

# 潜在ランク理論 はじめの一步と二歩

莊島宏二郎

大学入試センター 研究開発部

shoima@rd.dnc.ac.jp

# 潜在ランク理論の着想の経緯

- 2002年10月(26歳)大学入試センター助教に着任
  - しばらく現行のシステムなどに矛盾を感じず
- 2006年くらい(30歳前後)のとき、1点刻みの学力評価に意味があるのか、と疑いだす
  - たとえば、セ試英語は200点満点
  - 120点と121点の違いに差はある?(説明できる?)

# 精度(accuracy)

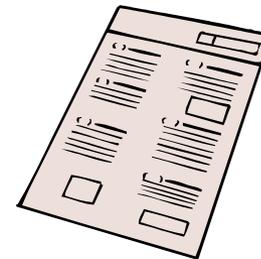
## • 体重計

- $A_1$  君は73 kgだった
- $f_W(A_1)=73$
- $f_W(A_1) \neq 74$
- $f_W(A_1) \neq 72$



## • 学力テスト

- $B_1$  君は73点だった
- $f_T(B_1)=73$
- $f_T(B_1) \neq 74 ?$
- $f_T(B_1) \neq 72 ?$



# 識別力(discrimination)

- 体重計

- $A_1$  君は73 kgだった
- $A_2$  さんは75 kgだった

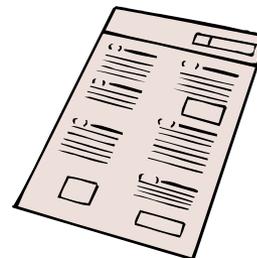
- $f_W(A_1) < f_W(A_2)$



- 学力テスト

- $B_1$  君は73点だった
- $B_2$  さんは75点だった

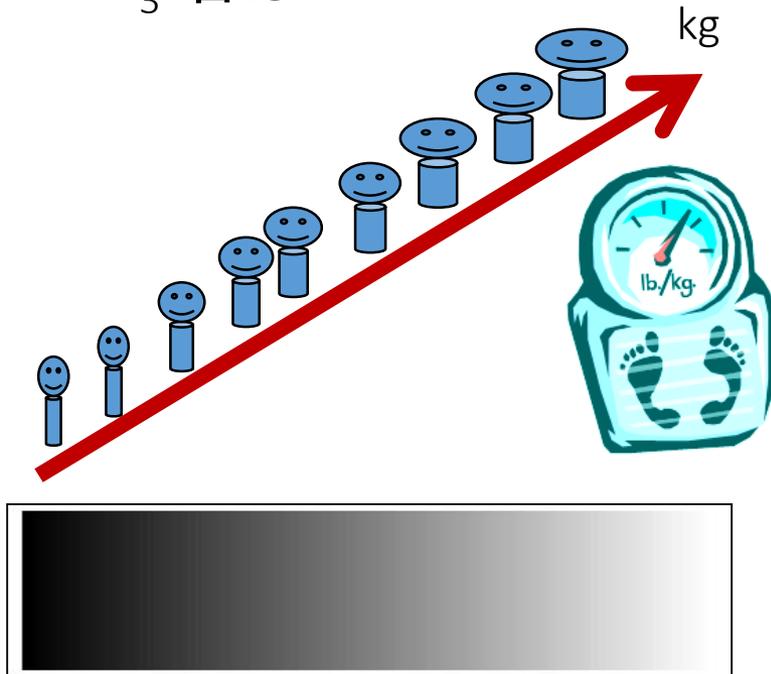
- $f_T(B_1) < f_T(B_2)$  ?



# 解像度(resolution)

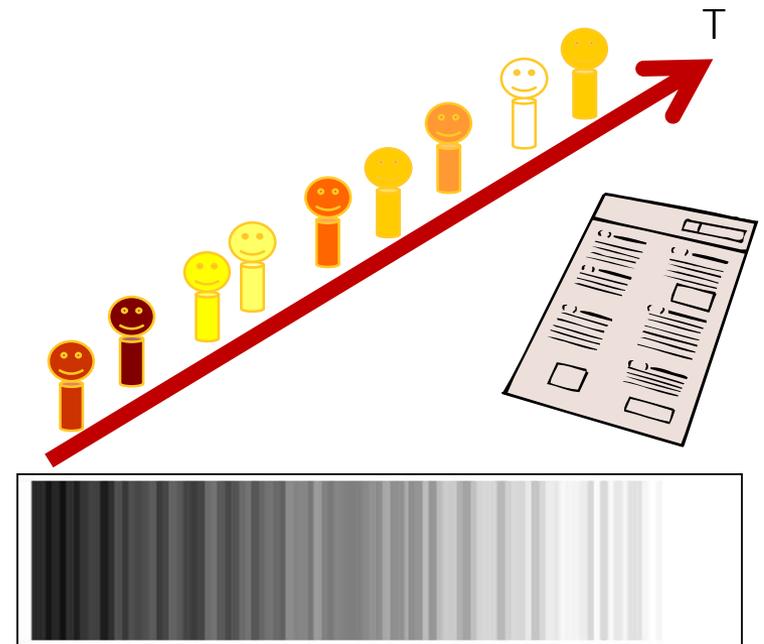
## • 体重計

- $A_1$  君は73 kgだった
- $A_2$  さんは75 kgだった
- $A_3$  君は...



## • 学力テスト

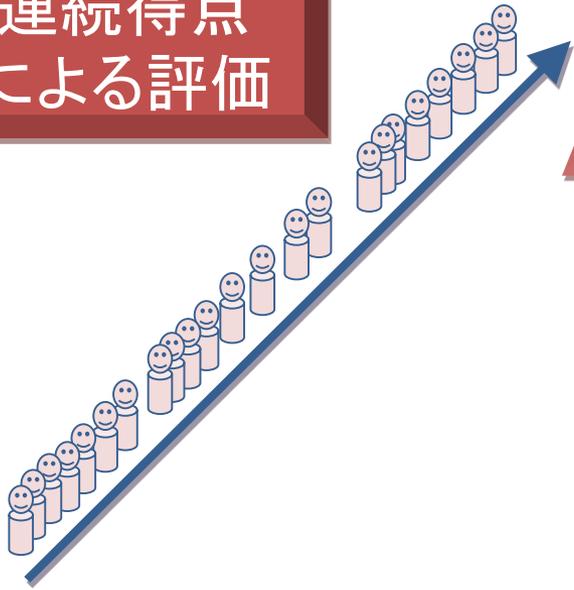
- $B_1$  君は73点だった
- $B_2$  さんは75点だった
- $B_3$  君は...



# 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
  - CTTの最大の貢献？
- 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

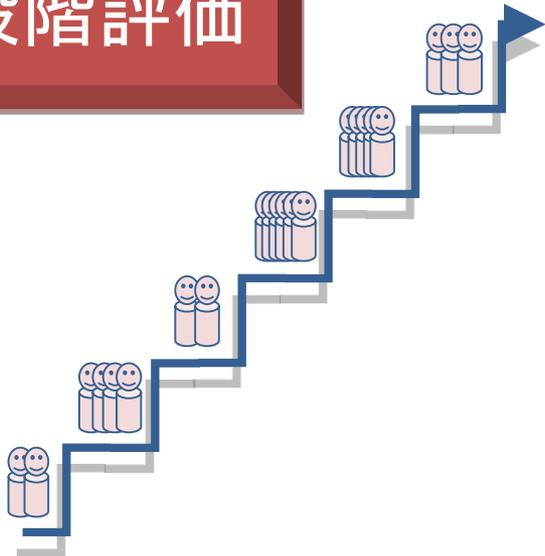
## 連続得点 による評価



It is difficult to explain the relationship between scores and abilities because individual abilities also change continuously

For Qualifying Tests

## 段階評価



- Ability to understand long sentences
- Ability to read charts and diagrams
- Ability to rearrange sentences

Because the individual abilities also change in stages, it is easy to explain the relationship between scores and abilities. This increases the test's accountability.

- Vocabulary skills

段階評価



アカウントビリティ



資格試験

# 展開

- 様々な分野で応用
  - 英語教育、日本語教育、心理学、情報、医療、看護、作業療法、福祉、保育、体育、マーケティング
  - 論文、書籍、学会発表など、博士論文も
- 業者テスト
  - 河合塾PROG
  - 奈良市テスト
  - 進行中のものも

# 教育社会学的理由

- 連続尺度の負の側面
  - 学生は日々一点でも高い得点をとるよう動機付けられる
  - 不安定な連続尺度の乱高下に一喜一憂させるのか
- 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**
  - 競争的場面では連続尺度が便利
  - 診断的場面では段階尺度で十分

# 潜在ランク理論

- 単に素点を分けるのではなく  
得点パターンごとに受検者を分類し順序づける
- 項目ごとの特徴を評価
  - 項目参照プロフィール
- 受験者ごとの特徴を評価
  - ランク・メンバーシップ・プロフィール
- 自己組織化マップあるいは生成トポグラフィック  
マッピングのメカニズムを利用

# LRTにおける統計的学習

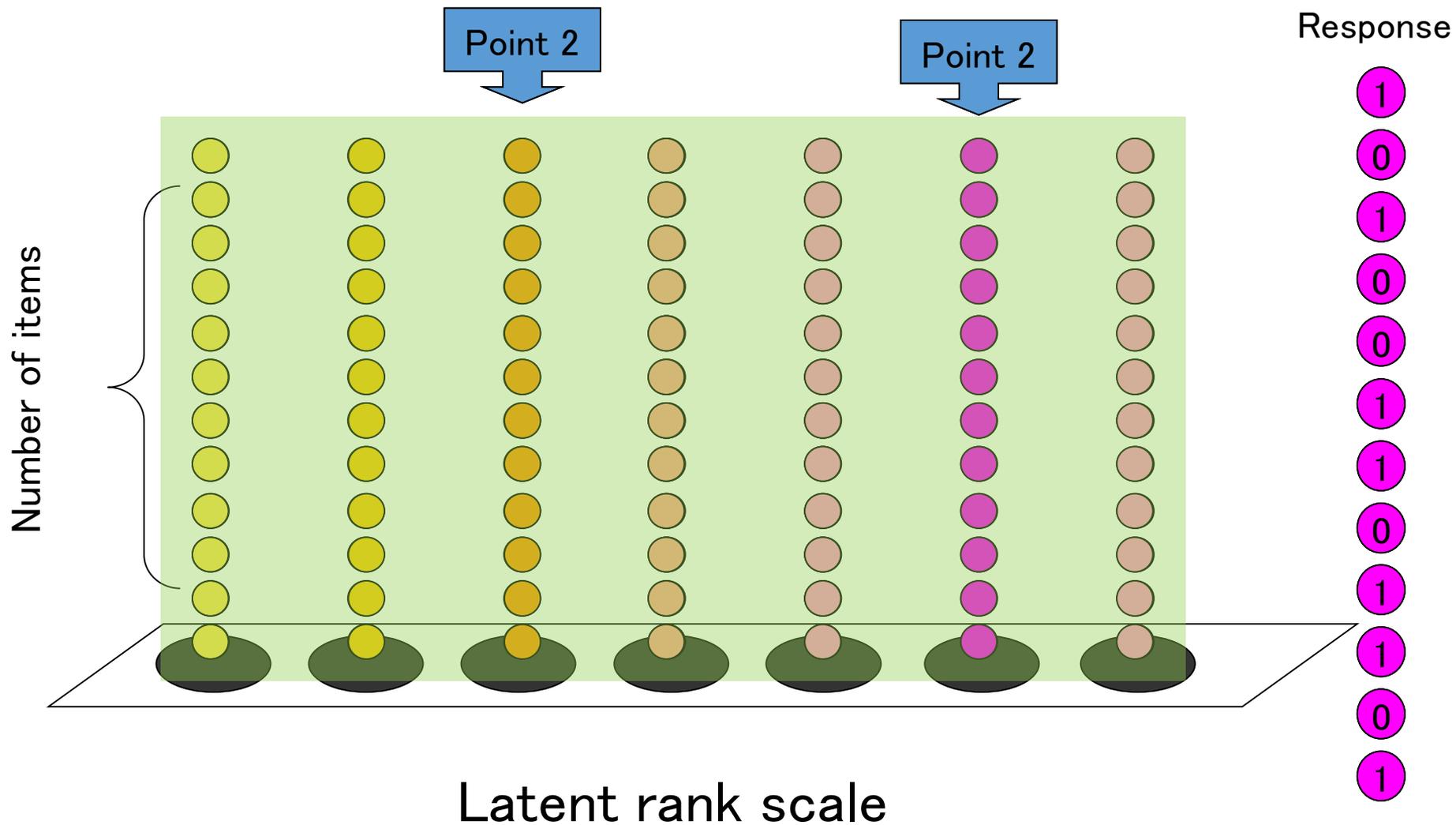
## • SOMメカニズム

- 逐次学習型
- 標本サイズ3000くらいまで(目安)
- 滑らかさ大

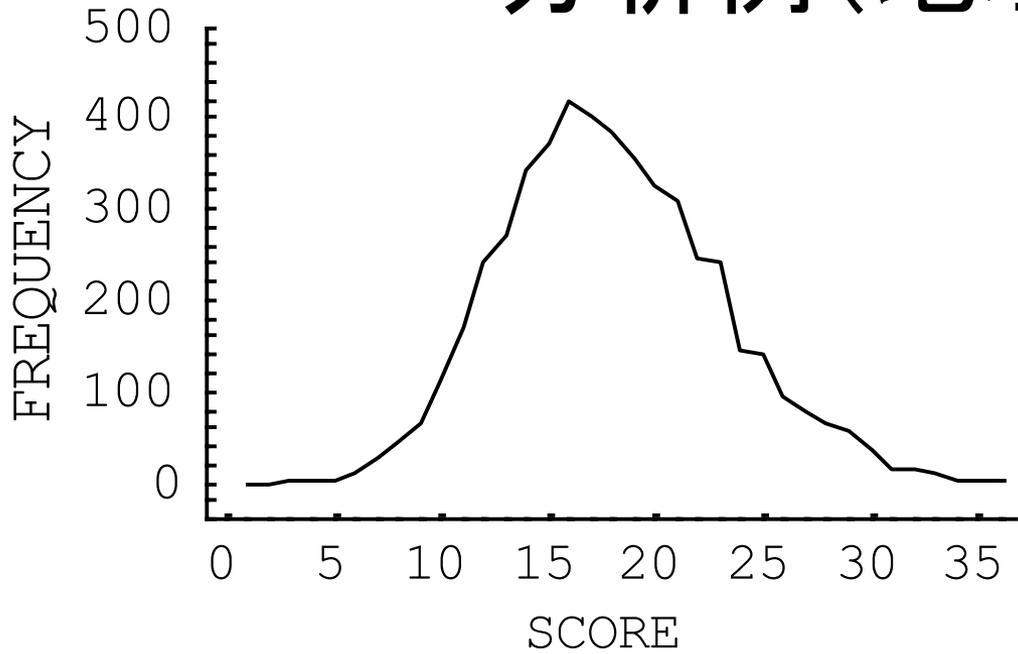
## • GTMメカニズム

- 一括学習型
- 滑らかさ小
- 標本サイズ3000以上(目安)
- 潜在クラス分析byEMアルゴリズム
- スムージング

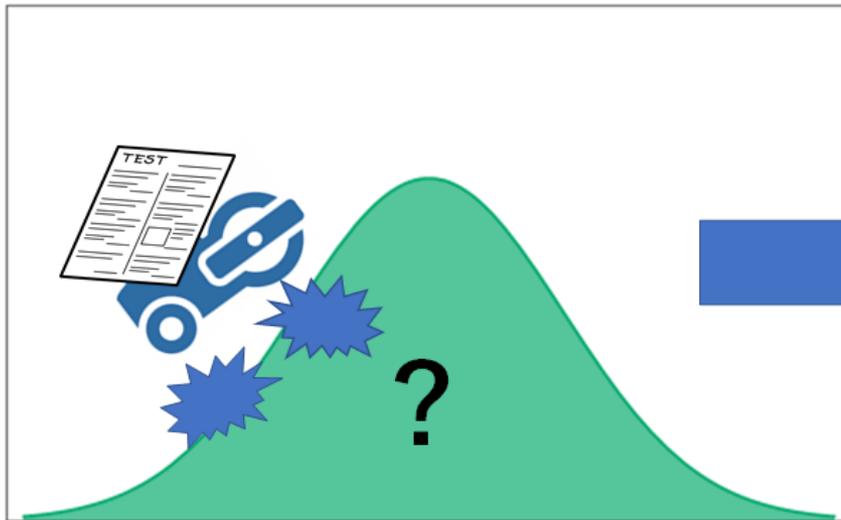
# LRTのメカニズム(SOMの場合)



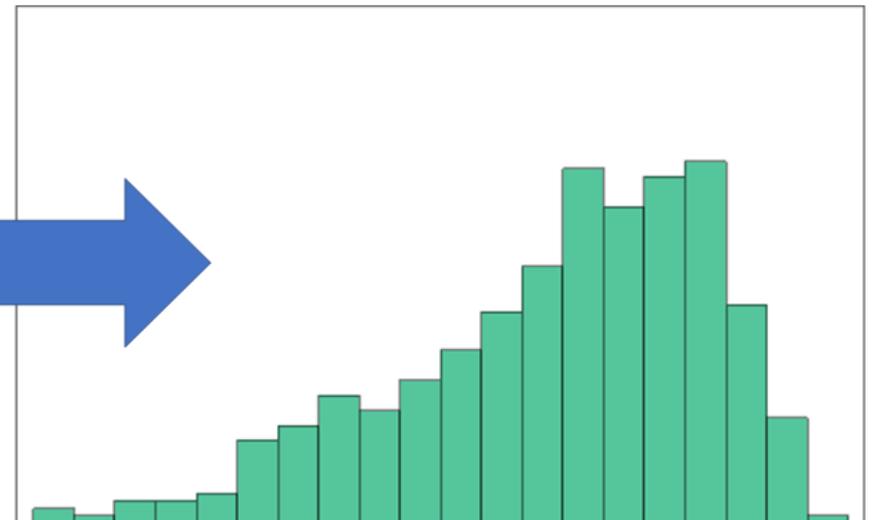
# 分析例(地理)



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



Ability Distribution



Score Distribution

# 適合度指標

ML, Q=10

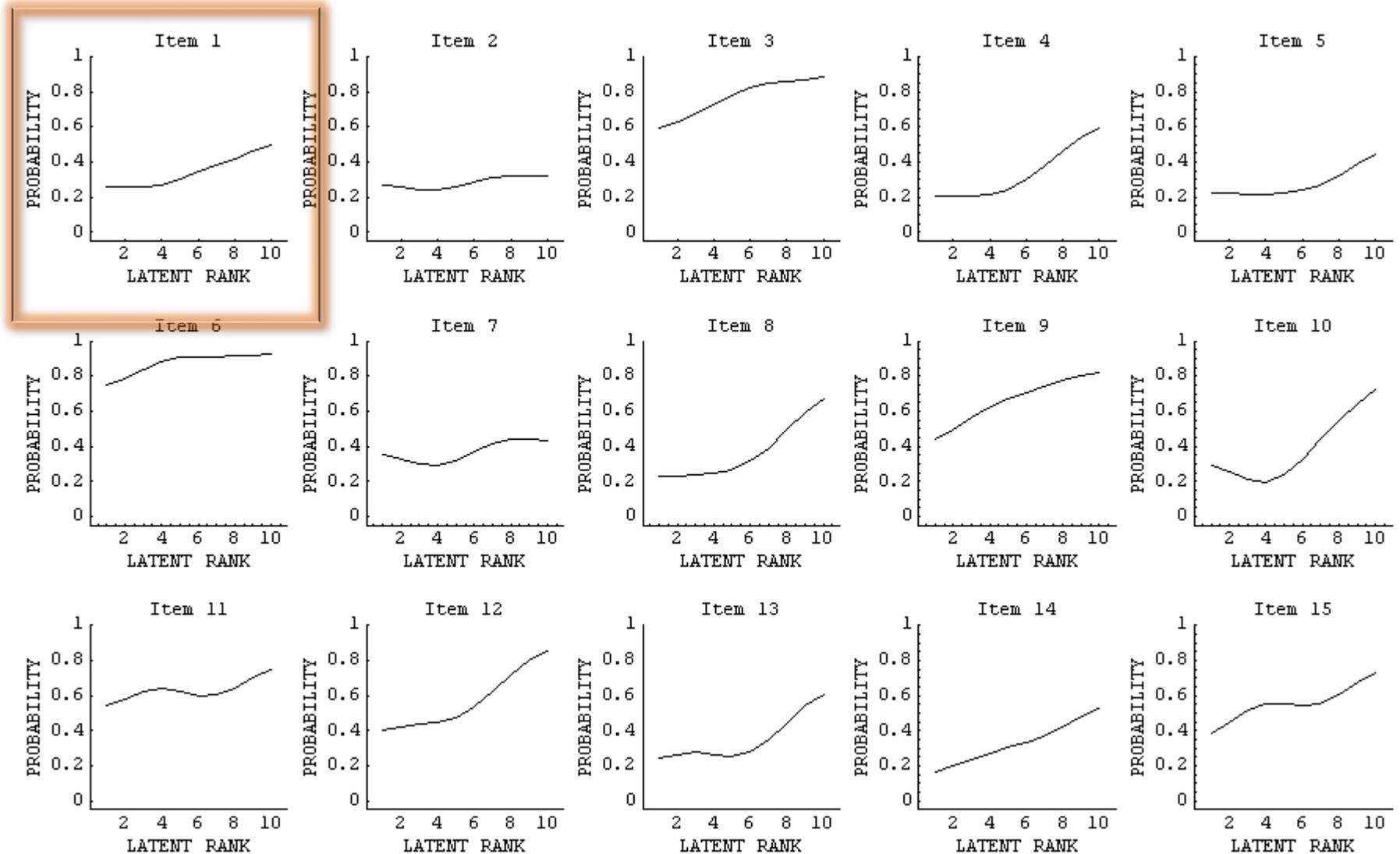
Index	Value
$\chi^2_{875}$	1596.01
NFI	0.889
RFI	0.848
IFI	0.946
TLI	0.925
CFI	0.945
RMSEA	0.013
AIC	-153.99
CAIC	-6731.53
BIC	-5856.53

ML, Q=5

Index	Value
$\chi^2_{1050}$	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

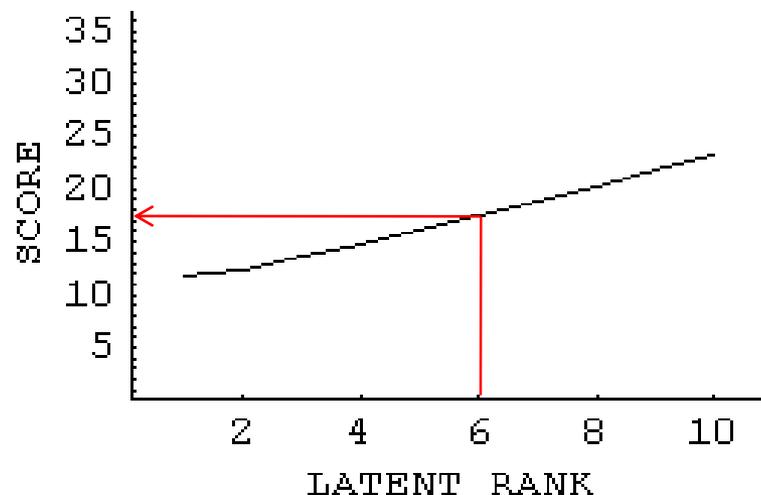
# 項目参照プロフィール



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

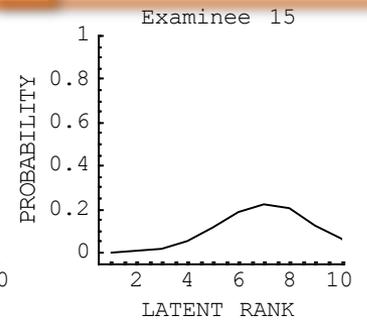
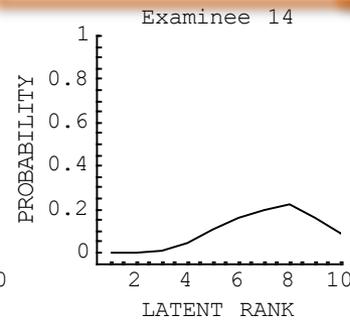
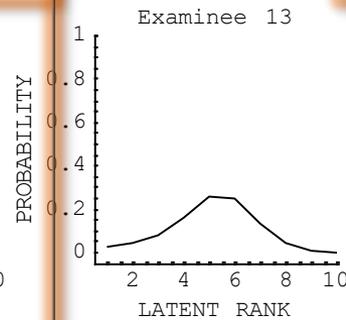
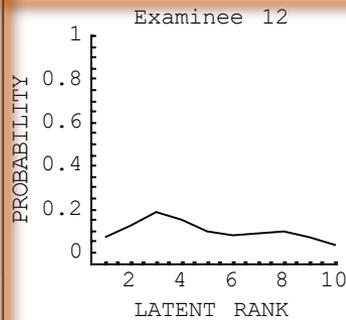
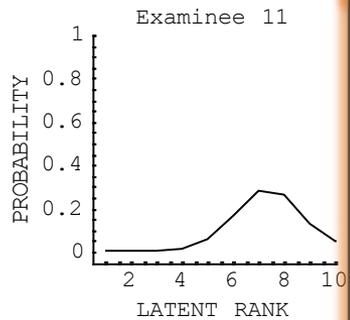
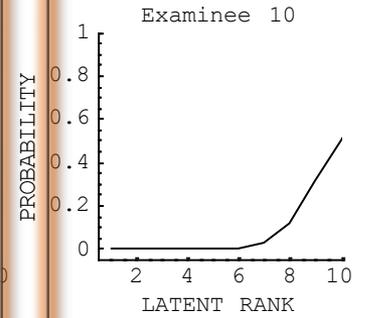
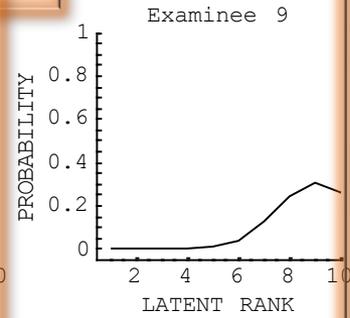
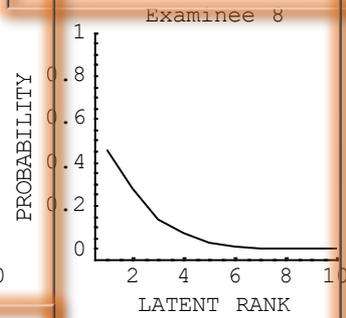
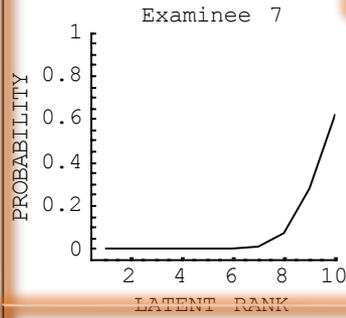
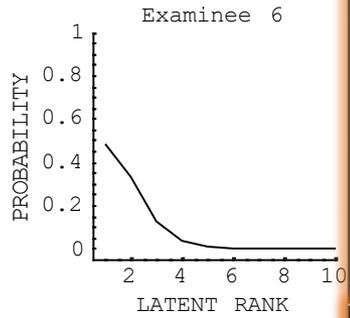
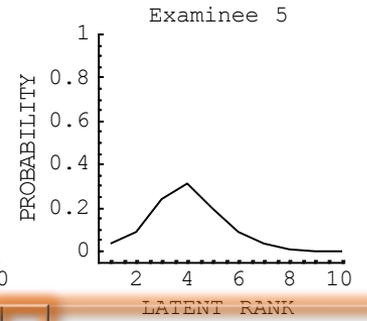
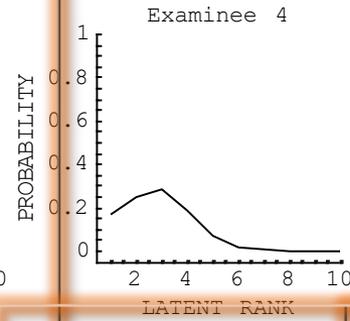
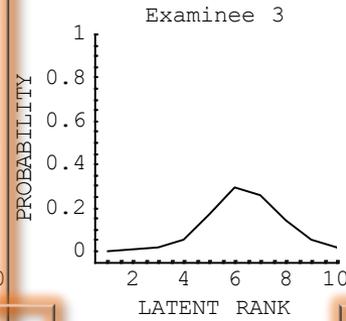
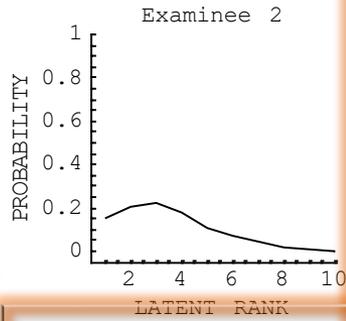
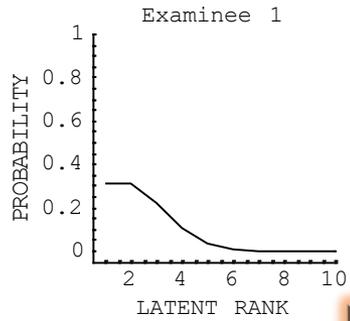
# テスト参照プロフィール(TRP)

- ・ 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WOACが必要

# RMPの例



## 潜在ランク分布

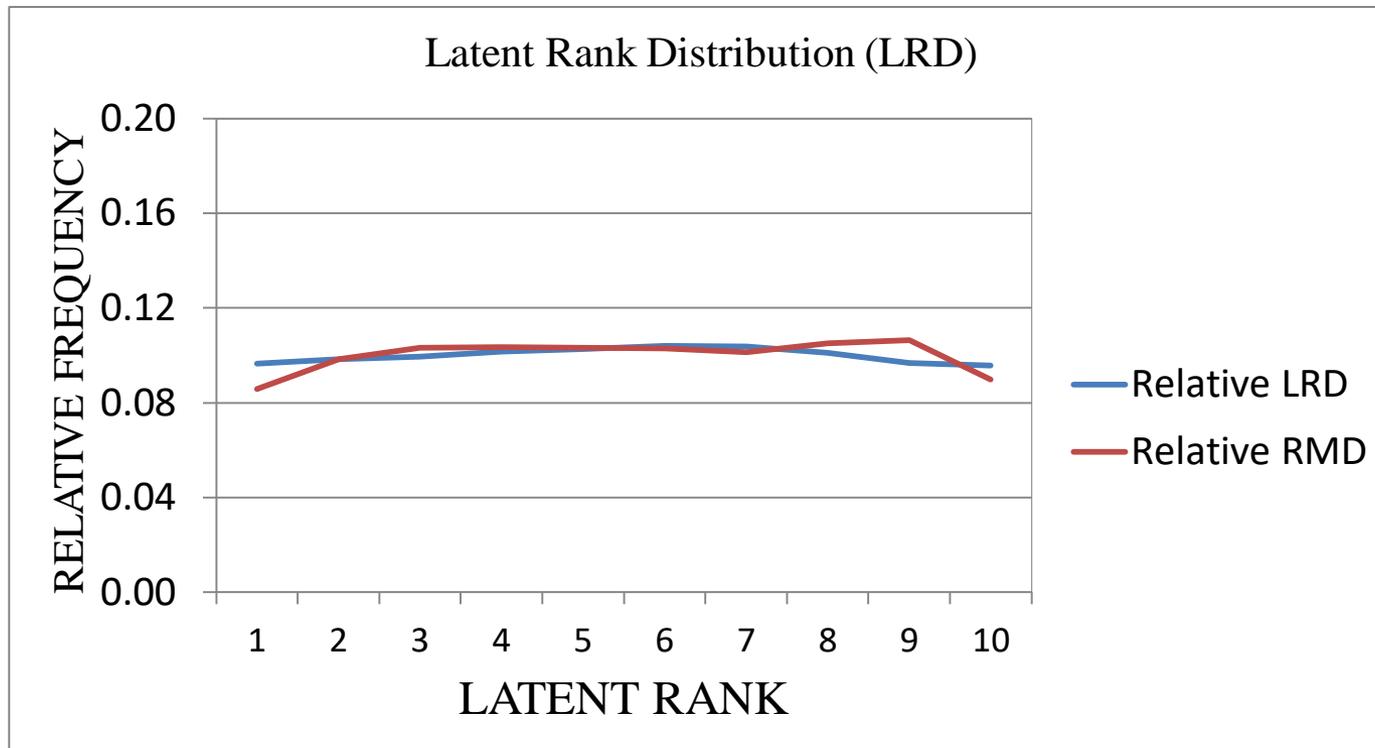
### Latent Rank Distribution (LRD)

- 受検者全体(sample)の潜在ランクの推定値の分布

## ランク・メンバーシップ分布

### Rank Membership Dist. (RMD)

- 受検者全体(population)の潜在ランク分布
- 受検者RMPの総和(平均)



## エグザメトリカ

exametrika

[English Site](#)

[トップ](#)

[潜在ランク理論](#)

[項目反応理論](#)

[非対称三角尺度法](#)

[カテゴリカルデータ解析](#)

[ダウンロード](#)

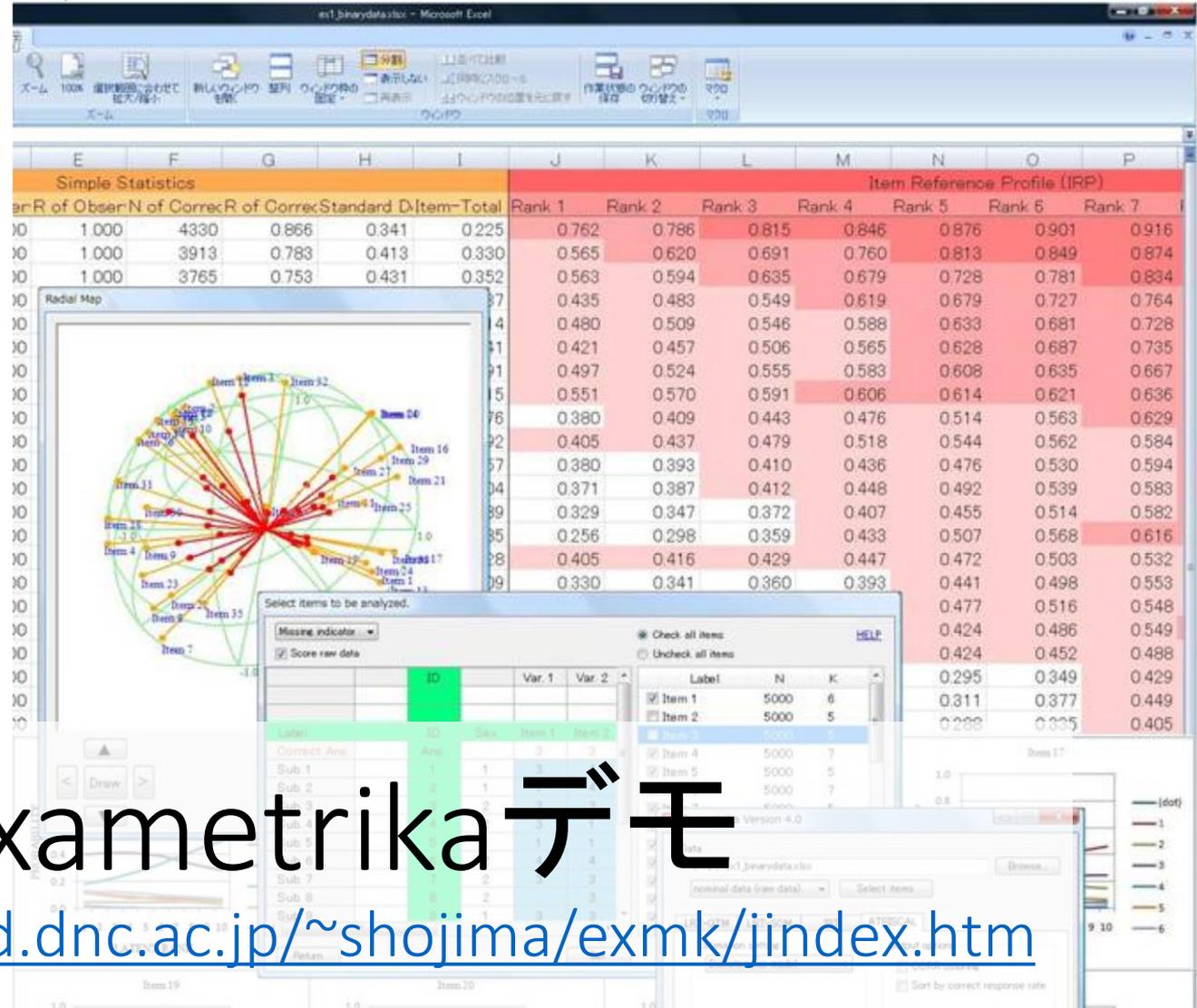
[ヘルプ](#)

[リンク](#)

[shojima@rd.dnc.ac.jp](mailto:shojima@rd.dnc.ac.jp)

## エグザメトリカ

エグザメトリカは、潜在ランク理論(ニューラルテスト理論)や項目反応理論, 非対称三角尺度法のためのソフトウェアです。



# exametrika デモ

<http://www.rd.dnc.ac.jp/~shojima/exmk/jindex.htm>

# 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

# 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
- 目標潜在ランク分布(target latent rank dist.)
  - 一様分布、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

# 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) :  $[0, \infty]$  0.0が良い適合

# 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に利用
    - 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

# Can-Do Chart (例)

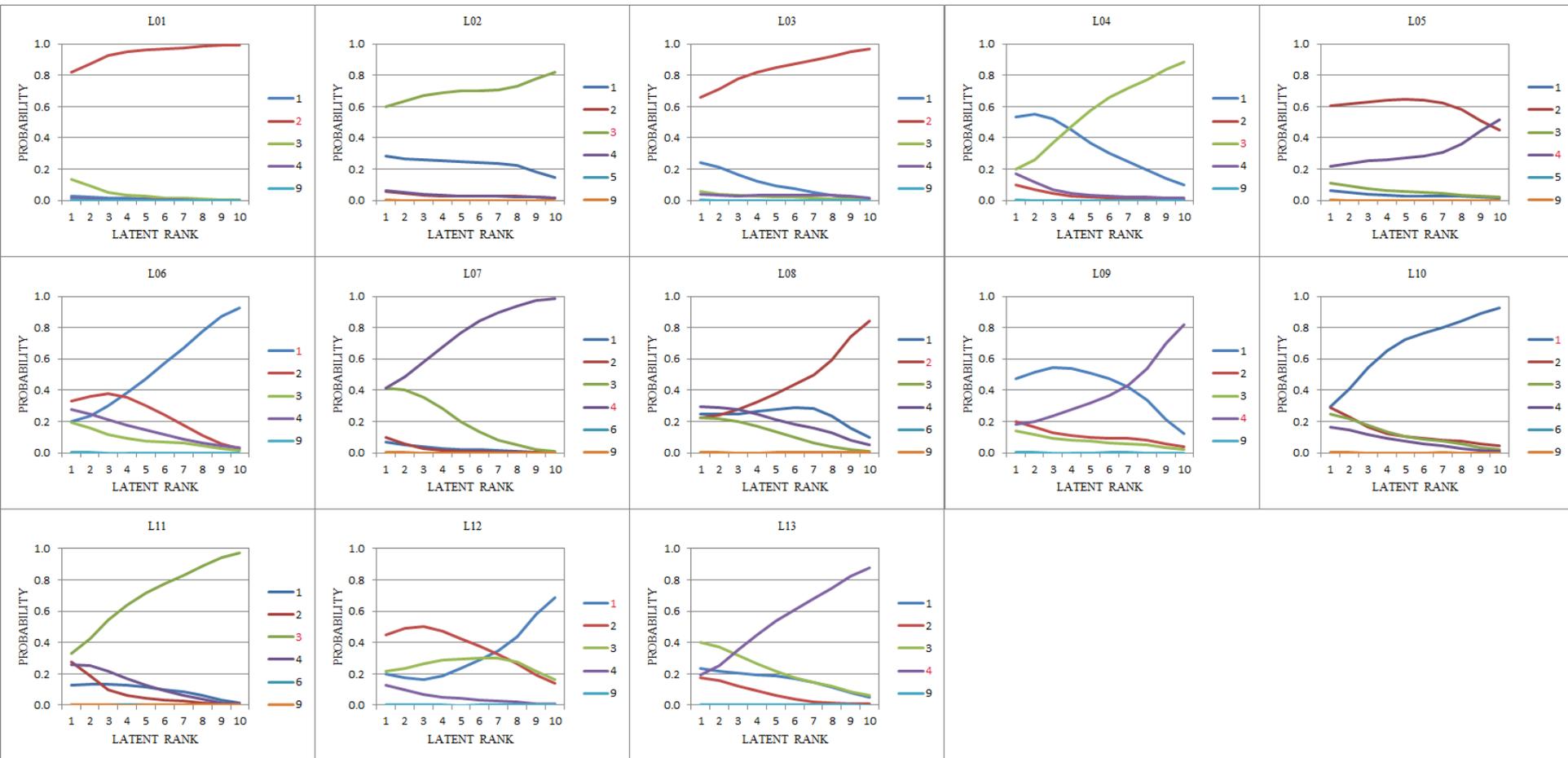


# 潜在ランク理論

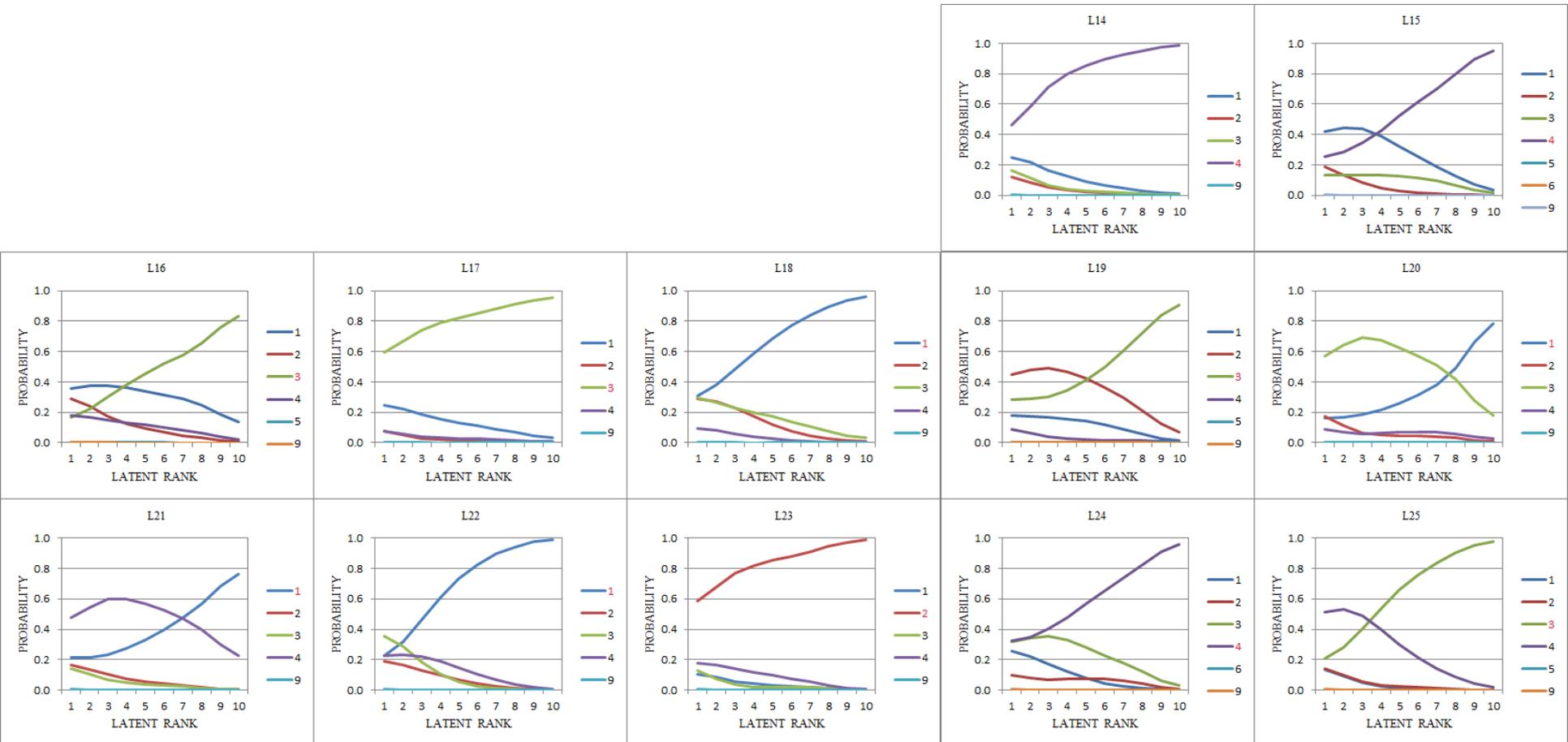
latent rank theory (LRT)

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

# Item Category Reference Profiles

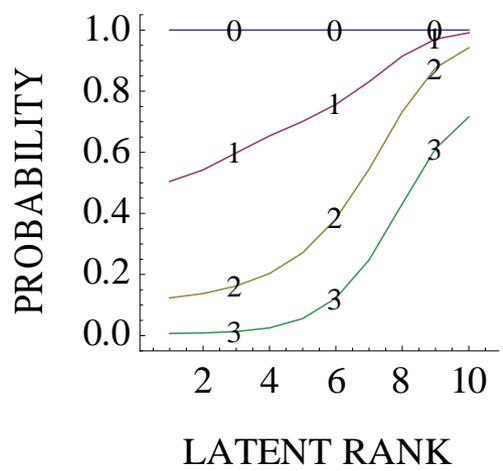
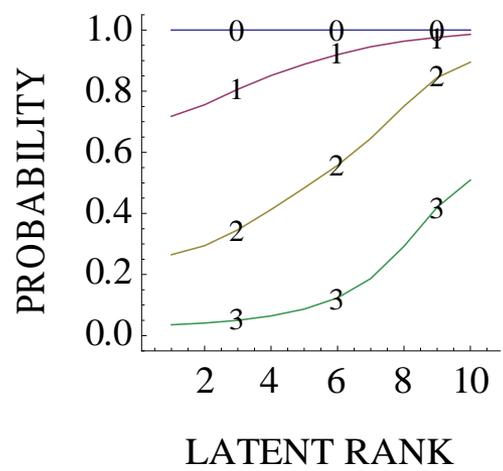
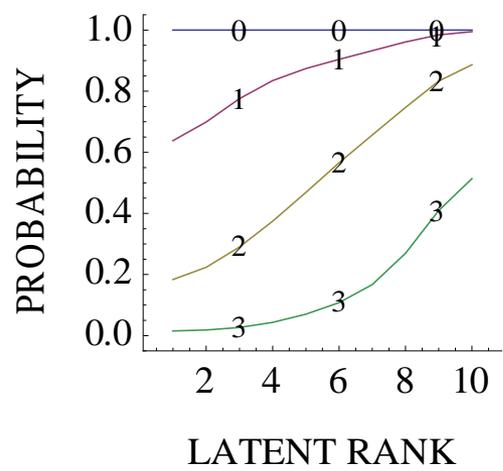
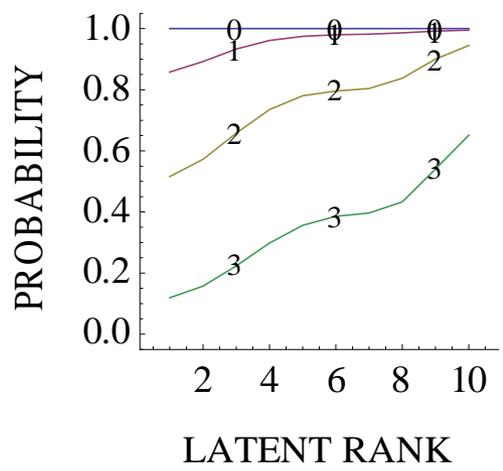
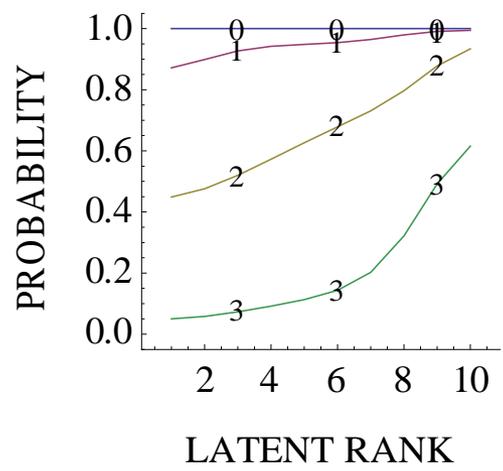
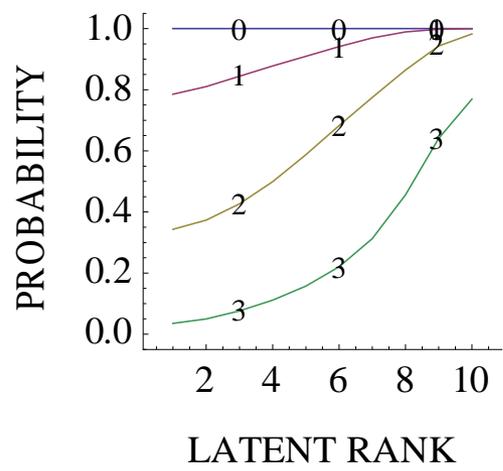


## Item Category Reference Profiles



# 境界カテゴリ参照プロファイル

## Boundary Category Reference Profile



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

