

項目反応理論を用いた設問解答率分析図の評価

菊地 賢一*

要 約

大学入試センターでは、問題作成等に役立てるために試験問題やその解答の評価を行っているが、そのための分析手法の一つとして設問解答率分析図がある。ある設問の設問解答率分析図は、受験者をテスト得点の順位により5群に分け各群における項目の解答率を求めて、横軸に5つの群、縦軸に各群における解答率をとり、プロットしたものである。正答率(正答選択肢の解答率)だけではなく、誤答選択肢の解答率についてもプロットすることにより、誤答選択肢の評価を行うことも可能である。

日本においては、項目反応理論における学力の特性値を用いて入試を行うことは少なく、得点を用いて入学者の選抜が行われることが多い。それゆえ受験者の学力の評価は得点そのものを用いて行われており、こういった観点からこの設問解答率分析図における横軸は受験者の学力を示しているとみなすことができる。しかしながら、この手法の欠点は得点分布を用いて評価を行っているため、その結果が得点分布の違いの影響を受けると思われる点である。

本論文では、項目反応理論を用いて、得点分布の違い(項目の組み合わせの違い)が設問解答率分析図にどの程度影響を与えるのかを評価した。その結果、得点分布が違っていても設問解答率分析図はそれほど影響を受けないことがわかった。分析のための計算の簡便性と得点分布によらない安定性を考えると、この分析方法は項目の評価のために十分に利用できると思われる。

キーワード：設問解答率分析図、項目反応理論

1 はじめに

大学入試センターでは、問題作成等に役立てるために試験問題やその解答の評価を行っているが、そのための分析手法の一つとして、各設問(項目)に対して以下のような解析を行っている。

1. 受験者をその科目の得点の順に5群に分ける。
2. それぞれの群における設問の解答率(正答率)を求める。
3. 横軸に5つの群、縦軸に各群における解答率をとり、プロットする。

大学入試センターでは、このような図を設問解答率分析図と呼び、共通第1次学力試験の初年度から分析(清水、1983等参照)を行っている。図1に設問解答率分析図の例を示した。正答率(正答選択肢の解答率)だけではなく、誤答選択肢の解答率についてもプロットすることにより、誤答選択肢の評価を行うことも可能である。また、設問単位だけではなく、いくつかの関連する設問の集まりである大問に関しても同様の分析を行っている。この分析方法は、多肢選択問題の評価のためにETS(Educational Testing Service)で古くから行われている方法(ETS、1963)を改良したものと思われ、テスト得点と項目の正答率との関係を求

* 大学入試センター 研究開発部 情報処理研究部門

〒153-8501 東京都目黒区駒場 2-19-23 e-mail: kikuchi@rd.dnc.ac.jp Telefax: 03-5478-1297

めているという観点では、item-test regressions (Lord, 1965, 1980, p27-30)とも関係がある。

日本においては、項目反応理論における学力の特性値を用いて入試を行うことは少なく、得点を用いて入学者の選抜が行われることが多い。それゆえ受験者の学力の評価は得点そのものを用いて行われており、こういった観点からこの設問解答率分析図における横軸は受験者の学力を示しているとみなすことができる。しかしながら、この手法の欠点は得点分布を用いて評価を行っているため、その結果が得点分布の違いの影響を受けると思われる点である。すなわち、ある設問について評価する場合に、他の設問との組み合わせによって評価結果が変わってしまうということである。ここで、設問解答率分析図において何がプロットされているのか考えてみよう。この図では受験者数を考えて便宜上5群に分けて正答率を計算しプロットしているが、これは受験者を得点で順位付けた場合のそれぞれ20パーセント毎の正答率を計算していることになる。そこで、受験者が無限におり分割も無限に行えると考えると、横軸はテスト得点分布の累積確率、縦軸はその累積確率に対応するテスト得点で条件付けた時の項目の正答率と解釈できる。ここで、このように横軸にテスト得点分布の累積確率、縦軸に項目正答率をとった図を項目正答率分析図と名付ける。本論文では、項目反応理論を用いて、得点分布の違い(項目の組み合わせの違い)が項目正答率分析図にどの程度影響を与えるのかを評価する。

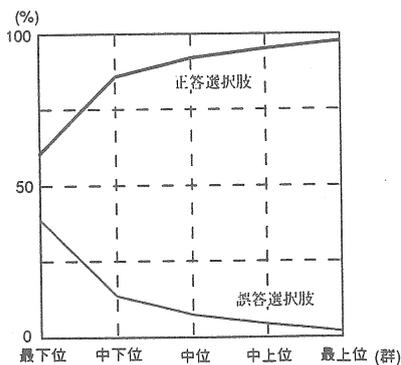


図1 設問解答率分析図

2 項目反応理論を用いた項目正答率分析図のモデル化

テストで測定しようとしている学力(特性)を、特性値 θ として表すことにする。この θ は、直接観測することができない潜在変数である。項目反応理論(芝, 1991 等参照)では、各項目に正答する確率が θ の関数として表され、この関数は項目特性関数と呼ばれる。この関数は通常、 θ に関して単調増加し、0 以上 1 以下の値をとる。また、 θ を固定した時の各項目の反応はそれぞれ独立であること(局所独立)を仮定する。ここで、

n : 項目数

U_j : j 番目の項目に対する反応

$$U_j = \begin{cases} 1 \cdots \text{正答} \\ 0 \cdots \text{誤答} \end{cases}$$

X : テスト得点

$$X = U_1 + U_2 + \cdots + U_n, \quad 0 \leq X \leq n$$

θ : 被験者の特性値

$P_j(\theta)$: j 番目の項目に関する項目特性関数

とする。ただし、特性値 θ の分布を $f(\theta)$ とする。また、3節で行ったシミュレーションでは、 $f(\theta)$ を標準正規分布の密度関数、 $P_j(\theta)$ を式(3)に示した2パラメータ・ロジスティックモデルの項目特性関数としている。なお、局所独立の仮定により、 θ を与えた時の条件付きの項目反応 U_j はそれぞれ独立である。以下では、このような項目反応理論の仮定が成り立つものとして、項目正答率分析図を解析的に求める。

まず、 θ を固定した時の X の分布を $g(x|\theta)$ とすると、 $g(x|\theta)$ は平均が $\sum_{j=1}^n P_j(\theta)$ の複合2項分布となることは良く知られている。例えば、 $g(1|\theta)$ は、

$$\begin{aligned} g(1|\theta) &= \Pr(X = 1|\theta) \\ &= P_1(\theta)(1 - P_2(\theta)) \cdots (1 - P_n(\theta)) + \cdots \\ &\quad + (1 - P_1(\theta)) \cdots (1 - P_{n-1}(\theta))P_n(\theta) \end{aligned}$$

と表される。ここで、 X の周辺分布を $h(x)$ とすると、

$$h(x) = \int_{-\infty}^{\infty} g(x|\theta)f(\theta)d\theta$$

となる。

次に、得点 X を固定した時の特性値 θ の条件付き分布を $s(\theta|x)$ とし、これを求める。得点 X と θ の同時分布は、 $g(x|\theta)$ を用いると $g(x|\theta)f(\theta)$ と書ける。一方、 $s(\theta|x)$ を用いると $s(\theta|x)h(x)$ である。よって、 $g(x|\theta)f(\theta) = s(\theta|x)h(x)$ であり、 $s(\theta|x)$ は、

$$s(\theta|x) = \frac{g(x|\theta)f(\theta)}{h(x)}$$

と表される。

項目正答率分析図にもどって考えてみると、 j 番目の項目の項目正答率分析図の y 軸の値は、 x 軸の値である累積確率に対応する得点 X を与えた時の U_j の条件付き平均と考えることができる。得点 X を固定した時の U_j の条件付き平均は、

$$E(U_j|X=x) = \int_{-\infty}^{\infty} P_j(\theta)s(\theta|x)d\theta \quad (1)$$

となり、項目正答率分析図は $x=0, \dots, n$ について、

$$(Pr(X \leq x), E(U_j|X=x))$$

をプロットしたものとなる。

最後に、本来の設問解答率分析図のように被験者の集団を 5 群に分けた場合を考える。得点分布 $h(x)$ の 5 分位点を求め、その中のある区間を $(\alpha, \beta]$ とする。ここで設問解答率分析図は、得点 X がこの区間の値をとる時の U_j の条件付き平均を求めればよい。よって、式 (1) に期待値の繰り返しの公式 (竹村、1991、p52 等参照) を適用して、

$$\begin{aligned} E(U_j|\alpha < X \leq \beta) &= \frac{\int_{\alpha}^{\beta} E(U_j|X=x)h(x)dx}{\int_{\alpha}^{\beta} h(x)dx} \\ &= \frac{\int_{\alpha}^{\beta} \left(\int_{-\infty}^{\infty} P_j(\theta)s(\theta|x)d\theta \right) h(x)dx}{0.2} \quad (2) \end{aligned}$$

を求めればプロットできる。ただし、ここでは便宜上得点分布は連続分布としたが、通常の離散分

布の場合は上記の積分を和として計算すればよい。(計算例は、付録参照。)

3 シミュレーション

3.1 方法

- 項目数 n
センター試験の項目数を考えて $n=30$ とした。
- 特性値 θ
 $\theta \sim N(0,1)$ とする。ただし、数値計算を行うため θ を区間幅 0.2 で区切り、区間の中心値が $-4, -3.8, \dots, 4$ となるような 41 の区間を用いて、計算を行った。
- 項目反応モデル
今回のシミュレーションにおいては、2パラメータ・ロジスティックモデル

$$P_j(\theta) = \frac{1}{1 + \exp(-1.7a_j(\theta - b_j))} \quad (3)$$

を用いた。なお、 a_j は識別力、 b_j は困難度を表すパラメータである。

3.2 結果

様々なケースについて数値計算を行ったが、代表的なケースとして以下のようなパラメータの結果を示す。

- Case 1
 $a_j = 0.5, b_j = 0, j = 1, \dots, 30$
- Case 2
 $\begin{cases} a_1 = 0.5, b_1 = 0 \\ a_j = \text{区間 } [0.3, 1.0] \text{ の一様乱数, } b_j = \text{区間 } [-3, -1] \text{ の一様乱数, } j = 2, \dots, 30 \end{cases}$
- Case 3
 $\begin{cases} a_1 = 0.5, b_1 = 0 \\ a_j = \text{区間 } [0.3, 1.0] \text{ の一様乱数, } b_j = \text{区間 } [1, 3] \text{ の一様乱数, } j = 2, \dots, 30 \end{cases}$

このように 1 番目の項目以外のパラメータを変化させ、1 番目の項目について項目正答率分析図の数値計算を行った結果を図 2 ~ 4 に示す。図 2 にはそれぞれのケースの得点分布を示した。項目

の難易度にあたるパラメータ b_j を大きく変化させているため、得点分布の中心は3つのケースでかなりずれる。図3には、1番目の項目に関して横軸に得点、縦軸に得点で条件付けた場合の条件付き項目正答率をプロットしたもの (item-test regressions) を示した。このように横軸に得点そのものをとってしまうと得点分布の偏りに影響されて、分析結果もかなり変化してしまう。図4には、図3の横軸を得点ではなくその得点に関する累積確率としたものをプロットした。この図は項目正答率分析図にあたるものだが、得点分布の偏りの影響をあまり受けない。他に様々なケースについても計算を行ったが同様の結果であった。なお、識別力パラメータの変化に比べて困難度パラメータの変化の方が、項目正答率分析図に、より影響を与える傾向が見られた。これは、困難度パラメータを変化させた方が、その得点分布に、より大きな違いが見られるためと考えられる。また、特性値 θ の分布として正規分布ではなくベータ分布を用いてシミュレーションを行ったが、この場合も同様に得点分布の偏りの影響をあまり受けなかった。

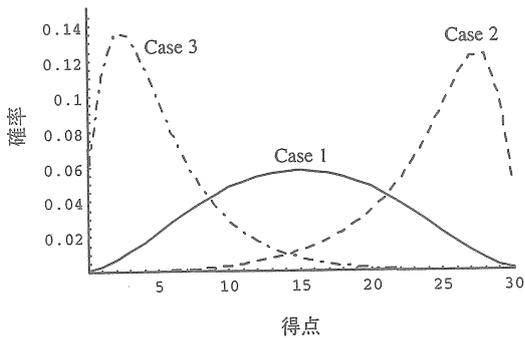


図2 得点分布

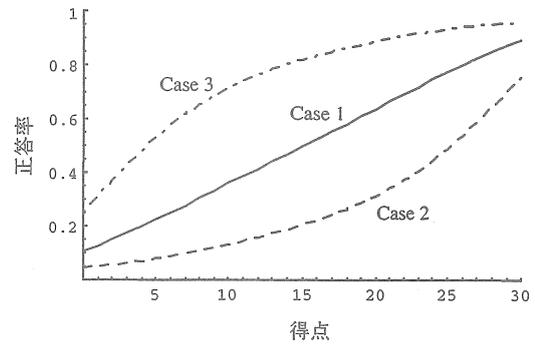


図3 得点 - 正答率 (Item-Test Regressions)

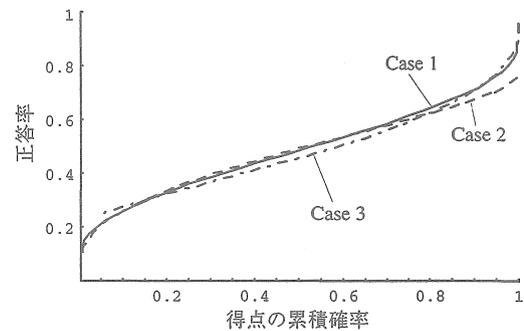


図4 累積確率 - 正答率 (項目正答率分析図)

4 まとめと考察

本論文では、項目正答率分析図について評価を行った。この方法は、得点分布を用いた方法であるので得点分布の違い、すなわち評価すべき項目以外の他の項目の組み合わせによって解析結果が変化することが予想された。しかしながら、項目反応理論の仮定を用いて理論的に数式化し、項目正答率分析図を数値計算により求めた結果、得点分布が違って項目正答率分析図はそれほど影響を受けないことがわかった。

この理由は、得点分布が異なるように項目パラメータを変えても、得点分布におけるある得点の相対的な順位と、その得点で条件付けた場合の特性値 θ の条件付き分布の位置の、 θ の周辺分布に対する相対的な順位があまり変わらないためであろう。例えば、得点分布で見た場合に上位 5% にあたる得点をとった受験者の特性値は、特性値

の分布からみた場合にも上位5%付近にあることが多いということである。得点における順位で条件付けた特性値の条件付き分布の中心の相対的な順位が大きく変動しなければ、項目正答率分析図もあまり変化しない。このように考えると、項目正答率分析図は、特性値の分布を一様分布に変換して、項目反応曲線を描いた場合に近いものとも、考えることができる。このような考え方は、石塚・山田(1987)でも議論されている。また、特性値の分布をいろいろな形に変換した場合の項目反応曲線の変化に関する議論は、Ramsay(1996)で行われている。

3節では、実際に行われているセンター試験の項目数を考慮して $n=30$ として、シミュレーションを行ったが、項目数が少ない場合はどうか。項目数を極端に減らし $n=5$ として、3節に示したものと同一パラメータでシミュレーションを行った結果を図5～7に示す。

このように項目数が極端に少ない場合でも、横軸に得点そのものをとった item-test regressions に比べると、得点分布の違いの影響が少ないことがわかる。他にも様々なケースについて計算を行ったが、極端に得点分布が偏ってしまう場合を除き同様の結果となった。このような項目数が少ないような場合は、設問解答率分析図の手法を大問に対して行う分析手法である大問得点率分析図に近いものと考えられるが、この結果から大問得点率分析図も得点分布の違いの影響をそれほど受けない分析方法であると思われる。

テストを行う側にとって行ったテストの項目を評価することは以後の作題に参考にできる等の理由で、非常に重要な課題である。評価の方法として項目反応理論を用いる場合はパラメータ推定を行う必要があり、そのための簡易推定の方法もあるが、通常は専用のソフトウェアパッケージを用いることが多い。しかし、本論文で評価した項目正答率分析図の考え方であれば、受験者を得点の順位によって何群かに分け、その群における項目の正答率を求めるだけで項目の評価を簡単に行うことができる。また、日本では学力の評価のために得点そのものを用いることが多い。この点から、潜在変数である特性値よりも、得点そのものを用いた分析方法の方がより解釈がしやすい場合もある。このような利便性と得点分布に

よらない安定性を考えると、この分析方法は項目の評価のために十分に利用できると考える。

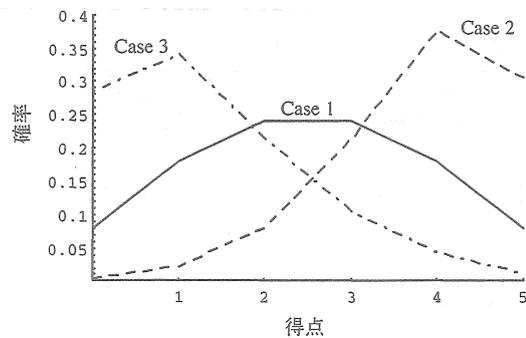


図5 得点分布 ($n=5$)

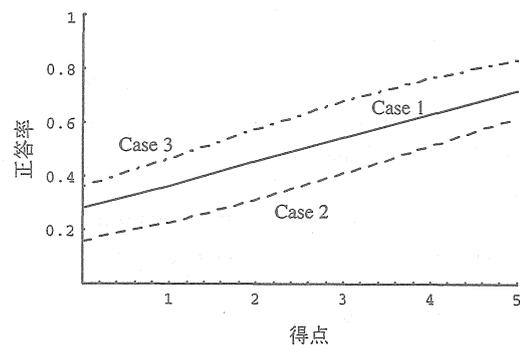


図6 Item-Test Regressions ($n=5$)

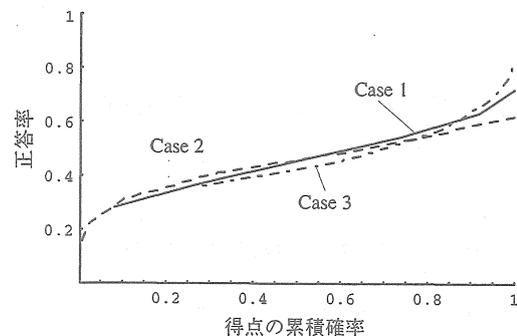


図7 項目正答率分析図 ($n=5$)

謝 辞

まず、本稿を執筆するきっかけとなった第7回IRTセミナーにおける発表の機会を与えてくださった東京大学の繁榊算男教授に感謝申し上げます。また、原稿の執筆段階では研究開発部の清水留三郎教授と前川眞一助教授に大変有益なコメントをいただいた。記して感謝申し上げます。最後に、建設的なコメントをいただいた査読者の方々にも、感謝申し上げます。

参考文献

- ETS (1963). *Multiple-choice questions: a close look*. Educational Testing Service.
- Lord, F.M. (1965). An empirical study of item-test regression. *Psychometrika*, **30**, 3, 373-376.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.
- Ramsay, J.O. (1996). A geometrical approach to item response theory. *Behaviormetrika*, **23**, 1, 3-16.
- 石塚智一・山田文康 (1987). 共通第1次学力試験の試験問題の分析—昭和57年度「化学I」を中心として—. 大学入試センター研究紀要, **16**, 1-51.
- 芝祐順 (編) (1991). 項目反応理論—基礎と応用—. 東京大学出版会.
- 清水留三郎 (1983). 共通第1次学力試験における解答の分析について. 大学入試フォーラム, **1**, 36-37.
- 竹村彰通 (1991). 現代数理統計学. 創文社.
- 野村祐次郎 (1983). 昭和54年度「化学I」解答の分析結果からみた問題の評価 (解答分析その1). 大学入試フォーラム, **1**, 38-47.

付 録

本節では、式(2)を数値計算し、設問解答率分析図をプロットしてみることにする。3節と同じモデルに従うものとし、 $n=30$ 、 $a_j=1$ 、 $j=1, \dots, 30$ 、 b_j は以下の通りとする。

- Type 1: $b_j = -2, j = 1, \dots, 6$
- Type 2: $b_j = -1, j = 7, \dots, 12$
- Type 3: $b_j = 0, j = 13, \dots, 18$
- Type 4: $b_j = 1, j = 19, \dots, 24$
- Type 5: $b_j = 2, j = 25, \dots, 30$

このような5つのタイプの項目に関する item-test regressions、項目正答率分析図、設問解答率分析図をそれぞれ図8～10に示す。ただし、項目正答率分析図にそろえて設問解答率分析図の横軸は、5つの群の累積確率の代表値として、 $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ にプロットした。図10に見られる数値計算による設問解答率分析図のプロットは、実際のデータから計算した設問解答率分析図(野村、1983、石塚・山田、1987等参照)に見られるものに近いことがわかる。

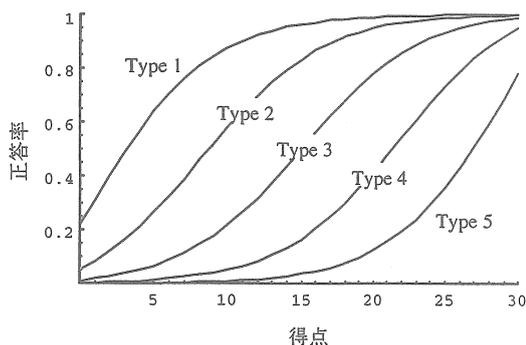


図8 Item-Test Regressions

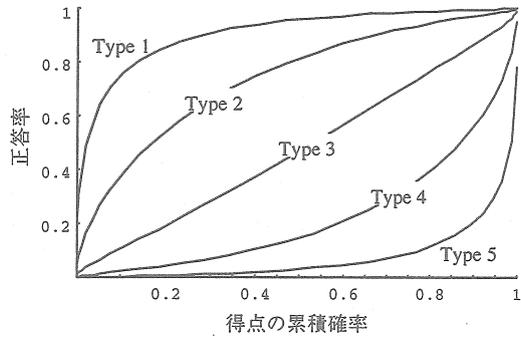


図 9 項目正答率分析図

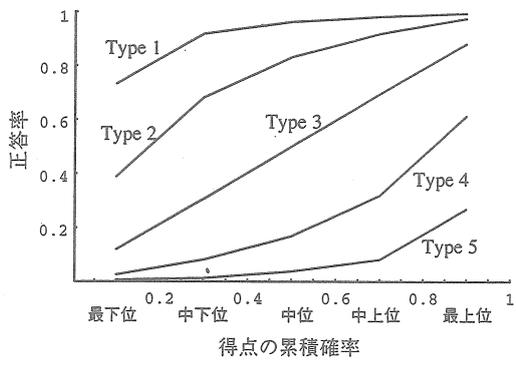


図 10 設問解答率分析図

Evaluation of Quintile Item Response Chart Using Item Response Theory

Kenichi KIKUCHI*

Abstract

At the National Center for University Entrance Examinations (NCUEE), we have analyzed the performance of items on NCUEE administered tests. To do this, we use a method called the quintile item response chart (QIR chart). QIR charts are plotted as follows:

1. Based on their test scores examinees are divided into five equally sized groups.
2. For each item we calculate and then plot the proportion of examinees in each of the five groups who returned a correct response.

In this paper, we use item response theory to evaluate how changes in test score distribution affect QIR charts. We conclude that QIR chart is not greatly affected by changes in the distribution of test scores. Since QIR charts are easy to plot and the charts are stable, they may be used to analyze item performance.

Keywords: quintile item response chart, item response theory

* Information Processing Section, Research Division,
The National Center for University Entrance Examinations.
2-19-23 Komaba, Meguro-ku, Tokyo 153-8501 Japan.
e-mail: kikuchi@rd.dnc.ac.jp, Telefax: 81-3-5478-1297