

項目間の依存性が等化係数ならびに受検者集団の能力推定値に及ぼす影響

○坪田 彩乃

名古屋大学大学院教育発達科学研究科

1 問題と目的

項目応答理論 (IRT) を用いたテストを行うとき、テストの項目間には、“局所独立性の仮定”が必要となる。しかし、実際のテストでは、様々な要因によって、そうした局所独立性の仮定が成り立たない場面も存在している。厳密には、こうした場面でのテストに、IRT を当てはめることは出来ない。そのため、このような局所依存性をもつ項目を含むテストについて、IRT を当てはめるために、項目間の局所依存性を考慮したモデルの検討がなされてきた (Jannarone, 1986; Ip, 2002; Wainer, Bradlow, & Wang, 2007)。また、テストに含まれる局所依存性を検出するための手法も数多く検討がなされてきた (Chen & Thissen, 1997; Houts & Edwards, 2013)。しかし、実際のテスト場面において、全てのテストについて、全ての局所依存性を検出し、更にそれを考慮してモデルを選択していくことは現実的には不可能である。それぞれのテストにおいて、あらゆる依存性を考慮して数多くあるモデルの中から適切であると考えられるただ一つのモデルを正しく選択する必要があるとすれば、IRT を用いたテストのもつ有用性をもってしても、その手続きの煩雑さから敬遠される恐れもある。

さて、IRT を用いたテストの利点として、項目の異なる複数のテスト間で、共通尺度を用いた評価することを可能とすることが挙げられる。これを可能にする方法の一つに、共通項目による等化係数を用いた等化が挙げられる。それぞれのテストに共通な項目を用いて、等化係数を推定する。そして推定された等化係数に基づき、複数のテストの項目パラメタ値ならびに受検者の潜在特性値

を共通尺度上の値に変換するという方法である。等化を行う前に、それぞれのテストについて項目パラメタならびに受検者の潜在特性値を推定する。項目間に局所依存性をもつ項目があるとき、項目パラメタ値の推定には影響を及ぼす (登藤, 2012a) 一方で、潜在特性値の推定には、適合度の観点からは局所依存性を考慮したモデルに劣るものの、最小二乗誤差や相関などといった指標では局所依存性をもつモデルに遜色のない結果となる (登藤, 2012b) という。しかし、等化係数を算出するには項目パラメタの情報を参照するため、テストに局所依存性をもつ項目を含んでいるとき、その影響を受ける可能性が考えられる。

そのため本研究では、テスト間等化を行うための共通項目、ならびにそれぞれのテストの独自項目に依存性を持つとき、等化後の受検者集団の潜在特性推定値にどのような影響を及ぼすかを明らかにする。

2 方法

共通項目と独自項目のうち、どの項目に局所依存性をもつかによって分類した 4 つのパターンに対し、それぞれのテストを規準集団と等化用集団が受験したと仮定し合計 4×2 種類のテストを想定し、それら全てに 2PLM を当てはめて項目パラメタの推定ならびに等化用集団の等化後の潜在特性値を推定した。

2.1 想定するテストモデル

テストに用いる項目数はそれぞれ 60 項目 (共通項目 18 項目, 独自項目 42 項目) とした。また、受検者の人数は、1000 名を想定した。

(i) 全ての項目で、項目間の局所独立性を仮定できるテスト (LI), (ii) 共通項目のみに局所依存

性を持ち、独自項目では局所独立性を仮定できるテスト (CLD), (iii) 全ての項目で、項目間に局所依存性をもつテスト (LD), (iv) 独自項目のみに局所依存性を持ち、共通項目では局所独立性を仮定できるテスト (ULD) の 4 種類のテストを想定した。

項目反応モデルは以下に述べる Two-Parameter Logistic Model(2PLM) を仮定する。

$$P_j(\theta_i) = \frac{1}{1 + \exp[-1.7a_j(\theta_i - b_j)]} \quad (1)$$

$$Q_j(\theta_i) = 1 - P_j(\theta_i) \quad (2)$$

$P_j(\theta_i)$ は潜在特性値が θ_i である個人が項目 j に正答する確率を表している。 θ_i は個人 i の潜在特性値を、 a_j は項目 j の識別力母数を、 b_j は項目 j の困難度母数を、それぞれ表している。

依存性をもつ項目は、3 項目を 1 セットとし、先の項目に誤答したとき、セット内の後の項目に必ず誤答する依存性をおいた。局所依存性をもつ 1 つのセットでの正誤パターンならびに反応確率は、以下の 4 通りとなる。

(1, 1, 1)	$P_1(\theta_i)P_2(\theta_i)P_3(\theta_i)$
(1, 1, 0)	$P_1(\theta_i)P_2(\theta_i)Q_3(\theta_i)$
(1, 0, 0)	$P_1(\theta_i)Q_2(\theta_i)$
(0, 0, 0)	$Q_1(\theta_i)$

2.2 識別力母数と困難度母数

項目識別力は、石塚・中畝・内田・前川 (2001) において多くの項目識別力をが 0.3 から 0.9 に収まっていたことより、全ての項目において $U(0.3, 0.9)$ とした。

項目困難度は、項目間で局所独立性を仮定できるときは $U(-1.0, 1.5)$ とし、依存性をもつときは、1 セットである 3 項目 (b_1, b_2, b_3) に対して、(i) b_1 を $U(-1.25, -0.75)$, b_2 を $U(-0.25, 0.25)$, b_3 を $U(0.75, 1.25)$, (ii) b_1 を $U(0.25, 0.75)$, b_2 を $U(0.75, 1.25)$, b_3 を $U(1.25, 1.75)$ とした。また、依存性をもつ項目では、(i), (ii) のどちらのパターンも同じセット数にした。

2.3 受検者の潜在特性値

受検者の潜在特性値は、規準集団と等化用集団で異なる集団であることを仮定した。具体的には、規準集団として、平均 0、標準偏差 1 の正規分布を、等化用集団として、平均 1、標準偏差 0.5 の正規分布を仮定した。

なお、それぞれの集団のみについて、潜在特性値を推定すると、平均 0、標準偏差 1 の正規分布に従うように推定される。そのため、等化用集団の潜在特性値を、規準集団を用いて等化すると、等化後の潜在特性値を算出するための等化係数ならびに一次変換式は、 $y = 1/2x + 1$ が理想の値となる。等化係数の算出には、Mean & Mean(MM) 法 (Loyd & Hoover, 1980), Mean & Sigma(MS) 法 (Marco, 1977), Haebara(1980) による方法 (HB), Stocking & Lord(1983) による方法 (SL) の 4 種について検討した。

2.4 シミュレーションの手続き

Step1 規準集団の受検者の潜在特性値の真値 θ_C を標準正規分布 $N(0, 1)$ より、等化用集団の受検者の潜在特性値の真値 θ_E を標準正規分布 $N(1, 0.5)$ より発生させた。

Step2 項目識別力と項目困難度を先に述べた一様分布に基いて発生させた。

Step3 それぞれのテストについて項目の正答確率を 2PLM で算出した。

Step4 回答パターン作成のための一様分布 $U(0, 1)$ に基づき、Step3 で求めた確率に応じて回答パターン行列を作成した。

Step5 こうして発生させたデータについて項目パラメタ値の推定を行い、その推定値を用いて規準集団の潜在特性値 $\hat{\theta}_C$ ならびに等化用集団の潜在特性値の推定を行った。

Step6 事前に設定した共通項目を用いて、等化係数ならびに等化用集団の等化後の潜在特性値 $\hat{\theta}_E$ を算出した。

Step7 Step1~6 を 100 回行った。

2.5 用いた指標

等化用集団の等化後の能力値について、 $\overline{Bias}$, $\overline{RMSE}$, $\overline{cor\theta_E}$, $\overline{cor\theta_A}$ を用いた。

$\overline{Bias}$ は、真値に対して推定値が全体として正負のどちらの方向にあるかを示す。 $\overline{RMSE}$ は、真値に対して推定値が全体的にどのくらい離れているのかを示す。それぞれの式は以下の通りとなる。

$$Bias(\hat{\theta}) = \frac{1}{100} \sum_{r=1}^{100} (\hat{\theta}_r - \theta) \quad (3)$$

$$RMSE(\hat{\theta}) = \sqrt{\frac{1}{100} \sum_{r=1}^{100} (\hat{\theta}_r - \theta)^2} \quad (4)$$

ここで、 θ は等化用集団における受検者の潜在特性値の真値を表し、 $\hat{\theta}$ は等化用集団における等化後の潜在特性値の推定値を表す。また、 $\hat{\theta}_r$ はそれらの r 回目の推定値をそれぞれ表す。

また、 $\overline{cor\theta_E}$ は、等化用集団の能力値真値と推定値の相関であり、真値と推定値の線型性を示す。 $\overline{cor\theta_A}$ は、規準集団と等化用集団を合わせた全ての受検者集団についての能力値真値と推定値の相関であり、真値と推定値の線型性を示す。

等化係数について、 $\bar{a}, a_{SD}, \bar{b}, b_{SD}$ を用いた。 a, b は変換式 $y = ax + b$ の係数ならびに切片に対応する。 $\bar{a}$ は a の値の 100 回分の平均を、 a_{SD} はその分散を、 $\bar{b}$ は b の値の 100 回分の平均を、 b_{SD} はその分散を表す。

3 結果と考察

結果を表 1、表 2 に示した。等化法について、MM 法では、テスト項目に局所依存性をもつ項目があるテスト条件において、推定精度が低くなった。また、等化係数についても、 $\bar{b}$ について一貫して 1.0 よりも大きな値を示しており、 b_{SD} も大きくなった。MM 法では、項目の識別力母数の平均を利用して等化係数を算出する。しかし、項目に局所依存性をもつとき、依存性をもつ項目に対して、項目識別力は過大推定される (登藤, 2012a)。このことより、依存性をもつテストでは MM 法は十分な機能を果たさないと考えられる。

3.1 等化後の能力値の推定について

表 1 より、 $\overline{RMSE}, \overline{cor\theta_E}, \overline{cor\theta_A}$ について、それぞれの条件ごとに見ていくと、LI で相対的に最も推定精度が良く、次いで ULD, CLD, LD となった。 $\overline{cor\theta_E}$ が全ての等化法で同一の値であることは、等化前の受検者の潜在特性値が全ての等化法で同一であるためである。また、 $\overline{cor\theta_A}$ に比べて $\overline{cor\theta_E}$ の値が小さくなっていることは、切断効果によるものだと考えられる。

3.2 全体的な観点より

MS 法では、共通項目に局所依存性をもつ CLD 条件と全ての項目に依存性をもつ LD 条件において、 $\overline{RMSE}$ が大きくなった。等化係数についても、MS 法では一貫して $\bar{a}$ は 0.5 に近いもの

の、 $\bar{b}$ の値が小さく、 b_{SD} が大きくなった。これにより、一貫して $\hat{\theta}_E$ を過小評価しており、それが $\overline{RMSE}$ を大きくした原因であると考えられる。

HB 法と SL 法では、 $\overline{Bias}$ に差はあるものの、 $\overline{RMSE}, \overline{cor\theta_A}$ では全ての条件で近い値となった。等化係数については、共通項目に局所依存性をもつ CLD 条件と全ての項目に依存性をもつ LD 条件において、 $\bar{a}$ について 0.5 よりやや大きくなっており、 $\bar{b}$ について 1.0 よりやや小さくなった。また、一貫して HB 法が SL 法よりも小さい値となった。しかし、MS 法に比べ推定精度は高く、 a_{SD}, b_{SD} も小さいことより、共通項目に依存性をもつテストではこれらの方法が安定していると考えられる。

3.3 今後の課題

本研究では、等化用集団についての等化後の $\hat{\theta}_E$ ならびに等化係数について検討を行った。その結果、共通項目に依存性をもつテストでは、推定精度は他の条件に劣るものの、HB 法や SL 法といった特性曲線を用いた等化係数の算出法が望ましいことが示唆された。しかしながら、これらの結果が現実場面での実用水準にあるかについては検討されていない。テスト項目が局所依存性をもっている以上、規準集団の $\hat{\theta}_C$ についてもその影響が生じているはずである。今回検討された $\hat{\theta}_E$ についての推定精度について、局所依存性をもつ項目を含むテストであることによって生じた誤差と、局所依存性をもつ項目が等化係数の推定に影響を及ぼしたことによって生じた誤差とを弁別することは難しい。そのため、規準集団での推定精度についても併せて検討する必要性が考えられる。加えて、これらについて局所依存性を仮定できるモデルとの比較検討を行い、等化場面における局所独立性の仮定の頑健性についての検討も必要であろう。

また、実際のテスト場面では、共通項目、独自項目のうち、全ての項目について依存性が生じていることは考えにくい。すなわち、共通項目の一部、独自項目の一部といった依存性の生じ方を考えた方がより現実的である。こうしたパターンについても、推定精度について検討を行うと同時に、実際のテストについても当てはめて検討していく必要があるだろう。

表 1 それぞれの条件における $\overline{Bias}$, $\overline{RMSE}$, $\overline{cor\theta_e}$, $\overline{cor\theta_a}$

	LI				CLD				LD				ULD			
	MM	MS	HB	SL	MM	MS	HB	SL	MM	MS	HB	SL	MM	MS	HB	SL
$\overline{Bias}$	0.078	0.009	0.027	0.039	1.178	-0.297	-0.224	-0.101	0.376	-0.234	-0.086	0.044	0.250	0.070	0.112	0.126
$\overline{RMSE}$	0.327	0.302	0.304	0.309	1.580	0.581	0.541	0.522	0.810	0.662	0.595	0.613	0.509	0.412	0.414	0.418
$\overline{cor\theta_E}$	0.861	0.861	0.861	0.861	0.767	0.767	0.767	0.767	0.695	0.695	0.695	0.695	0.732	0.732	0.732	0.732
$\overline{cor\theta_A}$	0.953	0.952	0.952	0.952	0.871	0.886	0.894	0.902	0.873	0.847	0.867	0.873	0.915	0.914	0.916	0.917

表 2 それぞれの条件における $\bar{a}$, a_{SD} , $\bar{b}$, b_{SD}

	LI				CLD				LD				ULD			
	MM	MS	HB	SL	MM	MS	HB	SL	MM	MS	HB	SL	MM	MS	HB	SL
$\bar{a}$	0.509	0.471	0.492	0.499	0.978	0.518	0.563	0.613	0.704	0.512	0.590	0.628	0.462	0.408	0.446	0.450
a_{SD}	0.031	0.063	0.037	0.035	0.426	0.162	0.139	0.121	0.067	0.166	0.100	0.084	0.036	0.095	0.042	0.040
$\bar{b}$	1.064	0.997	1.014	1.025	2.060	0.656	0.725	0.842	1.279	0.696	0.833	0.958	1.201	1.026	1.064	1.078
b_{SD}	0.087	0.083	0.062	0.059	1.690	0.261	0.194	0.140	0.269	0.350	0.153	0.119	0.555	0.128	0.067	0.059