２つのテスト理論
―項目反応モデルと認知診断モデル―

東京大学大学院 教育学研究科
岡田謙介

総括的評価と形成的評価

summative assessment　formative assessment

	総括的評価	形成的評価
時期	教育課程を終えた後で	教育課程の中途で
目的	学習目標が達成された程度の把握	学習状況の把握，学習プロセスの改善
結果	達成度・学習到達度（試験得点，合否）	学習改善のための働きかけ，フィードバック
例	期末試験，共通テスト	宿題・課題，診断テスト
統計モデル	**項目反応モデル** (item response theory, **IRT**)	**認知診断モデル** (cognitive diagnostic model, **CDM**)

Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365-379. https://doi.org/10.1080/0969594970040304

IMPS2021のプログラム

- IRTとCDMが現在の計量心理(≈テスト)学の2大潮流

1.1: Configural Complexity in Latent Variable Modeling	1.2: Scoring and Equating	1.3: Cognitive Diagnosis Models: the Q Matrix	1.4: Differential Item Functioning I
2.1: Bayesian Inference	2.2: Outlier Detection and Analysis	2.3: Item Response Theory	2.4: Structural Equation Modeling
3.1: Machine Learning	3.2: Factor Analysis/ Dimensionality Reduction	3.3: IRT with Noncognitive Measures	3.4: Cognitive Diagnostic Models
4.1: Response Times and Cognitive Diagnosis	4.2: Measurement Invariance	4.3: Multilevel Analysis	4.4: Mediation Analysis
5.1: Topics on Cross-Cultural Measurement	5.2: Big Data/Process Data	5.3: Mixture Models	5.4: Missing Data
6.1: Networks	6.2: Rasch Modeling	6.3: DIF and Anchor Selection Issues	6.4: Longitudinal Data Analysis
7.1: Identification Problems in Empirical Research	7.2: Response Times	7.3: Longitudinal Data Analysis	7.4: Time Series Analysis
8.1: Foundational Issues	8.2: Issues in IRT	8.3: Issues in Cognitive Diagnostic Modeling	8.4: Test Security and Aberrant Responding
9.1: Writing Assessment	9.2: Reliability	9.3: Causal Inference	

項目反応モデル(IRT)	テスト理論関連	認知診断モデル(CDM)	

https://www.psychometricsociety.org/imps-2021

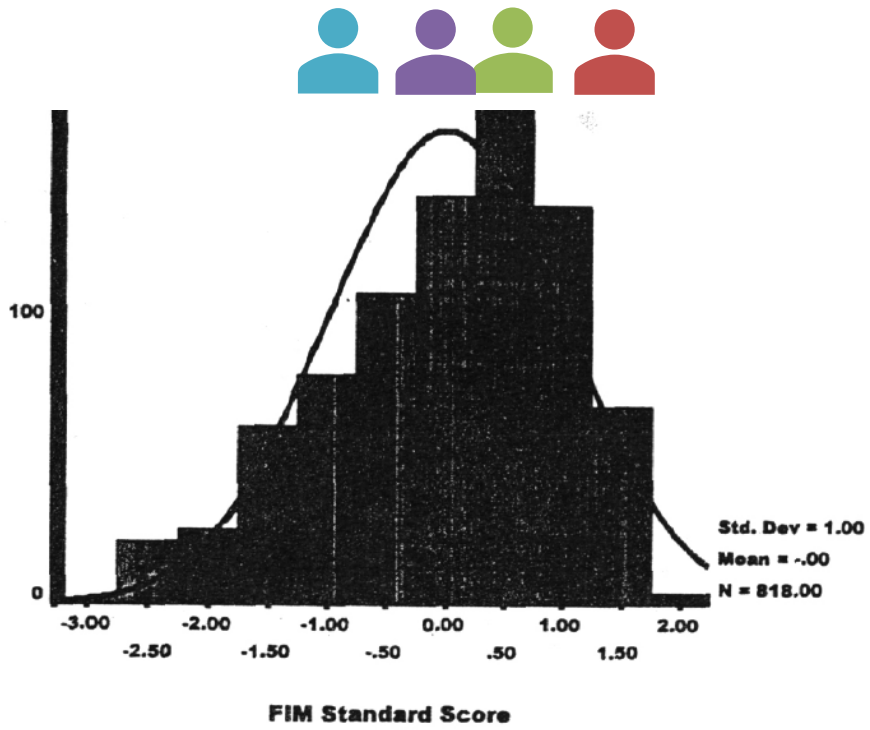
本発表の流れ

- IRTとCDMは現代における**計量心理学・テスト理論の2大モデル**群と言うことができ、両モデルとも活発な研究が進んでいる
- そこで本発表は以下の内容を扱う
  - IRT・CDMの**基本となるモデル**と**使われ方**
  - **両モデル間の関連・相違**と、そこから見えてくること
  - いくつかの**今後の研究の方向性**

古典的テスト理論と項目反応理論

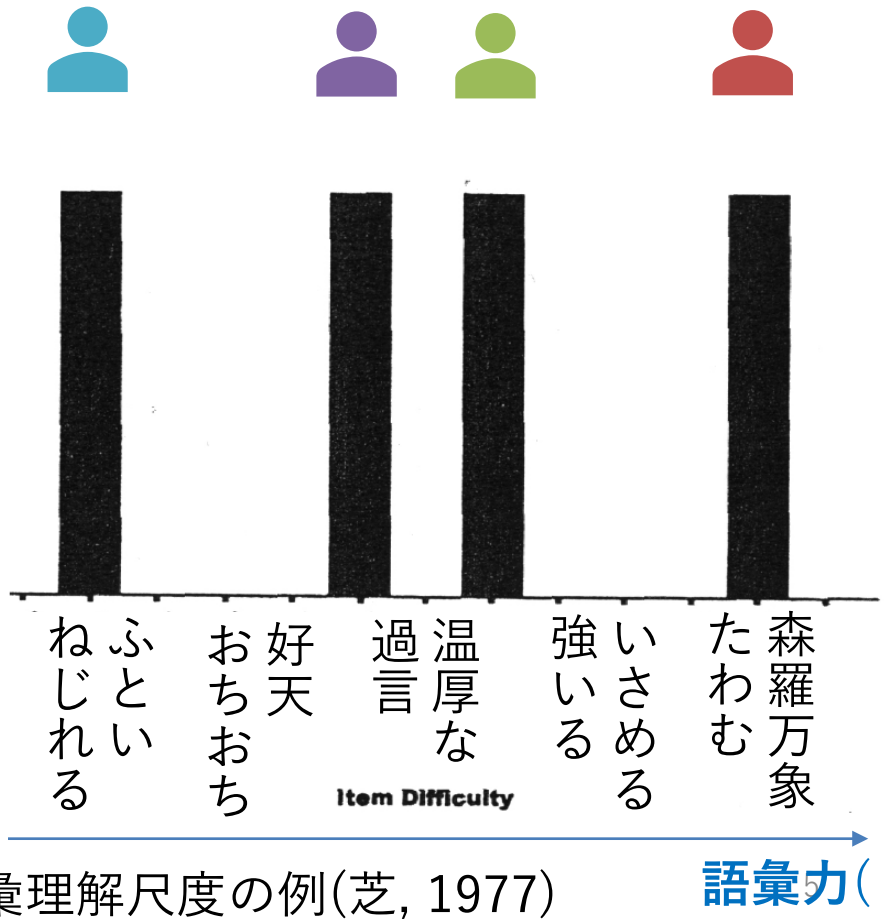
■ **古典的テスト理論**(CTT)

今回の解答者集団中での
相対比較



■ **項目反応理論(IRT)**

問題項目で規程される次元上
での**絶対評価**



語彙理解尺度の例(芝, 1977)

芝祐順. (1977). 語彙理解尺度作成の試み. 東京大学教育学部紀要, 17, 47-58. https://doi.org/10.15083/00030009

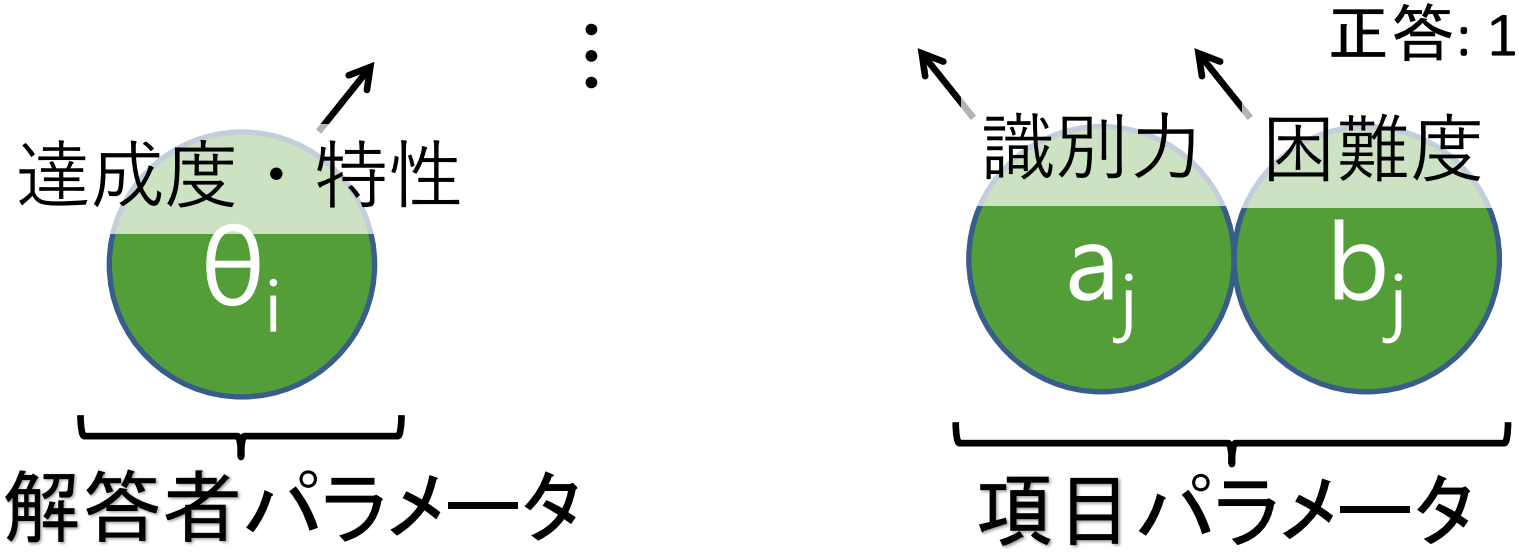
項目反応モデル

項目反応理論(item response theory, **IRT**)

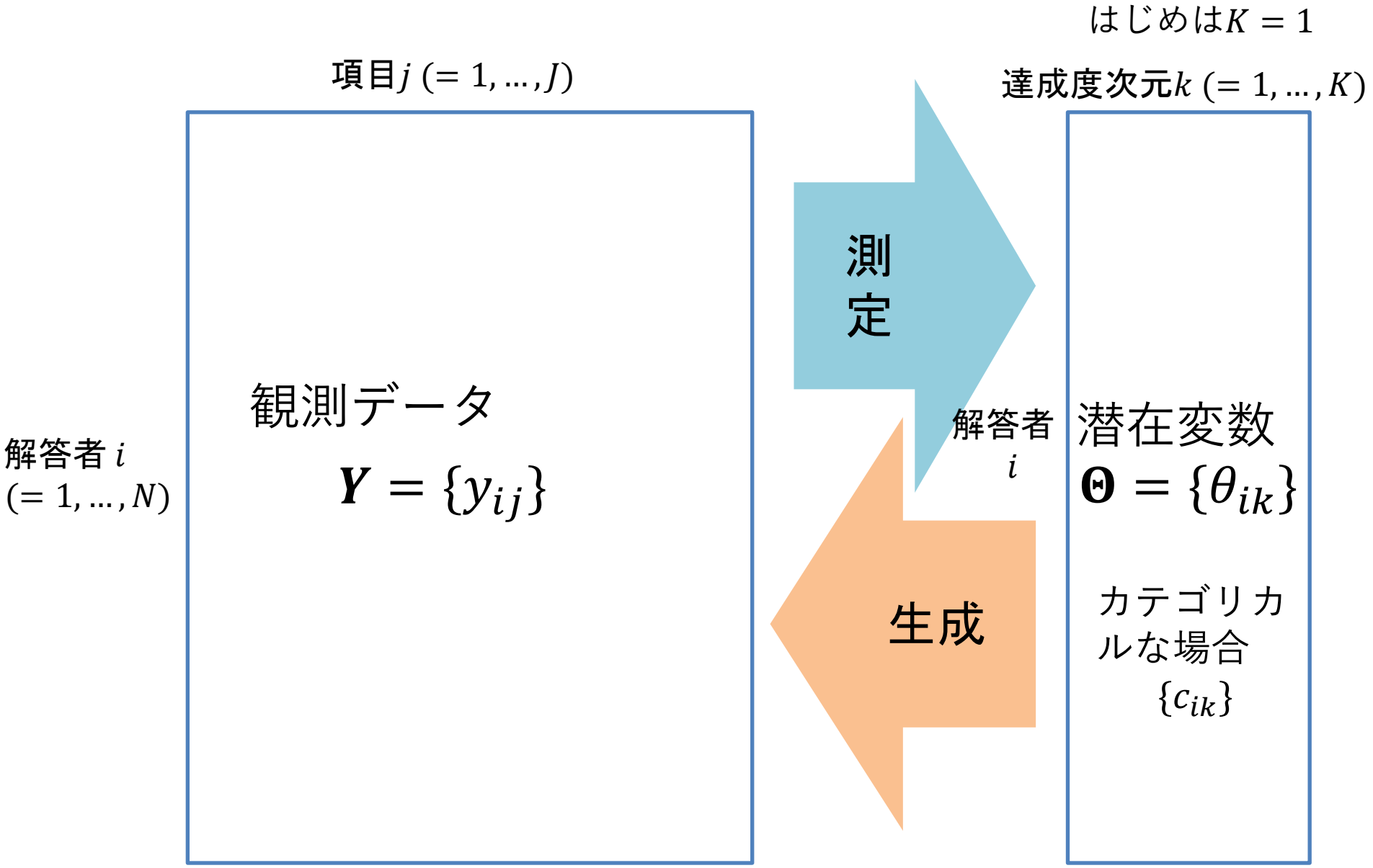
- テスト解答に基づき達成度を推定するために利用される
- 調査回答からの心理特性の推定にも応用される
- **解答者／項目パラメータの分離**が特徴

	項目1	項目2	項目3	項目4	項目5	項目6
Aさん	0	1	0	1	1	1
Bさん	1	1	0	0	0	0
Cさん	0	0	1	1	0	1

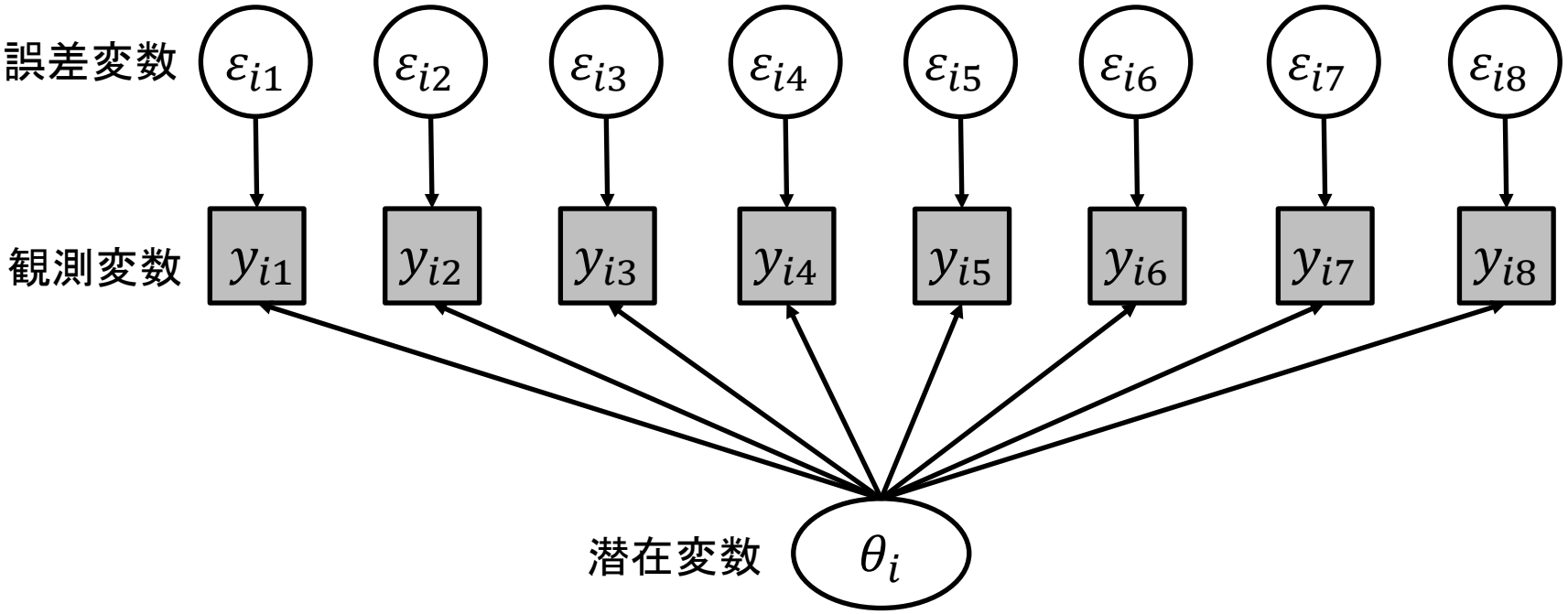
正答: 1，誤答: 0



想定するモデル構造


はじめは$K = 1$
項目$j\ (= 1, \ldots, J)$
達成度次元$k\ (= 1, \ldots, K)$
観測データ
$\boldsymbol{Y} = \{y_{ij}\}$
解答者 i
$(= 1, \ldots, N)$
測定
解答者
i
潜在変数
$\boldsymbol{\Theta} = \{\theta_{ik}\}$
生成
カテゴリカルな場合
$\{c_{ik}\}$

想定するモデル構造（1次元）



モデルの一般形

- 観測変数
  - y_{ij}: 項目解答
- 第一義的目的は観測値y_{ij}を生成する解答者パラメータθ_iの測定
- モデルの一般形

$$y_{ij} \sim \mathrm{ProbDist}(\mu_{ij})$$
$$\mu_{ij} = f(\theta_i, \kappa_j)$$

- パラメータ
  - θ_i: 解答者パラメータ（添え字i; 達成度・心理特性）
  - κ_j: 項目パラメータ（添え字j; 難易度など）

項目反応モデル

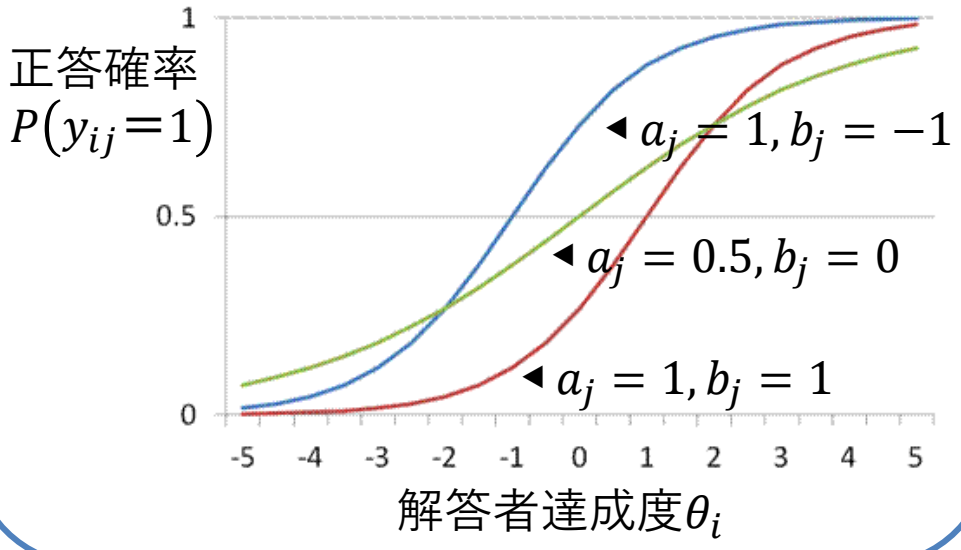
- 社会的役割の大きな達成度テストの多くは、**項目反応モデル**(IRT)を用いて運用されている

解答者iの項目jへの正答確率

$$P(y_{ij}=1) = \mathrm{logit}^{-1}(a_j(\theta_i - b_j))$$

- 解答者パラメータ：
  達成度θ_i
- 項目パラメータ：
  困難度b_j, 識別力a_j

項目jの特性曲線の例



項目反応理論モデル
（2パラメータロジスティックモデル, 2PL）

- 観測変数
  - $y_{ij} \in \{0,1\}$ （2値データ）
- モデル
  - $y_{ij} \sim \mathrm{Bernoulli}(\mu_{ij})$
  - $\mu_{ij} = \mathrm{logit}^{-1}(a_j(\theta_i - b_j)) = \frac{1}{1+\exp(-a_j(\theta_i - b_j))}$
  - $\theta_i \sim \mathrm{Normal}(0,1)$ （←入れないこともある）
- 解答者パラメータ
  - θ_i：**達成度**、潜在特性(achievement, trait)
- 項目パラメータ
  - a_j：**識別力**(discrimination)
  - b_j：**困難度**(difficulty)

ベルヌーイ分布

- $y \sim \mathrm{Bernoulli}(\theta)$: $y = 1$をとる確率がθである確率分布

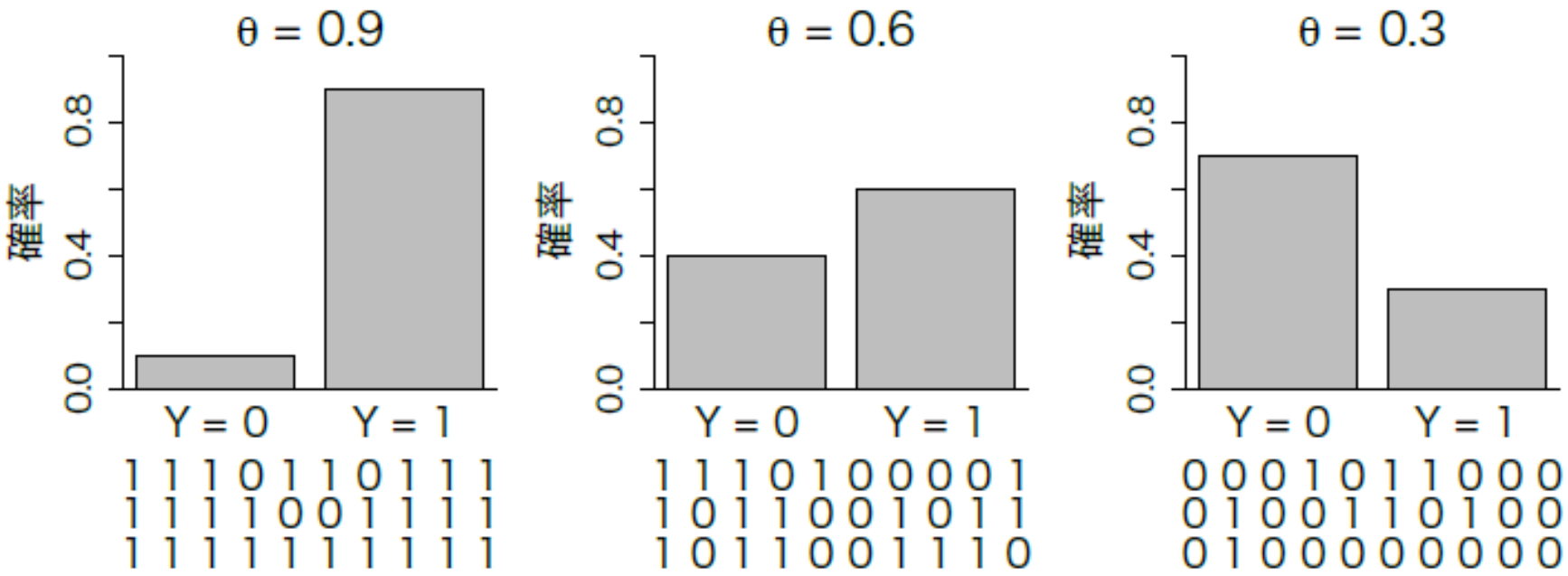


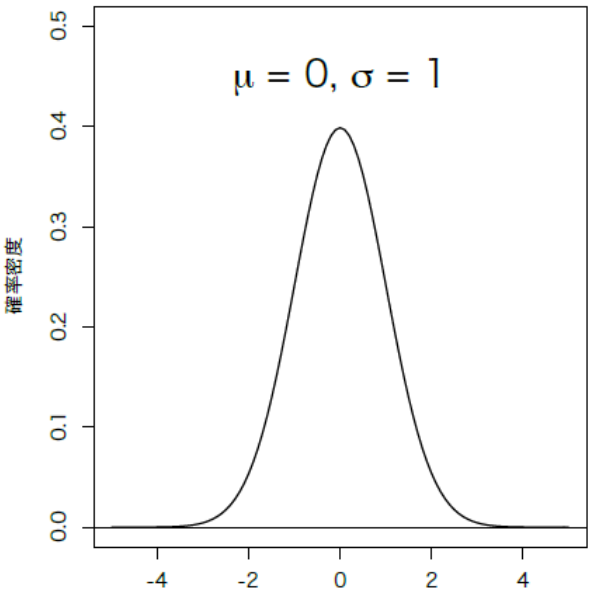
図 2.8　$\theta = 0.9$、0.6、0.3 の 3 つのベルヌーイ分布

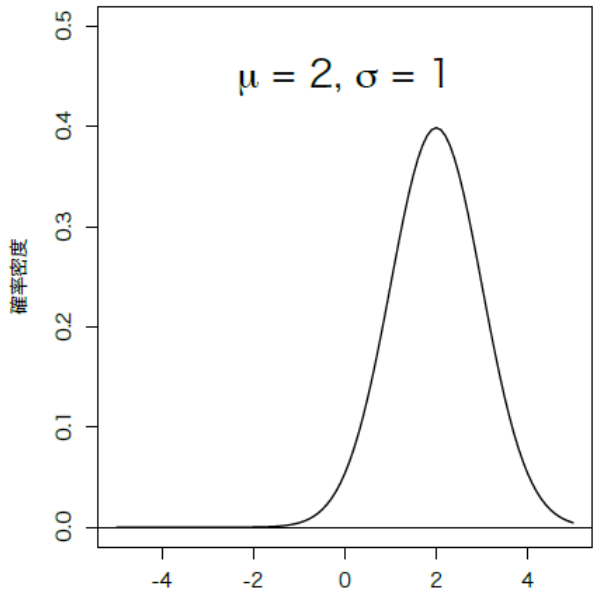
下の1,0は各ベルヌーイ分布から発生させた乱数の例

岡田謙介・矢内勇生(近刊) 社会科学のためのベイズ統計分析入門. 有斐閣

正規分布

- $y \sim \mathrm{Normal}(\mu, \sigma)$
- 左右対称、釣鐘型。最も代表的な連続型確率分布





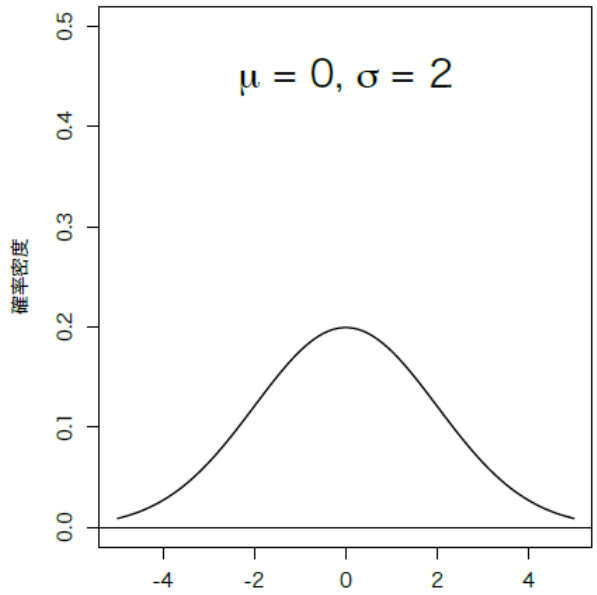


図 2.12　正規分布の例

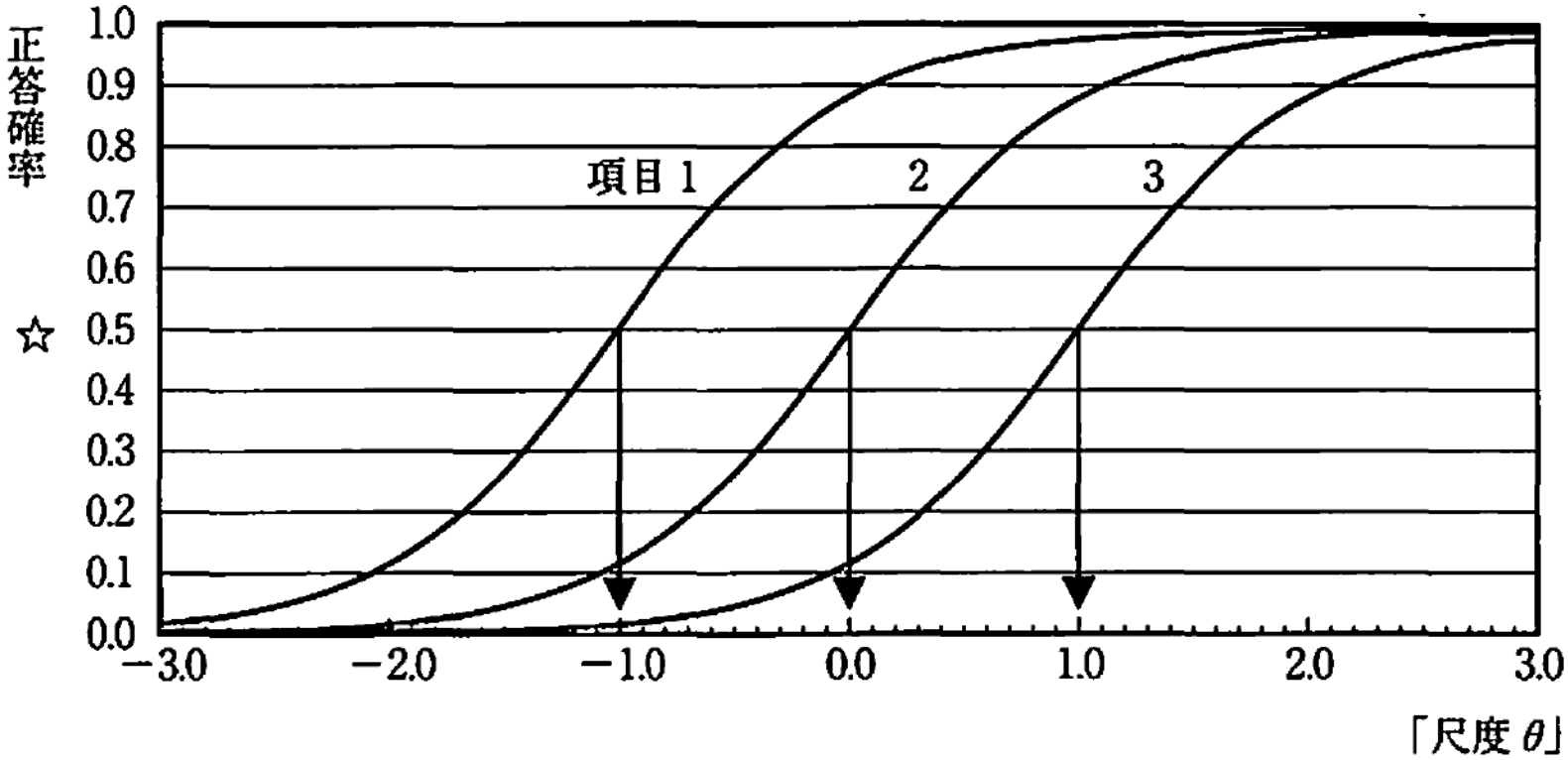
岡田謙介・矢内勇生(近刊) 社会科学のためのベイズ統計分析入門. 有斐閣

項目jの困難度b_j

- 解答者の達成度θ_iを固定したとき、困難度が高い項目は低い項目よりも正答確率が小さくなる

	項目1	項目2	項目3
識別力	1.2	1.2	1.2
困難度	−1.0	0.0	1.0

項目特性曲線

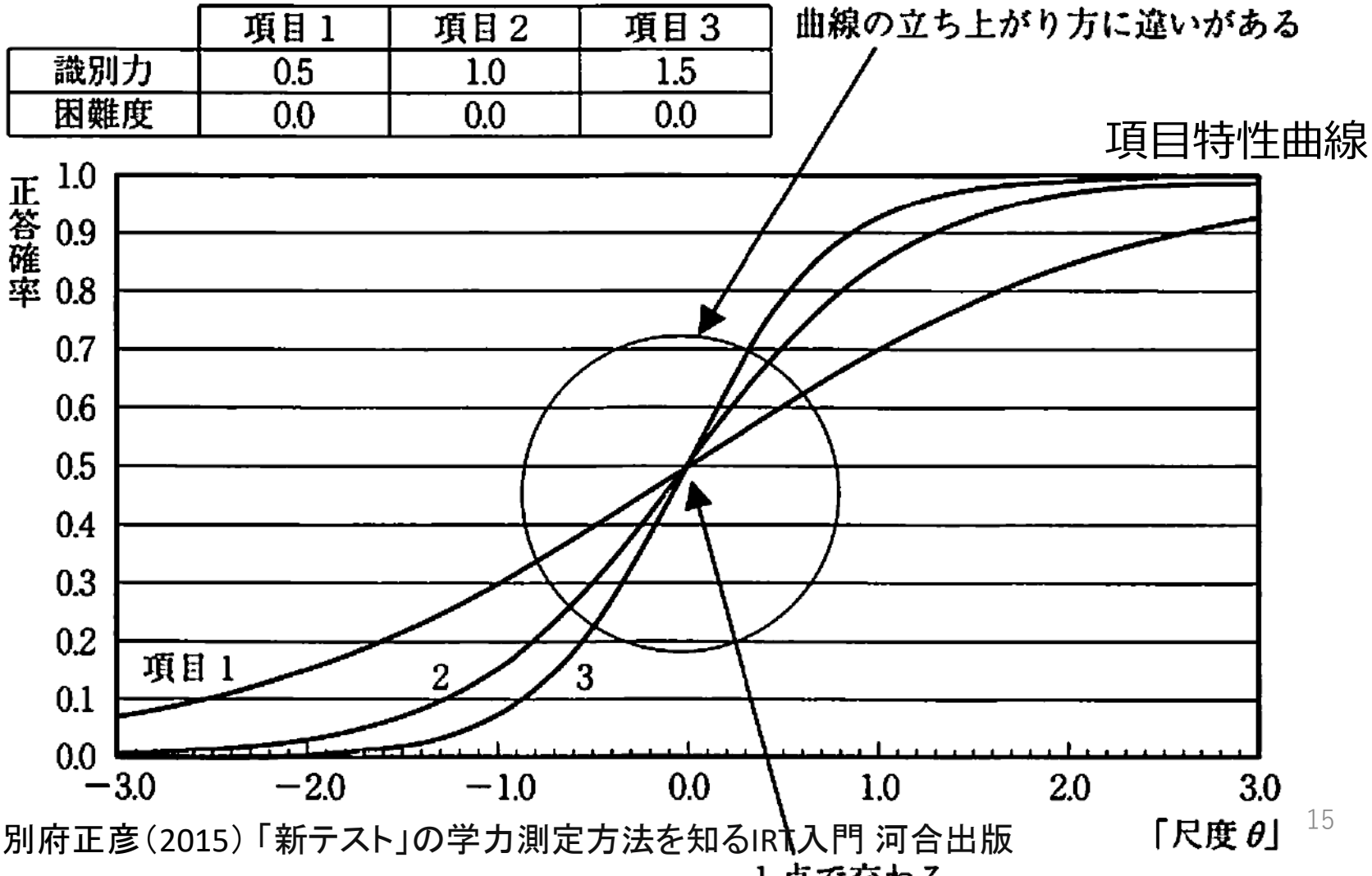


図：別府正彦（2015）「新テスト」の学力測定方法を知るIRT入門 河合出版

項目jの識別力a_j

- 識別力が高い項目は低い項目よりも、中央付近の達成度θ_iの変化によって大きく正答確率が変化する

	項目１	項目２	項目３
識別力	0.5	1.0	1.5
困難度	0.0	0.0	0.0



図：別府正彦（2015）「新テスト」の学力測定方法を知るIRT入門 河合出版

現在標準的な達成度推定の例

問　整式

$$P = (x-1)^2(y+5) + (2x-3)(y+4) - (x-1)$$

を考える。

(1)　整式 P を変形して

$$P = \left(x^2 - \boxed{\text{K}}\right)\left(y + \boxed{\text{L}}\right)$$

を得る。

(2)　$P=7$ となるような整数 x, y の組 (x, y) は

$$\left(\pm\boxed{\text{M}}, \boxed{\text{NOP}}\right), \quad \left(\pm\boxed{\text{Q}}, \boxed{\text{RS}}\right)$$

である。

(3)　a を有理数とする。$x = \sqrt{2} + 2\sqrt{3}$，$y = a + \sqrt{6}$ のとき，P の値が有理数となるような a の値は $\boxed{\text{TU}}$ である。

問題項目1

$a_1 = 0.30, b_1 = -0.40$

$y_{i1} = 1$（正答）

問題項目2

$a_2 = 0.60, b_2 = 0.10$

$y_{i2} = 1$（正答）

問題項目3

$a_3 = 0.50, b_3 = 0.30$

$y_{i3} = 0$（誤答）

達成度推定値

項目反応モデルによる推定

$\widehat{\theta_i} = 0.25\ [95\%CI = 0.17, 0.33]$

2段階推定

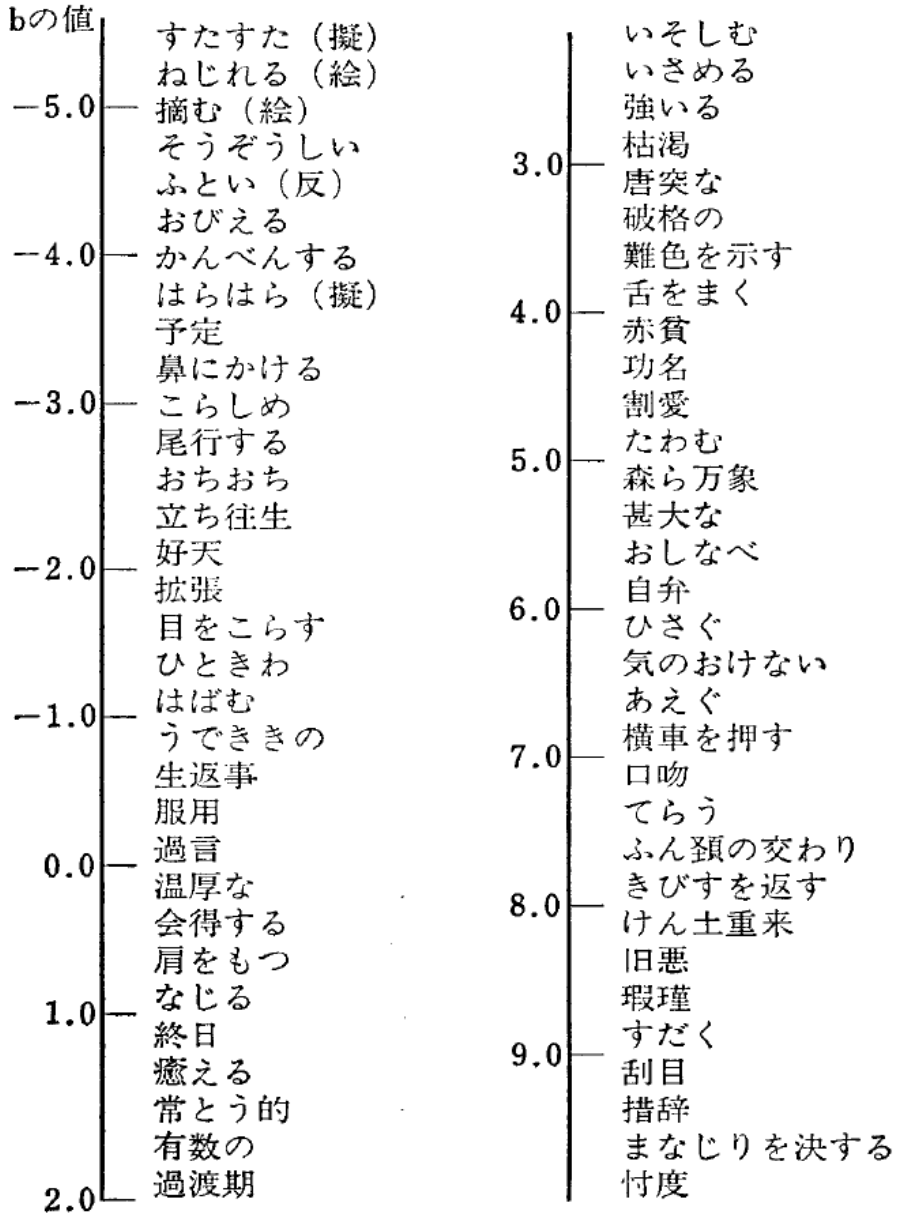
- 項目反応モデルにおける、解答者・項目パラメータの同時最尤推定量には一致性がない(Neyman-Scott問題)
  - 一致性： $N \to \infty$のとき、推定量が真値に確率収束すること
- そのため、実際のテスト運用では、次の2段階推定が用いられている

1. 予備試験データで
   項目パラメータ推定 $\hat{a}_j, \hat{b}_j$
2. 項目パラメータ値を固定して、本番試験データで
   解答者パラメータ推定 $\hat{\theta}_i \mid a_j, b_j$

- ここでの前提は問題項目の秘匿と繰り返し使用
  - 社会的役割の大きい達成度テストと違い、学術的な心理測定研究では項目を秘匿しないことも多い

語彙理解尺度の推定された項目パラメータ値

項目番号	項目	識別力	困難度
1	むずかしい	1.0	−4.5
2	あおむけになる	1.0	−3.6
3	さっそく	0.9	−2.7
4	軽蔑する	0.6	−1.6
5	やさきに	0.5	−0.6
6	年配の	0.4	0.7
7	なおざりにする	0.4	1.7
8	唐突	0.5	2.8
9	固執	0.5	3.9
10	森羅万象	0.3	4.7



芝祐順. (1977). 語彙理解尺度作成の試み. 東京大学教育学部紀要, 17, 47-58. https://doi.org/10.15083/00030009

項目パラメータの推定結果（UPI）

$N = 548$

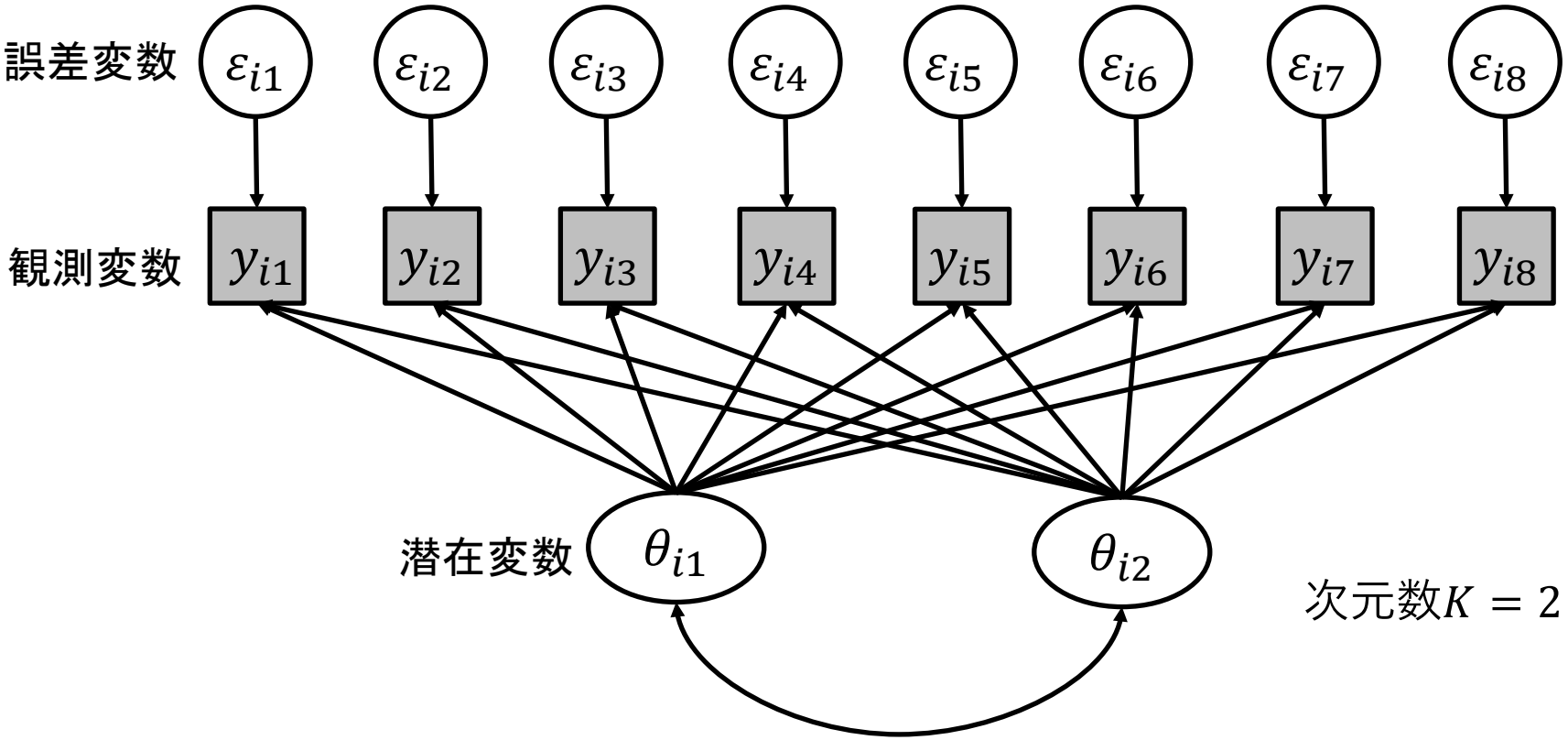
項目番号	識別力	困難度
1	1.01 (0.19)	1.39 (0.23)
2	0.94 (0.17)	1.37 (0.24)
3	0.77 (0.16)	1.27 (0.27)
4	1.09 (0.28)	2.61 (0.52)
6	1.46 (0.24)	1.34 (0.17)
7	0.80 (0.21)	2.97 (0.70)
8	1.02 (0.24)	2.93 (0.58)
9	1.61 (0.23)	1.22 (0.13)
10	1.35 (0.27)	2.11 (0.30)
11	1.54 (0.29)	2.07 (0.26)
12	1.25 (0.19)	0.99 (0.15)
13	2.26 (0.30)	0.58 (0.08)
14	1.47 (0.20)	0.54 (0.11)
15	1.43 (0.21)	0.43 (0.11)
16	0.76 (0.21)	3.08 (0.77)
17	0.83 (0.83)	1.53 (0.30)
18	0.52 (0.13)	0.97 (0.33)
19	1.61 (0.30)	1.89 (0.22)
21	1.29 (0.21)	1.12 (0.15)
22	1.51 (0.20)	0.27 (0.10)
23	1.24 (0.19)	0.96 (0.14)
24	0.93 (0.19)	2.06 (0.37)
25	2.16 (0.43)	2.31 (0.25)
26	1.78 (0.33)	2.23 (0.27)
27	0.97 (0.18)	1.27 (0.22)
28	0.92 (0.17)	1.25 (0.23)
29	0.92 (0.15)	0.40 (0.15)
30	0.88 (0.16)	0.79 (0.18)
31	0.59 (0.15)	2.03 (0.52)
32	1.20 (0.27)	2.27 (0.38)
33	0.87 (0.19)	1.69 (0.32)
34	0.80 (0.34)	4.74 (1.92)
36	1.70 (0.22)	−0.02 (0.09)
37	1.01 (0.24)	2.53 (0.50)
38	1.62 (0.21)	0.47 (0.10)
39	1.29 (0.18)	0.37 (0.11)
40	1.30 (0.26)	1.91 (0.27)
41	1.96 (0.38)	1.82 (0.19)
42	1.03 (0.17)	1.03 (0.18)
43	1.09 (0.20)	2.27 (0.37)
44	1.77 (0.24)	0.87 (0.11)
45	1.29 (0.21)	0.87 (0.14)
46	1.34 (0.20)	1.01 (0.14)
47	0.79 (0.19)	2.12 (0.45)
48	0.66 (0.15)	1.30 (0.32)
49	1.35 (0.64)	3.82 (1.32)
51	0.76 (0.15)	1.23 (0.26)
52	0.93 (0.16)	0.29 (0.14)
53	0.80 (0.21)	2.71 (0.62)
54	1.76 (0.25)	0.91 (0.11)
55	1.14 (0.29)	2.29 (0.42)
56	1.26 (0.33)	2.57 (0.49)
57	1.73 (0.23)	0.77 (0.10)
58	1.24 (0.17)	0.16 (0.11)
59	1.63 (0.36)	2.87 (0.45)
60	1.62 (0.24)	1.04 (0.12)
平均	1.23	1.57
標準偏差	0.40	0.96

注：（ ）内は標準誤差を示す。

酒井渉・野口裕之 (2015). 大学生を対象とした精神的健康度調査の共通尺度化による比較検討. 教育心理学研究, 63(2), 111-120. https://doi.org/10.5926/jjep.63.111

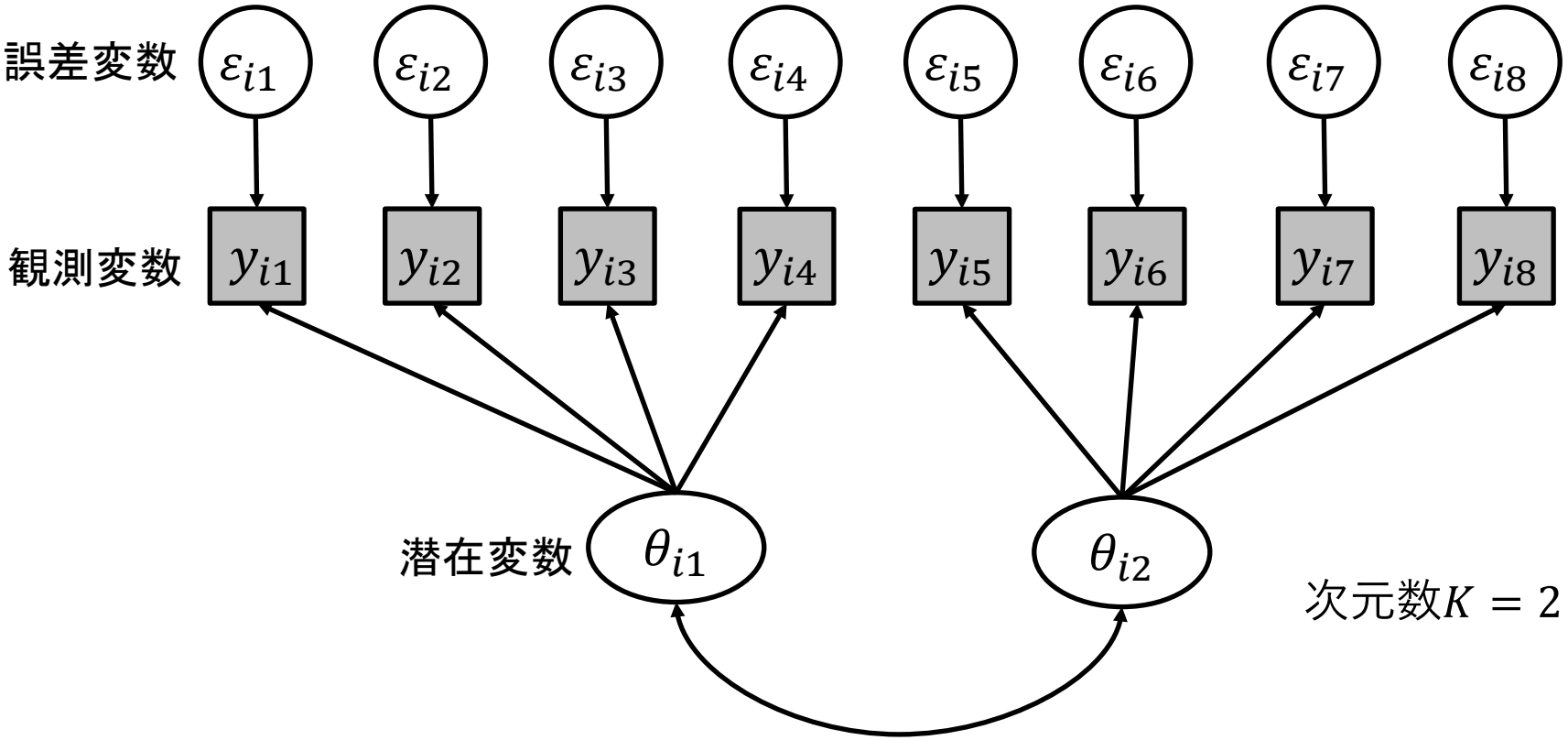
想定するモデル構造（多次元・探索的）

exploratory



想定するモデル構造（多次元・確認的）

confirmatory

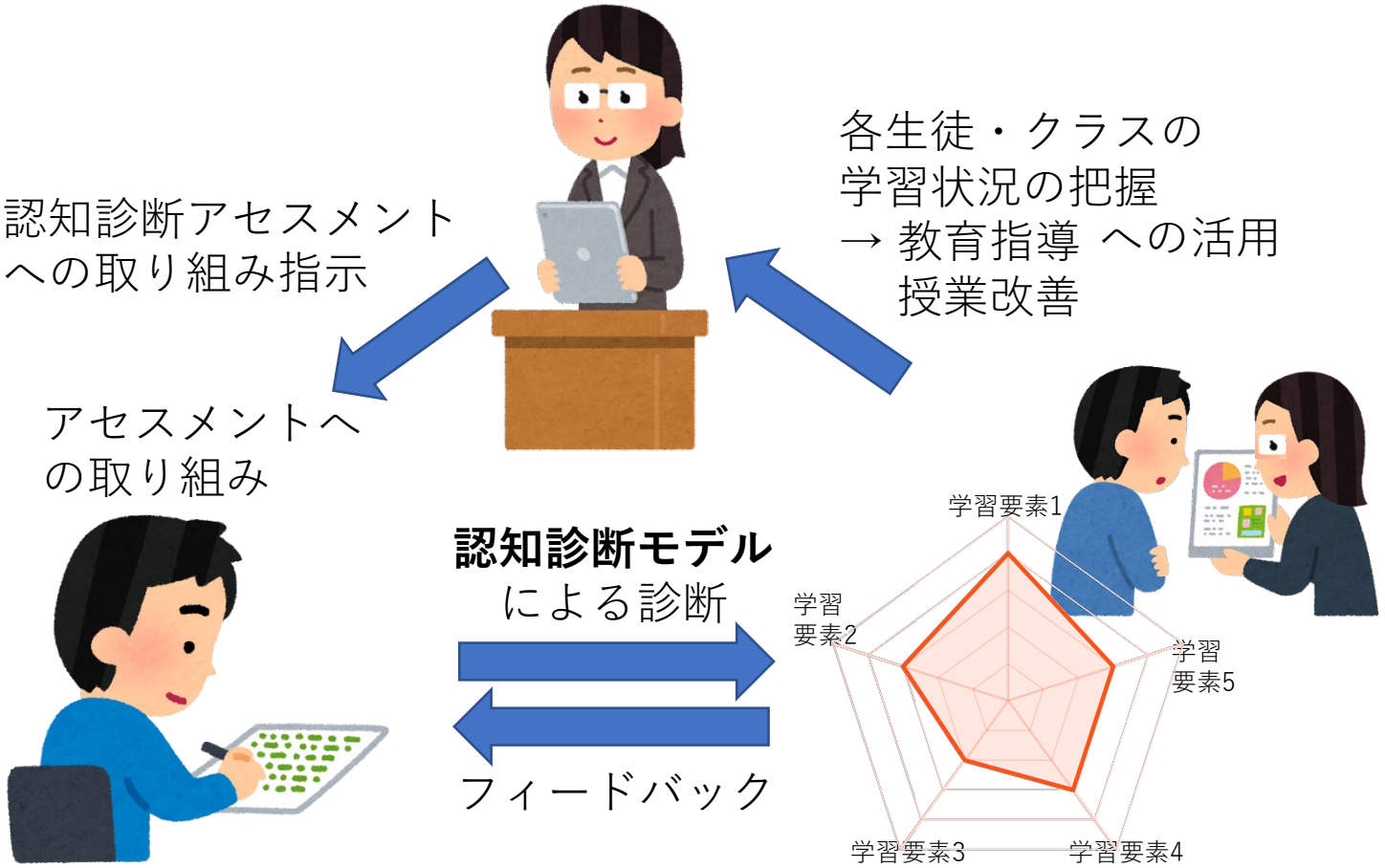


2次元以上の特性への拡張:多次元2PLモデル

- 観測変数
  - $y_{ij} \in \{0,1\}$ （2値データ）
- モデル
  - $y_{ij} \sim \mathrm{Bernoulli}(\mu_{ij})$
  - $\mu_{ij} = \mathrm{logit}^{-1}\left(\sum_{k=1}^{K} a_{jk}\theta_{ik} + d_j\right) = \frac{1}{1+\exp\left(-\left(\sum_{k=1}^{K} a_{jk}\theta_{ik}+d_j\right)\right)}$
  - $\boldsymbol{\theta}_i \sim \mathrm{MultiNormal}(\mathbf{0}, \boldsymbol{\Sigma})$
- 解答者パラメータ　　分散は1
  - $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iK})$：達成度、潜在特性
- 項目パラメータ
  - a_{jk}：負荷量、識別力
  - d_j：切片(intercept)

認知診断アセスメント

- 日々の学習達成状況をアセスメントを通じて教師が把握し、指導や授業改善に役立てる



CDMの研究例1(鈴木ら, 2015)

- 教研式標準学力検査NRTにおける，中学1年生数学の一部項目に基づく学習診断。アトリビュートは以下の4つ

計算力	負の数や文字式のある計算の遂行に関する能力．方程式を解く能力．なお，小学校で学習する計算スキルは含まない． 課題例：$-5^3+(-2)^2$ $7\chi-4-2\chi+3$
概念理解	数学用語の理解．公式の理解．なお，「正方形は，4つの辺の長さが等しい」など，小学校で学習する知識は含まない． 課題例：以下で使われている計算法則は何ですか． $15\times\left(\frac{5}{3}+\frac{3}{5}\right)-18$ $=15\times\frac{5}{3}+15\times\frac{3}{5}-18$
図形操作力	図形を心的に操作して問題解決をすることができる能力． 課題例：ある平面図形を，直線を軸にして一回転させてできる図形を選びなさい．
論理力	すじ道を立てて，解法を探索することができる力．なお，「ある数χに5を足した数は，χの3倍よりも小さい」を式で表す力など，文章と式を一対一対応させる力は含まない． 課題例：χの値が増加すると，yの値が減少する式を選びなさい．

鈴木雅之・豊田哲也・山口一大・孫媛. (2015). 認知診断モデルによる学習診断の有用性の検討: 教研式標準学力検査NRT「中学 1 年数学」 への適用. 日本テスト学会誌, 11(1), 81-97.

表6 Q行列

項目	アトリビュート				項目	アトリビュート			
	計算力	概念理解	図形操作力	論理力		計算力	概念理解	図形操作力	論理力
1	1	0	0	0	28	0	1	1	0
2	1	0	0	0	29	0	0	1	0
3	1	0	0	0	30	0	1	1	0
4	1	0	0	0	31	0	1	1	0
5	1	0	0	0	32	0	1	1	0
6	0	1	0	0	33	0	1	1	0
7	0	1	0	0	34	1	1	0	0
8	1	0	0	0	35	1	1	0	1
9	1	0	0	0	36	1	1	0	1
10	1	0	0	0	37	1	1	0	1
11	1	0	0	1	38	1	1	0	1
12	1	0	0	1	39	1	1	0	0
13	1	0	0	0	40	1	1	0	0
14	1	0	0	0	41	1	1	0	1
15	1	0	0	0	42	1	1	0	1
16	1	0	0	0	43	1	1	0	0
17	1	0	0	0	44	1	1	0	0
18	1	0	0	0	45	0	1	0	0
19	1	0	0	0	46	0	0	0	1
20	1	0	0	1	47	1	0	0	0
21	1	0	0	1	48	1	1	0	1
22	1	0	0	1	49	0	0	0	1
23	0	1	0	0	アトリビュートが関わる項目数				
24	0	1	0	0	数と式	20	2	0	5
25	0	1	0	1	図形	2	11	7	3
26	0	0	0	1	関数	11	11	0	8
27	0	1	1	0	計	33	24	7	16

注)項目 1～22 は「数と式」, 項目 23～35 は「図形」, 項目 36～49 は「関数」の領域からの出題

専門家の合議でQ行列を作成し，その後異なる独立な作成と評定者間一致率も確認(85-95%)

鈴木雅之・豊田哲也・山口一大・孫媛. (2015). 認知診断モデルによる学習診断の有用性の検討: 教研式標準学力検査NRT「中学 1 年数学」 への適用. 日本テスト学会誌, 11(1), 81-97.

$N = 948$のデータにおける診断結果

表 5.　数研式 NRT でのアトリビュート習得パタンおよびテスト得点

パタン	計算力	概念理解	図形操作力	論理力	人数	テスト得点
1	1	1	1	1	165	40.3
2	1	1	1	0	46	32.2
3	1	1	0	1	24	36.5
4	1	1	0	0	54	30.8
5	1	0	1	1	34	31.1
6	1	0	1	0	122	25.8
7	1	0	0	0	52	21.7
8	0	1	1	1	5	29.0
9	0	1	1	0	2	24.0
10	0	1	0	1	2	28.0
11	0	1	0	0	30	21.4
12	0	0	1	1	24	22.0
13	0	0	1	0	58	18.5
14	0	0	0	1	2	21.5
15	0	0	0	0	328	12.7

鈴木雅之・豊田哲也・山口一大・孫媛. (2015). 認知診断モデルによる学習診断の有用性の検討: 教研式標準学力検査NRT「中学 1 年数学」 への適用. 日本テスト学会誌, 11(1), 81-97.

CDMの研究例2(福島ら, 2021)

- 英文法の選択式空所補充形式テストに設定するアトリビュート($K = 4$)

表1　アトリビュートを構成する各種要素

時制	1-1	時制の選択
	1-2	適切な時制の形への変形
	1-3	動詞の基本的な形への変形
語彙	2-1	自動詞と他動詞の区別
	2-2	単語の意味
	2-3	動詞と前置詞の組み合わせ
	2-4	可算・不可算名詞の区別
分詞	3-1	現在分詞・過去分詞の使い分け
	3-2	受動態の知識・使用
関係詞	4-1	関係詞の選択
	4-2	不完全文とすること
	4-3	省略・語順等のルールの把握
	4-4	関係詞の必要性の判断

福島健太郎・内田奈緒・岡田謙介 (2021). Q 行列を付与した多枝選択形式テストの開発: 認知診断モデルのための英語の空所補充問題の作成. 日本テスト学会誌, 17(1), 45-59. https://bit.ly/3sXN4h5

設定したQ行列

- 選択枝の各々に対してQ行列を設定
  （これは拡張された特別なモデル用; e.g., Ozaki, 2015）

表２　項目の内容およびQ行列

item	question	option		時制	語彙	分詞	関係詞
1	How long have you (　) in bed?	1	be lying	0	1		
		2	been lying	1	1		
		3	lied	1	0		
		4	be laying	0	0		
2	The flag is (　) every morning by the person in charge.	1	rising		1	0	
		2	risen		0	1	
		3	raising		0	0	
		4	raised		1	1	
3	Show me the pictures (　) your brother took.	1	whose				0
		2	which				1
		3	and				0
		4	what				0
4	I (　) a shower when the telephone rang.	1	have been taking	0			
		2	have taken	0			
		3	was taken	0			
		4	was taking	1			
5	Juila lost the ring (　) her the day before.	1	that I gave	0			1
		2	and I had given	1			0
		3	and I gave	0			0
		4	which I had given	1			1
6	I've heard so (　) news about the scandal that I'm sick of it.	1	few		0		
		2	little		0		
		3	many		0		
		4	much		1		
7	Violet is not a coward (　) she was ten years ago.	1	that				1
		2	who				0
		3	when				0
		4	whom				0

Ozaki, K. (2015). DINA models for multiple-choice items with few parameters: Considering incorrect answers. *Applied Psychological Measurement*, *39*(6), 431-447. https://doi.org/10.1177/0146621615574693
福島健太郎・内田奈緒・岡田謙介 (2021). Q 行列を付与した多枝選択形式テストの開発: 認知診断モデルのための英語の空所補充問題の作成. 日本テスト学会誌, 17(1), 45-59. https://bit.ly/3sXN4h5

DINAモデル

deterministic inputs, noisy ‘‘and’’ gate model (Junker & Sijtsma, 2001)

- 観測変数
  - $y_{ij} \in \{0,1\}$ （2値データ）
- モデル
  - $\mu_{ij} = \left(1 - s_j\right)^{\eta_{ij}} g_j^{1-\eta_{ij}}$
  - $\eta_{ij} = \prod_{k=1}^{K} \alpha_{c_i k}^{q_{jk}}$ ：理想反応
  - $c_i \sim \text{Categorical}(\boldsymbol{\pi}), \boldsymbol{\pi} \sim \text{Dirichlet}(\mathbf{1})$
- 解答者パラメータ
  - c_i：アトリビュート習得パターン(潜在クラス)のうち1つへと解答者iを分類する際の、潜在クラス番号
- 項目パラメータ
  - s_j：slipパラメータ、 g_j：guessingパラメータ

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258-272. https://doi.org/10.1177/01466210122032064

2つの項目パラメータ

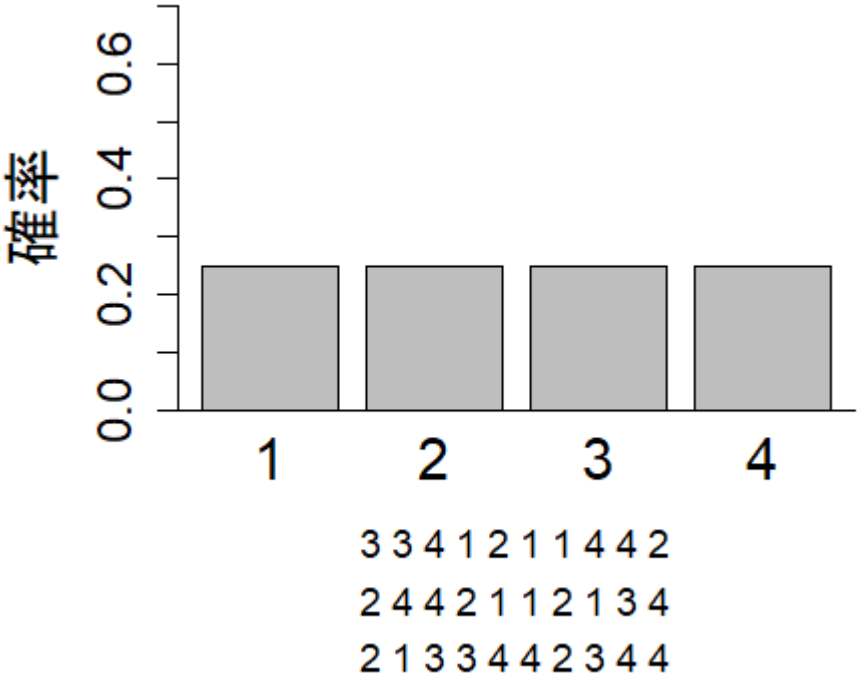
- **slipパラメタ**：$s_j = \Pr(y_{ij} = 0 | \eta_{ij} = 1)$
  - 項目に正答するために求められる学習要素を習得しているにもかかわらず，項目に誤答する確率
- **guessingパラメタ**：$g_j = \Pr(y_{ij} = 1 | \eta_{ij} = 0)$
  - 項目に正答するために求められる学習要素を習得していないにもかかわらず，項目に正答する確率
- 決定的である理想反応η_{ij}に付加される，モデルの偶然的（確率的）要素
- どちらもあまり大きな値の項目は診断に用いる上で適切でない可能性がある
- 単調性制約：$0 \leq g_j \leq 1 - s_j \leq 1$

  $\Leftrightarrow\ 0 \leq \Pr(y_{ij} = 1 | \eta_{ij} = 0) \leq \Pr(y_{ij} = 1 | \eta_{ij} = 1) \leq 1$

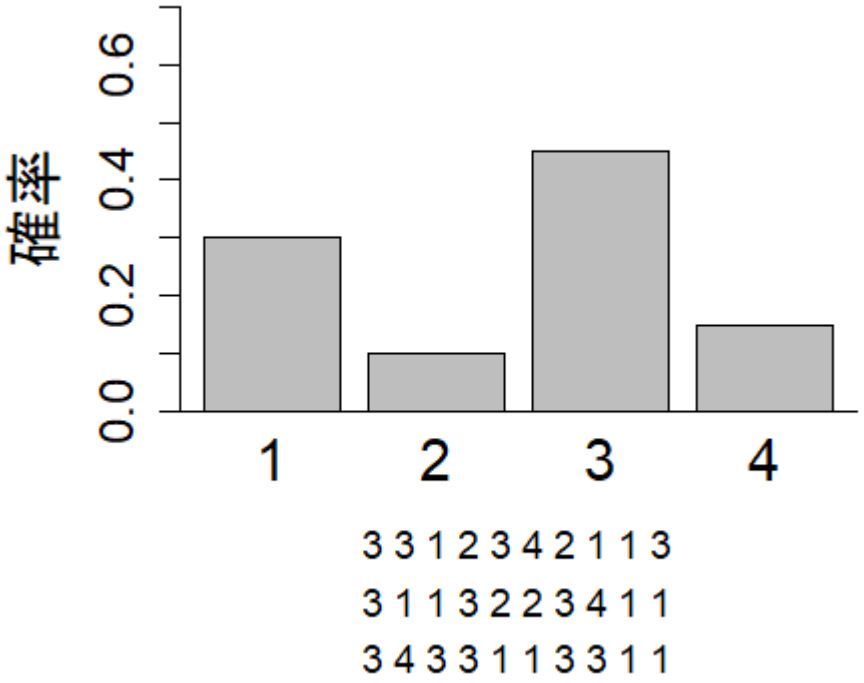
カテゴリカル分布

- $y \sim \mathrm{Categorical}(\boldsymbol{\theta})$: 質的変数$y$が$c$番目のカテゴリ（$c = 1, \ldots, C$）をとる確率が$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_C)$である確率分布

$\boldsymbol{\theta} = (0.25, 0.25, 0.25, 0.25)$



$\boldsymbol{\theta} = (0.3, 0.1, 0.45, 0.15)$



岡田謙介・矢内勇生(近刊) 社会科学のためのベイズ統計分析入門. 有斐閣

ディリクレ分布

- $\boldsymbol{\theta} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha})$: $\theta_c, \geq 0$ $\sum_c \theta_c = 1$を満たす確率変数$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_C)$の確率密度関数。ベータ分布を多変量に拡張したもの。パラメタ$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_C)$が$\boldsymbol{\alpha} = \mathbf{1} = (1, \ldots, 1)$のとき，台上の一様分布となる。

$C = 3$の場合

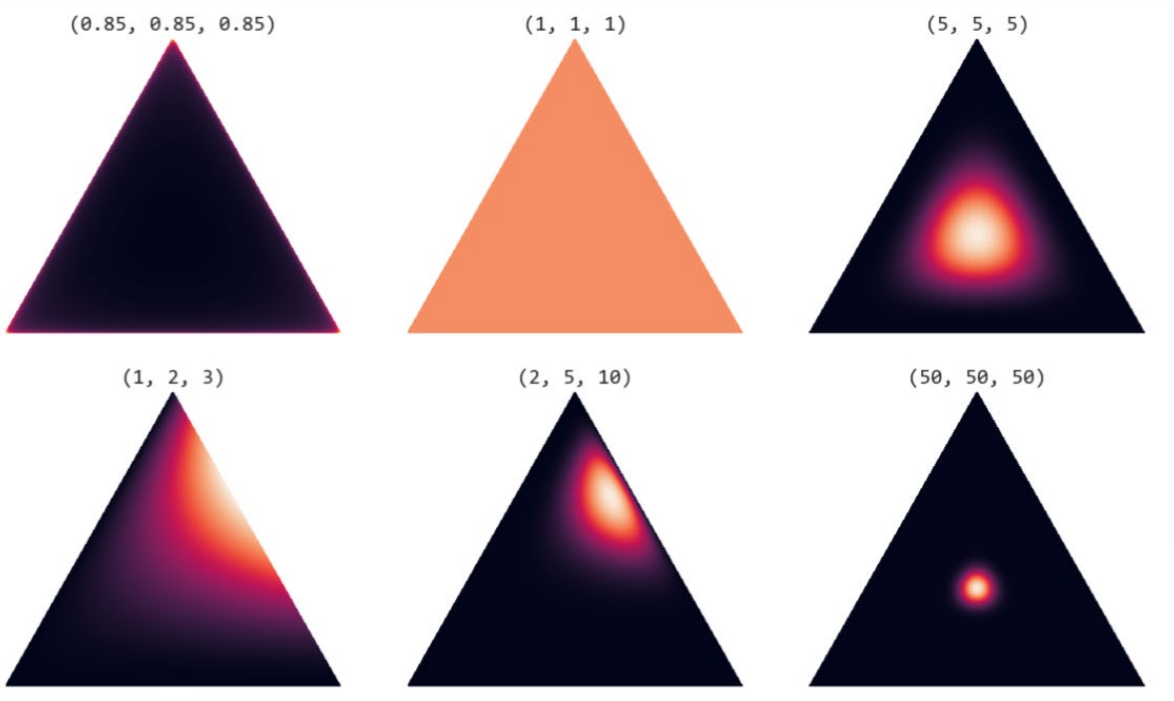


図: https://towardsdatascience.com/dirichlet-distribution-a82ab942a879

Q行列$\{q_{jk}\}$

- 各問題項目に正答するためにはどのアトリビュートが求められるのかを表す, {0, 1}の2値変数を要素に持つ行列

	アトリビュート $k = 1, \ldots, K(= 3)$		
$j = 1, \ldots, J(= 3)$ 項目	加減	乗除	通分
$1/3 + 4/3$	1	0	0
$1/2 \times 1/3$	0	1	0
$1/2 + 1/3$	1	0	1

- 専門家の知見や文献調査に基づいて事前に設定される
- 統計的に推定するアプローチも近年大きく研究が進んでいる

山口一大・岡田謙介 (2017). 近年の認知診断モデルの展開. 行動計量学, 44(2), 181-198. https://doi.org/10.2333/jbhmk.44.181

アトリビュート習得パターン$\{\alpha_{ck}\}$

- アトリビュート数がKのとき、可能なアトリビュート習得パターンは2^K通りある
- このようにアトリビュート習得パターン行列は決定的に決まる

$c = 1, \ldots, 2^K (= 8)$ 習得パタン (潜在クラス)	アトリビュート $k = 1, \ldots, K (= 3)$		
	加減	乗除	通分
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	1	0	1
7	0	1	1
8	1	1	1

山口一大・岡田謙介 (2017). 近年の認知診断モデルの展開. 行動計量学, 44(2), 181-198. https://doi.org/10.2333/jbhmk.44.181

クラスc_iに属する人iの項目jへの理想反応$\{\eta_{ij}\}$

$\{q_{jk}\}$

項目	アトリビュート		
	加減	乗除	通分
$1/3+4/3$	1	0	0
$1/2\times1/3$	0	1	0
$1/2+1/3$	1	0	1

$\{\alpha_{ck}\}$

習得パタン（潜在クラス）	アトリビュート		
	加減	乗除	通分
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	1	0	1
7	0	1	1
8	1	1	1

$$\eta_{ij}=\eta_{c_i j}=\prod_{k=1}^{K}\alpha_{c_i k}^{q_{jk}}$$

習得パタン（潜在クラス）	項目		
	1/3 +4/3	1/2 ×1/3	1/2 +1/3
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	0	1	1
7	1	0	1
8	1	1	1

DINAモデル 再掲

deterministic inputs, noisy ‘‘and’’ gate model (Junker & Sijtsma, 2001)

- 観測変数
  - $y_{ij} \in \{0,1\}$ （2値データ）
- モデル
  - $\mu_{ij} = \left(1 - s_j\right)^{\eta_{ij}} g_j^{1-\eta_{ij}}$
  - $\eta_{ij} = \prod_{k=1}^{K} \alpha_{c_i k}^{q_{jk}}$ ：理想反応
  - $c_i \sim \text{Categorical}(\boldsymbol{\pi}), \boldsymbol{\pi} \sim \text{Dirichlet}(\mathbf{1})$
- 解答者パラメータ
  - c_i：アトリビュート習得パターン(潜在クラス)のうち1つへと解答者iを分類する際の、潜在クラス番号
- 項目パラメータ
  - s_j：slipパラメータ、 g_j：guessingパラメータ

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258-272. https://doi.org/10.1177/01466210122032064

Log-linear CDM (LCDM; Henson et al., 2009)

- $\mu_{ij} = \text{logit}^{-1}(\lambda_{j,0} + \sum_{k=1}^{K} \lambda_{j,1,(k)} \alpha_{ck} q_{jk} + \sum_{k=1} \sum_{k'>1}^{K} \lambda_{j,2,(k,K')} \alpha_{ck} \alpha_{ck'} q_{jk} q_{jk'} + \cdots)$
- DINAモデル等は制約付きのLCDMとして表現できる

Latent class	Attribute profile	Probability of a correct response
1	[0,0,*,*]	$P(X_{ic}=1 \mid \alpha_{c1}, \alpha_{c2}) = \frac{\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)}(0)(0))}{1+\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)}(0)(0))} = \frac{\exp(\lambda_{1,0})}{1+\exp(\lambda_{1,0})} = g_1$
2	[0,1,*,*]	$P(X_{ic}=1 \mid \alpha_{c1}, \alpha_{c2}) = \frac{\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)}(0)(1))}{1+\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)}(0)(1))} = \frac{\exp(\lambda_{1,0})}{1+\exp(\lambda_{1,0})} = g_1$
3	[1,0,*,*]	$P(X_{ic}=1 \mid \alpha_{c1}, \alpha_{c2}) = \frac{\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)}(1)(0))}{1+\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)}(1)(0))} = \frac{\exp(\lambda_{1,0})}{1+\exp(\lambda_{1,0})} = g_1$
4	[1,1,*,*]	$P(X_{ic}=1 \mid \alpha_{c1}, \alpha_{c2}) = \frac{\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)}(1)(1))}{1+\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)}(1)(1))} = \frac{\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)})}{1+\exp(\lambda_{1,0} + \lambda_{1,2,(1,2)})} = (1-s_1)$

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. Psychometrika, 74, 191–210. https://doi.org/10.1007/s11336-008-9089-5

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press. https://www.guilford.com/books/Diagnostic-Measurement/Rupp-Templin-Henson/9781606235270

さまざまな認知診断モデル（山口・岡田,2017）

モデル	反応データ	アトリビュートのカテゴリ水準	アトリビュート以外の潜在変数	補償・非補償	応用研究例	1項目あたりの項目パラメータ数
DINA	2値	2値	無	非補償	Lee, Park, & Taylan (2011)	2
DINO	2値	2値	無	補償	Templin & Henson (2006)	2
NIDA	2値	2値	無	非補償	なし	$2K_j^*$
NIDO	2値	2値	無	補償	なし	$2K_j^*$
LLM	2値	2値	無	補償	Chen & de la Torre (2014)	$1+K_j^*$
A-CDM	2値	2値	無	補償	Chen & de la Torre (2014)	$1+K_j^*$
R-RUM	2値	2値	無	非補償	Jang (2009)	$1+K_j^*$
C-RUM	2値	2値	無	補償	山口 (2016)	$1+K_j^*$
G-DINA	2値	2値	無	飽和	鈴木他 (2015)	$2^{K_j^*}$
LCDM	2値	2値	無	飽和	Templin & Hoffman (2013)	$2^{K_j^*}$
GDM	多値	多値	無	飽和	von Davier (2008)	$1+K_j^*$
MC-DINA	多値	2値	無	非補償	なし	$(1+H_j^*)(H_j-1)$
HO-DINA	2値	2値	有	非補償	なし	2
MS-DINA	2値	2値	無	非補償	なし	$2M$
pG-DINA	2値	多値	無	飽和	Chen & de la Torre (2013)	$2^{K_j^*}$
H-GDM	多値	多値	有	飽和	von Davier (2007a)	$SG(K_j^*+1)$
H-DCM	2値	2値	無	飽和	Templin & Bradshaw (2014)	$2^{K_j^*}$
Multi-group CDM	2値	2値	無	非補償	Xu & von Davier (2008a)	$G(K_j^*+1)$
Random effect DINA	2値	2値	有	非補償	Huang & Wang (2014)	2
LTA-DINA	2値	2値	有	非補償	Li, Cohen, Bottge, & Templin (201	2
HO-R-DINA	2値	2値	有	非補償	Ayers, Rabe-Hesketh, & Nugent (20	2

項目反応モデルと認知診断モデル

	項目反応モデル IRT	認知診断モデル CDM
目的変数	2値観測変数(正答・誤答)	2値観測変数(正答・誤答)
説明変数	解答者の潜在変数	解答者の潜在変数
リンク関数	ロジット関数	ロジット関数
潜在変数	連続	離散
-の粒度	大きめ	小さめ
-の次元数	基本は1次元(多次元にも拡張可)	多次元(アトリビュート数)
目的	総括的評価	形成的評価・認知診断

- 両者は密接に関係する2大モデル群
- 中間に位置づけることが可能なモデルも複数ある

モデル適合

- 認知診断モデルよりも低次元項目反応モデルの当てはまりがよい (von Davier & Haberman, 2014; Bolt, 2019)
- 1次元の1～3PL IRTモデルよりもDINAモデルのほうが当てはまりがよい(Lee et al., 2011)。一般化CDMはさらによい(Yamaguchi & Okada, 2018)
- 中間に位置するモデルや一般化されたモデルを始め、実データに対する予測力やモデル適合に関する知見はさらに蓄積される必要がある(モデルの事前登録研究など)

von Davier, M., & Haberman, S. J. (2014). Hierarchical diagnostic classification models morphing into unidimensional 'diagnostic' classification models: a commentary. *Psychometrika*, *79*(2), 340-346. https://doi.org/10.1007/s11336-013-9363-z

Bolt, D. M. (2019). Bifactor MIRT as an Appealing and Related Alternative to CDMs in the Presence of Skill Attribute Continuity. In *Handbook of Diagnostic Classification Models* (pp. 395-417). Springer. https://doi.org/10.1007/978-3-030-05584-4_19

Lee, Y. S., Park, Y. S., & Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the US national sample using the TIMSS 2007. *International Journal of Testing*, *11*(2), 144-177. https://doi.org/10.1080/15305058.2010.534571

Yamaguchi, K., & Okada, K. (2018). Comparison among cognitive diagnostic models for the TIMSS 2007 fourth grade mathematics assessment. *PLOS ONE*, *13*(2), e0188691. https://doi.org/10.1371/journal.pone.0188691

両モデル共に研究されている拡張

- コンピュータ適応型テスト化
  - IRT (e.g., van der Linden, 2018)
  - CDM (e.g., Cheng, 2009)
- 変分ベイズ推定（ビッグデータ・スケーラビリティ）
  - IRT (e.g., Rijmen et al., 2016)
  - CDM (e.g., Yamaguchi & Okada, 2021)
- 説明的モデル（説明要因の検討）
  - IRT (e.g., De Boeck & Wilson, 2016)
  - CDM (e.g., Park et al., 2018)
- 一方のモデルで研究開発する価値のあることは、もう一方のモデルでもそうであることが多いと思われる

(前ページの文献)

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. Psychometrika, 74(4), 619-632. https://doi.org/10.1007/s11336-009-9123-2

De Boeck, P., & Wilson, M. R. (2016). Explanatory response models. In: W. J. van der Linden (ed) *Handbook of Item Response Theory, Volume One: Models.* pp. 593-608. Chapman and Hall/CRC. https://www.routledge.com/Handbook-of-Item-Response-Theory-Volume-1-Models/Linden/p/book/9780367220013

Rijmen, F., Jeon, M., & Rabe-Hesketh, S. (2016). Variational approximation methods. In: W. J. van der Linden (ed) *Handbook of item response theory, Volume 2: Statistical tools*. pp. 259-270. Chapman and Hall/CRC. https://www.routledge.com/Handbook-of-Item-Response-Theory-Volume-2-Statistical-Tools/Linden/p/book/9780367221041

Park, Y. S., Xing, K., & Lee, Y. S. (2018). Explanatory cognitive diagnostic models: Incorporating latent and observed predictors. *Applied Psychological Measurement, 42(5), 376-392*. https://doi.org/10.1177/0146621617738012

van der Linden, W. J. (Ed.). (2016). *Handbook of item response theory: Volume 1: Models.* Chapman and Hall/CRC. https://www.routledge.com/Handbook-of-Item-Response-Theory-Volume-1-Models/Linden/p/book/9780367220013

van der Linden, W. J. (2018). Adaptive testing. In: W. J. van der Linden (ed) *Handbook of item response theory: Volume 3: Applications.* pp. 197-227. Chapman and Hall/CRC. https://www.routledge.com/Handbook-of-Item-Response-Theory-Volume-3-Applications/Linden/p/book/9780367221188

IRT ➡ CDM 研究の可能性

- 等化・リンキング ：観測得点リンキング(Liu, 2020)等が提案されているものの、潜在変数が離散であるCDMでは単純ではない
- 解答時間：解答時間のモデルは提案されている(Jiao et al., 2019)が、Diffusion-IRT (Tuerlinckx & De Boeck, 2005)のような生成モデル化はされていない
- 多次元モデルでのパラドキシカル状況の生起条件： 単調性の導入にはパラメータ制約が必要であるため、生起条件の記述は単純でない(van Rijn & Rijmen, 2015)

Jiao, H., Liao, D., & Zhan, P. (2019). Utilizing process data for cognitive diagnosis. In M. von Davier, Y.-S. Lee (eds.) *Handbook of Diagnostic Classification Models,* pp. 421-436. Springer. https://doi.org/10.1007/978-3-030-05584-4_20

Liu, R. (2020). Addressing score comparability in diagnostic classification models: an observed-score equating and linking approach. *Behaviormetrika*, 47, 55–80. https://doi.org/10.1007/s41237-019-00102-7

Tuerlinckx, F., & De Boeck, P. (2005). Two interpretations of the discrimination parameter. *Psychometrika*, 70(4), 629–650. https://doi.org/10.1007/s11336-000-0810-3.

van Rijn, P., & Rijmen, F. (2015). On the explaining - away phenomenon in multivariate latent variable models. *British Journal of Mathematical and Statistical Psychology*, *68*(1), 1-22. https://doi.org/10.1111/bmsp.12046

CDM ➡ IRT 研究の可能性

- 確認的多次元順序回答モデルの実用化：本来、質問紙回答の生成モデルとして最も妥当なモデルと思われる。実際医学系ではIRTの活用が進んでいる(e.g., PROMISの項目バンク作成; Irwin et al., 2010)
- Neyman-Scott問題に対し、CDMでは同時最尤推定復権の動きが見られる(Chiu et al. 2016)がIRTでは？
- 構造推定：Q行列推定法は推測統計学的研究が進んでいる(e.g., Chen et al., 2015)。多次元IRTにおける構造推定の新たな展開は？(e.g., Doebler & Doebler, 2020)

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. Journal of the American Statistical Association, 110(510), 850-866. https://doi.org/10.1080/01621459.2014.934827

Chiu, C. Y., Köhn, H. F., Zheng, Y., & Henson, R. (2016). Joint maximum likelihood estimation for diagnostic classification models. *Psychometrika*, *81*(4), 1069-1092. https://doi.org/10.1007/s11336-016-9534-9

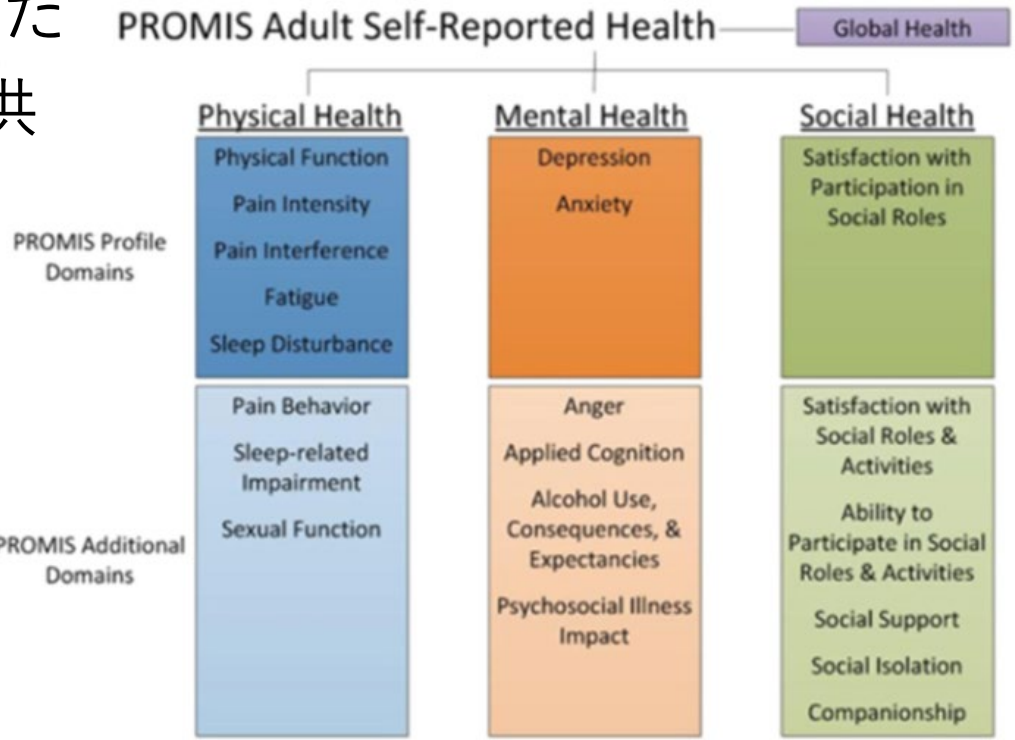
Doebler, A., & Doebler, P. (2020). Rotate and Project: Measurement of the Intended Concept with Unidimensional Item Response Theory from Multidimensional Ordinal Items. *Multivariate Behavioral Research*, Online Ahead of Print. https://doi.org/10.1080/00273171.2020.1794776

Irwin, D. E. et al. (2010). An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Quality of Life Research*, *19*(4), 595-607. https://doi.org/10.1007/s11136-010-9619-3

PROMIS

(Patient-Reported Outcomes Measurement Information System)

- 2004年に米国NIHが主導してはじめた、医療における患者報告式アウトカム尺度(PRO)の開発・運用にかかわる方法・技術の整備を目指したプロジェクト
- 項目反応モデルを活用した**項目バンク**の整備、ソフトウェア開発が行われてきた
- 項目バンクは3形式で提供
  - 短縮版
  - プロフィール版
  - コンピュータ適応型テスト(CAT)版



Tucker, C. A et al. (2014). Concept analysis of the Patient reported outcomes measurement information system (PROMIS®) and the international classification of functioning, disability and health (ICF). *Quality of Life Research*, *23*(6), 1677-1686. https://doi.org/10.1007/s11136-014-0622-y

まとめ

- IRTとCDMは現代における**計量心理学・テスト理論の2大モデル**群と言うことができ、いずれも活発に研究が進んでいる
- 両者は**目的変数が2値観測変数**であり、**説明変数が解答者の潜在変数**であり、**ロジットリンク関数**を用いるという基本的な点において共通している
- そのため、一方のモデルで研究する価値のあることは、他方のモデルでも然りであることが多い
- 一方、両者の違いは**潜在変数の設定**（連続／離散、1次元／多次元、粒度）にある。これは**応用上の目的**の違い（**総括的評価／形成的評価**）を反映したものである
- 両者の類似点・相違点を念頭に、方法論の開発・評価、および応用・実践研究を進めることが有用と思われる