Tsukuba Psychological Research
2005, **29**, 41-45

一般化部分採点モデルの項目母数の等化

筑波大学心理学系　服部　　環

Equating item parameters in the generalized partial credit model

Tamaki Hattori (*Institute of Psychology, University of Tsukuba, Tsukuba 305-8572, Japan*)

The generalized partial credit model is useful for analyzing test items, scored as either correct/wrong or grade-based. This paper presents five methods of estimating equating coefficients for the item parameters of the generalized partial credit model under the common-item design: namely, the minimum chi-square method, the least squares method, the category characteristic function method, the true score method and the Mean & Mean method. The results of numerical experiments, that implemented horizontal and vertical equatings, suggest that the minimum chi-square method, the characteristic curve, the true score and the Mean & Mean methods are preferable to the least squares method in terms of estimate bias.

Key words: item response theory, the generalized partial credit model, equating, numerical experimentation

1　はじめに

項目反応モデルでは受検者の特性値の分布に依存しないで項目母数が尺度化される．しかし，測定の原点と単位に不定性があるので，現実のテストを運用する際には，テスト冊子ごとに測定の原点と単位を定めた上で項目母数値を推定する．このため，異なるテスト冊子にある項目の母数値をそのまま比較することはできない．そのような項目の母数値を比較可能とするためには，項目母数値を共通の尺度上へ変換する必要がある．この手続きを等化と呼ぶ．等化は項目プール（項目バンク）を作成するためには必須の手続きであり，等化が行われていれば任意の項目を用いて共通の尺度上で受検者の特性値を推定することができ，便利である．

項目母数を等化する方法として無作為受検者法，共通受検者法，共通項目法がある．

無作為受検者法は等質集団に異なるテスト冊子を実施するだけでよい．これは，等質の受検者集団であればテスト冊子が異なっても既に共通の尺度上で項目母数値が推定されている，という論理に基づく．しかし，実際には等質集団を確保したり，それを保証することが難しい．また，共通受検者法は複数のテスト冊子を一部の受検者集団に実施して，異なるテスト冊子から推定した受検者母数値を利用して項目母数を等化する．共通項目法はテスト冊子に共通項目を入れておき，それぞれのテスト冊子を異なる受検者集団に実施して項目母数を推定する．そして，共通項目の母数推定値を利用して項目母数を等化する．

ところで，項目反応モデルの１つに一般化部分採点モデル（Muraki, 1992）がある．このモデルは正答・誤答（はい・いいえ）と採点したテスト項目，もしくは段階点（多段階評定）を用いて採点したテスト項目を解析するために有用である．本稿は共通受検者法を適用して得られた一般化部分採点モデルの項目母数値を等化する方法を提案する．

2　モデルとモデル母数の変換

2．1　モデル式と母数の変換

一般化部分採点モデルのカテゴリ特性関数，つまり，特性値 θ_i の受検者が項目 j のカテゴリ k に反応する確率は次式のように定義される．

$$P_{jk}(\theta_i)=\frac{\exp\sum_{v=0}^{k}a_j(\theta_i-\delta_{jv})}{\sum_{h=0}^{m_j}\exp\sum_{v=0}^{h}a_j(\theta_i-\delta_{jv})} \tag{1}$$
$$(k=0,1,2,...,m_j)$$

ここで，a_jは項目の識別力，δ_{jv}は項目jでカテゴリvを取る難しさを示す項目母数（ステップ母数と呼ばれる）である．また，$\sum_{v=0}^{0}a_j(\theta_i-\delta_{jv})\equiv 0$と置く．

2つの定数$A(A>0)$とKを用いて式（1）のモデル母数を式（2），式（3），式（4）のように変換しても，カテゴリ特性関数値は変わらない．

$$\theta_i^*=A\theta_i+K \tag{2}$$

$$a_j^*=a_j/A \tag{3}$$

$$\delta_{jv}^*=A\delta_{jv}+K(v=1,2,...,m_j) \tag{4}$$

共通項目の母数値にまったく誤差がなければ，式（3）および式（4）の左辺に一方のテスト冊子で推定した母数値（等化のターゲットとなる尺度値），右辺に他方のテスト冊子で推定した母数値（等化させたい尺度値）を代入すれば，等化係数$A(A>0)$とKを求めることができる．しかし，実際には種々の原因による誤差があるので，いずれの等式も成り立たない．そこで，本稿はDivgi（1985）によってロジスティック・モデルに適用された最小χ^2法，Haebara（1980）とStocking & Lord（1983）によって提案された特性曲線法を用いて，等化係数$A(A>0)$とKを求める方法を提案する．

2.2 想定される状況

2.2.1 新しいテスト項目のみからなる場合

テスト冊子1とテスト冊子2を作り，冊子の間に共通項目を入れる．そして，それぞれのテスト冊子を異なる受検者集団へ実施して，テスト冊子ごとに項目母数を推定する．等化係数$A(A>0)$とKの推定値は，共通項目のテスト冊子1で得られた項目母数の推定値を式（3）および式（4）の右辺に代入して，また，テスト冊子2で得られた推定値を式（3）および式（4）の左辺に代入して，両辺の差異を最小化するという条件で求める．テスト冊子1独自の項目の母数値は，その母数値と$A(A>0)$とKの推定値を式（3）および式（4）の右辺して，テスト冊子2の尺度へ変換することになる．

テスト冊子2独自の項目母数値も同様の手続きにより，テスト冊子1の尺度へ変換できる．あるいは，テスト冊子1の項目母数値を等化するための係数$A(A>0)$とKを用いて，テスト冊子2独自の項目の母数値を式（5）および式（6）の右辺に代入してもよい．

$$a_j=a_j^*A \tag{5}$$

$$\delta_{jv}=\delta_{jv}^*/A-K/A(v=1,2,...,m_j) \tag{6}$$

2.2.2 母数値が既知の項目がある場合

これは既に項目プールがあり，新しく作成した項目をその項目プールへ追加したいという状況である．そのためには，項目プールの項目の一部と新項目を用いて1つのテスト冊子を作成して実施する．このとき項目プールの項目は母数値が既知であるから，理論的には，その値を固定して新項目の母数値を推定できる．この結果，既に新項目の母数値は項目プールの尺度上で推定されたことになり，改めて等化係数を推定する必要はない．

また，項目プールの項目と新項目の母数値を同時に推定して，項目プールでの推定値を式（3）および式（4）の左辺に代入し，改めて推定した母数値を式（3）および式（4）の右辺に代入する．そして，両辺の差異を最小化するという条件で等化係数$A(A>0)$とKを求めることもできる．そして，新項目の母数値と$A(A>0)$とKの推定値を式（3）および式（4）の右辺に代入して新項目の母数値を項目プールの尺度へ等化する．項目プールを作成したときよりも多くの受検者を確保できるなら，この方法も利用できよう．

3 等化法

本稿では，以下の最小χ^2法，カテゴリ特性関数法，真の得点法，単純最小自乗法，Mean & Mean法を提案する．

3.1 最小χ^2法

共通項目法を用いて2つのテスト（T_1とT_2）を実施したものとする．共通項目jのテストT_pにおける母数推定値をa_{pj}, δ_{pjv}とする．最小χ^2法は式（7）を最小化基準として等化係数を推定する．

$$\begin{aligned}\chi^2=\sum_{j=1}^{n}[&(a_{1j}-a_{2j}/A,\delta_{1j1}-A\delta_{2j1}-K,...,\\&\delta_{1jm_j}-A\delta_{2jm_j}-K)\\&(\textstyle\sum_{1j}+\sum_{2j}^{*})^{-1}\\&(a_{1j}-a_{2j}/A,\delta_{1j1}-A\delta_{2j1}-K,...,\\&\delta_{1jm_j}-A\delta_{2jm_j}-K)']\end{aligned} \tag{7}$$

ここで，n は共通項目数，Σ_j は a_{1j}, δ_{1j1}, $\delta_{2j1}, \cdots$, δ_{1jm_j} の推定値の漸近分散共分散行列，Σ^*_{2j} は共通項目 j の T_2 で推定された値を T_1 の尺度へ等化したときの値，すなわち，a_{2j}/A，$A\delta_{2j1}+K, \cdots A\delta_{2jm_j}+K$ の漸近分散共分散行列である．

なお，後述の数値実験ではPARSCALE（Muraki & Bock, 1991）を用いて項目母数を推定したが，PARSCALEは推定値の漸近共分散を出力しない．そのため，最小 χ^2 法で定義される式（7）法の基準式における Σ_{1j} と Σ^*_{2j} には，母数推定値の漸近分散のみを代入した．そのときの $\Sigma_{1j}+\Sigma^*_{2j}$ は次式となる．

$$
\begin{aligned}
&\sum\nolimits_{1j}+\sum\nolimits^{*}_{2j} \\
&=\begin{pmatrix}
\hat{\sigma}^2(a_{1j})+\hat{\sigma}^2(a_{2j})/A^2 & 0 & \cdots & 0 \\
0 & \hat{\sigma}^2(\delta_{1j1})+\hat{\sigma}^2(\delta_{2j1})A^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \hat{\sigma}^2(\delta_{1jm_j})+\hat{\sigma}^2(\delta_{2jm_j})A^2
\end{pmatrix}
\end{aligned}
\tag{8}
$$

$\hat{\sigma}^2$ (a_{pj}) は a_{pj} ($p=1,2$) の推定値の漸近分散，$\hat{\sigma}^2$ (δ_{pjv}) は δ_{pjv} ($p=1,2$) の推定値の漸近分散である．以上により，最小化すべき基準式は次式となる．

$$
\begin{aligned}
&\chi^2 \\
&=\sum_{j=1}^{n}[(a_{1j}-a_{2j}/A)^2[\hat{\sigma}^2(a_{1j})+\hat{\sigma}^2(a_{2j})/A^2]^{-1} \\
&\quad+\sum_{v=1}^{m_j}(\delta_{1jv}-A\delta_{2jv}-K)^2[\hat{\sigma}^2(\delta_{1jv})+A^2\hat{\sigma}^2(\delta_{2jv})]^{-1}]
\end{aligned}
\tag{9}
$$

未知数に関する基準関数の偏微分は以下の通りである．

$$
\begin{aligned}
&\frac{\partial\chi^2}{\partial A} \\
&=2\sum_{j=1}^{n}\big[(a_{1j}-a_{2j})/A)a_{2j}A^{-2}[\hat{\sigma}^2(a_{1j})+\hat{\sigma}^2(a_{2j})/A^2]^{-1} \\
&\quad+2(a_{1j}-a_{2j}/A)^2[\hat{\sigma}^2(a_{1j})+\hat{\sigma}^2(a_{2j})/A^2]^{-2}\hat{\sigma}^2(a_{2j})A^{-3} \\
&\quad+\sum_{v=1}^{m_j}[(\delta_{1jv}-A\delta_{2jv}-K)(-\delta_{2jv})[\hat{\sigma}^2(\delta_{1jv})+A^2\hat{\sigma}^2(\delta_{2jv})]^{-1} \\
&\quad-2(\delta_{1jv}-A\delta_{2jv}-K)^2[\hat{\sigma}^2(\delta_{1jv})+A^2\hat{\sigma}^2(\delta_{2jv})]^{-2}A\hat{\sigma}^2(\delta_{2jv})]\big]
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
&\frac{\partial\chi^2}{\partial K} \\
&=-2\sum_{j=1}^{n}\sum_{v=1}^{m_j}[(\delta_{1jv}-A\delta_{2jv}-K)[\hat{\sigma}^2(\delta_{1jv})+A^2\hat{\sigma}^2(\delta_{2jv})]^{-1}
\end{aligned}
\tag{11}
$$

未知数の計算にはDavidon-Fletcher-Powell法（渡部・名取・小国，1989）を用いる．

3．2　カテゴリ特性関数法－Haebara（1980）の基準式 Q_1 を援用する方法－

Haebara（1983）の基準式 Q_1 を用いて最小化基準式を式（12）のように定義し，等化係数 A と K を推定する．

$$
Q_{HT}=\sum_{i=1}^{s}\sum_{j=1}^{n}\sum_{k=1}^{m_j}[P_{jk}(\theta_i)-P_{jk}(\theta_i)^*]^2h(\theta_i) \tag{12}
$$

ここで，$P_{jk}(\theta_i)$ と $P_{jk}(\theta_i)^*$ はそれぞれ等化前後のカテゴリ特性関数の値である．また，$h(\theta_i)$ は能力値 θ_i の相対度数，s は求積点の数である．

未知数の計算にはDavidon-Fletcher-Powell法（渡部・名取・小国，1989）を用いた．その際，未知数に関する基準関数の偏微分の計算には中心差分（$\delta=0.00001$）に基づく数値微分法（戸川，1991）を用いた．

3．3　真の得点法－Stocking & Lord（1983）の基準式を援用する方法－

Stocking & Lord（1983）はロジスティック・モデルの真の得点に着目した基準式を提案している．ここではその基準式を援用して，式（13）を最小化する等化係数 A と K を求める．

$$
Q_{SL}=\sum_{i=1}^{s}\left[\sum_{j=1}^{n}\sum_{k=1}^{m_j}k(P_{jk}(\theta_i)-P_{jk}(\theta_i)^*)\right]^2 \tag{13}
$$

さらに，Haebara（1983）に倣い，能力値の分布 $h(\theta_i)$ を重み付けすることもでき，このときの基準関数は次式となる．

$$
Q_{SL}=\sum_{i=1}^{s}\left[\sum_{j=1}^{n}\sum_{k=1}^{m_j}k(P_{jk}(\theta_i)-P_{jk}(\theta_i)^*)\right]^2 h(\theta_i) \tag{14}
$$

未知数の計算にはDavidon-Fletcher-Powell法（渡部・名取・小国，1989）を用いた．未知数に関する基準関数の偏微分の計算には上記の方法と同様に数値微分法を用いた．

3.4　単純最小自乗法

比較検討のために，最小 χ^2法における Σ_{1j} と Σ^*_{2j} を単位行列として A と K を推定した．これを単純最小自乗法と呼んでおく．

3.5　Mean & Mean 法

Mean & Mean 法を用いて A と K を求めた．等化係数の推定式は以下の通りである．$\bar{a}_2$と $\bar{b}_2$は等化したいテスト冊子で推定された共通項目の識別力とステップ母数の平均値，$\bar{a}_1$と $\bar{b}_1$はターゲットするテスト冊子で推定された共通項目の識別力とステップ母数の平均値である．

$$A = \frac{\bar{a}_2}{\bar{a}_1} \tag{15}$$

$$K = \bar{b}_1 - A\bar{b}_2 \tag{16}$$

4　数値実験

新しいテスト項目のみからなる場合を想定し，下記の4条件について100回のシミュレーション実験を行った．

4.1　方法

4.1.1　水平等化1

すべて4カテゴリとする20項目のテストを想定した．この条件では2つのテストが全く同一の項目から構成されているとしたが，便宜的に2つのテストを T_1と T_2とする．そして，母集団分布が等しい受検者群 S_1と S_2を用意し（各500名），受検者群 S_1へテスト T1，受検者群 S_2へテスト T_2を実施し，S_2で求めた T_2の推定値を S_1で求めた T_1の推定値へ等化した．

実験に用いた項目反応は，モデル母数の真値をカテゴリ特性関数へ代入してカテゴリの累積反応確率を計算し，その値と0～1の一様乱数の値とを比べることにより生成した．母数の真値 θ_i と b_{jv} として正規分布 $N(0, 1)$ に従う乱数，a_j として0.5～1.5の一様乱数を用いた．ここでは2つの母集団分布が等質であるから，等化係数の推定値は，$A=1$，$K=0$ となることが期待される．

4.1.2　垂直等化1

被験者群 S_2の能力値として正規分布 $N(0.7, 1.2^2)$ に従う乱数を用い，他の条件は水平等化1と同一とした．したがって，等化係数の推定値は $A=1.2$，$K=0.7$が期待される．

4.1.3　水平等化2

T_1と T_2にそれぞれ独自の項目が20項目あり，共通項目が10項目あるものとした．他の条件は水平等化1と同一とした．等化係数の推定値は $A=1$，$K=0$ が期待される．

4.1.4　垂直等化2

被験者群 S_2の能力値として正規分布 $N(0.7, 1.2^2)$ に従う乱数を用い，他の条件は水平等化2と同一とした．したがって，等化係数の推定値は $A=1.2$，$K=0.7$が期待される．

4.2　結果と結論

A と K の推定値の平均と標準偏差を表1に示す．

この種の実験は項目母数の推定精度の問題と切り離すことができないので，この結果のみから等化法の優劣を断定することは難しい．そのため，ここでは等化条件ごとに，等化係数の推定値の平均値と標

Table 1　A と K の推定値の平均と標準偏差

等化法		実験条件							
		水平等化1		垂直等化1		水平等化2		垂直等化2	
（A と K の計算方法）	等化係数	平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
最小 χ^2法	A	1.001	0.057	1.205	0.066	1.008	0.062	1.209	0.090
	K	0.022	0.037	0.698	0.052	0.039	0.037	0.718	0.054
単純最小自乗法	A	0.980	0.059	1.177	0.072	0.985	0.077	1.168	0.106
	K	0.018	0.037	0.688	0.052	0.037	0.039	0.695	0.063
カテゴリ特性関数法	A	1.014	0.057	1.216	0.064	1.007	0.057	1.217	0.094
	K	0.024	0.035	0.706	0.052	0.040	0.034	0.720	0.051
真の得点法	A	1.009	0.055	1.214	0.062	1.009	0.061	1.217	0.090
	K	0.023	0.035	0.709	0.051	0.040	0.035	0.724	0.052
Mean & Mean 法	A	0.994	0.058	1.201	0.066	1.005	0.060	1.208	0.088
	K	0.019	0.036	0.701	0.052	0.038	0.038	0.721	0.056

準偏差を吟味してみた．その結果，すべての実験条件において，本研究で提案した方法は期待通りの等化係数を得ることができていた．ただし，実用的には大きな問題とはならないと思われるが，単純最小自乗推定値のバイアスが他の方法のバイアスよりもわずかに大きい．したがって，さらに詳しい検討が必要であるが，最小 χ^2法，カテゴリ特性関数法，真の得点法，Mean & Mean 法を利用する方が安全であろう．

また，本実験は等化係数の推定値の標準誤差を推測するシミュレーション実験とも言え，その標準誤差が表１に示す標準偏差である．この標準誤差が小さいほど安定した推定値を与えることを意味する．水平等化と垂直等化との間では，垂直等化の方がいくらか標準誤差は大きい．

また，等化法の間で標準誤差を比較すると，単純最小自乗法は水平等化１を除き，等化係数Ａの標準誤差が他の等化法よりもいくらか大きい傾向を示す．最小 χ^2法，カテゴリ特性関数法，真の得点法，Mean & Mean 法を利用する方が良いであろう．

5　引用文献

Divgi, D.R. (1985). A minimum chi-square method for developing a common metric in item response theory. *Applied Psychological Measurement*, **9**, 413-415.

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, **22**, 144-149.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, **16**, 159-176.

Muraki, E. & Bock, D.R. (1996). *PARSCALE* Chicago IL: Scientific Software International, Inc.

Stocking, M. L. & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, **7**, 201-210.

戸川隼人 (1991). 数値計算　岩波書店

渡部　力・名取　亮・小国　力 (1989). Fortran77による数値計算ソフトウェア　丸善株式会社

補足　ステップ母数の漸近分散について

PARSCALE は式（1）のステップ母数 b_{jv} を次式のように分解し，b_j と d_{jv} の推定値とその標準誤差（$\hat{\sigma}^2(b_j)$, $\hat{\sigma}^2(d_{jv})$）を出力する．

$$b_{jv} = b_j - d_{jv} \quad (v = 1,2,\ldots,m_j) \tag{17}$$

しかし，b_j と d_{jv} の漸近共分散を出力しないので，本稿ではその漸近共分散を０とみなし，ステップ母数 b_{jv} の漸近分散 $\hat{\sigma}^2(b_{jv})$ を次式によって求めた．

$$\begin{aligned}\hat{\sigma}^2(b_{jv}) &= \hat{\sigma}^2(b_j - d_{jv})\\ &= \hat{\sigma}^2(b_j) + \hat{\sigma}^2(d_{jv})\\ &(v = 1,2,\ldots,m_j)\end{aligned}$$

（受稿９月30日：受理11月17日）