Tsukuba Psychological Research
2006, 31, 19-29

共通項目計画に基づくテストの等化

筑波大学心理学系　服部　　環

Test Equating using on common item design

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

Unidimensional item response models have been developed for tests that are intended to measure a single psychological trait. A number of test equating methods using a commom-item design for such models have been proposed over the last 25 years. In this paper, test equating methods using descriptive statistics for item parameters and characteristic curves are reviewed, and new methods using category, item and test information functions are proposed. Real and artificial data are used to illustrate the procedures.

Key words: item response models, equating, information curves

1　はじめに

テストが測定する心理的特性に1次元性を仮定する項目反応モデルとして，1・2・3母数ロジスティック・モデル，段階反応モデル，一般化部分採点モデル，名義反応モデル，多肢選択モデルなどがある（例えば，Baker, 2004)。これまで，それぞれのモデルについて，共通項目計画に基づいて等化係数を推定するための基準関数が提案されてきた (Baker, 1992, 1993ab, 1997; Haebara, 1980; 服部, 2004abc; Kim & Kolen, 2003; 前川・菊地，2001；Stocking & Lord, 1983)。本稿では，先行研究によって提案された主要な基準関数と等化係数の推定方法を概観し，さらに，先行研究には見られない基準関数を提案して，その計算例を示す。計算には自作プログラムを用いた。

2　1・2・3母数ロジスティック・モデル

2.1　モデル式と等化係数

3母数ロジスティック・モデル (Birnbaum, 1968) は，能力値θ_iの個人iが項目jに正答する確率を

$$P_j(\theta_i) = c_j + (1 - c_j)\frac{1}{1+\exp[-Da_j(\theta_i - b_j)]} \tag{1}$$

と定義する。本稿では個人の心理的特性値を能力値，反応確率を正答確率と記すが，能力検査以外にも項目反応モデルを適用できる。ここで，a_jは項目識別力，b_jは項目困難度，c_jは疑似チャンスレベルである。尺度係数Dの扱いはプログラムによって異なり，Thissen (2003) のMULTILOGは1・2母数ロジスティック・モデルの場合に$D=1$，3母数ロジスティック・モデルの場合に$D=1.7$とする。一方，Zimowski, Muraki, Mislevy & Bock (2003) のBILOG-MG3や服部 (2004d,e) のプログラムでは，利用者がDの値として1もしくは1.7を指定する。

等化係数A_{21}とK_{21}を用いて，冊子1の尺度値は

$$\theta_i^* = A_{21}\theta_i + K_{21} \tag{2}$$
$$a_j^* = a_j/A_{21} \tag{3}$$
$$b_j^* = A_{21}b_j + K_{21} \tag{4}$$
$$c_j^* = c_j \tag{5}$$

によって冊子2の尺度へ等化される。ここで，θ_i, a_j, b_j, c_jは冊子1の尺度における母数値，θ_i^*, a_j^*, b_j^*, c_j^*は冊子2の尺度における母数値である。共通項目法により等化係数を求める場合，冊子2で推

定された共通項目の項目母数値を左辺に代入し，冊子1で推定された共通項目の項目母数値を右辺に代入して，両辺の差異を最小化する等化係数を求める。

一方，冊子2の項目母数値を冊子1の尺度へ等化するための等化係数をA_{12}とK_{12}とすると，

$$A_{12}=1/A_{21} \tag{6}$$
$$K_{12}=-K_{21}/A_{21} \tag{7}$$

という関係がある。等化係数の推定値がこの2つの関係を満たすとき，推定値が対称性を持つことになる。

2．2　推定方法・基準関数

2．2．1　Mean & Mean 法

これはLoyd & Hoover（1980）の方法をロジスティック・モデルへ準用した方法である（Kolen & Brennan, 2004）。式（3）と（4）の左辺に冊子2で推定された共通項目の推定値を代入し，冊子1で推定された共通項目の推定値を右辺に代入して両辺の平均を求めると，

$$\bar{a}^{(2)}=\bar{a}^{(1)}/A_{21} \tag{8}$$
$$\bar{b}^{(2)}=A_{21}b^{(1)}+K_{21} \tag{9}$$

となる。ここで，$\bar{a}^{(t)}$ は共通項目の冊子tにおける識別力母数a_jの推定値$\hat{a}^{(t)}$ の平均，$\bar{b}^{(t)}$ は冊子tにおける困難度母数b_jの推定値$\hat{b}_j^{(t)}$ の平均である。Mean & Mean 法はこの関係式を利用して等化係数を求める。計算式は

$$\hat{A}_{21}=\bar{a}^{(1)}/\bar{a}^{(2)} \tag{10}$$
$$\hat{K}_{21}=b^{(2)}-\hat{A}_{21}b^{(2)} \tag{11}$$

である。この推定値は対称性を持つ。

2．2．2　Mean & Sigma 法

これはMarco（1977）の方法をロジスティック・モデルへ準用した方法である（Kolen & Brennan, 2004）。式（4）の左辺に冊子2で推定された共通項目の推定値を代入し，冊子1で推定された共通項目の推定値を右辺に代入して，両辺の平均と標準偏差を求めると，

$$\bar{b}^{(2)}=A_{21}\bar{b}^{(1)}+K_{21} \tag{12}$$
$$S(\hat{b}^{(2)})=A_{21}S(\hat{b}^{(1)}) \tag{13}$$

となる。ここで，$S(\hat{b}^{(t)})$ は共通項目の困難度母数の冊子tにおける推定値$\hat{b}_j^{(t)}$ の標準偏差を示す。Mean & Sigma 法はこの関係式を利用して等化係数を求める。したがって，等化係数の計算式は

$$\hat{A}_{21}=S(\hat{b}^{(2)})/S(\hat{b}^{(1)}) \tag{14}$$
$$\hat{K}_{21}=\bar{b}^{(2)}-\hat{A}_{21}\bar{b}^{(1)} \tag{15}$$

となる。等化係数の推定値は対称性を持つ。この方法は共通項目の識別力母数a_jを利用しない点に特徴がある。

2．2．3　項目特性曲線に基づく基準関数

Haebara（1980）の基準関数は，

$$Q=\sum_{i=1}^{s}\sum_{j=1}^{n}\ [[P_j(\theta_i)^{(2)}-P_j(\theta_i)^{(21)}]^2h(\theta_i)^{(2)} + [P_j(\theta_i)^{(1)}-P_j(\theta_i)^{(12)}]^2h(\theta_i)^{(1)}] \tag{16}$$

である。この基準関数は$1/\hat{A}_{21}$をA_{12}の推定値，また，$-\hat{K}_{21}/\hat{A}_{21}$を$K_{12}$の推定値として，冊子2の項目母数値を冊子1の尺度へ等化する際の等化誤差も考慮して基準関数を定義しているので，$\hat{A}_{21}$と$\hat{K}_{21}$は対称性を持つ。

ここで，$P_j(\theta_i)^{(t)}$ は冊子tで推定された項目母数値を用いて計算した項目特性関数値，$P_j(\theta_i)^{(21)}$ は冊子1で推定された項目母数値を冊子2の尺度へ等化した上で計算した項目特性関数値，$P_j(\theta_i)^{(12)}$ は$1/\hat{A}_{21}$と$-\hat{K}_{21}/\hat{A}_{21}$を用いて冊子2で推定された項目母数値を冊子1の尺度へ等化した上で計算した項目特性関数値である。$h(\theta_i)^{(t)}$ は冊子tの受検者群における能力値θ_iの確率密度，sは求積点の数である。

なお，第2項を削除したものが，対称性を持たない等化係数を求めるための基準式となる。

2．2．4　テスト特性曲線に基づく基準関数

Stocking & Lord（1983）はテスト特性曲線に着目して，

$$Q=\sum_{i=1}^{s}\left[\sum_{j=1}^{n}\left[P_j(\theta_i)^{(2)}-P_j(\theta_i)^{(21)}\right]\right]^2 h(\theta_i)^{(2)} \tag{17}$$

を提案した。Stockingらの基準関数は重み付け（$h(\theta_i)^{(2)}$）をしていないが，本稿では，重み付けが可能な基準式を用いた。この基準式は，冊子1の項目母数値を冊子2の尺度へ等化する際の誤差のみしか考慮していないので，等化係数の推定値は対称性を持たない。

等化係数の推定値に対称性を持たせるための基準式は，

$$Q=\sum_{i=1}^{s}\left[\left[\sum_{j=1}^{n}\left[P_j(\theta_i)^{(2)}-P_j(\theta_i)^{(21)}\right]\right]^2 h(\theta_i)^{(2)} +\left[\sum_{j=1}^{n}\left[P_j(\theta_i)^{(1)}-P_j(\theta_i)^{(12)}\right]\right]^2 h(\theta_i)^{(1)}\right] \quad (18)$$

となる（Kim & Kolen, 2003; 服部，2004b)。

2．2．5　項目情報量曲線に基づく基準関数

共通項目で定義される項目情報量の差異に着目する基準関数であり，これは先行研究には見あたらない。この方法は，冊子1の母数値を冊子2の尺度へ等化する差異の誤差と冊子2から冊子1の尺度へ等化する際の誤差を考慮した基準関数を

$$Q=\sum_{i=1}^{s}\sum_{j=1}^{n}\ \left[\left[I_j(\theta_i)^{(2)}-I_j(\theta_i)^{(21)}\right]^2 +\left[I_j(\theta_i)^{(1)}-I_j(\theta_i)^{(12)}\right]^2\right] \quad (19)$$

と定義する。この基準式は各共通項目ごとに，その情報量の差異を最小化するねらいがあり，等化係数の推定値は対称性を持つ。基準関数の第2項を削除した場合，等化係数は対称性を持たない。ここで，$I_j(\theta_i)^{(t)}$ はテスト冊子tで推定された項目母数値を用いて計算した項目情報量，$I_j(\theta_i)^{(21)}$ はテスト冊子1で推定された項目母数値をテスト冊子2の尺度へ等化した上で計算した項目情報量，$I_j(\theta_i)^{(12)}$ は式(6)と式(7)のA_{12}とK_{12}を用いてテスト冊子2で推定された項目母数値をテスト冊子1の尺度へ等化した上で計算した項目情報量である。なお，情報量の差異に重み付けをすることも可能であるが，本稿では重み付けをしなかった。

2．2．6　テスト情報量曲線に基づく基準関数

共通項目のテスト情報量曲線の差違を最小化する基準関数であり，先行研究には見あたらない。この基準関数は

$$Q=\sum_{i=1}^{s}\left[\left[\sum_{j=1}^{n}\left[I_j(\theta_i)^{(2)}-I_j(\theta_i)^{(21)}\right]\right]^2 +\left[\sum_{j=1}^{n}\left[I_j(\theta_i)^{(1)}-I_j(\theta_i)^{(12)}\right]\right]^2\right] \quad (20)$$

と定義される。この基準関数に基づく等化係数は対称性を持ち，基準関数から第2項を削除したときは対称性を持たない。

2．3　計算例

Kolen & Brennan (2004) のTable6.6に記載されている3母数ロジスティック・モデルの項目母数推定値を利用した。共通項目数は12項目である。等化係数の推定値をTable 1に示す。括弧内に非対称と記した基準は，冊子1から冊子2への等化によって生じる誤差のみを最小化する基準関数であり，推定値は対称性を持たない。また，対称と記した基準関数は冊子2から冊子1への等化によって生じる誤差も考慮するので，推定値は対称性を持つ。推定値を比較すると，情報量に着目して推定した等化係数の値が他の推定値とやや異なるが，シミュレーション実験をしていないので，結論を一般化することはできない。この点は，以下で示す他のモデルの計算例も同様である。

3　段階反応モデル

3．1　モデル式

段階反応モデル（Samejima, 1969）は，能力値θ_iの個人iが項目jでカテゴリk($k=1,2,\cdots,m_j$;m_jは項目jのカテゴリ最大値）を取る確率を

$$P_{jk}(\theta_i)=P^*_{jk}(\theta_i)-P^*_{jk+1}(\theta_i)$$
$$=\frac{1}{1+\exp[-a_j(\theta_i-b_{jk-1})]}-\frac{1}{1+\exp[-a_j(\theta_i-b_{jk})]} \quad (21)$$
$$(k=1,2,\cdots,m_j)$$

と定義する。ここで，a_jは勾配母数（識別力母数），b_{jk}は閾値母数（困難度母数）であり，b_{jk-1}は項目jにおいてカテゴリk以上をとる確率が0.5となる能力値を表す。そして，$P^*_{j1}(\theta_i)=1$，$P^*_{jk+1}(\theta_i)=0$と

Table 1　1・2・3母数ロジスティック・モデルのための計算例

推定方法・基準関数	$\hat{A}_{21}$	$\hat{K}_{21}$
Mean & Mean 法	1.14	−0.48
Mean & Sigma 法	1.18	−0.50
項目特性曲線（非対称）	1.07	−0.46
テスト特性曲線（非対称）	1.09	−0.47
項目特性曲線（対称）	1.06	−0.45
テスト特性曲線（対称）	1.08	−0.47
項目情報量曲線（非対称）	1.20	−0.68
テスト情報量曲線（非対称）	1.15	−0.72
項目情報量曲線（対称）	1.18	−0.61
テスト情報量曲線（対称）	1.14	−0.70

おく。

冊子1のモデル母数値は,

$$\theta_i^* = A_{21}\theta_i + K_{21} \tag{22}$$
$$a_j^* = a_j/A_{21} \tag{23}$$
$$b_{jk-1}^* = A_{21} b_{jk-1} + K_{21} \tag{24}$$

を用いて冊子2の尺度へ等化できる。したがって,共通項目法では冊子2で推定された共通項目の推定値を左辺に,冊子1で推定された共通項目の推定値を右辺に代入し,両辺の差異が小さくなるように等化係数を推定する。

3.2　推定方法・基準関数

3.2.1　Mean & Mean 法

Loyd & Hoover (1980) の方法を段階反応モデルへ準用した方法である (Kolen & Brennan, 2004)。等化係数の推定値は,

$$\hat{A}_{21} = \bar{\hat{a}}^{(1)}/\bar{\hat{a}}^{(2)} \tag{25}$$
$$\hat{K}_{21} = \bar{\hat{b}}^{(2)} - \hat{A}_{21}\bar{\hat{b}}^{(1)} \tag{26}$$

である。ここで,$\bar{\hat{a}}^{(t)}$ は共通項目の冊子tにおける勾配母数a_jの推定値$\hat{a}^{(t)}$の平均,$\bar{\hat{b}}^{(t)}$ は冊子tにおける閾値母数b_{jk}の推定値$\hat{b}_{jk}^{(t)}$の平均である。この推定値は対称性を持つ。

3.2.2　Mean & Sigma 法

Marco (1977) の方法を段階反応モデルへ準用した方法である (Kolen & Brennan, 2004)。等化係数の推定式は,

$$\hat{A}_{21} = S(\hat{b}^{(2)})/S(\hat{b}^{(1)}) \tag{27}$$
$$\hat{K}_{21} = \bar{\hat{b}}^{(2)} - \hat{A}_{21}\bar{\hat{b}}^{(1)} \tag{28}$$

である。ここで,$S(\hat{b}^{(t)})$ は共通項目の閾値母数の冊子tにおける推定値$\hat{b}_{jk}^{(t)}$の標準偏差を示す。

3.2.3　カテゴリ特性曲線に基づく基準関数

Haebara (1980) に倣い,カテゴリ特性曲線に着目して等化係数の推定値に対称性を持たせるための基準関数を

$$Q=\sum_{i=1}^{s}\sum_{j=1}^{n}\sum_{k=1}^{m_j}\left[\left[P_{jk}(\theta_i)^{(2)} - P_{jk}(\theta_i)^{(21)}\right]^2 h(\theta_i)^{(2)} + \left[P_{jk}(\theta_i)^{(1)} - P_{jk}(\theta_i)^{(12)}\right]^2 h(\theta_i)^{(1)}\right] \tag{29}$$

とする。この基準関数は共通項目のカテゴリ特性曲線を可能な限り一致させることを意図している。ここで,$P_{jk}(\theta_i)^{(t)}$ は冊子tで推定された項目母数値を用いて計算したカテゴリ特性関数値,$P_{jk}(\theta_i)^{(21)}$ は冊子1で推定された項目母数値を式 (23) と式 (24) を用いて冊子2の尺度へ等化した上で計算したカテゴリ特性関数値,$P_{jk}(\theta_i)^{(12)}$ は$1/\hat{A}_{21}$をA_{12}の推定値,$-K_{21}/\hat{A}_{21}$をK_{12}の推定値として冊子2で推定された項目母数値を冊子1の尺度へ等化した上で計算したカテゴリ特性関数値である。$h(\theta_i)^{(t)}$ は等化誤差の二乗に乗じる重みで,分析者が任意の値を与える。仮に冊子tの受検者群における能力値θ_iの分布が既知であるなら,その確率密度を利用することもできる。nは共通項目数,sは求積点の数である。

基準関数の第1項のみを最小化する場合,等化係数の推定値は対称性を持たない。この基準関数において,さらに$h(\theta_i)^{(2)}$ を定数とした基準関数は,Baker (1993b) が付録において示した基準関数と一致する。

3.2.4　項目特性曲線に基づく基準関数

項目特性曲線は能力値θ_iの受検者が項目jで取る得点の期待値$E(x_{ij}|\theta_i)$ を表し,

$$E(x_{ij}|\theta_i)=\sum_{k=1}^{m_j} kP_{jk}(\theta_i) \tag{30}$$

と定義される。項目特性曲線に着目した,等化係数の推定値に対称性を持たせるための基準関数は,

$$Q=\sum_{i=1}^{s}\sum_{j=1}^{n}\left[\left[\sum_{k=1}^{m_j} k\left(P_{jk}(\theta_i)^{(2)} - P_{jk}(\theta_i)^{(21)}\right)\right]^2 h(\theta_i)^{(2)} + \left[\sum_{k=1}^{m_j} k\left(P_{jk}(\theta_i)^{(1)} - P_{jk}(\theta_i)^{(12)}\right)\right]^2 h(\theta_i)^{(1)}\right] \tag{31}$$

となる。この基準関数は2つの冊子の間で共通項目の項目特性曲線を可能な限り一致させることを意図している。ここで,$P_{jk}(\theta_i)^{(t)}$,$P_{jk}(\theta_i)^{(21)}$,$P_{jk}(\theta_i)^{(12)}$,$h(\theta_i)^{(t)}$,sなどの意味は前述と同様である。先行研究では,項目特性曲線に着目した基準関数は提案されていないと思われる。

なお,この基準関数の第1項のみを用いた場合,等化係数の推定値は対称性を持たない。

3.2.5　テスト特性曲線に基づく基準関数

テスト特性曲線は能力値θ_iの受検者が取るテスト得点の期待値$E(x_i|\theta_i)$ を表し,

$$E(x_i|\theta_i)=\sum_{j=1}^{n}\sum_{k=1}^{m_j} kP_{jk}(\theta_i) \tag{32}$$

と定義される。テスト特性曲線に着目し,等化係数の推定値に対称性を持たせるための基準関数は,

$$Q=\sum_{i=1}^{s}\left[\left[\sum_{j=1}^{n}\sum_{k=1}^{m_j}k\left(P_{jk}(\theta_i)^{(2)}-P_{jk}(\theta_i)^{(21)}\right)\right]^2 h(\theta_i)^{(2)}\right.$$
$$\left.+\left[\sum_{j=1}^{n}\sum_{k=1}^{m_j}k\left(P_{jk}(\theta_i)^{(1)}-P_{jk}(\theta_i)^{(12)}\right)\right]^2 h(\theta_i)^{(1)}\right] \quad (33)$$

となる。この基準関数は，２つの冊子の間で共通項目が定義するテスト特性曲線を可能な限り一致させることを意図している。ここで，$P_{jk}(\theta_i)^{(t)}$，$P_{jk}(\theta_i)^{(21)}$，$P_{jk}(\theta_i)^{(12)}$，$h(\theta_i)^{(t)}$，s などの意味は前述と同様である。

この基準関数で第１項のみを用いた場合，等化係数の推定値は対称性を持たない。さらに，$h(\theta_i)^{(2)}$ を定数とした基準関数が Baker（1992）の基準関数と一致する。

3．2．6　カテゴリ情報量に基づく基準関数

共通項目のカテゴリ情報量の差違を最小化する基準関数を

$$Q=\sum_{i=1}^{s}\sum_{j=1}^{n}\sum_{k=1}^{m_j}\left[\left[P_{jk}(\theta_i)^{(2)}I_j(\theta_i)^{(2)}-P_{jk}(\theta_i)^{(21)}I_j(\theta_i)^{(21)}\right]^2\right.$$
$$\left.+\left[P_{jk}(\theta_i)^{(1)}I_j(\theta_i)^{(1)}-P_{jk}(\theta_i)^{(12)}I_j(\theta_i)^{(12)}\right]^2\right] \quad (34)$$

と定義する。この基準関数は先行研究に見あたらない。ここで，$P_{jk}(\theta_i)^{(t)}$ と $I_j(\theta_i)^{(t)}$ はテスト冊子 t で推定された項目母数値を用いて計算したカテゴリ特性関数値と項目情報量，$P_{jk}(\theta_i)^{(21)}$ と $I_j(\theta_i)^{(21)}$ はテスト冊子１で推定された項目母数値をテスト冊子２の尺度へ等化した上で計算したカテゴリ特性関数値と項目情報量，$P_{jk}(\theta_i)^{(21)}$ と $I_j(\theta_i)^{(21)}$ はテスト冊子２で推定された項目母数値をテスト冊子１の尺度へ等化した上で計算したカテゴリ特性関数値と項目情報量である。n は項目数，s は求積点の数である。

等化係数の推定値は対称性を持つが，第１項のみを用いたとき，等化係数の推定値は対称性を持たない。

段階反応モデルの項目情報量 $I_j(\theta_i)$ は

$$I_j(\theta_i)=D^2a_j^2\left[\sum_{k=1}^{m_j}\left[p^*_{jk}(\theta_i)q^*_{jk}(\theta_i)-p^*_{jk+1}(\theta_i)q^*_{jk+1}(\theta_i)\right]/p_{jk}(\theta_i)\right] \quad (35)$$

と定義される。ここで，$q^*_{jk}(\theta_i)=1-p^*_{jk}(\theta_i)$ である。

3．2．7　項目情報量に基づく基準関数

共通項目の項目情報量の差違を最小化する基準関数を

$$Q=\sum_{i=1}^{s}\sum_{j=1}^{n}\left[\left[I_j(\theta_i)^{(2)}-I_j(\theta_i)^{(21)}\right]^2+\left[I_j(\theta_i)^{(1)}-I_j(\theta_i)^{(12)}\right]^2\right] \quad (36)$$

と定義する。この基準関数も先行研究には見あたらない。記号の意味は前述の通りである。等化係数の推定値は対称性を持つが，第１項のみを用いたとき，等化係数の推定値は対称性を持たない。

3．2．8　テスト情報量に基づく基準関数

共通項目のテスト情報量の差違を最小化する基準関数を

$$Q=\sum_{i=1}^{s}\left[\left[\sum_{j=1}^{n}\left[I_j(\theta_i)^{(2)}-I_j(\theta_i)^{(21)}\right]\right]^2\right.$$
$$\left.+\left[\sum_{j=1}^{n}\left[I_j(\theta_i)^{(1)}-I_j(\theta_i)^{(12)}\right]\right]^2\right] \quad (37)$$

と定義する。これも先行研究には見あたらない基準である。記号の意味は前述の通りである。また，等化係数の推定値は対称性を持つが，第１項のみを用いたとき，等化係数の推定値は対称性を持たない。

3．3　計算例

ここでは，Kim & Kolen（2003）によって提供されている段階反応モデルの項目母数値を用いた。等化係数の推定値を Table 2 に示す。この計算例では，推定値はほぼ等しい。

Table 2　段階反応モデルのための計算例

推定方法・基準関数	$\hat{A}_{21}$	$\hat{K}_{21}$
Mean & Mean 法	1.50	−0.25
Mean & Sigma 法	1.50	−0.25
カテゴリ特性曲線（非対称）	1.50	−0.25
項目特性曲線（非対称）	1.50	−0.25
テスト特性曲線（非対称）	1.50	−0.25
カテゴリ特性曲線（対称）	1.50	−0.25
項目特性曲線（対称）	1.50	−0.25
テスト特性曲線（対称）	1.50	−0.25
カテゴリ情報量曲線（非対称）	1.50	−0.25
項目情報量曲線（非対称）	1.50	−0.25
テスト情報量曲線（非対称）	1.50	−0.25
カテゴリ情報量曲線（対称）	1.50	−0.25
項目情報量曲線（対称）	1.50	−0.26
テスト情報量曲線（対称）	1.50	−0.26

4　一般化部分採点モデル

4.1　モデル式

一般化部分採点モデル（Muraki, 1992）は Masters（1982）の部分採点モデルを一般化した項目反応モデルであり，多値的に採点された項目に適用する。このモデルは，特性値 θ_i の受検者が項目 j のカテゴリ k（$k=0,1,\cdots,m_j$）に反応する確率もしくは k 点を取る確率，つまりカテゴリ特性関数を

$$P_{jk}(\theta_i)=\frac{\exp\sum_{r=1}^{k}Da_j(\theta_i-b_{jr})}{1+\sum_{h=1}^{m_j}\exp\sum_{r=1}^{h}Da_j(\theta_i-b_{jr})} \quad (38)$$

$$(k=0,1,2,\cdots,m_j)$$

と定義する。ここで，a_j は項目 j の識別力，b_{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）であり，$a_j(\theta_i-b_{j0})\equiv 0$ とおく。D は尺度係数であり，1もしくは1.7とするが，本稿では1とした。

2つの定数 A_{21}（$A_{21}>0$）と K_{21} を用いてモデル母数を

$$\theta_i^*=A_{21}\theta_i+K_{21} \quad (39)$$

$$a_j^*=a_j/A_{21} \quad (40)$$

$$b_{jv}^*=A_{21}b_{jv}+K_{21}\ (v=1,2,\cdots,m_j) \quad (41)$$

と変換しても，カテゴリ特性関数値は変わらない。そこで，共通項目法はこの関係を利用して，テスト冊子1の尺度値をテスト冊子2の尺度へ等化するための等化係数 A_{21} と K_{21} を求める。

4.2　推定方法・基準関数

Mean & Mean 法と Mean & Sigma 法は段階反応モデルと同様であるから，ここでは，計算式を省略する。推定値は対称を持つ。

4.2.1　最小 χ^2 法

服部（1998, 2005a）は最小 χ^2 法を提案した。基準関数は，

$$Q=\sum_{j=1}^{n}\left[(a_{2j}-a_{1j}/A_{21},\ \delta_{2j1}-A_{21}\delta_{1j1}-K_{21},\cdots,\ \delta_{2jm_j}-A_{21}\delta_{1jm_j}-K_{21})(\boldsymbol{\Sigma}_{2j}+\boldsymbol{\Sigma}_{1j}^*)^{-1}(a_{2j}-a_{1j}/A_{21},\ \delta_{2j1}-A_{21}\delta_{1j1}-K_{21},\cdots,\ \delta_{2jm_j}-A_{21}\delta_{1jm_j}-K_{21})'\right] \quad (42)$$

である。ここで，n は共通項目数，$\boldsymbol{\Sigma}_{2j}$ は a_{2j}，δ_{2j1}，$\delta_{2j1},\cdots,\delta_{2jm_j}$ の推定値の漸近分散共分散行列，$\boldsymbol{\Sigma}_{1j}^*$ は共通項目 j の冊子1で推定された値を冊子2の尺度へ等化したときの漸近分散共分散行列である。推定値は対称性を持つ。PARSCALE は項目母数の漸近共分散を出力しないので，PARSCALE による項目母数推定値を用いる場合は漸近分散のみを利用する。また，漸近分散共分散行列を単位行列とした基準関数は最小2乗法の基準関数となる。

4.2.2　特性曲線に基づく基準関数

服部（2004a）はカテゴリ特性曲線，項目特性曲線，テスト特性曲線に着目する基準関数を提案した。基準式は段階反応モデルと同様であるから，ここでは省略する。

4.2.3　情報量に基づく基準関数

服部（2005b）は共通項目のカテゴリ情報量，項目情報量，テスト情報量に着目した基準関数を提案している。基準式は段階反応モデルと同様であるから省略する。ただし，一般化部分採点モデルの項目情報量 $I_j(\theta_i)$ は

$$I_j(\theta_i)=D^2a_j^2\left[\sum_{k=0}^{m_j}T_k^2p_{jk}(\theta_i)-\left[\sum_{k=0}^{m_j}T_k\,p_{jk}(\theta_i)\right]^2\right] \quad (43)$$

と定義される（Muraki & Bock, 1996）。ここで T_k は得点関数であり，$T_k=k+1$ とする。

4.3　計算例

Muraki & Bock（2003）の PARSCALE はステップ母数 b_{jv} を次式のように分解して，b_j と d_{jv} の推定値と標準誤差を出力する。

$$b_{jv}=b_j-d_{jv} \quad (44)$$

$$(v=1,2,\cdots,m_j)$$

そのため，計算例には Table 3 に示す b_j と d_{jv} を用いた。推定値を Table 4 に示すが，推定値に大きな相違はない。

5　名義反応モデル

5.1　モデル式

名義反応モデル（Bock, 1972）は，能力値 θ_i の個人 i が項目 j において選択肢 k（$k=1,2,\cdots,m_j$;m_j は項目 j の選択肢数）を選択する確率を

$$P_{jk}(\theta_i)=\frac{\exp[a_{jk}\theta_i+c_{jk}]}{\sum_{h=1}^{m_j}\exp[a_{jh}\theta_i+c_{jk}]} \quad (45)$$

$$(k=1,2,\cdots,m_j)$$

Table 3　一般化部分採点モデルの計算例に用いた母数値

項目番号	a_j	b_j	d_{j1}	d_{j2}	d_{j3}
冊子 2					
1	1.468	−0.241	1.580	−0.670	−0.909
2	0.641	−0.196	0.347	−0.492	0.144
3	1.090	−0.108	−0.348	−0.240	0.589
4	0.998	−0.384	−0.607	−0.002	0.610
5	0.550	−0.449	−0.661	0.328	0.333
6	1.011	−0.989	−0.130	−1.130	1.261
7	0.906	1.123	−1.374	0.876	0.497
8	0.605	−0.176	0.697	−0.419	−0.277
9	1.059	0.086	0.548	−0.921	0.373
10	1.179	0.808	1.006	−0.113	−0.893
冊子 1					
1	1.724	−0.797	1.206	−0.482	−0.723
2	0.792	−0.586	−0.035	−0.314	0.350
3	1.333	−0.690	−0.375	0.237	0.138
4	1.125	−1.046	−0.496	−0.087	0.584
5	0.734	−0.972	−0.708	0.502	0.205
6	1.200	−1.651	−0.300	−0.710	1.010
7	0.872	0.447	−1.714	1.156	0.557
8	0.749	−0.704	0.428	−0.510	0.081
9	1.400	−0.516	0.606	−1.065	0.458
10	1.412	0.051	0.895	0.001	−0.896

Table 4　一般化部分採点モデルのための計算例

推定方法・基準関数	$\hat{A}_{21}$	$\hat{K}_{21}$
Mean & Mean 法	1.19	0.72
Mean & Sigma 法	1.06	0.63
Normit χ^2法	1.16	0.71
最小二乗法	1.07	0.64
カテゴリ特性曲線（非対称）	1.15	0.72
項目特性曲線（非対称）	1.12	0.71
テスト特性曲線（非対称）	1.17	0.73
カテゴリ特性曲線（対称）	1.15	0.73
項目特性曲線（対称）	1.12	0.71
テスト特性曲線（対称）	1.17	0.73
カテゴリ情報量曲線（非対称）	1.18	0.78
項目情報量曲線（非対称）	1.18	0.79
テスト情報量曲線（非対称）	1.18	0.79
カテゴリ情報量曲線（対称）	1.16	0.77
項目情報量曲線（対称）	1.17	0.78
テスト情報量曲線（対称）	1.18	0.79

と定義する。ここで，a_{jh} と c_{jh} は項目 j のカテゴリ h の特性を表す。そして，項目母数を識別するために $\Sigma_{h=1}^{m_j} a_{jh}=0$，$\Sigma_{h=1}^{m_j} c_{jh}=0$ と制約する。MULTILOG はこの制約を課さずに項目母数を推定し，それを上記の制約を課した母数へ変換する。

冊子 1 の母数は

$$\theta_i^* = A_{21}\theta_i + K_{21} \tag{46}$$

$$a_{jk}^* = a_{jk}/A_{21} \tag{47}$$

$$c_{jk}^* = c_{jk} - (a_{jk}/A_{21})K_{21} \tag{48}$$

を用いて冊子 2 の尺度へ等化できる。共通項目法では冊子 2 で推定された共通項目の推定値を左辺に，冊子 1 で推定された共通項目の推定値を右辺に代入し，両辺の差異が小さくなるように等化係数を推定する。さらに冊子 2 から冊子 1 へ等化する際の誤差も考慮するなら，等化係数の推定値は対称性を持つ。

5．2　基準関数

5．2．1　カテゴリ特性曲線に基づく基準関数

名義反応モデルは項目得点とテスト得点の期待値

を定義できないので，カテゴリ特性曲線を用いて等化係数を推定する。そのための基準関数は

$$Q=\sum_{i=1}^{s}\sum_{j=1}^{n}\sum_{k=1}^{m_j}\left[\left[P_{jk}(\theta_i)^{(2)}-P_{jk}(\theta_i)^{(21)}\right]^2 h(\theta_i)^{(2)}+\left[P_{jk}(\theta_i)^{(1)}-P_{jk}(\theta_i)^{(12)}\right]^2 h(\theta_i)^{(1)}\right] \quad (49)$$

である。記号の意味は段階反応デルと同様である。この基準式に基づく推定値は対称性を持つが，Baker（1993）とKim & Hanson（2002）は重み付けをしない第1項のみによって基準関数を定義しているので，推定値は対称性を持たない。また，服部（2004b）はKim & Kolen（2003）のプログラムの存在を知らないまま，それとは独立して対称性を持たせる基準関数を提案した。

5．3 計算例

Kim & Kolen（2003）で提供されている名義反応モデルの項目母数値を計算例に用いた。MULTILOGは推定値の合計値に制約を置かない母数値も出力するので，それを用いた等化係数を推定するプログラムも作成したが，ここでは制約を置いたKimらの項目母数値を用いた。共通項目は10項目である。推定値をTable 5に示す。この計算例の場合，対称性の有無による推定値の差は小さい。なお，情報量を用いた基準関数も理論的には利用できるが，ここでは，情報量を用いた計算例を示していない。プログラムの作成は今後の課題である。

6 多肢選択モデル

6．1 モデル式

Thissen & Steinberg（1984）の多肢選択モデルは名義反応モデルを拡張して，当て推量による解答確率を説明するモデルである。このモデルは能力値θ_iの個人iが項目jにおいて選択肢k（$k=2,\cdots,m_j+1$;m_j+1は項目jの選択肢数）を選択する確率を

$$P_{jk}(\theta_i)=\frac{\exp\left[a_{jk}\theta_i+c_{jk}\right]+d_{jk}\exp\left[a_{j1}\theta_i+c_{j1}\right]}{\sum_{h=1}^{m_j+1}\exp\left[a_{jh}\theta_i+c_{jh}\right]} \quad (50)$$

$$(k=1,2,\cdots,m_j)$$

と定義する。ここで，$P_{j1}(\theta_i)$は能力値θ_iの個人iが項目jで当て推量解答を行う確率，d_{jk}（$k\geq2$）は項目jで当て推量解答を行う個人のうち，選択肢k（$k=2,\cdots,m_j+1$）を選ぶ確率である。ただし，$d_{j1}=0$である。また，母数を識別するために$\sum_{h=2}^{m_j+1}d_{jh}=1$，$\sum_{h=1}^{m_j+1}a_{jh}=0$，$\sum_{h=1}^{m_j+1}c_{jh}=0$と制約される。

6．2 基準関数

6．2．1 カテゴリ特性曲線に基づく基準関数

冊子1の尺度で推定された項目母数a_{jk}とc_{jk}は，名義反応モデルと同一の式を用いて冊子2の尺度へ等化できる。また，d_{jk}は等化係数の影響を受けない。多肢選択モデルも項目得点とテスト得点の期待値を定義できないので，選択肢の反応確率に着目して基準関数を定義する。基準関数は名義反応モデルと同様であるから，ここでは省略する。

6．3 計算例

計算例にはKim & Kolen（2003）で提供されている多肢選択式モデルの項目母数値を利用した。共通項目は10項目である。推定値をTable 6に示す。この計算例では，対称性を持たせた推定値と持たせない推定値との相違は小さい。

7 複数モデルの混在

7．1 基準関数

MULTILOGプログラムは1・2・母数ロジスティック・モデル，段階反応モデル，名義反応モデル，多肢選択モデルを同時に扱うことができる。そこで，複数のモデルを混在させてカテゴリ特性曲線に着目した基準関数を定義した（服部，2004b）。本稿では既にカテゴリ特性曲線に基づく基準関数を紹介しているので，ここでは省略する。Kim & Kolen（2003）のプログラムは複数のモデルの項目母数を同時に等化するための係数を求めることはできない。

7．2 計算例

Table 7に計算例に用いた母数値を示す。ここでは，MULTILOGで利用される無制約の推定値を示した。本稿のプログラムでは，この無制約の推定値

Table 5 名義反応モデルの計算例

基準関数	$\hat{A}_{21}$	$\hat{K}_{21}$
カテゴリ特性曲線（非対称）	1.15	0.72
カテゴリ特性曲線（対称）	1.15	0.73

Table 6 多肢選択モデルのための基準関数と計算例

基準関数	$\hat{A}_{21}$	$\hat{K}_{21}$
カテゴリ特性曲線（非対称）	1.08	0.00
カテゴリ特性曲線（対称）	1.02	0.01

を制約付きの母数値へ変換した上で基準関数を定義している。Table 8 に計算結果を示す。等化係数に対称性を持たせた場合と持たせない場合の相違は小さい。

8　まとめ

本稿は共通項目計画に基づいて，1・2・3母数ロジスティック・モデル，段階反応モデル，一般化部分採点モデル，名義反応モデル，多肢選択モデルの項目母数を等化するための基準関数を概観し，新しい基準関数を提案した。本稿で提案した情報量曲線を用いた基準関数は先行研究にはないように思われる。

段階反応モデル，一般化部分採点モデル，名義反応，多肢選択モデルの等化係数に対称性を持たせた基準関数を提案した論文は公刊されていないと思われるが，本稿で利用したプログラムを作成した後，Kim & Kolen（2003）のプログラムが対称性を持たせた等化係数を推定できることを知った。ただし，Kim & Kolen（2003）はカテゴリ特性曲線，そして，段階反応モデルと一般化部分採点モデルついてはテスト特性曲線を用いているが，本稿で提案した項目特性曲線と情報量曲線を用いた基準関数を利用できない。

さらに，Kim & Kolen（2003）は等化係数を出力するが，新項目の等化後の推定を出力しない。しかも，その等化係数の推定値を用いて項目母数を変換しても，そのままではMULTILOGとPARSCALEで利用することはできない。利用者がプログラムの仕様に合わせて項目母数を再変換する必要がある。一方，本稿で利用した自作プログラムは，等化後の項

Table 8　複数モデルの同時等化の計算例

基準関数	$\hat{A}_{21}$	$\hat{K}_{21}$
カテゴリ特性曲線（非対称）	1.15	0.72
カテゴリ特性曲線（対称）	1.15	0.73

Table 7　複数モデルの同時等化の例に用いた母数値

項目番号	モデル	M_j	母数値													
テスト冊子2																
1	L1		0.86	−0.11												
2	L1		0.86	0.82												
3	L2		0.52	2.96												
4	L2		0.80	−1.55												
5	L3		0.96	−0.65	0.20											
6	L3		0.90	0.78	0.25											
7	GR	5	1.48	−1.73	−0.37	0.89	2.24									
8	GR	5	1.39	−2.63	−1.16	0.04	1.44									
9	NO	4	1.77	−0.34	−0.08	−1.35	1.92	−0.36	0.06	−1.63						
10	NO	4	−0.54	−0.65	−0.32	1.51	−0.63	−0.96	−0.42	2.02						
11	NO	4	0.77	−0.27	−0.61	0.11	2.33	−0.83	−1.08	−0.42						
12	BS	5	−3.51	0.59	0.11	3.70	−0.89	3.73	−0.08	3.96	−7.99	0.37	0.24	0.33	0.32	0.12
13	BS	5	−2.03	−0.54	0.89	2.10	−0.42	−1.15	−0.23	0.11	1.41	−0.14	0.03	0.43	0.47	0.07
14	BS	5	−3.79	2.47	0.43	0.34	0.54	0.15	−0.31	0.50	0.01	−0.35	0.22	0.29	0.22	0.27
テスト冊子1																
1	L1		0.86	−0.22												
2	L1		0.86	0.90												
3	L2		0.58	2.43												
4	L2		0.72	−1.65												
5	L3		0.85	−0.65	0.24											
6	L3		0.71	0.75	0.26											
7	G	5	1.38	−1.89	−0.42	0.81	2.51									
8	GR	5	1.42	−2.35	−1.07	0.10	1.55									
9	NO	4	1.35	−0.25	−0.17	−0.92	1.75	−0.35	0.02	−1.42						
10	NO	4	−0.33	−0.83	−0.41	1.56	−0.30	−1.25	−0.49	2.04						
11	NO	4	0.79	−0.32	−0.16	−0.31	2.23	−0.65	−1.03	−0.56						
12	BS	5	−2.02	0.21	1.59	−1.88	2.11	0.30	0.39	1.97	0.11	−2.77	0.28	0.47	0.02	0.23
13	BS	5	−1.61	−0.43	0.94	1.63	−0.54	−0.33	−0.68	−0.29	1.44	−0.14	0.14	0.40	0.43	0.02
14	BS	5	−2.71	1.74	0.22	0.12	0.63	0.46	−0.05	0.31	−0.03	−0.69	0.26	0.26	0.25	0.23

M_j: カテゴリ数
L1: 1母数ロジスティック・モデル
L2: 2母数ロジスティック・モデル
L3: 3母数ロジスティック・モデル
GR: 段階反応モデル
NO: 名義反応モデル
BS: 多肢選択モデル

目母数も出力し，それをそのままMULTILOGとPARSCALEで利用することができる。本稿のプログラムの方がKim & Kolen (2003) のプログラムよりも汎用性は高いと言える。

引用文献

Baker, F.B. (1992). Equating tests under the graded response model. *Applied Psychological Measurement*, **16**, 87-96.

Baker, F.B. (1993a). EQUATE 2.0: A computer program for the characteristic curve method of IRT equating. *Applied Psychological Measurement*, **17**, 20.

Baker, F.B. (1993b). Equating tests under the nominal response model. *Applied Psychological Measurement*, **17**, 239-251.

Baker, F.B. (1997). Empirical sampling distributions of equating coefficients for graded and nominal response instruments. *Applied Psychological Measurement*, **21**, 157-172.

Baker, F.B. & Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques*. New York: Marcel Dekker.

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*. New York: Addison-Wesley.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two ormore nominal categories. *Psychometrika*, **37**, 29-51.

富士通株式会社 (1987). 科学用サブルーチンライブラリ．富士通株式会社.

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

服部　環 (1998). 一般化部分採点モデルの母数等化－最小χ自乗法と特性曲線法－　日本心理学会第62回大会発表論文集　417.

服部　環 (2004a). 共通項目法による一般化部分採点モデルの項目母数の等化－対称性を持たせた等化係数の推定－　日本応用心理学会第71回大会発表論文集，73.

服部　環 (2004b). MULTILOGで利用できるモデルの共通項目法による等化　第2回日本テスト学会発表論文抄録集，10-13.

服部　環 (2004c). 段階反応モデルにおいて対称性を有する等化係数の推定　日本言語テスト学会第8回全国研究大会配付資料.

服部　環 (2004d). 1・2・3母数ロジスティックモデル (1-,2-,3-Prameter Logistic Model) http://www.human.tsukuba.ac.jp/˜hattori/irt/irt.html

服部　環 (2004e). 段階反応モデル (Graded Response Model) http://www.human.tsukuba.ac.jp/˜hattori/irt/irt.html

服部　環 (2005a). 一般化部分採点モデルの項目母数の等化　学筑波大学心理学研究，**29**巻，41-45.

服部　環 (2005b). 情報量の差異を最小化する等化係数の推定－一般化部分採点モデルの場合－　日本応用心理学会第72回大会発表論文集，84.（福島学院大学，9月）

服部　環 (2005c). 段階反応モデルにおいて対称性を有する等化係数の推定　日本心理学会第69回大会発表論文集，417（慶応大学，9月）

Kim, S. & Kolen, M.J. (2003). *POLYST: A Computer Program for Polytomous IRT Scale Transformation Version 1.0*. Iowa City, IA: The University of Iowa.

Kolen, J.M. & Brennan, R.L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices*. (2nd Ed.) New York: Springer.

Loyd, B.H. & Hoover, H.D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, **17**, 179-193.

前川眞一・菊地賢一 (2001). 段階反応モデルのパラメタの共通尺度への等化方法　日本行動計量学会第29回大会発表論文集，200-203.

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

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, **47**, 149-174.

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

Muraki, E. & Bock, R.D. (2003). *PARSCALE 4: IRT Scaling, Item Analysis, and Scoring or Rating Scale Data*. Chicago, IL: Scientific Software International.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 17.

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

Thissen, D. & Steinberg, L. (1984). A response model for multiple-choice items. *Psychometrika*, 49, 501-519.

Thissen, D., Chen, W.-H. & Bock, R.D. (2003). *MULTILOG 7 : Multiple, Categorical Item Analysis and Test Scoring using Item Response Theory.* Chicago, IL: Scientific Software International.

Zimowski, M.F., Muraki, E., Mislevy, R.J. & Bock, R.D. (2003). *BILOG-MG 3: Multiple-Group IRT Analysis and Test Maintenance for Binary Items*. Chicago, IL: Scientific Software International.

（受稿９月30日：受理10月26日）