第 20 回統計教育の方法論ワークショップ

特異値分解から見た項目反応理論の新評価

— 大学数学のCBTによるテスト —

廣瀬英雄

久留米大学客員教授

中央大学研究開発機構教授

2023年3月13日

多数の問題と多数の受験者の正答誤答表から、
問題の難しさのレベルと受験者の能力のレベルの
両方を同時に推定できる。

項目反応理論

Item Response Theory

IRT

推定値の信頼度も高い

本日のテーマ

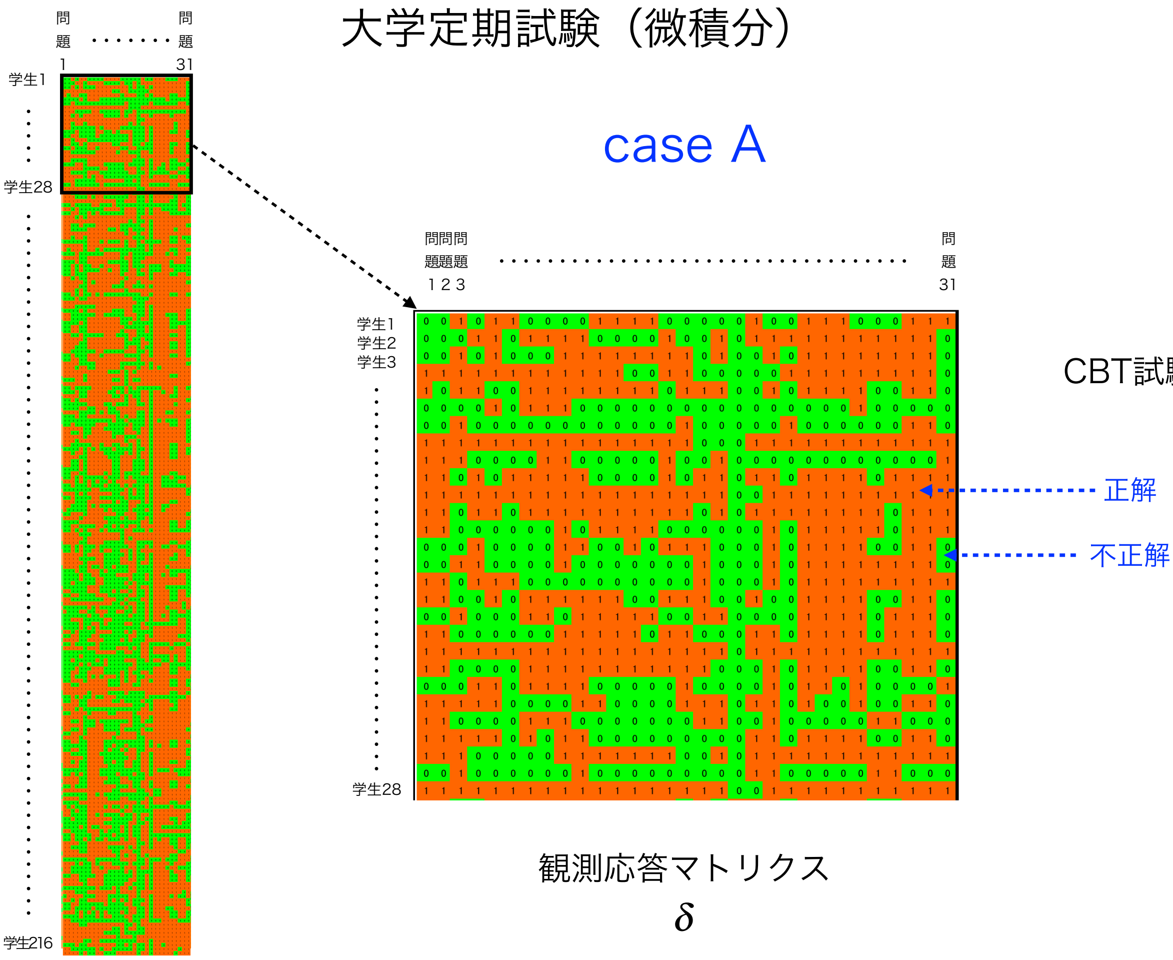
汎用性の高い評価法IRTそのものをどう評価するか

応答マトリクスから評価する

大学定期試験（微積分）

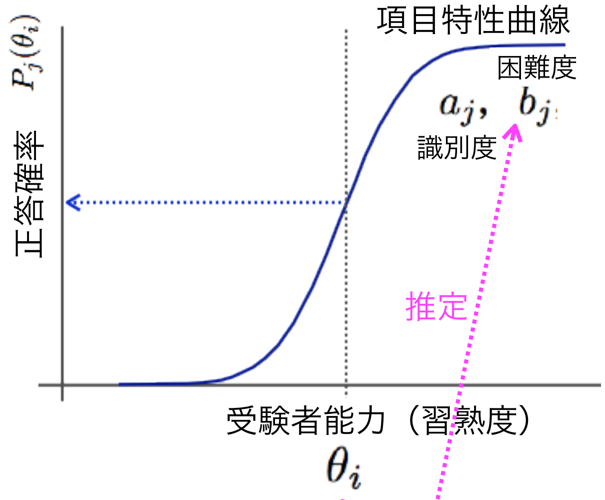
case A

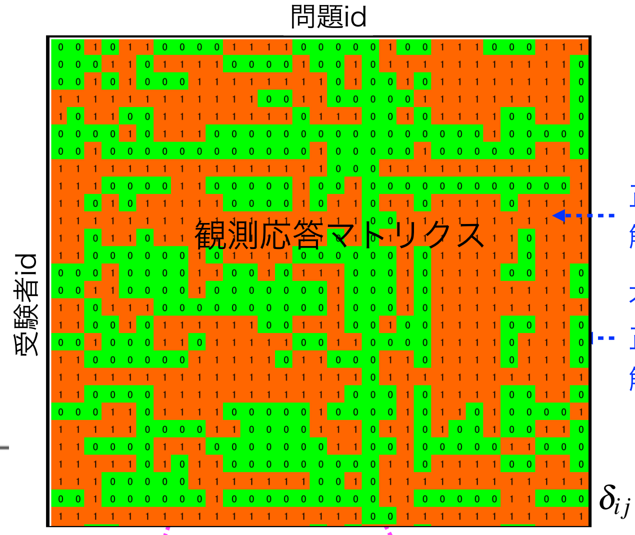
CBT試験



観測応答マトリクス
δ

項目反応理論（IRT）の数理モデル





入力データ

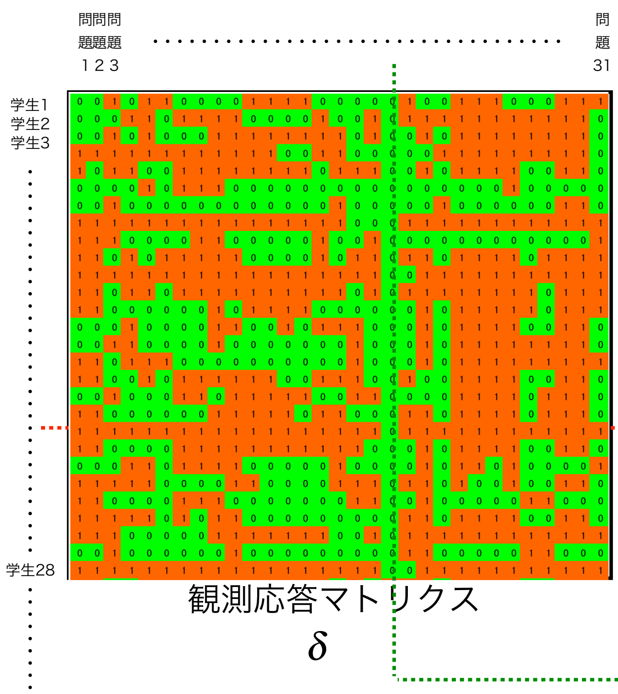
最尤推定法　推定

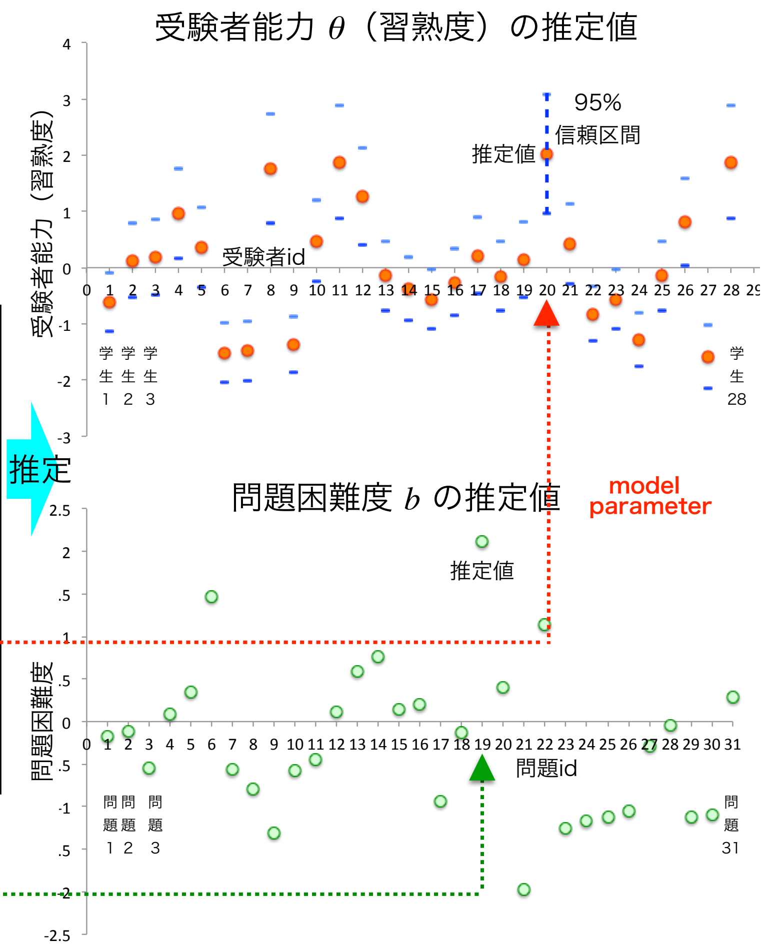
$$L = \prod_{i=1}^{N} \prod_{j=1}^{n} P_j(\theta_i)^{\delta_{i,j}} \cdot (1 - P_j(\theta_i))^{1-\delta_{i,j}}$$

ロジスティック確率分布　$P_j(\theta_i) = \dfrac{1}{1 + \exp\{-1.7 a_j (\theta_i - b_j)\}}$

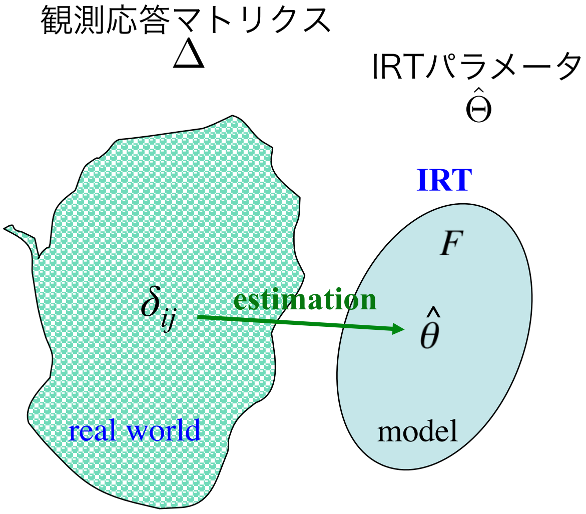
未知パラメータ　未知パラメータ

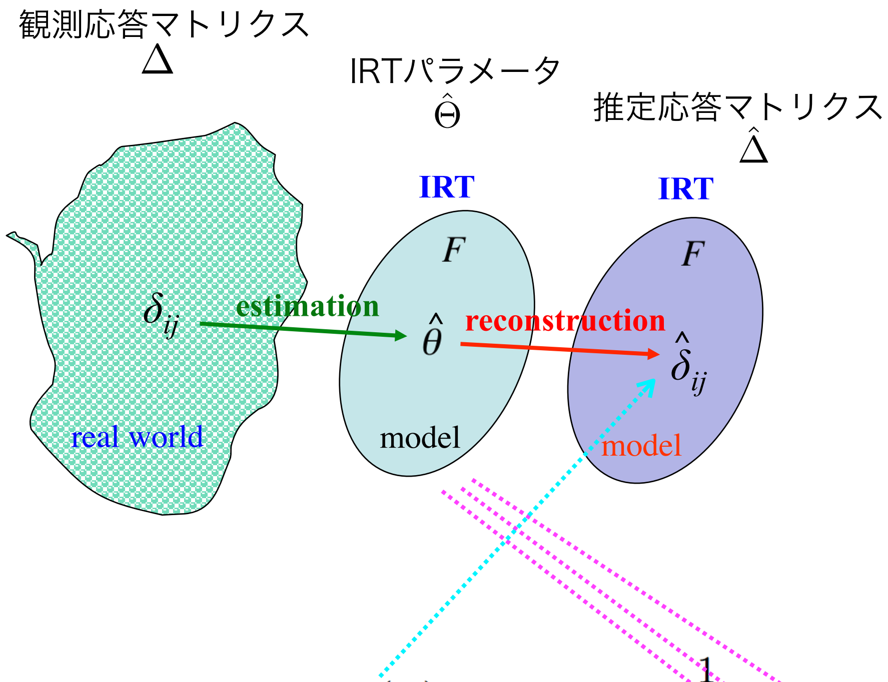
観測データ
case A





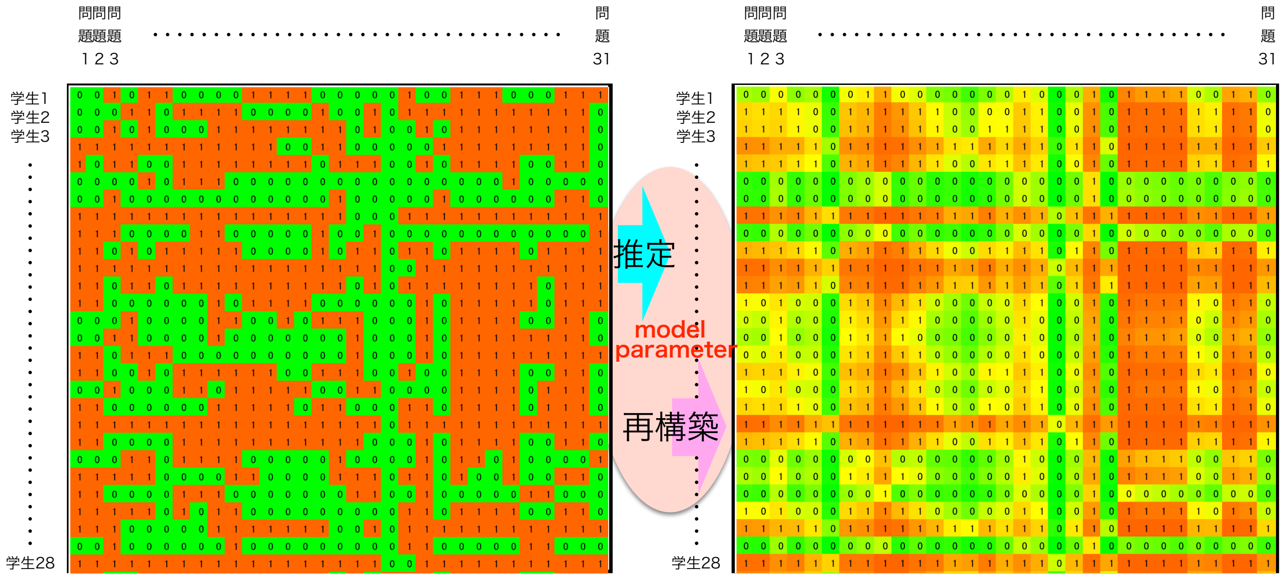
IRTによる推定





$$P_j(\theta_i) = \frac{1}{1 + \exp\{-1.7a_j(\theta_i - b_j)\}}$$

観測応答マトリクス　　IRT推定応答マトリクス



IRTの応答マトリクス再構築能力を評価したい

$$\mathrm{RMSE}(A,B) = \sqrt{\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}(a_{ij}-b_{ij})^2}$$
$$= \sqrt{\frac{1}{nm}(||A-B||_F)^2}.$$

IRTから再構成されたマトリクスと
観測マトリクスとの差のノルムを求める

比較

他の方法により再構成されたマトリクスの
観測マトリクスとの誤差を求める

$$
\begin{aligned}
\mathrm{RMSE}(A,B) &= \sqrt{\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}(a_{ij}-b_{ij})^2} \\
&= \sqrt{\frac{1}{nm}(||A-B||_F)^2}.
\end{aligned}
$$

IRTから再構成されたマトリクスの
観測マトリクスとの誤差を求める

比較

特異値分解(SVD)から再構成された低近似マトリクスの
観測マトリクスとの誤差を求める

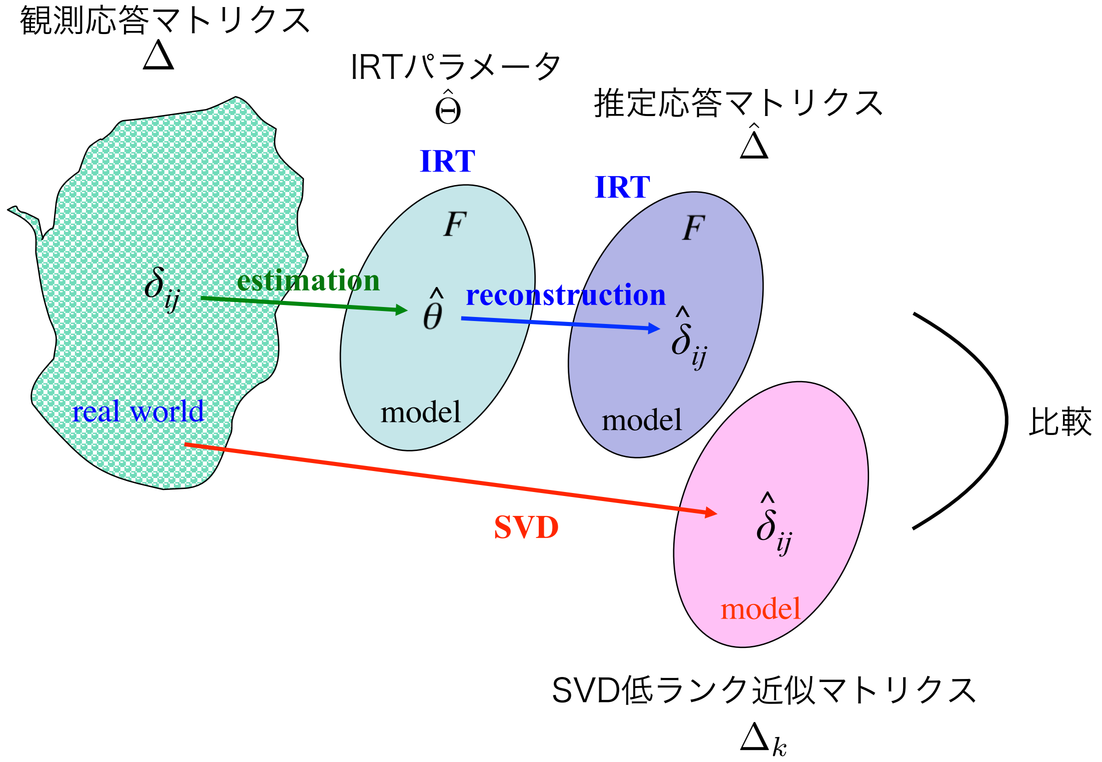
IRTと同程度の近似ができるランクkを求める

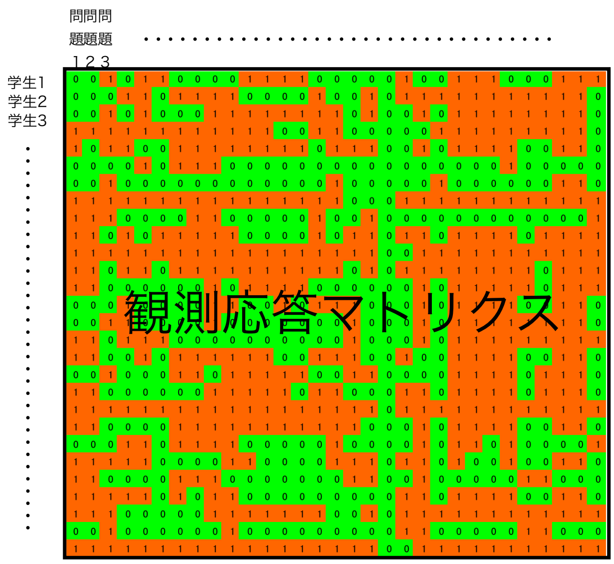
k（分解の複雑度）

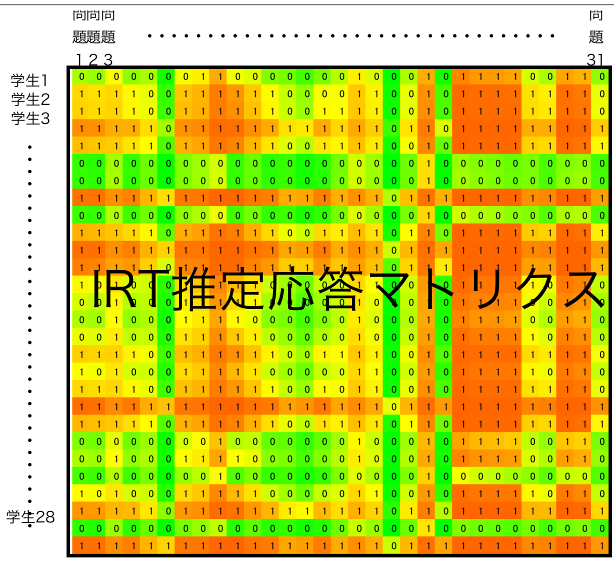
$$
A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m1} & \cdots & a_{mn} \end{pmatrix} = (a_{ij}) \xrightarrow{\text{SVD}} A = \sum_{l=1}^{r} \sigma_l \boldsymbol{u}_l \boldsymbol{v}_l^{\mathsf{T}}
$$

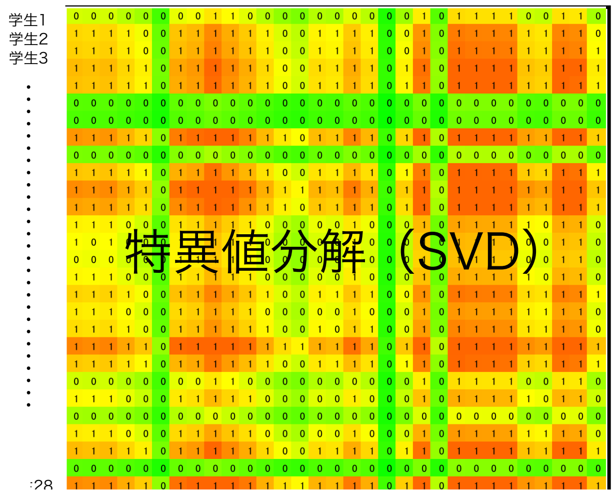
$$
\boldsymbol{u}_l = \begin{pmatrix} u_{l1} \\ u_{l2} \\ \vdots \\ u_{lm} \end{pmatrix} \quad \boldsymbol{v}_l = \begin{pmatrix} v_{l1} \\ v_{l2} \\ \vdots \\ v_{ln} \end{pmatrix}
$$

$$
A_k = \sum_{l=1}^{k} \sigma_l \boldsymbol{u}_l \boldsymbol{v}_l^{\mathsf{T}}
$$



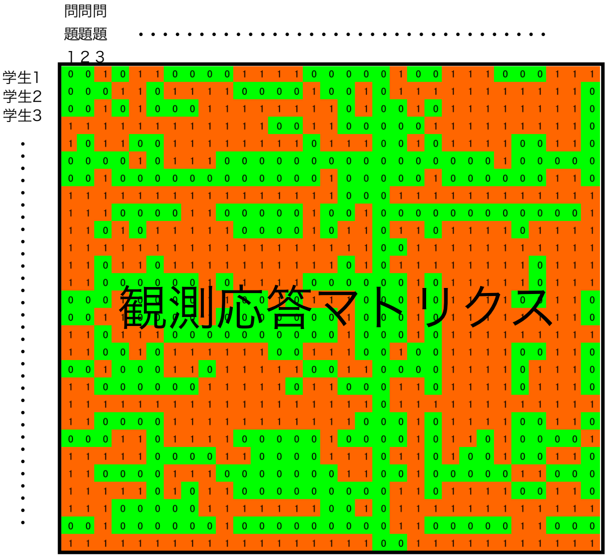


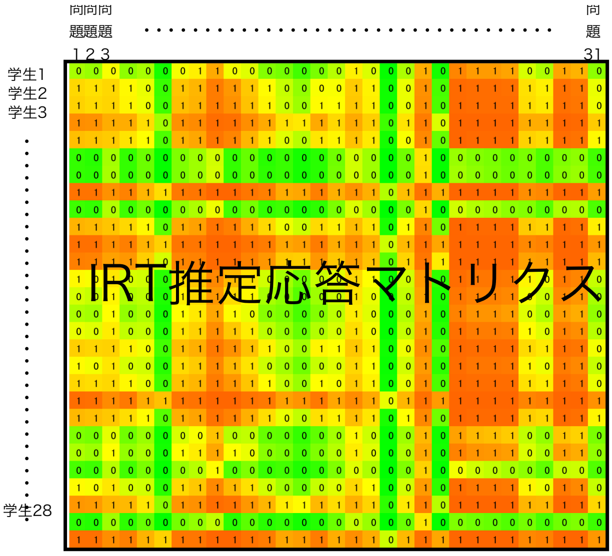


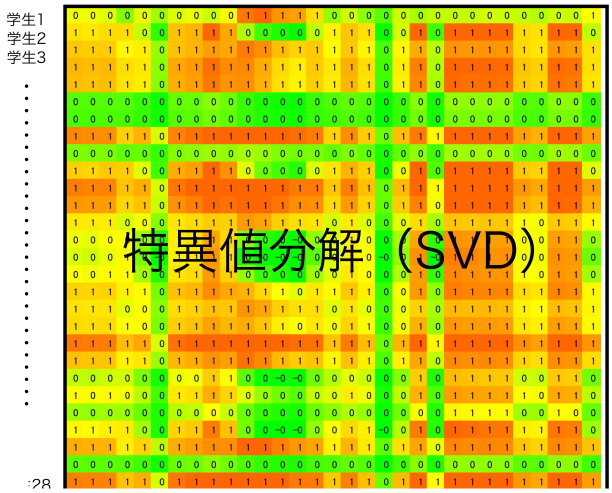


低ランクマトリクスのランク

$k=1$（分解の複雑度）

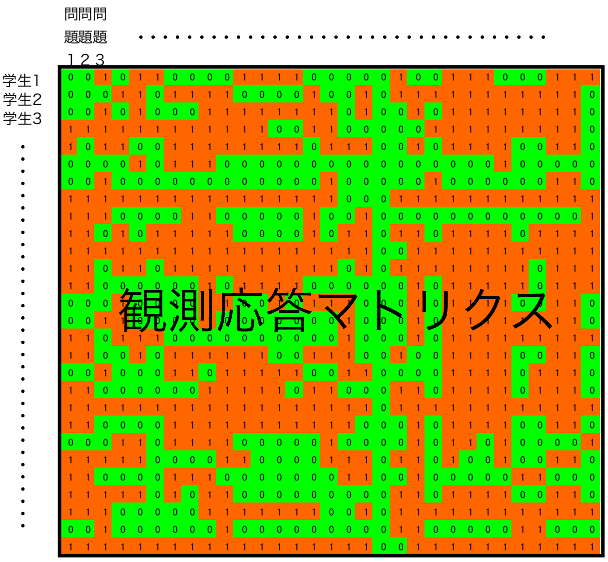


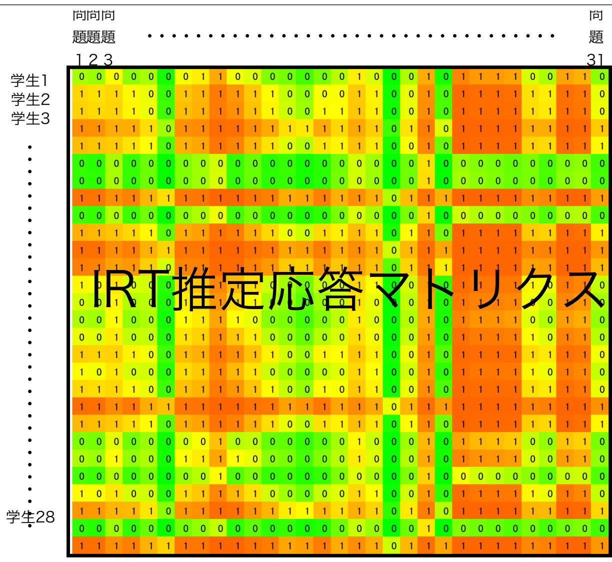


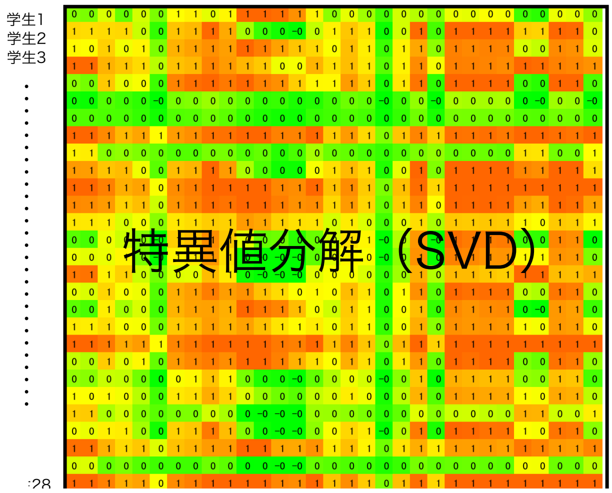


低ランクマトリクスのランク

$k = 2$（分解の複雑度）

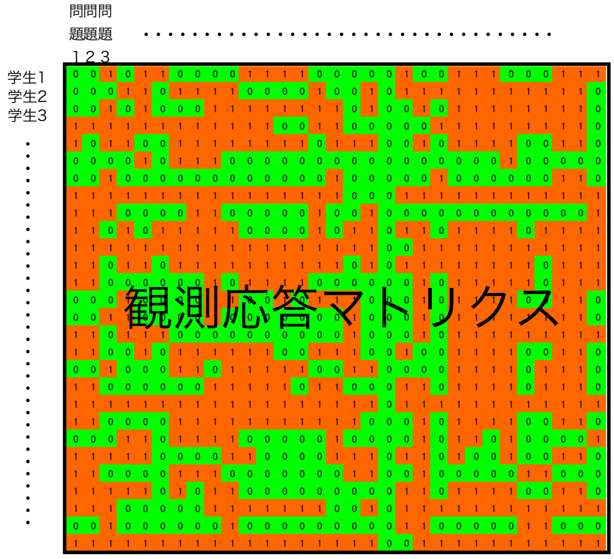


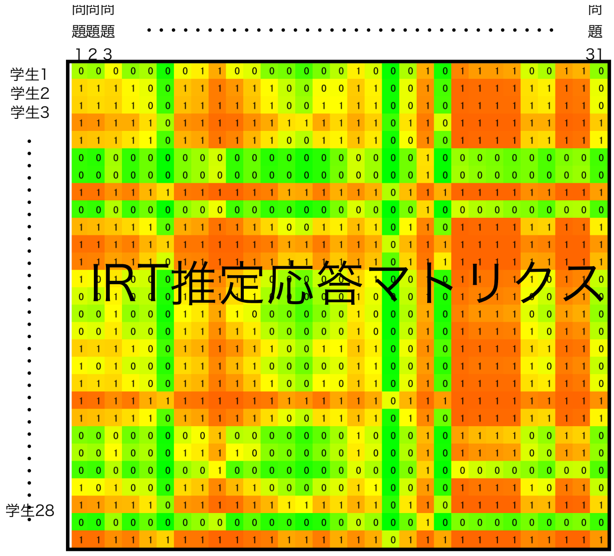


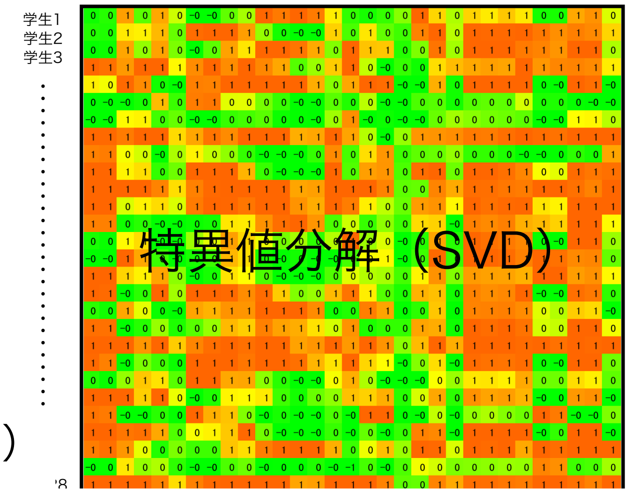


低ランクマトリクスのランク

$k = 3$（分解の複雑度）

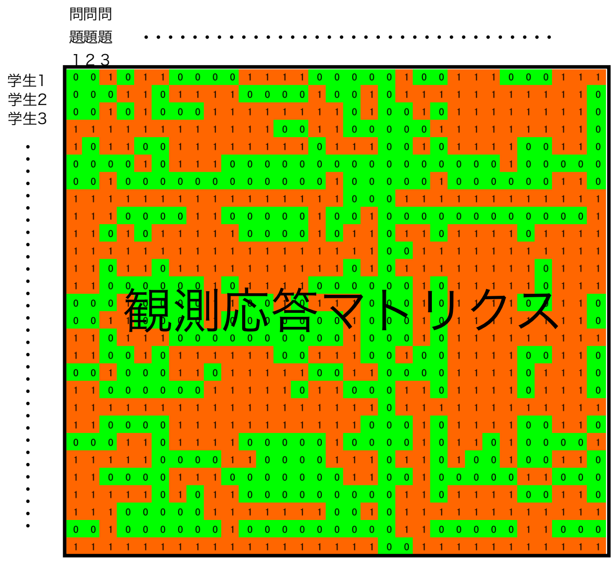


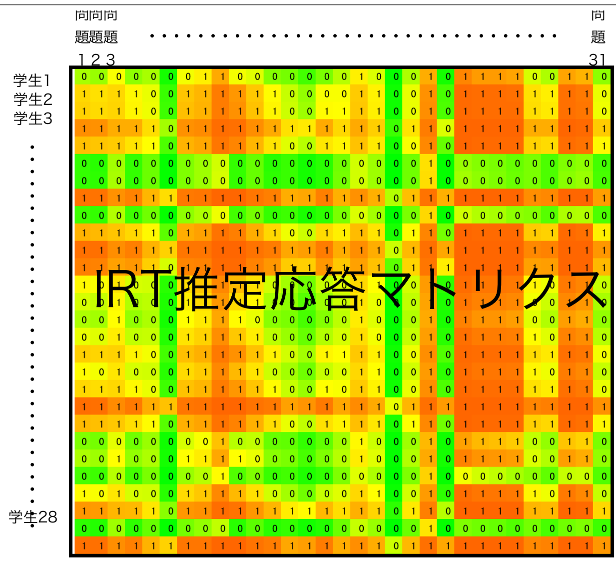


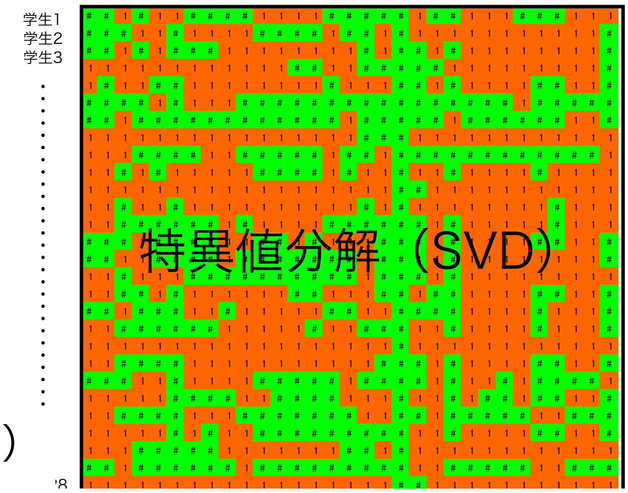


低ランクマトリクスのランク

$k = 10$（分解の複雑度）

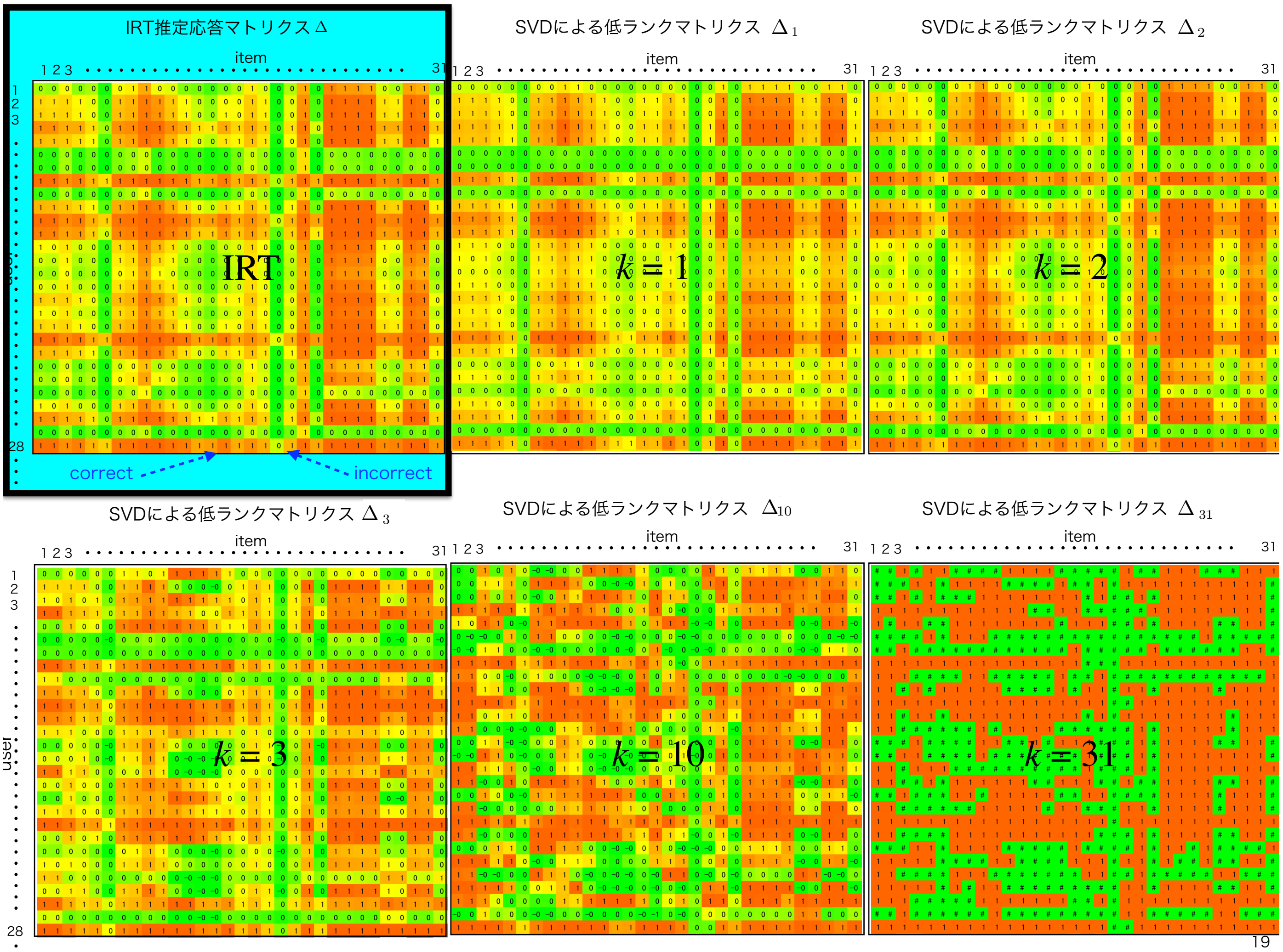






低ランクマトリクスのランク

$k = 31$（分解の複雑度）



低ランク近似マトリクスのランク数によるRMSEの違い

ランク数	SVD	IRT
k	$\mathrm{RMSE}(\Delta_k, \Delta)$	$\mathrm{RMSE}(\hat{\Delta}, \Delta)$
		0.3915
1	0.4066	
2	0.3851	
3	0.3652	
4	0.3479	
5	0.3306	
10	0.2562	
20	0.1325	
31	0	

IRTの近似精度は
ランクk=2以下の
SVDの近似マトリクスの精度以下

項目反応理論から

応答マトリクスを再現すると
$k < 2$ 程度の低ランクマトリクス近似の精度しか得られない

case A

42ケースのデータでも調べる

id	subject	n	m	id	subject	n	m	id	subject	n	m	
1	PS	44	14	15	C	45	42	29	C	1131	77	
2	PS	41	19	16	LA	36	39	30	LA	1101	84	
3	PS	34	17	17	LA	566	6	31	C	215	36	
4	P	97	32	18	C	66	33	32	LA	47	39	
5	S	57	15	19	S	97	12	33	C	209	36	
6	S	75	14	20	C	76	30	34	C	868	6	
7	C	40	19	21	LA	132	49	35	C	215	36	
8	PS	44	15	22	LA	132	84	36	LA	585	84	
9	PS	72	21	23	LA	177	49	37	LA	39	39	
10	ODE	41	13	24	LA	142	45	38	C	209	31	
11	ODE	49	25	25	LA	46	39	39	C	209	67	
12	PS	54	21	26	LA	39	45	40	C	216	31	← ケース A
13	C	70	26	27	LA	181	45	41	LA	585	49	
14	C	9	16	28	LA	229	84	42	C	145	34	

PS: 確率・統計, P: 確率論, S: 統計学,
ODE: 常微分方程式, C: 微積分, LA: 線形代数

項目反応理論から

応答マトリクスを再現すると
$k < 2$ 程度の低ランクマトリクス近似の精度しか得られない

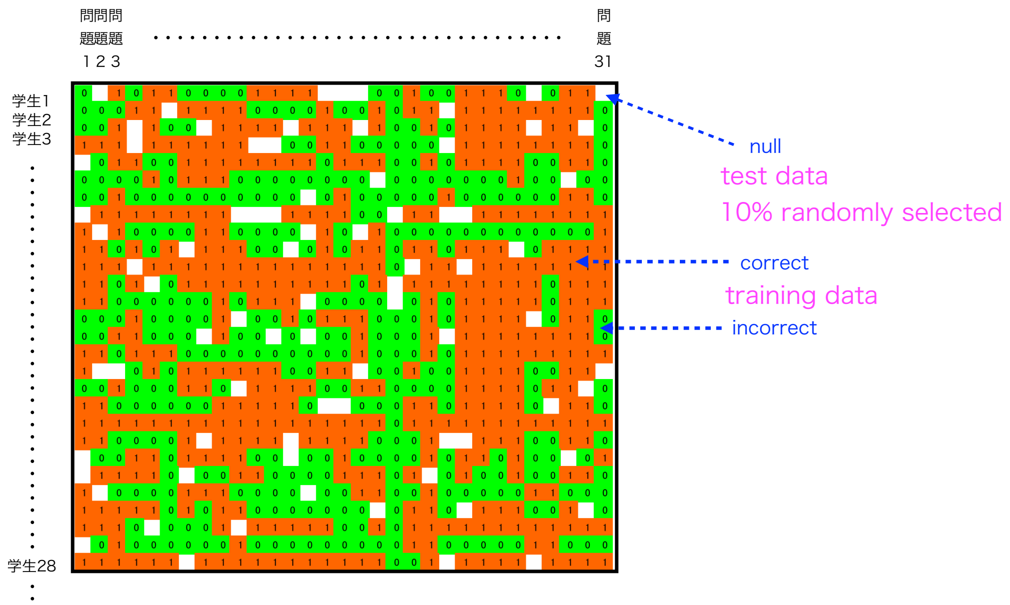
42ケースすべてに対して成立している

項目反応理論から

応答マトリクスを再現すると
$k < 2$ 程度の低ランクマトリクス近似の精度しか得られない

それはトレーニングデータでの結果であって
テストデータでもそういえるだろうか

予測誤差を調べるため



training data で予測モデルを作り

test data で予測誤差を調べてみる

項目反応理論から

応答マトリクスを再現すると

$k < 2$ 程度の低ランクマトリクス近似の精度しか得られない

テストデータでもそうだった

ただし、case Aのとき

42ケースから8ケースを選んで調べる

8ケース

case id	$\mu(\text{RMSE}(\tilde{T}^{\text{SVD}}_{k\_opt}, T))$	k_opt	$\mu(\text{RMSE}(\tilde{T}, T))$	case name
5	0.2935	1	0.2908	
10	0.3547	1	0.3591	
15	0.3457	1	0.3344	
20	0.3801	1	0.3726	
25	0.3701	5	0.3789	
30	0.3442	16	0.3771	case B
35	0.3982	2	0.3964	
40	0.4053	3	0.4068	case A

最適なkのときの
低ランク
近似マトリクスで

case id	$\mu(\mathrm{RMSE}(\tilde{T}^{\mathrm{SVD}}_{k\_opt}, T))$	k_opt	$\mu(\mathrm{RMSE}(\tilde{T}, T))$	case name
5	0.2935	1	0.2908	
10	0.3547	1	0.3591	
15	0.3457	1	0.3344	
20	0.3801	1	0.3726	
25	0.3701	5	0.3789	
30	0.3442	16	0.3771	case B
35	0.3982	2	0.3964	
40	0.4053	3	0.4068	case A

8ケース

変わらない

本日のテーマ

汎用性の高い評価法IRTそのものをどう評価するか

応答マトリクスから評価すると

IRTの推定能力は

応答マトリクスサイズが中規模以下（n=100, m=50）なら、IRTの推定応答マトリクスの再現能力は特異値分解のそれと同程度（マトリクスのランクはどちらもかなり小さい）

IRTによる推定能力は推定の限界近くまで達している

応答マトリクスサイズが大規模（n=1000，m=100）なら、IRTの推定応答マトリクスの再現能力は特異値分解のそれより悪い（マトリクス分解のマトリクスのランクが大きくなる）

IRTの推定能力はかなり高いが
IRT以外の推定方法を確立できる余地は残されている

特異値分解から見た項目反応理論の新評価

— 大学数学のCBTによるテスト —

thank you

廣瀬英雄
久留米大学客員教授
中央大学研究開発機構教授

2023年3月13日