

【個人研究】

項目反応理論による多値データの分析について 段階反応モデルと部分得点モデル

藤 森 進*

Analysis for Polytomous Data Based on Item Response Theory

Susumu FUJIMORI

Generally, item response models, including the two-parameter logistic model, can only deal with binary data, such as correct or incorrect responses - not polytomous data. In item response theory, polytomous statistics are obtained using special models such as Samejima's graded response model. However, since these models are complex, it is not easy to estimate their parameters. In contrast, Fujimori's partial test score model makes estimating and interpreting parameters easy, because this model is a natural expansion of the two-parameter logistic model. In the present study, the partial test score model having these features is compared to the graded response model, which is used widely as an analysis model for polytomous data. The comparison was performed through simulations and analysis of real data. The partial test score model yielded good results in terms of AIC (Akaike's information criterion), as well as model reproducibility. It also showed good results in terms of estimation of parameters. That is, in the case of the graded response model, bias was observed in estimators when the number of items was small, whereas no such biases were observed for the partial test score model, which is indicative of the advantage of using the proposed model.

1. 研究の目的

2 母数ロジスティックモデル (Birnbaum, 1968) のような一般の項目反応理論では、正答、誤答のような 2 値データを処理することが出来るだけであり、多値のデータを分析することは出来ない。しかし心理学一般の研究では質問紙データなどで多値データを利用す

ることも多い。項目反応理論でも多値データが全く扱えないわけではなく、幾つかの分析モデルが開発されている。例えば Samejima (1969) の段階反応モデル (graded response model)、評定尺度モデル (rating scale model; Andrich, 1978) や partial credit model (Masters, 1982) はこの種の代表的なものとされる。しかし現実問題としては分析するためのソフトウェア環境が十分整備されていない

*ふじもり すすむ 文教大学人間科学部人間科学科

こともあってか、これらのモデルが広く一般に利用されているとは言い難い。

このような状況に鑑み、本研究では藤森(2001)の部分得点モデル(partial test score model)と段階反応モデルとの比較をシミュレーション及び質問紙データの分析結果をもとに行い、項目反応理論を利用した多値データに関する分析方法の検討を行うことにする。

多値データへ項目反応理論を適用するとしても、どのようなモデルが良いのか検討する必要がある。このためには理論的側面の検討、モデル母数の再現性や能力母数の分布などの影響を検討するためのシミュレーションによる検討、そして現実のデータにどの程度うまくあてはまるのかという実証データに基づく検討が求められるであろう。このため、本研究ではを1.3節で、を2節で、そしてを3節で行うことにする。

1.1. 段階反応モデル

m 個の順序性を持つ段階反応を許容する Samejima(1969)の段階反応モデルを初めに説明する。ここではテストのある一つの項目だけを考えることにする。被験者の回答の各段階に対応して決まるダミー変数を u とし、ある段階より低い段階に反応したとき0、そうでないとき1となるダミー変数を X とする。能力を持つ被験者が $X=1$ となる確率を

$$P_u^*(\theta) = P(X=1|\theta) \quad (1)$$

で表すとき

$$P_u^*(\theta) = P_u^*(\theta) - P_{u+1}^*(\theta) \quad (2)$$

$$P_u^*(\theta) > 0 \quad (3)$$

が段階反応モデルである。ただし

$$P_0^*(\theta) = 1 \quad (4)$$

$$P_m^*(\theta) = 0 \quad (5)$$

である。(1)は段階の間を確定する境界反応

曲線であり、(2)は段階の反応確率を表現する段階反応曲線となる。前記の条件を満たす限り(1)の関数はどのようなものでも自由であるが、本研究では良く利用される2母数ロジスティックモデル

$$P_u^*(\theta) = \frac{1}{1 + \exp(-a(\theta - b_u))} \quad (6)$$

$u = 1, \dots, m-1$

を仮定する。(3)式の条件のため、 a は(6)の全ての曲線に共通した母数となり、識別力と呼ばれる。 b_u は各段階の閾値に関係した母数であり困難度と呼ばれる。すなわち段階反応モデルでは項目ごとに1つの識別力と各段階の閾値に対応する $m-1$ 個の困難度母数を持っていることになる。

1.2. 部分得点モデル

部分得点モデルでは部分得点を r ($0 \sim 1$ の範囲)とするとき母数推定のための対数尤度は次式で表される。

$$\ell = \sum_{i=1}^N \sum_{j=1}^n s_j \{ r_{ij} \ln(P_j(\theta)) + (1 - r_{ij}) \ln(Q_j(\theta)) \} \quad (7)$$

ここで i は被験者、 θ はその特性値を表す母数、 j は項目、 s_j は2値項目換算の繰返し数、 P は2母数ロジスティックモデル、 $Q = 1 - P$ である。また N は人数、 n は項目数である。

すなわち部分得点モデルは2母数ロジスティックモデルをその基礎に置いている。また同一あるいは類似項目母数を持つ項目を被験者に対し潜在的に繰返し実施することを想定している。このとき r は繰返し実施の正誤平均と考えることができる。このような部分得点モデルと繰返し実施を考えた2母数ロジスティックモデルの最尤解は一致する(類似母数の時は近似)ことが証明できる(藤森,2002a)。

1.3. 理論的側面からの検討

部分得点モデルの特徴は、以下のようなものである(藤森,2002b)。項目母数は識別力と困難度で一般の利用者にも比較的理解し

やすい。モデル母数の数が他の多値モデルと比較して少なく推定上の問題がない。どのような解答/回答結果でも値を0~1の部分得点に変換すれば適用できるので応用範囲が広い。質問紙の評定尺度にも容易に適用できる。解答の段階数が増えても母数が増加しないので、採点結果の変更や微調整にも対応が容易である。2母数ロジスティックモデルとシームレスであり、識別力や困難度母数の解釈もそのまま2母数ロジスティックモデルと同様に行えるので、2値と多値の混在データなどでの運用に便利である。test-1 et / 組問などの部分得点化により、2値項目からなるテストに組問が存在する場合、組問を1つの部分得点項目としてとしての運用することが可能になり、副次的にテスト全体での母数の数が減る。残差得点にモデルを再適用して多次元データへの対応もできる。

については本研究では取り上げていないが特徴としては大きなものである。

続いて段階反応モデルの特徴は、以下のようである。項目反応理論で多値データに対応したモデルとして著名なものである。

発表されてから30年以上の時間がたち応用研究も報告されている(例えば野口(1999)など)。公開されている分析用ソフトウェアMULTILOGが存在する。各段階反応曲線が得られるため、各段階に対する反応にまで関心を持つ場合は便利である。例えば境界反応曲線の母数を利用すれば、段階ごとの反応が上のどの水準で発生するのかについて予測も出来る。

両モデルを比較してみると、モデル化の考えの違いが明らかである。段階反応モデルは、モデルの母数の数が多いだけ、部分得点モデルに比較してデータとの当てはまりも良くなると考えられる。もちろんモデル母数の数が多ければ、母数の正確な推定にはデータ量が必要であるなどの問題が生じる危険性がある。逆に部分得点モデルは単純であるだけ、当てはまりは劣ると予想されるが推定値などの安定性は良いと考えられる。この点に関しては

赤池情報量規準AICなどの観点が評価に必要となろう。

また段階反応モデルの点については、部分得点モデルでは、モデル上これは困難である。行くとすれば、事後的には段階ごとに被験者の分布を作成して検討することになるだろう。項目母数は所与であってもテスト実施の事前であれば、被験者の能力分布について適当なる仮定を設けて、シミュレーションを行い段階ごとに被験者の分布を作成して検討することになるだろう。いずれにしても段階反応モデルより手間がかかることになる。

このような利点がある一方で段階反応モデルは(そして今まで提案された殆どの多値モデルも)、各段階にモデル母数を置くため段階の変化に柔軟性を欠く問題がある。たとえば、能力テストで教師が採点時に20点満点である項目を、0、5、10、15、20の粗い5段階で評価していたとしたら5段階の段階反応モデルでの分析となる。しかし、もし誤字脱字などである答案を1点減点することにしたなら、たちまちモデルの母数の値だけでなく、母数の数そのものが変化してしまうという問題点がある。質問紙でも評定を5段階で行っている項目を4段階にしたなら、段階反応モデルでは、それまでの項目母数をそのまま利用するわけにはいかなくなってしまうのである。

以上のように、どちらのモデルも一長一短はあるのだが、各段階や評定についてはそれほど関心を持たず、項目の全体的な特性と能力母数の推定だけに関心がある場合には、部分得点モデルで十分と思われる。

2. シミュレーションによる検討

実際のテストや質問紙において回答がどのようなメカニズムで生起しているのかは正確に知ることは出来ない。部分得点モデルと段階反応モデルのどちらがより現実に近いかは現時点で判断することは難しい。このため両モデルの比較をシミュレーションで行うにしても、データの作成にあたっては、どちらかのモデルに従ってデータを作成すべきかを

決めることは困難である。あるモデルを仮定してデータを作成し、当該モデルと別のモデルで分析しても、別のモデルは良い成績を發揮することは期待できず、比較もうまく行えないだろう。このため本研究では、部分得点モデルを前提としたシミュレーションと段階反応モデルを前提としたシミュレーションの2つを行うことにした。各モデルは自己のモデルに従ったデータと他のモデルのデータの2つを分析することになる。もちろん実際のデータの発生はこの2者以外のメカニズムによる可能性もあるわけであるが、両モデルの公平な比較を行うことはできる。すなわちデータがモデルに従っている場合は、モデルの再現性がシミュレーションによって検討でき、またデータがモデルに従っていない場合はモデルの頑健性が評価できる。

2.1. 部分得点モデルを前提としたシミュレーションデータ

部分得点モデルを前提としたシミュレーションデータは、以下のようにして作成した。まず部分得点モデルの構成要素として2母数ロジスティックモデルを仮定する。この2母数ロジスティックモデルの母数の分布型を以下のように定める。識別力母数は、平均0.65、標準偏差0.25、下限0.3、上限2.0の切断正規分布、また困難度母数は、平均0、標準偏差0.5の正規分布に従うと仮定する。能力母数

は平均0、標準偏差1.0の正規分布に従うと仮定する。能力母数を標準正規分布に従って作成し、2母数ロジスティックモデルから予想される正答確率を、範囲0～1の一様乱数と比較し、前者が下回る場合被験者の反応を正答1、上回る場合誤答0とする。2母数ロジスティックモデルに従う、この2値データパターンを、被験者数500、項目数200として各10回繰り返し作成した(データ1～10)。ただし5項目ずつ同一母数としている。続いて、このデータの同一母数の5項目ずつの正誤の和の平均をとり0,0.2,0.4,0.6,0.8,1.0の6段階の値をとる部分得点データとした。すなわち部分得点データは、被験者数は2値デー

タと同様に500であるが、項目数は40となる。同様にして、前記の項目母数を所与として新たに被験者の能力母数を500人分追加してクロスバリデーション用のデータを作成した。

2.2. 段階反応モデルを前提としたシミュレーションデータ

段階反応モデルを前提としたデータは以下のようにして作成した。段階反応モデルも、その構成要素として2母数ロジスティックモデルを仮定する。またその母数の分布形も2.1節と同様である。またデータの段階数としては1から5の値をとる5段階を仮定する。このため、モデルより段階間の境界反応曲線は4つ必要となる。この曲線群を確定するため、まず識別力母数を1つ分布に従って発生させ、これを各境界反応曲線に共通する識別力とする。続いて4つの困難度母数を作成し、最も小さいものを選択して、段階1と2の境界反応曲線の困難度とする。以下同様にして困難度の小さなものから順にとり各境界反応曲線の困難度を決定する。これらの境界反応曲線間の差をとって各段階反応曲線とする。標準正規分布に従う能力母数を1つ作成し、この値を固定して各段階反応曲線で予想される反応確率の区間(を固定したときの全段階反応曲線の大きさの和は1である)に、0～1の一様乱数が入ったとき当該反応が生じたことにする。以上の過程を500人分繰返し母数推定用のデータとした。また2.1節と同様に、以上で確定した項目母数を利用してクロスバリデーション用のデータを500人分作成した。

2.3. 母数の推定

母数の推定は、両モデルとも項目母数と潜在特性値の交互同時最尤推定による自作のFORTRANプログラムによった。部分得点モデル用では交互同時最尤推定だけでなく項目母数の周辺最尤推定が可能であるが、段階反応モデルの推定プログラムは交互同時最尤推定のみに対応しているため、比較の便を考慮して両モデルとも項目母数を最尤推定した(結果は省くが部分得点モデルに関しては周

辺最尤推定と交互同時推定の結果に大きな差はない。両モデルとも 困難度の推定値は-3.5 ~ 3.5の範囲と定め、識別力の推定値の範囲は0.02 ~ 2.0としている。

2.4. シミュレーションの結果と考察

Table 1 は、部分得点モデルにより作成したシミュレーションデータ（以下部分得点データと呼ぶ）の能力母数の真値、正答数得点と両モデルで推定した の推定値の相関である。ただし相関は、ケンドールの順位相関係数を求めている（以下特に明示しない場合は相関はケンドールの順位相関を指す）。一般によく利用されるピアソンの積率相関でなくて順位相関を求めたのは、項目反応理論の現実の運用場面では正答数得点などと推定値の順位の逆転現象が問題となるケースが多いためである。Table 1 より明らかなように真の との相関は、いずれも部分得点モデルによる推定値の方が高いものの段階反応モデルとの差はほとんどない。これに比し、正答数との相関は、若干差が拡大し部分得点モデルによる推定値が高い相関を与えている。

なお部分得点モデルの項目母数の推定値に関しては、例えばデータセット1については識別力の平均自乗誤差 (MSE) は、0.0014、困難度に関しては0.0017であり、特に問題はない精度で推定値が得られている。ちなみに、データ作成の元となったモデルとは異なるので値の良し悪しは判断できないが、段階反応モデルの識別力とのMSEは0.027、困難度は0.31となっている。ただし困難度に関しては、段階反応曲線の困難度の推定値平均によってMSEを算出している。

さて統計的な視点からは A I C (Akaike information criterion, 赤池情報量規準)

$$AIC = -2(\text{モデルの最大対数尤度}) + 2(\text{モデルの自由母数の数}) \quad (8)$$

がモデルよさと自由母数の数のバランスをどのようにとるかについて我々に示唆を与えてくれる（赤池情報量規準については、詳しくは坂元ら（1983）など参照のこと）。AICで

Table1 部分得点モデルを基礎としたシミュレーション結果

データ	真値との相関		正答数との相関	
	GS	PS	GS	PS
1	0.916	0.924	0.972	0.980
2	0.895	0.901	0.960	0.969
3	0.912	0.920	0.962	0.975
4	0.906	0.914	0.958	0.968
5	0.907	0.912	0.958	0.964
6	0.907	0.909	0.959	0.969
7	0.903	0.908	0.961	0.970
8	0.904	0.911	0.964	0.976
9	0.906	0.912	0.960	0.969
10	0.918	0.924	0.970	0.983

注：GS:段階反応モデル PS:部分得点モデル
相関はケンドールの順位相関

Table2 部分得点モデルを基礎としたシミュレーション結果 (クロスバリデーションデータ)

データ	真値との相関		正答数との相関	
	GS	PS	GS	PS
1	0.911	0.917	0.972	0.980
2	0.912	0.912	0.961	0.968
3	0.918	0.924	0.967	0.977
4	0.906	0.910	0.965	0.971
5	0.906	0.909	0.962	0.967
6	0.914	0.920	0.962	0.971
7	0.907	0.910	0.962	0.972
8	0.914	0.919	0.965	0.976
9	0.909	0.914	0.960	0.968
10	0.921	0.926	0.972	0.984

は値が小さいほど良いモデルとされるため、モデルの自由母数の数は抑制されることになる。試みにデータセット1の段階反応モデルのAICを求めたところ、46245.34であり、部分得点モデルは22142.46となっている。AICは小さい値であるほうが良いのであるから圧倒的に部分得点モデルが良いことになる。これはいずれのデータセットについても全く同様であり、個々の数値を問題にする必要は無いと判断し他の結果は省略する。なぜこのような圧倒的な差が生じるかと言えば、段階反応モデルの段階2 ~ 4の生起確率に小さくなるものがかかりあり、これが対数尤度の計算の際に影響しているためである。本研究では、AICの観点からは明確な差が両モデルであったと考えるが、理論的側面の検討も含め他の

指標たとえばカルバック・ライブラー情報量等による比較も行うことが適当であろう。

続いて部分得点データの推定値を利用してクロスバリデーションデータの成績を比較してみよう(Table 2)。クロスバリデーションにおいても様相は先と同様で、真値との相関はいずれも部分得点データがほんの僅か上回り、正答数得点に関してはややモデル間の差が拡大することが示されている。

さて段階反応モデルにより作成したシミュレーションデータ(以下段階反応データと呼ぶ)についての結果がTable 3である。また同データに基づき推定された項目母数をクロスバリデーションデータに適用した結果がTable 4である。どちらにおいても部分得点モデルに比べて段階反応モデルの方が真のとの相関がほんのわずが高いが、大きな差ではなく、データセット1, 5のように逆転しているケースも見受けられる。また全体的に真値との相関がTable 1、Table 2に比較してやや低くなっており、段階反応モデルが部分得点モデルに比較して再現性が難しいモデルであることを示唆していると言えよう。部分得点モデルに比べて段階反応モデルによるデータの段階数が1つ少ないことを考えれば、このことは、更に強調できる。段階反応データであるにもかかわらず、正答数得点との相関は段階反応モデルより部分得点モデルによる推定値の方が高くなっている。しかも部分得点データの場合よりその差が拡大していることが分る。合計点との順序関係の逆転は段階反応モデルに多くなっているのであるが、この逆転が生じる理由は、段階反応モデルでは項目によって評定の間部分の段階反応曲線が相対的にかなり低くなり、他項目の回答結果によっては、当該項目の回答が の推定に(推定誤差の大きさはともかくとして、その点推定値の決定には)殆ど影響力を持たなくなるためである。この現象は段階反応モデルだけではなく、選択肢やカテゴリーごとの反応確率曲線をモデル化する他の多値モデルにも共通するため、これらのモデルの利用の際

Table3 段階反応モデルを基礎としたシミュレーション結果

データ	真値との相関		正答数との相関	
	GS	PS	GS	PS
1	0.844	0.845	0.917	0.949
2	0.837	0.834	0.925	0.945
3	0.839	0.829	0.904	0.934
4	0.856	0.847	0.911	0.941
5	0.823	0.825	0.904	0.946
6	0.832	0.832	0.907	0.946
7	0.847	0.845	0.930	0.956
8	0.841	0.837	0.913	0.949
9	0.832	0.831	0.914	0.953
10	0.840	0.840	0.923	0.947

Table4 段階反応モデルを基礎としたシミュレーション結果(クロスバリデーションデータ)

データ	真値との相関		正答数との相関	
	GS	PS	GS	PS
1	0.835	0.838	0.905	0.949
2	0.838	0.838	0.925	0.949
3	0.831	0.827	0.900	0.934
4	0.849	0.844	0.909	0.941
5	0.818	0.819	0.898	0.941
6	0.835	0.832	0.910	0.948
7	0.848	0.841	0.929	0.956
8	0.830	0.825	0.914	0.952
9	0.830	0.826	0.912	0.952
10	0.846	0.841	0.927	0.947

Table5 段階反応データ1の項目母数の真値と推定値の例

項目	識別力	困難度1	困難度2	困難度3	困難度4
8	0.906	-0.271	-0.083	0.386	0.393
	0.125	-0.382	0.667	3.000	3.000
15	1.311	0.122	0.301	0.358	0.383
	0.324	0.880	1.376	1.752	1.766

注：上段は真値、下段は推定値。困難度kは、段階k-1とkの境界反応曲線の困難度である。

には注意を払う必要があると考えられる。もちろん部分得点モデルも2母数ロジスティックモデルをベースとしている以上、合計点との逆転も一部生じざるを得ないが、その程度は低く抑えられている。

次に段階反応モデルの項目母数の推定値について検討しよう。例えばデータセット1については識別力の平均自乗誤差(MSE)は、0.2993、困難度に関しては0.0636であり、やや部分得点モデルの場合に比較して推定精度

が悪いことがわかる。Table 5 に示した項目 8, 15 はデータセット 1 の中の推定成績が悪い項目であるが、いずれも真の境界反応曲線が互いに接近しすぎている場合に推定が出来なくなっていることがわかる。このような場合には、2つの境界反応曲線を1つとして処理する(Bond & Fox, 2001)などの対処が必要になると考えられるが、段階反応モデルでは、どのような基準でこれを行うかなどの問題もあり本研究ではこれを行っていない。ある意味ではこのような対処を必要とすることに段階反応モデルの問題点が存在するとも言えよう。なおモデルは異なるものの部分得点モデルの識別力のMSEは0.0136となり段階反応モデルよりも良い推定値を与えている。ただし段階反応モデルの項目母数の推定に関しては周辺最尤推定では成績が改善する可能性も残されているので、ここではこれ以上取り上げず、別の研究で報告することにしたい。

以上をまとめると本研究のシミュレーションの条件下では、部分得点データ、段階反応データのいずれであっても真値との相関は、どちらのモデルを利用しても大きな差は生じていない。これに対し正答数得点との相関は明らかに部分得点モデルの方が良い成績であり、特に予期に反し段階反応データで差が大きくなっている。また項目母数の数が多い段階反応モデルで危惧されたクロスバリデーションデータでの不適合は生じていない(ただし母数推定上の問題は残る)。これは一つには、本研究が特にノイズをのせていない単純なシミュレーションであることも影響しているかもしれない。

3. 質問紙データによる検討

3.1. 質問紙データ及び母数の推定

本研究で分析の対象としたデータは、0県の青少年基本調査の中高校生1849人のデータで、教師と生徒の心理的距離について調べた9項目尺度である(Table 6)。各項目は「1.よくあてはまる」から「4.まったくあてはまらない」の4段階で評定を求めるものであるが、

9項目とも回答は値が小さいほど心理的距離が小さくなるように処理されている。部分得点モデルでは、このデータを0~1の部分得点として処理するために、各回答結果を x とするとき $r = (x - 1) / 3$ と変換して「0, 1/3, 2/3, 1」の4段階をとる部分得点データとした。また推定方法はシミュレーションと同様の方法である。なお本調査の詳細については木原ら(1997)参照されたい。

3.2. 質問紙データの結果と考察

段階反応モデルの識別力は部分得点モデルより全体的に高くなっている(Table 7)。これと類似の現象はデータを2値化して通常の2母数ロジスティックモデルで分析した場合にも生じる。たとえば4段階評定の場合、2値化は実質的には段階2を1に、段階3を4に置き換えるために生じる現象である。Table 7の「2値モデル」はこのようなデータ変換をした結果である。部分得点モデルに比較してやや識別力が高くなっていることが分る。このような現象が発生することは、Figure 1の項目4の群別の平均回答結果を4値と2値の場合で比較してみると良く理解できる。ただし図の作成にあたっては、4値2値いずれの場合も全ての回答を0~1の範囲に変換している。多値データの2値化が識別力の推定値に大きな影響を与えることが図より理解できる。境界反応曲線が実質的には多値の回答を2値化して得られるデータに基づき定義され、推定されるため、この2値化と類似の現象が生じている可能性もある。しかしシミュレーションの結果では、部分得点モデルの識別力と段階反応モデルの識別力の推定値は大きく異なっていなかったのだから、2値化を原因とすることにはやや疑問も残る。

もう一つの可能性として、テスト項目数の違いがある。シミュレーションでは40項目であったが、本質問紙は9項目であり、この違いが影響した可能性がある。このため1つシミュレーションを追加した。段階反応データのデータセット1の初めの10項目だけを取り、段階反応モデルにより交互同時推定値を求めたと

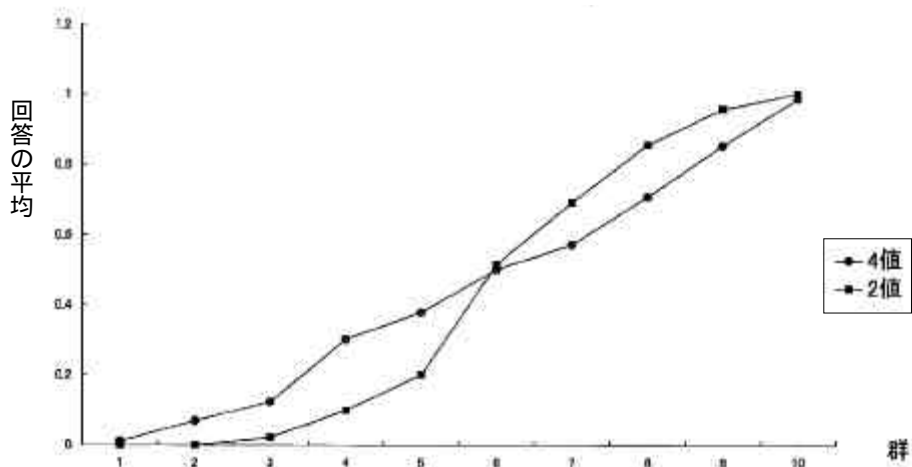


Figure 1 質問紙の項目4の2値及び4値による群別回答平均

Table6 質問項目

番号	項 目
1	先生は、自分の良いところをみつけてよくほめてくれる
2	先生は、友人との関係に気を配ってくれる
3	先生の考え方とあなたのお父さんやお母さんの考え方は、ほとんど同じだ
4	先生に話しかけると、いつも気持ちよく応じてくれる
5	先生は、本気で相談にのってくれる
6	先生は、自分のことを信頼してくれている
7	先生は、いつもいそがしそうだ
8	先生は、やさしく、あたたかい感じがする
9	先生は、いろいろなことを決めるとき、自分たちの言い分や希望をよく聞いてくれる

Table7 質問紙の項目母数の推定値 (交互同時推定)

項目	部分得点モデル		段階反応モデル			2値モデル		
	識別力	困難度	識別力	困難度 1	困難度 2	困難度 3	識別力	困難度
1	0.609	-0.608	1.047	-2.109	-0.686	0.865	0.856	-0.763
2	0.700	-0.560	1.016	-1.881	-0.656	0.716	0.751	-0.807
3	0.328	-0.374	0.561	-2.654	-0.394	1.908	0.416	-0.461
4	0.822	0.181	2.000	-0.991	0.204	1.363	1.486	0.332
5	1.052	0.086	2.000	-0.933	0.098	1.190	1.745	0.186
6	0.738	-0.211	1.879	-1.513	-0.195	1.034	1.386	-0.191
7	0.575	0.410	0.887	-1.064	0.395	1.964	0.671	0.594
8	1.149	-0.190	2.000	-1.230	-0.221	0.802	1.641	-0.208
9	1.037	0.032	1.985	-0.971	0.059	1.009	1.353	0.156

ころ、識別力の値はTable 8 のようになった。明らかに項目数が少なくなると識別力が大きくなる傾向があることが分る(MSE = 0.2647)。段階反応モデルでは少数項目での識別力の推定にバイアスがかかることは明らかとしても、その原因がモデルに存在するのか、それとも推定方法に問題があるのかは区別されなければならぬ。つまり段階反応モデルそのものではなくて、同時最尤推定という方法が推定に悪影響を及ぼしている可能性もある(同時最尤推定による項目母数の推定値が一致性を持たないことは良く知られている)。このため、MULTILOG (Thissen,1991) により周辺最尤推定値も算出したが、バイアスはほとんど変わらない(MSE = 0.2558)。このことは推定法の問題というよりはモデルに問題があることを示唆するものであるかもしれない。なおTable 8からは項目7は一見すると同時最尤推定の識別力が極端に低く、反対にMULTILOGの推定値は

妥当な値を得ているようであるが、実はMULTILOGは困難度において異常な推定値となっており、うまく推定できていないことには変わりはない。このためMSEの計算からは同項目を除外している。

部分得点データセット1の10項目を利用した部分得点モデルの分析では交互同時推定のMSEは、識別力が0.0094、困難度が0.0007であり、周辺最尤推定では識別力が0.0032、困難度が0.0014と

なった(Table 9)。同モデルでは交互同時推定、周辺最尤推定ともこのバイアスは認められないといって良いだろう。Table10には質問紙データを両モデルで分析し、周辺最尤推定を利用した推定値を示してある。同Tableより、シミュレーションと同様に、周辺最

Table8 段階反応モデル識別力推定値(段階データセット1)

項目	真値	同時推定 40項目	同時推定 10項目	周辺最尤推定 (MULTILOG)
1	0.529	0.528	0.580	0.700
2	0.550	0.559	0.579	0.310
3	0.768	0.877	1.079	1.200
4	0.817	0.856	1.555	1.420
5	0.559	0.543	0.590	0.730
6	0.334	0.358	0.396	0.570
7	1.122	1.389	2.000	2.120
8	0.906	0.125	0.112	0.720
9	0.727	0.928	1.707	1.480
10	0.575	0.521	0.562	0.700

Table9 少数項目の部分得点モデルによる推定値(部分得点データセット1)

項目	真値		同時最尤推定		周辺最尤推定	
	識別力	困難度	識別力	困難度	識別力	困難度
1	0.836	0.230	0.894	0.239	0.843	0.204
2	0.956	-0.756	1.009	-0.751	0.962	-0.711
3	0.696	-0.182	0.779	-0.170	0.766	-0.177
4	0.864	0.173	0.990	0.165	0.909	0.133
5	0.872	-0.388	0.950	-0.342	0.905	-0.338
6	0.769	-0.377	0.921	-0.392	0.885	-0.384
7	0.894	-0.314	1.035	-0.324	0.965	-0.324
8	0.879	-0.026	0.978	-0.078	0.918	-0.095
9	0.385	-0.140	0.399	-0.113	0.418	-0.120
10	0.639	0.167	0.717	0.151	0.697	0.123

Table10 質問紙の項目母数の推定値(周辺最尤推定)

項目	部分得点モデル		段階反応モデル (Multilog)			
	識別力	困難度	識別力	困難度 1	困難度 2	困難度 3
1	0.584	-0.611	1.390	-2.520	-0.800	1.100
2	0.643	-0.588	1.330	-2.280	-0.770	0.930
3	0.334	-0.372	0.790	-3.130	-0.440	2.280
4	0.733	0.215	2.440	-1.190	0.310	1.680
5	0.885	0.120	2.780	-1.060	0.170	1.430
6	0.672	-0.214	2.290	-1.830	-0.200	1.320
7	0.558	0.426	1.240	-1.230	0.510	2.310
8	0.932	-0.195	2.770	-1.410	-0.230	0.990
9	0.867	0.059	2.290	-1.190	0.130	1.310

尤推定を利用して段階反応モデルでは識別力が大きくなっていることが分る。

段階反応モデルにおいて少数項目では識別力の上方バイアスがかかることについては、慎重な検討を要するので、これについては別の研究で行うこととし、ここではTable 7の同時最尤推定の結果に基づいて検討を進める。Figure 2に項目1の部分得点モデルの項目反応曲線を、またFigure 3に段階反応モデルの反応曲線を示した。評定結果の合計点と両モデルの のケンドールの順位相関を求めたところ段階反応モデルとは0.9104、部分得点モデルとは0.9424の相関を得た(無回答の無いケース1588人のみで算出した。)部分得点モデルの方が、合計点との相関がやや高く、その再現には適していることはシミュレーションの結果と一致している。相関の値は段階反応データのシミュレーション結果から得られた両モデルの相関と近いものとなっているが、これだけで本データは段階反応モデルに従って発生していると決めるわけには行かないだろう。その理由としては、たとえばシミュレーションデータの作成の際に多次元性などの影響を加味すれば、真値と推定値の相関だけでなく正答数(質問紙の場合は合計点)との相関も低下すると容易に予想できることなどがあげられるからである。

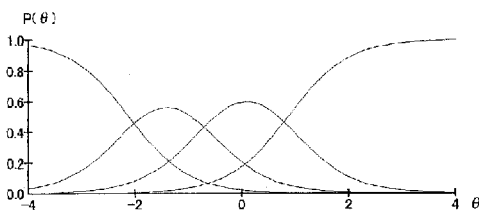


Figure2 項目1の段階反応曲線(段階反応モデル)

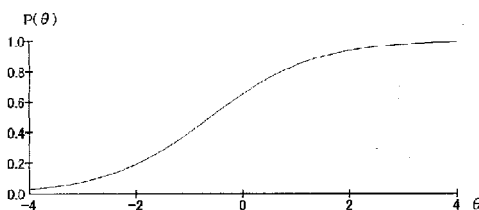


Figure3 項目1の項目反応曲線(部分得点モデル)

本節のデータは実際の調査であるため真値は不明であるからどちらのモデルが良いかを決定することは困難である。しかし段階反応モデルにおいて合計点との相関が低いことは、以下の様な問題を生じる。順位の逆転が十分説明がつくようなものであれば良いのであるが、一見で分るような性格のものではないため、モデルの利用者には逆転現象を納得しにくいということがおきる。いずれにしろシミュレーションの結果からは、(シミュレーションの条件に近い場合は)真の値の推定という点では大きな差は両モデルでないことが想像される。しかし、少数項目の場合の段階反応モデルの識別力推定値のバイアスがどのような影響をもたらすのかは、もう少し検討する必要があるだろう。いずれにしろ項目特性として識別力や項目困難度以外に大きな関心がない場合、つまり回答の各段階の困難度などに特別な関心がない場合(質問紙を利用した多くの研究はこれに該当する)には、複雑で多くの母数を推定しなければならない段階反応モデルより簡単な部分得点モデルの方が適していると言えるだろう。もちろん、個々の段階の反応にまで関心を持つ場合はこの限りではない。

文献

- Andrich, D. 1978 A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Birnbaum, A. 1968 Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds.), *Statistical theories of mental test scores* (pp.395-479) Reading, MA: Addison-Wesley.
- Bond, T.G. & Fox, C.M. 2001 *Applying the Rasch model: Fundamental measurement in the Human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- 木原孝博・田中治彦・藤森進 1997 中学生・高校生の学校不適応に関する研究 岡山県青少年基本調査(1994)に基づく分析 岡山大学教育学部研究集録, 104, 105-122.
- 藤森進 2001 項目反応理論における部分得点の処理について 日本教育心理学会第43回総会発

表論文集,394.

藤森進 2002a 項目反応理論におけるテストの部分得点の処理方法について 未発表論文.

藤森進 2002b 部分得点モデルとその応用 第1回心理測定研究会.

Masters,G.N. 1982 A Rasch model for partial credit scoring. *Psychometrika*, 47,149-174.

野口裕之 1999 適応型テストへの応用：CAT方式による識別性検査 渡辺直登・野口裕之編著「組織心理測定論」第8章 白桃書房.

坂元慶行・石黒真木夫・北川源四郎 情報量統計学 共立出版.

Samejima,F. 1969 Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, No.17.

Thissen, D 1991 *Multilog user's guide*. Chicago,IL:Scientific Software.