

テストデータを分析する:
古典的テスト理論・項目反応理論・
潜在ランク理論による学力評価

庄島宏二郎
大学入試センター研究開発部
東京工業大学大学院社会理工学研究科
shojima@rd.dnc.ac.jp

Classical Test Theory (CTT)

- 初めて、テストデータを科学的対象とした
- 狭義:正答数と和得点に対する標準化理論
- CTTの最大の貢献は信頼性係数
- 尺度化
 - テストの性能(狭義には統計的性質)を解明
 - テスト統計量
 - 項目統計量
 - 受検者統計量

データ行列(data matrix)

Sub. \ Item	1	2	...	j	...	J
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1J}
2	X_{21}	X_{22}	...	X_{2j}	...	X_{2J}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{iJ}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
I	X_{I1}	X_{I2}	...	X_{Ij}	...	X_{IJ}

X_{ij} : 受検者 i の項目 j に対する反応 (項目反応 item response)
 I : 受検者数 (被験者数 sample size)
 J : 項目数 (変量数 number of items/variables)

欠測指示子行列
(missing indicator matrix)

Sub. \ Item	1	2	...	j	...	J
1	Z_{11}	Z_{12}	...	Z_{1j}	...	Z_{1J}
2	Z_{21}	Z_{22}	...	Z_{2j}	...	Z_{2J}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	Z_{i1}	Z_{i2}	...	Z_{ij}	...	Z_{iJ}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
I	Z_{I1}	Z_{I2}	...	Z_{Ij}	...	Z_{IJ}

Z_{ij} : 受検者 i の項目 j に対する欠測指示子 (missing indicator)
 観測されれば1、欠測であれば0の2値データ
 (binary data/dichotomous data)
 欠測指示子は、フルデータとして得られる

受検者統計量(1)

回答数 (解答数) $f_i = \sum_{j=1}^J z_{ij} = \mathbf{1}_j' \mathbf{z}_i$

- 観測数(number of observations)
- 値域[0, J]: 0に近いほど、欠測している

回答率 (解答率) $\bar{z}_i = \frac{f_i}{J} = \frac{\sum_{j=1}^J z_{ij}}{J} = \mathbf{1}_j' \mathbf{z}_i / J$

- 観測率(observation rate)
- 値域[0, 1]: 0に近いほど、欠測率が高い

正答数 (和得点) $t_i = \sum_{j=1}^J z_{ij} w_j x_{ij} = \mathbf{z}_i' (\mathbf{w} \circ \mathbf{x}_i)$

- Number-right score (total score)
- 項目得点の総和

平均 (正答率) $\bar{x}_i = \frac{t_i}{f_i} = \frac{\sum_{j=1}^J z_{ij} w_j x_{ij}}{\sum_{j=1}^J z_{ij}} = \frac{\mathbf{z}_i' (\mathbf{w} \circ \mathbf{x}_i)}{\mathbf{1}_j' \mathbf{z}_i}$

- Mean (correct response rate)
- 正答率ならば値域[0, 1]
- 受検者の能力についての最重要資料の1つ

標準化得点 $z_i = \frac{t_i - \bar{t}}{s_t}$

- Standardized score
- Z-score
- 標準正規分布における位置を知る

偏差値得点 $y_i = z_i \times 10 + 50$

- 標準化得点の線形変換
- 東アジアのみ
- 諸外国では使われず

項目統計量

回答数 (解答数) $f_j = \sum_{i=1}^I z_{ij} = \mathbf{1}_i' \mathbf{z}_j$

- 観測数(number of observations)
- 値域[0, I]: 0に近いほど、欠測している

回答率 (解答率) $\bar{z}_j = \frac{f_j}{I} = \frac{\sum_{i=1}^I z_{ij}}{I} = \mathbf{1}_i' \mathbf{z}_j / I$

- 観測率(observation rate)
- 値域[0, 1]: 0に近いほど、欠測率が高い

正答数(和) $t_j = \sum_{i=1}^I z_{ij} x_{ij} = z_j' x_j$

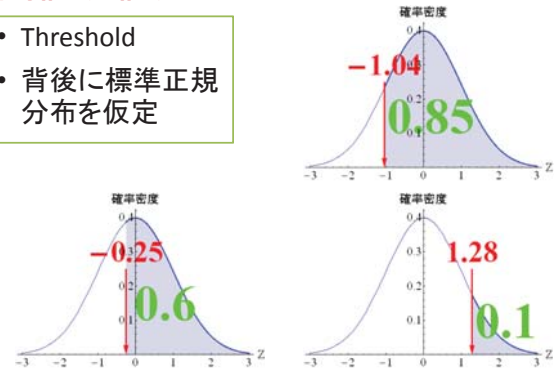
- Number of correct responses (sum)
- 項目得点の総和(意味がないことが多い)

正答率(平均) $\bar{x}_j = \frac{t_j}{f_j} = \frac{\sum_{i=1}^I z_{ij} x_{ij}}{\sum_{i=1}^I z_{ij}} = \frac{z_j' x_j}{\mathbf{1}_I' z_j}$

- Correct response rate (mean)
- 正答率ならば値域[0, 1]
- 項目の難易度に関する初歩だが重要資料
- さまざまな大きさの正答率(平均)をもつ項目を用意したい

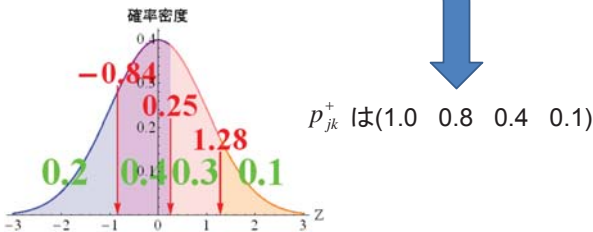
閾値(2値) $\tau_j = 1 - N^{-1}(p_j)$

- Threshold
- 背後に標準正規分布を仮定



閾値(多値) $\tau_{jk} = 1 - N^{-1}(p_{jk}^+)$

- Ex) 4件法リッカートデータ
- カテゴリ(0 1 2 3)
- 選択率 p_{jk} が (0.2 0.4 0.3 0.1)



分散 $s_j^2 = \frac{\sum_{i=1}^I z_{ij} (x_{ij} - \bar{x}_j)^2}{\sum_{i=1}^I z_{ij} - 1}$

- variance
- 平均的な正方形の面積

標準偏差 $s_j = \sqrt{\frac{\sum_{i=1}^I z_{ij} (x_{ij} - \bar{x}_j)^2}{\sum_{i=1}^I z_{ij} - 1}}$

- Standard deviation (SD)
- 平均的な正方形の面積の一辺
- 大きいほうが良い

項目平均情報量 $e_j = -\sum_{k=1}^{K_j} p_k \log_2(p_k)$

- Item entropy
- K_j : 項目 j のカテゴリ数
- p_k : カテゴリ k の選択率
- 値域 $[0, \log_2 K_j]$
 - 値が大きいほど、反応がバラけている
 - 0に近いほど、反応がどこかのカテゴリに集中
- 多少、分散やSDとかぶる指標だが、質的(名義・順序)データの分散の指標として使える

観測得点相関 $r(z_j, t) = \frac{\sum_{i=1}^I (z_{ij} - \bar{z}_j)(t_i - \bar{t})}{(I-1)s_{z_j}s_t}$

Observation-total correlation
 値域[-1, 1]: 1に近いほど、得点が高い人ほど観測し、低い人ほど欠測している、vice versa

項目得点相関 $r(x_j, t) = \frac{\sum_{i=1}^I z_{ij} (x_{ij} - \bar{x}_j)(t_i - \bar{t})}{(f_j - 1)s_{x_j}s_t}$

- Item-total correlation
- 値域[-1, 1]: 1に近いほど、和得点が高い人ほど項目得点が高かった(項目の識別力)
- 絶対値が高い項目を多くそろえたい

項目リマインダ相関

- Item-remainder correlation
- 値域[-1, 1]: 1に近いほど、得点が高い人ほど観測し、低い人ほど欠測している、vice versa
- IT相関より低めの値になる

例: 項目2のIR相関

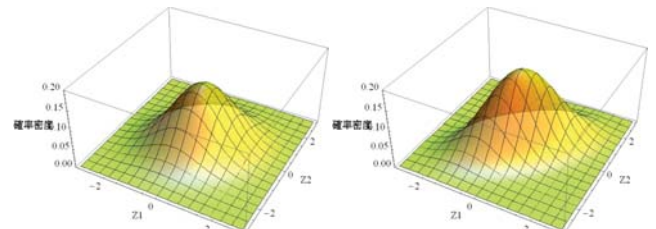
Sub. \ Item	1	2	...	j	...	J	t	Rem.
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1J}	t_1	$t_1 - X_{12}$
2	X_{21}	X_{22}	...	X_{2j}	...	X_{2J}	t_2	$t_2 - X_{22}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{iJ}	t_i	$t_i - X_{i2}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
I	X_{I1}	X_{I2}	...	X_{Ij}	...	X_{IJ}	t_I	$t_I - X_{I2}$

項目得点双列相関

- Item-total biserial correlation
- 2値のカテゴリカルデータと連続データの相関
- 2変量正規分布を仮定
- IT相関より絶対値が大きくなる傾向

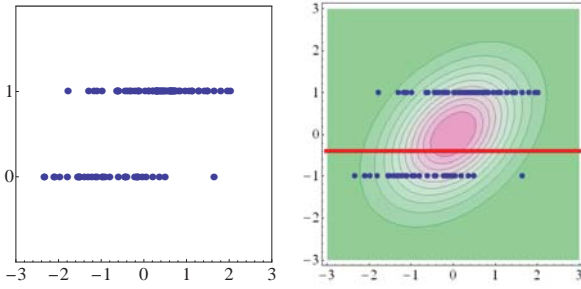
相関0.2の2変量正規分布

相関0.6の2変量正規分布



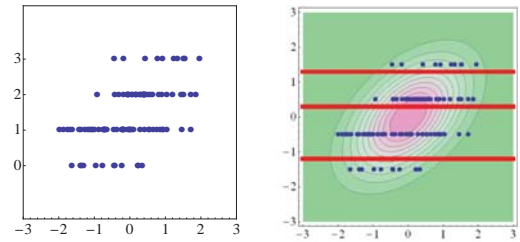
$$g(\rho) = \sum_{i=1}^N z_{ij} \{u_{ij} \log p_{ij} + (1-u_{ij}) \log(1-p_{ij})\} \quad \left(s.t. \quad p_{ij} = \int_{-\infty}^{\infty} N(x|z_{ij}\rho, \sqrt{1-\rho^2}) \right)$$

- 背後にどのような2変量正規分布が仮定できるか尤度を立てて解く



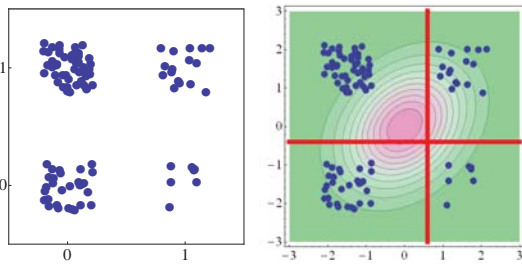
項目得点多列相関

- Item-total polyserial correlation
- 多値の順序カテゴリカルデータと連続データの相関
- 2変量正規分布を仮定
- IT相関より絶対値が大きくなる傾向



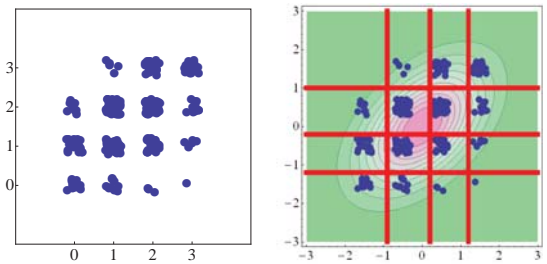
項目間四分相関

- Inter-item tetrachoric correlation
- 2値のカテゴリカルデータ同士の相関
- 2変量正規分布を仮定



項目間多分相関

- Inter-item polychoric correlation
- 多値のカテゴリカルデータ同士の相関
- 2変量正規分布を仮定



信頼性係数

- CTTの最大の貢献
- 測定値の分散に占める真値の分散説明率

$$X = T + E$$

$$V[X] = V[T] + V[E] \quad (\text{ただし, } Cov[T, E] = 0)$$

$$\rho = \frac{V[T]}{V[X]} \quad (\leq 1)$$

Cronbachのα係数 (1)

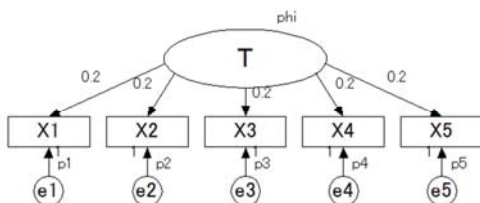
- 内部一貫性の指標
 - もっとも有名な信頼性係数
 - 2値データの場合はKuder-Richardsonの公式20 (KR20)

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{j=1}^n s_j^2}{s^2} \right) \quad \begin{array}{l} J: \text{項目数} \\ s^2: \text{和得点分散} \\ s_j^2: \text{項目}j\text{の分散} \end{array}$$

- 性質
 - 上限は1.0, 下限は無限小
 - 項目数が多いほど, 高くなりやすい

Cronbachのα係数 (2)

- 弱平行(weakly parallel)測定の仮定
 - 1因子の因子分析で, すべてのパス係数を1/nとおいた状況



Cronbachのα係数 (3)

共分散によるαの表現

$$\alpha = \frac{n}{n-1} \left(\frac{2 \sum_{j=2}^{J-1} \sum_{j'=j+1}^J s_{jj'}}{\sum_{j=1}^J s_j^2 + 2 \sum_{j=2}^{J-1} \sum_{j'=j+1}^J s_{jj'}} \right) \quad \begin{array}{l} J: \text{項目数} \\ s_j^2: \text{項目}j\text{の分散} \\ s_{jj'}: \text{項目}j\text{と項目}j'\text{の共分散} \end{array}$$

相関によるαの表現: Standardized Alpha

$$\alpha_{st} = \frac{n}{n-1} \left(\frac{2 \sum_{j=2}^{J-1} \sum_{j'=j+1}^J r_{jj'}}{J + 2 \sum_{j=2}^{J-1} \sum_{j'=j+1}^J r_{jj'}} \right) \quad \begin{array}{l} J: \text{項目数} \\ r_{jj'}: \text{項目}j\text{と項目}j'\text{の相関} \end{array}$$

McDonaldの ω 係数 (1)

- 1因子性の指標
 - もっとも現実的な信頼性係数
 - 1因子FAモデルの和得点に対する決定係数

$$\omega = \frac{\left(\sum_{j=1}^n \lambda_j\right)^2}{s^2} = \frac{\mathbf{1}'\boldsymbol{\lambda}\boldsymbol{\lambda}'\mathbf{1}}{s^2} = 1 - \frac{\sum_{j=1}^n e_j^2}{s^2}$$

λ_j : 項目 j の因子パターン

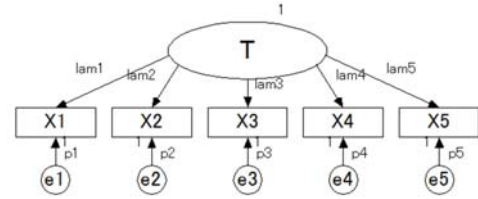
$\boldsymbol{\lambda}$: λ_j を要素とするベクトル

$\mathbf{1}$: 全要素が1のベクトル

e_j^2 : 項目 j の誤差分散

McDonaldの ω 係数 (2)

- 弱同族(weakly congeneric)測定の仮定
 - 1因子の因子分析で, すべてのパス係数が自由推定である状況



SKIP

多変量決定係数(1)

- 多因子FAモデルの信頼性の指標
 - もっとも将来性の高い信頼性係数
 - FAモデルと共分散行列の一般化分散の比

$$\rho = 1 - \frac{|\boldsymbol{\Psi}|}{|\boldsymbol{\Sigma}|}$$

$\boldsymbol{\Psi}$: 誤差共分散行列 (通常は対角行列)

$|\boldsymbol{\Psi}|$: $\boldsymbol{\Psi}$ の行列式 (体積)

$\boldsymbol{\Sigma}$: 共分散行列

$|\boldsymbol{\Sigma}|$: $\boldsymbol{\Sigma}$ の行列式 (体積)

SKIP

多変量決定係数 (2)

- 従属変数群に対する独立変数群の説明分散割合
 - データが張る空間の体積に対する, モデルが張る空間の体積の比

SKIP

一般化可能性理論

- 信頼性の問題を分散分析の応用問題として再定式
 - テスト得点を, 受検者・項目・採点者の要因等に分解
 - 分散分析の変量モデル
- G研究(generalizability study)
 - 実験計画を用いて変動要因を同定
- D研究(decision study)
 - G研究の情報をもとに, 特定状況(シナリオ)下での信頼性(一般化可能性係数)を検討

項目反応理論

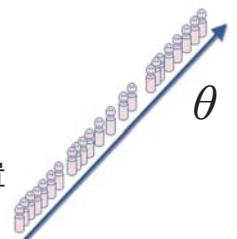
- Item response theory (IRT)
- テスト標準化理論のグローバルスタンダード
 - TOEFL
 - TOEIC
 - 医師国家試験(厚労省)
 - 情報処理技術者試験(通産省)
 - 日本留学試験(文科省)

IRTにおける尺度化

- テスト現象をパラメトリック(parametric)に
- 心理量と正答率の関係という現象を数学の言葉で構造化
- 以下の情報を得て, テストの性質を得たり, 項目選抜を行うこと
 - 項目反応関数(item response function)
 - テスト反応関数(test response function)
 - 情報関数(information function)
- 今日は2値データに対するIRTの尺度化

潜在尺度(latent scale)

- θ (theta): 能力尺度
 - 連続変数
 - 範囲 $(-\infty, \infty)$
 - 実用上の範囲 $[-3.0, 3.0]$
- θ 上に受験者の能力を布置
- 誤解を恐れない θ と偏差値の大体の対応

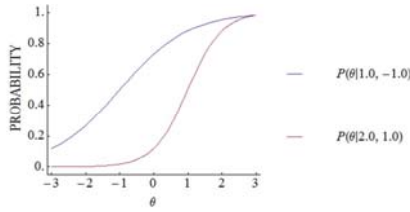


偏差値	20	30	40	50	60	70	80
θ	-3.0	-2.0	-1.0	0.0	1.0	2.0	3.0

2母数ロジスティックモデル 2-parameter logistic (2PL) model

$$P(\theta|A, B) = \frac{1}{1 + \exp\{-A(\theta - B)\}} \quad (-\infty < A < \infty, -\infty < B < \infty)$$

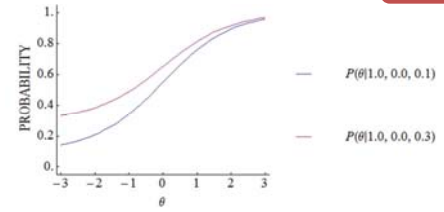
- A : 傾き母数 slope parameter (識別力母数)
- B : 位置母数 location parameter (困難度母数)



3母数ロジスティックモデル 3-parameter logistic (3PL) model

$$P(\theta|A, B, C) = C + \frac{1 - C}{1 + \exp\{-A(\theta - B)\}} \quad (0 \leq C \leq 1)$$

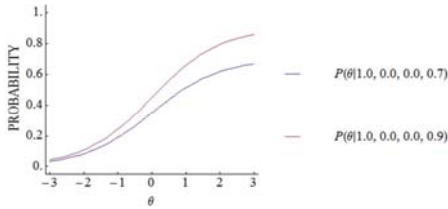
- C : 下方漸近母数 (当て推量母数) 表現として
適当でない
- lower asymptote parameter



4母数ロジスティックモデル 4-parameter logistic (4PL) model

$$P(\theta|A, B, C, D) = C + \frac{D - C}{1 + \exp\{-A(\theta - B)\}} \quad (0 \leq C < D \leq 1)$$

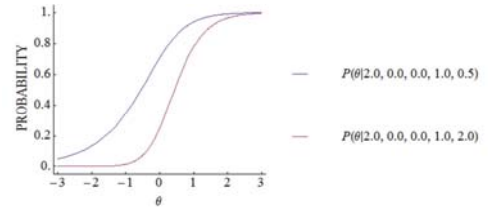
- D : 上方漸近母数
- upper asymptote parameter



5母数ロジスティックモデル 5-parameter logistic (5PL) model

$$P(\theta|A, B, C, D, E) = C + \frac{D - C}{[1 + \exp\{-A(\theta - B)\}]^E} \quad (0 < E)$$

- E : 非対称母数 asymmetry parameter



q PLモデルは $q-1$ PLモデルを含む

- 5PLモデル

$$P(\theta|A, B, C, D, E) = C + \frac{D - C}{[1 + \exp\{-A(\theta - B)\}]^E} \quad (0 < E)$$

- $E=1.0$ の5PLモデルが4PLモデル

$$P(\theta|A, B, C, D) = C + \frac{D - C}{1 + \exp\{-A(\theta - B)\}} \quad (0 \leq C < D \leq 1)$$

- $D=1.0$ の4PLモデルが3PLモデル

$$P(\theta|A, B, C) = C + \frac{1 - C}{1 + \exp\{-A(\theta - B)\}} \quad (0 \leq C \leq 1)$$

- $C=0.0$ の3PLモデルが2PLモデル

$$P(\theta|A, B) = \frac{1}{1 + \exp\{-A(\theta - B)\}} \quad (-\infty < A < \infty, -\infty < B < \infty)$$

傾き母数の前に定数があることしばしば もはや意味は喪失

- 2PLモデル

$$P(\theta|A, B) = \frac{1}{1 + \exp\{-1.7A(\theta - B)\}} \quad (-\infty < A < \infty, -\infty < B < \infty)$$

- 3PLモデル

$$P(\theta|A, B, C) = C + \frac{1 - C}{1 + \exp\{-1.7A(\theta - B)\}} \quad (0 \leq C \leq 1)$$

- 4PLモデル

$$P(\theta|A, B, C, D) = C + \frac{D - C}{1 + \exp\{-1.7A(\theta - B)\}} \quad (0 \leq C < D \leq 1)$$

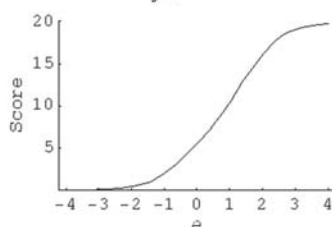
- 5PLモデル

$$P(\theta|A, B, C, D, E) = C + \frac{D - C}{[1 + \exp\{-1.7A(\theta - B)\}]^E} \quad (0 < E)$$

テスト反応関数 Test Response Function (TRF)

- θ が与えられたときのテスト得点の予測値

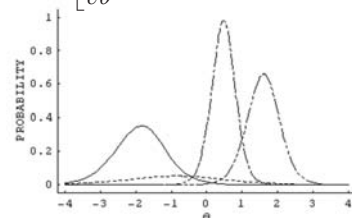
$$T(\theta) = \sum_{j=1}^n w_j P_j(\theta)$$



項目情報関数 (Item Information Function (IIF))

- 項目ごとの θ の推定精度
- フィッシャー情報量
- 対数尤度の2階微分の負の期待値

$$I_j(\theta) = -E \left[\frac{\partial^2}{\partial \theta^2} \{u_j \ln P_j(\theta) + (1 - u_j) \ln Q_j(\theta)\} \right]$$

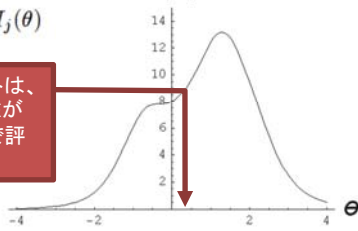


テスト情報関数 Test Information Function (TIF)

- テストの能力値に対する推定精度
- θ の標準誤差の2乗の逆数

$$I(\theta) = \sum_{j=1}^n I_j(\theta)$$

$I(\theta)$



たとえば、このテストは、 $\theta=0.4$ くらいの人をSEが0.333くらいの精度で評価できる

IRFスライド IIFスライド

IRTの代表的な仮定

- **単一次元性の仮定(unidimensionality)**
 - どんなデータも単一次元であることなどない
 - でも、単一次元でいくんだ、という**決意**
 - 統計的な仮定は脳内なので現象として不成立
 - 正規分布するデータを見たことがないのと同じ
- **局所独立の仮定(local independence)**
 - 絶対に満たさないが、満たすと思って**がんばる**
 - IRTでやたらよく言われるが、この仮定がないと因子分析も潜在クラス分析も破綻

局所独立の仮定

(assumption of local independence)

- 被検者*i*の尤度の一般形

$$p_i(\theta_i|A) = \prod_{j=1}^n P(\theta_i|\lambda_j)^{u_{ij}} \{1 - P(\theta_i|\lambda_j)\}^{1-u_{ij}}$$

- θ が所与のとき、各項目に対する反応が独立
 - 単なる掛け算となっているという意味
- 一般に、数学や理科(特に、物化)で局所独立の仮定は成立していない。
 - ある項目の正誤が、別の項目の正誤に影響

よく言われること

- N について1母数は500、2母数は1000、3母数は5000が必要とかよく言われる
 - $N=500$ で3母数だってやってよい
 - Lordは $N=1000$ くらいで3母数とか言っている
 - SEでは、100もあれば複雑なモデルを推定
- (理由)項目母数の推定精度のため
 - 彗星の軌道を推定しているわけではない
 - 項目母数の推定値は生もの、1週間後には変化
 - 項目の(統計的)性質は可変的なもの
 - 静的でないものを厳密に測定する意味がない

SKIP

多値データ

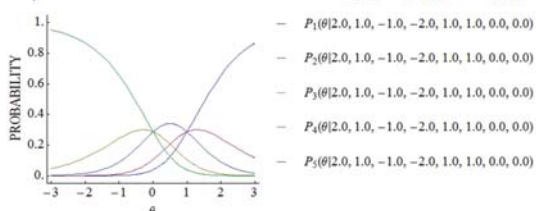
- **polytomous data** \approx polychotomous data
- 2つのタイプ
 - 名義的多値データ
 - 順序的多値データ

SKIP

名義カテゴリモデル

\square ominal Categories \square odel (\square C \square)

$$\begin{cases} P_k(\theta|A_1, \dots, A_{K-1}, B_1, \dots, B_{K-1}) = \frac{\exp\{A_k(\theta - B_k)\}}{1 + \sum_{c=1}^{K-1} \exp\{A_c(\theta - B_c)\}} \\ (k = 1, \dots, K-1; -\infty < A_k, B_k < \infty) \\ P_K(\theta|A_1, \dots, A_{K-1}, B_1, \dots, B_{K-1}) = \frac{1}{1 + \sum_{c=1}^{K-1} \exp\{A_c(\theta - B_c)\}} \end{cases}$$

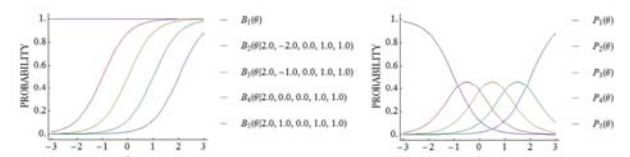


SKIP

段階反応モデル

Graded Response \square odel (GR \square)

$$\begin{cases} P_k(\theta) = B_k(\theta) - B_{k+1}(\theta) \\ B_1(\theta) = 1 \\ B_k(\theta|A, B, C, D, E) = C + \frac{D - C}{1 + [\exp\{-A(\theta - B)\}]^E} \\ B_{K+1}(\theta) = 0 \end{cases} \quad (k = 2, \dots, K)$$



潜在ランク理論 latent rank theory (LRT)

- ニューラルテスト理論(neural test theory, NTT)
- 段階評価を行うためのテスト理論
 - 自己組織化マップ(SOM)あるいは、生成トポグラフィックマッピング(GTM)のメカニズムを利用
 - 潜在尺度が**順序尺度**
- Shojima, K. (2009) Neural test theory.
 - K. Shigemasu et al. (Eds.) *New Trends in Psychometrics*, Universal Academy Press, Inc., pp.417-426.

なぜ順序尺度？

- 項目反応理論・古典的テスト理論は**連続尺度**
- 2つの主な理由
 - 方法論的理由
 - 教育社会学的理由

テストの精度 (accuracy)

- | | |
|--|---|
| <ul style="list-style-type: none"> 体重計で測定しました <ul style="list-style-type: none"> A₁君は73kgでした $f_W(A_1)=73$ $f_W(A_1) \neq 74$ $f_W(A_1) \neq 72$ | <ul style="list-style-type: none"> テストで測定しました <ul style="list-style-type: none"> B₁君は73点でした $f_T(B_1)=73$ $f_T(B_1) \neq 74$? $f_T(B_1) \neq 72$? |
|--|---|



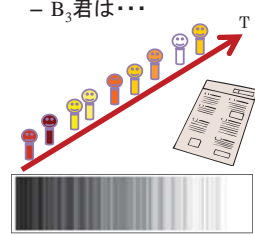
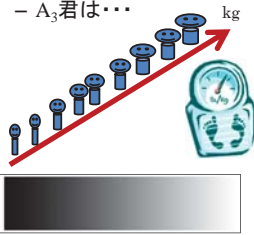
テストの識別力 (discrimination)

- | | |
|---|---|
| <ul style="list-style-type: none"> 体重計で測定しました <ul style="list-style-type: none"> A₁君は73kgでした A₂君は75kgでした $f_W(A_1) < f_W(A_2)$ | <ul style="list-style-type: none"> テストで測定しました <ul style="list-style-type: none"> B₁君は73点でした B₂君は75点でした $f_T(B_1) < f_T(B_2)$? |
|---|---|



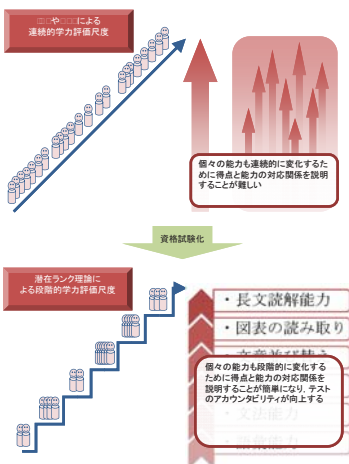
テストの解像度 (resolution)

- | | |
|--|--|
| <ul style="list-style-type: none"> 体重計で測定しました <ul style="list-style-type: none"> A₁君は73kgでした A₂君は75kgでした A₃君は... | <ul style="list-style-type: none"> テストで測定しました <ul style="list-style-type: none"> B₁君は73点でした B₂君は75点でした B₃君は... |
|--|--|



段階評価しよう

- テストは大事な社会の公具 (public tool)
- 高い精度の測定は困難
 - 測定内容の10%は誤差
- テストは、学力を5~20レベルに段階評価するくらいがせいぜい
 - 60点と61点とか、違いを説明できない評価をすることの意味は？



段階評価
↓
アカウンタビリティ
↓
資格試験

教育社会学的理由

- 連続尺度の負の側面
 - 生徒たちは、日々、一点でも高い得点をとるよう動機付けられている
 - 不安定な連続尺度の乱高下に一喜一憂させるべきではない
- 順序尺度の正の側面
 - 段階評価は、連続尺度上での評価よりも頑健
 - 継続して努力しないと上位ランクに進めない
- ただし、どちらのやり方にも正負の側面
 - 分析者のニーズに合わせてさまざまな方法論で世が満てばよい。

LRTにおける統計的学習

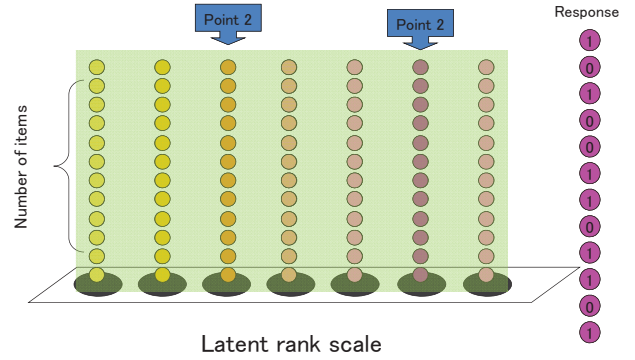
SOメカニズム

- 逐次学習型
- 標本サイズ10000以下
- 滑らかさ大

GTメカニズム

- 一括学習型
- 滑らかさ小
- 標本サイズ10000以上
- 潜在クラス分析byEアルゴリズム
- スムージング

LRTのメカニズム(SOMの場合)



Point 1: 勝者ランク選択

Likelihood

$$p(\mathbf{u}_h^{(t)} | \mathbf{V}^{(t,h-1)}) = \prod_{j=1}^n z_{hj}^{(t)} \left\{ u_{hj}^{(t)} \ln v_{qj}^{(t,h-1)} + (1 - u_{hj}^{(t)}) \ln(1 - v_{qj}^{(t,h-1)}) \right\}$$

ML $R_w^{(ML)} : w = \arg \max_{q \in Q} \ln p(\mathbf{u}_h^{(t)} | \mathbf{V}^{(t,h-1)})$

Bayes $R_w^{(MAP)} : w = \arg \max_{q \in Q} \{ \ln p(\mathbf{u}_h^{(t)} | \mathbf{V}^{(t,h-1)}) + \ln p(f_q) \}$

Point 2: ランク参照行列の更新

$$\mathbf{V}^{(t,h)} = \mathbf{V}^{(t,h-1)} + (\mathbf{1}_n \mathbf{h}^{(t)}) \odot (\mathbf{z}_h^{(t)} \mathbf{1}_Q) \odot (\mathbf{u}_h^{(t)} \mathbf{1}_Q - \mathbf{V}^{(t,h-1)})$$

$$\mathbf{h}^{(t)} = \{h_{qw}^{(t)}\} \quad (n \times 1)$$

$$h_{qw}^{(t)} = \frac{\alpha_t Q}{N} \exp \left\{ -\frac{(q-w)^2}{2Q^2 \sigma_t^2} \right\}$$

$$\alpha_t = \frac{(T-t)\alpha_1 + (t-1)\alpha_T}{T-1}$$

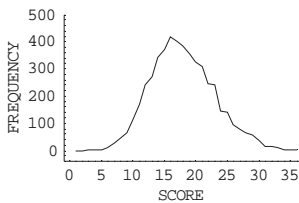
$$\sigma_t = \frac{(T-t)\sigma_1 + (t-1)\sigma_T}{T-1}$$

- The nodes of the ranks nearer to the winner are updated to become closer to the input data
- h : tension
- α : size of tension
- σ : region size of learning propagation

分析例

- 地理テスト

N	5000
n	35
Median	17
Max	35
Min	2
Range	33
Mean	16.911
Sd	4.976
Skew	0.313
Kurt	-0.074
Alpha	0.704



適合度指標

ML, Q=10

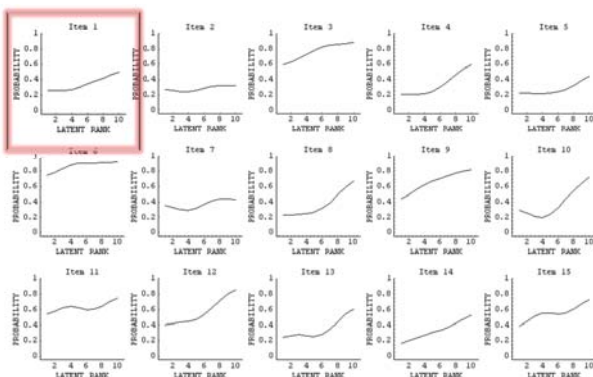
Index	Value
χ^2_{275}	1596.01
NFI	0.889
RFI	0.848
IFI	0.946
TLI	0.925
CFI	0.945
RMSEA	0.013
AIC	-153.99
CAIC	-6731.53
BIC	-5856.53

ML, Q=5

Index	Value
χ^2_{100}	2184.61
NFI	0.847
RFI	0.827
IFI	0.914
TLI	0.902
CFI	0.914
RMSEA	0.015
AIC	84.61
CAIC	-7808.44
BIC	-6758.44

- 潜在ランク数を決定するのに便利

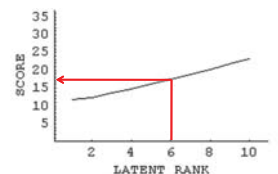
項目参照プロファイル



単純増加制約 (monotonic increasing constraint) を課することも可能

テスト参照プロファイル(TRP)

- IRPの重み付き和
- 各潜在ランクの期待値



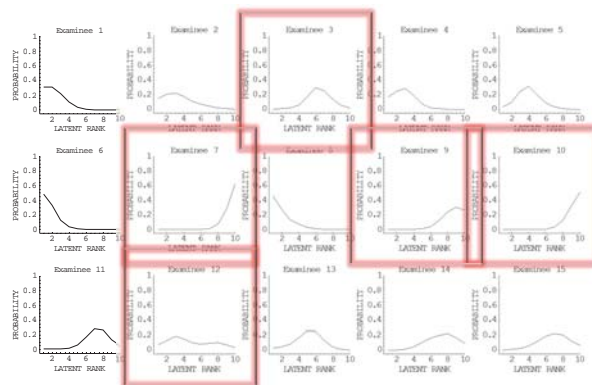
- 弱順序配置条件 (Weakly ordinal alignment condition)
 - TRPが単調増加だけど、全てのIRPが単調であるわけではない
- 強順序配置条件 (Strongly ordinal alignment condition)
 - 全てのIRPが単調増加 → TRPも単調増加
- 潜在尺度が順序尺度であるためには、少なくともWOACを満たす必要

ランク・メンバーシップ・プロファイル $\square\square M \square\square$

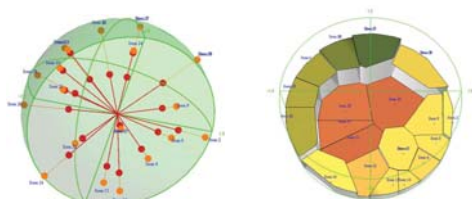
- 各受検者がどの潜在ランクに所属するかについての事後分布

$$\text{RMP} \quad p_{iq} = \frac{p(\mathbf{u}_i | \mathbf{v}_q) p(f_q)}{\sum_{q'=1}^Q p(\mathbf{u}_i | \mathbf{v}_{q'}) p(f_{q'})}$$

RMPの例



Can-Do Chart (例) Exametrikaデモ



非対称三角尺度法(ATRISCAL)
Asymmetric Triangulation Scaling

SKIP

多値モデル polytomous model

- 名義モデル(nominal model)
- 段階モデル(graded model)

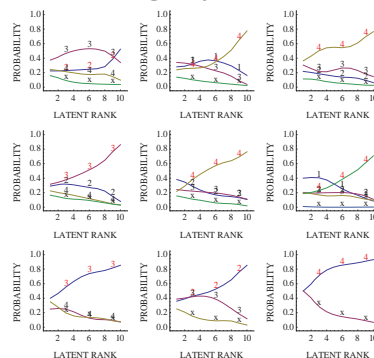
SKIP

名義潜在ランクモデル

- 名義カテゴリデータのためのLRTモデル
- 主要出力
 - 項目カテゴリ参照プロファイル
 - 項目参照プロファイル
 - テスト参照プロファイル
 - ランク・メンバーシップ・プロファイル
 - ランク・メンバーシップ分布
- 選択数が少ないカテゴリはマージしたほうがよい

SKIP

項目カテゴリ参照プロファイル Item Category Reference Profile



SKIP

その他の出力

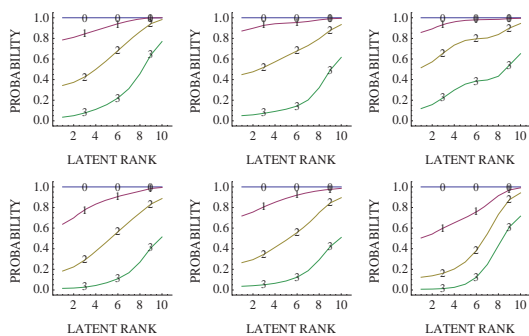
- 項目参照プロファイル(item reference profile)
 - 正答選択肢のICRPがIRP
- テスト参照プロファイル(test reference profile)
 - IRPの和(2値モデルと同じ)
 - 少なくとも弱順序配置条件を満たすこと
- Rank membership profile
 - 2値モデルと同じ(各人の各ランクへの所属確率)
- Rank membership distribution
 - 受検者全体がどのランクに分けられたかの分布

SKIP

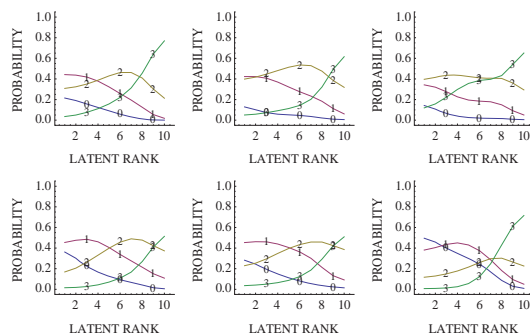
段階潜在ランクモデル

- 段階(リッカート)データのためのLRTモデル
- 主要出力
 - 境界カテゴリ参照プロファイル
 - 項目カテゴリ参照プロファイル
 - 項目参照プロファイル
 - テスト参照プロファイル
 - ランク・メンバーシップ・プロファイル
 - ランク・メンバーシップ分布

SKIP 境界カテゴリ参照プロファイル
Boundary Category Reference Profile



SKIP 項目カテゴリ参照プロファイル
Item Category Reference Profile



SKIP その他の出力

- 項目参照プロファイル(item reference profile)
 - BCRPの和
 - 各ランクに対する項目得点の期待値
- テスト参照プロファイル(test reference profile)
 - IRPの和(2値モデルと同じ)
 - 少なくとも弱順序配置条件を満たすこと
- Rank membership profile
 - 2値モデルと同じ(各人の各ランクへの所属確率)
- Rank membership distribution
 - 受検者全体がどのランクに分けられたかの分布

SKIP Exametrikaの仕様
(潜在ランク理論)

SKIP Exametrikaの推定の設定①

- LRT-GTM
 - generative topographic mappingのメカニズム
 - 一括学習型(計算時間が早い)
 - 標本サイズが3000以上(目安)
- LRT-SOM
 - Self-organizing mapのメカニズムを利用
 - 逐次学習型(計算時間がかかる)
 - 標本サイズが3000以下(目安)
 - **GTMのメカニズムより滑らかなIRPが得られる**
- 潜在ランク数(Number of latent ranks)
 - 項目数やNが小さいときはランク数も小さく

SKIP Exametrikaの推定の設定②

- 単調増加制約(monotonicity constraint)
 - IRP(名義モデルでは正答選択肢のICRP、段階モデルではBCRP)を強制的に単調増加にしたいとき
- 事前分布(prior distribution)
 - 正答数の高さが高ランクに強く反映されるように
 - Ver. 5.3でデフォルトにしちゃったけど外してよいかも
- 目標潜在ランク分布(target latent rank dist.)
 - 様分布、正規分布が選べる
 - 事前分布をチェックすると指定できない

SKIP Exametrikaで等化

- 2値モデル
 - 固定したいIRPのシートを用意
- 多値モデル
 - 固定したいICRPのシートを用意
- 項目ごとにコンプリートしている必要はない
 - ところどころ抜けていてもよい
 - ランク数を増減させて等化したいときにも使える

SKIP 適合度指標(絶対指標)

- Chi-square (df, p -value)
- NFI (normed fit index): [0, 1] 1.0が良い適合
- RFI (relative fit index): [0, 1] 1.0が良い適合
- IFI (incremental fit index): [0, 1] 1.0が良
- TLI (Tucker-Lewis index): [0, 1] 1.0が良い適合
- CFI (comparative fit index): [0, 1] 1.0が良
- RMSEA (root mean square error of approximation): [0, ∞] 0.0が良い適合

SKIP

適合度指標 (相対指標)

- AIC (Akaike information criterion)
- CAIC (consistent AIC)
- BIC (Bayesian information criterion)
- 複数のモデルを比較するとき利用
 - LRTではランク数の異なるモデルを比較するとき
- モデルの効率の良さを表す
- 小さいDFで高い適合を示すモデルを高く評価
- AICはランク数の多いモデルを採用しやすい
- CAICかBICで判断するほうが無難

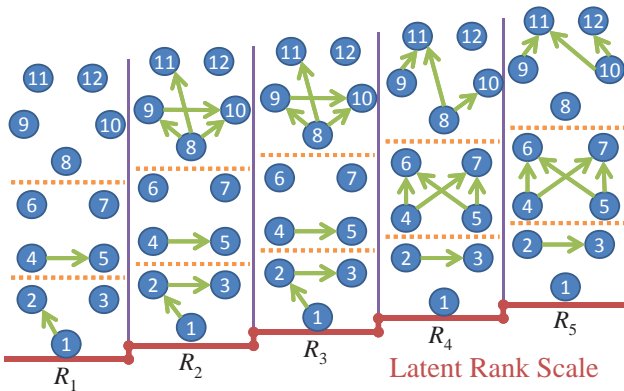
SKIP

局所従属潜在ランク理論

- 局所従属を扱えるLRTモデルの開発
 - 潜在ランク理論+ベイジアンネットワーク
 - ランクごとに異なる従属関係構造を考察可能
 - モデル同定は難しいがモデリングの楽しさ
- 主に数学や理科が主眼
 - 語学(文法構造)や歴史(因果構造)など文系科目でもよく構成された項目間の分析は可能
- Shojima (2011) 応用統計学, 印刷中

SKIP

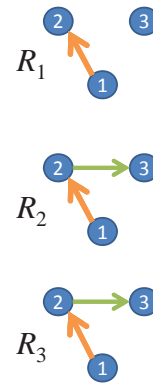
学力段階ごとに従属関係をモデリング



SKIP

CI's with One PI**

Rank	CI	PI	Response of PI	
			0	1
1	2	1	0.029	0.431
1	5	4	0.096	0.231
2	2	1	0.033	0.561
2	3	2	0.019	0.456
2	5	4	0.142	0.346
2	9	8	0.257	0.727
2	11	9	0.034	0.400
3	2	1	0.038	0.730
3	3	2	0.030	0.623
3	5	4	0.223	0.562
3	9	8	0.322	0.847
3	11	8	0.053	0.640
4	3	2	0.057	0.721
4	10	8	0.148	0.850
5	3	2	0.099	0.796
5	12	10	0.154	0.348



Item 1	Item 2
0	0.029
1	0.431

Item 1	Item 2
0	0.033
1	0.561

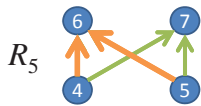
Item 1	Item 2
0	0.038
1	0.730

SKIP



Item 4-5	Item 6
0-0	0.005
0-1	0.101
1-0	0.020
1-1	0.697

Rank		CI		Response Pattern of PIs			
				PI1	PI2	00	01
2	10	8	9	0.078	0.127	0.152	0.742
3	10	8	9	0.088	0.156	0.205	0.846
4	6	4	5	0.005	0.101	0.020	0.697
4	7	4	5	0.007	0.026	0.036	0.555
4	11	8	9	0.059	0.123	0.097	0.831
5	6	4	5	0.008	0.123	0.038	0.799
5	7	4	5	0.009	0.028	0.063	0.662
5	11	9	10	0.078	0.174	0.222	0.919



Item 4-5	Item 6
0-0	0.008
0-1	0.123
1-0	0.038
1-1	0.799