# 数式自動採点システムの解答データへの多段階項目反応理論の適用の試み

伊藤 可子[*1]，中村 泰之[*2]
Kako Ito[*1]，Yasuyuki Nakamura[*2]
[*1] 名古屋大学大学院情報学研究科
[*1,*2] Graduate School of Informatics, Nagoya University
Email: [*1]itou.kako.w9@s.mail.nagoya-u.ac.jp

**あらまし**：多段階項目反応理論は,正答と誤答の二値で構成されたデータに適用する二値項目反応理論とは異なり,誤答をいくつかのカテゴリに分類し,多段階の値で構成されたデータに対して項目反応理論を適用するものである.本研究では,数式自動採点システム STACK で行ったオンラインテストの解答データに多段階項目反応理論の適用を行い,項目推定量について,二値項目反応理論との比較を行った.
**キーワード**：STACK，項目反応理論

## 1. はじめに

近年，教育の ICT 化が進み e ラーニングに注目が集まっている．数学のオンラインテスト教材の解答形式の多くは多肢選択式で, 数式入力形式は少ない. しかし，多肢選択式や数値入力式の場合，選択肢から解答を推測できる場合があり，個々の問題に対する学生の実際の能力を測ることが難しい．一方，数式入力形式の場合，解答が推測できないため，学生の実際の能力を測ることができると考えられる．また，数式入力形式は誤答分析により，受験者の理解不足箇所の情報が得られることも期待できる.

本研究は STACK の解答データに多段階項目反応理論を適用した. STACK は数式入力形式の問題の出題，正誤評価が可能であり,部分点を用いた採点を擬似的に行うことが可能である．オンラインテストの受験した学生の実際の解答データへ多段階項目反応理論の適用を行った.

## 2. STACK について

STACK(1)は英国バーミンガム大学の Christopher Sangwin 氏らが開発した数学オンラインテストシステムである．学習管理システムの一つである Moodle の小テストで数式入力の解答が可能な問題タイプとして提供され，入力された数式を数式処理システムによって代数的に等価かを評価し，正誤評価，自動採点を行う．また，ポテンシャル・レスポンス・ツリーと呼ばれる機構を用いて学習者の解答に応じたフィードバックを返すことができる.

## 3. 解析手法

### 3.1 IRT(項目反応理論)

項目反応理論(2)(Item Response Theory 以下 IRT)は 1952 年米国の Lord が提出した現代テスト理論である．項目は試験の各問題を示し，反応は項目に対する正誤状況を示す.

IRT はある項目の困難度，識別力が判明している場合，項目に対する反応から測定できる受験者の能力を推定する．項目の識別力は受験者の能力推定の正確さを示し，受験者の能力が正確に推定される項目ほど値が大きく，項目の困難度は，正答に受験者の能力値が高い必要があるほど値が大きくなる.

IRT は学生の能力値の推定を目標としているが，直接能力値を求めることは不可能であるため，能力値の推定に必要な項目の識別力と困難度を先に推定する．この時，周辺最尤推定法を用い，項目の識別力と困難度を推定する．周辺最尤推定法は能力値パラメタを積分によって消去し，周辺尤度関数が最大となるように能力値に直接依存しない識別力と困難度を推定する手法である.

項目の識別力と困難度を推定した後，学生の能力値を推定する．学生の能力値の推定にはベイズ推定法を用いる. 能力値の事前確率分布を正規分布とし，尤度関数を用いて事後分布を求め，能力値に関する推定を行う.

また，IRT では横軸を能力値，縦軸を項目に正答する確率とした項目特性曲線(Item Characteristic Curve 以下 ICC)により，項目の難易度と受験者の能力を分離して表す．ICC は正規累積モデルで表されていたが，積分を含むため，Birnbaum がロジスティック分布を利用したロジスティックモデルを提案した.

2 パラメタ・ロジスティックモデルは項目$j$の識別力を$a_j$，困難度を$b_j$とした時，能力値$\theta$の学生が項目$j$に正答する確率$p_j(\theta)$を表す.

$$p_j(\theta) = \frac{1}{1 + \exp\left(-Da_j(\theta - b_j)\right)}$$

$D$は Lord と Birnbanm が導入した$D = 1.7$で全ての$\theta$において誤差が 0.01 以下となる尺度因子を表す. 項目特性曲線の変曲点での傾きが識別力によって表現される.

### 3.2 多値型 IRT モデル

多値型 IRT モデルは受験者の反応が段階によって三つ以上のカテゴリに表される場合に適用される.

代表的な多値型 IRT モデルの一つに 1969 年に鮫島が提案した段階反応モデル(3)(Graded Response Model 以下 GRM)がある.

カテゴリ数を K とし,カテゴリを 1 から K までの整数で表した場合にある項目$i$のカテゴリは$X_i = 1, \cdots, K$となる.この時,あるカテゴリ$k$の選択確率を以下のようにおく.

$$P\left(X_i = k \middle| \theta_p, a_i, b_i, C_1, \cdots, C_{K-1}\right) = P_{k-1}\left(\theta_p, a_i, b_i, C_1, \cdots, C_{K-1}\right) - P_k\left(\theta_p, a_i, b_i, C_1, \cdots, C_{K-1}\right)$$

$$P_k(\theta, a_i, b_i, C_1, \cdots, C_{K-1}) = \frac{1}{1 + \exp\left(-1.7a_i(\theta - b_i - C_k)\right)}$$

$a_i$と$b_i$は項目特性を表すパラメータであり,$\theta$は学生の能力値を表すパラメータである.また,$C_k$は各カテゴリの境界となる値のパラメータであり,項目 1 を基準として設定する.

今回の計算では,パラメータの同定のために$b_1 = 0$として計算されるが,モデル式中で$b_i + C_k$となっているため,任意の値$d$を用い,$b_i + d$や$C_k + d$の 2 式により,$b_i$や$C_k$の値を任意に変更することができる.

## 4. 適用結果

全 5 回のオンラインテストの問題に上記の理論を適用した. 部分点を利用してカテゴリ分けを行うにあたり,明確な採点基準を設けるために,対象とする問題を全て不定積分の問題とし,以下のようにカテゴリ分けを行った.二値項目反応理論は正答を 1,誤答を 0 としたデータに適用した.多段階項目反応理論は正答を 5,積分定数忘れを 4,誤答を 1 とした 5 段階のカテゴリに分類したデータに適用した.このカテゴリ分けにおいて,2 と 3 の値を取る項目は存在しないが,正答,積分定数忘れの解答に重みづけをすると言う観点からこのようなカテゴリ分けとした.今後,誤答の種類に応じてさらに細かく分類し,段階数を増やすことが可能である.

また,今回は$b_i$や$C_k$の値を変更するにあたって,正答,積分定数忘れのカテゴリと誤答のカテゴリを分けるために,$C_3$の値を用いることとし,$b_k$を$b_k = b_k + C_3$の式で算出された値に変更した.

図 1,2 はそれぞれ二値 IRT,多段階 IRT の結果である.これらの図より,多段階 IRT の項目特性曲線は二値 IRT のものよりも変曲点付近での傾きが緩やかになり,全体的に識別力が低くなっていることが窺える. このことから,積分定数忘れを誤答に含めず,一つのカテゴリとして計算を行うことで,学生の入力し忘れなどが計算結果に影響し,その結果二値データで計算した場合よりも識別力が低くなったのではないかと考えられる.

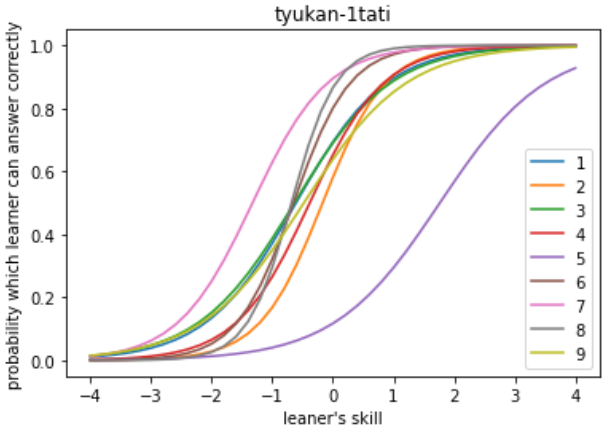


図 1 二値での項目特性曲線

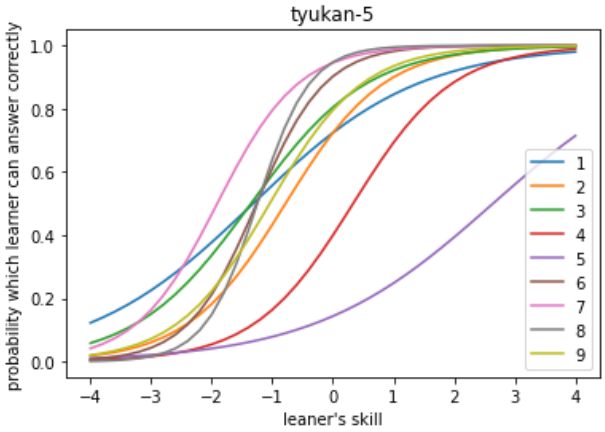


図 2　多段階での項目特性曲線

## 5. まとめ

正答,誤答といった二値データではなく,部分点を利用してカテゴリ分けを行うことで,二値データで行われた計算結果よりも全体的に識別力が低くなっていることがわかった.今回はカテゴリ分けの容易さの観点から不定積分の問題のみを対象としたが,今後はカテゴリ分けの条件の再検討や,不定積分以外の問題へ適用する際のカテゴリ分けの条件を検討したい.

### 参考文献

(1) STACK, https://www.ed.ac.uk/maths/stack, 参照日 2021 年 2 月 8 日
(2) 豊田秀樹,「項目反応理論[入門編] 第 2 版」, 朝倉書店, (2012)
(3) 岡本安晴,「Python プログラミング備忘録」, http://y-okamoto-psy1949.la.coocan.jp/Python/misc/IRTpol/, 参照日 2022 年 2 月 7 日