# 認知診断モデルの基本の基

山口一大
筑波大学人間系
Mail: yamaguchi.kazuhir.ft@u.tsukuba.ac.jp

2022年3月5日

日本テスト学会 第15回記念講演ワークショップ
＠アルカディア市ヶ谷（私学会館）

# 本日の内容・目標

- 内容

1. 認知診断モデルの基礎
2. Rによる実データ解析
3. 発展的な話題

- 目標
  - 認知診断モデルの仕組みを理解する。
  - 認知診断モデルをRで実行できるようになる。
  - （推定方法の詳細などは，興味深い話題もあるが，今回は扱わない）

# 認知診断モデル（Cognitive diagnostic models, CDMs）

- CDM：複数の認知要素を仮定して学習者の到達度を調べるための教育測定モデル (e.g., Leignton & Gierl, 2007)
  - 認知要素：アトリビュート
  - Diagnostic classification models (DCMs)とも呼ばれる
- **CDMの主目的：学習者のアトリビュートプロフィールを推定すること**
  - アトリビュートプロフィール：アトリビュートの習得・未習得の組み合わせ
    - アトリビュート習得パタン，アトリビュートパタンなどとも呼ばれる
- → 学習者の認知要素の強み・弱みを発見できる

# CDMの主要な構成要素

- アトリビュートプロフィールを推定するため必要な要素
  1. アトリビュートの集合
     - テストの認知理論や文献研究に基づいて決定される
  2. Q-matrix (Tatsuoka, 1983)
     - アトリビュートと項目の関連を示した行列
  3. 項目反応関数（item response function, IRF）
     - 測定モデル (measurement model)
     - アトリビュートの認知理論に依存して決定される
  4. 構造モデル (structural model)
     - アトリビュートの間の関係のモデル
  5. 解答データ
     - 実際の項目反応のデータ

# 認知診断モデルの模式図

Q行列
（要設定，1：解答に必要，0：解答に不要）

| 問題項目 | アトリビュート | | |
|---|---|---|---|
| | 文字式の計算 | 一次方程式の計算 | 連立方程式の計算 |
| a + 3a | 1 | 0 | 0 |
| 2a + 3 = 0 | 1 | 1 | 0 |
| 2a + b = 2<br>a - b = 1 | 1 | 1 | 1 |

解答データ（1：正答，0：誤答）

| 解答者 | 問題項目 | | |
|---|---|---|---|
| | a + 3a | 2a + 3 = 0 | 2a + b = 2<br>a - b = 1 |
| Aさん | 1 | 0 | 0 |
| Bくん | 0 | 0 | 0 |
| Cくん | 1 | 1 | 1 |
| Dさん | 1 | 1 | 0 |

推定

測定モデル

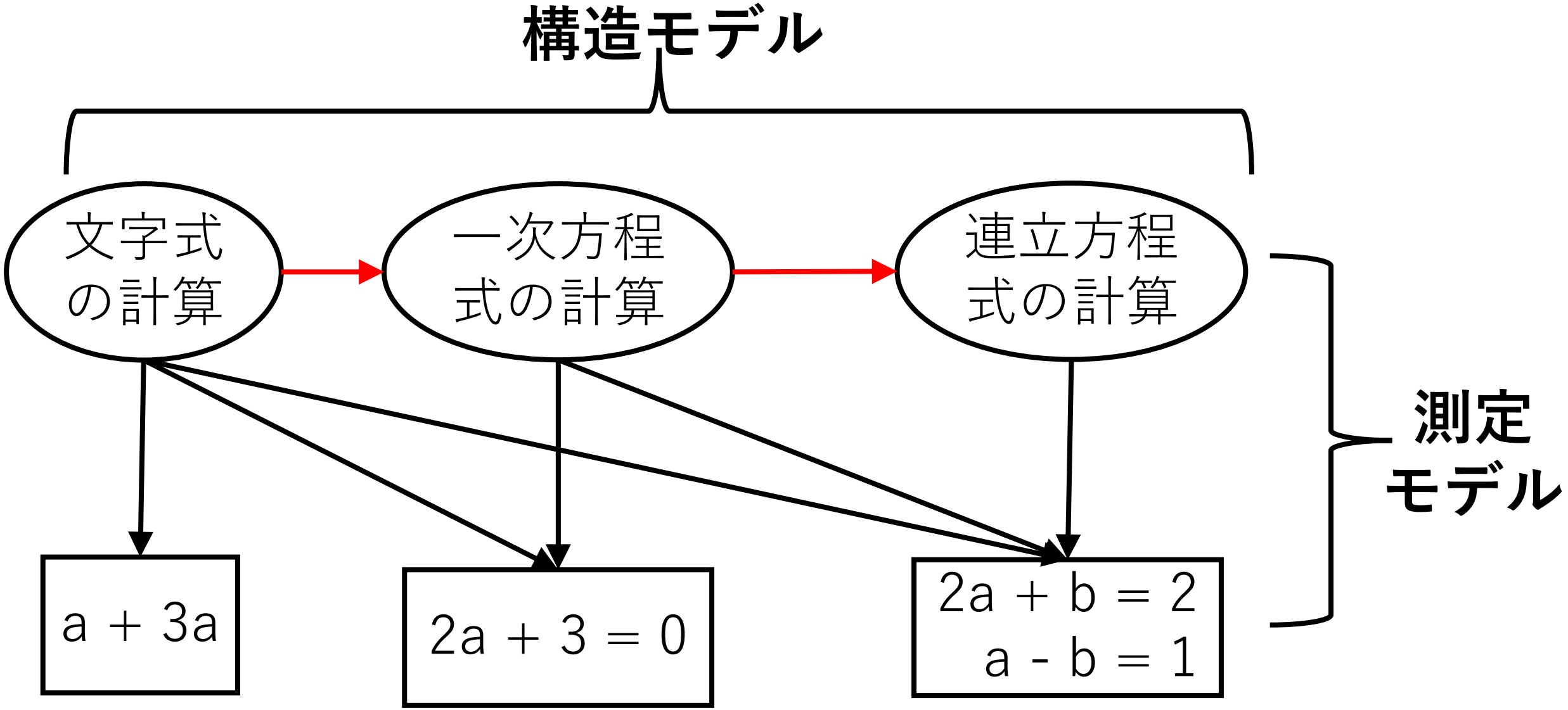
構造モデル

項目パラメタ

アトリビュート習得パタン
（1:習得，0：未習得）

| 解答者 | アトリビュート | | |
|---|---|---|---|
| | 文字式の計算 | 一次方程式の計算 | 連立方程式の計算 |
| Aさん | 1 | 0 | 0 |
| Bくん | 0 | 0 | 0 |
| Cくん | 1 | 1 | 1 |
| Dさん | 1 | 1 | 0 |

# 測定モデルと構造モデルの模式図



# CDMの発展の系譜

- Templin (2016, in Fritz ed; Technology and Testing)の図に詳しい。
- クラスタリングが今は主流だが，いくつかの流れがある。
  - 数理心理，項目反応理論，etc…
  - 山口（2018,博論）の序文などでもIRTモデルからの発展の流れについて整理してある。

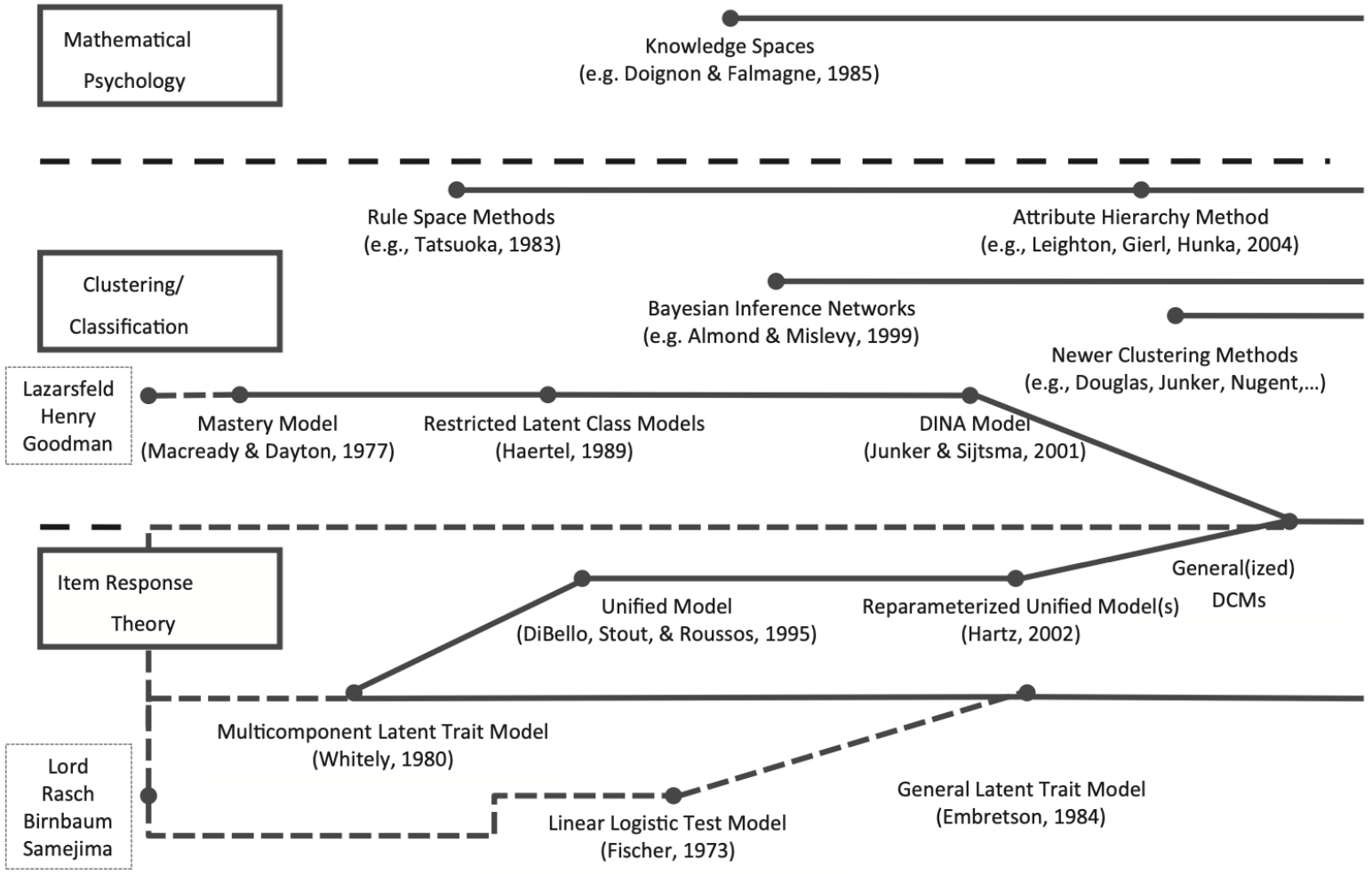



**Figure 13.1** Lineage of modern DCMs

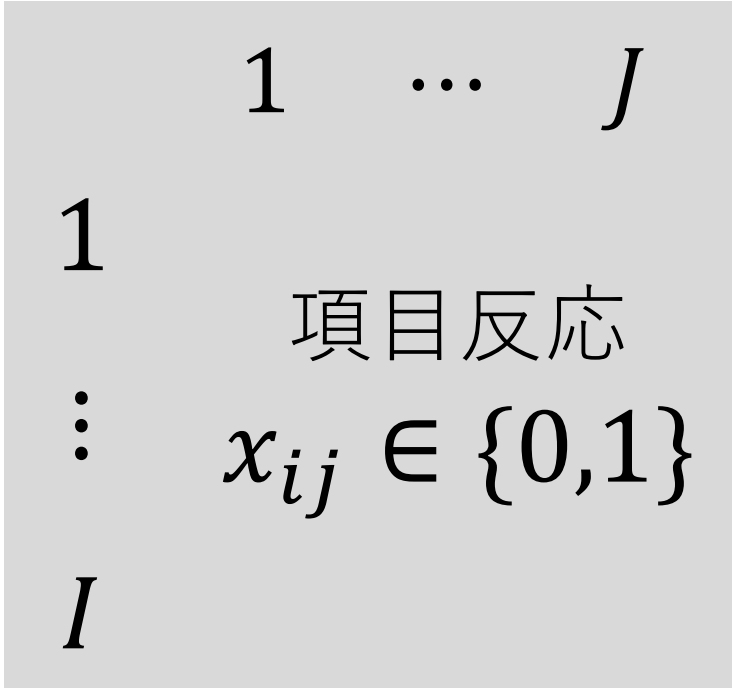
# CDMの統計モデルとしての実際

- CDM＝制約付き潜在クラスモデル
  - 「正答・誤答」に対応する従属変数のみならず，「ある能力の習得・非習得」のように説明変数にもカテゴリカル変数を仮定するモデル
  - Rupp & Templin（2008, p.226）の定義から
  - Restricted Latent Class model (RCL)などと呼ばれることもある
- 潜在クラスごとに観測変数への反応確率が異なる。
  - Q行列（Tatsuoka, 1983）によって，アトリビュート習得パタンごとの反応確率を定義できる
- さらに，単調制約（monotonicity constraint; Henson, Templin, & Willse, 2009; Yamaguchi & Templin, 2021)も重要な要素
  - あるアトリビュート習得パタンに対して，追加でアトリビュートを習得する場合のほうが正答確率が高くなる，という制約。
  - アトリビュート習得パタンの意味づけやラベルの反転を防ぐ意味で大切。
  - Q行列の識別の上でも重要。
  - 実際には，部分的な制約でも似たような推定結果が得られることも多い。

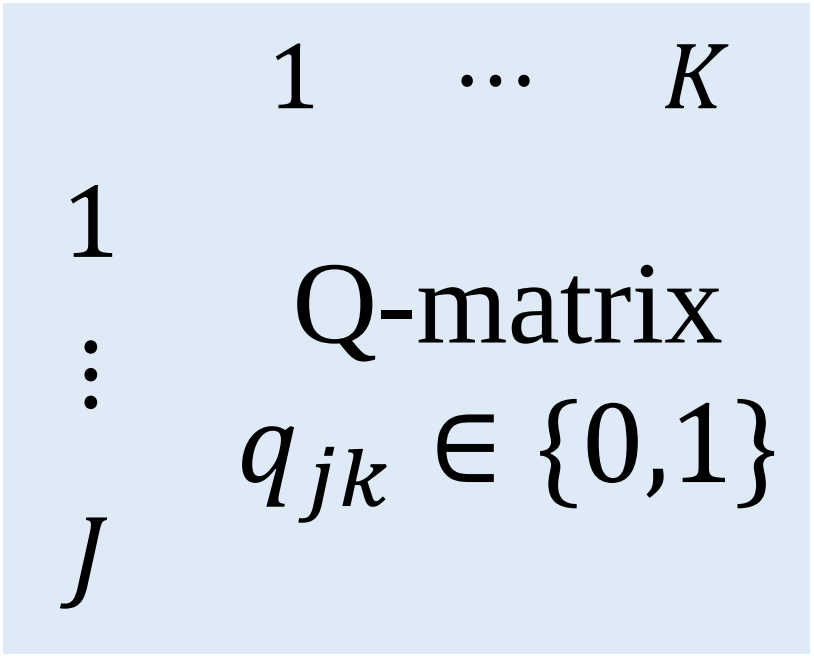
# もっとも素朴なDCM：DINAモデル

- Deterministic input noisy “and” gate model
  - 名前の由来は後ほど説明
- 基本的な発想：ある項目に正答するためには，その項目に必要な能力をすべて習得していないといけないだろう。
- 例：方程式「2a + 3 = 0」を解くためには，「文字式の計算」と「一次方程式の計算」という２つのアトリビュートを**どちらも**習得している必要がある。
- 言い換え：「文字式の計算」と「一次方程式の計算」を習得している場合には，「2a + 3 = 0」に正答できる確率は，そうでない場合よりも高くなる。
  - 一つでもアトリビュートを習得していない場合には，正答確率は低いはず。
  - １つ足りないのも全部足りないのも同じくらい低くなるのではないか。

# DINAモデルの定式化：記号の準備




- $x_{ij} = 1$: 解答者 $i$ が項目$j$に正答する。
- $x_{ij} = 0$: 誤答する
- 確率変数$X_{ij} \in \{0,1\}$の実現値




- $q_{jk} = 1$: アトリビュート$k$が項目$j$に必要
- $q_{jk} = 0$: 不要

# DINAモデルの定式化：記号の準備

| | $1$ | $\cdots$ | $L$ |
|---|---|---|---|
| $1$ | | | |
| $\vdots$ | | | |
| $I$ | | | |

クラス所属インディケータ行列

$$z_{il} \in \{0,1\}$$

| | $1$ | $\cdots$ | $K$ |
|---|---|---|---|
| $1$ | | | |
| $\vdots$ | | | |
| $L$ | | | |

アトリビュート習得パタン行列

$$\alpha_{lk} \in \{0,1\}$$

- $\mathbf{z}_i = (z_{i1}, \ldots, z_{iL})$:個人$i$がどのアトリビュート習得パタンであるかを示した指示ベクトル
- $\sum_l z_{il} = 1$で，要素のうち1つだけが 1
- $\alpha_{lk} = 1$: $l$番目のアトリビュート習得パタンにおいてアトリビュート $k$を習得している
- $\alpha_{lk} = 0$: 未習得

# DINAモデルの設定

- $z_{il}$：個人$i$があるアトリビュートパタン$l$であれば1，そうでなければ0を取る潜在変数ベクトル$\boldsymbol{z}_i = (z_{i1}, \cdots, z_{iL})^{\mathsf{T}}$の要素
  - $\sum_l z_{il} = 1$である。
- $\alpha_{lk}$：$l$番目のアトリビュートパタン$\boldsymbol{\alpha}_l = (\alpha_{l1}, \cdots, \alpha_{lK})^{\mathsf{T}}$の$k(= 1, \cdots, K)$番目のアトリビュートの習得の有無（1であれば習得，0であれば未習得）を表す
  - $q_{jk}$は項目$j$に$k$番目のアトリビュートが必要であれば1，不要ならば0を示すQ行列の要素である。ただし$0^0 = 1$とする。
- 理想反応（ideal response）$\eta_{ij} = \boldsymbol{\eta}_j^{\mathsf{T}} \boldsymbol{z}_i = \sum_l \eta_{lj} z_{il}$：（誤差がないとして）個人$i$が項目$j$へ正答できるか否かを表す
  - ただし，$\eta_{lj} = \prod_{k=1}^{K} \alpha_{lk}^{q_{jk}}$
  - $\eta_{lj}$をまとめたベクトル$\boldsymbol{\eta}_j = (\eta_{1j}, \cdots, \eta_{Lj})^{\mathsf{T}}$

# アトリビュートが2個の場合の例

アトリビュート
パタン行列

| パタン番号 | アトリビュート | | $\boldsymbol{\alpha}_l^{\mathsf{T}}$ |
|---|---|---|---|
| | 1 | 2 | |
| 1 | 0 | 0 | $\boldsymbol{\alpha}_1^{\mathsf{T}}$ |
| 2 | 1 | 0 | $\boldsymbol{\alpha}_2^{\mathsf{T}}$ |
| 3 | 0 | 1 | $\boldsymbol{\alpha}_3^{\mathsf{T}}$ |
| 4 | 1 | 1 | $\boldsymbol{\alpha}_4^{\mathsf{T}}$ |

$Z$行列

| 解答者番号 | アトリビュートパタン番号 | | | | $\boldsymbol{z}_i^{\mathsf{T}}$ |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| 1 | 1 | 0 | 0 | 0 | $\boldsymbol{z}_1^{\mathsf{T}}$ |
| 2 | 0 | 0 | 1 | 0 | $\boldsymbol{z}_2^{\mathsf{T}}$ |
| 3 | 0 | 1 | 0 | 0 | $\boldsymbol{z}_3^{\mathsf{T}}$ |
| 4 | 0 | 0 | 0 | 1 | $\boldsymbol{z}_4^{\mathsf{T}}$ |

解答者1は，アトリビュートパタン1であることを意味している。

# 理想反応行列の生成

アトリビュートパタン行列

| パタン番号 | アトリビュート | | $\boldsymbol{\alpha}_l^{\mathrm{T}}$ |
|---|---|---|---|
| | 1 | 2 | |
| 1 | 0 | 0 | $\boldsymbol{\alpha}_1^{\mathrm{T}}$ |
| 2 | 1 | 0 | $\boldsymbol{\alpha}_2^{\mathrm{T}}$ |
| 3 | 0 | 1 | $\boldsymbol{\alpha}_3^{\mathrm{T}}$ |
| 4 | 1 | 1 | $\boldsymbol{\alpha}_4^{\mathrm{T}}$ |

Q行列

| 項目番号 | アトリビュート | | $\boldsymbol{q}_j^{\mathrm{T}}$ |
|---|---|---|---|
| | 1 | 2 | |
| 1 | 1 | 0 | $\boldsymbol{q}_1^{\mathrm{T}}$ |
| 2 | 0 | 1 | $\boldsymbol{q}_2^{\mathrm{T}}$ |
| 3 | 1 | 1 | $\boldsymbol{q}_3^{\mathrm{T}}$ |

$$\eta_{lj} = \prod_{k=1}^{K} \alpha_{lk}^{q_{jk}}$$

理想反応行列

| アトリビュートパタン番号 | 項目番号 | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 |
| 3 | 0 | 1 | 0 |
| 4 | 1 | 1 | 1 |
| $\boldsymbol{\eta}_j$ | $\boldsymbol{\eta}_1$ | $\boldsymbol{\eta}_2$ | $\boldsymbol{\eta}_3$ |

★どのアトリビュートパタンが，項目の正答に必要なアトリビュートをすべて習得しているパタンなのかを示している。

# 理想反応行列の生成

アトリビュートパタン行列

| パタン番号 | アトリビュート | | $\boldsymbol{\alpha}_l^{\mathrm{T}}$ |
|---|---|---|---|
| | 1 | 2 | |
| 1 | 0 | 0 | $\boldsymbol{\alpha}_1^{\mathrm{T}}$ |
| 2 | 1 | 0 | $\boldsymbol{\alpha}_2^{\mathrm{T}}$ |
| 3 | 0 | 1 | $\boldsymbol{\alpha}_3^{\mathrm{T}}$ |
| 4 | 1 | 1 | $\boldsymbol{\alpha}_4^{\mathrm{T}}$ |

Q行列

| 項目番号 | アトリビュート | | $\boldsymbol{q}_j^{\mathrm{T}}$ |
|---|---|---|---|
| | 1 | 2 | |
| 1 | 1 | 0 | $\boldsymbol{q}_1^{\mathrm{T}}$ |
| 2 | 0 | 1 | $\boldsymbol{q}_2^{\mathrm{T}}$ |
| 3 | 1 | 1 | $\boldsymbol{q}_3^{\mathrm{T}}$ |

$$\eta_{lj} = \prod_{k=1}^{K} \alpha_{lk}^{q_{jk}}$$

理想反応行列

| アトリビュートパタン番号 | 項目番号 | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 |
| 3 | 0 | 1 | 0 |
| 4 | 1 | 1 | 1 |
| $\boldsymbol{\eta}_j$ | $\boldsymbol{\eta}_1$ | $\boldsymbol{\eta}_2$ | $\boldsymbol{\eta}_3$ |

★どのアトリビュートパタンが，項目の正答に必要なアトリビュートをすべて習得しているパタンなのかを示している。

# $\eta_{ij}$の計算

理想反応行列

| アトリビュートパタン番号 | 項目番号 | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 |
| 3 | 0 | 1 | 0 |
| 4 | 1 | 1 | 1 |
| $\boldsymbol{\eta}_j$ | $\boldsymbol{\eta}_1$ | $\boldsymbol{\eta}_2$ | $\boldsymbol{\eta}_3$ |

$Z$行列

| 解答者番号 | アトリビュートパタン番号 | | | | $\boldsymbol{z}_i^{\top}$ |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | |
| 1 | 1 | 0 | 0 | 0 | $\boldsymbol{z}_1^{\top}$ |
| 2 | 0 | 0 | 1 | 0 | $\boldsymbol{z}_2^{\top}$ |
| 3 | 0 | 1 | 0 | 0 | $\boldsymbol{z}_3^{\top}$ |
| 4 | 0 | 0 | 0 | 1 | $\boldsymbol{z}_4^{\top}$ |

$$\eta_{ij} = \boldsymbol{\eta}_j^{\top}\boldsymbol{z}_i = \sum_l \eta_{lj} z_{il}$$

個人ごとの理想反応

$\eta_{ij}$

| 解答者番号 | 項目番号 | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 |
| 3 | 1 | 0 | 0 |
| 4 | 1 | 1 | 1 |

# DINAモデルの項目反応関数

- 項目パラメタ
  - slipパラメタ：$s_j = \Pr(x_{ij} = 0 | \eta_{ij} = 1)$
    - $s_j$は項目に正答に必要な能力を全て習得している個人が誤答する確率
  - guessingパラメタ：$g_j = \Pr(x_{ij} = 1 | \eta_{ij} = 0)$
    - $g_j$は正答に必要な能力が欠けている個人が偶然問題に正答する確率
  - 単調性制約：$0 \leq g_j \leq 1 - s_j \leq 1$
    - 直感的説明：項目の正答に必要な能力を習得している場合には，当て推量より高い正答確率になる。
- 項目反応関数（item response function: IRF）
  - $\mathrm{P}(X_{ij} = 1 | s_j, g_j, \boldsymbol{z}_i, \boldsymbol{q}_j, \boldsymbol{\alpha}_l) = (1 - s_j)^{\boldsymbol{\eta}_j^{\top} \boldsymbol{z}_i} g_j^{1 - \boldsymbol{\eta}_j^{\top} \boldsymbol{z}_i}$
  - $\rightarrow \mathrm{P}(X_{ij} = x_{ij} | s_j, g_j, \boldsymbol{z}_i, \boldsymbol{q}_j, \boldsymbol{\alpha}_l) = \left( (1 - s_j)^{\boldsymbol{\eta}_j^{\top} \boldsymbol{z}_i} g_j^{1 - \boldsymbol{\eta}_j^{\top} \boldsymbol{z}_i} \right)^{x_{ij}} \left( s_j^{\boldsymbol{\eta}_j^{\top} \boldsymbol{z}_i} (1 - g_j)^{1 - \boldsymbol{\eta}_j^{\top} \boldsymbol{z}_i} \right)^{1 - x_{ij}}$
  - ★項目について，２つのクラス（理想反応が0か1か）を想定した潜在クラスモデルになっている

# 項目反応確率の表

$\eta_{ij}$

| 解答者番号 | 項目番号 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 |
| 3 | 1 | 0 | 0 |
| 4 | 1 | 1 | 1 |

$$s_j = \Pr(x_{ij} = 0 | \eta_{ij} = 1)$$
$$g_j = \Pr(x_{ij} = 1 | \eta_{ij} = 0)$$

理想反応は決定的入力（deterministic input）で，“かつ”という演算で算出される（and-gate）が，
項目反応は理想反応にノイズが乗った出力（noisy output）となっている。ゆえにDINAモデルという名前になる。

$\mathrm{P}(x_{ij} = 1 | s_j, g_j, \boldsymbol{z}_i)$

| 解答者番号 | 項目番号 1 | 2 | 3 |
|---|---|---|---|
| 1 | $g_1$ | $g_2$ | $g_3$ |
| 2 | $g_1$ | $1 - s_2$ | $g_3$ |
| 3 | $1 - s_1$ | $g_2$ | $g_3$ |
| 4 | $1 - s_1$ | $1 - s_2$ | $1 - s_3$ |

$\mathrm{P}(x_{ij} = 0 | s_j, g_j, \boldsymbol{z}_i)$

| 解答者番号 | 項目番号 1 | 2 | 3 |
|---|---|---|---|
| 1 | $1 - g_1$ | $1 - g_2$ | $1 - g_3$ |
| 2 | $1 - g_1$ | $s_2$ | $1 - g_3$ |
| 3 | $s_1$ | $1 - g_2$ | $1 - g_3$ |
| 4 | $s_1$ | $s_2$ | $s_3$ |

# DINAモデルの完全データの尤度・周辺尤度

- $\pi_l \geq 0$: $l$番目のアトリビュート習得パタンの混合比率（$\sum_l \pi_l = 1$を満たす）を導入する。
  - つまり，$\boldsymbol{z}_{il} = 1$となる確率：$\mathrm{P}(\boldsymbol{z}_i|\boldsymbol{\pi}) = \prod_l {\pi_l}^{z_{il}}$
- 完全データの尤度（局所独立と個人のランダムサンプリングを仮定）

$$\mathrm{P}(\mathbf{X}, \mathbf{Z}|\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{\pi}, \mathbf{Q}, \mathbf{A}) = \prod_i \prod_j \left\{ \prod_l \left\{ \pi_l \mathrm{P}\left(X_{ij} = x_{ij} \middle| s_j, g_j, z_{il} = 1, \boldsymbol{q}_j, \boldsymbol{\alpha}_l\right) \right\}^{z_{il}} \right\}$$

  - $\mathbf{X}, \mathbf{Z}$は$x_{ij}$と$z_{il}$をまとめた行列
  - $\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{\pi}$はそれぞれパラメタ$s_j, g_j, \pi_l$をまとめたベクトル
  - $\mathbf{Q}$は要素が$q_{jk}$のQ行列，$\mathbf{A}$は要素が$\alpha_{lk}$のアトリビュートパタン行列
- 周辺尤度

$$\mathrm{P}(\mathbf{X}|\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{\pi}, \mathbf{Q}, \mathbf{A}) = \sum_{\mathbf{Z}} \mathrm{P}(\mathbf{X}, \mathbf{Z}|\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{\pi}, \mathbf{Q}, \mathbf{A}) = \prod_i \left\{ \sum_l \left( \pi_l \prod_j \mathrm{P}\left(X_{ij} = x_{ij} \middle| s_j, g_j, z_{il} = 1, \boldsymbol{q}_j, \boldsymbol{\alpha}_l\right) \right)^{z_{il}} \right\}$$

- EMアルゴリズムによってパラメタ推定が可能
  - 潜在変数が$\mathbf{Z}$,パラメタが$\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{\pi}$

# EM algorithm（単調性の制約がない場合；山口, 2019）

- $\Theta^{\mathrm{old}} = [\boldsymbol{s}^{\mathrm{old}}, \boldsymbol{g}^{\mathrm{old}}, \boldsymbol{\pi}^{\mathrm{old}}]$とする
- E-step: $\mathrm{P}(z_{il}|\boldsymbol{x}_i, \Theta^{\mathrm{old}}, \mathbf{Q}, \mathbf{A}) = \frac{\mathrm{P}(z_{il}, \boldsymbol{x}_i|\Theta^{\mathrm{old}}, \mathbf{Q}, \mathbf{A})}{\sum_{l'} \mathrm{P}(z_{il'}, \boldsymbol{x}_i|\Theta^{\mathrm{old}}, \mathbf{Q}, \mathbf{A})} = \gamma(z_{il})$をすべての$i, l$で計算
- M-step: 関数$\mathrm{Q}_{\mathrm{EM}}(\Theta|\Theta^{\mathrm{old}}) = \mathbb{E}_{\mathrm{P}(\mathbf{Z}|\mathbf{X}, \Theta^{\mathrm{old}})}[\mathrm{P}(\mathbf{X}, \mathbf{Z}|\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{\pi}, \mathbf{Q}, \mathbf{A})]$を最大化する$\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{\pi}$を求める。
  - $\mathrm{Q}_{\mathrm{EM}}(\Theta|\Theta^{\mathrm{old}}) = \sum_i \sum_j \sum_l \gamma(z_{il}) [x_{ij}\{\eta_{lj} \log(1 - s_j) + (1 - \eta_{lj}) \log g_j\} + (1 - x_{ij})\{\eta_{lj} \log s_j + (1 - \eta_{lj}) \log(1 - g_j)\}] + \sum_i \sum_l \gamma(z_{il}) \log \pi_l$
  - $s_j^{\mathrm{new}} = \frac{\sum_i \sum_l \gamma(z_{il}) \eta_{lj} (1 - x_{ij})}{\sum_i \sum_l \gamma(z_{il}) \eta_{lj}}$, $g_j^{\mathrm{new}} = \frac{\sum_i \sum_l \gamma(z_{il}) (1 - \eta_{lj}) x_{ij}}{\sum_i \sum_l \gamma(z_{il}) (1 - \eta_{lj})}$, $\pi_l^{\mathrm{new}} = \frac{\sum_i \gamma(z_{il})}{I}$
- 収束基準を満たさない場合，$\Theta^{\mathrm{new}} = [\boldsymbol{s}^{\mathrm{new}}, \boldsymbol{g}^{\mathrm{new}}, \boldsymbol{\pi}^{\mathrm{new}}]$を$\Theta^{\mathrm{old}}$に代入し，E-step・M-stepを繰り返す。そうでなければ，$\Theta^{\mathrm{new}}$を出力して終了。
- 単調性の制約を考慮する場合には，M-stepで非線形最適化が必要。
- （漸近）標準誤差は，周辺尤度を使ってフィッシャー情報行列を計算して推定する。

# DINO model

**(Templin & Henson, 2006; Deterministic input noisy "or" gate model)**

- DINOモデルは DINAモデルの補償版
  - 理想反応の定義がDINAモデルとDINOモデルが異なっている。
- 理想反応： $\omega_{ij} = 1 - \prod_{k=1}^{K}(1 - \alpha_{ik})^{q_{jk}}$
  - $\omega_{ij}$ は解答者$i$が項目$j$に必要なアトリビュートを少なくとも1つ習得していれば1，そうでなければ0。
  - DINOモデルのslipとguessingの定義はDINAモデルと同じ
- DINOモデルのIRF： $\mathrm{P}(x_{ij} = 1 | s_j, g_j, \omega_{ij}) = (1 - s_j)^{\omega_{ij}} g_j^{1-\omega_{ij}}$
- 関連モデルとして，NIDA (noisy input deterministic "and" gate model)モデル，NIDO (noisy input deterministic "or" gate model) モデルなどもある。

# CDMの分類

- 2つの主要な分類: 補償（compensatory）モデルと非補償（ non-compensatory）
  - それぞれ，disjunctive，conjunctive (DiBello, Roussos, & Stout, 2006; Rupp, Templin, & Henson, 2010)と呼ばれることもある。

1. 補償モデル：特定のアトリビュートが未習得であっても，他のアトリビュート習得済のアトリビュートが正答確率を補償するモデル
   - e.g., DINO model (Templin & Henson, 2006), NIDO model (Templin, 2006), A-CDM (de la Torre, 2011), LLM (Maris, 1999), C-RUM (Templin, 2006 ).
2. 非補償モデル：項目の正答に必要なアトリビュートをすべて習得していなければ，正答確率が低くなってしまうモデル
   - e.g., DINA model (e.g., Haertel, 1989), NIDA model , RRUM (Roussos, DiBello, Stout, Hartz, Henson, & Templin, 2007).

- 飽和モデル（saturated model）.
  - GDINA (de la Torre, 2011) model, LCDM (Henson, et al., 2009 ), GDM (von Davier, 2008).

# 統合的モデル（e.g., G-DINAモデル, Generalized DINAモデル；de la Torre, 2011）

$$\Pr\left(X_j = 1|\boldsymbol{\alpha}_{jh}^*\right) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{jhk}^* + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\alpha_{jhk}^*\alpha_{hjk'}^* + \cdots + \delta_{j12\cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{jhk}^*$$

- 飽和モデル：補償・非補償の両方の性質を持つ
  - 特徴：交互作用の部分
  - $\delta_{j0}$：切片項，基本的な正答率
  - $\delta_{jk}$：主効果項，アトリビュート$k$がある場合の正答率の変化
  - $\delta_{jkk'}$：一次の交互作用項，アトリビュート$k, k'$が同時にある場合の正答率の変化
  - $\delta_{j12\cdots K_j^*}$：最高次の交互作用，項目の正答に必要なアトリビュートが全てある場合の正答率の変化
  - 分散分析と同様の考え方
- 非補償モデル（noncompensatory model）の例：DINAモデル：$\Pr\left(X_j = 1|\boldsymbol{\alpha}_{jh}^*\right) = \delta_{j0} + \delta_{j12\cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{jhk}^*$
- 補償モデル（compensatory model）の例：A-CDM：$\Pr\left(X_j = 1|\boldsymbol{\alpha}_{jh}^*\right) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{jhk}^*$
- リンク関数として，log, logit, identityを想定できる。
- $\boldsymbol{\alpha}_{jh}^* = \left[\alpha_{jh1}^*, \ldots, \alpha_{jhK_j^*}^*\right]$：その項目に特有のアトリビュート習得パタン（詳細は後述）
- $K_j^* = \sum_k q_{jk}$で，ここでは1~$K_j^*$番目のアトリビュートが必要な項目を想定

# 補償・非補償関係の決定

- IRFはデータ分析の前に，認知理論とアトリビュートの交互作用の点から決定される必要がある。
- しかし，先行研究から，アトリビュートの交互作用を事前に決めることが難しいことが示唆されている。
  - e.g., Yamaguchi & Okada (2018)：G-DINAモデルとその下位モデルを TIMSS (Trends in International Mathematics and Science Study) のデータ・セットの7つの国を用いて，AICとBICの観点から比較した。
  - 測定しているアトリビュートが同じにもかかわらず，国によって，最適なモデルが異なっており，補償モデルと非補償モデルの両方が選択された。
- 実際に使われるアトリビュートの交互作用として，補償・非補償の関係を事前に決定することが，難しい可能性
  - データにモデルを当てはめて，比較を行うことも必要。
- CDMは項目ごとに項目反応関数が違う可能性も考慮に入れる。
  - →（統計モデル的には）ある項目は補償的IRF，別の項目は非補償的IRF，としてもよい。

# パラメタ数からみたモデルの再分類

| モデル | 分類 | 制約 | アトリビュートの事前知識 | 特徴 |
|---|---|---|---|---|
| DINAモデル<br>DINOモデル | 倹約 | 強 | 多 | ×：データの複雑な構造に適合しない可能性がある<br>◯：モデルが適合すれば診断の結果も理論的背景を強く反映<br>◯：項目パラメタ（slip・guessing）の意味は理解しやすい |
| R-RUM<br>A-CDM<br>LLM | 主効果 | 中 | 中 | ◯：項目パラメタを参照して，Q行列を確認し修正するための簡便な方法として利用可能<br>×：アトリビュートの交互作用が表現できない<br>◯：パラメタ数が多く，データに適合しやすい |
| G-DINAモデル<br>LCDM<br>GDM | 飽和 | 弱 | 少 | ×：必要なサンプルサイズが相対的に大きい<br>◯：制約が弱く，アトリビュートの交互作用をすべて含む<br>◯：アトリビュートの組み合わせの効果について事前に予測ができない場合に探索的に使用可能 |

# 参考：（通常の）潜在クラスモデル

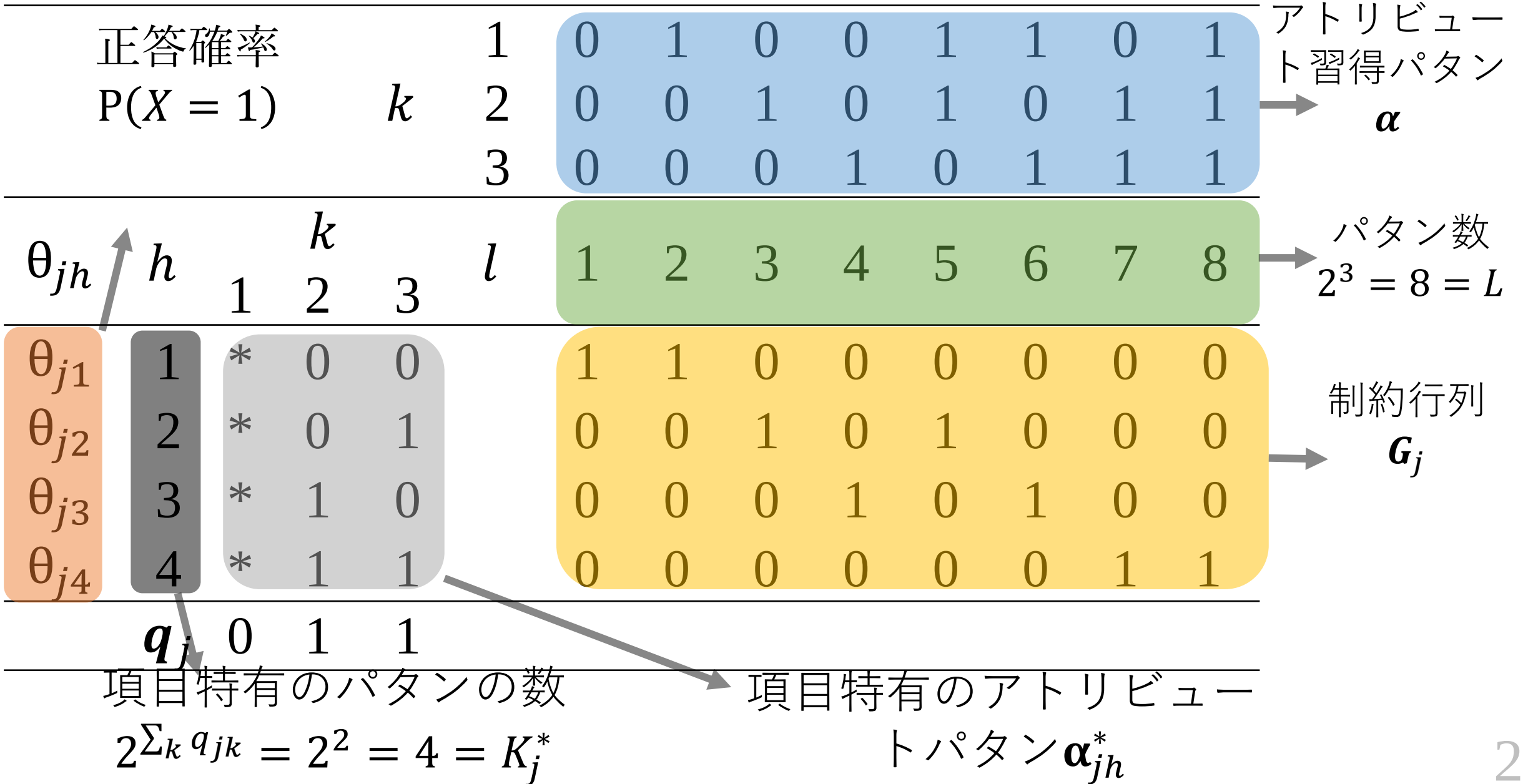
- $z_{il}$:個人$i$が$l\ (=1,\ldots,L)$番目の潜在クラスに所属しているかどうかを示す潜在変数。
- $\theta_{jl}$: $l$番目の潜在クラスの項目$j$の正答確率
- 潜在クラスモデルは以下のように定式化できる

$$\mathrm{P}\left(X_{ij}=x_{ij}\middle|\boldsymbol{z}_i,\boldsymbol{\theta}_j\right)=\prod_l\left\{\theta_{jl}^{x_{ij}}\left(1-\theta_{jl}\right)^{1-x_{ij}}\right\}^{z_{il}}$$

- $\pi_l \geq 0$: $l$番目の潜在クラスの母集団での混合比率で，$\sum_l \pi_l = 1$とする。

$$\mathrm{P}\left(X_{ij}=x_{ij}\middle|\boldsymbol{z}_i,\boldsymbol{\theta}_j,\boldsymbol{\pi}\right)=\prod_l\left\{\pi_l\left(\theta_{jl}^{x_{ij}}\left(1-\theta_{jl}\right)^{1-x_{ij}}\right)\right\}^{z_{il}}$$

# 制約付き潜在クラスモデルとしての定式化



# 一般的なCDMの定式化

- $g_{jhl}$：項目$j$ の制約行列（$G_j$）の要素.
  - 項目特有のアトリビュート習得パタン: $\boldsymbol{\alpha}_{jh}^{*}, h\left(h=1\cdots,2^{\sum_k q_{jk}}=K_j^{*}\right)$.
    - 例：$\boldsymbol{q}_{\mathrm{j}}=[0,1,1]$に対して４パタン $\boldsymbol{\alpha}_{jh}^{*}$:$[*,0,0]$, $[*,0,1]$, $[*,1,0]$, $[*,1,1]$
    - “$*$”：情報がないので，無視すべき要素.
  - Gの要素は $\boldsymbol{\alpha}_{jh}^{*}$が $\boldsymbol{\alpha}_l$ のうち$\{k;q_{jk}=0\}$となる$k$ を除いて，同じであれば1, そうでなければ 0.
- $\theta_{jh}$ :項目$j$特有のアトリビュート習得パタン$h$の正答確率
- 項目反応関数は

$$\mathrm{P}\left(X_{ij}=x_{ij}\middle|\boldsymbol{z}_i,\boldsymbol{\theta}_j\right)=\prod_l\prod_h\left\{\left[\theta_{jh}^{x_{ij}}\left(1-\theta_{jh}\right)^{1-x_{ij}}\right]^{g_{jhl}}\right\}^{z_{il}}$$

- Q行列は$G_j$を規定し，$G_j$は等値制約を表している。
  - 例：$G_j$の要素を理想反応$\eta_{lj}$とすればDINAモデルを表現できる，

# 項目反応理論モデルとの関係からの理解

- LCDM（Log-linear cognitive diagnostic model; Henson, Templin, & Willse, 2009）とIRTモデルの異同
- LCDM の項目反応：$\mathrm{P}\left(X_{ij}=1\middle|\lambda_{j,0},\boldsymbol{\lambda}_j,\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l,\boldsymbol{q}_j\right)=\frac{1}{1+\exp\left(-\left(\lambda_{j0}+\boldsymbol{\lambda}_j^{\top}h\left(\boldsymbol{\alpha}_l,\boldsymbol{q}_j\right)\right)\right)}$
- ただしヘルパー関数は,

  $\boldsymbol{\lambda}_j^{\top}h\left(\boldsymbol{\alpha}_l,\boldsymbol{q}_j\right)=\sum_{k=1}^{K}\lambda_{j,1,(k)}\alpha_{lk}q_{jk}+\sum_{k=1}^{K-1}\sum_{k'>k}^{K}\lambda_{j,2,\left(k,k'\right)}\alpha_{lk}\,\alpha_{lk'}q_{jk}q_{jk'}+\cdots.$
- LCDM：潜在特性が多次元でカテゴリカル（2値or多値），交互作用を含む
  - 厳密には単調性制約のために項目パラメタ$\boldsymbol{\lambda}_j$に制約が必要。
- 多次元IRT：潜在特性が多次元で連続値，基本は交互作用を含まない（補償的多次元IRT，非補償的なモデルもある）
- ★切片，主効果，交互作用を用いた表現は，G-DINAモデルより先にHenson et al. (2009)で提案されている。

# CDMとIRTの違い・使い分け？（所感）

- 潜在特性の性質
  - CDM：離散的な潜在変数でやや多い個数（３～１０数個程度）。カテゴリカルな多少荒っぽい潜在特性の表現。
  - IRT：連続的で比較的少数の潜在変数（１～３個程度）。連続量による細かい潜在特性の表現。
- 能力の交互作用の有無
  - CDM：交互作用あり->潜在特性の組み合わせの効果
  - IRT：交互作用なし->潜在特性の線形和の効果
- テスト利用場面
  - CDM：ローステイクスなテストへの利用（理解度確認テストなど）
  - IRT：ハイステークスなテストに利用（大規模テストなど）
- 個人へのフィードバックの内容
  - CDM：学習者がどのような強み・弱みがあるのか，学習改善のためのフィードバック
  - IRT：個人が有している能力保証（e.g., これくらいの能力がある）

# LCDMは潜在クラスモデルに変換可能

- 例：項目$\boldsymbol{q}_j = (1,1,0)$の項目反応確率は,

$$\theta_{j1} = \frac{\exp(\lambda_{j0})}{1+\exp(\lambda_{j0})},$$

$$\theta_{j2} = \frac{\exp(\lambda_{j0}+\lambda_{j,1,(2)})}{1+\exp(\lambda_{j0}+\lambda_{j,1,(2)})},$$

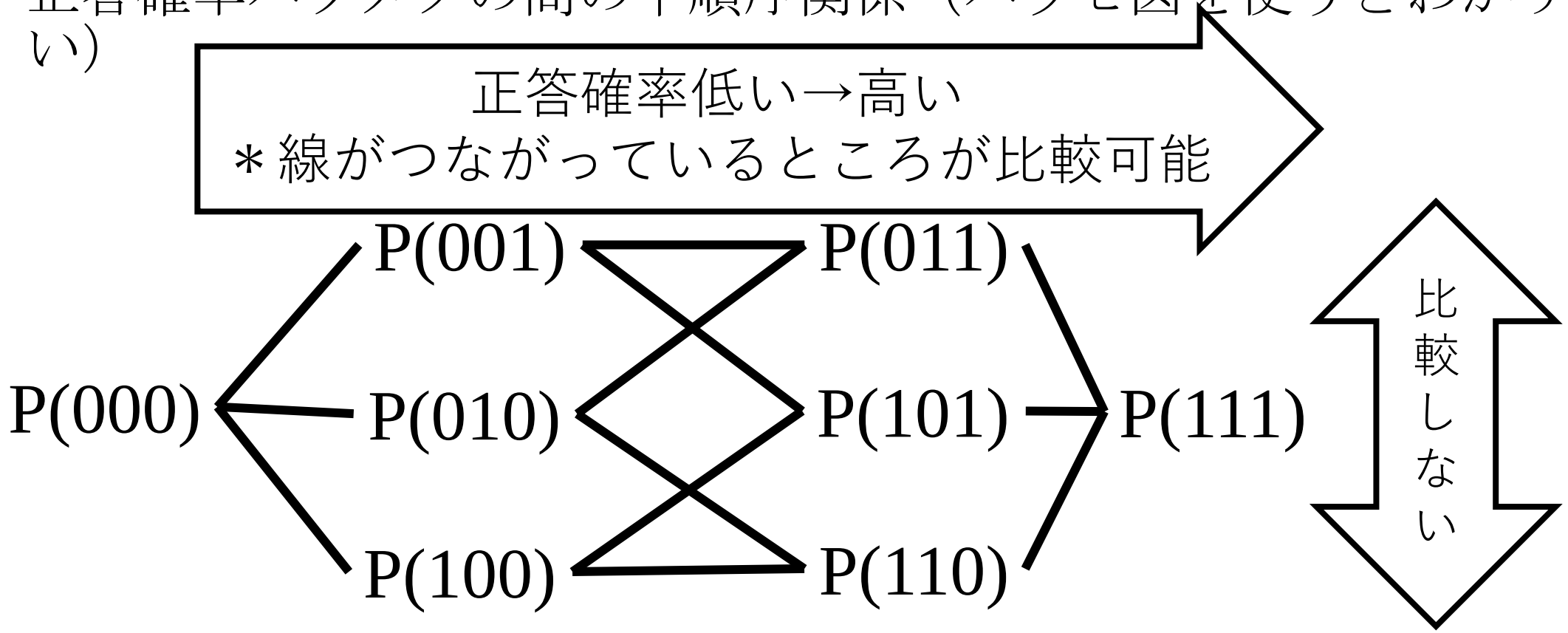
$$\theta_{j3} = \frac{\exp(\lambda_{j0}+\lambda_{j,1,(1)})}{1+\exp(\lambda_{j0}+\lambda_{j,1,(1)})},$$

$$\theta_{j4} = \frac{\exp(\lambda_{j0}+\lambda_{j,1,(1)}+\lambda_{j,1,(2)}+\lambda_{j,2,(1,2)})}{1+\exp(\lambda_{j0}+\lambda_{j,1,(1)}+\lambda_{j,1,(2)}+\lambda_{j,2,(1,2)})},$$

- と表す事ができる。先の潜在クラスモデルの表現と一致させられる。
  - GDINAモデルとも一致する。尤度も同じになる。
- 単調性の制約は，直接θの間に課せばよくなる。
- ★CDMには同値なモデルの表現がたくさんある。

# 単調性の制約

- Monotonicity constraints are “defined as the property such that for any examinee that masters additional skills his or her probability of a correct response must be equal to or greater than the probability of a correct response prior to learning the additional skills” (Henson et al., 2009, p. 198).
- 正答確率パラメタの間の半順序関係（ハッセ図を使うとわかりやすい）



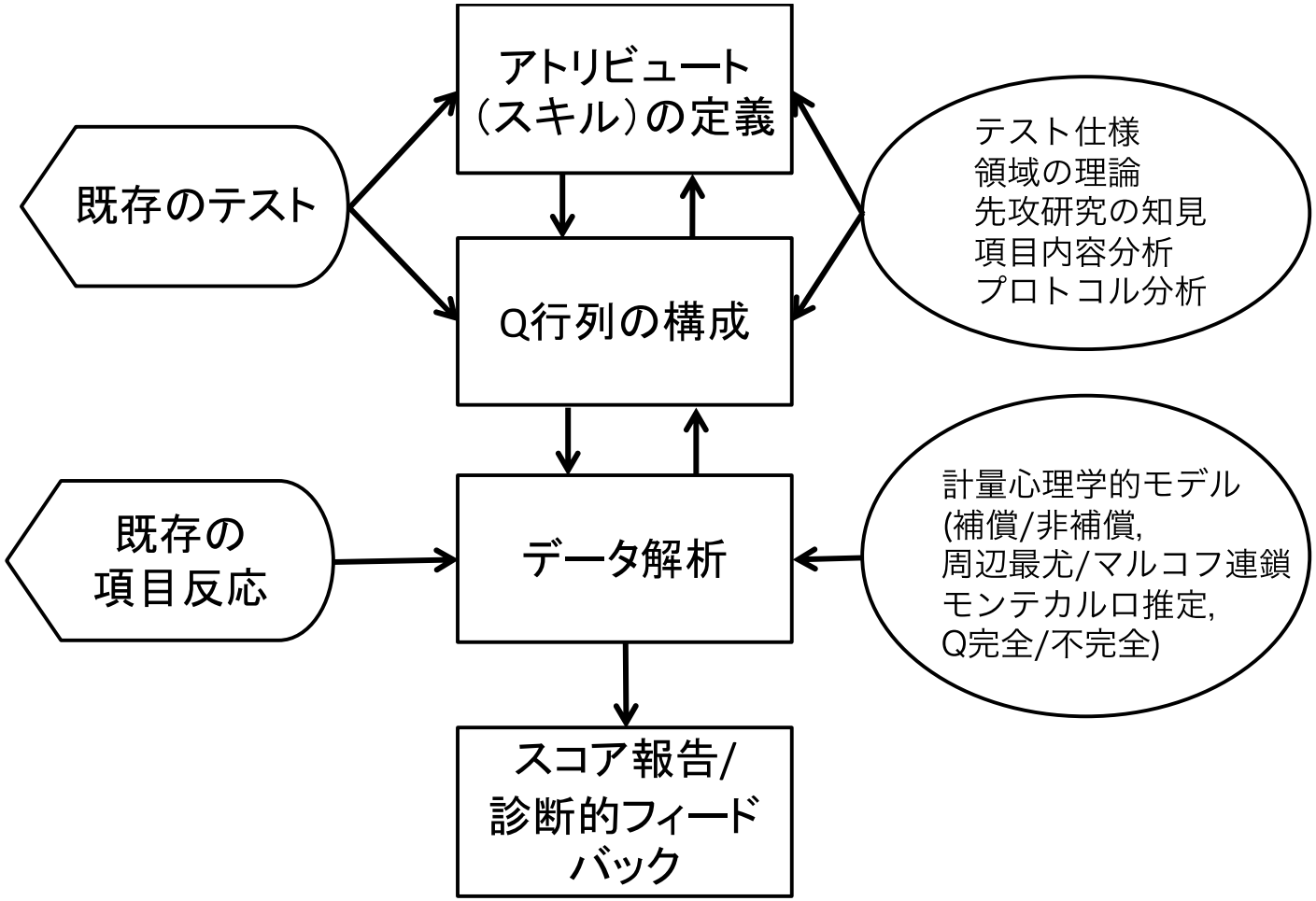
# CDMまとめ

1. テストデータから個人を分類する手法
   - どのようなアトリビュートが習得・未習得であるかを示すアトリビュート習得パタンを推定
2. 分析に必要なもの
   - アトリビュートの集合，Q行列，項目反応モデル，構造モデル，解答データ
3. Q行列の設定と項目反応モデルの指定がCDMを特徴づける
   - アトリビュート習得パタンの推定に対して重要
   - 構造モデルは，ソフトのデフォルトを使うことも多い
4. IRTモデル，潜在クラスモデルの両方に関係するモデルの表現
   - アトリビュートについての背景理論の表現
   - 推定のための表現（潜在クラス的）と解釈しやすい表現（IRT的）の違い

# CDMの運用方法（DiBello, Roussos, & Stout, 2006）

1. 検査目的の記述
2. 診断的興味のある潜在的アトリビュート（アトリビュート空間）の記述,
3. テスト項目の分析と開発,
4. 利用するモデルの特定,
5. 統計モデルの選択と結果の評価,
6. 解答者・教師・その他のテスト結果を利用する人に対する検査結果の報告

# 既存のテストに対する認知診断モデルの適用過程（Lee & Sawaki, 2009を改変）



# RによるCDMの実行

- いくつかのソフトウェアが存在
  - CDM（George, Robitzsch, Kiefer, Groß, & Ünlü, 2016）
    - GDINA,GDINAを含めたモデルパラメタの推定が可能
    - 適合度指標の出力も多い。
  - GDINA（Ma & de la Torre, 2020b）
    - GDINAができる範囲のことができる。
    - GUIでモデル指定ができるモードもある
    - Q行列の修正もできる
  - Jagsと呼ばれる汎用的なMCMCソフトを用いた各種CDMの実装方法については，Zhan, Jiao, Man, and Wang (2019)に詳しい。類似の内容は，岡田（2020, https://onl.tw/veZqxhF）でも紹介されている。
- LCDMはMplusを利用しているので，今回は扱わない。
- 今回は，Shi, Ma, Robitzsch, Sorrel Man (2021)のチュートリアルに沿った分析手順を紹介する。他にもチュートリアル論文がいくつか存在する。
- Rスクリプト：https://osf.io/3nqey/

# 例題データ

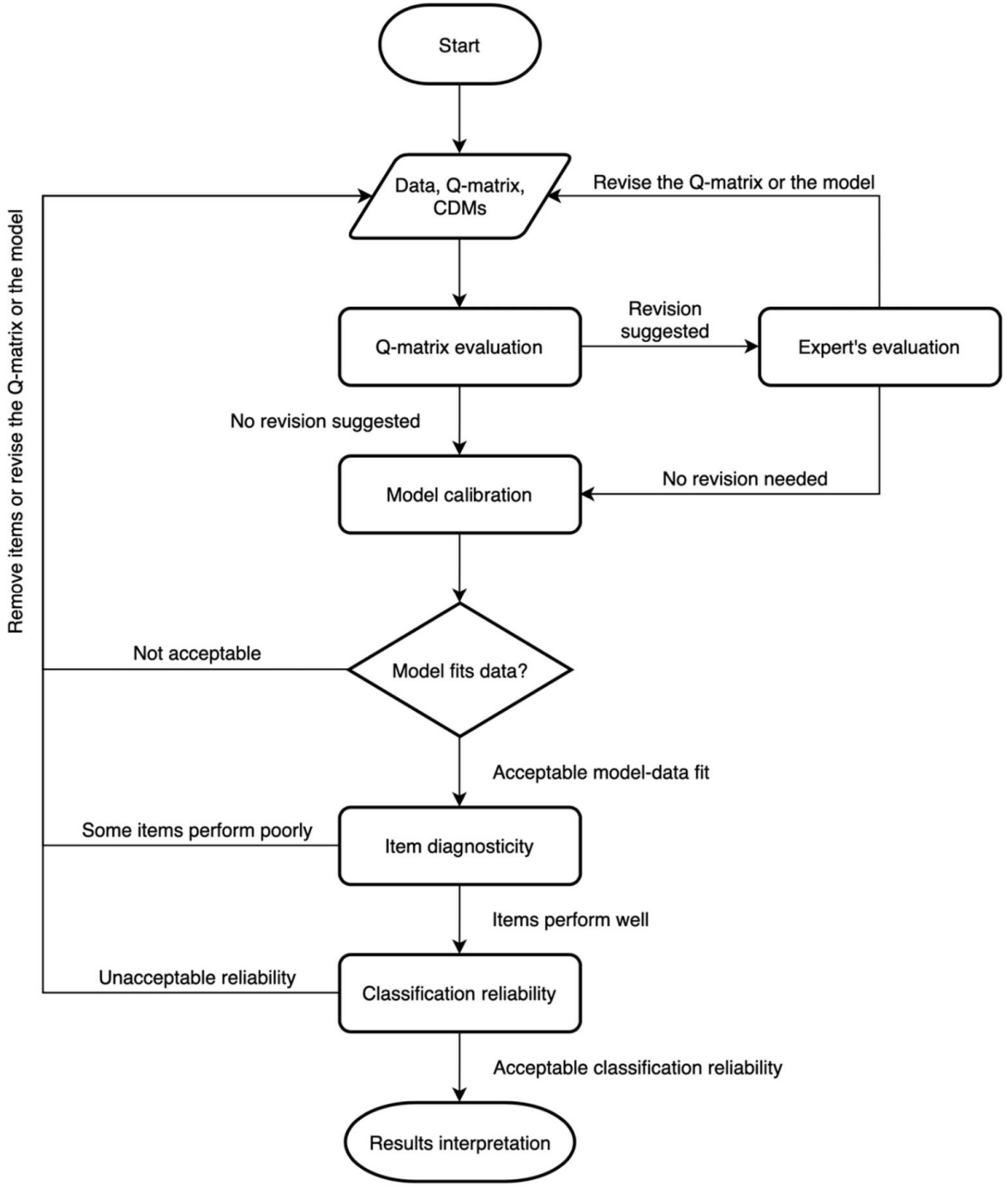
- The grammar section of Examination for the Certificate of Proficiency in English (ECPE)
  - e.g., Templin & Bradshaw (2014)
- Q-matrix（右参照）
  - 28項目
  - 3アトリビュート
  - morphosyntactic rules（形態的統語的規則）
  - cohesive rules（一貫性規則）
  - lexical rules（語彙規則）
- サンプルサイズ：2922

| Item | Morphosyntactic Attribute | Cohesive Attribute | Lexical Attribute |
|---|---|---|---|
| 1 | 1 | 1 | 0 |
| 2 | 0 | 1 | 0 |
| 3 | 1 | 0 | 1 |
| 4 | 0 | 0 | 1 |
| 5 | 0 | 0 | 1 |
| 6 | 0 | 0 | 1 |
| 7 | 1 | 0 | 1 |
| 8 | 0 | 1 | 0 |
| 9 | 0 | 0 | 1 |
| 10 | 1 | 0 | 0 |
| 11 | 1 | 0 | 1 |
| 12 | 1 | 0 | 1 |
| 13 | 1 | 0 | 0 |
| 14 | 1 | 0 | 0 |
| 15 | 0 | 0 | 1 |
| 16 | 1 | 0 | 1 |
| 17 | 0 | 1 | 1 |
| 18 | 0 | 0 | 1 |
| 19 | 0 | 0 | 1 |
| 20 | 1 | 0 | 1 |
| 21 | 1 | 0 | 1 |
| 22 | 0 | 0 | 1 |
| 23 | 0 | 1 | 0 |
| 24 | 0 | 1 | 0 |
| 25 | 1 | 0 | 0 |
| 26 | 0 | 0 | 1 |
| 27 | 1 | 0 | 0 |
| 28 | 0 | 0 | 1 |

# 分析フロー

- Shi et al. (2021) Figure 1

1. Q行列の評価
2. モデルのキャリブレーション
   - モデル-データフィットの評価
3. 項目の診断
4. 分類の信頼性
5. 結果の報告



# Q行列の妥当化

- 統計的な方法によるQ行列の妥当化の手続は，領域の専門家がQ行列作成を促進するツールとして利用されるべき。
  - 統計的な側面だけでなく，修正されたQ行列が内容的にも妥当なものとなっているのかどうかを慎重に判断する必要がある。
  - 構造モデル（アトリビュート間の関係）も理論的な点からモデルに組み込む
    - 高次の特性があるかどうか，アトリビュートの階層構造があるかどうか，など。
- 今回紹介する方法以外にも，探索的なCDM（exploratory CDM)として現在活発に方法論の開発が続いている。
  - E.g., Chen, Culpepper, Chen, & Douglas (2018)
- 個人的な注：本当は，当該テストと別のテストや指標などとアトリビュートとの関係も調べる必要があるはず。

# モデルデータフィット

- テストレベルのフィットと項目レベルでのフィット
  - Absolute fit indices vs. relative model fit indices
- 絶対適合度の指標
  - ピアソンカイ二乗値，limited information statistics, root mean square error of approximation, the standardized root mean square residualなど
- 相対指標（統計量）
  - AIC, BIC，尤度比検定，Wald検定など
- 項目レベルのフィットや後述の識別力が悪い項目は，Q行列の修正や削除が必要になることも。

# 項目の分析・分類の信頼性

- アトリビュートパタンごとの正答確率の図示
- 項目の識別力の計算
- その項目がよい特性を有していそうかをチェックする。
- CDMの最終目標は，個人のアトリビュート習得状況の推定
  - ＝アトリビュート習得パタンへの分類
- 分類の正確性や一貫性を，信頼性とみなして評価する。
  - CDMの信頼性については，Yamaguchi and Templin (2022)などでまとめられている。

# Rによる分析フロー

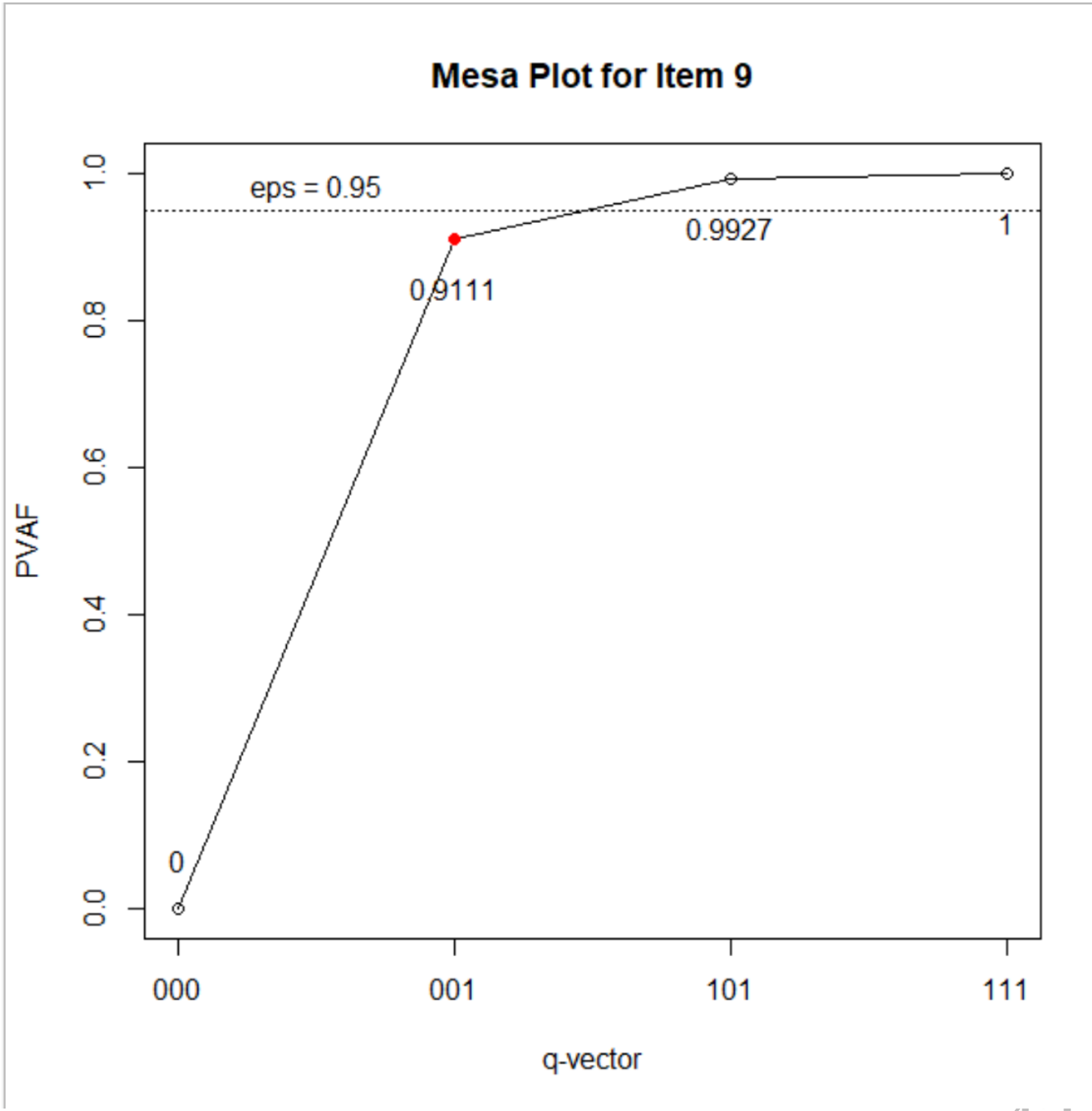
1. Qval関数でQ行列の要素を修正する。
   - 今回は，アトリビュート数はすでにわかっているとする。
2. GDINA関数で推定
3. modelcomp関数により，項目反応関数を探索
4. summary関数，modelfit関数, itemfit関数で，適合の確認
5. coef関数で，項目パラメタの抽出
6. 識別力の検査
7. CA関数による，分類の正確性の診断
8. personparm関数，個人のアトリビュート習得パタンを推定

# Qval関数

- データによるQ行列の修正
- Qval(res_gdinda,method = "wald")
- Waldテストによる，Q行列の要素の修正
- さらに， mesa plotと呼ばれる図を見て，どのQ行列が良いのかを確認してみる必要がある。
  - X軸：Qベクトル
  - Y軸：PVAF (the proportion of variance accounted for)値,
  - 説明された分散の比率が事前に決めたカットオフ値を超えているQベクトルのうち，最小のものがよいとする。

# Mesaプロット

- 赤い点がオリジナルQベクトル
  - [001]という設定
- eps(=0.95)はカットオフ
- PVAFが0.95を超えるQベクトル
- のうち最も小さいのは，[101]
- [001]→[101]と修正する



# GDINA関数

- モデルのパラメタの推定
- res_gdinda <- GDINA( dat =data, Q = Q, model = "GDINA", mono.constraint = T,control = list(conv.crit = 1e-4))
- 基本引数
  - dat：データ行列
  - Q：Q行列
  - model：推定に利用するモデル
    - model="logitGDINA"とするとLCDMパラメタと同等の推定が可能。
  - mono.constraint：単調性の制約，デフォルトはF。理論的には制約を課したほうがよい。ただし，推定にやや時間がかかる。
  - control：リスト形式で，推定のための詳細な設定ができる。ここでは，収束基準を指定している。

# （発展的な話）パラメタの固定について

- Delta パラメタ自体を固定する機能はGDINA関数にはない。
- そのかわり，各項目に対して，アトリビュート習得パタンごとにパラメタを指定して，固定することができる。
- また，構造モデルの部分も，適宜固定することができる。
- 具体的な方法については，Shi et al. (2021) p.820~821に記述がある。

# modelcomp関数

- modelcomp(GDINA.obj = res_gdinda) 2つ以上のアトリビュートが必要な項目に対して，Waldテストによるモデル選択が可能。
- 出力の“models”列は，推奨されたCDM
  - よりシンプルなモデルで，*p*値が最も大きいもの
  - 詳細はMa & de la Torre (2019)
- ここでの帰無仮説は，シンプルなモデルがGDINAモデルと同程度に適合している，というもの。
  - なので，帰無仮説が棄却されないなら，よりシンプルなモデルを採用することになる。
- GDINAモデルにネストしているモデルしか検討できない。
- ネストしていないモデルの比較には，別の指標が必要。

# summary関数・modelfit関数・ itemfit関数

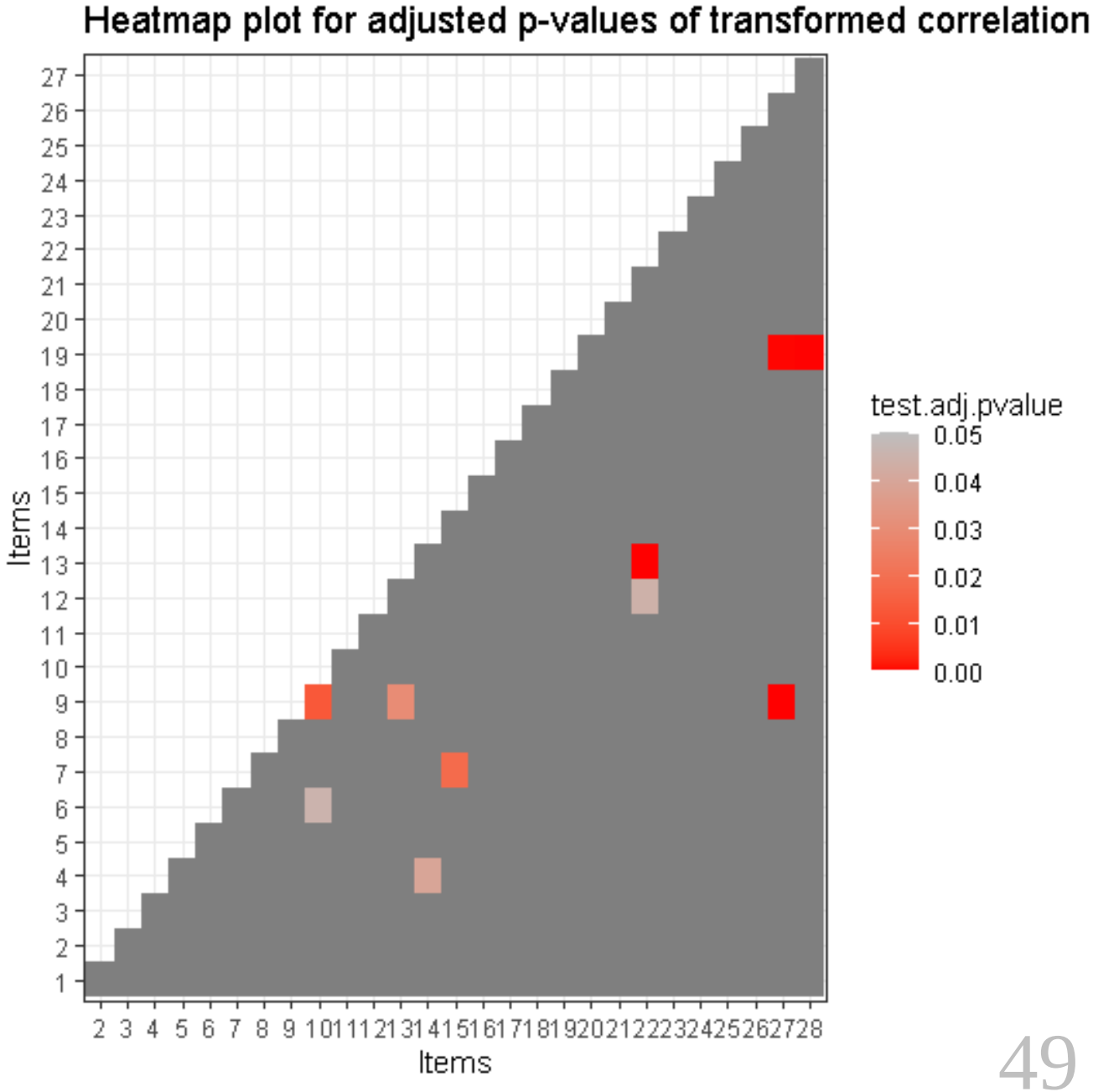
- summary(res_gdinda), modelfit(res_gdinda), or itemfit(res_gdinda)
- 対数尤度（Loglik），AIC，BIC,など確認できる。
- また，アトリビュート習得の周辺比率も表示される（Attribute Prevalence）
  - どのアトリビュートが比較的習得されていたかなどがわかる。
- Itemfit関数は，そのままでは集計された項目レベル・項目ペアレベルの絶対適合指標を返す。
  - 詳細な値については，itemfitオブジェクトにextract関数を作用させて，値を取り出す。

# 適合度指標（の一部）

- $M_2$ : second order marginal statistics
- RMSEA2: limited information RMSEA
- SRMSR: standardized means square root of squared residuals
- MaxAD.r: maximum absolute difference between observed and predicted Fisher transformed correlations
- MaxAD.LOR: maximum absolute deviation between observed and predicted log odds ratio
- 詳細は，Shi et al. (2021)など参照

# ヒートマップ

- 変換された項目間相関
- adjusted-p valueが小さい項目のペアは，Misfitを示唆
- 例えば項目9と10の間など。

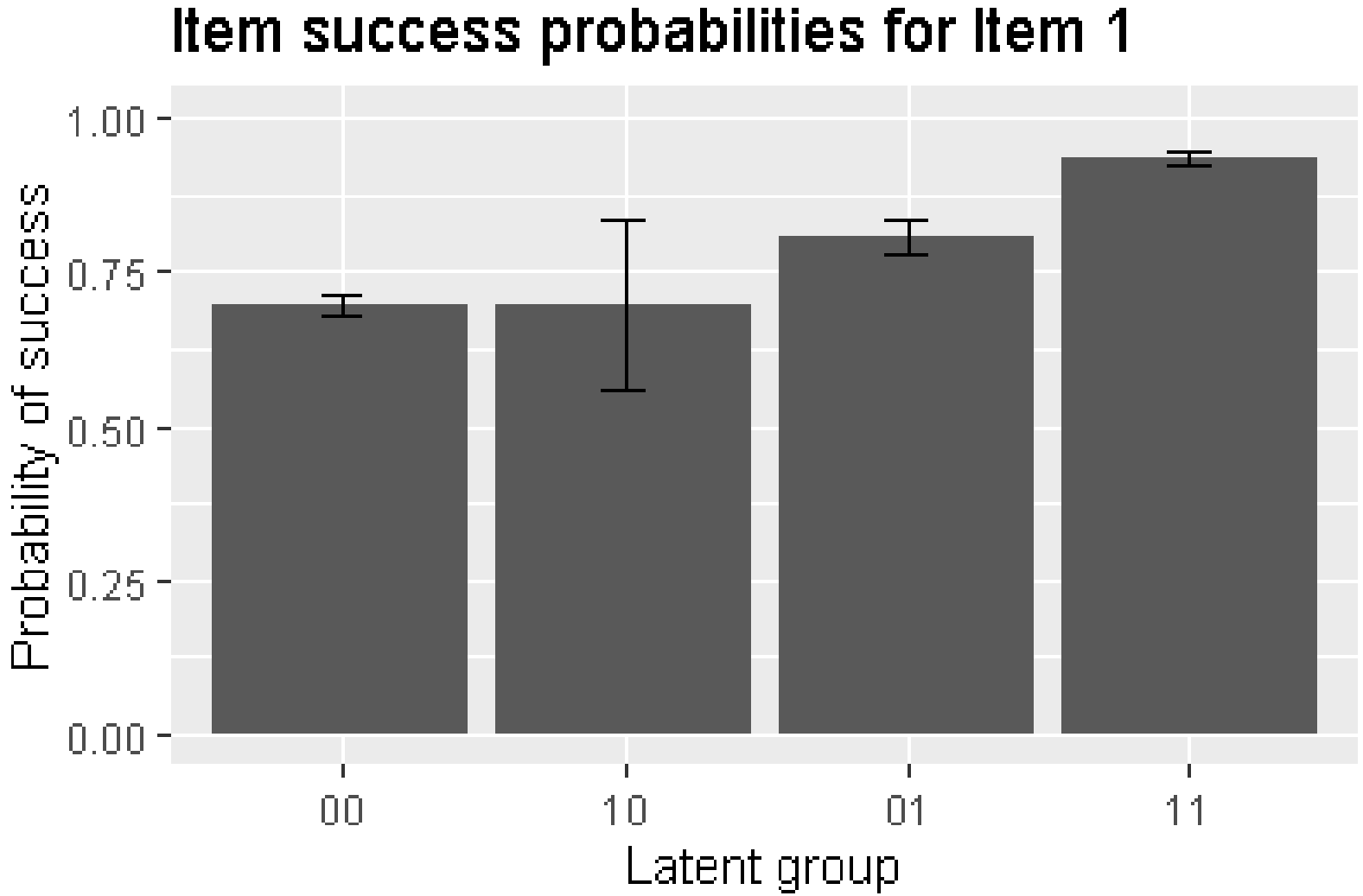


# coef関数

- 項目パラメタの推定値を表示
- coef(res_gdinda, what = "delta",withSE=TRUE)
- 引数
  - GDINAオブジェクト,
  - what: パラメタの種類, “catprob”か “delta”が比較的わかりやすい。
  - -> deltaはGDINAモデルのδパラメタの推定値
  - withSE：標準誤差の出力，標準誤差のタイプ（SE.type引数で変更可能）は，不完全情報行列による。

# 項目識別力のチェック

- plot(res_gdinda,item = 1,withSE = TRUE)によって，アトリビュート習得パタンごとの正答確率を棒グラフにすることができる。



# 項目識別力

- extract(res_gdinda, what=“discrim”)により，算出
- P(1)-P(0)：項目に必要なアトリビュートをすべて習得している場合の正答確率から，必要なアトリビュートを全く習得していない正答確率
- GDI：G-DINA discrimation index
  - 正答確率の分散に関する尺度
- どれくらいの値だと良いのかは統一的な見解はない。
- 高い値は重要だが，Qベクトルを過剰に指定（1とする箇所が多すぎる）可能性もある。

# 分類の正確性

- CA(res_gdinda, what="MAP"), デフォルトは，MLによる。
- パターンレベルとアトリビュートレベルの２種類を出力。

- 一貫性についての指標は，GDINAパッケージでは提供していない。
- 他の信頼性の指標もあり，CDMパッケージでは算出可能。

# personparm関数

- 個人のアトリビュート習得パタンを表示
- personparm(res_gdinda, what = "EAP")
- 引数
  - GDINAオブジェクト
  - what：推定方法の種類，”EAP”（期待事後推定）,”MAP”（事後確率最大化推定）,”MLE”（最尤推定）が選択可能。デフォルトはEAP
- 最終的に，個々人のテスト受験者が得る結果に近い。

# 発展的な話題

- モデルの拡張
- Q行列推定の最前線
- コンピュータアダプティブテスト
- 応用研究・社会実装に向けて

# 拡張的なモデル

- 日本語では，山口・岡田（2017b）でレビューがあるように，様々な発展がなされている。
- 縦断的CDM（e.g., Madison & Bradshaw, 2018）：アトリビュート習得の変化を追跡可能
- 多値型アトリビュート・多値反応CDM（e.g., de la Torre, 2009;Zhan, Bian, & Wang, 2016）：アトリビュートに習得段階を想定したり，様々なテスト項目にCDMが適用可能となっている。
- 外部情報を利用したモデル（e.g., Park, Xing, & Lee, 2018）：Explanatory CDMなどと呼ばれることも。他のテストや個人の背景情報を用いて，アトリビュートの習得確率の推定精度を高めたりする試み。
  - 反応時間を含めたモデルもある。
  - アトリビュートの妥当化のためにも必要なように思われる。
- アトリビュートの階層構造を考慮したモデル（Leighton, Gierl, & Hunka, 2004）：アトリビュートの習得順序を考慮して，学習マップを作成可能。アトリビュート数が多くても，実質的なアトリビュート習得パタンが少なくなる。
- ★アイデア次第で様々な拡張が可能。

# Q行列推定

- Q行列の設定は非常に難しい。
  - CDMの結果が妥当で納得できるかどうかを左右する。
- アトリビュートの背景理論だけでなく，データからQ行列を推定する試みが非常に活発に研究されている。
- モデルの識別性との関連もある話題で，CDMの応用をすすめる上でも重要。
- 統計モデルの理論的検討は非常に重要であるものの，本当に項目反応関数がCDM的でないと，Q行列のデータからの推定はあまり意味がない可能性がある。
- 実質科学的に考えるのであれば，介入等によって特定のアトリビュートが習得される様になるのかどうかといった検討が必要になる。

# コンピュータアダプティブテスト

- コンピュータアダプティブテスト（CAT）も，効率的にアトリビュートの習得パタンを推定するために必要。
  - Personalized learningのためには，テストの個人化も必要。
- 様々な項目選択アルゴリズムが提案されている。
  - E.g., Cheng (2009)
- IRTモデルを用いたCATと同様に，様々な制約要因を考慮している。
- オンライン学習が一般的になってきているので，こうした技術の発展も期待できる。
  - アルゴリズムがやや乱立している気もするので，様々テスト場面に利用できる，統一的かつ柔軟な方法が望まれる。

# 応用研究・社会実装に向けて

- CDMの応用研究も年々増えてきている。
- しかし，実際にCDMを用いた学習改善などまで到達している例はあまりない。
- フィードバックがどれほど学習を促進できるのか，という点は今後深めていく必要がある。
- 長期的な学習改善の方法など，“CDMをどのように使っていくと良いのか”という点には，様々な工夫の余地がある。
- IRTモデル研究ではあまり取り上げられていない，CDM研究の独自の研究関心となる可能性。
  - 個人の研究だけでは難しい問題。

# 読書案内

- von Davier, M., & Lee, Y. S. (2019). Handbook of diagnostic classification models. *Cham: Springer International Publishing*.
  - 最近のCDMの話題が網羅されている。理論的な話題からソフトウェアの使い方まで記載がある。CDMの最新の話題に手早く追いつきたい場合は，この本を熟読するのが近道。
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). Diagnostic measurement. *Theory, Methods, and Applications, New York: Guilford*.
  - LCDMを中心にCDMのまとまったテキスト。今も研究されているCDMのアイデアも多く含まれており，読み返すたびに発見がある。
- Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
  - 数理的にテクニカルな本ではないが，CDMに関しての妥当性の議論や，アトリビュート階層法などについて話題もよく記述されている。
- Nichols, P. D., Chipman, S. F., & Brennan, R. L. (1995). *Cognitively diagnostic assessment*. Routledge.
  - やや記述が古い部分もあるが，unifiedモデルについての記述や，IRTモデルを用いた認知能力の測定との関係なども勉強になる。

# 引用文献


- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika, 74*(4), 619-632.
- Chen, Y., Culpepper, S. A., Chen, Y., & Douglas, J. (2018). Bayesian estimation of the DINA Q matrix. *Psychometrika, 83*(1), 89-108.
- DiBello, L. V., Roussos, L. A., & Stout, W. (2006). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. Rao & S. Sinharay (Eds.), *Handbook of Statistics, Vol. 26: Psychometrics*. North Holland: Elsevier, pp. 979–1030. doi: https://doi.org/10.1016/S0169-7161(06)26031-0.
- Templin, J. (2016). Diagnostic Assessment: Methods for the reliable measurement of multidimensional abilities. In Drasgow, F. (Ed.), *Technology and Testing Improving Educational and Psychological Measurement*. New York: NY, Routledge, pp. 285-304.
- de la Torre, J. (2009). A cognitive diagnosis model for cognitively-based multiple-choice options. *Applied Psychological Measurement, 33*, 163–183.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*, 179-199. doi: http://dx.doi.org/10.1007/s11336-011-9214-8.
- de la Torre, J. & Chiu, C-Y. (2016). A General Method of Empirical Q-matrix Validation. Psychometrika, 81, 253-273.
- George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software*, *74*(1), 1-24. DOI: 10.18637/jss.v074.i02

# 引用文献


- Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191-210. https://doi.org/10.1007/s11336-008-9089-5
- Lee, Y. W., & Sawaki, Y. (2009). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly, 6*(3), 172-189.
- Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of educational measurement, 41*(3), 205-237.
- Ma, W., & de la Torre, J. (2019). Category-Level Model Selection for the Sequential G-DINA Model. Journal of Educational and Behavioral Statistics. 44, 61-82.
- Ma W, de la Torre J (2020a). An Empirical Q-Matrix Validation Method for the Sequential G-DINA Model. *British Journal of Mathematical and Statistical Psychology, 73(1*), 142–163. doi:10.1111/bmsp.12156
- Ma, W., & de la Torre, J. (2020b). GDINA: an R package for cognitive diagnosis modeling. *Journal of Statistical Software*, *93*(14), 1-26. DOI: 10.18637/jss.v093.i14
- Madison, M. J., & Bradshaw, L. P. (2018). Assessing growth in a diagnostic classification model framework. *Psychometrika, 83*(4), 963-990.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64*, 187-212. doi: https://doi.org/10.1007/BF02294535.

# 引用文献


- 岡田 謙介（2020）.情報教育の指導と授業改善に役立つベイズ統計, 日本教育工学会 第19回情報教育研究会, Available from: https://www.dropbox.com/s/aqinooe28elyhbu/cdm_20201219c.pdf?dl=0
- 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.
- Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. (2007). The fusion model skills diagnosis system. In J. Leighton & M. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge: Cambridge University Press, pp. 275–318.
- Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement, 6*, 219-262. doi: http://dx.doi.org/10.1080/15366360802490866.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). Diagnostic measurement: Theory, methods, and applications. New York, NY: Guilford Press.
- Shi, Q., Ma, W., Robitzsch, A., Sorrel, M. A., & Man, K. (2021). Cognitively Diagnostic Analysis Using the G-DINA Model in R. *Psych, 3*(4), 812-835. https://doi.org/10.3390/psych3040052
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. Journal of educational measurement, 345-354.
- Templin, J. L. (2006). *CDM Cognitive diagnosis modeling with Mplus user guide*., Mplus. Available from: http://jtemplin.coe.uga.edu/files/dcm/cdm/CDM_user_guide.pdf.

# 引用文献


- Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika, 79,* 317–339. https://doi.org/10.1007/s11336-013-9362-0
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*, 287-305. doi : 10.1037/1082-989X.11.3.287.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology, 61*, 287-307. doi: 10.1348/000711007X193957.
- 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
- 山口 一大（2018）.認知診断モデルの理論的および経験的検討.東京大学博士論文 DOI: info:doi/10.15083/00077834
- 山口 一大 （2019）. DINAモデルの再定式化と推定アルゴリズムの直接的導出 日本テスト学会誌, 15(1), 21-43. https://doi.org/10.24690/jart.15.1_21
- Yamaguchi, K., & Templin, J. (2021). A Gibbs sampling algorithm with monotonicity constraints for diagnostic classification models. *Journal of Classification*. https://doi.org/10.1007/s00357-021-09392-7
- Yamaguchi, K., & Templin, J. (2022). Observed score reliability indices in diagnostic classification models. *Behaviormetrika*. *49*, 47–68. https://doi.org/10.1007/s41237-021-00153-9
- Zhan, P., Bian, Y., & Wang, L. (2016). Factors affecting the classification accuracy of reparametrized diagnostic classification models for expert-defined polytomous attributes. *Acta Psychologica Sinica, 48*, 318–330.
- Zhan, P., Jiao, H., Man, K., & Wang, L. (2019). Using JAGS for Bayesian cognitive diagnosis modeling: A tutorial. *Journal of Educational and Behavioral Statistics*, *44*(4), 473-503. https://doi.org/10.3102/1076998619826040