

# 適応型テストの回答データを用いた項目特性値推定

## —本番テストの回答情報の利用—

○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 $\theta$ 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 $\theta$ 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<sup>[1]</sup>。これに対し、杉山ら(2016)では本番テストの結果である受検者の特性値の推定値 $\hat{\theta}$ 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<sup>[2]</sup>。また杉山(2019)では受検者の本番テストの推定結果として、 $\hat{\theta}$ だけでなく推定時の標準誤差も加味する試みについて検証したが、いずれの手法も困難度(b 値)の絶対値が大きな項目の推定精度には課題があった<sup>[3]</sup>。

本発表では、本番テストの回答状況を加味することが項目特性値の推定精度にどのような影響を与えるかについて2種類の方法について確認し、その結果を報告する。

### 2 本番テストの回答状況を加味する2種類の方法

#### 2.1 新作項目の特性値推定の初期値を決める方法

本番テスト結果である $\hat{\theta}$ が低い受検者の多くが正答している項目はb値が低く、 $\hat{\theta}$ が高い受検者の多くが誤答している項目はb値が高いと考えられる。受検者の $\hat{\theta}$ と新作項目への正誤の関係からb値の高低をある程度予測し、項目特性値推定の初期値とすることでb値の収束を早める効果が期待できる。

本研究ではb値の初期値の決め方として以下の3種類を比較した。

- ① 初期値0 (本番テストの正誤状況を加味しない。以下従来手法と呼ぶ)
- ② 新作項目への正答率と、新作項目の正誤と $\hat{\theta}$ の相関係数から算出 (Lord, F. M. (1980) <sup>[4]</sup>)
- ③ 正答者の $\hat{\theta}$ 平均と誤答者の $\hat{\theta}$ 平均から算出

②ではb値の初期値 $b_0$ を以下の式で求める。

$$b_0 = \frac{z}{r}$$

ただし、 $r$ は $\hat{\theta}$ と正誤の点双列相関係数、 $z$ 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。

③では正答者の $\hat{\theta}$ 平均と誤答者の $\hat{\theta}$ 平均の平均を $b$ 値の初期値とする。

## 2.2 各受検者の能力特性値 $\theta$ と新作項目の項目特性値を同時に推定する方法

新作項目の項目特性値を推定する時点では、全受検者について本番テストで出題された項目の項目特性値と正誤がわかっている。適応型テストの場合、受検者により出題項目が異なることには留意が必要だが、本番テストの項目特性値と正誤、そして新作項目の正誤を既知として、各受検者の能力特性値 $\theta$ と新作項目の項目特性値を同時に推定することが可能である。

ある新作項目が出題された $J$ 人の受検者に、適応型出題により全部で $I$ 個の本番テスト項目が出題されていた場合、対数尤度 $LL$ は以下の式で表すことができる。

$$LL = \sum_{j=1}^J \left( \left( u_{trj} P_{trj} + (1 - u_{trj})(1 - P_{trj}) \right) + \sum_{i=1}^I k_{ij} \left( u_{ij} P_{ij} + (1 - u_{ij})(1 - P_{ij}) \right) \right)$$

ここで $u_{trj}$ はトライアル項目への受検者 $j$ の正誤、 $u_{ij}$ は本番テスト項目 $i$ へ受検者 $j$ の正誤、 $P_{trj}$ および $P_{ij}$ はそれぞれトライアル項目および本番項目への正答確率である。また $k_{ij}$ は適応型出題で本番テスト項目 $i$ が受検者 $j$ に出題されたとき1、そうでないときに0の値を取るものとする。

なお2PLMのもとでは、 $P_{trj}$ および $P_{ij}$ は以下の式で表される。

$$P_{trj} = \frac{1}{1 + \exp[-Da_{tr}(\theta_j - b_{tr})]}, \quad P_{ij} = \frac{1}{1 + \exp[-Da_i(\theta_j - b_i)]}$$

本番テスト項目の項目特性値 $a_i$ 、 $b_i$ は既知として、MCMCを用いて対数尤度が最大となるような $a_{tr}$ 、 $b_{tr}$ および $\theta$ を推定する。

この方法では本番テスト項目で $b$ 値の高い項目に正答した受検者がトライアル項目は誤答した、といった情報がトライアル項目の項目特性値推定に加味されることにより $b$ 値の収束を早める効果が期待される。ただし受検者人数が増えていくと推定すべきパラメータも増えていくため計算に時間がかかる点には注意が必要である。

## 3 仮説およびシミュレーション実験

### 3.1 仮説

以下の2つの仮説を立て、シミュレーション実験により確認する。

仮説1	本番テスト結果と新作項目の正誤の傾向を利用して新作項目の $b$ 値の初期値を決めて推定することにより、少ない人数で $b$ 値の推定が安定する
仮説2	受検者の能力特性値 $\theta$ の推定と新作項目の特性値推定を同時に行うことにより、少ない人数で $b$ 値の推定が安定する

### 3.2 シミュレーションの手順

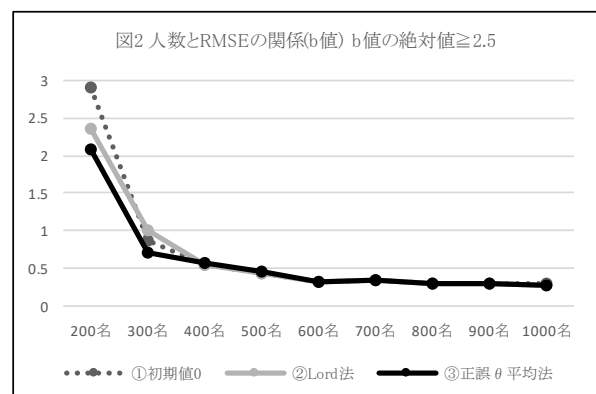
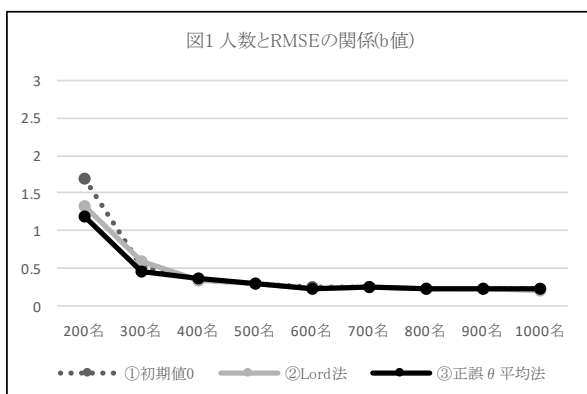
以下の手順でシミュレーション実験をおこなった

Step1		受検者の特性値 $\theta$ 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 ( $0.4 \leq a \leq 0.8$ ), $b$ 値 ( $-3.0 \leq b \leq 3.0$ ) 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 $0 \sim 1$ の一様分布の大小を使って, 受検者の正誤を決定する
Step5		65 項目のうち 1 項目を新作項目 (項目特性値未知) とし, 残りの 64 項目を本番テスト項目として, 64 項目の項目特性値と正誤から受検者の能力特性値 $\hat{\theta}$ を推定する
仮説 1	Step6	3 つの方法で新作項目の項目特性値を推定する ① $b$ 値の初期値を 0 とする ② 正答率, $\hat{\theta}$ と正誤の相関係数から求めた値を $b$ 値の初期値とする ③ 正答者の $\hat{\theta}$ 平均と誤答者の $\hat{\theta}$ 平均の平均値を $b$ 値の初期値とする
仮説 2	Step6	Step5 で決めた本番テスト項目の項目特性値と正誤, および新作項目の正誤から受検者の能力特性値 $\theta$ と新作項目の項目特性値を同時推定する
Step7		新作項目とする項目を入れ替えて Step5 以降を繰り返す

### 3.3 シミュレーションの結果

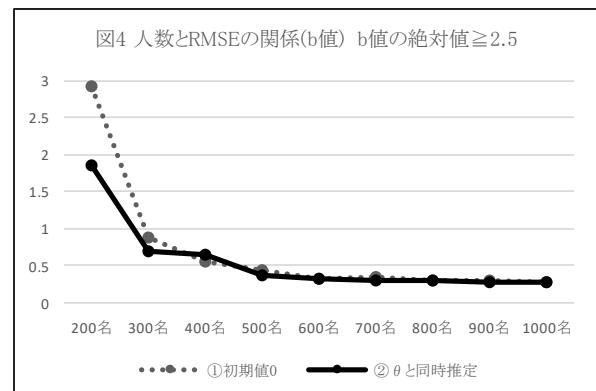
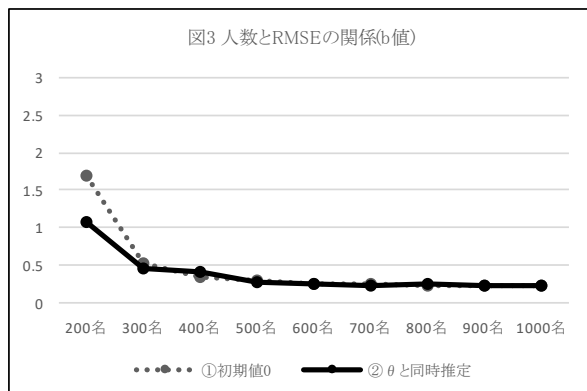
受検者の人数を 200 名から 1000 名に変化させながらシミュレーションをおこない, 項目特性値の推定結果と Step2 で設定した項目特性値の RMSE について確認した。

まず本番テスト項目の正誤の傾向を利用して新作項目の  $b$  値の初期値を決めた場合 (仮説 1) の  $b$  値の推定結果が受検人数によりどう変化するかを図 1, 図 2 にまとめた。



$b$  値の推定に関して, 全項目でみると 200 名のときに③の誤差がわずかに小さいくらいで手法による差はほとんどなかった.  $b$  値の絶対値が大きな項目に絞ってみると 200 名, 300 名のときに③の誤差が小さくなっている. なお  $a$  値については, 手法間の差はほとんどなく 500 名程度で推定は安定する (紙幅の都合で図表を省略).

次に受検者の能力特性値 $\theta$ と新作項目の項目特性値を同時推定した場合の  $b$  値の推定結果が人数によりどう変化するかを図 3, 図 4 にまとめた。



$b$  値の推定に関して、全項目でみても  $b$  値の絶対値が大きな項目に絞ってみても 200 名のとくに誤差がかなり小さくなっている一方、300 名以上では  $b$  値の初期値を 0 として推定する従来手法（仮説 1 ①）とほとんど差がなかった。なお  $a$  値については、受検者人数の多寡にかかわらず推定値が従来手法より高めになる傾向がみられた（紙幅の都合で図表を省略）。

#### 4 まとめと考察

仮説 1, 2 のいずれについても、受検者人数が 200 名のとくに誤差が小さくなっており、人数が少ないときの誤差が小さくなる効果はみられたが推定が安定する人数について効果は確認できなかった。  $a$  値が安定する 500 名以上では従来手法との差はほとんどなく、実用的には従来手法で十分であるということがわかった。

なお仮説 2 において、  $a$  値がやや高めに出る点については、受検者の能力特性値 $\theta$ の推定に新作項目の正誤の情報が加味されてしまうことが原因と考えられる。

一連の研究により、推定手法を工夫しても  $b$  値の絶対値が大きな項目の推定精度を上げることは困難であることがわかった。 今後は新作項目の項目特性値を簡易的に推定し、推定した  $b$  値近辺の能力特性値をもつ受検者に出題するといった出題方法の工夫について検討していきたい。

#### 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
- [3] 杉山 剛 (2019) : 「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—能力特性値の標準誤差情報の利用—」 日本テスト学会第 17 回大会発表資料集
- [4] Lord, F. M. (1980). Applications of item response theory to practical testing problems. Routledge.