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

○杉山 剛　仁田光彦
株式会社リクルートキャリア測定技術研究所

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]． θ をいくつかのレベルに分け，各レベルの正答率をもとに回帰計算により項目特性値を推定する方法であるが，安定した正答率を得るために一定数以上の回答情報が必要となる．藤田らによると 1000 名以上で安定するとされている．新作項目の場合は，十分吟味して作成したとしても項目特性値がよくないというケースも起こりうるが，回帰推定法の場合 1000 名分のデータを集めた上でないとその判断をおこなうことができないという意味においてやや非効率な面は否めない．

このほか，新作項目の項目特性値を随時推定し，その項目特性値をもとにどのような特性値 θ の受検者に出題するかをコントロールする方法については韓ら(2014)[2]の取り組みがある．

この発表では，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と，この推定を毎日行うことで最新の推定状況をもとに新作項目の出題をコントロールする手法について報告する．提案手法により，適応型テストにおける新作項目のデータ収集および項目特性値推定の効率化に寄与することが期待できる．

2　擬似最尤法による項目特性値の推定

2.1　対数尤度関数の導出

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．

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

また u と θ は既知なので定数とみなすと、尤度関数も同じ式となる。

$$L(a,b|\mathrm{u},\theta) = f(a,b|\mathrm{u},\theta)$$

日本テスト学会第 14 回大会発表論文抄録集　P150-153

ここで正答の確率 $p(\theta_i)$，誤答の確率 $q(\theta_i)$ 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。

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

尤度関数の両辺の対数をとって $p(\theta_i)$，$q(\theta_i)$ を展開することで以下の対数尤度関数が得られる．

$$
\begin{aligned}
log(L(a,b|\mathrm{u},\theta)) &= \sum_{i=1}^{N}(u_i log(p(\theta_i)) + (1-u_i)log(q(\theta_i))) \\
&= \sum_{i=1}^{N}\left(u_i log(\frac{1}{1+exp[-Da(\theta_i-b)]}) + (1-u_i)log(\frac{exp[-Da(\theta_i-b)]}{1+exp[-Da(\theta_i-b)]})\right) \\
&= \sum_{i=1}^{N}((1-u_i)(-Da(\theta_i-b)) - log(1+exp[-Da(\theta_i-b)]))
\end{aligned}
$$

この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る．実際には知ることのできない受検者の特性値 θ に本番テストの受検結果を使い，定数として扱っているため，擬似最尤法による項目特性値の推定となる．

2.2　初期値の算出

ニュートン・ラフソン法を使う場合，まず初期値を最適解の近くに設定する必要がある．

古典的テスト理論で項目の性能を示す点双列相関係数と正答率を使って，項目反応理論の a 値，b 値を求める以下の式が知られている[3][4]．

$$
\begin{aligned}
aa_0 &= \frac{Dr}{\sqrt{1-r^2}} \\
bb_0 &= \frac{z}{r}
\end{aligned}
$$

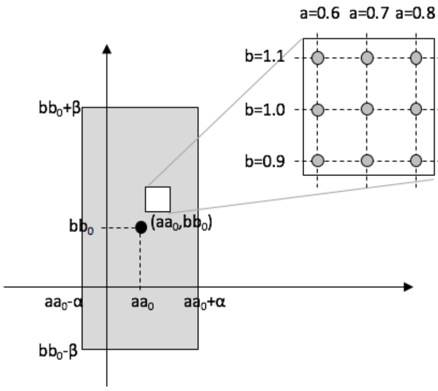
ただし，r は θ と u の点双列相関係数，D は定数(1.7)，z は正答率を標準正規累積確率に対応させたときの z スコア(正答率 $p=P(Z>z)$ となる z である．

しかしニュートン・ラフソン法の初期値とするには，この aa_0, bb_0 は最適解の十分近くにあるとは限らない．そこで aa_0, bb_0 を中心に一定範囲を 0.1 刻みに格子状に調べる（グリッドサーチ）ことで，その中で最も対数尤度関数が大きくなる a 値と b 値の組み合わせを探し，その値をニュートン・ラフソン法の初期値とする(図 1)．

なおここで相関係数の絶対値が 0 や 1 だったり，全員が正答，全員が誤答の場合は aa_0, bb_0 は計算できない．計算できない場合は，基本的には人数が少なすぎるか項目の性能に問題がある（易しすぎ，難しすぎ，識別力が低すぎ)ことが原因なので，この時点においては推定不能とし，回答

データが増えた後にあらためて推定することとする．

図 1　グリッドサーチのイメージ図



※網掛け部の中のすべての格子点について対数尤度を計算する

2.3　ニュートン・ラフソン法による推定

ニュートン・ラフソン法は，求めたい値のベクトル β があるとき，f(β)を β で偏微分した勾配ベクトル Δf(β)と 2 階偏微分したヘッセ行列 Δ^2f(β)を使って，

$$\beta[k+1] = \beta[k] - \Delta^2 f(\beta)^{-1} \cdot \Delta f(\beta)$$

という式で β [k]を順次更新していき，f(β)の変化量が閾値よりも小さくなったらそこで計算を終了，というアルゴリズムである．

今回求めたい値のベクトルは(a, b)なので，上記の更新式は以下となる．

$$\begin{pmatrix} a[k+1] \\ b[k+1] \end{pmatrix} = \begin{pmatrix} a[k] \\ b[k] \end{pmatrix} - \begin{bmatrix} \frac{\partial^2}{\partial a \partial a} LL(a,b) & \frac{\partial^2}{\partial a \partial b} LL(a,b) \\ \frac{\partial^2}{\partial b \partial a} LL(a,b) & \frac{\partial^2}{\partial b \partial b} LL(a,b) \end{bmatrix}^{-1} \begin{pmatrix} \frac{\partial}{\partial a} LL(a,b) \\ \frac{\partial}{\partial b} LL(a,b) \end{pmatrix}$$

※LL(a,b)は対数尤度関数

また，この推定結果の標準誤差はヘッセ行列の逆行列の対角成分の平方根をとって以下の式となる．

$$StdErr\_a = \sqrt{\frac{-\frac{\partial^2}{\partial b \partial b} LL(a[k],b[k])}{\frac{\partial^2}{\partial a \partial a} LL(a[k],b[k]) \frac{\partial^2}{\partial b \partial b} LL(a[k],b[k]) - \frac{\partial^2}{\partial a \partial b} LL(a[k],b[k]) \frac{\partial^2}{\partial b \partial a} LL(a[k],b[k])}}$$

$$StdErr\_b = \sqrt{\frac{-\frac{\partial^2}{\partial a \partial a} LL(a[k],b[k])}{\frac{\partial^2}{\partial a \partial a} LL(a[k],b[k]) \frac{\partial^2}{\partial b \partial b} LL(a[k],b[k]) - \frac{\partial^2}{\partial a \partial b} LL(a[k],b[k]) \frac{\partial^2}{\partial b \partial a} LL(a[k],b[k])}}$$

なお，この推定では当該受検で推定した θ を定数としているが，実際にはこの θ は本番テストの推定結果であり測定誤差を含んでいる．そのため算出された項目特性値の標準誤差は θ の測定誤差を含まない分，過小評価されたものであることに留意する必要がある．

日本テスト学会第 14 回大会発表論文抄録集　P150-153

3　出題のコントロール

適応型テストを実施している大規模コンピュータシステム上に前述の算出ロジックを組み込むと，たとえば毎日の夜間処理ですべての新作項目の項目特性値および標準誤差を計算することができる.

その推定結果を用いることで以下のような出題コントロールが可能となり，効率よく新作項目のデータ収集をおこなうことができる.

① 項目特性値推定時の標準誤差が十分小さくなった項目はこれ以上の回答データは不要なので出題を打ち切る
② 一定以上の人数のデータが集まったにもかかわらず標準誤差が小さくならない項目は項目自体の性能に問題があると判断して出題を打ち切る
③ 前日の推定結果をもとに，b値から大きくかけ離れたθの受検者には，得られる情報量が少なく効率が悪いので出題しない（当該受検者には別の新作項目が割り当てられる）

4　まとめ

擬似最尤法による項目特性値の推定を行うことで，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，データ収集の途中であっても最新の回答情報をもとに項目特性値と標準誤差を推定することが可能になる.

さらに得られた最新の推定結果をもとに適切な出題コントロールをおこなうことで，同程度の受検者数に対して，より多くの新作項目のデータ収集をおこなうことができるようになり項目プールの拡充が促進される.

今後は，標準誤差がどの程度であれば十分な精度といえるか，またどれくらいの人数をもってそれ以上推定精度があがらないと判断できるかなどについては，実際の出題を通じて慎重に吟味・調整していく必要がある. シミュレーションの結果とそれを踏まえた考察については当日発表予定である.

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] 韓 太哲，村木英治，吉川 厚，小林夏子，林 規生(2013)：「コンピュータ適応型テストの中で、項目パラメーターを適応的に推定（３）－２パラメーターモデルについて－」日本テスト学会第11回大会発表資料集
[3] Lord, F. M. (1980). Applications of item response theory to practical testing problems. Routledge.
[4] 加藤健太郎，山田剛史，川端一光 (2014)：「Rによる項目反応理論」オーム社

日本テスト学会第14回大会発表論文抄録集　P150-153