適応型テストの回答データを用いた項目特性値の推定

～データの偏りが項目特性値推定に及ぼす影響～

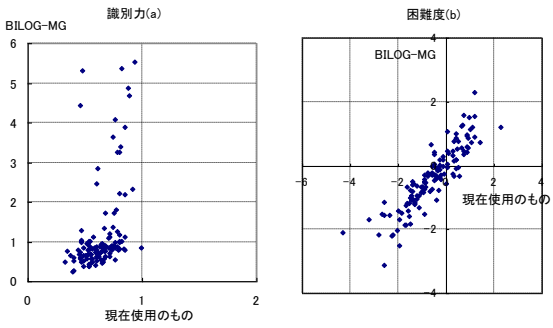
○藤田彩子　舛田博之
㈱リクルートマネジメントソリューションズ組織行動研究所

1.　背景

情報技術（Information Technology；IT）の発達に伴い，従来型の紙筆版のテストがコンピュータテスト（Computer Based Test；CBT）に置き換えられることが多くなってきている。CBTでは，項目反応理論（Item Response Theory；IRT）を使った適応型テストを実現することも可能であり，十分大きな項目プールがあれば，回答者それぞれに適した異なる問題を提示し，効率よく個人特性を測定することができるというメリットがある。

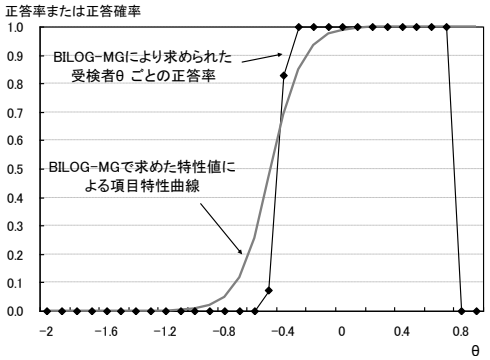
小社は 40 年来，紙筆版の基礎能力検査(語彙や読解などを中心とする言語検査と，数量の計算や推論などを中心とする非言語検査からなる)を主に企業の新卒採用向けに提供しているが，2004 年から，開始した総合テスティングサービスにおける CBT 適応型基礎能力検査は，ここ数年で紙筆版を超える受検者数を数えている。

CBT の受検者が増加すれば，項目プールの一層の充実が求められるようになる。新しい問題項目を追加する場合，テストの中に新作の項目を採点除外項目として紛れ込ませる形でデータ収集を行うことができれば,通常の運用の中で無理なく項目プールを拡充していくことも可能である。ただ，一方で，紙筆版から CBT に移行する際，往々にして紙筆版のテストを実施する中で蓄積された回答データを分析した項目特性値しか手元になく，これを用いている場合もある。この項目プールに後から CBT を運用する中で得られたデータを用いて推定した特性値をもった項目を混在させていいのか,という問題が生じる。



【図 1】現在使用中の紙筆版の項目特性値と適応型データを BILOG-MG で分析した項目特性値の比較

前田他(2004)によれば，同じ問題項目でも紙筆版の中で出題した場合と CBT の中で出題した場合では，項目の識別力・困難度がかなり異なる。よって，紙筆版で得られたデータから推定された特性値をもつ項目と CBT で得られたデータから推定された特性値をもつ項目を混在させることは望ましくない。項目プール拡充のためのデータ収集を CBT で行う運用を考えた場合，それまで使っていた紙筆版回答データから算出した項目特性値を，どこかのタイミングで CBT 回答データのものに置き換える必要があるだろう。このような背景から，本研究では，適応型テストを運用していく中で得られたデータを用いた項目特性値算出に取り組んだ。



【図 2】BILOG-MG で求めた項目特性値(a=5.52, b=-0.49)による項目特性曲線と θ のレベルごとの実際の正答率

まず，BILOG-MG で，2PT ロジスティックモデルによる非言語検査の項目特性値推定を試みた。算出した値の妥当感を見るため，紙筆版の特性値と比較したところ(図 1)，困難度 b は紙筆版での特性値と比べても妥当な数値が出ていると思われ

たが，識別力 a は極端に高い項目が散見された。このような項目の多くは，回答者の θ の分布が狭い傾向にあり， BILOG-MG による識別力および θ の推定が，ある少数の項目群に過剰に最適化されてしまっているため，それらの項目の識別力が異常に高くなり，その結果，その項目の正答率が θ の狭い範囲で急激に変化しているのではないかと推測された(図 2 に 1 例を示す)。これは，BILOG-MG が，どの項目についても能力分布がほぼ同じ集団が回答するという従来型の紙筆版テストを想定していることによるものかもしれない。そうだとすれば，項目によって回答者集団の能力分布が異なる適応型テストのデータの場合は，別の手法で項目特性値を算出する必要がある。そこでこのような偏りのあるデータでも妥当な項目特性値を推定できる方法を考案し，実際に項目特性値を推定して，この方法の妥当性や課題を検証した。

2. 「回帰推定法」の考案

そもそも項目特性曲線は，θ のレベルごとの正答確率を表すモデルである。各問題項目について，θ のレベルごとにその問題の正答率を求め，その θ のレベルにおける正答確率とみなして，最もよく再現するロジスティック曲線を求めることができれば，識別力と困難度のパラメータを推定することができるだろう。通常は，項目特性値も受検者の θ も未知であるが，すでにテストが運用されている状況では，受検者の θ が得られている。項目 1 つ 1 つに対し実際の正答率を θ レベルごとに求めれば，識別力や困難度の再推定ができると考えられる。θ は紙筆版の特性値をもとに求めたものを使用することになるが，あえてこれを用いることで，項目特性値の再推定の前後で θ の等化が行われることになり，検査結果の継続性を保つ上では，好都合である。

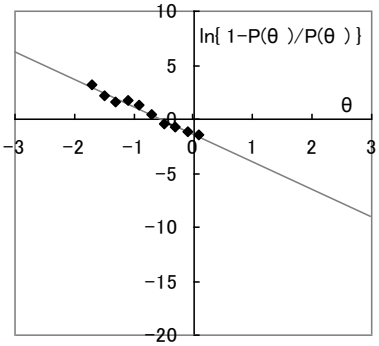
具体的には，特性値を推定したい項目に対する正誤が，θ の推定に与える影響を除くために，当該項目以外の回答データを用いて θ を再算出し，その値(最小－3 最大 3)の 0.2 ごとに受検者を 30 のグループに分け，グループごとにその項目に対する正答率を求めた。このとき，20 人に満たないグループは，正答率に対する誤差が大きい可能性を考えて使用しないこととした。2PT モデルの場合，ロジスティック曲線(式①)は，変形すると式②のような識別力パラメータ a と困難度パラメータ b の一次式となるので，単回帰分析により a，b 両パラメータを推定することができる(図 3)。

$$P(\theta)=\frac{1}{1+\exp(-1.7a(\theta-b))} \qquad \cdots ①$$

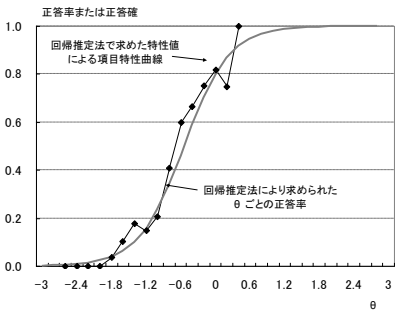
$$-1.7a(\theta-b)=\ln\left\{\frac{1-P(\theta)}{P(\theta)}\right\} \qquad \cdots ②$$

この方法を「回帰推定法」と呼ぶこととする。回帰推定法は，θ が既知の条件下で実際に起こっている受検者の回答状況に合うように特性値を推定するため，項目ごとの受検者分布に関する懸念が不要である。ただし，図 3 からもわかるように，1 つ 1 つのデータ(グループごとの正答率)が回帰係数に与える影響は大きい。安定した推定を行うには各グループに一定以上の人数が必要であると考えられる。図 4 に，回帰推定法で推定した項目特性値による項目特性曲線(図 2 と同じ項目)と θ のレベルによる正答率を示した。

回帰推定法を用いて非言語検査の 131 項目の項目特性値を算出し，紙筆版の項目特性値と比較した(図 5)。2 つの方法で算出した特性値の相関係数は，識別力で 0.567，困難度で 0.880 となっており，特に困難度については，適応型テストでのデータでも紙筆版とほぼ同様な値が得られ



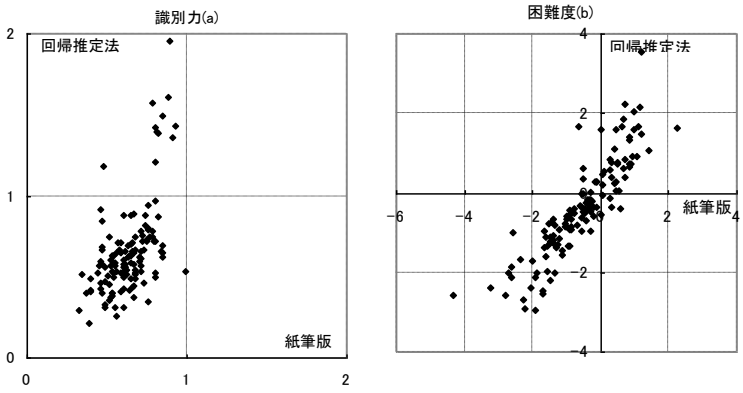
【図 3】回帰推定法による項目特性値の推定　直線の傾きと切片から項目特性値が推定できる



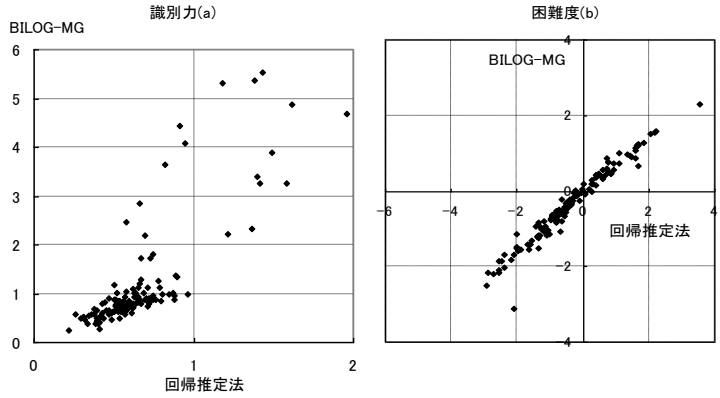
【図 4】回帰推定法によって推定した項目特性値による項目特性曲線と実際の正答率

ることがわかる(表 1)。

BILOG-MG による値と回帰推定法による値と比較したところ，BILOG-MG で識別力が極端に高く推定された項目についても，回帰推定法による識別力はそこまで高くならない(図 6)。回帰推定法が，これまでと継続性のある θ の軸をよりどころにしているためと思われる。回帰推定法は，①これまでのテストとの継続性を保つことができ，②適応型テストで起こりがちな回答者の能力分布が狭くなるという現象が，項目特性値の推定に与える影響を緩和できる方法といえる。



【図 5】回帰推定法によって推定した項目特性値と紙筆版データによる項目特性値の比較



【図 6】回帰推定法によって推定した項目特性値と BILOG-MG で推定した項目特性値の比較

3. 回帰推定法による項目特性値の推定に必要なデータ数

回帰推定法は，θ ごとのグループを細かく分けるほど回帰分析に用いることができるデータ(グループごとの正答率)が多くなるが，一方で，各グループに属する回答者が少なくなるため，グループごとの正答率が安定しなくなったり，人数が少なすぎてデータとして採用できないグループが増加したりする。本研究では，グループをつくるための θ の刻みを 0.2 ごととしたが，この場合にどのくらいの回答者のデータがあれば安定して推定ができるのかを検証した。回答者 6 万人規模(2008 年)と，15 万人規模(2009 年)のテストデータを用いて推定した特性値を比較した。適応型テストであるため，項目によって回答者の人数が大きく異なる。項目ごとの回答者数の概要を，表 2 に示す。

【表 1】回帰推定法によって推定した項目特性値と紙筆版データによる項目特性値の相関

		紙筆版	回帰推定法
識別力	平均	0.633	0.658
	相関係数	0.567	
困難度	平均	−0.578	−0.375
	相関係数	0.880	

【表 2】2 つの適応型テストデータの項目ごと回答者数

データの規模	約6万人	約15万人
平均	1950.8	8391.6
最小値	468	2583
最大値	4004	15334

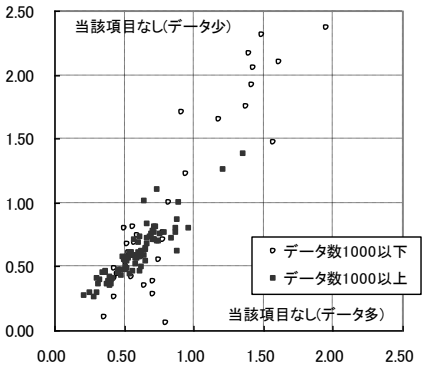
2 種類のデータによる識別力 a の値を比較したところ，約 6 万人のデータで 1000 人以上の回答者がいる項目では，約 15 万人のデータでの識別力 a の値と大きく違わない推定値を得ることができた。一方で，データ数が 1000 以下の場合は，15 万人のデータでの識別力との乖離が大きい項目が多く見られた(図 7)。今回の分析においては，1000 人以上の回答者がいる場合は安定した推定ができるということがいえる。

4. 対象項目の回答データも含めた θ を用いた場合

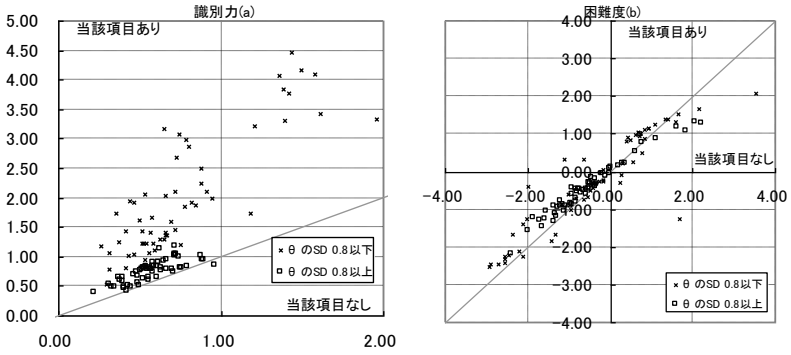
本研究では，特性値の推定を行う項目に対する回答を除いた上で，残りの項目に対する回答データのみを用いて受検者の θ を再推定した。ただ，もし当該項目の回答情報も含めて求められた θ を使って項目特性値の推定ができるとすれば，本研究のような状況の場合，すでに入手している θ を利用することができるので，大変便利である。そのような観点から，テストで得られている

θを用いた場合，識別力や困難度がどのように変わるか確認することとした。

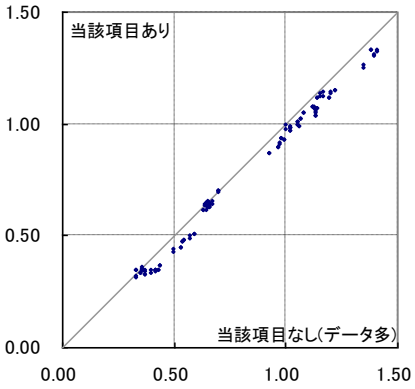
15 万人規模のデータを用いて，テストの結果としてすでに得られているθを用いて回帰推定法により推定した項目特性値と，当該項目を抜いて再算出したθを用いて推定した項目特性値を比較したのが図 8 である。すでに得られているθを用いて推定した識別力は，全体的に当該項目抜きのデータで算出した値より高くなっているが，中でもかなり高くなっている項目群がある。その項目に対する正誤と最終的な結果であるθの関連性が強くなることが伺える。これらの項目には，推定に用いている回答者のθの分布が狭い(0.8 程度より小さい)という共通点があり，回答者のθの分布が狭い場合，当該項目を含めて推定されたθを使うことで，識別力が本来より高く推定される可能性があることが示唆される。当該項目を除くことで，多くの場合，その項目の回答者集団の分布が広くなる。ただし，その分布の広がり方は，もともとの回答者集団のθの標準偏差の大きさによってあまり違いがない(図 9)。θの標準偏差が小さい場合は，その分，当該項目を抜くことが識別力の推定に与える影響が大きいことが考えられる。一方で，θの標準偏差がある程度以上大きければ，当該項目が含まれる影響はかなり小さくなる。困難度 b の値も当該項目なしの場合と大きな違いはなく，テスト結果としてすでに得られているθを用いて推定した項目特性値が，十分実用的である可能性がある。



【図 7】回帰分析法により識別力を 6 万人規模のデータおよび 15 万人規模のデータで推定した場合の推定値の比較



【図 8】回帰分析法で特性値を推定する際に，当該項目を抜いて算出したθをもとにした場合と当該項目を含めたまま算出したθをもとにした場合の比較



【図 9】当該項目なしの場合と当該項目ありの場合での，項目ごとの回答者のθの標準偏差の比較

5. 今後の課題

本研究では，適応型テストを運用する中で得られた特性値を推定するやり方を考案した。今後は，推定した特性値を実際に使った際に，どこまで結果の等化が実現されるか，確認していきたい。

引用文献

前田純子 他(2004)『一般知的能力検査における紙筆版と CBT の項目特性の比較』 日本テスト学界第２回大会発表論文抄録集