283 Japanese Journal of Educational psychology, 1989, 37, 283—285

資　料

項目反応モデルにおける信頼性係数の推定法

豊　田　秀　樹*

THE METHODS FOR ESTIMATING THE RELIABILITY COEFFICIENT UNDER ITEM RESPONSE MODEL

Hideki TOYODA

The purpose of this study is to propose the three methods for estimating the reliability coefficient under Item Response model. The first method may be used for the item selection, if the latent trait distribution is assumed to be known. The second method gives the upper bound of the reliability coefficient estimate. By the third method, the estimate of reliability coefficient may be calculated from the two following sample statistics only : the sample mean and variance of the estimated values on the latent trait. Some examples using the Scale for Word Meaning Comprehension (Shiba & Noguchi, 1982) are given as application.

Key words : Item Response Model, reliability coefficient, item selection, the Scale for Word Meaning Comprehension (Shiba & Noguchi, 1982).

目　　的

テスト理論には古典的テスト理論と項目反応理論という2つの大きな理論体系があり，テストの精度を表わす指標として，古典的テスト理論ではテスト得点の分散に対する真の得点の分散の比である信頼性係数を用い，項目反応理論では能力尺度 θ の関数でその最尤推定量 $\hat{\theta}$ の漸近分散の逆数を与えるテスト情報関数

$$I(\theta)=\sum_{j=1}^{n}\left[(P_j(\theta)(1-P_j(\theta)))^{-1}\left\{\frac{\partial}{\partial\theta}P_j(\theta)\right\}^2\right] \quad (1)$$

を用いるのが一般的である (Lord, 1980)。n はテストを構成する項目の数であり，$P_j(\theta)$ は尺度値が θ である被験者が項目 j に正答する確率を表わす。両者を比較すると，信頼性係数はテストを受ける被験者の特性値の分布に依存して定義されるのに対して，テスト情報関数は独立に定義することができる。あるいは，信頼性系数は被験者集団全体に対する精度の指標であるのに対して，テスト情報関数は θ のレベルごとに精度を知ることができるなど，従来は信頼性係数に対するテスト情報関数の有用性ばかりが論じられてきた．

しかしテストを行う主な目的のひとつは被験者間の特性値の差を識別することであり，測定誤差の分散が一定であっても特性値の分散が大きい場合は小さい場合よりもその識別が容易であることは事実であるから，同一のテストに対して前者の精度を高く，後者の精度を低く評価する信頼性係数は実用的に妥当な指標であり，テスト情報関数が提供し得ない知見を与えてくれる．そこで本研究では項目反応モデルにおける項目固定型テストの信頼性係数を推定する方法を提案し，その有用性を考察する。

I　潜在特性の分布に基づく信頼性の推定

テストを構成する項目のパラメータと，テストを実施する被験者集団の潜在特性 θ の分布 g（平均 μ_θ，分散

* 大学入試センター研究開発部 (Research Devision : National Center for University Entrance Examination)

$\sigma^2{}_\theta$）が所与であるときに，そのテストの信頼性係数を推定する方法を述べる。

θ と $\hat{\theta}$ の関係を，推定誤差 E を用いて

$$\hat{\theta}=\theta+E \tag{2}$$

と表わすと，Eは漸近的に平均0・分散 $I(\theta)^{-1}$ の正規分布に従う (Hambleton & Swaminathan, 1985)。ここで n がある程度以上大きいときは，

$$\varepsilon(E|\theta)=0 \tag{3}$$

$$\sigma^2{}_E|\theta=I(\theta)^{-1} \tag{4}$$

が成り立つと仮定すると，E と θ は統計的に独立ではないが，E の θ への回帰直線の傾きは0となるから，両者の共分散は

$$\sigma_{\theta E}=0 \tag{5}$$

が成り立つ。(3)・(4)式の仮定の近似は $n\geqq 20$ のときに十分よいものであることが報告されている (Samejima, 1978)。(5)式より $\hat{\theta}$ の分散は古典的テストモデルと同様に

$$\sigma_{\hat{\theta}}{}^2=\sigma_\theta{}^2+\sigma_E{}^2 \tag{6}$$

と，θ と E の分散の和で表わすことができる。E の分散

$$\sigma_E{}^2=\int(E-\mu_E)^2 f(E)dE \tag{7}$$

は（ただし μ_E は E の分布の平均），(3)，(4)より

$$=\int E^2\int f(E|\theta)g(\theta)d\theta dE \tag{8}$$

$$=\int I(\theta)^{-1}g(\theta)d\theta \tag{9}$$

で得られるから，所与である $\sigma_\theta{}^2$ を用いて信頼性係数は

$$\rho=\frac{\sigma_\theta{}^2}{\sigma_{\hat{\theta}}{}^2}=\frac{\sigma_\theta{}^2}{\sigma_\theta{}^2+\sigma_E{}^2} \tag{10}$$

によって推定される。

以上の議論は $g(\theta)$ の分布型によらず常に成り立つが，本論文ではとりあえず正規分布を仮定し，(9)式の評価はガウスエルミート積分公式（分点数は９）によって行う。

数　値　例

芝・野口 (1982) の語彙理解力尺度項目特性一覧表中の項目を用いて，項目数の異なる７つのテストを作製したと想定し，$\mu_\theta=0.0$，$\sigma_\theta{}^2=1.0$ として標準誤差 σ_E と信頼性係数 ρ を推定し，TABLE 1 に示した。７つのテストは，それを構成する項目の困難度が -1.8 から 1.8 の間に一様に分布しているが，項目数が増えるに従って標準誤差は小さくなり，信頼性係数は高くなっている。

TABLE 1　項目数の変化に伴う標準誤差と信頼性係数の変化

項目数	20	25	30	37	50	75	150
σ_E	.518	.511	.443	.401	.341	.281	.197
$\rho_{\hat{\theta}}$	.789	.793	.836	.861	.896	.927	.962

本方法の長所は，テストの実施データではなく項目パラメータを入力データとして利用できることであり，本節の例の様に項目プールが用意されている場合は，十分な信頼性係数の値が得られるまで項目の取捨選択と信頼性係数の推定を交互にくり返すことが可能である。このため本方法は，テストを実施する前の段階で，項目選択に利用することができる。

II　信頼性係数の最大値の推定

テストを構成する項目のパラメータのみが所与であるときに，信頼性が最大になる θ の分布の平均 $\mu_{\theta M}$ と分散 $\sigma^2{}_{\theta M}$ と，その際の信頼性係数 ρ_{MAX} を推定するためには，(10)式を関数

$$\rho=h(\mu_\theta,\ \sigma_\theta{}^2) \tag{11}$$

と見なし，ρ を最大にする μ_θ と $\sigma_\theta{}^2$ を求めればよい。(11)式の最大化は豊田 (1986) で示された数値微分 ($d=10^{-8}$) を用いた最急降下法（パラメータの初期値は $\mu_\theta=0.0$，$\sigma_\theta{}^2=1.0$，ステップサイズの初期値 $\alpha=1$，収束基準は微係数の絶対値の最大値が 10^{-4} 以下）により行う。

III　標本平均・分散による信頼性の推定

項目パラメータが所与である項目固定型テストを実施し，$\hat{\theta}$ の標本平均 $\hat{\mu}_{\hat{\theta}}$ と標本分散 $\hat{\sigma}_{\hat{\theta}}{}^2$ が得られたときに信頼性係数を推定する方法を述べる。

σ_E は μ_θ と $\sigma_\theta{}^2$ の関数であるが，(2)式より

$$\mu_\theta=\mu_{\hat{\theta}} \tag{12}$$

が成り立つから，$\hat{\mu}_{\hat{\theta}}$ を μ_θ の推定値 $\hat{\mu}_\theta$ として用い，値を固定すると，(6)式は $\sigma_\theta{}^2$ のみの関数

$$\sigma_{\hat{\theta}}{}^2=t(\sigma_\theta{}^2) \tag{13}$$

と見なすことができる。(13)式左辺を $\hat{\sigma}_{\hat{\theta}}{}^2$ に固定し，$\sigma_\theta{}^2$ について方程式を解き，その値を $\sigma_\theta{}^2$ の推定値 $\hat{\sigma}_\theta{}^2$ とする。$\hat{\mu}_\theta$ と $\hat{\sigma}_\theta{}^2$ を用いて(10)式を評価し，信頼性係数を推定する。(13)式の方程式は二分法（収束基準は10^{-4}）によって解く。解をはさむ２つの初期値には $\hat{\sigma}_{\hat{\theta}}{}^2$ と $\hat{\sigma}_{\hat{\theta}}{}^2/10.0$ を用いる。

適　用　例

IIで提案した方法によって，芝・野口 (1982) で作製されたテストＢ１版からＢ５版までの $\mu_{\theta M}$・$\sigma_{\theta M}$・ρ_{MAX}，その際の標準誤差 σ_{EM} を求め，項目数と共に TABLE 2 に示す。Ｂ１版からＢ５版はそれぞれ小学校１年生から５年生が受験することを想定したテストであるから，$\mu_{\theta\mathrm{M}}$ と $\sigma_{\theta\mathrm{M}}$ がＢ１版からＢ５版まで順に大きくなっているのは妥当な結果といえる。この様に信頼性係数の最

TABLE 2 信頼性を最大にする θ の分布の平均と標準偏差およびその時の標準誤差と信頼性係数

版	B 1	B 2	B 3	B 4	B 5
項目数	34	36	38	40	45
$\mu_{\theta M}$	−4.155	−3.602	−3.075	−2.176	−1.153
$\sigma_{\theta M}$	1.021	1.181	1.280	1.363	1.366
σ_{EM}	.321	.358	.377	.388	.376
ρ_{MAX}	.910	.916	.920	.925	.930

TABLE 3 $\hat{\theta}$ の標本平均と標本標準偏差から推定された θ の分布の標準偏差と標準誤差と信頼性係数

版	B 1	B 2	B 3	B 4	B 5
$\hat{\sigma}_{\hat{\theta}}$	.53	.60	.63	.72	.99
$\hat{\mu}_{\hat{\theta}}=\hat{\mu}_{\theta}$	−3.83	−3.23	−2.54	−1.71	−0.99
$\hat{\sigma}_{\theta}$	.472	.538	.561	.658	.853
$\hat{\sigma}_{E}$	.241	.267	.286	.293	.287
$\hat{\rho}$	.793	.803	.794	.834	.899

大値を推定することによって，当該テストを如何なる被験者集団に実施するのが最適であるかを知ることができる。

芝・野口(1982)の表4で与えられた $\hat{\theta}$ の標本平均 $\hat{\mu}_{\hat{\theta}}$ と標本標準偏差 $\hat{\sigma}_{\hat{\theta}}$ を用い，Ⅲで提案した方法によって θ の標準偏差と標準誤差と信頼性係数を推定し(それぞれ $\hat{\sigma}_{\theta}\cdot\hat{\sigma}_{E}\cdot\hat{\rho}$)，TABLE 3 に示す。$\hat{\rho}$ は ρ_{MAX} に比べて.03〜.11低く推定されている。全版に亘って $\hat{\mu}_{\theta}$ は $\mu_{\theta M}$ より高く，$\hat{\sigma}_{\theta}$ は $\sigma_{\theta M}$ より小さいが，$\hat{\sigma}_{E}$ は σ_{EM} より小さい。Ⅲで提案した方法は，項目に対する被験者の反応パタンや $\hat{\theta}$ の個々の値ではなく，$\hat{\mu}_{\hat{\theta}}$ と $\hat{\sigma}^2_{\hat{\theta}}$ のみを入力データとして用い，実施データを十分に利用する方法とはいえないが，それ等は項目パラメータと共に基本統計量として示されることが多く，実施データが手元にない場合にも信頼性係数を推定することができるという長所がある。

結　論

項目固定型テストは通常それを受ける特定の被験者集団を想定して作製されるのだから，精度の評価を特性値の分布を考慮して(分布に依存させて)行うことは信頼性係数の欠点とはならない。たとえば TABLE 3 で示した様に B 1 版から B 5 版の精度をそれぞれのテストの対象学年である小学校1年生から5年生の被験者集団を用いて評価することは実用的に意味がある。また，θ によらない無条件誤差分散である $\sigma_E{}^2$ を推定することは，想定された母集団から無作為に選ばれたひとりの被験者の測定誤差の期待値を確認することを意味し，テスト情報関数を用いて θ のレベルごとにそれを確認するのと同程度に有用である。

以上のことから，項目反応モデルにおける項目固定型テストの精度を評価する際にはテスト情報関数を用いるばかりでなく，本研究で提案した方法により信頼性係数と標準誤差を推定し，併用することは有益であろう。

付　記

本研究の計算は東京大学大型計算機センター M680—Hシステムで最適化パスカルを用いて行った。

引用文献

Hambleton, R.K. & Swaminathan, H. 1985 Item Response Theory Principles and Applications. Kluwer Nijhoff Publishing.

Lord, F.M. 1980 Applications of Item Response Theory to Practical Testing Problems. Lawrence Erlbaum Associate.

Samejima, F. 1977 A use of the information function in tailored testing. Applied Psychological Measurement, 233—247.

芝 祐順・野口裕之 1982 語彙理解力尺度の研究 I 東京大学教育学部紀要，22，31—42.

豊田秀樹 1986 被験者の推定尺度値とテスト情報関数を利用した潜在特性尺度の等化法，教育心理学研究，34，163—167.

謝　辞

本研究をまとめる際にいろいろと指導して下さった芝祐順先生(東京大学教育学部教授)に深謝します。

(1988年9月20日受稿)