

一般項目反応モデルにおける項目パラメタの周辺最尤推定法

中 村 知 靖* 前 川 眞 一**

MARGINAL MAXIMUM LIKELIHOOD ESTIMATION OF ITEM PARAMETERS
IN A GENERALIZED ITEM RESPONSE MODEL

Tomoyasu NAKAMURA AND Shin-ichi MAYEKAWA

According to the Thurstonean point of view, a usual two-parameter normal ogive model can be derived as a special case of the Law of Categorical Judgement, where all subjects have common dispersion parameter. The generalized item response (GIRT) model, first proposed by Torgerson (1958), is an extension of the IRT model, in which each subject is characterized not only by the ability parameter (θ) but also by the dispersion parameter (ϕ). For a subject characterized by (θ, ϕ) , the probability that he / she answers the item correctly is given by $\Pr(U=1|\theta, \phi) = \Phi\left(\frac{a}{\sqrt{1+\phi^2 a^2}}(\theta-b)\right)$, where Φ is the standard normal or logistic distribution function and (a, b) constitutes the set of usual item parameters. In this article, an item parameter estimation method maximizing the marginal likelihood where the subject parameters (θ, ϕ) are integrated out, is presented.

Key words : item response theory (IRT), generalized item response theory (GIRT), marginal maximum likelihood estimation, EM algorithm, law of categorical judgement.

問 題

一般項目反応モデル

項目反応理論では特性値 θ を持つ個人の項目 j への正答確率が $P_j(\theta)$ という形の反応関数 (response function) で書き表わされる事を仮定している。具体的には、 $P_j(\theta)$ は、2パラメタの累積正規分布関数やロジスティック分布関数の形をとることが多いが、最近の項目反応理論の入門書において、反応関数が累積分布関数の形を持つ根拠が示されていないことが多い。この反応関数の形の説明は、項目反応理論の応用におい

ては必要ないが、モデルの一般化を考える場合には、その理論的根拠に対する知識も必要であろう。通常、反応関数が2パラメタの累積正規分布関数で表わされる理由は、以下の様にして説明されることが多い。すなわち、

1. ある項目に正答するための能力 θ (連続量) を考え、その θ への回帰は線形である。
2. 正答・誤答の形の項目反応は、その項目に正答するための能力値がある閾値を越えるか否かによって2分化されたものである。

このような説明は、通常の項目反応理論を多次元に拡張する場合(2値データの因子分析)には適したものである (Bartholomew, 1987; Takane & de Leeuw, 1987)。

他方、反応関数の形は Torgerson (1958) のカテゴリー判断の法則 (Law of Categorical Judgement) を用いて説明する事も可能である。カテゴリー判断の法則は

* 東京大学大学院教育学研究科 (Department of Educational Psychology, Faculty of Education, The University of Tokyo)

** 大学入試センター研究開発部 (Research Division, The National Center for University Entrance Examination)

刺激と刺激の比較を取り扱う Thurstone (1927) の弁別過程を、刺激と質問項目のカテゴリーの境界値との比較へと拡張したものであるが、いわゆる正答・誤答の形の項目反応は、個人を刺激とみなし、また、各項目をカテゴリー数が2の質問項目として取り扱うことになる。すなわち、

1. 個人 i の特性値 θ_i と項目 j の特性値 b_j は、ある連続体上 (以下これを能力尺度と呼ぶ) に順序づけられている。

2. 様々な理由から実際の尺度値は、これらの特性値に誤差が加わった形、すなわち、

$$X_i \sim N(\theta_i, \phi_i^2) \quad (1)$$

$$Y_j \sim N(b_j, \psi_j^2) \quad (2)$$

で実現する。ただし、 $X_i \sim N(a, b)$ は確率変数 X_i が平均 a 、分散 b の正規分布に従うことを示す。

また、 ϕ_i^2 は尺度値の実現値 X_i の分散、 ψ_j^2 は Y_j の分散であるとし、さらに X と Y は独立であるとする。

3. 個人 i が項目 j に正答するためには、回答を試みる時点で、個人 i の能力尺度値 (X_i) が項目 j の尺度値 (Y_j) を越える必要がある。

以上の3つを仮定する事により、個人 i の項目 j への正答確率は

$$\begin{aligned} \Pr(u_{ij}=1 | \theta_i, \phi_i) &= \Pr(X_i > Y_j) \\ &= \Phi\left(\frac{\theta_i - b_j}{\sqrt{\phi_i^2 + \psi_j^2}}\right) \end{aligned} \quad (3)$$

と導かれる。ただし、 u_{ij} は個人 i の項目 j への反応であり、 Φ は標準累積正規分布関数である。したがって、

$$a_j = \frac{1}{\sqrt{\phi_i^2 + \psi_j^2}} \quad (4)$$

というパラメタの変換を行えばこれが2パラメタ累積正規モデルの反応関数となる。なお、この(4)式を見るとわかるように、2パラメタモデルでは ϕ_i が各個人に共通の値 (ϕ) をとらなければならない。ところで(3)式においては (θ, b) と (ϕ, ψ) の間には乗算的不定性が、また ϕ と ψ との間には、両者が正の値をとるという条件の下で、加算的不定性が存在する。したがって、この場合 $\phi=0$ としてもよい。また、 ψ_j を共通としてロジスティック関数を使用すれば Rasch モデルとなる。

ところで、項目反応モデルの応用場面においては、ある特定集団において θ が平均0、分散1の正規分布からのランダムサンプルであること、すなわち

$$\theta_i \sim N(0, 1), i=1, \dots, N \quad i.i.d. \quad (5)$$

を仮定して特性値の尺度を定義することが多いが、い

わゆる識別力のない項目とは、 X の級間分散 [(5)式で定義した分散1] に比べ級内分散 (ϕ^2) が大きいために、級間すなわち各個人の識別が困難な場合であると考えることもできる。

本研究で扱う一般項目反応モデルは、上記のカテゴリー判断の法則に基づくアプローチにおいて、個人の誤差分散 ϕ^2 の均一性に関する仮定をはずしたものである。これは誤差分散の個人差を認める、すなわち、各個人の能力特性値の変動を考慮することにほかならないが、このことにより従来の項目反応モデルでは当てはまりの悪い被験者のデータに対してもモデルを適用することが可能となる (孫・芝, 1990)。この項目反応データのモデルとしては、Torgerson (1958) の一般累積正規モデルや Shiba (1979) のロジスティック関数を用いた一般項目反応モデルがある。また、中村・豊田 (1991) では、Thurstone の比較判断の法則と一般項目反応モデルとの対応関係が論じられている。

ところが、このようにモデルそのものは以前から提案されているが、Torgerson (1958) ではパラメタの推定方法については論じられておらず、一般項目反応モデルを用いた適用例は長い間報告されていなかった。その後、孫・芝 (1990) は外国人に対して日本語語彙テストを実施し、通常の項目反応モデルではデータとモデルとの適合度が悪くなる状況で一般項目反応モデルを適用し、そのモデルの有効性を示した。しかしながら、孫・芝 (1990) のアプローチは、通常の項目反応モデルの下で推定された項目パラメタを一般項目反応モデルのパラメタであるとみなし、これを既知とした場合に個人パラメタ (θ_i と $\frac{1}{\phi_i}$) を最尤推定するというもの

であり、依然として一般項目反応モデルにおける項目パラメタの推定法は確立していない。

そこで本研究では、一般項目反応モデルのパラメタ推定法として、項目パラメタについては周辺尤度を利用した最尤推定法を、個人パラメタについては事後平均を利用したベイズ推定法を提案する。

周辺最尤推定法

通常の項目反応モデルにおけるパラメタ推定法としては、長い間 Birnbaum (1968) の同時最尤推定法が用いられ、この方法を取り入れたコンピュータプログラムとして LOGIST (Wingersky, Barton, & Lord, 1982) がある。この推定法は、理論的に平易であるために、項目反応モデルのパラメタ推定法として一般的となった。しかしながら、項目反応モデルにおいて構造パラメタ (項目パラメタ) と付随パラメタ (個人パラメタ) を同時に

推定すると、推定された構造パラメタは一致推定量にならないという問題がある (Anderson, 1972 ; Wright, 1977 ; Hambleton & Swaminathan, 1985 ; Harwell & Baker, 1991)。すなわち、パラメタの精度を上げるため、被験者数 (標本数) を増やしても、それに伴って推定すべき個人パラメタの数が増え、さらに項目パラメタの推定値が真の値に収束しないといった問題である。

そこで、同時推定法における統計学的に望ましくない性質を取り除くため、Bock & Lieberman (1970) は 2 パラメタロジスティックモデルにおいて、個人パラメタがある母集団からの独立なランダムサンプルであると考え、尤度関数から個人パラメタを積分消去し、周辺尤度関数を最大とするような項目パラメタを推定する方法を提案した。Bock & Aitkin (1981) は EM アルゴリズムを用いた周辺最尤推定法を提案し、計算の効率の面で Bock & Lieberman (1970) より改善された。Thissen (1982) は Rasch モデルにおける周辺最尤推定法を提案し、Mislevy (1984) は 3 パラメタロジスティックモデルにおいて、項目パラメタに関して事前分布を導入し周辺尤度関数を利用したベイズ推定法を提案している。さらに、Mislevy & Bock (1990) は、Mislevy (1984, 1986) の方法をコンピュータプログラム化し、BILOG として利用されている。

上記に示したように、通常の項目反応モデルの推定法としては、周辺尤度を利用した方法の方が同時推定法より優れていることがわかる。さらに、一般項目反応モデルの場合、通常の項目反応モデルに比べ、パラメタの数が多いため、周辺最尤推定法が有効であると考えられる。そこで、本研究では、計算の容易な面も考慮して、Bock & Aitkin (1981) に従って一般項目反応モデルにおける EM アルゴリズムを利用した周辺最尤推定法を提案する。

方 法

モデル

先に示したように、一般項目反応モデルでは個人の誤差分散 ϕ^2 に均一性を仮定しない。したがって、(3)式において $\psi_j = \frac{1}{a_j}$ とし、さらにロジスティック関数を用いると、項目特性関数は、

$$P_j(\theta_i, \phi_i | a_j, b_j) = \frac{1}{1 + \exp \left[-D \frac{a_j}{\sqrt{1 + \phi_i^2 a_j^2}} (\theta_i - b_j) \right]} \quad (6)$$

によって与えられる。ただし D は定数で標準累積正規

分布に近似する場合 1.7 とする。また、 θ_i は個人 i の能力分布の平均であり、 ϕ_i はその分布の標準偏差である。孫・芝 (1990) では、 θ_i を β_i として、また ϕ_i の逆数を α_i としてパラメタを設定しているが、本研究では ϕ_i の逆数をとっていない。その理由として、 ϕ_i の逆数をとらないほうが、パラメタの意味をとらえやすいからである。すなわち、 $\phi_i = 0$ のとき (6) 式は従来の項目反応モデルの項目特性関数に一致し、 ϕ_i の値が大きくなるにつれて (6) 式は通常の項目反応モデルから逸脱するものとする。このことから、本研究では ϕ_i を通常の項目反応モデルからの逸脱パラメタとして考えている。

項目パラメタの推定

周辺尤度関数

個人 i の項目 j に対する反応を u_{ij} とする。ただし、 u_{ij} は個人 i が項目 j に正答した時 1 を、誤答した時 0 を値とする変数である。項目数は m 個、被験者数は N 人とする。個人の特性値 θ_i 、 ϕ_i を固定したとき、各項目反応が互いに独立であると仮定し、 $\underline{u}_i = [u_{i1}, \dots, u_{im}]$ とすると、個人 i が反応した $\underline{u}_i$ をとる条件付き分布は、

$$f(\underline{u}_i | \theta_i, \phi_i; \underline{a}, \underline{b}) = \prod_{j=1}^m [P_j(\theta_i, \phi_i | a_j, b_j)]^{u_{ij}} [1 - P_j(\theta_i, \phi_i | a_j, b_j)]^{1-u_{ij}} \quad (7)$$

となる。ただし、 $\underline{a}$ 、 $\underline{b}$ はそれぞれ項目パラメタ a 、 b を要素とするベクトルである。

次に、母集団における個人特性値の分布としては、 θ_i に関しては正規分布を、 ϕ_i に関しては逆カイ分布を仮定すると、

$$t_1(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(\theta - \mu)^2}{2\sigma^2} \right), \quad -\infty < \theta < \infty \quad (8)$$

$$t_2(\phi) = \frac{\tau^{\frac{v}{2}}}{2^{\frac{v}{2}-1} \Gamma\left(\frac{v}{2}\right)} \phi^{-(v+1)} \exp \left(-\frac{\tau}{2\phi^2} \right),$$

$$0 < \phi < \infty \quad (9)$$

となる。ここで、 μ 、 σ 、 v 、 τ は母集団分布を規定する定数である。ところで、 ϕ の母集団分布の形が逆カイ分布である必然性は全くないが、ここではベイズ統計学における標準偏差の分布として標準的に用いられる逆カイ分布を利用した。さらに、個人特性値 θ 、 ϕ は従属としなければならない根拠がないので、互いに独立であると仮定し、個人が上記のように分布するパラメタを持つ母集団からランダムに抽出されたとする。そして、(7)式から θ_i 、 ϕ_i を積分消去とすると、

$$g(\underline{u}_i | \underline{a}, \underline{b}) = \int_0^\infty \int_{-\infty}^\infty f(\underline{u}_i | \theta, \phi; \underline{a}, \underline{b}) t_1(\theta) \times t_2(\phi) d\theta d\phi \quad (10)$$

となる。これは、個人特性値 θ_i , ϕ_i に依存しない反応パタンの周辺確率である。したがって(10)式より周辺尤度関数は、

$$L(\underline{a}, \underline{b} | \underline{U}) = \prod_{i=1}^N g(\underline{u}_i | \underline{a}, \underline{b}) \quad (11)$$

となる。

EM アルゴリズム

(10)式のように、積分を含む尤度関数を最大にするような最尤推定値を求める数値解析法として、EM アルゴリズムが提案されている (Dempster, Laird, & Rubin 1977; 前川 1991)。本研究では、この EM アルゴリズムを用いて項目パラメタを推定することとする。

まず、(10)式の積分を離散近似する。そのためには、 θ については Q 個の離散点 $X_q (q=1, \dots, Q)$ とそれに対応する重み $A(X_q)$ を、 ϕ については R 個の離散点 $Y_r (r=1, \dots, R)$ とそれに対応する重み $B(Y_r)$ を考え、これらを用いて(10)式の積分を以下のように近似する。

$$\int_0^\infty \int_{-\infty}^\infty f(\underline{u}_i | \theta, \phi; \underline{a}, \underline{b}) t_1(\theta) t_2(\phi) d\theta d\phi \\ \approx \sum_{r=1}^R \sum_{q=1}^Q f(\underline{u}_i | X_q, Y_r; \underline{a}, \underline{b}) A(X_q) B(Y_r) \quad (12)$$

$$= \sum_{r=1}^R \sum_{q=1}^Q k(\underline{u}_i, X_q, Y_r | \underline{a}, \underline{b}) \quad (13)$$

ただし、 $k(\underline{u}_i, X_q, Y_r | \underline{a}, \underline{b})$ は観測されたデータと欠測データを合わせた完全データ同時分布である。EM アルゴリズムは積分によって尤度から消される変数があったかも観測されていたものとして扱われることから、これを欠測データと呼ぶ。本研究において欠測データは、 $\theta_i, \phi_i (i=1, \dots, N)$ である。重み A, B は、Mislevy (1984, 1986) によれば、 t_1 や t_2 の値を規準化したもの

$$A(X_q) = \frac{t_1(X_q)}{\sum_{q=1}^Q t_1(X_q)}, \quad B(Y_r) = \frac{t_2(Y_r)}{\sum_{r=1}^R t_2(Y_r)} \quad (14)$$

を用いれば十分であることが示されている。

[E ステップ] 期待対数完全データ尤度の計算

E ステップは、期待対数完全データ尤度を計算するステップで、期待値 (Expectation) の頭文字 E をとって E ステップと呼ぶ。項目パラメタの仮の推定値 $\underline{a}^{(0)}$, $\underline{b}^{(0)}$ が与えられた時の期待対数完全データ尤度は、観測されたデータを与えたときの欠測データの条件付き分布による対数完全データ尤度の期待値である。前述のように積分を近似すると、

$$E \ln k \approx \sum_{i=1}^N \sum_{q=1}^Q \sum_{r=1}^R \ln [k(\underline{u}_i, X_q, Y_r | \underline{a}^{(0)}, \underline{b}^{(0)})] \\ \times h(X_q, Y_r | \underline{u}_i, \underline{a}^{(0)}, \underline{b}^{(0)}) \quad (15)$$

$$= \sum_{j=1}^M \sum_{r=1}^R \sum_{q=1}^Q [r_{jqr} \ln P_j(X_q, Y_r | a_j, b_j) + (N_{qr} - r_{jqr}) \\ \times \ln Q_j(X_q, Y_r | a_j, b_j)] + C \quad (16)$$

となる。但し、

$$Q_j(X_q, Y_r | a_j, b_j) = 1 - P_j(X_q, Y_r | a_j, b_j) \quad (17)$$

$$r_{jqr} = \sum_{i=1}^N u_{ij} h(X_q, Y_r | \underline{u}_i, \underline{a}^{(0)}, \underline{b}^{(0)}) \quad (18)$$

$$N_{qr} = \sum_{i=1}^N h(X_q, Y_r | \underline{u}_i, \underline{a}^{(0)}, \underline{b}^{(0)}) \quad (19)$$

$$h(X_q, Y_r | \underline{u}_i, \underline{a}^{(0)}, \underline{b}^{(0)}) \\ = \frac{k(\underline{u}_i, X_q, Y_r | \underline{a}^{(0)}, \underline{b}^{(0)})}{\sum_{r=1}^R \sum_{q=1}^Q k(\underline{u}_i, X_q, Y_r | \underline{a}^{(0)}, \underline{b}^{(0)})} \quad (20)$$

であり、 C は $\underline{a}, \underline{b}$ を含まない定数である。したがって、E ステップにおける実質的な計算は、 r_{jqr}, N_{qr} を求めることにある。

[M ステップ] 期待対数完全データ尤度の $\underline{a}, \underline{b}$ に関する最大化

M ステップでは、E ステップで求められた期待対数完全データ尤度を最大にするようなパラメタを求める。このことから、最大化 (Maximization) の頭文字 M をとって M ステップと呼ぶ。M ステップにおける(16)式の最大化は各項目ごとに行われ、

$$E \ln k_j = \sum_{r=1}^R \sum_{q=1}^Q r_{jqr} \ln P_j(X_q, Y_r | a_j, b_j) \\ + (N_{qr} - r_{jqr}) \ln Q_j(X_q, Y_r | a_j, b_j) \quad (21)$$

を最大化する a_j, b_j を求める。そのためには、期待対数完全データ尤度を各パラメタ a_j, b_j で偏微分して得られた方程式、

$$\frac{\partial E \ln k_j}{\partial \lambda_j} = 0 \quad (22)$$

を解けばよい。ただし、 $\lambda_j = (a_j, b_j)'$ である。なお、偏微分をとる際に、 N_{qr} と r_{jqr} の中に現れる $\underline{a}^{(0)}$ と $\underline{b}^{(0)}$ は定数として取り扱う (付録参照)。この(22)式は解析的に解けないため数値解法としてフィッシャーのスコア法を用いる。M ステップにおける繰返し計算のパラメタ更新式は、

$$\lambda_j^{(L+1)} = \lambda_j^{(L)} + J^{-1}(\lambda_j^{(L)}) g(\lambda_j^{(L)}) \quad (23)$$

となる。ただし、 $\lambda_j^{(L)}$ は L 回目のパラメタの値であり、 g は(21)式の 1 次微分ベクトルで、 J は(21)式の 2 次微分行列において r_{jqr} を $N_{qr} P_j(X_q, Y_r | a_j, b_j)$ で置き換えたものである (付録参照)。この(23)式により λ_j の値を更新した後、

$$E \ln k_j(\lambda_j^{(L+1)}) - E \ln k_j(\lambda_j^{(L)}) \leq \text{eps} \quad (24)$$

を評価し、(24)式が満たされていれば M ステップを終了

し、そうでなければ $\lambda^{(l)} = \lambda^{(l+1)}$ として、さらに(23)式によってパラメタを更新する。ただし、eps は所望の精度である。

標準誤差

周辺尤度に基づく情報行列は、

$$I_N(\lambda) = N \cdot E_u \left[\left(\frac{\partial \ln g(u|\lambda)}{\partial \lambda} \right) \times \left(\frac{\partial \ln g(u|\lambda)}{\partial \lambda'} \right) \right] \quad (25)$$

と定義され、これを

$$I_N(\lambda) = \sum_{i=1}^N \left[\left(\frac{\partial \ln g(u_i|\lambda)}{\partial \lambda} \right) \left(\frac{\partial \ln g(u_i|\lambda)}{\partial \lambda'} \right) \right] \quad (26)$$

と近似する (Mislevy & Sheehan, 1989)。ただし、

$$\frac{\partial \ln g(u|\lambda)}{\partial \lambda} = \sum_{r=1}^R \sum_{q=1}^Q \frac{\partial \ln f(u|X_q, Y_r, \lambda)}{\partial \lambda} \times h(X_q, Y_r|u, \lambda) \quad (27)$$

である。したがって、パラメタの精度の指標である標準誤差は

$$SE(\lambda) = \text{diag}(\sqrt{I_N^{-1}(\lambda)}) \quad (28)$$

によって得られる。

以上Eステップ・Mステップにおけるアルゴリズムを示したが、これらのステップを用いた全体のアルゴリズムは、

- 1) 項目パラメタの初期値の計算
- 2) Eステップ
- 3) Mステップ
- 4) 収束の判定 収束していないとき2)へ
- 5) 標準誤差の計算
- 6) 終了

となる。2), 3), 4)を一つにまとめてステージと呼ぶと、EM アルゴリズムでは各ステージごとで(11)式にある周辺尤度関数を増大させ、ステージ数が増すにつれてパラメタの値の変化が小さくなっていく。4)では前回のステージと今回のステージのパラメタ値の変化量を求め、最大変化量が所望の精度以内であれば収束したと判断し、そのときのパラメタの値を最尤推定値とする。

個人パラメタの推定

ここでは、先に示した方法などで項目パラメタの値が既知の場合における個人パラメタの推定法を論じる。この場合、前節までに示した項目パラメタの周辺最尤推定法には既に母集団分布という形で事前分布が取り入れられているため、ベイズ推定法を用いるのが自然であろう。個人パラメタの θ, ϕ の事前分布をそれぞれ (8), (9)式とすると、ベイズの定理より反応パターン u が

得られるときの θ, ϕ の同時事後分布は、

$$h(\theta, \phi|u) = \frac{f(u|\theta, \phi; a, b) t_1(\theta) t_2(\phi)}{g(u|a, b)} \quad (29)$$

である。したがって、ベイズ推定量として θ, ϕ の事後平均を用いると、

$$\bar{\theta} = \int_{-\infty}^{\infty} \theta \left[\int_0^{\infty} h(\theta, \phi|u) d\phi \right] d\theta \quad (30)$$

$$\bar{\phi} = \int_0^{\infty} \phi \left[\int_{-\infty}^{\infty} h(\theta, \phi|u) d\theta \right] d\phi \quad (31)$$

となるが、実際の計算には Bock & Mislevy (1982) に従い、積分の離散近似を用いればよい。また、これらの事後分布の標準偏差をもって推定の精度の目安とする。なお、 ϕ の値そのものや、 θ の事後標準偏差は Tatsuoka (1985, 1986) の ECI 指標と高い正の相関をもつことが明らかになっているが、その件に関しては別途報告することを考えている。

数 値 例

シミュレーション

本研究で提案した推定法が有効であることを示すため、シミュレーションを行った。データは、(6)式にあるモデルのもとで発生させた。パラメタの真の値は、以下のように決定した。項目パラメタ a は $\chi_{(2,1)}$ 乱数から得、項目パラメタ b は標準正規乱数から得た。個人パラメタについては、 ϕ を $\chi_{(3,1)}$ 乱数から得、 θ を標準正規乱数から得た。

また、推定場面における制約条件は以下のように設定した。積分近似の分割点の数は ϕ, θ とも15に決め、その範囲は $0.001 \leq \phi \leq 5, -4 \leq \theta \leq 4$ である。母集団分布を規定する定数は、 $v=3, \tau=1, \mu=0, \sigma=1$ とした。このようにパラメタ ϕ, θ の母集団分布を設定することで、全てのパラメタの不定性を取り除くことができる。また、 ϕ の上限を5とした理由として、一般項目反応モデルの節で示したように、これ以上の値をとった場合には X の級間分散に比べ級内分散 (ϕ^2) が大きくなりすぎて、モデルそのものが意味をもたなくなるからである。さらに、Mステップでは a の推定において推定値の上限値を3、下限値を0.001とした。また、収束条件は精度が0.001で、最大反復回数を100とした。

以上の条件のもと、項目数が50、被験者数が1000および500の計2種類のデータを発生させ、項目パラメタの推定に関し、本研究で提案した方法及び通常の項目反応モデルのもとでの周辺最尤推定法を適用し、推定精度について2つの方法を比較した。精度の指標とし

て、推定値と真値との相関係数 (r) と平均平方誤差の平方根 (d) を用いた。

結果

被験者数1000の場合、項目パラメタ a について一般項目反応モデル(GIRT)での周辺最尤推定値と真値との相関関係を示したものが FIG. 1 であり、他方通常の項目反応モデル (IRT) のもとでの周辺最尤推定値と真値との相関関係を示したものが FIG. 2 である。通常の項目反応モデルの推定と比較して一般項目反応モデルの推定の方が、相関係数 (r) は0.908と高く、平均平方誤差の平方根 (d) の値も0.217と小さく、推定値と真値とのずれが少ない。また、FIG. 2 をみると各点が x 軸に偏っていることから、通常の項目反応モデルを用いて a パラメタを推定すると全体的に低めに推定されるといえる。

同様に、項目パラメタ b に関して、一般の項目反応モデルの周辺最尤推定法と通常の項目反応モデルの周辺最尤推定法それぞれについて推定値と真値との相関関係を示したものが FIG. 3 と FIG. 4 である。 b パラメタについては、相関係数 (r) ならびに平均平方誤差の平方根 (d) とともに a パラメタのときのように2つのモデル間で大きな違いはない。

被験者数500の場合についても、先の被験者数1000と同様に、項目パラメタ a に関して、一般項目反応モデルの推定値と真値との相関関係を示したものが FIG. 5 であり、通常の項目反応モデルのもとでの推定値と真値との相関関係を示したものが FIG. 6 である。これらの結果を見ると、被験者数が500の場合も一般項目反応モデルの推定法が通常のモデルの推定法と比べて高い相関 (0.813) を示している。また、一般項目反応モデルの平均平方誤差の平方根 (d) の値も0.335と通常のモデルの推定法と比べ小さい。しかしながら、被験者数1000の場合と比べると図中の点のばらつきが大きく、推定値と真値とのずれは大きくなっていると言える。さらに、被験者数1000のときと同様に、通常の項目反応モデルの推定結果をみると、推定値が真値に比べ全体的に低めに推定されていることがわかる。

最後に、項目パラメタ b に関しては、一般項目反応モデルの推定値と真値との相関関係を示したものが FIG. 7 であり、通常の項目反応モデルの推定値と真値との相関関係を示したものが FIG. 8 である。相関係数 (r) をみると、両モデルの間に大きな違いはない。また、平均平方誤差の平方根 (d) において一般項目反応モデルのもとでの推定法が小さい値をとっているものの、いずれの方法もほとんど真値と大きなずれも

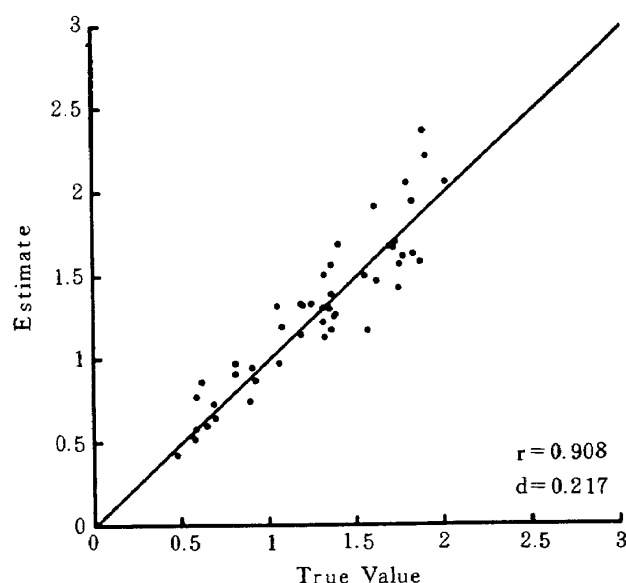


FIG. 1 Plot of estimated values and true values of the discrimination parameter in GIRT (N=1000).

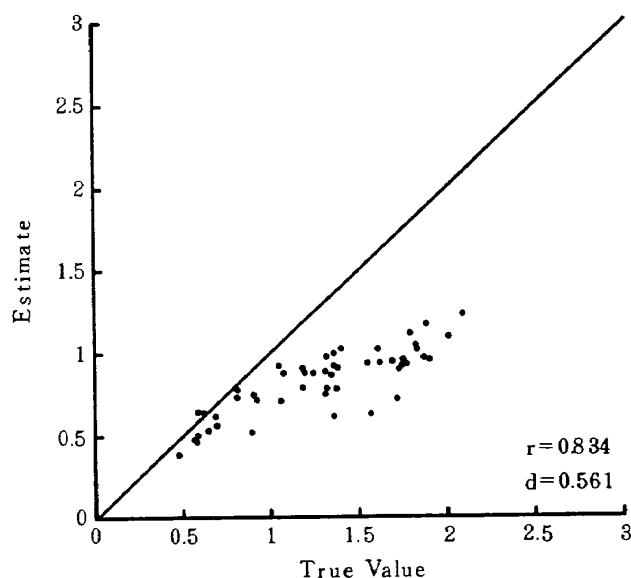


FIG. 2 Plot of estimated values and true values of the discrimination parameter in IRT (N=1000).

なく推定されているといえる。

考察

シミュレーションの結果より、今回提案した推定法は真値とのずれも少なく一般項目反応モデルの項目パラメタを推定する方法として適切であると考えられる。他方、従来行われてきたような通常の項目反応モデルの推定法では、推定された項目パラメタ a の値が真値と大きくずれたり、真値と比べ値が低めに推定されたりする。したがって、一般項目反応モデルに適合する

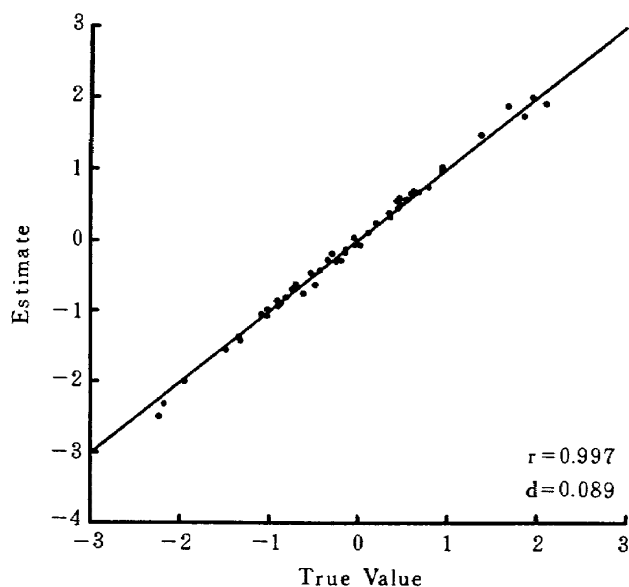


FIG. 3 Plot of estimated values and true values of the difficulty parameter in GIRT (N=1000).

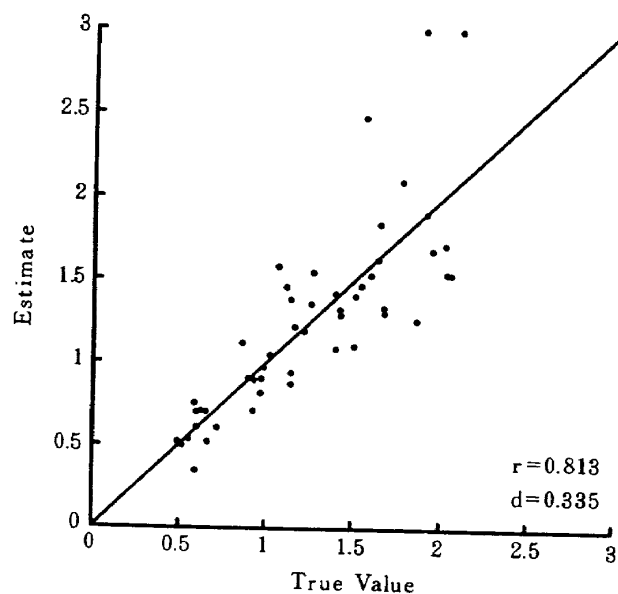


FIG. 5 Plot of estimated values and true values of the discrimination parameter in GIRT (N=500).

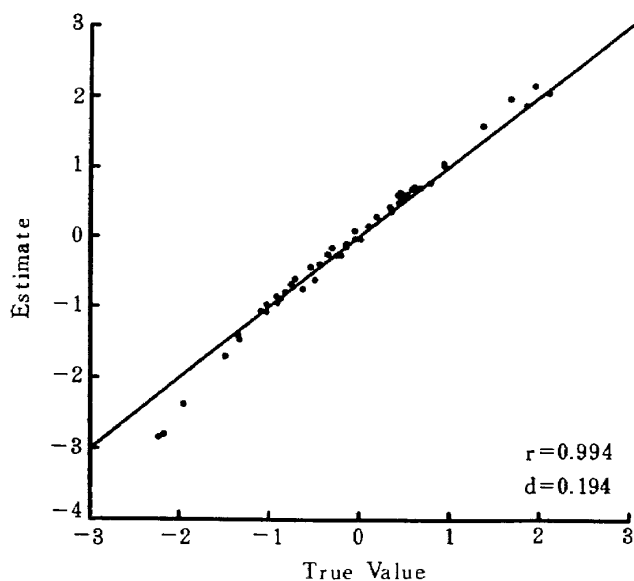


FIG. 4 Plot of estimated values and true values of the difficulty parameter in IRT (N=1000).

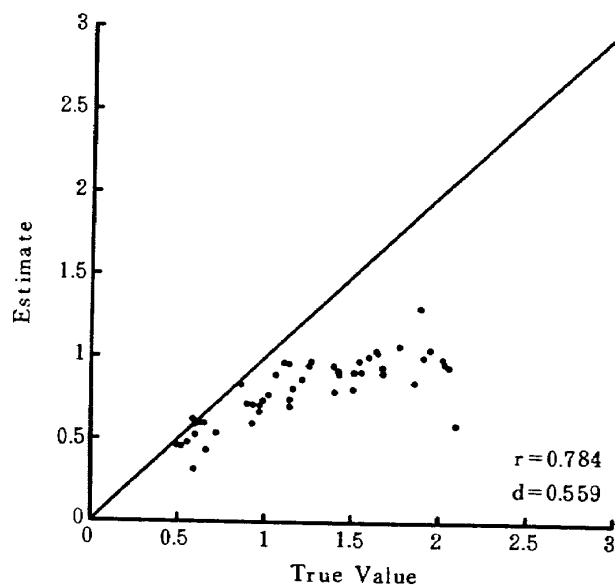


FIG. 6 Plot of estimated values and true values of the discrimination parameter in IRT (N=500).

データを分析する際には、通常の項目反応モデルのもとでの推定法を用いるのは適切ではない。確かに、従来一般項目反応モデルの項目パラメタ推定法が提案されていなかったため、簡易的に通常の項目反応モデルで項目パラメタを推定せざるをえなかったが、今後本研究で提案した方法を使えば、精度の高いパラメタ推定を行うことができると考えられる。

一般項目反応モデルを導入することの利点とは、通

常の項目反応モデルにない被験者の情報、すなわち個人パラメタ ϕ の情報を得られることであるため、このモデルの関心事は個人パラメタのみに向けられがちである。そのため、従来は通常の項目反応モデルから得られた項目パラメタを既知として個人パラメタを推定していた。しかしながら、本研究により一般項目反応モデルの項目パラメタ推定法が提案されたことで、完全な一般項目反応理論の下で個人パラメタを推定する

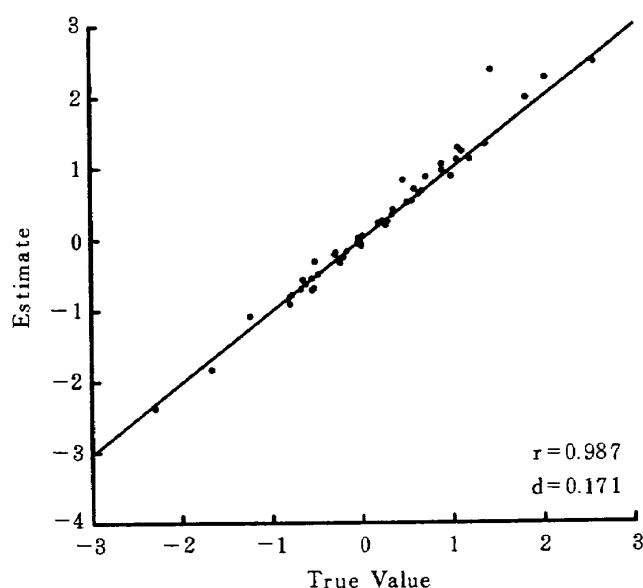


FIG. 7 Plot of estimated values and true values of the difficulty parameter in GIRT (N=500).

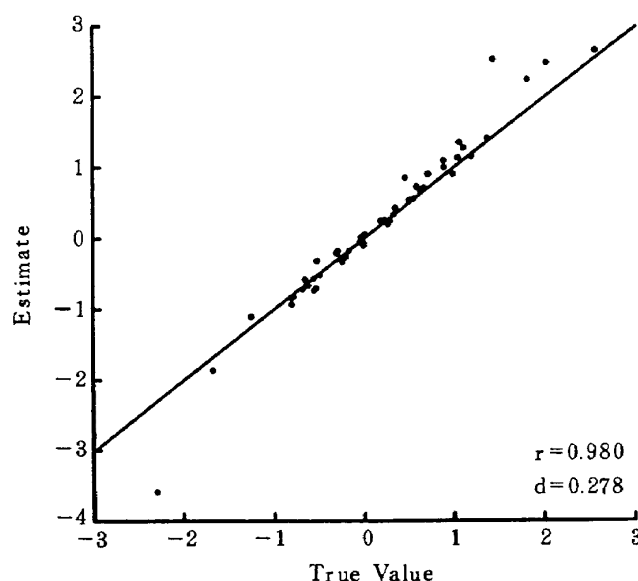


FIG. 8 Plot of estimated values and true values of the difficulty parameter in IRT (N=500).

ことが可能となった。

最後に、本研究の推定法を利用するにあたって注意する点をあげると、今回提案した方法は精度は良いものの、計算時間をかなり要する。また、シミュレーションの結果をみるとわかるように、パラメタの精度を上げようとするとき、標本数を増やさないとはいえないが、今回の結果からすると、項目数50ならば被験者数は、少なくとも500人は必要ではないかと考える。これよりも

少ない人数でも項目パラメタを推定することはできるが、現実データに適用した場合、項目パラメタ a の推定がかなり不安定になると考えられる。したがって、現実データを分析する場合、通常の項目反応モデルのときと比べ標本数をかなり増やす必要があるだろう。これは、通常の項目反応理論の場合は各 u_i が単に θ の分布を反映するようにサンプルされていればよいのに対して、一般項目反応理論の場合は θ と ϕ の同時分布を反映しなければならないことに由来している。

このように、本研究の方法は、いくつかの問題点を抱えているものの、一般項目反応モデルのもとでパラメタを推定できる点が評価できると考えられる。今後、計算の効率化や高速化そして適切な標本数に関する研究が必要であろう。

引用文献

- Anderson, E.B. 1972 The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society (Series B)*, 34, 42—54.
- Bartholomew, D.J. 1987 *Latent variable models and factor analysis*. Griffin.
- Birnbaum, A. 1968 Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores* (Pp. 395—479). MA : Addison-Wesley.
- Bock, R.D., & Aitkin, M. 1981 Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, 46, 443—459.
- Bock, R.D., & Lieberman, M. 1970 Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179—197.
- Bock, R.D., & Mislevy, R.D. 1982 Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431—444.
- Dempster, A.P., Laird, N.M., & Rubin, D.B. 1977 Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society (Series B)*, 39, 1—38.
- Hambleton, R.K., & Swaminathan, H. 1985 *Item response theory : principles and applications*. Boston : Kluwer-Nijhoff.

- Harwell, M.R., & Baker, F.B. 1991 The use of prior distribution in marginalized Bayesian item parameter estimation : A didactic. *Applied Psychological Measurement*, **15**, 375—389.
- 前川眞一 1991 パラメタの推定 芝 祐順(編) 項目反応理論：基礎と応用 Pp. 87—127. 東京大学出版会
- Mislevy, R.J. 1984 Estimating latent distributions. *Psychometrika*, **49**, 359—381.
- Mislevy, R.J. 1986 Bayes modal estimation in item response models. *Psychometrika*, **51**, 177—195.
- Mislevy, R.J., & Bock, R.D. 1990 PC-BILOG3 : *Item analysis and test scoring with binary logistic models*. Mooresville IN : Scientific Software Inc.
- Mislevy, R.J., & Sheehan, K.M. 1989 Information matrices in latent-variable models. *Journal of Educational Statistics*, **14**, 335—350.
- 中村知靖・豊田秀樹 1991 比較判断の法則と項目反応理論 芝 祐順(編) 項目反応理論：基礎と応用 Pp. 201—209. 東京大学出版会
- Shiba, S. 1979 Parameter Estimation in the general item response theory. *Personal communication*.
- 孫 媛・芝 祐順 1990 特異な反応パターンを示す被験者の能力推定 教育心理学研究, **38**, 360—368.
- Takane, Y., & de Leeuw, J. 1987 On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, **52**, 393—408.
- Tatsuoka, K.K. 1985 A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics*, **10**, 55—73.
- Tatsuoka, K.K. 1986 Diagnosing cognitive errors : Statistical pattern classification based on item response theory. *Behaviormetrika*, **19**, 73—86.
- Thissen, D. 1982 Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, **47**, 175—186.
- Thurstone, L.L. 1927 A law of comparative judgement. *Psychological Review*, **34**, 273—286.
- Torgerson, W.S. 1958 *Theory and methods of scaling*. New York : Wiley.
- Wright, R.D. 1977 Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, **14**, 97—116.
- Wingersky, M.S., Barton, M.A., & Lord, F.M. 1982 *LOGIST user's guide*. Princeton, NJ : Educational Testing Service.

付 録

Mステップにおける $\underline{g}$ ベクトルと $\underline{J}$ 行列の式を示す。

$$\underline{g}_a = \sum_{q=1}^Q \sum_{r=1}^R 1.7[r_{jqr} - N_{qr}P_j(X_q, Y_r)] \times \frac{X_q - b_j}{\sqrt{(1 + Y_r^2 a_j^2)^3}} \quad (32)$$

$$\underline{g}_b = \sum_{q=1}^Q \sum_{r=1}^R -1.7[r_{jqr} - N_{qr}P_j(X_q, Y_r)] \times \frac{a_j}{\sqrt{1 + Y_r^2 a_j^2}} \quad (33)$$

$$\underline{J}_{aa} = \sum_{q=1}^Q \sum_{r=1}^R 1.7^2 N_{qr} P(X_q, Y_r) Q(X_q, Y_r) \times \frac{(X_q - b_j)^2}{(1 + Y_r^2 a_j^2)^3} \quad (34)$$

$$\underline{J}_{bb} = \sum_{q=1}^Q \sum_{r=1}^R 1.7^2 N_{qr} P(X_q, Y_r) Q(X_q, Y_r) \times \frac{a_j^2}{1 + Y_r^2 a_j^2} \quad (35)$$

$$\underline{J}_{ab} = \sum_{q=1}^Q \sum_{r=1}^R -1.7^2 N_{qr} P(X_q, Y_r) Q(X_q, Y_r) \times \frac{a_j(X_q - b_j)}{(1 + Y_r^2 a_j^2)^2} \quad (36)$$

ただし, $Q(X_q, Y_r) = 1 - P(X_q, Y_r)$ である。

付 記

本研究は柴山・繁樹 (1992)*とは独立に行われたことを付記しておく。

本研究で提案した推定法を実行するために、パーソナルコンピュータ上で動作するコンピュータプログラムを作成した。プログラムの開発言語は Turbo Pascal 6.0a で、現在 NEC PC-9801 上で動作する。動作に関しては、CPU が i80286 以上、EMS および数値演算コプロセッサが必要、解析できるデータの制限は、項目数 100、被験者数 1500 である。入手希望の方は、筆者まで連絡のこと。

(1992.9.2 受稿, 12.21 受理)

* 一般項目反応モデルにおける母数の推定方法。日本行動計量学会第 20 回大会発表論文抄録集, 262—265.