【個人研究】

能力分布の位置と散布度の推定について
─小規模データの場合─

藤　森　　　進*

Estimation of the Location and Dispersion of Ability Distribution: In Cases of Small Data Sets

Susumu FUJIMORI

In the present study, suitable methods to estimate the location and dispersion of ability distribution were investigated in accordance with the equalization of tests based on the common-examinee design. In particular, with regard to cases that involve little data, where the form of the ability distribution does not necessarily follow a normal distribution, using robust methods of estimation — such as median and quartile deviation — or Mislevy's methods of estimation — such as population mean and population variance — may be more appropriate than using the arithmetic mean and variance of the estimates of ability parameters. The results of the simulations confirmed that the estimation results obtained from the arithmetic mean of the estimates of ability parameters were favorable for estimating the location of ability distribution, while the results obtained from quartile deviation and Mislevy's estimation of population variance were favorable for the dispersion of ability distribution.

Key words : item response theory, ability distribution, common-examinee design, equating

1. 研究の目的

同じ能力特性や性格特性を測定するテスト版が２つあるときに、両者を同一人物が受験したとする。同一人物の同一特性を測定しているのであるから２つのテストの結果は同一とならなければならない。しかし、２つのテストは問題の違いがあるため、同一の特性値に対して与える得点が異なることが一般的である。このような２つの異なる項目からなるテスト結果を比較するためには、等化（equating）と呼ばれる調整を施す必要がある。

このようなテスト得点の処理に関しては、大きく分けて古典的テスト理論（classical test theory）の枠組みに立つものと、項目反応理論（item response theory）の枠組みに立つものの２系統がある。

この研究では、項目の困難度などの特性とテスト受験者の能力などの特性を相互に依存

*ふじもり　すすむ　　文教大学人間科学部人間科学科

しないで定義できるなどの特徴を持つ項目反応理論を議論の対象とする。

さて、２つのテスト版を受験する集団が同一のときの等化法に関しては、等パーセンタイル法（equipercentile equating method）と線形等化法（linear equating method）の２種類に分類できる（池田, 1994）が、ここでは２つの得点間に線形１次式を仮定する線形等化法を検討の対象とする。

等パーセンタイル法は、２つのテスト得点のパーセンタイル順位が一致するように２つのテスト得点を変換するものであり、２つの得点分布が同一の形状でない限り１つの変換式で得点を換算することができず、このため一般的には換算表によらねば等化ができないなどの不便さがある。このような理由により、ここでは検討の対象としなかった。

さて、項目反応理論では、同一受験者の２つのテストによる結果を θ と θ^* とするとき、両者の間に

$$\theta^* = k\theta + \ell \tag{1}$$

なる関係が仮定される（野口, 1991）。ここで k と ℓ は等化係数であり、前者はテスト得点の分散、後者は平均値にかかわっている。すなわち項目反応理論では、２つのテスト得点には平均と分散の違いがあるだけで、線形的な関係を持つことが仮定されることから線形等化法を用いるのが自然である。

本研究では、項目反応理論を利用して小規模データのテスト結果を線形等化法に従い等化する際に、中央値や四分偏差（四分領域とも呼ぶ）などのような順序統計量を利用する可能性について検討する。線形等化であるにもかかわらず順序統計量を用いることを検討するのは、小規模データにおける順序統計量のロバスト性に関心を持つからである。

本研究での「小規模データ」とは、テストの受験者数として50～100程度を想定している。項目反応理論を利用したテストの等化作業において、この程度の規模のデータが出現するケースとしては、２つの異なったテスト版を、等化のため特別に集めた受験者に実施する共通受験者デザイン（common-examinee design）において発生することが考えられる。このように特別に集められた集団の特徴は、当然のことながら人数も比較的少数であり、当該のテストの本来の対象とする母集団からの標本とは見なしえない、すなわち偏りのあるケースもあるであろう。共通受験者デザインを利用して２つのテストの等化を行おうとする場合、受験者はテストを２個受験することになり、他の一般的な受験者の倍の負担が掛かることになり、この点からも少数にならざるを得ないのである。また、受験者の集め方によっては、能力分布として正規分布というよりは比較的狭い範囲の一様分布に近いものとなっている可能性もある。たとえば、特定の学校からだけ受験者を募集したりすれば、輪切り現象に象徴される学校の層別化された今日においては、一様分布となる可能性は小さくない。このような状況において共通受験者デザインを利用して等化を行おうとする場合、順序統計量のロバスト性は等化の成績に良い効果があると期待できよう。本研究は、以上のような目的をもってシミュレーションにより行われた。

2. 方法

2-1. 項目反応モデル

(2)式は、項目反応モデルに属するものの中で最も良く利用される２母数ロジスティックモデル（Birnbaum,1968）である。

$$P\left(x_{ij}=1\middle|\theta_i,a_j,b_j\right)=\frac{1}{1+\exp\left(-Da_j\left(\theta_i-b_j\right)\right)} \tag{2}$$

ここで i は受験者、θ はその能力の大きさを表す能力母数であり、D は1.7の定数、j はテスト項目の番号、a_j は項目が異なる能力水準を弁別する力と関係する識別力母数、b_j は項目の困難度を表す困難度母数である。項目の特性に関する母数が２つあるため、２

母数ロジスティックモデルと呼ばれる。x_{ij} は、受験者 i の項目 j に関する正誤を表し、正答のとき１、誤答のとき０をとるダミー変数である。(2)式は、項目 j に受験者 i が正答する確率を表現している。項目反応理論に関してより詳しくは、例えば藤森（2002）などを参照のこと。

本研究では、学力テストを想定して論を進めるが、質問紙に関しても、１次元を仮定して回答結果による特性の大きくなる方向を適宜揃えれば項目反応モデルを適用することに大きな困難さは存在しない。

2-2. 能力母数の推定

本研究のシミュレーションでは、後述するように項目母数は所与として能力母数の推定についてのみ検討する。従ってここでは、能力母数の推定についてのみ述べる。能力母数の推定には、色々な方法があるが、一般的には､最尤法あるいはベイズ法による。項目母数を所与としたとき能力母数を最尤推定するためには、(3)式の対数尤度を最大にする θ を求める。

$$\ell_i = \sum_{j=1}^{n} \left\{ x_{ij} \ln\left(P_j\left(\theta_i\right)\right) + \left(1 - x_{ij}\right) \ln\left(Q_j\left(\theta_i\right)\right) \right\} \quad (3)$$

ここで n はテストの項目数、P は(2)式の２母数ロジスティックモデルで表現される正答確率であり、Q=1－P は誤答確率である。本研究では､事前分布を必要とするベイズ推定は、特に各受験者に関して事前分布を定める情報が存在しないため採用しなかった。最尤推定で能力母数の推定を行う場合、全問正答や誤答などの受験者の能力母数の推定値を求めることはできないが、ここでは該当のケースに関してはプログラムの能力母数の制限（上限3.5、下限－3.5）により推定値を充てている。なお能力母数の推定は、自作のfortranプログラムを利用して行った。

2-3. 能力推定値を利用した等化法

２つのテストを等化する場面を想定する。同一受験者の２つのテストによる結果を θ と θ^* とするとき、両者の間に(1)式が成立する。モデルが完全にデータと当てはまっており、能力母数の推定値に誤差がなければ、能力水準の異なる２人のデータが存在すれば２元連立方程式を解くことにより等化係数 k と ℓ を求めることができるが、一般にはモデルの完全な当てはまりも推定値に誤差がないことも期待できない。この問題の解決方法には様々なものがあるが、ここでは簡単にMarcoの方法（Marco,1977）

$$\hat{k} = \frac{sd\left(\theta^*\right)}{sd\left(\theta\right)} \quad (4)$$

$$\hat{\ell} = \overline{\theta}^* - \hat{k}\overline{\theta} \quad (5)$$

を利用した。ここで、$sd\left(\theta^*\right)$ と $sd\left(\theta\right)$ は、テスト受験者の θ^* あるいは θ の推定値の標準偏差であり、$\overline{\theta}^*$ と $\overline{\theta}$ は、テスト受験者の θ^* あるいは θ の推定値の平均値である。

すなわち、等化係数 k と ℓ を求めるには、推定値の平均、標準偏差を知る必要がある。本研究では、これを中央値、四分偏差に置き換える可能性を検討する。

中央値は漸近的に

$$N\left(\mu_m, \frac{pq}{nf\left(\mu_m\right)^2}\right) \quad (6)$$

に従うことが知られている。ここで、μ_m は、母集団中央値であり、$f\left(\mu_m\right)$ は、その密度関数値、n は標本の大きさ、p は μ_m 以下の確率=0.5、$q = 1 - p$ である。データが正規分布の場合、平均値は標本分布の標準偏差 $\sigma/\sqrt{n}$ が中央値などと比較して小さいわけであるから、中央値などを利用する必要はない。

しかし、分布型が非正規である場合は、その限りでないこともあり、共通受験者デザインでは能力の一様性があるケースが存在するのではないかという観点から本研究は行われ

ている。

なお、標準偏差の不偏推定値は、

$$\sqrt{\frac{n}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\boldsymbol{s} \tag{7}$$

となり、これも等化係数の推定に利用することが考えられる。ここでsは、標本標準偏差である。四分偏差の場合は、(6)式でμ_mを第１四分位数、第３四分位数と置き換えれば漸近分布が得られるので、そのまま

$$\frac{Q_3-Q_1}{2} \tag{8}$$

を推定値とできるであろう。

また、項目反応理論の枠組みでは、Mislevy(1984)による母集団分布の母数推定の方法がある。N人のテストデータが得られたとき、能力母数θを積分によって排除したμ、σの周辺化された尤度は、

$$L\left(\mu,\sigma\middle|x_1,x_2,\cdots,x_N\right)=\prod_{i=1}^{N}\left\{\int_{-\infty}^{\infty}f\left(x_i\middle|\theta\right)h\left(\theta\middle|\mu,\sigma\right)d\theta\right\} \tag{9}$$

で表現される。ここで、f は、

$$f\left(x_i\middle|\theta\right)=\prod_{j=1}^{n}P^{x_{ij}}\left(1-P\right)^{1-x_{ij}} \tag{10}$$

であり、x_iは、受験者iの正誤を表すn個の成分x_{ij}からなる列ベクトルである。(9)式を最大にするμ、σを数値計算によって求め、母集団平均と標準偏差の推定値とするのがMislevyの方法である。この方法は、θを介さずに、正誤パタンから直接推定するものである。これも等化係数の推定に利用することが考えられよう。

3. シミュレーション

3-1. データの作成

シミュレーションのデータは、以下のようにして作成した。

項目数は、全て40項目とする。学力検査としては標準的な項目数であろう。テスト項目の正誤は全て２母数ロジスティックモデルに従って発生し、各項目母数は以下の分布型に従っていると仮定する。

識別力母数aは、平均0.8、標準偏差0.25、下限0.3、上限2.0の切断正規分布、また困難度母数bは、平均０、標準偏差0.5の正規分布に従うと仮定した。

被験者数は50、100、あるいは500人とする。研究の対象としているのは50から100人規模のデータであるが、人数の影響を確認するため500を追加した。能力θの分布は、標準正規分布$\boldsymbol{N}(0,1^2)$あるいは正規分布$\boldsymbol{N}(0.5,1.25^2)$に従うと仮定した。$\theta$の能力分布の平均0は、困難度$b$の平均と一致している。また、分布型の影響を見るため、区間［－0.5, 0.5］と［0, 1］の一様分布に従う能力分布を作成した。区間［－0.5, 0.5］の一様分布の平均は、困難度bの平均と一致している。一様分布の区間幅を1とした理由は、経験上、共通受験者の能力分布として、テストの一般的な受験者の標準偏差の1/4程度となる例が少なからず存在することによる。

能力母数θの者の各項目に対する正誤は、２母数ロジスティックモデルから予想される正答確率を、範囲０～１の一様乱数と比較し、前者が下回る場合に正答１、上回る場合誤答０とする。

以上のような方式で、２値データパターンを、各５回繰り返し作成し(データＤ1～Ｄ5)、テストデータとした。

3-2. 分析の指標

各データとも項目母数は所与として真値をそのまま充て、能力母数を推定し、その平均値、標準偏差を求める。これと比較するため、中央値及び四分偏差を求めた。(7)式による標準偏差の不偏推定値は、不偏分散及び標本分散の平方根より大きい。このため結果では推定値の標本標準偏差のみ示している。これ

らの値の平均二乗誤差の平方根（root mean square error）を求めて、推定の成績の指標としている。

Marcoの方法の(4)式が標準偏差の比となっていることを考慮し、θの推定値$\hat{\theta}$の標準偏差と真値θ_tの標準偏差の比

$$r_{sd} = \frac{sd\left(\hat{\theta}\right)}{sd\left(\theta_t\right)} \tag{11}$$

及び、θの推定値$\hat{\theta}$の四分偏差$\hat{Q}$と真値の四分偏差Q_tの比

$$r_Q = \frac{\hat{Q}}{Q_t} \tag{12}$$

を求めた。推定値に誤差がなければ、これらは１となるのであるから、これらの比が１から離れる程度を推定成績の指標とすることにした。

4. 結果と考察

表１から表６は、θの真値の平均が0.0である場合のシミュレーション結果である。すなわち、θの平均と困難度の平均が一致しており、テストの実施の上で望ましい状況にあると考えられる。表１から表３は、θの母集団分布が標準正規分布であるのに対して、表４から表６は区間［－0.5,0.5］の一様分布の結果である。

表１は、受験者数50の場合であり、表中の平均とあるのはθの真値あるいは推定値の平均であり、SDは同じくθの標準偏差、Q２とあるのは中央値、Qは四分偏差である。また、表中の平均（M）、SD（M）とあるのは、Mislevyによるθの母平均と母標準偏差の推定値である。D１からD５は、データセットを表し、データ平均とあるのは、５つのデータセットの平均をとっていることを意味する。分析指標は、表の最下段にまとめてあり、$\sqrt{\mathrm{MSE}}$とあるのは、表中の真値欄の平均、SD、Q２、Qなどの値と推定値欄の値の差を各データごとに２乗して和を作り、５つのデータセットの平均を取って、平方根を求めたものである。またrは、各データごとに(11)式、あるいは(12)式を求め、５つのデータセットの平均をとったものである。表中の値の意味は、表12まで同様である。

さて表１の分析指標に注目すれば、分布の位置の指標に関してはθの推定値の単純な平均値の$\sqrt{\mathrm{MSE}}$が最も小さく、Mislevyによるθの母平均推定値、中央値の順となる。

分布のバラツキの指標rに関しては、四分偏差が最も成績がよく、Mislevyによるθの母標準偏差推定値とθの推定値の単純な標準偏差が、ほぼ同程度で続いている。

表２は、人数が100人となった場合であるが、分布の位置の指標に関しては表１と同様であるが、分布のバラツキに関しては、Mislevyが最もよく、四分偏差がほぼ同程度であり、θの推定値の単純な標準偏差は最も成績が悪い。

表３は、人数500となった場合であり、分布の位置の指標に関しては表１、表２と同様であるが、分布のバラツキに関しては、Mislevyが最もよく、四分偏差、θの推定値の単純な標準偏差の順となる。

以上が分布型が正規分布の場合であるが、続いて一様分布の場合を検討してみよう。表４は、人数が50の場合であるが、分布の位置の指標に関しては表１と同様の結果である。分布のバラツキに関しては、表３と同様の傾向である。

続いて、表５の人数が100の場合であるが、分布の位置の指標に関しては表１と同様の結果である。分布のバラツキに関しては、四分偏差が最も成績がよく、Mislevyによるθの母標準偏差推定値、θの推定値の単純な標準偏差の順となる。

表６は、人数が500となった場合であるが、分布の位置の指標に関しては表１と同様の結果である。分布のバラツキに関しては、表５

と同様である。

表１から表６の結果をまとめると、分布の位置の指標に関しては、分布型や人数の違いにかかわらず、θの推定値の単純な平均値の成績が良く、Mislevyによるθの母平均推定値、中央値の順となる。分布のバラツキの指標に関しては、四分偏差とMislevyによるθの母標準偏差推定値が同程度で成績がよく、θの推定値の単純な標準偏差の成績は悪いことが分かる。

表１．　θ標準正規分布、人数50の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	−0.101	0.083	0.192	0.102	−0.020	0.051
	SD	1.056	1.061	1.012	0.840	0.758	0.946
	Q2	−0.119	0.023	0.325	0.074	−0.073	0.046
	Q	0.579	0.769	0.738	0.552	0.481	0.624
推定値	平均	−0.056	0.117	0.160	0.078	0.026	0.065
	SD	1.138	1.069	1.202	0.930	0.788	1.025
	Q2	−0.182	0.225	0.256	0.072	−0.171	0.040
	Q	0.498	0.706	0.737	0.670	0.515	0.626
	平均(M)	−0.073	0.095	0.144	0.090	0.010	0.053
	SD（M）	1.039	0.953	1.063	0.856	0.704	0.923
		平均	SD	Q2	Q	平均(M)	SD（M）
	√MSE	0.014	0.080	0.109	0.072	0.037	0.152
	r		1.084		1.013		0.923

表２．　θ標準正規分布、人数100の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	−0.047	0.149	0.109	−0.054	−0.011	0.029
	SD	0.979	1.080	0.962	1.023	0.984	1.005
	Q2	−0.023	0.187	0.000	−0.133	−0.030	0.000
	Q	0.684	0.645	0.633	0.684	0.706	0.670
推定値	平均	−0.043	0.207	0.094	−0.092	−0.011	0.031
	SD	1.122	1.131	1.048	1.090	1.098	1.098
	Q2	−0.082	0.162	−0.001	−0.202	−0.014	−0.027
	Q	0.665	0.609	0.626	0.682	0.719	0.660
	平均(M)	−0.052	0.195	0.078	−0.104	−0.011	0.021
	SD（M）	1.013	1.009	0.962	0.994	0.997	0.995
		平均	SD	Q2	Q	平均(M)	SD（M）
	√MSE	0.002	0.092	0.043	0.020	0.032	0.018
	r		1.093		0.984		0.995

表３．　θ標準正規分布、人数500の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	−0.010	−0.018	−0.005	−0.041	0.023	−0.010
	SD	0.993	1.002	1.028	1.004	1.017	1.009
	Q2	0.020	−0.011	−0.001	−0.092	0.048	−0.007
	Q	0.667	0.689	0.709	0.729	0.659	0.690
推定値	平均	−0.030	−0.042	−0.003	−0.042	0.030	−0.017
	SD	1.095	1.122	1.136	1.078	1.099	1.106
	Q2	−0.005	−0.063	−0.039	−0.110	0.122	−0.019
	Q	0.674	0.759	0.759	0.712	0.725	0.726
	平均(M)	−0.018	−0.030	−0.011	−0.047	0.034	−0.014
	SD（M）	0.993	1.023	1.050	0.995	1.000	1.012
		平均	SD	Q2	Q	平均(M)	SD（M）
	√MSE	0.007	0.097	0.046	0.049	0.014	0.025
	r		1.096		1.052		1.012

表４．　θ一様分布［-0.5，0.5］、人数50の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	0.000	0.104	0.036	0.044	0.094	0.055
	SD	0.292	0.320	0.300	0.307	0.267	0.297
	Q2	−0.007	0.214	0.036	0.090	0.153	0.097
	Q	0.256	0.279	0.286	0.303	0.225	0.270
推定値	平均	−0.015	0.087	0.057	0.049	0.078	0.051
	SD	0.432	0.410	0.426	0.427	0.469	0.433
	Q2	−0.041	0.145	0.135	−0.020	0.103	0.064
	Q	0.339	0.266	0.372	0.273	0.401	0.330
	平均(M)	−0.007	0.087	0.061	0.041	0.074	0.051
	SD（M）	0.342	0.324	0.344	0.327	0.383	0.344

		平均	SD	Q2	Q	平均(M)	SD（M）
	√MSE	0.004	0.136	0.078	0.097	0.016	0.059
	r		1.465		1.253		1.191

表５．　θ一様分布［-0.5，0.5］、人数100の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	−0.055	0.001	0.001	0.018	−0.024	−0.012
	SD	0.287	0.287	0.303	0.281	0.310	0.294
	Q2	−0.093	0.005	−0.016	0.034	−0.049	−0.024
	Q	0.259	0.252	0.270	0.255	0.285	0.264
推定値	平均	−0.060	0.032	−0.039	0.012	−0.041	−0.019
	SD	0.415	0.387	0.385	0.377	0.429	0.399
	Q2	−0.097	0.081	−0.069	0.048	−0.109	−0.029
	Q	0.303	0.279	0.274	0.252	0.303	0.282
	平均(M)	−0.048	0.032	−0.039	0.016	−0.049	−0.018
	SD（M）	0.323	0.275	0.295	0.285	0.347	0.305

		平均	SD	Q2	Q	平均(M)	SD（M）
	√MSE	0.007	0.105	0.050	0.025	0.024	0.031
	r		1.358		1.069		1.057

表６．　θ一様分布［-0.5，0.5］、人数500の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	−0.002	0.006	−0.004	−0.005	0.018	0.003
	SD	0.288	0.291	0.285	0.287	0.286	0.287
	Q2	−0.007	0.004	−0.004	0.000	0.037	0.006
	Q	0.258	0.262	0.233	0.253	0.248	0.251
推定値	平均	−0.005	0.006	−0.004	−0.007	0.023	0.002
	SD	0.366	0.394	0.387	0.385	0.359	0.378
	Q2	0.015	0.008	−0.013	−0.014	0.011	0.001
	Q	0.264	0.282	0.263	0.273	0.243	0.265
	平均(M)	−0.001	0.006	0.000	−0.003	0.023	0.005
	SD（M）	0.258	0.304	0.297	0.283	0.261	0.280

		平均	SD	Q2	Q	平均(M)	SD（M）
	√MSE	0.000	0.091	0.017	0.019	0.003	0.020
	r		1.316		1.058		0.971

続いて、表７から表12のθの平均と困難度の平均に差がある場合について検討しよう。

分布の位置の指標に関しては、表７から表12の結果をまとめると、分布型や人数の違いにかかわらず、θの推定値の単純な平均値の成績が良く、Mislevyによるθの母平均推定値、中央値の順となり、これまでの結果と一致する。

表７．　θ正規分布（μ=0.5、σ=1.25）、人数50の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	0.194	0.463	0.489	0.728	0.566	0.488
	SD	1.149	1.242	1.193	1.194	1.208	1.197
	Q2	0.253	0.554	0.645	0.782	0.515	0.550
	Q	0.935	0.779	0.814	0.797	0.875	0.840
推定値	平均	0.211	0.404	0.602	0.682	0.646	0.509
	SD	1.319	1.347	1.279	1.161	1.464	1.314
	Q2	0.223	0.431	0.630	0.757	0.369	0.482
	Q	0.962	0.792	0.889	0.798	0.836	0.856
	平均(M)	0.188	0.376	0.590	0.674	0.596	0.485
	SD（M）	1.228	1.275	1.220	1.058	1.371	1.230
		平均	SD	Q2	Q	平均(M)	SD（M）
	$\sqrt{}$MSE	0.021	0.117	0.087	0.040	0.071	0.104
	r		1.098		1.019		0.984

表８．　θ正規分布（μ=0.5、σ=1.25）、人数100の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	0.437	0.375	0.415	0.319	0.672	0.444
	SD	1.302	1.312	1.175	1.188	1.313	1.258
	Q2	0.395	0.356	0.477	0.462	0.590	0.456
	Q	0.977	1.054	0.703	0.821	0.975	0.906
推定値	平均	0.449	0.392	0.424	0.334	0.713	0.462
	SD	1.356	1.400	1.229	1.326	1.448	1.352
	Q2	0.373	0.325	0.448	0.517	0.690	0.471
	Q	0.991	0.924	0.774	0.797	1.018	0.901
	平均(M)	0.446	0.371	0.405	0.348	0.687	0.451
	SD（M）	1.290	1.334	1.156	1.222	1.383	1.277
		平均	SD	Q2	Q	平均(M)	SD（M）
	$\sqrt{}$MSE	0.019	0.094	0.056	0.070	0.022	0.085
	r		1.075		1.001		1.022

表９．　θ正規分布（μ=0.5、σ=1.25）、人数500の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	0.437	0.500	0.481	0.523	0.500	0.488
	SD	1.294	1.280	1.198	1.211	1.259	1.248
	Q2	0.501	0.500	0.437	0.522	0.511	0.494
	Q	0.858	0.768	0.835	0.833	0.902	0.839
推定値	平均	0.431	0.513	0.504	0.549	0.510	0.502
	SD	1.384	1.342	1.293	1.280	1.340	1.328
	Q2	0.442	0.464	0.392	0.586	0.479	0.473
	Q	0.916	0.810	0.825	0.831	0.921	0.861
	平均(M)	0.430	0.491	0.479	0.524	0.500	0.485
	SD（M）	1.295	1.247	1.203	1.177	1.255	1.235
		平均	SD	Q2	Q	平均(M)	SD（M）
	$\sqrt{}$MSE	0.013	0.079	0.049	0.033	0.018	0.044
	r		1.064		1.026		0.988

分布のバラツキの指標に関しては、Mislevyによるθの母標準偏差推定値の成績がよく、続いて四分偏差、θの推定値の単純な標準偏差の順となる。Mislevyの方法は、正規分布を仮定した上での母標準偏差の推定であるが、能力分布が一様分布であっても良い推定成績を示したことは意外である。

以上をまとめると、分布の位置の指標に関しては、θの推定値の単純な平均値の成績が良く、Mislevyによるθの母平均推定値、中央値の順となり、中央値を利用する意味はないことが分かる。分布のバラツキの指標に関しては、Mislevyの方法、あるいは四分偏差の利用が望ましく、θの推定値の単純な標準偏差の利用は適当でないことが分かる。

以上の結論は、等化の成績を直接問題にしていないため、直ぐに実践的に適用できるものではないが、ひとつの可能性として今後検討する意味はあるものと思われる。なおシミュレーションのテスト項目数が40と比較的大きいためか、最尤推定であるにもかかわらず全問正答や誤答などのために能力母数の推定上の制限に抵触したケースは存在しなかった。本研究では問題とならなかったものの、仮にテストの項目数が少なくなるなどの事情により全問正答や誤答などのケースが生じた場合、θの推定値の標準偏差の算出などには影響が避けられないが、中央値や四分偏差の算出には不都合がないのは順序統計量の利点と言えよう。

表10.　θ一様分布［0.0, 1.0］、人数50の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	0.516	0.510	0.551	0.511	0.477	0.513
	SD	0.317	0.314	0.278	0.272	0.282	0.292
	Q2	0.472	0.444	0.599	0.487	0.444	0.489
	Q	0.316	0.310	0.196	0.217	0.247	0.257
推定値	平均	0.481	0.468	0.571	0.549	0.455	0.505
	SD	0.431	0.432	0.330	0.356	0.373	0.385
	Q2	0.475	0.365	0.619	0.501	0.483	0.488
	Q	0.385	0.332	0.217	0.258	0.242	0.287
	平均(M)	0.470	0.453	0.555	0.533	0.443	0.490
	SD（M）	0.345	0.334	0.226	0.248	0.281	0.287
		平均	SD	Q2	Q	平均(M)	SD（M）
	√MSE	0.008	0.092	0.041	0.038	0.032	0.047
	r		1.312		1.113		0.994

表11.　θ一様分布［0.0, 1.0］、人数100の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	0.500	0.485	0.459	0.492	0.435	0.474
	SD	0.303	0.296	0.290	0.278	0.274	0.288
	Q2	0.521	0.496	0.459	0.510	0.386	0.474
	Q	0.269	0.273	0.223	0.226	0.227	0.243
推定値	平均	0.537	0.471	0.487	0.525	0.445	0.493
	SD	0.397	0.417	0.388	0.353	0.359	0.383
	Q2	0.533	0.416	0.444	0.490	0.386	0.454
	Q	0.282	0.293	0.281	0.212	0.254	0.265
	平均(M)	0.521	0.455	0.467	0.509	0.425	0.476
	SD（M）	0.297	0.309	0.278	0.231	0.235	0.270
		平均	SD	Q2	Q	平均(M)	SD（M）
	√MSE	0.019	0.095	0.038	0.031	0.027	0.037
	r		1.327		1.089		0.935

表12.　θ一様分布［0.0, 1.0］、人数500の推定結果

		D１	D２	D３	D４	D５	データ平均
真値	平均	0.505	0.490	0.497	0.506	0.494	0.498
	SD	0.277	0.279	0.289	0.299	0.288	0.286
	Q2	0.518	0.493	0.488	0.505	0.492	0.499
	Q	0.222	0.236	0.245	0.270	0.245	0.244
推定値	平均	0.527	0.514	0.506	0.522	0.525	0.519
	SD	0.378	0.401	0.398	0.411	0.391	0.396
	Q2	0.521	0.503	0.498	0.507	0.537	0.513
	Q	0.251	0.285	0.264	0.287	0.268	0.271
	平均(M)	0.511	0.494	0.486	0.502	0.509	0.500
	SD（M）	0.268	0.271	0.276	0.309	0.289	0.283
		平均	SD	Q2	Q	平均(M)	SD（M）
	√MSE	0.020	0.110	0.021	0.030	0.022	0.016
	r		1.383		1.114		0.979

文献

Birnbaum, A.　1968 Some latent trait models and their use in inferring an examinee's ability. In F.M.Lord & M.R.Novick (Eds.), *Statistical theories of mental test scores*（pp. 395-479）. Reading, MA:Addison-Wesley.

藤森進　2002　テスト得点を統計的枠組みで分析する―項目反応理論―　渡部洋編「心理統計の技法」第７章　福村出版.

池田　央　1994　現代テスト理論　朝倉書店.

Marco,G.L.　1977 Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*,14, 139-160.

Mislevy,R.J.　1984 Estimating latent distributions. *Psychometrika*,49, 359-381.

野口　裕之　1991　項目反応理論にもとづくテストの作成法　芝祐順編「項目反応理論」第３章　東京大学出版会.