

# 設問別解答率分析図について

大阪大学大学院基礎工学研究科 白旗 慎吾

## 1. はじめに

試験問題の特性を検討するためにはその正解率を分析しなければならない。分析には理論だけではなく、グラフに表現して考察するグラフィカルな方法も有益であり、その有力な手段として設問別解答率分析図がある（例えば菊地（1999））。分析図は受験者群を試験総得点によってグループ分けし、試験総得点を横軸に、試験総得点と同じグループ内でのある設問の正解率を縦軸に表現したものであり、そこから試験問題の検討や項目特性曲線等の考察を行っている。

この方法は多くのデータに適用されており（例えば、石塚，平，清水（1992）等），受験生の学力を反映した素直な設問では図は右上がりの傾向を示すと期待される。ただしこの方法では、総得点でグループ分けしているために高得点者の正解率が自然に高くなる、という欠点がある。例えば、ある多岐選択が難しすぎて正解するかどうかは全くの偶然の産物になったとしても分析図は右上がりの傾向を示す。ある設問が、理解が深いほど逆に誤解して誤答する傾向の強い悪問だったとしても（この場合は右下がりが期待される）、右下がりになるとは限らない。つまり当然期待される傾向がグラフに現れるとは限らない。

適用例では右下がりになることは稀であるが、常に右上がり傾向が顕著とは限らず、その場合上記の性質を考えると、試験問題の適切性の検討資料として使えるかどうか疑問を持つことも多い。そこで、本論では設問別解答率分析図の修正法を提案する。

## 2. 設問別解答率は何を測っているか

試験は  $K$  個の項目（設問）から成り、各項目

に対しては正解，不正解の2通り，かつ試験得点は正解の個数とする。すなわち第  $k$  番目の設問に，正解の場合  $X_k = 1$ ，不正解なら  $X_k = 0$  とする。試験得点は  $X = X_1 + \dots + X_k$  である。

学力（得点能力）を表すパラメータを  $\theta$  とする。 $\theta$  は受験者を固定すれば定数であるが，受験者全体では確率変数と見なす。かつ，受験者ごとに独立であり，さらに， $\theta$  が与えられたときの条件付き分布の意味では  $X_1, \dots, X_k$  は独立と仮定する。これは項目反応理論では通常仮定されている（芝（1991），池田（1994））。以下の記号を用いる。

- $\theta$  の密度関数を  $f(\theta)$  とおく。
- $X$  の確率関数を  $P(X=k)=g(k)$  とする。
- $\theta$  が与えられたときの  $X$  の条件付き確率関数を  $P(X=k|\theta)=g(k|\theta)$  とおく。
- $X=k$  のときの  $\theta$  の条件付き密度を  $f(\theta|k)$  とおく。
- $X$  と  $\theta$  の同時分布を  $h(k, \theta)$  とおく。
- $\theta$  が与えられたときの設問  $i$  への正解確率を  $P(X_i=1|\theta)=P_i(\theta)$  とおく。項目反応理論では  $P_i(\theta)$  にロジスティック曲線等を仮定するが，ここでは具体形は仮定しない。
- 設問  $i$  を除いたときの正解数  $X_{(i)}$  に関する確率関数や密度関数については添え字として  $(i)$  をつける。例えば  $g_{(i)}(k)=P(X_i=k)$ ， $f_{(i)}(\theta|k)$  等。

このとき設問別解答率は

$$(1) \quad E(X_i = 1|X = k)$$

$$(2) \quad = \frac{P(X_i = 1, X_{(i)} = k - 1)}{P(X = k)}$$

$$(3) \quad = \frac{g_{(i)}(k - 1)}{g(k)} \int P_i(\theta) f_{(i)}(\theta|k - 1) d\theta$$

を推定している。ただし、 $k=0$ のときは0となる。式(3)の積分部分は $X_{(i)}$ が与えられたときの $\theta$ の事後分布による正解率の平均であるが、その前にかかっている確率の比のために解釈がしにくい。 $k$ が大きくなると自然に大きくなってしまふ現象もこの確率の比のいたずらである。

正解数	$i$ に正解	$i$ に不正解	計
0	0	$n_{00}$	$n_0$
1	$n_{11}$	$n_{10}$	$n_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$k$	$n_{k1}$	$n_{k0}$	$n_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$K-1$	$n_{K-1,1}$	$n_{K-1,0}$	$n_{K-1}$
$K$	$n_{K1}$	0	$n_K$
計	$m_1$	$m_0$	$n$

$n$ 人の $X_{k1}$ を下のように表示する。

式(1)は  $\frac{n_{k1}}{n_k}$  で推定されるので、設問  $i$  に対

$$(4) \quad (k, n_{k1}/n_k), \quad k = 0, \dots, K$$

する設問別正解率分析図は  
を图示して与えられる。

### 3. 分析図の修正 (修正1)

通常の分析図の欠点は、合計点は考える設問

$$(5) \quad E(X_i | X_{(i)} = k)$$

$$(6) \quad = \frac{P(X_i = 1, X_{(i)} = k)}{P(X_{(i)} = k)}$$

$$= \int P_i(\theta) f_{(i)}(\theta|k) d\theta$$

の得点を含んでいることによる。そこで  
を考える。これは学力に事前分布を想定し、データによってその事前分布を修正した事後分布を求め、その事後分布で正解確率を平均したものであり、ベイズ統計学的に通常の分析図より解釈しやすく、上述の欠点がない。設問  $i$  に対

$$(7) \quad (k, \frac{n_{k+1,1}}{n_{k+1,1} + n_{k0}})$$

する修正設問別正解率分析図は

を  $k=0, 1, \dots, K-1$  に対してプロットして得られる。

なお、式(3)の前にかかっている確率の比を除いて图示すれば、ここで提案された図を右に1ずらした図が得られる。

### 4. 別の修正 (修正2)

2, 3節の図は積分の中が設問  $i$  を除いた得点で条件づけられた事後平均から得られる。ただし条件付けは全得点で行う方が自然であろう。

$$(8) \quad \int P_i(\theta) f(\theta|k) d\theta$$

全得点にするには  
が推定できればよい。

$$\frac{1}{g(k)} \int P_i(\theta) g(k|\theta) f(\theta) d\theta$$

ただし、このままでは推定できない。式(8)は

$$g(k|\theta) = P_i(\theta) g_{(i)}(k-1|\theta) + (1-P_i(\theta)) g_{(i)}(k|\theta)$$

と変形されるが、関係

$$\frac{g_{(i)}(k-1)}{g(k)} \int P_i^2(\theta) f_{(i)}(\theta|k-1) d\theta + \frac{g_{(i)}(k)}{g(k)} \int P_i(\theta)(1-P_i(\theta)) f_{(i)}(\theta|k) d\theta$$

を用いると

を得る。2, 3節での結果を $P_i$ の推定と見なし、その2乗で $P_i^2$ を推定し、 $P_i(1-P_i)$ も同様に推定すると設問別解答率分析図は横軸の値  $k = 0, 1,$

$$(9) \quad \frac{n_{k1}^2}{n_k(n_{k1} + n_{k-1,0})} + \frac{n_{k+1,0}n_{k0}}{n_k(n_{k+1,0} + n_{k0})}$$

$\dots, K$  に対して

をプロットして得られる。ただし  $k=0$  の場合は第2項のみ、 $k=K$  の場合は第1項のみである。これは式(8)のかなり荒い近似であり、多数の受験者がいてもより正確とは限らない。実際この推定量は一致性を持たない。

もし推定量の構成がうまくいけばこの修正により最初に述べた分析図の欠点は大幅に減少す

る。

## 5. シミュレーション結果

式(4), (7), (9)による3種の設問別解答率分析図をシミュレーションにより比較する。簡単のため以下のように設定する。項目反応理論のロディスティックモデルなどの通常のモデルは仮定しない。

- $K$ とし標本数は10,000とする。
- 簡単のため  $\theta$  の事前分布は区間  $(0, 1)$  の一様分布, すなわち  $f(\theta) = \langle \theta < 1$  とする。
- $P_1(\theta) = \frac{1}{2}$ ,  $P_2(\theta) = 1 - \theta$ ,  $P_3(\theta) = P_4(\theta) = P_5(\theta) = \theta$  とする。すなわち, 設問1は正解するかどうかは全くの偶然であり, 設問3-5は得点能力を学力とし, 設問2は設問3-5と逆の傾向にあり, 学力のある者が間違える悪問である。

このとき式(2), (6), (8)の真値を計算し, かつコンピュータ・シミュレーションを行った。各設問の真値を求めシミュレーション結果と比較した結果では, 修正法2の誤差がやや大きい, 従来法と修正法1ではデータ数が多いため推定精度は十分であった。真値の図は省略する。設問1-3のシミュレーション結果を図1-3に示す。その結果によれば, 従来法は設問1(図1), 設問2(設問2)でも顕著な右上がりの傾向を示す。これは不合理な傾向である。一方, 提案された2つの修正法にはその欠点が見られない。また設問3(図3)に対しては従来法に比べ提案された方法の右上がり傾向は小さい。理論的には, 推定は困難であるが式(8)が最も合理的な指標である。ただし推定精度に問題があるため常にそう言えるかどうかは疑問である。

## 6. 連続の場合

各  $X$  が連続の場合を考える。  $X_i$  の  $\theta$  が与えられたという条件の下での条件付き密度を  $g_i(x|\theta)$  とおく。このように  $X$  に関連した密度等は添え字に  $i$  をつける。前節までの確率関数は密度関数と読み替える。また, 各  $X$  が離散ではあるが,

複数個の値を取りうる場合は以下で積分を和に置き換えればよい。

$I$  を適当な区間とする。  $I$  が1点のみであるときは以下の積分で, を省略すればよい。得点の可能な範囲をいくつかの区間に分割して, そこにおける平均得点率を図示するのが, 項目別正解率分析図である。  $(X_i, X_m)$  および  $(X_i, X)$  の同

$$\int g_i(x_i|\theta)g_{(i)}(x_{(i)}|\theta)f(\theta)d\theta,$$

$$\int g_i(x_i|\theta)g_{(i)}(x - x_i|\theta)f(\theta)d\theta$$

時密度関数がそれぞれ

$$(10) \quad E(X_i|X \in I)$$

で与えられることを用いて

を積分の比で表現できる。通常の項目別正解率分析図はこれを推定しているが, その式の解釈は困難である。実際の分析図ではしばしば  $P(X \in I)$  が一定(可能な得点の範囲を  $L$  個に分ける場合は  $1/L$ ) となるような分割を考える。

$$(11) = \frac{E(X_i|X \in I)}{L} = \frac{\int \int \int x_i g_i(x_i|\theta) g_{(i)}(x - x_i|\theta) \times f(\theta) d\theta dx_i}{\int \int \int f(\theta) d\theta dx_i}$$

その場合上式は

と表されるが, 解釈の困難さに変わりはない。

$$(12) = \frac{E(X_i|X_{(i)} \in I)}{\int \int x_i g_i(x_i|\theta) f(\theta|X_{(i)} \in I) d\theta dx_i}$$

3節にならって修正すると

を得る。これは  $X_m$  が区間  $I$  に入ったという情報があるときの  $\theta$  の事後分布を考え, その分布で  $\theta$  が与えられたときの  $X$  の平均であると解釈される。グラフに表すことも容易であり, 設問  $i$  に対する分析図は, 設問  $i$  を除いた教科の合計点が区間  $I$  に入った受験者の平均点をプロットして得られる。

$$(13) \quad \int \int x_i g_i(x_i|\theta) f(\theta|X \in I) d\theta dx_i$$

4節にならって,

を考えるのも合理的と考えられるが, 推定量の

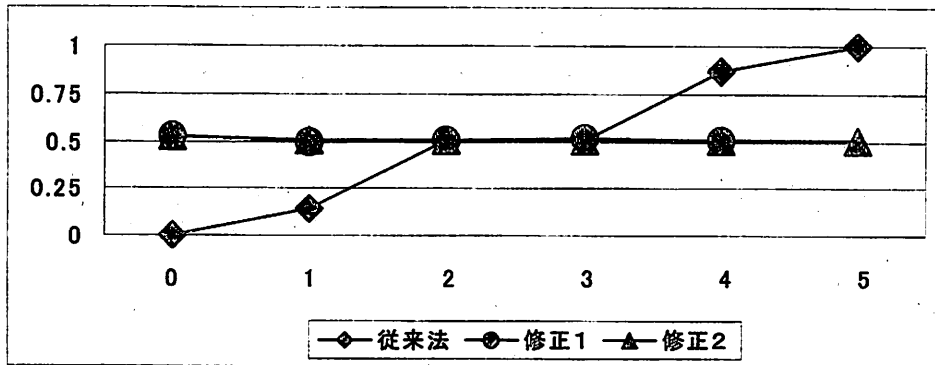


図1. 離散の場合の設問1に対するシミュレーション結果

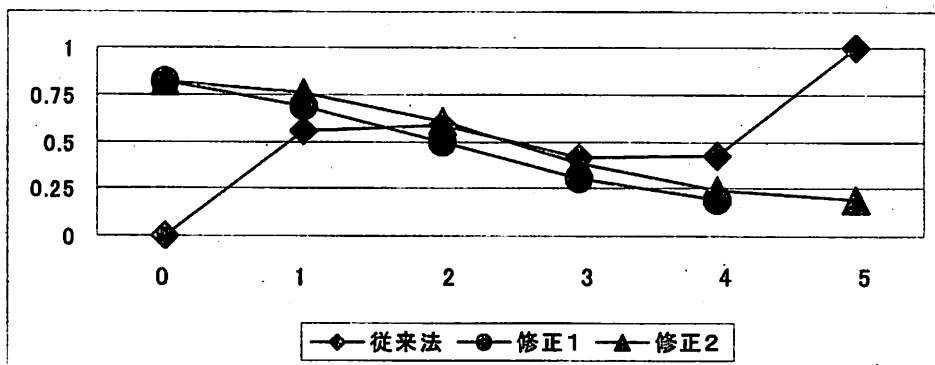


図2. 離散の場合の設問2に対するシミュレーション結果

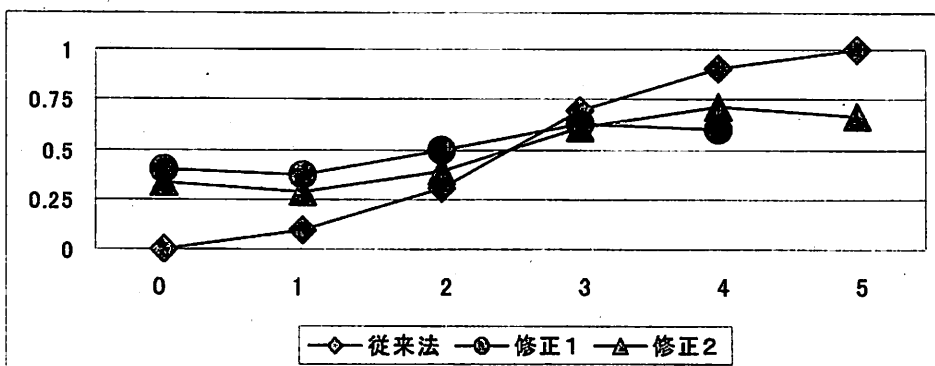


図3. 離散の場合の設問3に対するシミュレーション結果

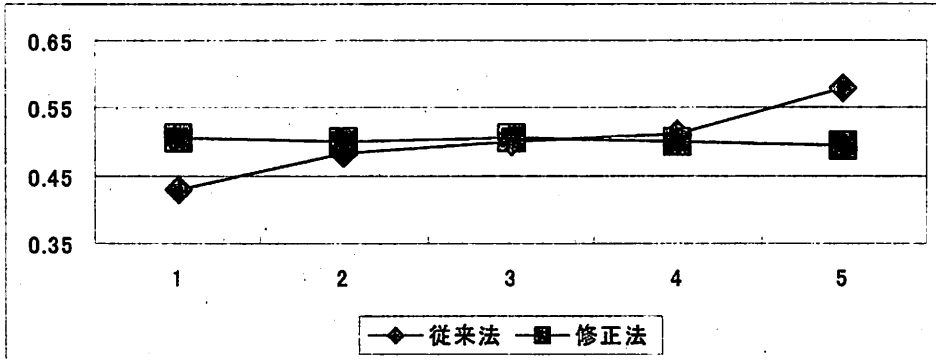


図4. 連続の場合の設問1に対するシミュレーション結果

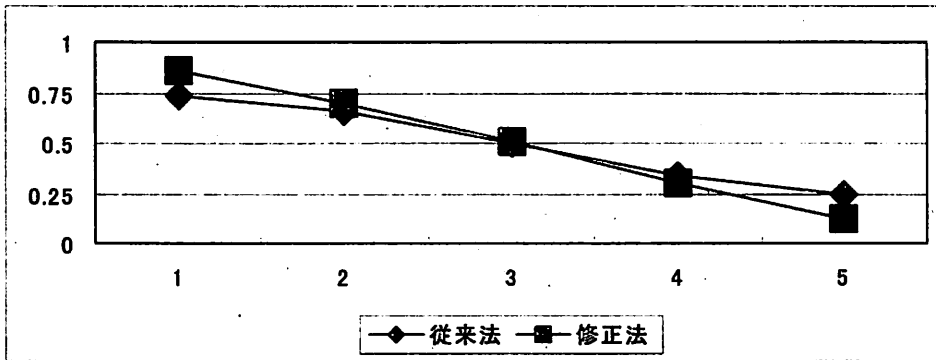


図5. 連続の場合の設問2に対するシミュレーション結果

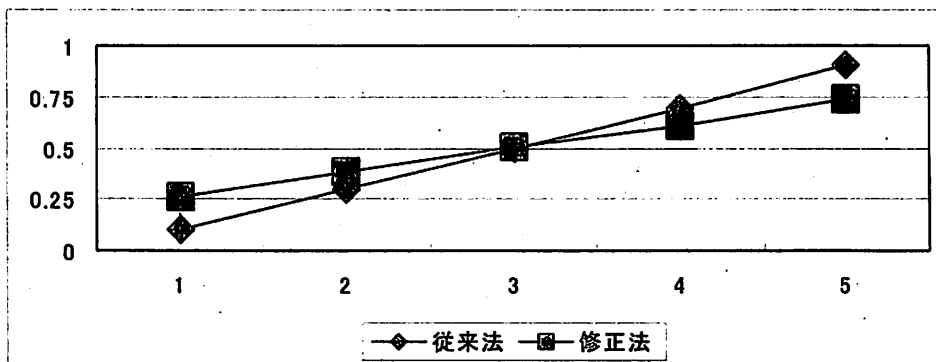


図6. 連続の場合の設問3に対するシミュレーション結果

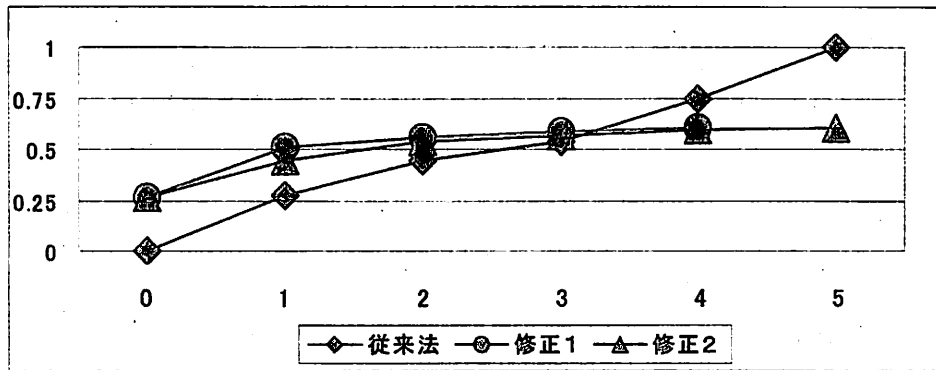


図7. センター試験国語のセンター試験における分析図

構成が困難である。

式(10), (12)はその真値を求めるのは困難である。そこでシミュレーションによりその性質を探った。モデルは5節と同様で以下のように設定した。

- $K=5$  とし標本数は10,000, 各設問の得点は区間(0, 1)に入る, とする。
- $\theta$  の事前分布は区間(0,1)の一様分布とする。
- 設問1の得点は $\theta$ の値に関わらず, 平均0.5, 標準偏差0.2の正規分布とする。設問2の得点は上と同じ分布の正規分布に $1-\theta$ を加えたものとする。設問3-5はやはり同じ分布の正規分布に $\theta$ を加えたものとする。すなわち, 設問1は正解するかどうかは学力に無関係であり, 設問3-5は得点能力は学力とほぼ等しく, 設問2は設問3-5と逆の傾向にあり, 学力のある者が間違える傾向の強い悪問である。

シミュレーションの結果を図4-6に示す。シミュレーション・データからは, 特に $\theta$ や他の設問に無関係な設問1で式(10)の動きが異常である。ただし離散の場合に比較すれば異常の程度は小さいが, 修正法の方が従来法より良いことは明らかであろう。

### 7. 実際例

図7は大阪大学のある学部受験生のセンター試験得点とその中の国語のデータに適用した設問別解答率分析図である。ただし得点を偏差値に変換した後, 偏差値50以上を1, 50未満を0として離散化した。グラフからは従来法では顕著な右上がりの傾向が見られるが, 修正法1, 2では中位以上の受験生ではほとんど差がないことが読みとれる。したがって, センター試験国語の得点は従来の分析図から受ける印象ほど受験生間で差がないことが分かる。

### 参考文献

- 池田 央 (1994) : 現代テスト理論。朝倉書店
- 石塚智一, 平 直樹, 清水留三郎 (1992) : 社会の試験問題の出題形式に関する比較研究。大学入試センター研究紀要, 21, 1-33.
- 菊地賢一 (1999) : 項目反応理論を用いた設問解答率分析図の評価。大学入試センター研究紀要, 29, 1-8.
- 芝 祐順 (1991) : 項目反応理論 — 基礎と応用。東京大学出版会