適応型テストの回答データを用いた項目特性値推定

―能力特性値の標準誤差情報の利用―

○杉山 剛
株式会社リクルートマネジメントソリューションズ測定技術研究所

1　はじめに

適応型テストの運用においては項目プールの拡充は重要なテーマのひとつである．項目プールの拡充の方法として，予備テストを行って新作項目の項目特性値をあらかじめ算出する方法のほかに，本番テストの中に新作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おこなって項目特性値を算出する方法がある．本番テストの中に新作項目を紛れ込ませる方法には，本番テストで得られた受検者の特性値 θ を新作項目の項目特性値推定に使用することで本番テストの項目との等化済みの項目特性値が算出されるという利点や，本番テストの運用の延長上で項目プールの拡充が可能になるという利点がある．一方で，予備テストを行う方法の場合は実務上莫大なコストがかかるほか，運用上の負荷も大きい．

受検者の特性値 θ と項目への正誤を用いて項目特性値を推定する方法として，藤田ら(2009)が提案した回帰推定法がある[1]．これに対し，杉山ら(2016)では受検者の特性値 θ と新作項目への正誤反応のみを用いて疑似最尤法による項目特性値の推定を行う方法を提案した[2]．この方法では受検者の特性値 θ の値として，同時に受検している本番テストの推定結果を用いているが，この推定結果には測定誤差を含んでいる．測定誤差を含んだ特性値 θ を定数として項目特性値の推定をおこなう場合，項目特性値の推定には θ の誤差の分が加味されていないため推定誤差が過小評価されている点に注意が必要であった．

本発表では，受検者の本番テストの推定結果として，特性値 θ だけでなく推定時の標準誤差も加味して項目特性値を推定することで，項目特性値の推定結果に変化がみられるかどうかについて確認した結果を報告する．

2　2つの方法の比較

2.1　受検者の特性値 θ と正誤のみを使った項目特性値推定

適応型のテストに新作項目を紛れ込ませる形でＮ人が受検した場合を考える．

受検の結果，新作項目以外の項目（項目特性値は既知）への正誤により推定された受検者の特性値 θ と，当該受検者の新作項目への正誤ｕのＮセットの組合せが得られる．

この組み合わせが得られる確率は，以下の式で表すことができる．

$$f(a,b,u,\theta)=\prod_{i=1}^{N}p(\theta)^{u_i}q(\theta)^{1-u_i}$$

ここで正答の確率 $p(\theta)$，誤答の確率 $q(\theta)$ は 2 母数ロジスティックモデルでは以下のように表される。

$$
\begin{aligned}
p(\theta) &= \frac{1}{1 + exp[-Da(\theta - b)]} \\
q(\theta) &= 1 - p(\theta) \\
&= \frac{exp[-Da(\theta - b)]}{1 + exp[-Da(\theta - b)]}
\end{aligned}
$$

杉山ら(2016)の提案手法では u と θ を既知の定数として扱い, 対数尤度関数を最大化する a 値と b 値を，ニュートン・ラフソン法を用いて数値的に求めている.

2.2 受検者の特性値 θ と正誤を使う際に測定誤差を加味した項目特性値推定

上記の方法では θ の値として，本番テストの測定結果をそのまま定数として用いているが，実際には受検者の特性値 θ は実際には知ることができず，同時に実施している本番テストの推定結果 $\hat{\theta}$ には測定誤差 ε が含まれている.

$$
\theta = \hat{\theta} + \epsilon
$$

ここで測定誤差の値 ε も実際には知ることができないため，推定の際に同時に算出される標準誤差 $StdErr\_\hat{\theta}$ を使って， ε は平均 0，標準偏差 $StdErr\_\hat{\theta}$ の正規分布に従うと想定する.

$$
\epsilon \sim N(0, StdErr\_\hat{\theta})
$$

この場合，正答確率 $p(\theta)$ は以下のようになり， $\hat{\theta}$，$StdErr\_\hat{\theta}$ および正誤 u を用いて項目特性値を推定すればよい.

$$
\begin{aligned}
p(\theta) &= \frac{1}{1 + exp[-Da(\theta - b)]} \\
\theta &\simeq \hat{\theta} + N(0, StdErr\_\hat{\theta})
\end{aligned}
$$

3　仮説およびシミュレーション実験

3.1 仮説

2 つの手法の違いは項目特性値の推定にあたり，受検者の能力特定値 θ に本番テストの推定値 $\hat{\theta}$ のみを扱うか，推定の標準誤差 $StdErr\_\hat{\theta}$ も加味して扱うかにある．しかし項目特性値の推定には少なくとも数百名の受検結果を使用するため，人数が多くなることで誤差の影響が相互に打ち消され，項目特性値の推定結果への影響が小さくなることが想定される．また，ある人数で誤差の影響が打ち消されるとすれば，それ以上に人数が増えても新たに誤差の影響が大きくなるとは考えにくい

そこで以下の2つの仮説を立ててシミュレーション実験により確認する．

仮説1	2つの手法の項目特性値の推定結果の違いは大きくない
仮説2	仮説1の傾向は一定の人数以上になると変わらない

3.2　シミュレーションの手順

以下の手順でシミュレーション実験をおこなった．

Step1	受検者の特性値θを標準正規分布N(0,1)により発生させる
Step2	項目特性値を a 値(0.4≦a≦0.8)，b 値(-3.0≦b≦3.0)の範囲で変化させた項目を 65 項目用意する
Step3	各受検者の各項目についての正答確率を2母数ロジスティックモデルに基づいて計算する
Step4	Step3で求めた各受検者の各項目についての正答確率と0〜1の一様分布の大小を使って，受検者の正誤を決定する
Step5	各受検者，項目ごとに，当該項目を除く正誤と Step2 で作成した項目特性値を使って最尤推定法により$\hat{\theta}$と$StdErr_\hat{\theta}$を推定する
Step6	①　各受検者の特性値$\hat{\theta}$および Step4 で生成した正誤を用いて各項目の項目特性値をMCMCにより推定し，Step2で設定した項目特性値と比較する ②　各受検者の特性値$\hat{\theta}$と$StdErr_\hat{\theta}$，およびStep4で生成した正誤を用いて各項目の項目特性値をMCMCにより推定し，Step2で設定した項目特性値と比較する

※Step1〜Step4：シミュレーションをおこなうための受検者，項目，回答データの準備

※Step5：当該項目以外を本番テストと見立てた，$\hat{\theta}$と$StdErr_\hat{\theta}$の推定

※Step6：項目特性値推定のシミュレーション．手法により①の②のいずれかをおこなう
　　　　比較のため，①についてもニュートン・ラフソン法ではなくMCMCを使う

3.3　シミュレーションの結果

受検者の人数を200名から1000名に変化させながらシミュレーションを行い，項目特性値の推定結果とStep2で設定した項目特性値のRMSEについて確認した．

表1:項目特性値のMAP推定値のRMSE

①	200名	400名	600名	1000名
a値	0.17	0.10	0.08	0.07
b値	4.68	0.40	0.23	0.20

②	200名	400名	600名	1000名
a値	0.18	0.11	0.08	0.08
b値	3.08	0.27	0.22	0.21

人数が200名では項目特性値の推定結果そのものが安定していない．400名以上では推定が安定し，600名以上になると項目特性値の推定結果の手法による差はかなり小さくなる．

特に200名においてb値のRMSEが大きくなっているのはb値の絶対値が大きい（易しすぎ，難しすぎの）項目ではb値近辺のθをもつ受検者が少なく項目特性値の推定が安定しないためである．そこで，b値の絶対値が2.0以下の項目に絞ってあらためて集計した（表2）．

表 2:項目特性値の MAP 推定値の RMSE（b 値の絶対値が 2.0 以下の項目のみ）

①	200名	400名	600名	1000名
a値	0.15	0.09	0.08	0.06
b値	0.38	0.21	0.19	0.18

②	200名	400名	600名	1000名
a値	0.17	0.10	0.08	0.07
b値	0.30	0.21	0.19	0.18

こちらからは 400 名以上で手法による差が小さくなり，人数が増えてもその傾向は変わらないことがわかる．

なおシミュレーションの詳細については当日発表する予定である．

4　まとめと考察

適応型テストにおいて，本番テストに新作項目を紛れ込ませる形でデータ収集を行う場合に，受検者の能力特性値 θ の値として，同時に実施している本番テストの推定結果 $\hat{\theta}$ を定数として扱うか，推定時の標準誤差を加味するかについて 2 種類の手法の比較をおこなった．

その結果，項目特性値の推定が安定する人数においては，推定時の標準誤差を加味してもしなくても項目特性値の推定結果は同様であることがわかった．これは仮説 1 の通り，$\hat{\theta}$ の誤差の影響が相互に打ち消されたものと考えられる．また，推定結果が同様になる人数を超えると，人数が増えてもその傾向は変わらないことから，仮説 2 についても支持された．

これらにより，杉山ら(2016)の提案手法である疑似最尤推定法による項目特性値の推定は，項目特性値の推定が安定する人数を扱うという前提において，θ の誤差を考慮する必要がないということが確認された．

一方で，b 値の絶対値が大きい項目（易しすぎる項目や難しすぎる項目）において，項目特性値の推定誤差が大きいという課題については解決が必要である．これは，同じ受検者数であっても b 値近辺の θ をもった受検者が相対的に少ないことが原因と考えられる．今後に向けてはデータ収集の観点から，いかにして b 値近辺の θ をもった受検者の回答データを集めるかについて検討を進めたい．

参考文献

[1] 藤田彩子，舛田博之(2009)：「適応型テストの回答データを用いた項目特性値の推定」日本テスト学会第 7 回大会発表資料集

[2] 杉山　剛, 仁田光彦(2016)：「適応型テストの回答データを用いた新作項目の項目特性値推定」日本テスト学会第１４回大会発表資料集