていないため(全ての k に関して $M \times t$ のデータのうち観測されているのは $2t$ 個のみである。)、これを不完備ブロックデザインとみなして、各要因の効果を推定しなければならず、この値を直接(5.4)及び(5.2)で定義されるIRTによる難易度、及び学

表5 昭和59年度理科の各科目の比較

a. 各科目の難易度

方法 \ 科目	物	化	生	地
IRT	.724	.752	.671	.760
順位*	2	3	1	4
線形モデル	2.82	-.66	.94	-6.20
順位*	1	3	2	4

*最もむずかしい科目の順位が1

b. 選択科目による学力のちがい

方法 \ 選択科目	物 化	物 生	物 地	化 生	化 地	生 地
IRT	.188	-.322	-.185	-.046	-.215	-.279
順位**	1	6	3	2	4	5
線形モデル	4.88	-6.01	-3.78	-2.03	-4.27	-6.44
順位**	1	5	3	2	4	6

**最も学力が高い群の順位が1

表6 昭和59年度社会の各科目の比較

a. 各科目の難易度

方法 \ 科目	倫・社	政・経	日本史	世界史	地理A	地理B
IRT	.731	.605	.627	.605	.662	.724
順位*	6	1.5	3	1.5	4	5
線形モデル	-4.78	7.10	-1.35	6.37	1.48	-2.73
順位*	6	1	4	2	3	5

*最もむずかしい科目の順位が1

b. 選択科目による学力のちがい

方法 \ 選択科目	倫 日	倫 世	倫 A	倫 B	政 B	政 世	政 A	政 日	日 世	日 A	日 B	世 B	世 A
IRT	-.123	.025	-.157	-.075	-.128	.120	-.072	-.223	.258	-.165	-.038	.401	.176
順位**	9	5	11	8	10	4	7	13	2	12	6	1	3
線形モデル	-.89	1.88	-1.49	-.56	-.92	2.15	-.23	-2.71	4.87	-1.05	-1.05	6.99	4.31
順位**	8	5	12	7	9	4	6	13	2	10.5	10.5	1	3

**最も学力が高い群の順位が1

力の平均値と比較することはむずかしい。したがって、ここでは、以下のように直観的な議論を進める。

今、 $E(y_{mjk})$ を t_{mj} と書き、それが全ての m と j の組み合わせに関して与えられているとする。この場合、通常の制約条件を持つ効果は、

$$A_m = \sum_{j=1}^t t_{mj} / t - \sum_{m=1}^M \sum_{j=1}^t t_{mj} / Mt \quad (5.6)$$

及び、

$$B_j = \sum_{m=1}^M t_{mj} / M - \sum_{m=1}^M \sum_{j=1}^t t_{mj} / Mt \quad (5.7)$$

である。ここで t_{mj} はテスト j を受けた者が科目 m でとる得点の期待値であるから、これはIRTの立場では、各項目の配点を無視すれば、

$$t_{mj} = S_m(\bar{\theta}_j) \times m_m \quad (5.8)$$

として表現されるところとする。そうすると(5.7)及び(5.4)から、定数部分を除けば、

$$\begin{aligned} B_j &= \sum_{m=1}^M S_m(\bar{\theta}_j) \times m_m / M \\ &= \sum_{m=1}^M \sum_{i \in m} P(\bar{\theta}_j | \mu_i) / M \\ &= \sum_{i=1}^n P(\bar{\theta}_j | \mu_i) / M \end{aligned}$$

となり、各ICCが θ に関して単調増加関数であることから、 B_j は $\bar{\theta}_j$ と単調増加関数にあることが言える。

また、 A_m に関しては、 $\min_j(\bar{\theta}_j)$ と $\max_j(\bar{\theta}_j)$ の間での科目特性曲線の線形性を仮定すると、定数部分を除けば、

$$\begin{aligned} A_m &= \sum_{j=1}^t S_m(\bar{\theta}_j) \times m_m / t \\ &= S_m(\sum_{j=1}^t \bar{\theta}_j / t) \times m_m \\ &= S_m(0) \times m_m \end{aligned} \quad (5.10)$$

となる。したがって、各テストの満点と同じになるように基準化してあれば、 A_m と $S_m(0)$ の間には、線形の関係が存在することにな

る。

この方法を、昭和59年度の理科と社会のデータに適応したものを表5-a, b及び表6-a, bの第3行に示す。この値は、清水(内部資料)を拝借した。ただし、 A_m の値(表5-aと表6-aの第3行)は慣例により、その符号が逆になっていることに注意されたい。

上記の期待値に関する直観的議論から、これらの値は各表の第1行に示すIRTによる値

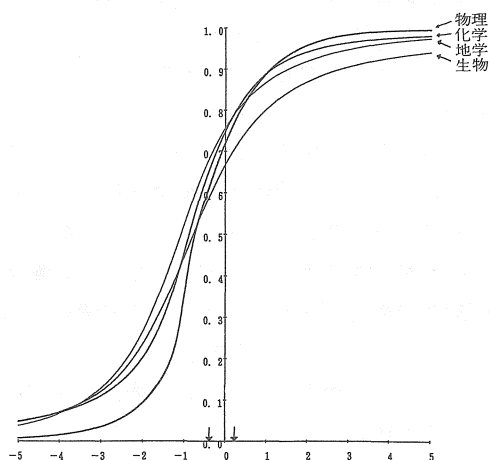


図-6 昭和59年度理科の科目特性曲線
(図中横軸の矢印は $\bar{\theta}_j$ の範囲を示す。)

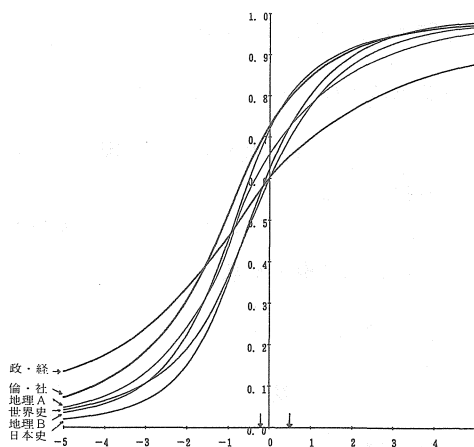


図-7 昭和59年度社会の科目特性曲線
(図中横軸の矢印は $\bar{\theta}_j$ の範囲を示す。)

と単調関係にあることが予想されるが、表の第2行と4行に示した順位はかなりの一致度を示していると言えよう。これは、図6と図7に示すように科目特性曲線が、局所的にかなり線形であることに由来しているものと思われる。

6. 考 察

第2節に示した4つの最小2乗基準は、3パラメータモデルが2パラメータモデルかにより、また、その重みの取り方によりいろいろな形をとる。

重みの選択に関しては、一般的には自由であるが、 w_j としては例えば N_j に比例するものを取りることが考えられる。 D_1 の W_{ij} に関しては x_{ij} が最尤推定値である場合には、その漸近分散共分散行列が項目パラメータの情報行列 $I(x_{ij})$ の逆行列として推定できるため、それを用いることが考えられ、この場合、 Q_1 を最小にする解は、

$$x_{ij} \sim N(\mu_{ij}, W_{ij}) \quad (6.1)$$

を仮定した時の最尤解となる。また W_{ij} を対角行列とするとも考えられるが、特に W_{ij} の(2, 2)要素を1とし、それ以外を全てゼロと置けば、 Q_1 は、 D_4 の第2項、

$$\sum_{j=1}^t w_j \sum_{i \in j} [b_{ij} - (u_j + v_j \beta_i)]^2 \quad (6.2)$$

となり、これは(欠測値を含む)因子分析モデルと合致する。

また、2パラメータモデルの場合、第4節に示したように重みを取ると、

$$Q_1 = \sum_{j=1}^t \sum_{i \in j} [(a_{ij} - \alpha_{ij})^2 + (b_{ij} - \beta_{ij})^2] \quad (6.3)$$

$$Q_3 = \sum_{j=1}^t \sum_{i \in j} [(a_{ij} - \alpha_{ij})^2 + (a_{ij}b_{ij} - \alpha_{ij}\beta_{ij})^2] \quad (6.4)$$

$$Q_4 = \sum_{j=1}^t \sum_{i \in j} [(1/a_{ij} - 1/\alpha_{ij})^2 + (b_{ij} - \beta_{ij})^2] \quad (6.5)$$

となり、その関係がより明確になる。すなわち、 Q_1 と Q_3 の差異はその第2項の b パラメータを含む部分であり、 Q_1 と Q_4 とはその第1項における a パラメータを含む部分において異なっている。芝(1982, 私信)は、データとモデルのズレを項目パラメータ間の差の2乗和で定義する場合、 a パラメータのズレと b パラメータのズレとを(6.3)のように同等に取り扱うことに関して疑問を呈しているが、(6.4)の第2項に現れる α_{ij} を a_{ij} で置き換えたものの、

$$Q_3' = \sum_{j=1}^t \sum_{i \in j} [(a_{ij} - \alpha_{ij})^2 + a_{ij}(b_{ij} - \beta_{ij})^2] \quad (6.6)$$

を考えると、これは D_1 において、 W_{ij} の(1.1)要素を1に、その(2.2)要素を a_{ij} に、そして他の全てを0と置いたものとなり、識別力の高い項目における困難度パラメータの差異を重視するという納得のいくものになる。 x_{ij} の漸近分散共分散行列が与えられていない場合、上記のように Q_1 の重みを取ることとも考えられよう。

最後に、 D_2' の $w_{ij}(\theta_l)$ に関しては、やはり x_{ij} が最尤推定値である場合は、それを用いて $P(\theta_l | x_{ij})$ のテイラー展開による1次近似の分散、

$$\Delta_{ijl}' [I(x_{ij})]^{-1} \Delta_{ijl} \quad (6.7)$$

ただし、

$$\Delta_{ijl} = \left(\frac{\partial P(\theta_l | \mu_{ij})}{\partial \mu_{ij}} \right)_{\mu_{ij} = x_{ij}}$$

を求め、その逆数に比例するものを用いることも考えられる〔Linn, Levine, Hastings & Wardrop (1980)〕。

また、Stocking & Lord (1983) のように、

TCCを用いた,

$$Q_5 = \sum_{j=1}^t w_j \int_{-\infty}^{+\infty} \left[\sum_{i \in j} P(\theta | x_{ij}) - \sum_{i \in j} P(\theta | \mu_{ij}) \right]^2 f_j(\theta) d\theta \quad (6.8)$$

という基準も考えられるが, この場合 μ_i の値を一意に定めることはできない。これは, TCCを与えた時にその成分であるICCが一意に定められないことに由来している。

本論文で提起された方法の特徴のひとつは等化を共通尺度から各尺度への変換係数の推定として捉えることであるが, テスト対の等化係数はその値を用いて,

$$A_{j(k)} = v_k / v_j = r_j / r_k \quad (6.6)$$

$$B_{j(k)} = u_k - v_k u_j / v_j \\ = (q_j - q_k) / r_k \quad (6.7)$$

として与えられる。

最後に, 方法の優劣に関しては, Stocking & Lord (1983) や Hambleton & Swaminathan (1985) が推す, IRTの基礎となるICCのズレを最小とする, characteristic curve methodとしての Q_2 が推奨されるが, 最尤解としての Q_1 も興味深い。今後の検討の課題である。

参 考 文 献

- Baillie, R. (1980) *GETAB: A Computer Program for Estimating Item and Person Parameters of the one-and two-parameter logistic model on the PLATO system*. Computer-based Education Research Laboratory. Urbana, University of Illinois.
- Haebara, T. (1980) 'Equating Logistic Ability Scales by a Weighted Least Squares Method', *Jap. Psychol. Reas.* Vol.22, pp144-9.
- Haebara, T. (1981) 'Least Squares Method for Equating Logistic Ability Scales' *Iowa Testing Program Occasional Paper No.30*. Iowa Testing Program, Iowa City, IA.
- Hambleton, R.K. & Swaminathan, H. (1985), *Item Response Theory. Principles and Applications*. Kluwer-Nijhoff Publishing. Boston.
- Linn, R.L., Levine, M.V., Hastings, C.N. & Wardrop, J.L. (1980) 'An Investigation of Item Bias in a Test of Reading Comprehension.' *Technical Report No.163*, Center for the Study of Reading. University of Illinois. (ERIC No.# ED184091)
- Linn, R.L., Levine, M.V., Hastings, C.N. & Wardrop, J.L. (1981) Item Bias in a Test of Reading Comprehension. *Applied Psychological Measurement*, vol.5 No.2, pp159-173.
- Marco, G. (1977) Item Characteristic Curve Solutions to Three Intractable Testing Problems. *Journal of Educational Measurement*, Vol.14, pp139-160.
- 野口裕之 (1983) 被験者の推定尺度値を利用した潜在特性尺度等化法 教育心理学研究第31巻 第3号 233-238
- 野口裕之 (1983) 共通被験者の反応ボタンを利用した潜在特性尺度等化法 教育心理学研究第34巻 第4号 315-323
- 芝 祐順 (1978) 語彙理解尺度作成の試み 東大教育学部紀要第17巻
- 清水留三郎 (1983) 共通第1次学力試験の「社会」と「理科」の選択科目における差異の統計解析 大学入試センター研究紀要 No.6
- Stocking, M.L. & Lord, F.M. (1983), 'Developing a Common Metric in Item Response Theory' *Applied Psychological Measurement* Vol.7 No.2, pp201-210.
- 高根芳雄 (1980) 多次元尺度法 東大出版会

付録：最適化の方法

第3節に述べたように, 各 Q_k の $\mu_i = (\alpha_i, \beta_i, \gamma_i)$, $i=1, 2, \dots, n$, 及び (u_j, v_j) , $j=1, 2, \dots, t$, に関する最小化は, ALSの観点から, (1)全ての (u_j, v_j) が与えられた時の, Q_k の μ_i に関する最小化, 及び, (2)全ての μ_i が与えられた時の, Q_k の (u_j, v_j) に関する最小化, を交互に繰り返して行うことにする。また, (1)の場合, Q_k がある特定の項目 i に関して,

$$Q_k = \sum_{j \ni i} w_j D_k(x_{ij}, \mu_{ij}) + \text{const.} \quad (\text{A.1})$$

ただし, const. は μ_i を含まない, という形に書けることから, 各 μ_i は独立に更新することができる. 同様に, (2) の場合, ある特定のテスト j に関して,

$$Q_k = w_j \sum_{i \in j} D_k(x_{ij}, \mu_{ij}) + \text{const.} \quad (\text{A.2})$$

ただし, const. は (u_j, v_j) を含まない, と書けるため, 各 (u_j, v_j) も独立に更新することができる. したがって実際の数値計算においては, たかだか 3 変数関数の最小化を取り扱えばよいことになる.

以下に各 Q_k に関する最小化のアルゴリズムを記す.

Q_1 の最小化

Q_1 の μ_i に関する偏微分は,

$$\begin{aligned} \frac{\partial Q_1}{\partial \alpha_i} = & -2 \sum_{j \ni i} [(a_{ij} - \alpha_i / v_j) w_{ij(aa)} \\ & + e_{ij}(\beta) w_{ij(ab)} + e_{ij}(\gamma) w_{ij(ac)}] w_j / v_j \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} \frac{\partial Q_1}{\partial \beta_i} = & -2 \sum_{j \ni i} [(b_{ij} - u_j - v_j \beta_i) w_{ij(bb)} \\ & + e_{ij}(\alpha) w_{ij(ab)} + e_{ij}(\gamma) w_{ij(bc)}] w_j v_j \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} \frac{\partial Q_1}{\partial \gamma_i} = & -2 \sum_{j \ni i} [(c_{ij} - \gamma_i) w_{ij(cc)} \\ & + e_{ij}(\alpha) w_{ij(ac)} + e_{ij}(\beta) w_{ij(bc)}] w_j \end{aligned} \quad (\text{A.5})$$

ただし, $w_{ij(pq)}$ は W_{ij}^{-1} の (p, q) 要素, $p = a, b, c, q = a, b, c$, を表し, また,

$$e_{ij}(\alpha) = (a_{ij} - \alpha_i / v_j)$$

$$e_{ij}(\beta) = (b_{ij} - u_j - v_j \beta_i)$$

$$e_{ij}(\gamma) = (c_{ij} - \gamma_i)$$

である,

となる. したがって, Q_1 の μ_i に関する最小値

は, (A.3)~(A.5) を同時にゼロとする μ_i に関する線形の正規方程式の解によって与えられるが, ここでは, (A.3)~(A.5) が, 各パラメータごとに解析的に解けることを利用し, (β_i, γ_i) が与えられた時の Q_1 の α_i に関する条件付き最小化, (α_i, γ_i) が与えられた時の Q_1 の β_i に関する条件付き最小化, 及び (α_i, β_i) が与えられた時の Q_1 の γ_i に関する条件付き最小化を繰り返すことにする. これは ALS に各ステップにおける最小化を, また ALS により行うことに相当するが, 第 3 節に述べた ALS の収束性に関する議論を再度適用することにより, その収束を保証することができる. したがって, α_i, β_i , 及び γ_i の更新式,

$$\begin{aligned} \alpha_i^{(New)} = & \left[\sum_{j \ni i} (a_{ij} w_{ij(aa)} + e_{ij}(\beta) w_{ij(ab)} \right. \\ & \left. + e_{ij}(\gamma) w_{ij(ac)}) w_j / v_j \right] / \left(\sum_{j \ni i} w_{ij(aa)} w_j / v_j^2 \right) \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} \beta_i^{(New)} = & \left\{ \sum_{j \ni i} [(b_{ij} - u_j) w_{ij(bb)} + e_{ij}(\alpha) w_{ij(ab)} \right. \\ & \left. + e_{ij}(\gamma) w_{ij(bc)}] w_j v_j \right\} / \left(\sum_{j \ni i} w_{ij(bb)} w_j v_j^2 \right) \end{aligned} \quad (\text{A.7})$$

$$\begin{aligned} \gamma_i^{(New)} = & \left[\sum_{j \ni i} (c_{ij} w_{ij(cc)} + e_{ij}(\alpha) w_{ij(ac)} \right. \\ & \left. + e_{ij}(\beta) w_{ij(bc)}) w_j \right] / \left(\sum_{j \ni i} w_{ij(cc)} w_j \right) \end{aligned} \quad (\text{A.8})$$

を得る. ただし, $e_{ij}(\alpha)$ 等の計算は常に最新のパラメータの更新値を用いて計算をすることにする. すなわち, (A.6) から (A.7) へ移る時には, (A.6) で得られた $\alpha_i^{(New)}$ の値を用いて (A.7) に現れる $e_{ij}(\alpha)$ を計算することになる.

次に, Q_1 の (u_j, v_j) に関する最小化は, Q_1 の u_j に関する偏微分をゼロと置くことにより,

$$u_j = A_j - B_j v_j - C_j / v_j \quad (\text{A.9})$$

ただし,

$$A_j = \left(\sum_{i \in j} (b_{ij} w_{ij(bb)} + e_{ij}(\gamma) w_{ij(bc)} + a_{ij} w_{ij(ab)}) \right) / \left(\sum_{i \in j} w_{ij(bb)} \right)$$

$$B_j = \left(\sum_{i \in j} \beta_i w_{ij(bb)} \right) / \left(\sum_{i \in j} w_{ij(bb)} \right)$$

$$C_j = \left(\sum_{i \in j} \alpha_i w_{ij(ab)} \right) / \left(\sum_{i \in j} w_{ij(bb)} \right)$$

を得るため、これを Q_1 に代入することにより、 Q_1 を v_j のみの関数として書くことができる。したがって、

$$\frac{\partial Q_1}{\partial v_j} = D_3/v_j^3 + D_2/v_j^2 + D_0 + D_1 v_j \quad (\text{A.10})$$

ただし、

$$D_3 = - \sum_{i \in j} (\alpha_i^2 w_{ij(aa)} + C_j^2 w_{ij(bb)}) - 2 C_j \alpha_i w_{ij(ab)}$$

$$D_2 = \sum_{i \in j} \{ a_{ij} \alpha_i w_{ij(aa)} - C_j (b_{ij} - A_j) w_{ij(bb)} - [a_{ij} C_j - \alpha_i (b_{ij} - A_j)] w_{ij(ab)} + e_{ij}(\gamma) \alpha_i w_{ij(ac)} - C_j e_{ij}(\gamma) w_{ij(bc)} \}$$

$$D_0 = \sum_{i \in j} \{ (b_{ij} - A_j) (B_j - \beta_i) w_{ij(bb)} + a_{ij} (B_j - \beta_i) w_{ij(ab)} + (B_j - \beta_i) e_{ij}(\gamma) \alpha_i w_{ij(bc)} \}$$

$$D_1 = \sum_{i \in j} (B_j - \beta_i)^2 w_{ij(bb)}$$

及び、

$$\frac{\partial^2 Q_1}{\partial v_j^2} = -3 D_3/v_j^4 - 2 D_2/v_j^3 + D_1 \quad (\text{A.11})$$

を用いて1変数の最小化を行えばよいことになる。ニュートン・ラフソン法を用いれば、ステップサイズを s 、(A.10)を d 、(A.11)を h

と置けば、

$$v_j^{(New)} = v_j^{(Old)} - s(d/h), \quad 0 < s \leq 1 \quad (\text{A.12})$$

という更新式を得るが、ステップサイズ、 s 、としては、 $v_j^{(New)}$ 及びそれを用いて(A.9)を評価した、

$$u_j^{(New)} = A_j - B_j v_j^{(New)} - C_j / v_j^{(New)} \quad (\text{A.13})$$

で Q_1 を評価した時に、その値が前回の繰り返しより小さくなるようなものを選べばよい。なお、ALSによる Q_1 の最小化という観点からは、(A.10)を完全にゼロとするような v_j を求める必要はなく、単に各繰り返して Q_1 の値が減少するように v_j を更新するだけで充分である。

なお、 μ_i の初期値としては例えば、後述する Q_4 の更新式の一部、(A.32)～(A.34)を用いればよい。

Q_2 の最小化

Q_2' の (u_j, v_j) が与えられた時の μ_i に関する最小化は、正規方程式が Q_1 の場合のように簡単にならないため、

$$\frac{\partial P(q_i + r_j \theta_i | \mu_i)}{\partial \alpha_i} = 1.7(\theta_{ij} - \beta_i) Q(\theta_{ij} | \mu_i) \times [P(\theta_{ij} | \mu_i) - \gamma_i] / (1 - \gamma_i) \quad (\text{A.14})$$

$$\frac{\partial P(q_i + r_j \theta_i | \mu_i)}{\partial \beta_i} = -1.7 \alpha_i Q(\theta_{ij} | \mu_i) \times [P(\theta_{ij} | \mu_i) - \gamma_i] / (1 - \gamma_i) \quad (\text{A.15})$$

$$\frac{\partial P(q_i + r_j \theta_i | \mu_i)}{\partial \gamma_i} = Q(\theta_{ij} | \mu_i) / (1 - \gamma_i) \quad (\text{A.16})$$

$$\text{ただし、} \theta_{ij} = q_j + r_j \theta_i \quad (\text{A.17})$$

$$Q(\theta_{ij} | \mu_i) = 1 - P(\theta_{ij} | \mu_i)$$

を用いて、3変数のガウス・ニュートン法により行う。この場合、 μ_i の更新式は(A.14)～(A.16)をそれぞれ、 $S_{ij(a)}$ 、 $S_{ij(b)}$ 、 $S_{ij(c)}$ と書けば、

$$\mu_i^{(New)} = \mu_i^{(Old)} - s H_i^{-1} g_i, \quad 0 < s \leq 1$$

$$(A.18)$$

ただし,

$$g_i = (g_{i(p)}), \quad p = a, b, c, \quad (A.19)$$

$$g_{i(p)} = \sum_{j \ni i} \sum_{l=1}^m s_{ijl(p)} w_{ij}(\theta_l) \times \\ [P(\theta_{ij} | \mu_i) - P(\theta_l | x_{ij})]$$

及び,

$$H_i = (h_{i(pq)}), \quad p = a, b, c, \quad q = a, b, c,$$

$$h_{i(pq)} = \sum_{j \ni i} \sum_{l=1}^m s_{ijl(p)} s_{ijl(q)} w_{ij}(\theta_l) \quad (A.20)$$

となり, ステップサイズ, s , は Q_2' を $\mu_i^{(New)}$ で評価したものが前回の値を上まわらないように選ぶものとする. なお, μ_i の初期値としては例えば, 後述する Q_4 の (A.32) ~ (A.34) を用いればよい.

Q_2' の (q_j, r_j) に関する最小化は, 同様に,

$$\frac{\partial P(q_j + \gamma_i \theta_i | \mu_i)}{\partial q_j} = P'(\theta_{ij} | \mu_i) \quad (A.21)$$

$$\frac{\partial P(q_j + r_i \theta_i | \mu_i)}{\partial r_j} = P'(\theta_{ij} | \mu_i) \cdot \theta_{ij} \quad (A.22)$$

ただし,

$$P'(\theta_{ij} | \mu_i) = \frac{1.7 \alpha_i (1 - \gamma_i) \exp[1.7 \alpha_i (\theta_{ij} - \beta_i)]}{\{1 + \exp[1.7 \alpha_i (\theta_{ij} - \beta_i)]\}^2}$$

を用いて 2 変数のガウス・ニュートン法により行う. この場合, $q_j = (q_j, r_j)'$ と書き, (A.21) と (A.22) をそれぞれ $s_{ijl(u)}$, $s_{ijl(v)}$ と書けば, q_j の更新式は,

$$q_j^{(New)} = q_j^{(Old)} - s H_j^{-1} g_j, \quad 0 < s \leq 1 \quad (A.23)$$

ただし,

$$g_j = (g_{j(p)}), \quad p = u, v$$

$$g_{j(p)} = \sum_{i \in j} \sum_{l=1}^m s_{ijl(p)} w_{ij}(\theta_l) [P_i(\theta_{ij} | \mu_i) - P(\theta_l | x_{ij})] \quad (A.24)$$

$$H_j = (h_{j(pq)}), \quad p = u, v, \quad q = u, v$$

$$h_{j(pq)} = \sum_{i \in j} \sum_{l=1}^m s_{ijl(p)} s_{ijl(q)} w_{ij}(\theta_l) \quad (A.25)$$

となる.

Q_3 の最小化

Q_3 は (3.16) を用いることにより,

$$Q_3 = \sum_{j=1}^t w_j \sum_{i \in j} (a_{ij} - \alpha_i r_j)^2 \\ + \sum_{j=1}^t w_j \sum_{i \in j} [a_{ij} + \alpha_i (q_j - \beta_i)]^2 \quad (A.26)$$

となり, 全てのパラメータに関して線形となる. ここで, 第 1 項には α_i と r_j のみが, 第 2 項には q_j と $\alpha_i \beta_i$ が入っていることを考慮すれば, 線形の回帰分析の理論より, それぞれの条件付きの更新式は,

$$\alpha_i^{(New)} = (\sum_{j \in i} a_{ij} r_j w_j) / (\sum_{j \in i} r_j^2 w_j) \quad (A.27)$$

$$r_j^{(New)} = (\sum_{i \in j} a_{ij} \alpha_i) / (\sum_{i \in j} \alpha_i^2) \quad (A.28)$$

$$\beta_i^{(New)} = [\sum_{j \ni i} (a_{ij} b_{ij} + \alpha_i q_j) w_j] / (\sum_{j \ni i} w_j) \quad (A.29)$$

$$q_j^{(New)} = -[\sum_{i \in j} (a_{ij} b_{ij} - \alpha_i \beta_i) \alpha_i] / (\sum_{i \in j} \alpha_i^2) \quad (A.30)$$

となる. Q_4 の場合と同様に, 右辺の評価は最も新しいパラメータの更新値を用いて行うものとする.

Q_4 の最小化

Q_4 は, $1/\alpha_i$ をパラメータと見れば,

$$Q_4 = \sum_{j=1}^t w_j \sum_{i \in j} [(1/a_{ij} - v_j/\alpha_i)^2 \\ + (b_{ij} - u_j - v_j \beta_i)^2 + (c_{ij} - \gamma_i)^2] \quad (A.31)$$

となり, 全てのパラメータに関して線形となる. したがって, その更新式は $(1/a_{ij}, b_{ij})$ の, (u_j, v_j) または $(1/\alpha_i, \beta_i)$ への線形回帰という観点から

$$\alpha_i^{(New)} = (\sum_{j \ni i} v_j^2 w_j) / (\sum_{j \ni i} v_j w_j / a_{ij}) \quad (A.32)$$

$$\beta_i^{(New)} = (\sum_{j \in i} v_j w_j (b_{ij} - u_i)) / (\sum_{j \in i} v_j^2 w_j) \quad (A.33)$$

$$\gamma_i^{(New)} = (\sum_{j \in i} c_{ij} w_j) / (\sum_{j \in i} w_j) \quad (A.34)$$

$$u_i^{(New)} = \{ \sum_{j \in i} [(1/\alpha_i)^2 + \beta_i^2] - \sum_{i \in j} [(1/a_{ij})(1/\alpha_i) + b_{ij}\beta_i] \} (\sum_{i \in j} b_{ij}) / d_j \quad (A.35)$$

$$v_j^{(New)} = \{ n_j \sum_{i \in j} [(1/a_{ij})(1/\alpha_i) + b_{ij}\beta_i] - (\sum_{i \in j} b_{ij})(\sum_{i \in j} \beta_i) \} / d_j \quad (A.36)$$

ただし,

$$d_j = n_j \sum_{i \in j} [(1/\alpha_i)^2 + \beta_i^2] - (\sum_{i \in j} \beta_i)^2 \quad (A.37)$$

で得られる。

Shin-ichi Mayekawa*, Norio Suzuki** (1988) Simultaneous Equating of Several Sets of Separately Calibrated Item Parameters.

Res. Bull. Nat. Cent. Univ. Ent. Exam. No. 17, 249-271

Given several sets of separately calibrated item parameters, methods to place all the item parameters simultaneously on a common ability scale were developed.

Denoting the estimate of the item parameter by

$$\mathbf{x}_{ij} = (a_{ij}, b_{ij}, c_{ij}), \quad i=1, 2, \dots, n, \quad i \in j, \\ j=1, 2, \dots, t,$$

the first method finds the values of item parameters on the common scale

$$\underline{\mu}_i = (\alpha_i, \beta_i, \gamma_i), \quad i=1, 2, \dots, n,$$

and the transformation coefficients from the common scale to the individual test scale

$$(u_j, v_j), \quad j=1, 2, \dots, t,$$

by minimizing

$$Q_1 = \sum_{j=1}^t w_j \sum_{i \in j} (\underline{x}_{ij} - \underline{\mu}_{ij})' W_{ij}^{-1} (\underline{x}_{ij} - \underline{\mu}_{ij})$$

where $\underline{\mu}_{ij} = (\alpha_{ij}, \beta_{ij}, \gamma_{ij})$,

$$\alpha_{ij} = \alpha_i / v_j, \quad \beta_{ij} = u_j + v_j \beta_i, \quad \gamma_{ij} = \gamma_i$$

with respect to $\underline{\mu}_i$ and (u_j, v_j) .

The second method minimizes

$$Q_2 = \sum_{j=1}^t w_j \sum_{i \in j} \sum_{l=1}^m w_{ij}(\theta_l) [P(\theta_l | \underline{x}_{ij}) - P(\theta_l | \underline{\mu}_{ij})]^2$$

where $P(\theta | x)$ is the item characteristic curve of the item with the parameter x , with respect to $\underline{\mu}_i$ and (u_j, v_j) .

The weight denoted by w_j and W_{ij}^{-1} must be chosen so that they reflect the precision of the estimate, $\underline{x}_{ij}$, whereas $w_{ij}(\theta_l)$ reflects the distribution of the ability in the j -th test group.

Two additional methods, which also minimize similar least squares criterions

* Scholastic Aptitude Section, Research Division, The National Center for University Entrance Examination

** Information Processing Section, Research Division, The National Center for University Entrance Examination

with respect to the item parameters on the common scale and the transformation coefficients, were proposed and compared.

Key words

IRT Equating

Item Response Theory (IRT)

Test Equating

