

項目反応理論

2021 年 10 月 6 日

1 項目反応理論による評価を加味した数学テストと e-learning システムへの実装の試み

今回は教職教養の問題を題材に項目特性図を用いて、テスト分析を行った。その最後で数学のテストへの適用に関心を示した。そこで今回は、九州工業大学の広瀬英雄教授や月原由紀、鈴木敬一らの論文をもとに、実際に数学のテストで応用した例を紹介する。

1.1 はじめに

項目反応理論は、英語能力判定として TOEFL や TOEIC テストに使われ定着している。大学数学などでは、正解不正解といった 2 値応答の問題を作りにくいこともあってか、IRT を積極的に用いてテスト分析に応用した実例は少ない。

1.2 IRT(Item Response Theory) とは

従来の古典的テスト評価法は、単に問題に配点を与えておき、各問題における得点を合計したもので総合評価が与えられるので、個々の問題の特性や受験者の能力などは加味されていないといえる。IRT は問題の特性と学生の能力を同時に取り扱うことができる総合的な評価になっている。従って、古典的評価法では同点であっても、難問に解答した割合が多いほど評価値は高くなる傾向がある。

1.2.1 IRT によるもでるについて

今回は 2 母数ロジスティックモデルを用いて考える。学生 i が能力 θ_i を持ち、テストの j 番目の項目に正解できる確率 $P_j(\theta_i)$ を

$$P_j(\theta_i) = \frac{1}{1 + \exp -1.7a_j(\theta_i - b_j)} \quad (1)$$

と表すことができた。ここで a_j は問題の識別力、 b_j は問題の困難度とする。この $P_j(\theta_i)$ を縦軸に、横軸を能力 θ としたグラフを ICC(項目特性曲線) と呼ぶ。

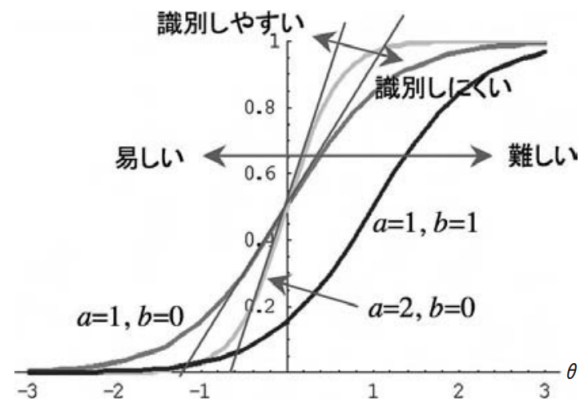


図 1: ICC の見方

1.3 項目パラメータの推定

個々の学生のパラメータが未知のままでは、サンプルが増えるだけパラメータも増えることになる。

1.4 IRT 評価と素点評価の比較

テストは被験者が紙に答案を書く一般的な形式で行われている。次の表はある学科での正答、誤答の反応パターンでありグラフは推定された項目特性曲線である。



図 2: 反応パターン

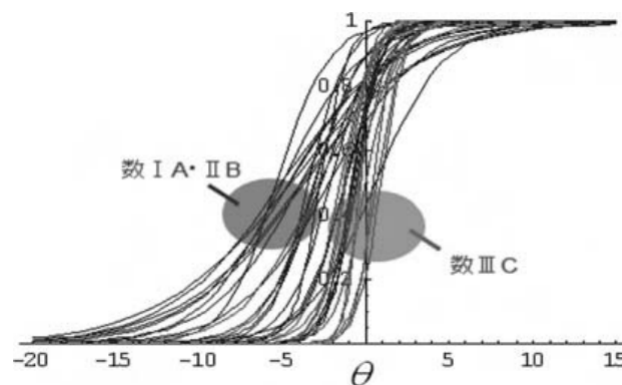


図 3: ICC

図3、図4より数IIIの難易度は高く、問題の識別力も高いことが示されている。また、以下の図により、同じ素点から得られる能力値の幅が観測されるべきがわかる。

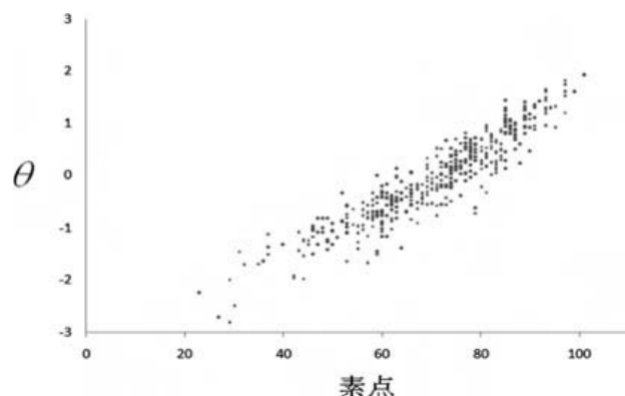


図 4: ICC

例えば、素点 60 点では、 θ は 0 から -1.5 までばらついている。逆に $\theta = 0$ でも素点は 60 点から 80 点までばらついている。このことから、能力と点数が逆転することもあると考えられる。1 回だけのテストでは決定的な判断をするのが難しくなる。誤差が確率的にどのくらいの割合で観察されるのかを検証しておく必要がある。

そこで次は真の能力値 $\hat{\theta}$ を持った学生が今回のテストを受けたとして、IRT での評価と従来の素点評価とが真の能力値にどのように反映されるのかを考える。