

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成25年5月30日現在

機関番号：12608
 研究種目：挑戦的萌芽研究
 研究期間：2012～2012
 課題番号：24650548
 研究課題名（和文） 項目反応理論に基づく能力値の簡易推定法に関する研究
 研究課題名（英文） Quick method for ability estimation based on item response theory

研究代表者
 前川 真一 (Maekawa Shinichi)
 東京工業大学・大学院社会理工学研究科・教授
 研究者番号：70190288

研究成果の概要（和文）：

項目反応理論（IRT）には、テスト項目が異なっても得点が比較可能であるという利点があるため、多くのテストで用いられている。しかし、項目反応理論に基づく得点（能力値）の推定は簡単ではない。本研究では、最適重みをより一般的な項目反応モデルへ拡張した。そして、これを能力値の簡易推定法に応用し、その有効性を、シミュレーションデータと実データを用いて示した。

研究成果の概要（英文）：

Item response theory (IRT) is used in many testing applications, since this provides major advantage: comparability of test scores from different test forms. However, IRT scaled scores are complicate for students to calculate. In this study, we extended globally optimal scoring weights to more general IRT models. We also applied these weights to a quick method for ability estimation and showed the efficiencies using simulated data and real data.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
交付決定額	1,100,000	330,000	1,430,000

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学・教育工学

キーワード：教育学習支援システム・テスト理論

1. 研究開始当初の背景

項目反応理論(Item Response Theory)は、テスト理論の一つであり、SAT や PISA (OECD 生徒の学習到達度調査) など、欧米で開発された大規模試験の多くで用いられている。その理由は、項目反応理論を用いて作成されたテストの成績（通常、 θ という能力値尺度を用いて表示される）が、その算出に用いられたテスト項目が異なっても比較可能であるという利点を持つという性質、つまり、 θ 尺度の項目に対する非依存性にある。 θ 尺度の項目非依存性により、異なる項目からなるテスト冊子を多数用意して

試験を行うことが可能となり、テストのセキュリティを高めることが出来る。また、同じ個人に同じ試験を複数回受験させることが可能となり、個人の能力の伸びを測定することが出来る。

しかし、項目反応理論は教室や個人の単位ではほとんど利用されていない。その理由として、項目反応理論においては、受験者の能力値 θ は通常、最尤推定値(MLE)や事後平均(EAP)として推定されるため、受験者や教師などの個人が計算するには難しいということが挙げられる。また、素点や配点付き得点を重視する我が国のテスト文化においては、その計算過程が単純ではない θ 尺度の利用

をはばかる傾向がある。

この解決策として、日本で広く用いられる点数の付け方、すなわち、正答した項目の配点を足し合わせた合計点（重み付き合計点）を利用して、項目反応理論に基づく能力値を推定する方法が提案された（荒井・Chen・前川, 2008）。そこでは、まず、 θ を与えた時の重み付き得点の条件付き分布を計算し、その後、重み付き得点を与えた時の θ の事後平均を計算することで θ の推定を行っているが、反応モデルが二値反応モデルに限定されており、多値反応モデルを取り扱っていない。また、重みの選択に際しても、十分統計量を持たないモデルに対しては理論的に優れた方法が用いられているとは言い難い。

2. 研究の目的

(1) 問題項目に割り当てる重み（配点）について

問題項目に割り当てる重み（配点）については、Mayekawa(2008)が提案した最適重みを応用する。本研究では、この最適重みを多値反応モデル（部分採点モデル、一般化部分採点モデル、名義反応モデル）に拡張することを目的とする。

(2) 重み付き合計点から能力値を推定について

重み付き合計点から能力値を推定する際には、多項分布に従う独立な確率変数の和の分布の計算を必要とする。この計算については、Mayekawa & Arai (2008)が効率的な計算アルゴリズムを示している。本研究では、この計算アルゴリズムを利用して重み付き得点を与えた時の θ の事後平均の計算を行う。この計算について、(1)で示したモデルに拡張することを目的とする。

3. 研究の方法

(1) 問題項目に割り当てる重みについて

まず、Mayekawa(2008)が提案した最適重みを多値反応モデル（部分採点モデル、一般化部分採点モデル、名義反応モデル）に拡張する。

続いて、拡張したモデルのそれぞれに対して、重みの計算を実行するためのRプログラムを開発する。二値型項目反応モデルのためのRプログラムは、Arai & Mayekawa(2010)によって提示されているため、それを応用する。

最後に、拡張した最適重みが、問題項目の配点として一般的によく用いられている他の重みに比べて、テスト受験者の能力をより良く推定するかどうか、重みの有効性について検証する。

(2) 重み付き合計点から能力値を推定について

まず、Mayekawa & Arai (2008)で示された計算アルゴリズムを利用して重み付き得点の条件付き分布の計算方法を、(1)で拡張したモデルに拡張する。続いて、拡張したアルゴリズムをRで実行するためのプログラムの開発を行う。

4. 研究成果

研究方法に述べたように、最適重みおよび重み付き得点から能力値を推定する方法を多値反応モデルに対して拡張し、それを実行するRプログラムを開発した。

このRプログラムを用いて、(1)シミュレーション研究および(2)実データを用いた研究を行った。

(1) シミュレーション研究

表1のような二値反応モデルと多値反応モデルが混在した6項目を想定し、シミュレーションを行った。

表1 シミュレーションに用いた項目

項目	モデル
1	2パラメタロジスティックモデル
2	3パラメタロジスティックモデル
3	段階反応モデル
4	段階反応モデル
5	名義反応モデル
6	名義反応モデル

これらに対して項目パラメタを指定し、最適重みを計算したところ、表2のような重みが得られた。実際の配点に近づくため、この重みを四捨五入し、一桁の整数としたのが表3である。表2、表3でのタイトル行にある「0, 1, 2, 3」は、重み付けを考えない場合に通常用いる配点を示している。たとえば、項目3は3カテゴリの項目であるが、重みを考えない場合には、カテゴリ1を答えたら0点、カテゴリ2を答えたら1点、カテゴリ3を答えたら2点を与えるという意味である。

表2 最適重みの計算結果

	0	1	2	3
item1	0	4.404		
item2	0	3.861		
item3	0	3.861	8.268	
item4	0	3.345	6.552	9.849
item5	0	2.649	5.262	
item6	0	2.658	5.286	6.582

表3 最適重み (四捨五入後)

	0	1	2	3
item1	0	4		
item2	0	4		
item3	0	4	8	
item4	0	3	6	9
item5	0	3	5	
item6	0	3	5	6

最適重みの有効性を検証するために、表3の重みを用いた場合 (GO weight) と、重みを考えない場合 (No weight) とで、重み付き合計点から能力値を推定し、そのときの推定値 (事後平均; EAP) と事後標準偏差 (PSD) を求め、比較を行った (図1)。なお、表3の重みを用いた場合の満点が $4+4+8+9+5+6=36$ 点であるのに対し、重みを考えない場合の満点は $1+1+2+3+2+3=12$ 点であるので、図示する際に、重みを考えない場合の得点を3倍して表示した。

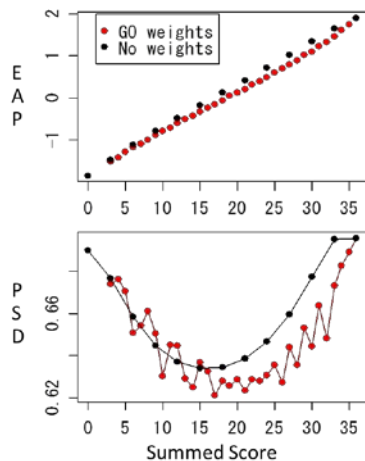


図1 合計点に対する、能力値の推定値と事後標準偏差

図1上から、能力値の推定値は、GO weights がわずかに低めであるものの、両重みで近い値になることが分かった。また、図1下において事後標準偏差はGO weights の方が特に点数の高いところで低いことから、

GO weights の方が推定精度が高く、その傾向は点数の高いところで強く出ることが分かった。

(2) 実データを用いた研究

実データとして、ある年の法科大学院適性試験第1部の問題を利用した。項目数は18であり、二値項目と多値項目から構成される。反応モデルとしては段階反応モデルを仮定した。これらの項目に対して、重みを考えない場合 (No weights)、もともとの配点 (Built-in weights) および最適重み (GO weights) の3種類の重みを用い、結果を比較した。

図2は、3種類の重みを用いたときのテスト情報量を比較したものである。横軸のThetaは能力値 θ であり、 θ に対してテスト情報を描いたものである。図から、最適重みを用いた場合は他の二つの重みに比べてテスト情報量が高いこと、重みを考えない場合ともともとの配点は近い値であることが分かる。テスト情報量が高いほど、その θ での推定精度が高いことを意味することから、最適重みを用いることで、推定精度を高めることができることが示された。

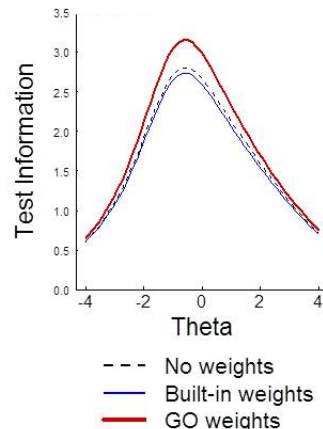


図2 テスト情報関数の比較

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計0件)

[学会発表] (計2件)

- ① Sayaka Arai and Shin-ichi Mayekawa.
Globally optimal scoring weights and their application for test score equating in IRT model. 5th International Conference of the ERCIM Working Group on Computing & Statistics. Oviedo, Spain.
2012年12月2日

- ② Sayaka Arai and Shin-ichi Mayekawa.
Globally optimal scoring weights and their application in item response theory models. Analysis and Modeling of Complex Data in Behavioural and Social Sciences. Capri, Italy.
2012年9月4日.

[図書] (計0件)

[産業財産権]

○出願状況 (計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況 (計0件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

前川 眞一 (MAEKAWA SHINICHI)
東京工業大学・大学院社会理工学研究科・
教授
研究者番号：70190288

(2) 研究分担者

荒井 清佳 (ARAI SAYAKA)
独立行政法人大学入試センター・その他部
局等・助教
研究者番号：00561036