項目反応理論における新しい多肢選択モデルの母数推定

繁 枡 算 男*, 藤 森 進*

PARAMETER ESTIMATION IN
A NEW MULTIPLE CHOICE ITEM RESPONSE MODEL

Kazuo SHIGEMASU, Susumu FUJIMORI

We propose a new response model for multiple choice items, which is an extension of the three-parameter logistic model. This model enables us to make use of the information provided by distractors as well as correct item alternatives.

The essential idea of this model is the further partition of the incorrect response curves in terms of the weighted average of the "ideal" correct curve and incorrect curve. By using artificial data, we showed that this model can produce better estimates of the ability parameters than the usual three parameter logistic model.

1. 目 的

多肢選択形式のテストの利点の一つは、テスト結果を数量的に扱うことが容易であることである。このことを利用して、テストの特徴を明らかにするために、あるいは潜在する能力や学力の推定の精度を高めるために、テスト得点の統計的取り扱いのための理論が発展してきたと言える。テスト理論には、大別して、古典的テスト理論と項目反応理論 (Item Response Theory) がある。前者が、ある被験者集団に実施されたテスト得点の一種の記述統計学であるのに対し、後者は、テスト得点のデータ発生モデルを想定して、そのモデルの母数を推定しようとするものである。モデルを特定することによって、各被験者が異なるテスト項目から成るテストを受けたときにも各人の能力の推定が可能になり、また、将来、テストを受けに来る被験者のための最適な項目の選択が可能になる (Lord and Novick, 1968, 池田, 1977, Lord, 1980 等参照)。しかし、通常の項目反応モデルの場合、多肢選択テストに対する反応は、正答か誤答かという 2 分変数 (dichotomous variable) として処理される。このとき、被験者が、正答ではないがある意味で正答に近い選択肢に反応した場合と、一見して間違いとわかる選択肢に

反応した場合が区別されず、一括して誤答として処理される。これはテスト反応が本来持っている情報の損失である。本論文は、誤答の選択肢に対する反応の持つ情報を利用することができモデルを構成することによって、(1) テスト項目の特徴をより明確にすること、および、(2) 能力の推定の精度を高めることを目的とする。すなわち、各テスト項目の特徴を、正答項目の反応曲線で表わすだけでなく、各選択肢への反応をモデル化することによって、各項目の特徴はより明らかになり、テスト項目の最適な選択にも有益な情報をもたらす。また、目的 (2) に関しては、我々はすでに、統計的推論の方法を改良することによって、通常の 3 母数ロジスティックモデルの母数の推定の精度を高めるという結果を得ているが (繁枡, 藤森, 1984), 本論文では、各選択肢への反応を情報として組入れることのできるモデル構成によって、能力の母数の推定の精度を高めようと意図するものである。

2. モ デ ル

前節で述べた目的のための多肢選択形式の各選択肢に対する反応モデルは、すでにいくつか提案されている。これらのモデルは、各選択肢にそれぞれ異なる潜在的な反応傾向を仮定する多項モデル (multinomial model, Bock, 1972), また、単一な反応傾向を仮定し、それを各選択肢に区分する段階的反應モデル (graded

* 東京工業大学 (Tokyo Institute of Technology)

response model, Samejima, 1969, Andrich, 1978 等)の2つがある。さらに、各々のモデルにおいて、潜在的反応傾向に正規分布を仮定する場合と、正規分布の分布関数をロジスティック関数で近似する場合に分けることができる。また、当て推量 (random guessing) を説明するために多項モデルを一般化したモデル (Samejima モデル, Samejima, 1979), この Samejima モデルをさらに修正したモデル (Thissen, 1984) などが見受けられる。

本論文は、これらと別に、簡明であり、かつ、当て推量を含めた複雑な要因によって決定される項目反応を近似するための十分な柔軟性を持つモデルを提出する。このモデルを説明するために、次のような記号の用い方をする。

記号 (Notation)

i : i 番目の被験者

j : j 番目の項目

k : k 番目の選択肢

N : 被験者の総数

n : 項目の総数

m_j : j 番目の項目の選択肢の数 (以降誤解の余地がない場合には j を省略)

θ_i : 被験者 i の能力母数

X_{ijk} : i 番目の被験者の j 番目の項目の k 番目の選択肢に対する反応を表わすダミー変数, 反応するとき, $X_{ijk}=1$, それ以外, $X_{ijk}=0$.

簡単のために、一般性を失わず、第1番目の選択肢を正答と仮定する。このとき、3母数ロジスティックモデルは、

$$P(X_{ij1}=1) = c_j + \frac{1-c_j}{1+\exp[-Da_j(\theta_i-b_j)]} \quad (1)$$

と表わされる。ここで、 $D=1.7$, a_j は項目 j の識別力母数, b_j は困難度母数, c_j は偶然正答確率を示す。テスト得点の項目反応モデルは、もともと、潜在的な反応傾向に正規分布を仮定することによって得られる。ロジスティックモデルはこの正規分布モデルの近似として正規分布の母数から導かれる a_j と b_j を固定したうえでロジスティックモデルの近似を最大とするように開発されたので、 $D=1.7$ とおくことが習慣となっている (テスト得点のモデル化の基礎については、Lord and Novick, 1968 参照)。ロジスティックモデルは、医学や生物学において、ある現象の出現比率を計量デー

タによって予測するために、非線形回帰分析の1つとしてよく用いられる (たとえば、フライス, 1975 参照)。ただし、モデル (1) や本論文で開発されるモデルでは、ロジスティック回帰分析における予測変数がデータとして与えられず能力母数 θ という潜在変数であり、この θ や他の項目母数から正答確率を予測しようとする構造になっている。関数

$$\begin{aligned} \zeta(\theta) &= \{1 + \exp[-Da_j(\theta - b_j)]\}^{-1} \\ \eta(\theta) &= 1 - \zeta(\theta) \end{aligned}$$

を定義すると、(1)式は、

$$P(X_{ij1}=1) = \zeta(\theta) + c_j \eta(\theta) \quad (2)$$

と書くことができる。 $\zeta(\theta)$ は、2母数ロジスティックモデルであり、当て推量で正答する確率を考慮しない場合の正答曲線を示す。逆に、 $\eta(\theta)$ は、本来誤答するはずの確率を示す誤答曲線である。

モデルA

(2)式を一般化して、選択肢 ($k=1, 2, \dots, m$) に対する反応を

$$P(X_{ijk}=1) = f_k \zeta(\theta_i) + g_k \eta(\theta_i) \quad (3)$$

とする。正しい選択肢 ($k=1$) に反応する全ての者が、問題項目の内容を理解していると限らないことを反映して、 $k=1$ に対する反応確率にも $\eta(\theta_i)$ が関与する。また、 $\theta \rightarrow \infty$ の場合も、 $f_k < 1$ とすれば、不注意等が原因で正答確率は1にならない。逆に、誤答の選択肢 ($k=2, \dots, m$) に反応する者の中にも、問題項目の内容を部分的に理解する者がいることを反映して、これらに対する反応にも $\zeta(\theta_i)$ が関与する。(3)式において、 $\sum_k^m P(X_{ijk}=1) = 1$ であるから、 $\sum_k^m f_k = 1$, $\sum_k^m g_k = 1$ という制約条件を課す。このモデルでは、 $f_k > g_k$ のとき、 $P(X_{ijk}=1)$ は、 θ の単調増加関数、 $f_k < g_k$ のとき $P(X_{ijk}=1)$ は、 θ の単調減少関数となる。

モデルB

3母数ロジスティックモデルを一般化する際に、正答選択肢に対する反応モデルは(1)式で十分表現できると仮定する。このとき、残されたモデル化の問題は、誤答選択肢全体の確率 $1 - P(X_{ij1}=1)$ をいかに $k=2, \dots, m$ に配分するかということである。誤答選択肢にも正答の要素があることから、基本的には、(3)式をモデルとし、かつ、誤答の選択肢 ($k=2, \dots, m$) に対する反応確率の総和が、 $1 - P(X_{ij1}=1)$ となるという条件を課すモデルを次のように定める。すなわち、 $k=2, \dots,$

m に対し,

$$P(X_{ijk}=1)=[f_k\zeta(\theta_i)+g_k\eta(\theta_i)]\xi(\theta_i) \quad (4)$$

ただし, $\xi(\theta_i)=1-P(X_{ij1})=1-[\zeta(\theta_i)+c_j\eta(\theta_i)]$

$$\text{また, } \sum_{k=2}^m f_k = \sum_{k=2}^m g_k = 1$$

である.

モデル A とモデル B は, 基本的には, $\zeta(\theta)$ と $\eta(\theta)$ の加重平均によるモデルである. モデル A は, 非常に簡明なモデルである. 各選択肢の特徴は, f と g の大きさによって見ることができる. すなわち, f が大であるほど, その選択肢は正答に近いと判断できる. 明確な正答が存在しないようなテスト項目においては, データによって正答を決めることも可能である. ある種のテストの応用分野において, この簡単なモデルで十分であろう. しかし, 先に述べたように, モデル A の場合, 各選択肢の反応は θ の単調増加関数か単調減少となる. ところが, テスト項目によっては, 能力や知識が一定以上であるほうが反応し易い誤答選択肢がある. モデル A は, このようなテストのモデルとしては適しない. これに対して, モデル B は, 多様な曲線を描く柔軟なモデルである. たとえば, 図 1 は, モデル (B) の以下に示すケースを図に示したものである. すなわち, $a_j=.8$, $b_j=-.87$, $c_j=.29$, $f_2=.6$, $f_3=.4$, $f_4=0$, $f_5=0$, $g_2=0$, $g_3=0$, $g_4=.4$, $g_5=.6$. モデル (A) と異なり, 能力 θ が一定以上の者のほうが反応しやすい誤答選択肢をモデル化することができることが図 1 からわかる.

以降, 本論文では, 学力テストへの応用を目的として, モデル B について論ずる. モデル B は, 正答への反応と, その他の誤答選択肢への反応を区別して段階的に推定しているので, 正答反応曲線が十分正確に推

定されている限り, 正答か誤答かの 2 分変数による推定よりも能力母数の推定の精度が高いことが期待される.

モデルの推定法

モデル B の母数は次のように推定することができる. ① 3 母数ロジスティックモデルの推定のためのプログラム BAYIRT によって (1) 式の $P(X_{ij1}=1)$ および $\xi(\theta)$ を推定する (ここでは推定方法の相違, たとえば, 最尤推定法 (MLE) かベイズの推定方法か等を問題としていないので, わかりやすいように推定方法は全て最尤推定法を用いる). ② (2) 式に従って, 正答曲線 $\zeta(\theta)$ と, 誤答曲線 $\eta(\theta)$ を推定する. ③ 正答選択肢以外のデータを用いて f_k と g_k の推定値を求める. 項目 j において最適化すべき尤度関数は, (4) 式を P_{ijk} と表わすと,

$$L(f, g | a, b, c, \theta) = \prod_{i,k} P_{ijk}^{x_{ijk}} \quad (5)$$

となる. ここで,

$$f = (f_2, f_3, \dots, f_{mj})^t$$

$$g = (g_2, g_3, \dots, g_{mj})^t$$

$$a = (a_1, a_2, \dots, a_n)^t$$

$$b = (b_1, b_2, \dots, b_n)^t$$

$$c = (c_1, c_2, \dots, c_n)^t$$

$$\theta = (\theta_1, \theta_2, \dots, \theta_N)^t$$

である. a, b, c, θ は, 3 母数ロジスティックモデルの推定値によって求められているので, 所与とみなす. 自由母数は, f と g であるが, 最適化の過程を簡単にするために次のような制約条件を課す. すなわち, 能力母数 θ の母集団分布を密度関数 $p(\theta)$ で表わす. このとき, 選択肢 k に反応する集団全体の確率 $P(X_{*jk}=1)$ は, $k=2, \dots, m$ に対し,

$$P(X_{*jk}=1) = \int_{-\infty}^{\infty} [f_k\zeta(\theta) + g_k\eta(\theta)]\xi(\theta)p(\theta)d\theta \quad (6)$$

となる. $P(X_{*jk}=1)$ とその観測値 $\hat{P}_{*jk}$ (全被験者の内 k 番目の選択肢に反応した者の割合) を等しいとおく.

$$\text{また, } w_1 = \int_{-\infty}^{\infty} \zeta(\theta)\xi(\theta)p(\theta)d\theta$$

$$w_2 = \int_{-\infty}^{\infty} \eta(\theta)\xi(\theta)p(\theta)d\theta$$

とおくとき, (6) 式は,

$$\hat{P}_{*jk} = f_k w_1 + g_k w_2 \quad (7)$$

図 1. 多肢選択項目反応モデル

注. 図中の番号は選択肢番号を表す.

となる。母数 a, b, c, θ を所与とすると、 w_1, w_2 は所与となり、(7)式によって g_k は f_k の関数となる。本論文では、被験者の母集団分布を標準正規分とみなす。能力母数 θ の原点と尺度は任意であるので、このことは無理のない仮定であろう。結局 (5) 式の尤度の自由母数の数は、 $(m-1)$ 個となり、尤度の最大値を見いだすことが簡単になる。

3. シミュレーション

ここでは前述のモデル B (以下多肢選択モデルと呼ぶ) について数値的に検討を加える。分析するデータは、A, B の 2 種類であり、いずれもコンピュータによって計算された人工データである。

(I. データ)

データ A は、次のような手続きで産出された。項目の識別力母数 a はカイ分布 (自由度 20, 尺度母数 0.03), 困難度母数 b は切断正規分布 (平均 0.3, 分散 1, 区間 $[-2, 2]$), 偶然正答確率の母数 c は、区間 $[0.02, 0.4]$ の一様分布に従って作られた。各項目の誤答選択肢の重み f と g は、特徴的なパターン (例えば、表 2 参照) を示すように主観的に定めた (この際選択肢数は 5 とし、正答は選択肢番号 1 に割り当てた)。このようにして母数が決定された 20 項目の仮想的テストを、100 人の仮想的な被験者に実施した。被験者の能力母数 θ は、標準正規分布に従う乱数を用いられた。正誤の反応パターンは、0 から 1 の一様乱数の値と、項目母数及び各被験者の能力母数を (1) 式に代入した値を比較し、前者が後者をこえるとき誤答、こえないとき正答、つまり選択肢番号 1 を選んだとした。さらに誤答のときには、誤答のどの選択肢に反応したかを決定するため、まず、(1) 式に、 $k=2$ としたときの (4) 式の値を加えた値と乱数値を比較し、乱数値が小さければ選択肢番号 2 に反応したとする。乱数値が大きかった場合には、 $k=3$ 以降も同様に、反応確率曲線を積み上げて、乱数値と比較する作業を繰り返し、選択肢番号を決定する。この反応パターンが各母数を推定する為のデータである。またクロスバリデーション用のデータとして、同じテストを用いて、更に 200 人の別の仮想的被験者の反応パターンを作成した。

データ B は、データ A と同一の手続きで定めた。ただしデータ A とテスト項目の母数は異なる。また、データ A の場合と異なり、各母数を推定する為の計算に利用した人数は、200 人、クロスバリデーション用のデータは 100 人である。

(II. 母数の推定)

反応パターンからの 3 母数ロジスティックモデルによる母数 a, b, c 及び θ の推定は、コンピュータプログラム BAYIRT の改訂版によった (BAYIRT については、繁樹・藤森 1984 を参照)。この推定値をもとに重み f と g を、推定し、この結果から各被験者の能力母数を推定し多肢選択モデルによる能力母数の推定値とした。

(III. 重み f と g の推定)

まず (5) 式対数を最大にする項目 j の選択肢番号 2 の f の推定値を求める (選択肢番号 1 は正答であり、 f と g の推定に関係ない)。この f に対応する g は、(7) 式を利用して f の関数として求める。このとき当該の選択肢も含め、各 f と g は、 $0 \leq f_k \leq 1, 0 \leq g_k \leq 1, \sum_{k=2}^m f_k = 1, \sum_{k=2}^m g_k = 1$ の制約条件により変更する。続いて選択肢番号 3, 4, 5 について、順にこの手続きを繰り返して第 1 回目の推定値とする。再び選択肢 2 にもどり前記の手続きを繰り返す。以上のようにして、前回の推定値との差が全て .001 以下になったときに推定を終了する。

(IV. クロスバリデーション)

3 母数ロジスティックモデル、多肢選択モデルとも、以上のようにして得られたデータ A, B に関する項目母数 (及び重み f と g) を利用して、クロスバリデーションを行なった。すなわち、データ A のテストの場合は、母数の推定に利用しなかった 200 人 (データ B の場合は 100 人) の被験者に対して、推定された項目母数 (及び f と g) を用いて、能力母数の推定を行ない、その成績を調べた。

4. 結 果

表 1 は、データ A, B の多肢選択モデルによる各選択肢の重み f と g の推定値と真値との相関係数の値である。

表 2 は、多肢選択モデルによる各選択肢の重み f と g の推定値と真値の具体的な例である。データは、データ A であり、項目番号 1 は、推定の成績が、比較的よ

表 1. 重み f と g の推定値と真値との相関

	f	g
データ A	0.726	0.835
データ B	0.758	0.896

表2. 重み f と g の例
項目番号 = 1 (データ A) 項目番号 = 4 (データ A)

	識 別 力	困 難 度	偶 然 正 答			識 別 力	困 難 度	偶 然 正 答	
真 値	. 53	－. 34	. 16		真 値	. 69	. 61	. 35	
推 定 値	. 874	－. 617	. 034		推 定 値	. 849	－. 272	. 02	
選択肢番号	2	3	4	5	選択肢番号	2	3	4	5
f 真値	. 9	. 1	. 0	. 0	f 真値	. 5	. 4	. 1	. 0
f 推定値	. 799	. 073	. 122	. 007	f 推定値	. 207	. 006	. 268	. 520
g 真値	. 0	. 0	. 1	. 9	g 真値	. 0	. 1	. 4	. 5
g 推定値	. 0	. 051	. 005	. 943	g 推定値	. 222	. 091	. 593	. 094
反応度数	12	2	2	16	反応度数	9	2	18	13
真の θ の平均	0.03	－. 30	－. 29	－1.0	真の θ の平均	－. 03	－1.1	－. 52	－. 17
推定の θ の平均 1	－. 22	－. 55	－. 33	－1.1	推定の θ の平均 1	－. 47	－2.0	－. 67	－. 37
推定の θ の平均 2	－. 11	－. 60	－. 36	－1.1	推定の θ の平均 2	－. 40	－1.8	－. 51	－. 19

注意 推定 θ の平均 1 は、3 母数ロジスティックモデルによる能力母数の推定値の平均であり、推定 θ の平均 2 は、多肢選択モデルによるものである。

表3. 項目母数の真値と推定値の相関

	識別力	困難度	偶然正答
データ A	0.405	0.884	0.191
データ B	0.174	0.917	0.433

表4. 能力母数の真値と推定値の相関

	データ A	データ B
多肢選択モデル	0.905	0.884
3 母数ロジスティック	0.883	0.854
(正答数得点)	0.886	0.847

かった例であり、項目番号 4 は、推定の最も悪かったものである。なお能力母数 θ の平均の欄は、各選択肢に反応した被験者の θ の平均値である。

表3は、データ A, B の 3 母数ロジスティックモデルによる項目母数の推定値と真値との相関係数の値である。本研究で報告する多肢選択モデルの項目母数 a , b , c は、3 母数ロジスティックモデルによって得られた結果をそのまま利用している。

表4は、データ A, B に於ける、能力母数の真値と、3 種の推定値(多肢選択モデルによる推定値、3 母数ロジスティックモデルによる推定値、及び正答数得点)との相関係数である。

表5は、クロスバリデーションの結果得られた能力母数の真値と推定値の相関係数である。

図2は、データ A に於ける、能力母数の真値と、多肢選択モデルによる推定値との散布図である。

表5. 能力母数の真値と推定値の相関
(クロスバリデーション)

	データ A	データ B
多肢選択モデル	0.879	0.887
3 母数ロジスティック	0.859	0.873
(正答数得点)	0.869	0.876

図2

図3は、データ A に於ける、能力母数の真値と、3 母数ロジスティックモデルによる推定値との散布図である。

図3

5. 考 察

表1より, f, g どちらも推定の成績は, かなり良いことがわかる. ただし, 直接推定した f の方が, (7)式を利用して f から間接的に推定した g よりも推定の成績が劣っている. これは, 次のような理由によると考えられる. f, g の推定には, 誤答の選択肢の情報を, 主に利用している. ここで, f が, 項目に正答する確率 $\zeta(\theta)$ にかかる重みであるのに対して, g は, 項目に誤答する確率 $\eta(\theta)$ にかかる重みである. 誤答の選択肢には, 正答する確率 $\zeta(\theta)$ に関する情報が, 誤答する確率 $\eta(\theta)$ と比較して少ないと考えられる. このことが g と f の推定の成績の差をもたらしたと考えられる.

また, f と g の推定の成績には, 推定に利用する標本数の違いが, 反映されていると考えられる. すなわち表1においてデータB (200人) の f と g の推定値と真値との相関は, データA (100人) の場合を上回っている. この標本数の差は, f, g の推定に利用するデータが, 本質的には, 該当項目に誤答した被験者に限られていることから, 更に強調されていると考えられる (表2参照).

表2より, 項目1においては, f, g ともに, 真値の大小のパターンと推定値のそれとがほぼ一致していると言える. 項目4においては, f, g ともに, 真値の大小のパターンと推定値のそれとが一致していない. 両項目とも, 識別力の推定に関しては余り差はないが, 困難度・偶然正答の推定の成績は, 項目4は, 項目1に

比べ, かなり悪い. 項目4では, 困難度と偶然正答確率を低く推定し過ぎていた. 基本的には, f の真値の大きい選択肢に反応した被験者の能力母数の平均値は, f の小さい選択肢に反応した被験者の平均値よりも, 高くしてかかるべきである. 実際, 表2によれば, 項目1では, ほぼこの通りになっている. しかし項目4では, このようにはなっていない (特に推定された能力母数). このような現象が生じたのは, 真の偶然正答確率 (つまりシミュレーションデータ産出に利用した確率である) が項目4で大きく, データの中に存在する偶然的正答とそうでない正答とを区別できていないことによると考えられる. このことは, 項目母数及び重み f と g の場合だけでなく, 能力母数の推定にも影響を与えていると考えられる. 図2, 3は, 真の値と推定値との関連の程度を視覚的にとらえるためのものであるが, この図からも, 相対的に能力母数の低い場合の推定は, 高い場合に比べて悪いことがうかがえる.

表4より, データA, B どちらの場合も, 通常の3母数ロジスティックモデルによる能力母数の推定よりも, 多肢選択モデルによる能力母数の推定の方が, 良い成績であることが分かる.

表4の結果は, 項目母数 (及び重み f と g) の推定に利用した被験者の能力母数を推定するという状況の為, 項目母数 (及び重み f と g) の推定に利用しない別の被験者の能力母数の推定の際には, 推定の成績が悪くなる可能性がある. この傾向はモデルの母数の数が f と g だけ増加している多肢選択モデルには, より強く生じる可能性がある. しかし, 表5のクロスバリデーションの結果に於いても (データA, B どちらの場合も) 多肢選択モデルの方が推定の成績は良くなっていることから, 多肢選択モデルは, 3母数ロジスティックモデルに比べて, 能力母数の推定の精度を実際の意味に於いて向上させていると結論できよう.

なおデータBに関するクロスバリデーションの結果 (表5) が, 母数の推定に利用した標本の結果 (表4) より良くなっているのは単なる偶然であると考えられる. 正答数得点と真値の相関を表4と表5で比較してみると, この相関は, 母数の推定に利用したデータよりもクロスバリデーションのデータにおいて高いことがわかる. データBのクロスバリデーションの結果が良いのはこのようなデータの違いに起因していると考えるのが妥当であろう.

一般的にモデルの母数の数を増加させれば推定の成績が良くなるのは当然という印象があるかもしれない

い。しかし、ここで推定の精度を競っているのは、モデルの（項目母数の推定に利用しない別の被験者の能力母数も含めて）母数そのものの推定であり、項目母数の推定に利用した被験者の反応パターンの予測にあるのではない点に注意したい。

本研究で報告した多肢選択モデルは、全ての母数を同時に推定するのではなく、3母数ロジスティックモデルの結果をそのまま利用する形で簡単に適用可能であり、能力母数の推定の成績の向上を期待し得ることから、実際の適用にも有効であると考えられる。

6. 今後の課題

今後の課題としては、以下の3点が重要なものとしてあげられる。

① ここで示した多肢選択モデルが、通常の項目反応モデルより、テスト情報量や項目情報量を、どの程度増加させているかを比較検討する。多肢選択モデルの導入によって、テスト作成の過程に大きな影響を与える可能性があり（例えば、単なる項目の通過率では、テスト作成の際の項目選択の指標として適切でなくなる）、この点の考察が、実際のテストにおける多肢選択モデルの検討と合わせて、必要となろう。また、選択肢の持つ情報量の検討も必要である。

② f と g の推定の精度を向上させ、テスト項目の各選択肢の分析の検討の材料として耐え得るものとする。

③ 実際のデータを用いて、多肢選択モデルAとモデルB及び通常のロジスティックモデルとを相互に比較検討する。

参考文献

- Andrich, D. (1978) A rating formulation for ordered response categories. *Psychometrika*, **43**, 561-573.
- Bock, R.D. (1972) Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, **37**, 29-51.
- フライス, J.L. (佐久間 昭訳) (1975) 計数データの統計学. 東京大学出版会.
- 藤森 進 (1984) 項目反応理論に基づく制約条件付3母数ロジスティックモデル. 東京工業大学人文論叢, **10**, 21-32.
- Lord, F.M. and Novick, M.R. (1968) Statistical Theories of Mental Test Scores. Addison Wesley.
- Samejima, F. (1968) Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, **34**.
- Samejima, F. (1979) A new family of models for the multiple-choice item. *Research report* 79-4. University of Tennessee.
- 繁樹算男・藤森 進 (1984) 項目反応モデルにおける母数の同時推定, 教育心理学研究, **32**, 266-275
- Thissen, D. and Steinberg, L. (1984) A Response Model for Multiple Choice Items. *Psychometric Technical Report* No.1. Unive. of Chicago.

附 記

* BAYIRT の実行にあたっては、東京大学大型計算機センターを利用した。

—1985年9月初稿受付, 1985年12月改稿受付—