第３章　IRT（項目反応理論）による調査結果の分析

３．１　IRT の概要

（１）IRT とは

IRT（項目反応理論）とは，Item Response Theory の略で，項目応答理論とも呼ばれるテスト理論である．IRT では，ある分野（例えば高校数学）の複数の問題（IRT では「項目」という）で構成される問題セットによる試験を，ある集団（例えば理系の大学受験生）の多数の受験者に受けさせて，各問題への正答・誤答（反応）から，問題の特性（困難度，識別力）と受験者の能力値を推定できる．IRT の利点としては次が挙げられる．

- 試験結果から問題の困難度等の特性と受験者の能力を同時に推定できる（項目パラメータと能力値の推定）．
- 異なる問題セットを用いた試験結果の間でも同一尺度の能力値で比較できる（項目パラメータの等化により可能）．
- 受験者集団に適した問題セットを作成できる（テスト情報量等の指標により，対象集団の能力値範囲を精度良く推定できる問題セットを作成可能）．

（２）IRT 適用の前提条件

IRT を学力調査に適用するためには次の前提条件が必要となる．

- IRT で推定できる対象分野（例えば高校数学）の能力は，単一の能力値 θ で評価できる（一次元性が成り立つ）．
- 受験者の各問題への反応は正答または誤答のどちらかであり，能力値 θ の受験者が問題 j に正答する確率は θ の単調増加関数 $P_j(\theta)$で表させる．従って問題 j に誤答する確率は $1-P_j(\theta)$ となる（反応はベルヌーイ分布）．
- 能力値 θ の受験者の各問題への反応は互いに独立である（局所独立性が成り立つ）．

本学力調査の試験結果においても，以上の条件が成り立つと仮定した．

（３）項目特性曲線

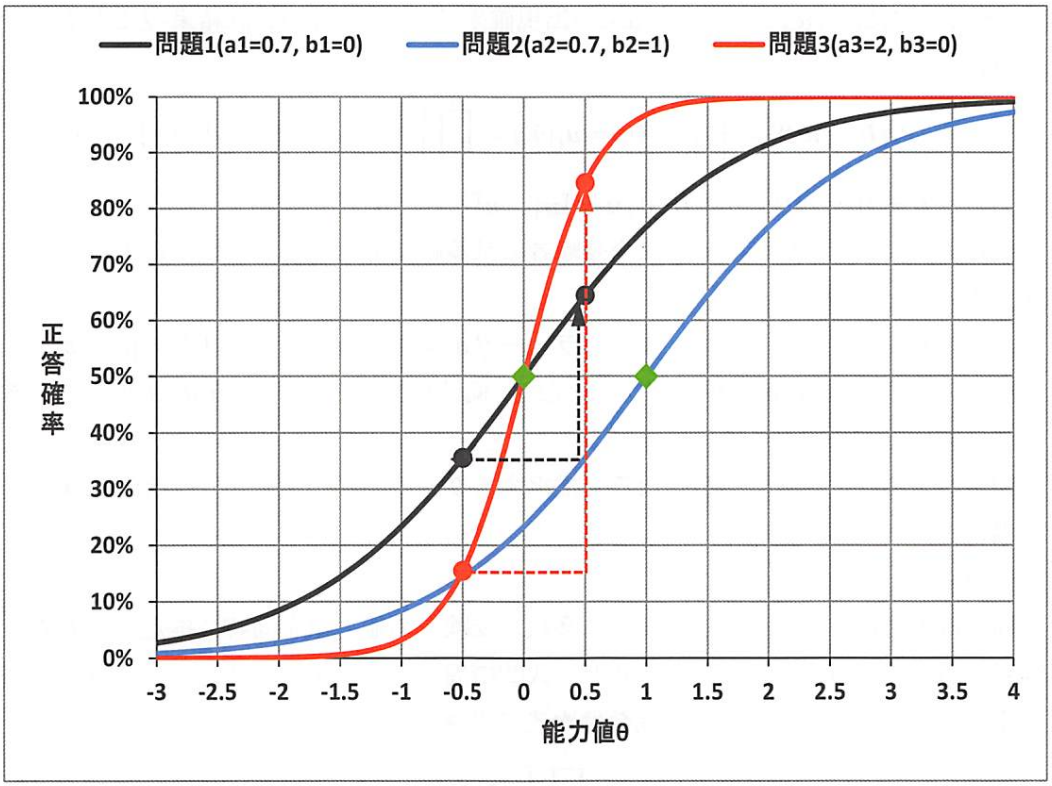
能力値 θ に対する問題 j の正答確率 $P_j(\theta)$ を，θ の関数曲線とみなし「項目特性曲線」(Item Characteristic Curve, ICC)という．ICC として当分析では以下の２パラメータロジスティック（2PL）モデルを用いた．

$$P_j(\theta)=\frac{1}{1+\exp\left(-Da_j\left(\theta-b_j\right)\right)},D=1.7$$

ここで，a_j は「識別力パラメータ」$(a_j>0)$，b_j は「困難度パラメータ」という問題固有のパラメータで，２つのパラメータをまとめて「項目パラメータ」という．

ロジスティックモデルの ICC の形状は，変曲点の位置と変曲点での傾きで規定される．2PL モデルの場合，変曲点の位置は正答確率が 50%となる能力値である．2PL モデル ICC の形状と項目パラメータの関係を以下に示す．

図 3.1　2PL モデル ICC の形状と特性



- 困難度パラメータ：b_j

  ICC の変曲点の位置，すなわち正答確率が 50%になる能力値．この値が大きいほど問題が難しい．図 3.1 では問題 1 より問題 2 の方が困難度が高い(緑色の点)．

- 識別力パラメータ：a_j

  ICC の変曲点での傾きに比例するパラメータ．この値が大きくなると傾きが急になり，変曲点付近で正答確率が急激に増大．能力値推定でこの値が大きい問題を用いるほど精度が向上．図 3.1 では，識別力の低い問題 1(黒い曲線)では，能力値が-0.5～0.5 の範囲で正答確率が 35～65%しか変化しないが，識別力が高い問題 3(赤い曲線)では，15～85%も変化している．

（4）項目パラメータの推定

受験者 i の問題 j への反応（正答または誤答）の確率変数を U_{ij}，その実現値を u_{ij} とし

$$u_{ij} = \begin{cases} 1\,(\text{正答}) \\ 0\,(\text{誤答}) \end{cases}$$

とすると，受験者 i の問題 j への反応確率は次式になる．

$$\Pr\left(U_{ij} = u_{ij}|\theta\right) = P_j(\theta)^{u_{ij}}\left(1 - P_j(\theta)\right)^{1-u_{ij}}$$

受験者 i が問題 $1, 2, \cdots, m$ で構成される問題セットに受験した場合，これらの問題への反応の確率変数ベクトルを $\boldsymbol{U}_i = [U_{i1} \quad U_{i2} \quad \cdots \quad U_{im}]$，その実現値ベクトル（試験結果）を $\boldsymbol{u}_i = [u_{i1} \quad u_{i2} \quad \cdots \quad u_{im}]$ とすると，局所独立性の仮定から試験結果 $\boldsymbol{u}_i$ の発生確率は次式になる．

$$\Pr(\boldsymbol{U}_i = \boldsymbol{u}_i|\theta_i) = \prod_{j=1}^{m} \Pr\left(U_{ij} = u_{ij}|\theta_i\right) = \prod_{j=1}^{m} P_j(\theta_i)^{u_{ij}}\left(1 - P_j(\theta_i)\right)^{1-u_{ij}} \cdots ①$$

- 同時最尤法

全受験者 $1,2,\cdots,n$ による試験結果の出現確率は，各受験者の試験結果の発生確率の積で式①を用いて次式になる．

$$\Pr(\boldsymbol{U}=\boldsymbol{u}|\boldsymbol{\theta})=\prod_{i=1}^{n}\Pr(\boldsymbol{U}_i=\boldsymbol{u}_i|\theta_I)=\prod_{i=1}^{n}\prod_{j=1}^{m}P_j(\theta_i)^{u_{ij}}\left(1-P_j(\theta_i)\right)^{1-u_{ij}}\cdots②$$

ここで，$\boldsymbol{U}=[\boldsymbol{U}_1^T\quad\boldsymbol{U}_2^T\quad\cdots\quad\boldsymbol{U}_n^T]^T,\boldsymbol{u}=[\boldsymbol{u}_1^T\quad\boldsymbol{u}_2^T\quad\cdots\quad\boldsymbol{u}_n^T]^T$ とする．

採点後，式②で既知であるのは全受験者の試験結果 $\boldsymbol{u}$ だけで，式②に含まれる項目パラメータ，能力値は未知である．

式②は $\boldsymbol{u}$ を与えられた場合，項目パラメータ $\boldsymbol{a}=[a_1\quad a_2\quad\cdots\quad a_m],\boldsymbol{b}=[b_1\quad b_2\quad\cdots\quad b_m]$ および $\boldsymbol{\theta}$ の尤度関数（$L(\boldsymbol{a},\boldsymbol{b},\boldsymbol{\theta})$ とする）となる．同時最尤法とは，$L(\boldsymbol{a},\boldsymbol{b},\boldsymbol{\theta})$ が最大となる $\boldsymbol{a},\boldsymbol{b},\boldsymbol{\theta}$ を同時に求める方法である．

同時最尤法は，受験者が多くなると推定する $\boldsymbol{\theta}$ の要素数が多くなり，数値計算として解けない場合がほとんどである．

- 周辺最尤法

周辺最尤法は，同時最尤法のように多数の受験者の能力値を同時に推定するのではなく，受験者の能力値を事前分布として仮定し，式②の周辺分布を尤度関数として，項目パラメータを推定する方法である．その尤度関数は次式で表される．

$$L_M(\boldsymbol{a},\boldsymbol{b})=\prod_{i=1}^{n}\int_{-\infty}^{\infty}\Pr(\boldsymbol{U}_i=\boldsymbol{u}_i|\theta_i)\,g(\theta_i)d\theta_i$$

ここで，$g(\theta_i)$ は能力値の事前分布の確率密度関数で，事前分布は標準正規分布が仮定されることが多い．

同時最尤法は EM（Expectation-Maximization）アルゴリズムにより多くのケースで項目パラメータを推定できる．本分析でも事前分布に標準正規分布を仮定し，周辺最尤法により項目パラメータの推定を行った．

（5）能力値の推定

求めた項目パラメータを用いて，問題 $1\sim m$ の問題セットを受けた受験者 k の能力値を最尤法により推定する（k は受験者集団 $1\sim n$ に含まれている必要はない）．

受験者 k の試験結果を $[u_{k1}\quad u_{k2}\quad\cdots\quad u_{km}]$，能力値を θ とすると，尤度関数 $L_T(\theta)$ は

$$L_T(\theta)=\prod_{j=1}^{m}P_j(\theta)^{u_{kj}}\left(1-P_j(\theta)\right)^{1-u_{kj}}$$

であり，$L_T(\theta)$ が最大となる θ を受験者 k の能力推定値とする．

（6）項目パラメータの等化(Equating)

項目パラメータの等化の目的は，異なる問題セット，異なる受験者の集団でも，同一尺度の能力値で評価を可能にすることである．そのために全ての問題セットに共通する項目パラメータを定める．

本分析では，年度間で共通問題が複数あるので共通項目法の Mean & Sigma 法により次の手順で等化を行った．

A) ベースとなる問題セット(ここではベースセットという)を定める（本分析では比較的受験者が多い 2007 年度の問題セット）．

B) ベースセット以外の問題セットで共通問題以外のパラメータは，両者の共通問題のパラメータの平均と標準偏差を用いて，線形変換により推定し，ベースセットのパラメータに加える．

C) ベースセット以外の年度も問題セットについて，B)の操作を繰り返し，全年度のパラメータを等化する．

3．2　基礎学力調査結果の分析

（1）IRT による分析手順

基礎学力調査結果から次の手順で項目パラメータと能力値を推定した．

A) 2005～2014 年度の結果を，年度毎に周辺最尤法で項目パラメータを推定

B) 2007 年度のパラメータを基準に，全 86 問のパラメータを等化

C) 等化したパラメータを用いて，最尤法により全問正解・全問不正解を除いた受験者 42,350 人の能力値を推定

（2）項目パラメータの推定結果

図 3.2 は，ある年度の問題セットごとの項目パラメータを２次元座標(困難度，識別力)上にプロットしたものである．この図から記述式問題（図 3.2 の赤い点）が比較的困難度が高いことが判る．

図 3.2　問題毎の項目パラメータ

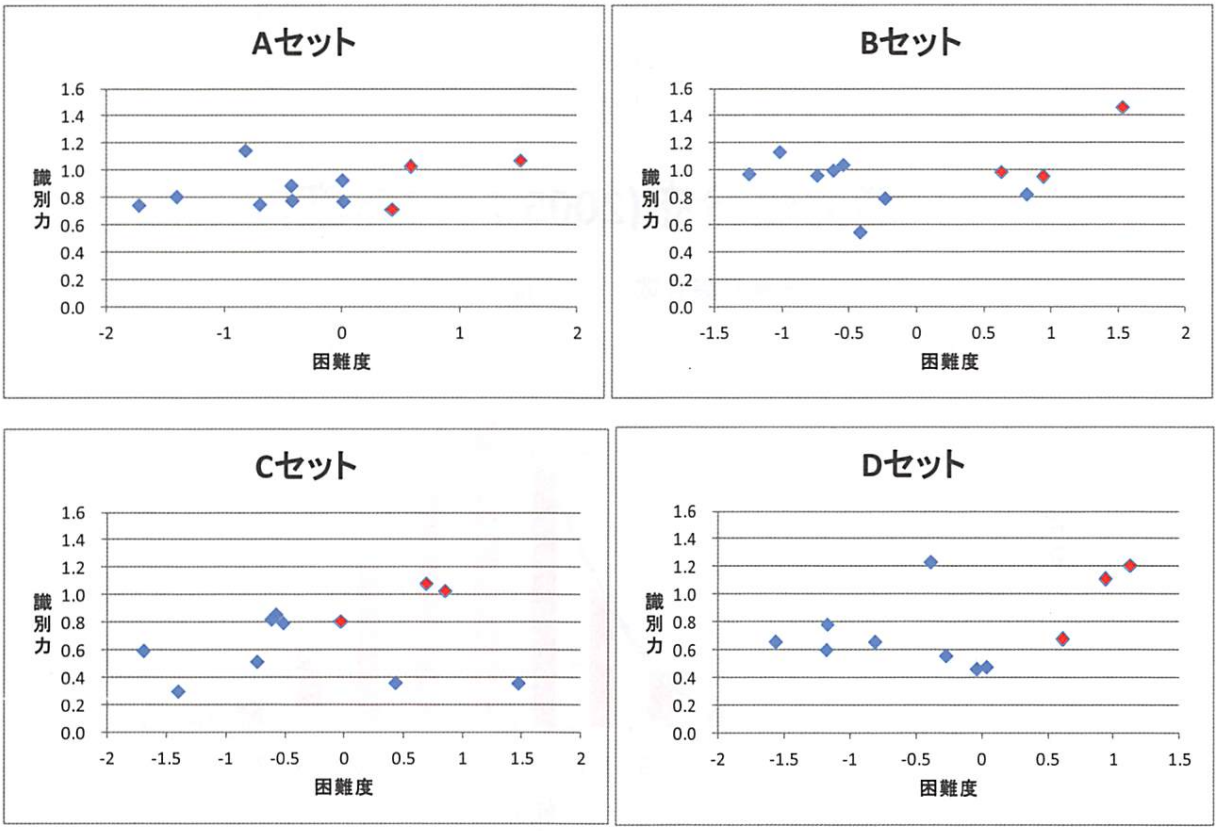
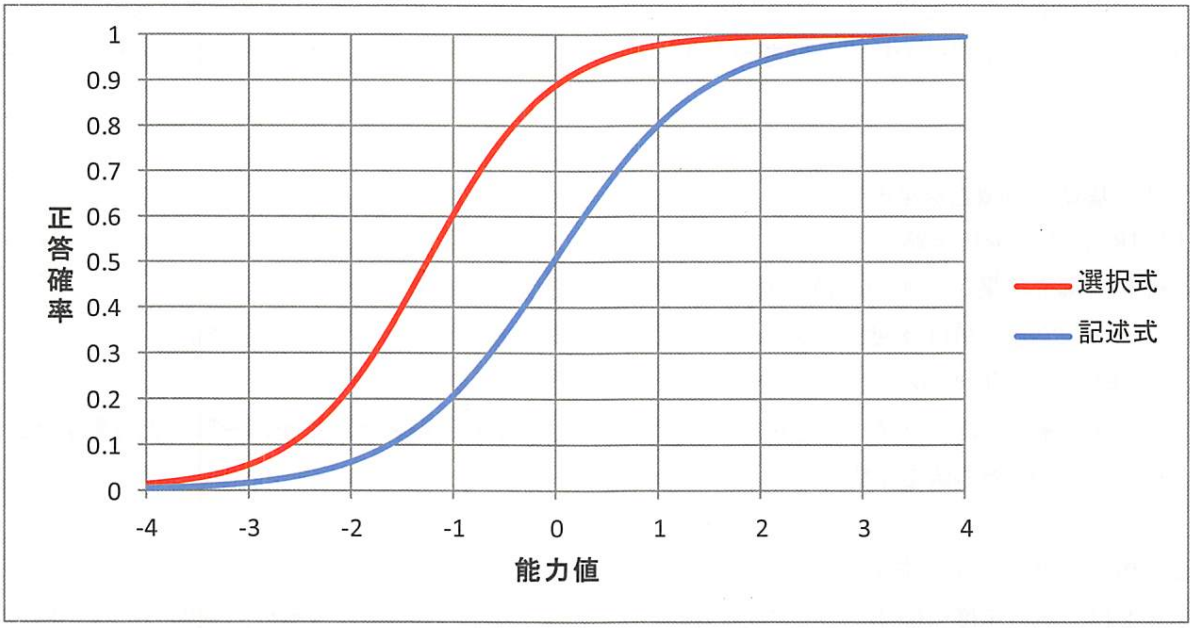


図 3.3 は，同じ内容の問題で出題形式を選択式にした場合と記述式にした場合の ICC である．同じ内容の問題でも出題形式が選択式よりも記述式のほうが困難度が高くなることが判る．

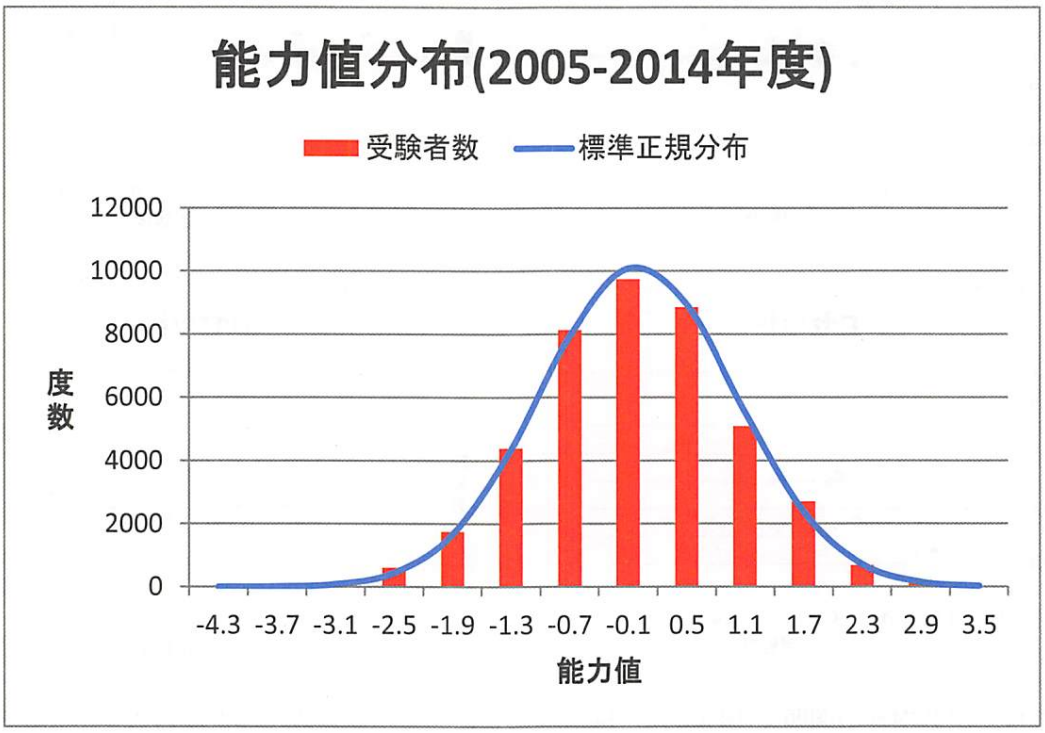
図 3.3　選択式問題と記述式問題の ICC



（3）能力値の推定結果

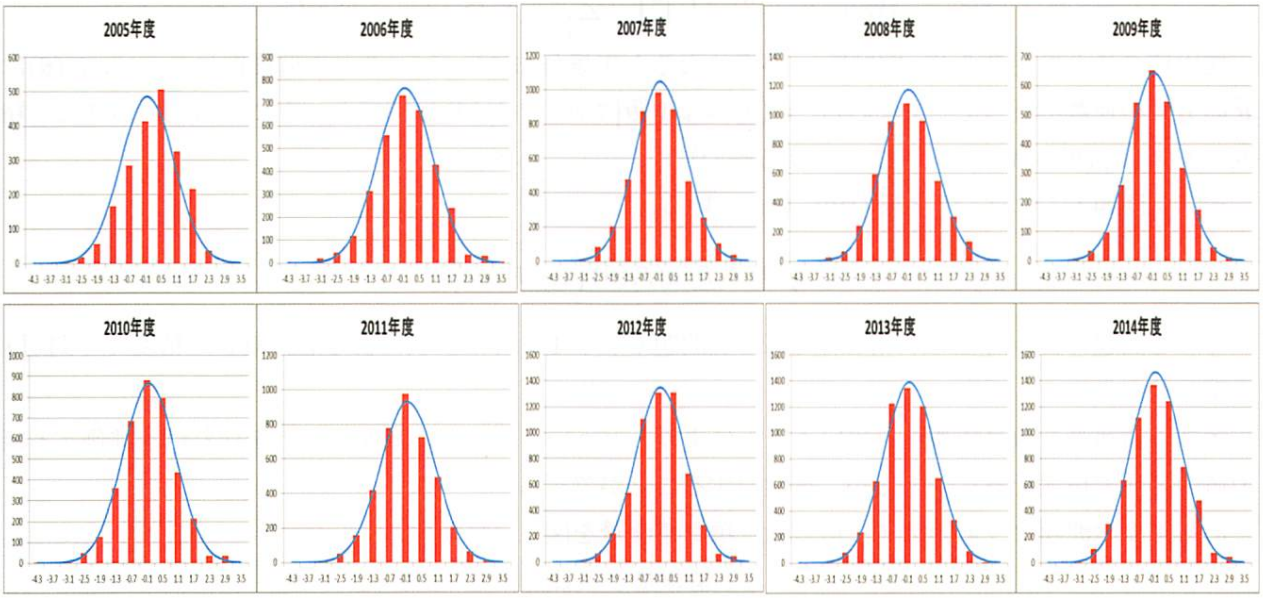
図 3.4 に 2005 年度から 2014 年度の全年度の全問正解・全問不正解を除く受験者の能力値分布を示す．この分布の平均は-0.02，標準偏差が 1.014 であり標準正規分布に近い．

図 3.4　能力値分布（全年度）



各年度の能力値分布は図 3.5 の通りであり，2005 年度を除き標準正規分布に近い．

図 3.5　各年度の能力値分布



全問正解・全問不正解を除く受験者の平均能力値と全年度受験している 9 校（「共通 9 校」と記す）の平均能力値の推移を図 3.6 に示す．平均能力値は 2005 年度が高いが，2006 年度以降は大きな変動がないことが判る．共通 9 校の平均は 2008 年度に落ちこみがあるが，その他の年度に大きな変動はないことが判る．

図 3.6　全受験者の平均能力値と共通 9 校の平均能力値の推移

	2005年度	2006年度	2007年度	2008年度	2009年度	2010年度	2011年度	2012年度	2013年度	2014年度	全年度
全受験者数（満点と0点は除く）	2,033	3,199	4,393	4,913	2,698	3,622	3,894	5,648	5,816	6,134	42,350
全受験者の平均能力値	0.225	0.036	-0.049	-0.080	-0.003	-0.014	-0.056	-0.016	-0.075	-0.012	-0.021
共通9校の平均能力値	0.057	0.072	0.008	-0.246	-0.075	-0.098	-0.016	0.021	-0.040	-0.085	-0.046

（4）出題問題セットの評価（テスト情報量と項目情報量による評価）

テスト情報量とは，能力値 θ のフィッシャー情報量であり，能力値を θ としたときに確率変数である試験結果ベクトル $\boldsymbol{u}$ がもつ情報量である．2PL モデルの場合，テスト情報量 $I_T(\theta)$ は次式で表される．

$$I_T(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log L_T(\boldsymbol{u}|\theta)\right)^2 \middle| \theta\right] = \sum_{j=1}^{m} D^2 a_j^2 P_j(\theta)\left(1 - P_j(\theta)\right) \cdots ③$$

能力値の最尤推定量 $\hat{\theta}$ が，問題数 m が多くなると正規分布に近づき，$\hat{\theta}$ が不偏推定量ならば，$I_T(\theta)$は $\hat{\theta}$ の分散の逆数に近づく．すなわち，$I_T(\theta) \fallingdotseq 1/V[\hat{\theta}]$であり，$I_T(\theta)$ は問題セットによる能力値 θ での精度を表す（$I_T(\theta)$ が高いほどθ での精度を高い）．

式③は問題 j についての次式の和になっている．

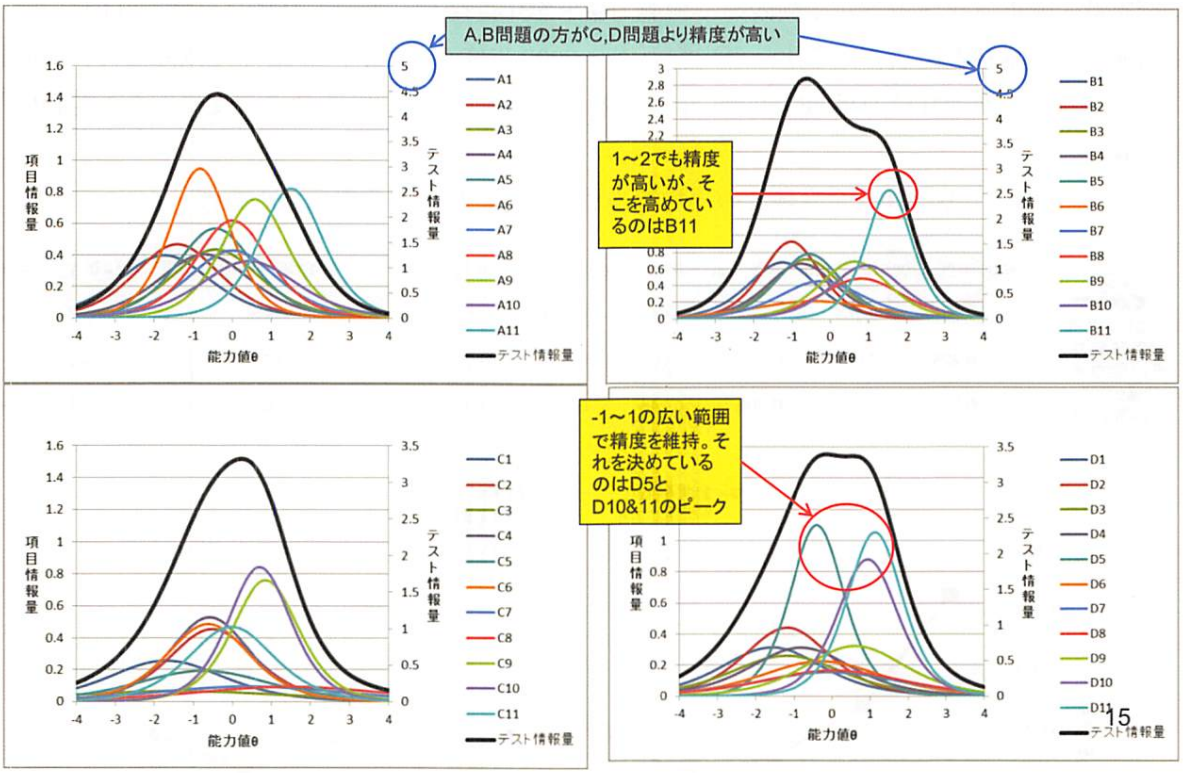
$$I_j(\theta) = D^2 a_j^2 P_j(\theta)\left(1 - P_j(\theta)\right)$$

これを問題（項目）j の項目情報量といい，問題セット $1 \sim m$ の試験による能力値 θ の精度での問題 j の寄与量を表す．

テスト情報量と項目情報量を用いることにより，どういう組合せの問題セットを作れば，評価する能力値の範囲で精度を高くできるか，を検討することができる．

ある年度の問題セットのテスト情報量と項目情報量を図 3.7 に示す．

図 3.7　テスト情報量と項目情報量の計算例



3．3　まとめ

2005 年度から 2014 年度までの学力調査結果を IRT で分析し，次の知見を得た．

- IRT により問題の特性を把握できる．例えば
  - 選択式問題より記述式問題の方が困難度が高い．
  - 同じ設問でも出題形式を変えると特性が変わる．
- 能力値の傾向
  - 能力値は，2005 年度を除き標準正規分布に従う．
  - 平均能力値は，2005 年度が高いが，2006～2014 年度で大きな変化はなし．
- テスト情報量と項目情報量により，問題セットの評価ができる．