統計数理（2024）
第 72 巻 第 1 号 93–119

特集「心理統計学の新展開」
［研究詳解］

近年の診断分類モデルの推定法の展開

山口 一大†

（受付 2023 年 6 月 23 日；改訂 12 月 7 日；採択 12 月 12 日）

要　旨

診断分類モデルあるいは認知診断モデルは学力テストの解答に必要な一連の認知能力（アトリビュート）を想定し，テスト受験者をアトリビュート習得の有無のパタンに分類する統計モデルである．診断分類モデルは，教育テスト分析において有用なツールであり，応用も広がりつつあるものの，パラメタ推定についての包括的な日本語のレビューは未だ存在しない．そこで本稿では診断分類モデルのパラメタ推定法についての展開を展望し，診断分類モデルの応用の促進と理論的発展に資することを目指した．レビューの結果，最尤推定法，ベイズ推定法，ノンパラメトリック推定法の 3 種類の方法での発展がみられた．最尤推定法では正則化法の利用，ベイズ推定法では変分ベイズといった比較的新しい方法も用いられていた．レビューの結果を踏まえて，診断分類モデルの推定法について残された問題と今後の展望について議論した．

キーワード：診断分類モデル，認知診断モデル，パラメタ推定法．

1. はじめに

学力テストからテスト解答者の強みや弱み推定することができる統計分析モデルとして，診断分類モデル（diagnostic classification models [DCMs]）あるいは認知診断モデル（cognitive diagnostic models [CDMs]）と呼ばれるモデル（Rupp et al., 2010; von Davier and Lee, 2019）への注目が集まっている．DCM では，例えば中学生における「数学能力」という包括的な構成概念を分解し，「計算力」，「概念理解」，「図形操作力」，「論理力」などといったテストの正答に必要なより細かい認知要素（アトリビュート；attribute）を想定することで，個々のテスト解答者がどのアトリビュートを習得しているのかを推定することができる（応用例として，鈴木 他, 2015 などが参考になる）．このように，DCM はテスト解答者がどのような認知的なつまづきを抱えているのかを，学力テストへの項目反応パタンから推定し，個別のテスト解答者への学習の診断情報の提供を可能にするモデルである．

DCM は，統計モデルとしては制約付き潜在クラスモデル（restricted latent class models; Rupp and Templin, 2008b）として考えることができる．DCM では，アトリビュートの習得の有無の組み合わせによって定義されるアトリビュート習得パタンが潜在クラスに対応する．和文で参照できる DCM のレビューとして，山口・岡田（2017）があり，DCM の主要なモデルについては網羅的に解説されている．

しかし，山口・岡田（2017）では，DCM の推定法の詳細は記載されていない．この他，山口（2019）は DCM の中でも最も基本的な DINA モデル（deterministic input noisy “AND”-gate model; e.g., Junker and Sijtsma, 2001）の推定法を示しているが，特定の DCM の特定の推

† 筑波大学 人間系心理学域：〒 305–0006 茨城県つくば市天王台 1–1–1

定法について述べているにすぎず，記述が限定的であると言わざるを得ない．DCM についての成書 (e.g., Rupp et al., 2010; von Davier and Lee, 2019) においても，DCM のパラメタ推定について包括的にレビューされていない．DCM を含めた潜在変数モデルにおいては，一般にモデルの特定のみならずパラメタ推定法の選択も重要であるものの，包括的に DCM のパラメタ推定法をレビューした日本語で参照できる文献がないという現状がある．

DCM のパラメタ推定法をレビューした論文がない理由として，DCM 定式化の複雑さがあると考えられる．DCM は，アトリビュートを多次元のカテゴリカル変数とみなすか，アトリビュート習得パタンを潜在クラスとみなすかという定式化の任意性があり，結果の解釈をしやすい定式化と推定を行いやすい定式化に隔たりがある．すなわち，推定法を統合的に理解する際にわかりやすい定式化とモデルを応用していく際に解釈しやすい定式化が一致していないということである．このような DCM の定式化の複雑さから，DCM の推定方法のレビューが行われてこなかった可能性がある．この結果として，DCM の専門家以外にはパラメタの推定法を統合的に理解することが難しい現状があると考えられる．本稿の意義は，DCM の 2 種類の定式化とその関係を示し，現在提案されている DCM の推定法を概観し，それぞれの推定法の特色について整理することにある．

上記を踏まえて，本稿ではパラメタ推定法を定式化しやすい定式化として潜在クラスモデルベースのものを採用し，一般性の高い DCM がどのように制約つき潜在クラスモデルとして定式化できるのかを示す．その上で，多数の推定法の異同や，DCM に特有のノンパラメトリック推定法や単調性制約を満足するような推定法についても合わせてレビューする．これにより，DCM の特徴を，推定法の観点からも理解できると期待できる．DCM の推定法をレビューすることにより，DCM の応用のみならず DCM の理論的研究の発展に寄与することを本稿の目的とする．なお，本研究では Q 行列と呼ばれる要素が既知である確証的な DCM のパラメタ推定について扱い，近年盛んに研究がなされている Q 行列もデータから推定を行う探索的 DCM (e.g., Culpepper, 2019) についての詳細は扱わないこととした．

以下に本論文の構成を示す．第 2 章では，潜在クラスモデルベースの診断分類モデルの定式化および単調性制約 (monotonicity constraints) について説明する．本稿では，最も一般性の高い DCM の一つである対数線形認知診断モデル (log linear cognitive diagnostic model [LCDM]; Henson et al., 2009) から，潜在クラスモデルベースの診断分類モデルの定式化を行う．LCDM のモデルパラメタは DCM の応用に際して理解しやすいパラメタであり，DCM というモデル族の特徴を把握しやすいと考えられる．なお，本稿では von Davier and Lee (2019) などの最新の成書に準じて，CDM ではなく DCM という用語を用いるが，LCDM はすでに定着したモデルの名称であるため本稿でもその名称を変更せずに用いる． 第 3 章では，最尤推定法に関係する推定法を概観する．特に，Expectation-Maximization (EM) アルゴリズムによる推定アルゴリズムおよび単調性制約を満足するための方法について概観する．更に，正則化推定法 (regularized estimation method) についても示す．第 4 章ではベイズ推定法について概観する．古典的な MAP (maximum a posteriori) 推定のみならず，MCMC による事後分布の近似法と変分ベイズによる事後分布の近似法についても述べる．第 5 章ではノンパラメトリック推定法について示す．第 6 章では，本稿の推定法のレビューにもとづき，DCM の推定法の選択について議論する．最後の第 7 章では，レビューした推定法から DCM のパラメタ推定について残された問題と今後の研究について議論を行う．

2. 診断分類モデルの定式化

2.1 記号の準備と測定モデル

本節では項目反応が 2 値である DCM の基礎的な定式化を行う．まず，あるテスト解答者 $i = 1, 2, \ldots, I \in \mathbb{N}$ の項目 $j = 1, 2, \ldots, J \in \mathbb{N}$ への反応を $X_{ij} \in \{0, 1\}$ とする．ここで，$X_{ij} = 1$ は正答反応，$X_{ij} = 0$ を誤答反応とする．DCM においては，テスト全体で $K \in \mathbb{N}$ 個のアトリビュート，すなわち問題の正答に必要な認知要素を想定する．アトリビュートは $\alpha_k \in \{0, 1\}(k = 1, 2, \ldots, K)$ であり，$\alpha_k = 1$ は k 番目のアトリビュートを習得している状態，$\alpha_k = 0$ は未習得の状態を意味する．項目反応およびアトリビュートは多値を想定することもある．K 個のアトリビュート習得の有無を並べたベクトル $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)^\top \in \{0, 1\}^K$ をアトリビュート習得パタンと呼ぶ．アトリビュート数が K であることから，ありうるアトリビュートの習得パタンの数は 2^K であり，アトリビュート習得パタンを添え字 $l = 1, 2, \ldots, L = 2^K$ によって区別して $\boldsymbol{\alpha}_l$ と書き，その要素を α_{lk} とする．また，テスト解答者 i のアトリビュート習得パタンを $\boldsymbol{\alpha}_i$ とする．

DCM においては，項目 j の正答に関係するアトリビュートを示す長さ K のベクトル $\boldsymbol{q}_j = (q_{j1}, q_{j2}, \ldots, q_{jK})^\top \in \{0, 1\}^K \backslash \mathbf{0}_K$（ただし，$\mathbf{0}_K$ は 0 を K 個並べたベクトル）を $\boldsymbol{q}$ ベクトルと呼び，その k 番目の要素 $q_{jk} \in \{0, 1\}$ は $q_{jk} = 1$ ならば項目 j の正答にアトリビュート k が必要，$q_{jk} = 0$ であれば不要であることを示す．また，$\boldsymbol{q}_j^\top$ を第 j 行に持ち，サイズが $J \times K$ である行列 $\mathrm{Q} = (\boldsymbol{q}_1^\top, \boldsymbol{q}_2^\top, \ldots, \boldsymbol{q}_J^\top)^\top$ を Q 行列（Tatsuoka, 1985）と呼ぶ．

この $\boldsymbol{q}$ ベクトルに，アトリビュート習得パタンに加え，項目パラメタベクトル $\boldsymbol{\lambda}_j = (\lambda_{j0}, \lambda_{j1}, \ldots, \lambda_{j1\ldots K})^\top$ を用いて，LCDM の項目反応関数は，

$$P(X_{ij} = 1 | \boldsymbol{\lambda}_j, \boldsymbol{q}_j, \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l) = \frac{\exp(f(\boldsymbol{\lambda}_j, \boldsymbol{q}_j, \boldsymbol{\alpha}_i))}{1 + \exp(f(\boldsymbol{\lambda}_j, \boldsymbol{q}_j, \boldsymbol{\alpha}_i))} \tag{2.1}$$

と表される．ここで，$f(\boldsymbol{\lambda}_j, \boldsymbol{q}_j, \boldsymbol{\alpha}_i)$ は

$$\begin{aligned} f(\boldsymbol{\lambda}_j, \boldsymbol{q}_j, \boldsymbol{\alpha}_i) &= \log \frac{P(X_{ij} = 1 | \boldsymbol{\lambda}_j, \boldsymbol{q}_j, \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)}{1 - P(X_{ij} = 1 | \boldsymbol{\lambda}_j, \boldsymbol{q}_j, \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)} \\ &= \lambda_{j0} + \sum_{k=1}^{K} \lambda_{jk} q_{jk} \alpha_{ik} + \sum_{k=1}^{K} \sum_{k' < k} \lambda_{jkk'} q_{jk} q_{jk'} \alpha_{ik} \alpha_{ik'} + \cdots + \lambda_{j1\ldots K} \prod_{k=1}^{K} q_{jk} \alpha_{ik} \end{aligned} \tag{2.2}$$

と表される関数である．LCDM の項目パラメタベクトル $\boldsymbol{\lambda}_j$ は以下のパラメタを含んでいる．まず，λ_{j0} は切片パラメタであり，切片パラメタは当該の項目に必要なアトリビュートを全く習得していない場合のベースラインの正答確率を規定している．次に，λ_{jk} はアトリビュート k の主効果を表し，あるアトリビュートの正答確率への単独の効果を規定するパラメタである．さらに，$\lambda_{jkk'}$ はアトリビュート k と $k'(\neq k)$ の間の 1 次の交互作用パラメタおよび $\lambda_{j1\ldots K}$ は K 個のアトリビュートの間の $K-1$ 次の交互作用パラメタであり，複数のアトリビュートを同時に習得している場合の効果を表している．実際には，すべての項目について 2^K 個のパラメタが想定されるわけではなく，$q_{jk} = 0$ を含む項の λ パラメタは推定できないため，項目 j のパラメタの個数は $2^{\sum_{k=1}^{K} q_{jk}} \leq 2^K$ である．

例として，$\boldsymbol{q}_j = (1, 1, 0)^\top, K = 3$ という項目に対する，LCDM の項目反応関数を具体的に書くことにする．まず，式(2.2)から，LCDM の $f(\boldsymbol{\lambda}_j, \boldsymbol{q}_j, \boldsymbol{\alpha}_i)$ 関数は

$$f(\boldsymbol{\lambda}_j, \boldsymbol{q}_j, \boldsymbol{\alpha}_i) = \lambda_{j0} + \sum_{k=1}^{3} \lambda_{jk} q_{jk} \alpha_{ik} + \sum_{k=1}^{3} \sum_{k' < k} \lambda_{jkk'} q_{jk} q_{jk'} \alpha_{ik} \alpha_{ik'} + \lambda_{j123} \prod_{k=1}^{3} q_{jk} \alpha_{ik} \tag{2.3}$$

である．ここに，$q_{j1} = q_{j2} = 1$ および $q_{j3} = 0$ を代入すると，項目反応関数は

$$P(X_{ij}=1|\boldsymbol{\lambda}_j,\boldsymbol{q}_j=(1,1,0)^\top,\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l)=\frac{\exp(\lambda_{j0}+\lambda_{j1}\alpha_{i1}+\lambda_{j2}\alpha_{i2}+\lambda_{j12}\alpha_{i1}\alpha_{i2})}{1+\exp(\lambda_{j0}+\lambda_{j1}\alpha_{i1}+\lambda_{j2}\alpha_{i2}+\lambda_{j12}\alpha_{i1}\alpha_{i2})} \tag{2.4}$$

と縮約した形式で記述できる．この項目 j の項目パラメタの個数は $2^2=4$ となる．

さらに，ここで LCDM のモデルパラメタの数値例を示す．Templin and Hoffman (2013) は The Examination for the Certificate of Proficiency in English (ECPE) という英語力を測定するテストに LDCM を適用した例を示している．ECPE データには 28 の多枝選択項目が含まれており，アトリビュート数は $K=3$ で，1. 形態的統語的規則(morphosyntactic rules)，2. 一貫性規則(cohesive rules)，3. 語彙規則(lexical rules)が想定されていた．また，分析のためのサンプルサイズは 2,922 であった．ECPE データの 1 番目の項目には，形態的統語的規則および一貫性規則が必要とされているため，$\boldsymbol{q}_1=(1,1,0)^\top$ であった．Templin and Hoffman (2013) の Table 1 から，項目 1 の項目パラメタの推定値は，$\lambda_{10}=0.835$，$\lambda_{11}=0.000$，$\lambda_{12}=0.600$ および $\lambda_{112}=1.222$ であった．これらの結果から，例えば，$\lambda_{11}=0.000$ であるため，形態的統語的規則の単独の習得は項目の正答に寄与しないこと，$\lambda_{112}=1.222$ であることから形態的統語的規則と一貫性規則のどちらも習得していることによる交互作用効果が項目の正答に寄与しているといったことがわかる．

LCDM は項目パラメタに制約をかけることによって，さまざまな DCM のサブモデルを表現できる．例えば，最も基本的な DCM である DINA モデルは切片と当該項目で測定しているアトリビュートの添字集合 $\mathcal{K}=\{k;q_{jk}=1,k=1,2,\ldots,K\}$ についての最高次の交互作用パラメタのみを含めて，

$$f(\boldsymbol{\lambda}_j,\boldsymbol{q}_j,\boldsymbol{\alpha}_i)=\lambda_{j0}+\lambda_{j\mathcal{K}}\prod_{k\in\mathcal{K}}\alpha_{ik} \tag{2.5}$$

と表される．DINA モデルでは，2 つの項目パラメタを

$$g_j=\frac{\exp(\lambda_{j0})}{1+\exp(\lambda_{j0})}, \tag{2.6}$$

$$1-s_j=\frac{\exp(\lambda_{j0}+\lambda_{j\mathcal{K}}\prod_{k\in\mathcal{K}}\alpha_{ik})}{1+\exp(\lambda_{j0}+\lambda_{j\mathcal{K}}\prod_{k\in\mathcal{K}}\alpha_{ik})} \tag{2.7}$$

と表す．ここで，g_j は当て推量(guessing)パラメタ，s_j はスリップ(slipping)パラメタと呼ばれる．すなわち，g_j は項目 j の正答に必要なアトリビュートのうち，少なくとも一つが未習得であるアトリビュート習得パタンのテスト解答者が当該項目に正答する確率であり，s_j は項目 j の正答に必要なアトリビュートをすべて習得しているアトリビュート習得パタンのテスト解答者が当該項目に誤答する確率である． また，切片と主効果項のみを含めたモデルは compensatory reduced unified model (C-RUM; Rupp et al., 2010) と呼ばれ，

$$f(\boldsymbol{\lambda}_j,\boldsymbol{q}_j,\boldsymbol{\alpha}_i)=\lambda_{j0}+\sum_{k=1}^{K}\lambda_{jk}q_{jk}\alpha_{ik} \tag{2.8}$$

と表される．

さて，LCDM においては，$\boldsymbol{\lambda}$ がモデルパラメタ，$\boldsymbol{\alpha}$ がカテゴリカルで多次元の潜在変数とみなすことができる．この意味では，LDCM はで連続的な多次元の潜在変数を持つ多次元項目反応理論モデル(multidimensional item response theory models) と関係があるモデルとしても考えることができる．しかし，$\boldsymbol{\lambda}$ ではなく，アトリビュート習得パタンごとの項目反応確率をモデルのパラメタとみなすことで，本稿の主眼であるパラメタ推定法を表現しやすくなる．さきほどの $\boldsymbol{q}_j=(1,1,0)^\top$ という項目について，4 つの(条件付き)正答反応確率パラメタを $0\le\theta_{jh}\le 1(h=1,2,3,4)$ と書くとすると，

$$\theta_{j1} = \frac{\exp(\lambda_{j0})}{1+\exp(\lambda_{j0})}, \tag{2.9}$$

$$\theta_{j2} = \frac{\exp(\lambda_{j0}+\lambda_{j1})}{1+\exp(\lambda_{j0}+\lambda_{j1})}, \tag{2.10}$$

$$\theta_{j3} = \frac{\exp(\lambda_{j0}+\lambda_{j2})}{1+\exp(\lambda_{j0}+\lambda_{j2})}, \tag{2.11}$$

$$\theta_{j4} = \frac{\exp(\lambda_{j0}+\lambda_{j1}+\lambda_{j2}+\lambda_{j12})}{1+\exp(\lambda_{j0}+\lambda_{j1}+\lambda_{j2}+\lambda_{j12})} \tag{2.12}$$

となる．ここで，θ_{j1} はアトリビュート習得パタン $(0,0,0)^\top$ と $(0,0,1)^\top$ の正答確率に対応するパラメタと考えることができる．同様に，$\theta_{j2}, \theta_{j3}, \theta_{j4}$ はそれぞれ，$(1,0,0)^\top$ と $(1,0,1)^\top$，$(0,1,0)^\top$ と $(0,1,1)^\top$，$(1,1,0)^\top$ と $(1,1,1)^\top$ に対応する正答確率パラメタを表している．このことから，DCM はアトリビュート習得パタンを潜在クラスとみなした制約付き潜在クラスモデルと考えることができる．

上記の例からわかるように，一般的な DCM において，$q_{jk}=0$ は項目 j においてアトリビュート k の習得の有無の情報を区別することができないことを意味している．この意味で，$\boldsymbol{q}$ ベクトルはアトリビュート習得パタンを項目特有のアトリビュート習得パタンに縮約する役割を持っていると考えることができる．ここで，$\boldsymbol{q}$ ベクトルが区別できるアトリビュート習得パタンを $\boldsymbol{\alpha}^*_{jh}$ と書くことにする．$\boldsymbol{q}_j=(1,1,0)^\top$ という $\boldsymbol{q}$ ベクトルから，$\boldsymbol{\alpha}^*_{j1}=(0,0,*)^\top$，$\boldsymbol{\alpha}^*_{j2}=(0,1,*)^\top$，$\boldsymbol{\alpha}^*_{j3}=(1,0,*)^\top$，$\boldsymbol{\alpha}^*_{j4}=(1,1,*)^\top$ という 4 種類のアトリビュート習得パタンが区別できることがわかる．“$*$” は，そのアトリビュートの習得の有無を無視することを意味する．ここから，ある項目 j において同じ正答確率を持つパタンを示す

$$\mathcal{I}(\boldsymbol{\alpha}^*_{jh}=\boldsymbol{\alpha}_l) = \begin{cases} 1 & \alpha^*_{jhk}=\alpha_{lk}, \forall k \in \{k; q_{jk}=1, k=1,\ldots,K\} \\ 0 & \text{Otherwise} \end{cases} \tag{2.13}$$

という指示関数 $\mathcal{I}(\cdot)$ を導入する．この記号と条件付き正答確率パラメタ $\boldsymbol{\theta}_j=\{\theta_{jh}\}_{h=1}^{H_j}, H_j=2^{\sum_{k=1}^K q_{jk}}$ をもちいることで，DCM の条件付き尤度関数は

$$P(X_{ij}=x_{ij}|\boldsymbol{\theta}_j, \boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l) = \prod_{h=1}^{H_j}\prod_{l=1}^{L}[\theta_{jh}^{x_{ij}}(1-\theta_{jh})^{1-x_{ij}}]^{\mathcal{I}(\boldsymbol{\alpha}^*_{jh}=\boldsymbol{\alpha}_l)\mathcal{I}(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l)} \tag{2.14}$$

と書くことができる．ただし，$\mathcal{I}(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l)$ は $\boldsymbol{\alpha}_i$ と $\boldsymbol{\alpha}_l$ の要素がすべて等しい場合に 1，そうでなければ 0 をとる関数として，先に導入した指示関数とは引数の違いで区別する．式(2.14)から，テスト解答者 i の項目反応パタンベクトル $\boldsymbol{X}_i=(X_{i1}, X_{i2}, \ldots, X_{iJ})^\top$ の実現値 $\boldsymbol{x}_i=(x_{i1}, x_{i2}, \ldots, x_{iJ})^\top$ を観測したときのパラメタの尤度は

$$P(\boldsymbol{X}_i=\boldsymbol{x}_i|\boldsymbol{\Theta}, \boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l) = \prod_{j=1}^{J}\prod_{h=1}^{H_j}\prod_{l=1}^{L}[\theta_{jh}^{x_{ij}}(1-\theta_{jh})^{1-x_{ij}}]^{\mathcal{I}(\boldsymbol{\alpha}^*_{jh}=\boldsymbol{\alpha}_l)\mathcal{I}(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l)} \tag{2.15}$$

となる．ただし，$\boldsymbol{\Theta}=\{\boldsymbol{\theta}_j\}_{j=1}^J$ は項目パラメタの集合である．さらに，$\boldsymbol{\alpha}_i$ が混合比率パラメタ $\boldsymbol{\pi}=(\pi_1, \pi_2, \ldots, \pi_L)^\top$（ただし，$0 \le \pi_l \le 1, \sum_l \pi_l = 1$）であるカテゴリカル分布

$$P(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l|\boldsymbol{\pi}) = \prod_{l=1}^{L}\pi_l^{\mathcal{I}(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l)} \tag{2.16}$$

に従うとする．式(2.15)および(2.16)から，テスト解答者 i の項目反応パタンベクトルを得たときの周辺尤度は

$$
\begin{aligned}
(2.17) \qquad P(\boldsymbol{X}_i = \boldsymbol{x}_i|\boldsymbol{\Theta}, \boldsymbol{\pi}) &= \sum_{l=1}^{L} P(\boldsymbol{X}_i = \boldsymbol{x}_i|\boldsymbol{\Theta}, \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l) P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l|\boldsymbol{\pi}) \\
&= \sum_{l=1}^{L} \left\{ \pi_l \prod_{j=1}^{J} \prod_{h=1}^{H_j} [\theta_{jh}^{x_{ij}} (1-\theta_{jh})^{1-x_{ij}}]^{\mathcal{I}(\boldsymbol{\alpha}_{jh}^* = \boldsymbol{\alpha}_l)} \right\}
\end{aligned}
$$

と書くことができる．さらに，すべてのテスト解答者が独立にサンプリングされているとし，項目反応行列を $\mathrm{X} = (\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top, \ldots, \boldsymbol{x}_I^\top)^\top$ とすると，周辺尤度関数は

$$
\begin{aligned}
(2.18) \qquad L(\boldsymbol{\Theta}, \boldsymbol{\pi}) = P(\mathrm{X}|\boldsymbol{\Theta}, \boldsymbol{\pi}) &= \prod_{i=1}^{I} P(\boldsymbol{X}_i = \boldsymbol{x}_i|\boldsymbol{\Theta}, \boldsymbol{\pi}) \\
&= \prod_{i=1}^{I} \sum_{l=1}^{L} \left\{ \pi_l \prod_{j=1}^{J} \prod_{h=1}^{H_j} [\theta_{jh}^{x_{ij}} (1-\theta_{jh})^{1-x_{ij}}]^{\mathcal{I}(\boldsymbol{\alpha}_{jh}^* = \boldsymbol{\alpha}_l)} \right\}
\end{aligned}
$$

となる．DCM のモデルパラメタの推定は上記の周辺尤度関数を用いる方法が基本となる．

具体的な推定法について記述を行う前に，DCM を利用する際の一般的な注意点を述べる．まず，Q 行列はテストで測定したいアトリビュートについての専門知識と項目の分析から設定される．ただし，Q 行列の誤特定は，アトリビュート習得パタンの推定にバイアスを生じさせることが知られているため（e.g., Rupp and Templin, 2008a），慎重に設定する必要がある．そのため，項目反応データから当該のテストの Q 行列をデータから推定する方法も近年活発に開発されている（e.g., Chen et al., 2018）．さらに，Q 行列の設定はパラメタの識別性（identification; Xu and Shang, 2018）を規定するため重要である．また，アトリビュートの間の習得の順序関係を表すアトリビュート階層構造（attribute hierarchy structure; e.g., Leighton et al., 2004）を想定する場合には，アトリビュート習得パタンの数は 2^K よりも少なくなる．本稿のモデルの定式化はアトリビュートの間の関係を特段仮定しないものであった．これをもとに，モデルのパラメタ $\{\boldsymbol{\Theta}, \boldsymbol{\pi}\}$ の推定法を示す．

2.2 単調性制約

単調性制約は，あるアトリビュート習得パタンにおける項目の正答確率よりも，そのアトリビュート習得パタンで未習得のアトリビュートを習得している別の習得パタンの正答確率のほうが高いか少なくとも等しい，という制約である．前述の $\boldsymbol{q}_j = (1,1,0)^\top$ という項目のパラメタの具体的な単調性制約は以下のようになる．式(2.9)から(2.12)で表される正答確率は，それぞれ $\boldsymbol{\alpha}_{j1}^* = (0,0,*)^\top$，$\boldsymbol{\alpha}_{j2}^* = (1,0,*)^\top$，$\boldsymbol{\alpha}_{j3}^* = (0,1,*)^\top$ および $\boldsymbol{\alpha}_{j4}^* = (1,1,*)^\top$ という習得パタンに対する正答確率であるため

$$
\begin{aligned}
0 \le \theta_{j1} \le \theta_{j2} \le \theta_{j4} \le 1, \\
0 \le \theta_{j1} \le \theta_{j3} \le \theta_{j4} \le 1
\end{aligned}
$$

という順序関係を満足する必要があり，このパラメタの順序関係が単調性制約である．単調性制約から，当該の項目の正答に関連するアトリビュートを一つも習得していないパタン（ここでは $\boldsymbol{\alpha}_{j1}^* = (0,0,*)^\top$）は最も低い正答確率であり，すべてのアトリビュートを習得しているパタン（$\boldsymbol{\alpha}_{j4}^* = (1,1,*)^\top$）は最も高い正答確率になることがわかる．

単調性を形式的に表現するために，アトリビュート習得パタンの「順序」を以下のように定義する．すべての $k = 1, \ldots, K$ に対して，$\alpha_{lk} \ge \alpha_{l'k}$ であれば $\boldsymbol{\alpha}_l \succeq \boldsymbol{\alpha}_{l'}$ とし，$\boldsymbol{\alpha}_l \succeq \boldsymbol{\alpha}_{l'}$ を満足し，かつ，いくつかの k に対して $\alpha_{lk} > \alpha_{l'k}$ であれば $\boldsymbol{\alpha}_l \succ \boldsymbol{\alpha}_{l'}$ とする．$\boldsymbol{\alpha}_{jh}^*$ についても同様に順序を考えることとする．このとき，上記の $\boldsymbol{q}_j = (1,1,0)^\top$ に対して

$$\boldsymbol{\alpha}^*_{j4} \succ \boldsymbol{\alpha}^*_{j2} \succ \boldsymbol{\alpha}^*_{j1},$$
$$\boldsymbol{\alpha}^*_{j4} \succ \boldsymbol{\alpha}^*_{j3} \succ \boldsymbol{\alpha}^*_{j1}$$

である．この表記を用いることで，単調性制約は $\boldsymbol{\alpha}^*_{jh} \succ \boldsymbol{\alpha}^*_{jh'}$ であれば，$\theta_{jh} \geq \theta_{jh'}$ を満足することと表現できる．

上記の単調性制約はやや強い関係であり，理論研究においては Xu (2017) や Xu and Shang (2018) による

$$\max_{\boldsymbol{\alpha};\boldsymbol{\alpha}\succeq\boldsymbol{q}_j} \theta_{j,\boldsymbol{\alpha}} = \min_{\boldsymbol{\alpha};\boldsymbol{\alpha}\succeq\boldsymbol{q}_j} \theta_{j,\boldsymbol{\alpha}} \geq \theta_{j,\boldsymbol{\alpha}'} \geq \theta_{j,\mathbf{0}_K}, \forall\boldsymbol{\alpha}' \not\succeq \boldsymbol{q}_j, \tag{2.19}$$

という単調性の定義を用いることもある．この定義においては，ある項目 j で測定されているアトリビュートをすべて測定しているアトリビュート習得パタンの正答確率が全て同じで最も高く，また，当該項目の正答に関係するアトリビュートを一つも習得していないパタンが最も低い正答確率になり，それ以外のアトリビュート習得パタン($\boldsymbol{\alpha}'$ であらわされる)の正答確率はそれらの 2 つの正答確率の間であるという関係を表している．ここで，$\boldsymbol{\alpha}'$ に対応する複数のアトリビュート習得パタンの間の正答確率の間には，大小関係を想定していない点に注意が必要である．後述の単調性を満足するためのアルゴリズムは，前述のより厳しい単調性を満足するものであるが，こちらのややゆるい単調性を満足するように変更することも容易にできる．

3. 最尤推定法

本章では，モデルの尤度を用いた推定法として，同時最尤推定法(joint likelihood estimation method)，周辺最尤推定法(marginal likelihood estimation method)，正則化推定法を示す．周辺最尤推定法については，EM アルゴリズムを用いた方法を示し，パラメタの推定の標準誤差の推定法および単調性制約を満足する推定値を得るための方法も示す．

3.1 同時最尤推定法

DCM においては，アトリビュート習得パタン $\boldsymbol{\alpha}_i$ が潜在変数であり，テスト解答者のアトリビュート習得パタンベクトルが観測されていれば，$\mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)$ が計算でき，完全データの尤度を構成できる．後述の周辺最尤推定法は，尤度から潜在変数を周辺化して消去するため，アトリビュート数が大きくなったりサンプルサイズが大きくなった場合に計算の負荷が大きくなる傾向がある．一方，同時最尤推定法は単純である完全データの尤度を用いるため，周辺最尤推定法と比較して推定の計算アルゴリズムが簡単になり，効率的な推定が可能である．しかし，同時最尤推定法は一般にパラメタ推定値に一致性がないことから，潜在変数モデルにおいてはあまり用いられてきていない方法である．Chiu et al. (2016) は，外部で推定された一致性のあるアトリビュート習得パタンを初期値として用いることにより，同時最尤推定でも一致性のある項目パラメタを得ることができることを証明し，この問題を解決した．これにより，効率的な計算アルゴリズムを活用して統計的に望ましい性質をもった推定値を得ることができるため，大規模データやアトリビュート数が多い状況で DCM を利用しやすくなったと考えられる．

同時最尤推定法の具体的な方法を以下に示す．式(2.16)，(2.15)から，完全データの尤度は

$$L(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}) = P(\mathrm{X}, \boldsymbol{\mathcal{A}}|\boldsymbol{\Theta}, \boldsymbol{\pi}) = \prod_{i=1}^{I}\prod_{l=1}^{L}\left\{\pi_l \prod_{j=1}^{J}\prod_{h=1}^{H_j}[\theta_{jh}^{x_{ij}}(1-\theta_{jh})^{1-x_{ij}}]^{\mathcal{I}(\boldsymbol{\alpha}^*_{jh}=\boldsymbol{\alpha}_l)}\right\}^{\mathcal{I}(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l)} \tag{3.1}$$

である．ただし，I 人分のアトリビュート習得パタンの集合を $\boldsymbol{\mathcal{A}} = \{\boldsymbol{\alpha}_i\}_{i=1}^{I}$ とした．同時最尤

推定法は

$$\{\hat{\boldsymbol{\mathcal{A}}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}\} = \underset{\boldsymbol{\mathcal{A}},\boldsymbol{\Theta},\boldsymbol{\pi}}{\operatorname{argmax}}\{\log L(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})\} \tag{3.2}$$

と式(3.1)の完全データの尤度あるいはその対数尤度を最大化する潜在変数 $\hat{\boldsymbol{\mathcal{A}}}$ およびモデルパラメタ $\{\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}\}$ を推定値とする方法である．

同時最尤推定法においては，モデルパラメタとテスト解答者パラメタの更新を反復する方法で完全データの尤度を最大化することで，パラメタの推定値を得ることができる．より具体的には，まず後述のノンパラメトリック推定法で得られたアトリビュート習得パタンの一致推定値を初期値とし，アトリビュート習得パタンを既知としたもとで，項目ごとの尤度(式(2.14)をテスト解答者について積をとることで得られる尤度)を最大化する項目パラメタを求める．また，混合比率パラメタは式(2.16)を用いて更新する．さらに，これらのパラメタを既知として，式(2.15)を最大化するアトリビュート習得パタンを得る．これらの項目パラメタとアトリビュート習得パタンの更新を適当な収束基準を満足するまで反復する，というのが同時最尤推定法のパラメタ推定アルゴリズムである．Chiu et al. (2016)は，アトリビュート習得パタンが既知の状況で，LCDM パラメタを切片パラメタから高次の交互作用に向けて再帰的に更新する比較的単純なパラメタ更新規則を示した．

上記のパラメタ更新アルゴリズムにおいては，モデルパラメタの更新とテスト解答者パラメタの更新の際に，項目ごとあるいはテスト解答者ごとのパラメタに注目した更新が可能であり，更新の並列化が可能となる．このように，効率的なパラメタの更新アルゴリズムを利用することができる点が同時最尤推定法の利点である．Chiu et al. (2016)は，シミュレーションにより，同時最尤推定法のほうが周辺最尤推定法よりも高速にパラメタ推定が完了することと，同時最尤推定法の推定値が一致的であることを確認した．ただし，Chiu et al. (2016)の推定アルゴリズムはパラメタの単調性を満足するものではない．しかし，その推定アルゴリズムにおいて，項目パラメタの更新を行う際に，後述する周辺最尤推定法で述べる EM アルゴリズムの M ステップにおいて単調性を満足するための最適化法が利用できる．

3.2 周辺最尤推定法

周辺最尤推定法は，欠測データを含むデータ分析や潜在変数モデルのパラメタ推定法として標準的に用いられる方法であり，DCM のパラメタ推定においても基礎的な方法として位置づけられる．式(2.18)の周辺尤度と式(3.1)の完全データの尤度の間には

$$L(\boldsymbol{\Theta}, \boldsymbol{\pi}) = \sum_{\boldsymbol{\mathcal{A}}} L(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}) \tag{3.3}$$

という関係が成り立つ．ここで $\sum_{\boldsymbol{\mathcal{A}}}$ は $\boldsymbol{\mathcal{A}}$ が取りうるすべてのパタンについての和を取ることを意味する．実際には，局所独立の仮定があるため，各テスト解答者について L 個のアトリビュート習得パタンごとに式(2.14)を用いて式(2.15)を評価することで，周辺尤度の値を評価する．

周辺最尤推定法は上記の周辺尤度を用いて，

$$\{\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}\} = \underset{\boldsymbol{\Theta},\boldsymbol{\pi}}{\operatorname{argmax}}\{\log L(\boldsymbol{\Theta}, \boldsymbol{\pi})\} = \underset{\boldsymbol{\Theta},\boldsymbol{\pi}}{\operatorname{argmax}}\left\{\log \sum_{\boldsymbol{\mathcal{A}}} L(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})\right\} \tag{3.4}$$

と周辺尤度を最大化するモデルパラメタを推定量とする方法である．周辺最尤推定法におけるモデルパラメタは，最尤推定法の一般論から一致的である．周辺最尤推定法においてはアトリビュート習得パタンが周辺化されているため，モデルパラメタの推定値 $\{\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}\}$ を用いて事後

的に推定される．

このため，アトリビュート習得パタン $\hat{\boldsymbol{\mathcal{A}}} = \{\hat{\boldsymbol{\alpha}}_i\}_{i=1}^I$ の推定法にはいくつかの選択肢がある．まず，式(2.15)の条件付き尤度を用いた方法として，テスト解答者 i の項目反応パタン $\boldsymbol{x}_i$ から

$$\hat{\boldsymbol{\alpha}}_i^{ML} = \underset{\boldsymbol{\alpha}_i}{\operatorname{argmax}}\{P(\boldsymbol{X}_i = \boldsymbol{x}_i|\hat{\boldsymbol{\Theta}}, \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)\} \tag{3.5}$$

はアトリビュート習得パタンの最尤推定値である．

アトリビュート習得パタンの事後確率を

$$\hat{\boldsymbol{\alpha}}_i^{MAP} = \underset{\boldsymbol{\alpha}_i}{\operatorname{argmax}}\{P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l|\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}})\} \tag{3.6}$$

と最大化するアトリビュート習得パタンが事後確率最大化(MAP)推定値である．ここで，$P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l|\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}})$ はテスト解答者 i のアトリビュート習得パタン $\boldsymbol{\alpha}_i$ が $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l$ である事後確率であり，

$$\begin{aligned} P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l|\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}) &= \frac{P(\boldsymbol{X}_i = \boldsymbol{x}_i|\hat{\boldsymbol{\Theta}}, \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l|\hat{\boldsymbol{\pi}})}{\sum_{l'=1}^L P(\boldsymbol{X}_i = \boldsymbol{x}_i|\hat{\boldsymbol{\Theta}}, \boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{l'})P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{l'}|\hat{\boldsymbol{\pi}})} \\ &= \frac{\hat{\pi}_l \prod_{j=1}^J \prod_{h=1}^{H_j} [\hat{\theta}_{jh}^{x_{ij}}(1-\hat{\theta}_{jh})^{1-x_{ij}}]^{\mathcal{I}(\boldsymbol{\alpha}_{jh}^* = \boldsymbol{\alpha}_l)}}{\sum_{l'=1}^L \{\hat{\pi}_{l'} \prod_{j=1}^J \prod_{h=1}^{H_j} [\hat{\theta}_{jh}^{x_{ij}}(1-\hat{\theta}_{jh})^{1-x_{ij}}]^{\mathcal{I}(\boldsymbol{\alpha}_{jh}^* = \boldsymbol{\alpha}_{l'})}\}} \end{aligned} \tag{3.7}$$

と計算される．期待事後推定(expected-a-posteriori[EAP])法は，各テスト解答者の各アトリビュート習得の式(3.7)による事後確率を用いた事後期待値を用いる．アトリビュート習得の事後期待値は

$$\begin{aligned} \mathbb{E}_{P(\boldsymbol{\alpha}_i|\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}})}[\alpha_{ik}] &= \sum_{l=1}^L \alpha_{lk} P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l|\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}) \\ &= P(\alpha_{ik} = \alpha_{lk}|\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}) \end{aligned} \tag{3.8}$$

である．ここで，$\mathbb{E}_{P(\boldsymbol{\alpha}_i|\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}})}[\cdot]$ はアトリビュート習得パタンの事後確率による期待値である．これはアトリビュート習得確率としても解釈でき，アトリビュート習得確率を DCM の結果として報告することもある．ここから，テスト解答者 i の k 番目のアトリビュート習得の EAP 推定値は，

$$\alpha_{ik}^{EAP} = \begin{cases} 1 & P(\alpha_{ik} = \alpha_{lk}|\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}) \geq 0.5 \\ 0 & P(\alpha_{ik} = \alpha_{lk}|\boldsymbol{x}_i, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}) < 0.5 \end{cases} \tag{3.9}$$

であり，アトリビュート習得パタンの EAP は $\boldsymbol{\alpha}_i^{EAP} = (\alpha_{i1}^{EAP}, \ldots, \alpha_{ik}^{EAP}, \ldots, \alpha_{iK}^{EAP})^\top$ である．EAP 推定を行う場合には，閾値を設定する必要があるが，ここでは 0.5 とした．

3.2.1 EM アルゴリズム

周辺尤度の最大化を実行するためのアルゴリズムとして，EM アルゴリズムが利用できる．EM アルゴリズムを用いた周辺尤度の最大化による最尤推定値を得る方法を MML-EM (marginal maximum likelihood with EM algorithm)と呼ぶ．EM アルゴリズムは，まずモデルパラメタの適当な初期値 $\{\boldsymbol{\Theta}^{(0)}, \boldsymbol{\pi}^{(0)}\}$ を定め，以下の E-step (expectation step)と M-step (maximization step)を反復して，対数周辺尤度を最大化するパラメタ値を求める．本節では上付き添字 $(t), (t = 1, 2, \ldots, T)$ をアルゴリズムの反復時点を表すものとする．

まず，すべてのテスト解答者 i について，アトリビュート習得パタン $\boldsymbol{\alpha}_i$ が $\boldsymbol{\alpha}_l$ である事後分布による期待値 $\gamma_{il}^{(t)} = \mathbb{E}_{P(\boldsymbol{\alpha}_i|\boldsymbol{x}_i, \boldsymbol{\Theta}^{(t-1)}, \boldsymbol{\pi}^{(t-1)})}[\mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)]$ を求める．これは，式(3.7)で計算さ

れる事後確率に等しい．

次に式(3.1)の完全データの尤度を用いて，t 回目の反復において M-step で最適化する関数 $Q(\{\boldsymbol{\Theta},\boldsymbol{\pi}\}|\{\boldsymbol{\Theta}^{(t-1)},\boldsymbol{\pi}^{(t-1)}\})$ は

$$
\begin{aligned}
(3.10)\qquad Q(\{\boldsymbol{\Theta},\boldsymbol{\pi}\}|\{\boldsymbol{\Theta}^{(t)},\boldsymbol{\pi}^{(t)}\}) &= \mathbb{E}_{P(\boldsymbol{\mathcal{A}}|\boldsymbol{X},\boldsymbol{\Theta}^{(t)},\boldsymbol{\pi}^{(t)})}[\log L(\boldsymbol{\mathcal{A}},\boldsymbol{\Theta},\boldsymbol{\pi})] \\
&= \sum_{i=1}^{I}\sum_{l=1}^{L}\left[\gamma_{il}^{(t)}\left\{\log \pi_l + \sum_{j=1}^{J}\sum_{h=1}^{H_j}\mathcal{I}(\boldsymbol{\alpha}_{jh}^{*}=\boldsymbol{\alpha}_l)(x_{ij}\log\theta_{jh}+(1-x_{ij})\log(1-\theta_{jh}))\right\}\right]
\end{aligned}
$$

と計算される．ここで，$\mathbb{E}_{P(\boldsymbol{\mathcal{A}}|\boldsymbol{X},\boldsymbol{\Theta}^{(t-1)},\boldsymbol{\pi}^{(t-1)})}[\cdot]$ はデータ X 及びパラメタ $\boldsymbol{\Theta}^{(t-1)},\boldsymbol{\pi}^{(t-1)}$ を所与としたときの，アトリビュート習得パタン集合 $\boldsymbol{\mathcal{A}}$ の事後確率による期待値である．

$Q(\{\boldsymbol{\Theta},\boldsymbol{\pi}\}|\{\boldsymbol{\Theta}^{(t-1)},\boldsymbol{\pi}^{(t-1)}\})$ を各パラメタについて偏微分し，$\sum_l \pi_l = 1$ という条件に注意しつつ，0 とおいた方程式を解くことで，M-step における $\theta_{jh}^{(t)}$ と $\pi_l^{(t)}$ の更新式

$$
(3.11)\qquad \theta_{jh}^{(t)} = \frac{\sum_{i=1}^{I}\sum_{l=1}^{L}\mathcal{I}(\boldsymbol{\alpha}_{jh}^{*}=\boldsymbol{\alpha}_l)\gamma_{il}x_{ij}}{\sum_{i=1}^{I}\sum_{l=1}^{L}\mathcal{I}(\boldsymbol{\alpha}_{jh}^{*}=\boldsymbol{\alpha}_l)\gamma_{il}},
$$

$$
(3.12)\qquad \pi_l^{(t)} = \frac{\sum_{i=1}^{I}\gamma_{il}}{I}
$$

を得る．更新されたパラメタを用いて，E-step として新たに $\gamma_{il}^{(t+1)}$ を計算することで Q 関数を更新し，さらにモデルパラメタを更新するというステップを適当な収束基準を満足するまで繰り返す．なお，この更新式においては，正答確率パラメタの単調性制約を満たしていない点に注意が必要である．

3.2.2 単調性制約を満足するためのアルゴリズム

Hong et al. (2016) は単調性制約を満足するためのアルゴリズムとして，Upward アルゴリズムと Downward アルゴリズムの 2 つを提案している．これは，EM アルゴリズムなどの最適化の過程において単調性を犯す推定値が得られてしまった場合に，それらを補正するアルゴリズムである．まず，Upward アルゴリズムは，M-step の際に，$\boldsymbol{\alpha}_{jh}^{*} \succ \boldsymbol{\alpha}_{jh'}^{*}$ に対して $\theta_{jh} < \theta_{jh'}$ となった場合に，θ_{jh} を

$$
(3.13)\qquad \theta_{jh}^{(t)} = \max\{\theta_{jh'}^{(t)}|\boldsymbol{\alpha}_{jh}^{*} \succ \boldsymbol{\alpha}_{jh'}^{*}\}
$$

とする操作を導入することで単調性を満足するように推定を行う．Downward アルゴリズムは，逆に，$\boldsymbol{\alpha}_{jh}^{*} \succ \boldsymbol{\alpha}_{jh'}^{*}$ に対して $\theta_{jh} < \theta_{jh'}$ となった場合に，$\theta_{jh'}$ を

$$
(3.14)\qquad \theta_{jh'}^{(t)} = \min\{\theta_{jh}^{(t)}|\boldsymbol{\alpha}_{jh}^{*} \succ \boldsymbol{\alpha}_{jh'}^{*}\}
$$

とする操作を用いる．

この他，Ma and Jiang (2021) は単調性制約を満足するために，制約を表す行列 $\boldsymbol{C}$ を導入し，

$$
(3.15)\qquad \boldsymbol{C}\boldsymbol{\theta}_j = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \times \begin{pmatrix} \theta_{j1} \\ \theta_{j2} \\ \theta_{j3} \\ \theta_{j4} \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \boldsymbol{0}_{H_j}
$$

というパラメタの差の非負値制約を課して，M-step の項目パラメタについての最適化を実行する方法を提案している．項目反応確率への直接的な単調性制約は，LCDM パラメタにおけるパラメタ制約に変換することができる．LCDM およびその下位モデルについての制約の具体的

な表現は Rupp et al.（2010）で詳細に解説されている．さらに，Rupp et al.（2010）や Templin and Hoffman（2013）では一般的な潜在変数モデルのパラメタ推定が可能な Mplus というソフトウェアを用いて単調性制約を満足する推定値を得るためモデルの記述方法を示している．

3.2.3 標準誤差の推定

標準誤差はフィッシャー情報行列を用いて推定される．Philipp et al.（2018）はフィッシャー情報行列 $\mathcal{I}_{\boldsymbol{\vartheta}}$（ただし，$\boldsymbol{\vartheta}$ は $\vartheta_p \in \{\boldsymbol{\Theta}, \boldsymbol{\pi}\}, p = 1, \ldots, P$ を要素とする長さが P のベクトル）の逆行列により，パラメタ推定値 $\hat{\boldsymbol{\vartheta}}$ の漸近分散共分散行列 $V_{\boldsymbol{\vartheta}} = \mathcal{I}_{\boldsymbol{\vartheta}}^{-1}$ を推定する方法を示した．パラメタの推定値の漸近標準誤差は，$V_{\boldsymbol{\vartheta}}$ の対角要素の平方根により推定される．さて，式（2.17）の尤度から得られるスコア関数 $\psi(\boldsymbol{\vartheta}; \boldsymbol{x}) = \left(\frac{\partial \log(P(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{\vartheta}))}{\boldsymbol{\vartheta}_1}, \ldots, \frac{\partial \log(P(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{\vartheta}))}{\boldsymbol{\vartheta}_p}, \ldots, \frac{\partial \log(P(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{\vartheta}))}{\boldsymbol{\vartheta}_P}\right)^\top$ から，フィッシャー情報行列は

$$\mathcal{I}_{\boldsymbol{\vartheta}} = E(\psi(\boldsymbol{\vartheta}; \boldsymbol{x})\psi(\boldsymbol{\vartheta}; \boldsymbol{x})^\top) \tag{3.16}$$

と定義される．ここで，期待値は式（2.17）の同時確率を用いて，すべての項目反応パタンについて取る．

このフィッシャー情報行列の推定法として，対数尤度の外積を用いた方法とヘシアンを用いた方法がよく用いられる．勾配の外積推定量は

$$\mathcal{I}_{\boldsymbol{\vartheta}} \approx \frac{1}{I} \sum_{i=1}^{I} \psi(\boldsymbol{\vartheta}; \boldsymbol{x}_i)\psi(\boldsymbol{\vartheta}; \boldsymbol{x}_i)^\top \bigg|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} = \mathcal{I}_{\boldsymbol{\vartheta}, OPG} \tag{3.17}$$

であり，対数尤度の負のヘシアンによる近似は

$$\mathcal{I}_{\boldsymbol{\vartheta}} \approx -\frac{1}{I} \sum_{i=1}^{I} \frac{\partial^2 \log(P(\boldsymbol{X} = \boldsymbol{x}_i|\boldsymbol{\vartheta}))}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \bigg|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} = \mathcal{I}_{\boldsymbol{\vartheta}, Hess} \tag{3.18}$$

である．これらの具体的な要素の計算方法については，Philipp et al.（2018）で示されている．また，フィッシャー情報行列として，構造パラメタ（$\boldsymbol{\pi}$）を考慮しない項目レベルでの情報行列が de la Torre（2009）と de la Torre（2011）によって提案されている．さらに，Liu et al.（2019）はモデルを誤設定したもとでのサンドイッチ型の漸近分散共分散行列を得る方法として，フィッシャー情報行列を

$$\mathcal{I}_{\boldsymbol{\vartheta}} \approx \mathcal{I}_{\boldsymbol{\vartheta}, SW} = \mathcal{I}_{\boldsymbol{\vartheta}, Hess}^{-1} \mathcal{I}_{\boldsymbol{\vartheta}, OPG} \mathcal{I}_{\boldsymbol{\vartheta}, Hess}^{-1} / I \tag{3.19}$$

とし，シミュレーションによりサンドイッチ型フィッシャー情報行列を用いた標準誤差の頑健性を示した．

また，Yamaguchi（2023b）は EM アルゴリズムの枠組みを活用した supplemental EM アルゴリズム（Cai, 2008; Meng and Rubin, 1991）を活用し，完全データのフィッシャー情報行列から観測フィッシャー情報行列を得る方法を提案した．Supplemental EM アルゴリズムはパラメタ推定後に追加の EM 計算が必要であり，計算量が増えるものの，完全データの情報行列は観測データの情報行列よりも計算がしやすく，実装が容易である．Supplemental EM アルゴリズムには，推定されたパラメタ $\hat{\boldsymbol{\vartheta}}$ の漸近分散共分散行列が

$$\mathcal{I}_{\boldsymbol{\vartheta}, Obs}^{-1}|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} = \mathcal{I}_{\boldsymbol{\vartheta}, Comp}^{-1}[\boldsymbol{I}_P - \Delta(\boldsymbol{\vartheta})]^{-1}|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}} \tag{3.20}$$

と書くことができることを利用する．ここで，$\mathcal{I}_{\boldsymbol{\vartheta}, Obs}$ は観測データの情報行列，$\mathcal{I}_{\boldsymbol{\vartheta}, Comp}$ は完全データの情報行列であり，$\boldsymbol{I}_P$ は $P \times P$ の単位行列，そして $\Delta(\boldsymbol{\vartheta})$ は EM 写像のヤコビアンである．$\Delta(\boldsymbol{\vartheta})$ の要素 $d_{ij}, i, j = 1, \ldots, P$ は 1 回の EM サイクルの出力の j 番目の要素 $M_j(\boldsymbol{\vartheta})$ を用いて，

$$d_{ij} = \frac{\partial M_j(\boldsymbol{\vartheta})}{\partial \vartheta_i} \tag{3.21}$$

と表すことができ，EM アルゴリズムを用いた数値微分により求めることができる（Jamshidian and Jennrich, 2000）．

3.3 正則化推定法

DCM のモデルパラメタは，Q 行列と項目反応関数を適切に仮定することで，潜在クラスとしてのアトリビュート習得パタンやアトリビュートの項目への影響やアトリビュート間の交互作用効果としての項目パラメタといったような解釈しやすいものになっていた．しかし，DCM は，場合によっては制約が強すぎてデータに適合しにくいことや，数多くの DCM からどのモデルを選択すればよいのか明確ではないことが問題となる．その一方，制約がない探索的な潜在クラスモデルは比較的制約が緩いためデータにモデル適合しやすいものの，解釈性が高いパラメタ推定値が得られるとは限らない．また，DCM を用いるためには Q 行列を適切に設定したり，項目反応関数を設定する必要があり，その設定はかならずしも容易ではない．実際，前述のように Q 行列に誤設定があることによって項目パラメタの推定値にバイアスがかかるといった問題が生じる．こうしたことから，制約ができるだけ少なく，DCM 的な推定結果が得られるモデルが応用上は使いやすいと考えられる．

このような目標を達成するために，Chen et al.（2017）は部分的に併合された潜在クラス（partially merged latent class）という考え方を提案し，DCM 的な解釈性の高いパラメタ推定と探索的潜在クラスモデルの柔軟性をもつ正則化潜在クラス分析（regularized latent class analysis）を提案した．DCM の項目反応関数（式（2.1））からは，アトリビュート習得パタンで条件づけた正答確率が同じになる複数のパタンがあることがわかる．これは，言い換えると，潜在クラスの一部が併合されていると見ることができるため，DCM は正則化潜在クラスモデルとみなすことができる．このようなモデルのパラメタ推定には正則化推定法が利用できる．正則化推定法はスパースモデリングなどの文脈で頻繁に利用される方法であり，（対数）周辺尤度の後に正則化項を追加した目的関数を最適化するパラメタを推定値とする推定法である．

Chen et al.（2017）はあるクラス l の項目 j に対する正答反応確率を θ_{jl} とする通常の潜在クラスモデルの周辺尤度

$$L(\boldsymbol{\Theta}, \boldsymbol{\pi}) = \prod_{i=1}^{I} \sum_{l=1}^{L} \left\{ \pi_l \prod_{j=1}^{J} \theta_{jl}^{x_{ij}} (1-\theta_{jl})^{1-x_{ij}} \right\} \tag{3.22}$$

に対して項目反応関数に対する正則化項 $\kappa_\xi(\boldsymbol{\Theta})$ を仮定し

$$\{\hat{\boldsymbol{\Theta}}^\xi, \hat{\boldsymbol{\pi}}^\xi\} = \underset{\boldsymbol{\Theta}, \boldsymbol{\pi}}{\operatorname{argmax}} \{\log L(\boldsymbol{\Theta}, \boldsymbol{\pi}) - I\kappa_\xi(\boldsymbol{\Theta})\} \tag{3.23}$$

をパラメタ推定値とする方法を開発した．正則化項 $\kappa_\xi(\boldsymbol{\Theta})$ により，同じ正答確率を示すクラスを推定することができる．DCM は前述のように，ある項目に対して同じ正答確率を示すアトリビュート習得パタンを想定するため，上記の正則化推定法により，通常の潜在クラスモデルから DCM と同じような推定を得ることができる．

正則化項 $\kappa_\xi(\boldsymbol{\Theta})$ は $\kappa_\xi(\boldsymbol{\Theta}) = \sum_{j=1}^{J} p_\xi(\boldsymbol{\theta}_j), \boldsymbol{\theta}_j = (\theta_{j1}, \ldots, \theta_{jL})$ と項目ごとの正則化項 $\kappa_\xi(\boldsymbol{\theta}_j)$ の和で表される．さらに，スパース性を決定するハイパーパラメタ p_ξ と項目反応確率パラメタの順序統計量 $\theta_{j(1)} \leq \theta_{j(2)} \leq \cdots \leq \theta_{j(L)}$ を用いて，項目ごとの正則化項は

$$p_\xi(\boldsymbol{\theta}_j) = \sum_{l=1}^{L-1} p_\xi^{SCAD}(\theta_{j(l+1)} - \theta_{j(l)}) \tag{3.24}$$

と表される．ここで，p_{ξ}^{SCAD} は

$$p_{\xi}^{SCAD}(x) = \begin{cases} \xi|x| & \text{if } |x| \leq \xi \\ -\left(\frac{|x|^2-2a\xi|x|+\xi^2}{2(a-1)}\right) & \text{if } \xi < |x| < a\xi \\ \frac{(a+1)^2\xi^2}{2} & \text{if } |x| \geq a\xi \end{cases} \tag{3.25}$$

と定義される．また，スパース性を決めるパラメタの決定方法も Chen et al. (2017) で示されている．適当な条件のもと，この正則化項を用いた推定により，一致的なパラメタ推定とモデル選択ができることが示されている（Chen et al., 2017, pp. 668–670）．

また，正則化推定法はアトリビュート階層構造の推定（Wang and Lu, 2021; Ma et al., 2023b）や縦断的 DCM（e.g., Yamaguchi and Martinez, 2024）におけるアトリビュート習得パタンの遷移で表される学習軌跡の推定（Wang, 2021），Q 行列の推定（Chen et al., 2020, 2015; Xu and Shang, 2018）にも活用されているように，DCM におけるパラメタ推定・構造推定のための標準的な方法として定着しつつある．

4. ベイズ推定法

本章では DCM におけるパラメタのベイズ推定法を示す．本章で扱うベイズ推定法は，モデルパラメタに事前分布を仮定し，事後分布やその近似分布などからパラメタの推定値や予測分布を得る方法として広く捉える．

4.1 MAP 推定

周辺尤度に加えパラメタの事前分布を考慮し，点推定を得る方法は MAP 推定法として知られている．MAP 推定法は式 (2.18) に事前分布の対数を加えた目的関数を最適化するパラメタを推定値とする方法であり，

$$\{\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\pi}}\} = \underset{\boldsymbol{\Theta},\boldsymbol{\pi}}{\operatorname{argmax}}\{\log L(\boldsymbol{\Theta},\boldsymbol{\pi}) + \log(p(\boldsymbol{\Theta})) + \log(p(\boldsymbol{\pi}))\} \tag{4.1}$$

とパラメタの推定値を得る方法である．ここで，$\log(p(\boldsymbol{\Theta}))$ と $\log(p(\boldsymbol{\pi}))$ はそれぞれ項目パラメタと混合比率パラメタの対数事前確率密度である．例えば，θ_{jh} の事前分布としてパラメタが $a^0_{jh} > 0, b^0_{jh} > 0$ であるベータ分布，$\boldsymbol{\pi}$ の事前分布としてパラメタが $\boldsymbol{\delta}^0 = (\delta^0_1, \ldots, \delta^0_L), \delta^0_l > 0, \forall l$ であるディリクレ分布を仮定することができる．具体的な 2 つの事前確率密度は

$$p(\boldsymbol{\Theta}) = \prod_{j=1}^{J}\prod_{h=1}^{H_j} \frac{\Gamma(a^0_{jh})\Gamma(b^0_{jh})}{\Gamma(a^0_{jh}+b^0_{jh})}\theta_{jh}^{a^0_{jh}-1}(1-\theta_{jh})^{b^0_{jh}-1}, \tag{4.2}$$

$$p(\boldsymbol{\pi}) = \frac{\Gamma(\sum_{l=1}^{L}\delta^0_l)}{\prod_{l=1}^{L}\Gamma(\delta^0_l)}\prod_{l=1}^{L}\pi_l^{\delta^0_l-1} \tag{4.3}$$

である．

この事前分布の設定のもと，MAP 推定のための EM アルゴリズムの Q 関数は，式 (3.10) に対数事前分布の定数項を除いたものを足し，

$$\begin{aligned} &Q_{MAP}(\{\boldsymbol{\Theta},\boldsymbol{\pi}\}|\{\boldsymbol{\Theta}^{(t)},\boldsymbol{\pi}^{(t)}\}) \\ &= Q(\{\boldsymbol{\Theta},\boldsymbol{\pi}\}) + \sum_{j=1}^{J}\sum_{h=1}^{H_j}\{(a_{jh}-1)\log(\theta_{jh}) + (b_{jh}-1)\log(1-\theta_{jh})\} + \sum_{l=1}^{L}(\delta_l-1)\log(\pi_l) \end{aligned} \tag{4.4}$$

となる．M-step ではこの $Q_{MAP}(\{\boldsymbol{\Theta},\boldsymbol{\pi}\}|\{\boldsymbol{\Theta}^{(t)},\boldsymbol{\pi}^{(t)}\})$ を $\{\boldsymbol{\Theta},\boldsymbol{\pi}\}$ について最大化する．このときの更新式は，周辺最尤推定法の場合の導出と同様に導出でき，

$$\theta_{jh}=\frac{\sum_{i=1}^{I}\sum_{l=1}^{L}\mathcal{I}(\boldsymbol{\alpha}_{jh}^{*}=\boldsymbol{\alpha}_l)\gamma_{il}x_{ij}+a_{jh}^{0}-1}{\sum_{i=1}^{I}\sum_{l=1}^{L}\mathcal{I}(\boldsymbol{\alpha}_{jh}^{*}=\boldsymbol{\alpha}_l)\gamma_{il}+a_{jh}^{0}+b_{jh}^{0}-2}, \tag{4.5}$$

$$\pi_l=\frac{\sum_{i=1}^{I}\gamma_{il}+\delta_l^{0}-1}{I+\sum_{l=1}^{L}\delta_l^{0}-L} \tag{4.6}$$

となる．ただし，この更新式は周辺最尤推定法のための EM アルゴリズムと同様にパラメタの単調性制約を課していない点に注意が必要である．最尤推定の場合と同様に，MAP 推定の M-step においても，単調性制約を課した推定を行うこともできる．

4.2 マルコフ連鎖モンテカルロ法

MAP 推定法は最尤推定法と同様に点推定値を得る方法であった．一方，マルコフ連鎖モンテカルロ（Markov chain Monte Carlo [MCMC]）法を用いて，パラメタの事後分布を近似することで，点推定値だけではなく推定の不確かさを事後標準偏差などの統計量によって評価することができる．MCMC 法は MAP 推定や後述の変分ベイズ推定法よりも計算負荷が高いものの，単調性制約などモデルの制約も柔軟に取り込んだ推定が比較的容易に実行できるという利点がある．本節では単調性を満足するギブスサンプリング法（Liu and Johnson, 2019; Yamaguchi and Templin, 2022b）を示す．事前分布は，MAP 推定の際に用いたものと同様の分布を仮定する．

まず，上付き添字 $(t),(t=1,\ldots,T)$ は MCMC サンプリングのイテレーション番号を表すとする．$\boldsymbol{x}_i,\boldsymbol{\Theta}^{(t-1)},\boldsymbol{\pi}^{(t-1)}$ が得られているもとでのテスト解答者のアトリビュート習得パタン $\boldsymbol{\alpha}_i^{(t)}$ は，完全条件付き確率が

$$\begin{aligned}&P(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l|\boldsymbol{x}_i,\boldsymbol{\Theta}^{(t-1)},\boldsymbol{\pi}^{(t-1)})\\&=\frac{\pi_l^{(t-1)}\prod_{j=1}^{J}\prod_{h=1}^{H_j}[(\theta_{jh}^{(t-1)})^{x_{ij}}(1-\theta_{jh}^{(t-1)})^{1-x_{ij}}]^{\mathcal{I}(\boldsymbol{\alpha}_{jh}^{*}=\boldsymbol{\alpha}_l)}}{\sum_{l'=1}^{L}\{\pi_{l'}^{(t-1)}\prod_{j=1}^{J}\prod_{h=1}^{H_j}[[(\theta_{jh}^{(t-1)})^{x_{ij}}(1-\theta_{jh}^{(t\hat{-}1)})^{1-x_{ij}}]^{\mathcal{I}(\boldsymbol{\alpha}_{jh}^{*}=\boldsymbol{\alpha}_{l'})}\}}\end{aligned} \tag{4.7}$$

であるカテゴリカル分布からのサンプリングにより生成する．

次に，混合比率パラメタ $\boldsymbol{\pi}^{(t)}$ の完全条件付き事後分布はパラメタが

$$\delta_l^{*,(t)}=\sum_{i=1}^{I}\mathcal{I}(\boldsymbol{a}_i^{(t)}=\boldsymbol{a}_l)(1-x_{ij})+\delta_l^{0} \tag{4.8}$$

であるディリクレ分布であり，この分布からサンプリングを行う．

最後に，項目パラメタ $\theta_{jh}^{(t)}$ の完全条件付き事後分布は，パラメタが

$$\begin{cases}a_{jh}^{*,(t)}=\sum_{i=1}^{I}\sum_{l=1}^{L}\mathcal{I}(\boldsymbol{a}_{jh}^{*}=\boldsymbol{a}_l)\mathcal{I}(\boldsymbol{a}_i^{(t)}=\boldsymbol{a}_l)x_{ij}+a_{jh}^{0}\\ b_{jh}^{*,(t)}=\sum_{i=1}^{I}\sum_{l=1}^{L}\mathcal{I}(\boldsymbol{a}_{jh}^{*}=\boldsymbol{a}_l)\mathcal{I}(\boldsymbol{a}_i^{(t)}=\boldsymbol{a}_l)(1-x_{ij})+b_{jh}^{0}\end{cases} \tag{4.9}$$

であるベータ分布である．ただし，項目パラメタのサンプリングの際には，単調性制約を満足するために，切断ベータ分布を用いた逆関数法を利用したアルゴリズム用いる（e.g., Hoijtink, 1998; Laudy et al., 2004）．上限と下限が Upp と Low である切断ベータ分布から乱数をサンプリングする逆関数法は以下の 2 つの手続きを踏む．まず，下限 0 から上限 1 の一様分布に従う変数 u をサンプリングし，次にパラメタが a,b のベータ分布の分位点 $Low+u\times(1-Upp-Low)$ を計算する．この分位点が，目標とする切断ベータ分布からのランダムサンプリングになっており，この手続きを項目パラメタの単調性制約を満足するために利用する．

項目パラメタ θ_{jh} をサンプリングする際には，以下の手続きを踏む．まず，$\{h|\sum_k \alpha^*_{hjk}=0\}$ に対して，下限 0, 上限 $\min\{\theta^{(t-1)}_{jh'}|\boldsymbol{\alpha}^*_{h'j}\succ\boldsymbol{\alpha}^*_{hj}\}$ で，パラメタが $a^{*,(t)}_{jh}, b^{*,(t)}_{jh}$ の切断ベータ分布から MCMC サンプルをサンプリングする．次に，$\{h|\sum_k q_{jk}>\sum_k \alpha^*_{hjk}>0\}$ に対して，$\sum_k \alpha^*_{hjk}=1,2,\ldots,\sum_k q_{jk}-1$ の順に，下限 $\max\{\theta^{(t)}_{jh}|\boldsymbol{\alpha}^*_{hj}\succ\boldsymbol{\alpha}^*_{h'j}\}$, 上限 $\min\{\theta^{(t-1)}_{jh'}|\boldsymbol{\alpha}^*_{h'j}\succ\boldsymbol{\alpha}^*_{hj}\}$ で，パラメタが $a^{*,(t)}_{jh}, b^{*,(t)}_{jh}$ の切断ベータ分布から MCMC サンプルをサンプリングする．最後に，$\{h|\alpha^*_{hjk}=\sum_k q_{jk}\}$ に対して，下限 $\max\{\theta^{(t)}_{jh}|\boldsymbol{\alpha}^*_{hj}\succ\boldsymbol{\alpha}^*_{h'j}\}$, 上限 1 で，パラメタが $a^{*,(t)}_{jh}, b^{*,(t)}_{jh}$ の切断ベータ分布から MCMC サンプルをサンプリングする．以上の手続きを，項目パラメタおよび混合比率パラメタの初期値 $\boldsymbol{\Theta}^{(0)}, \boldsymbol{\pi}^{(0)}$ を定めたもとで，予め決めた回数 T だけ反復することで，MCMC サンプルを得る．

この他，後述のノンパラメトリック推定と関連して，モデルのパラメタを周辺化し，アトリビュート習得パタンを直接サンプリングする崩壊型ギブスサンプリング（collapsed Gibbs sampling）も提案されている（Yamaguchi and Templin, 2022a）．さらに，スライスサンプリングを用いた方法（Xu et al., 2020）や，DINA モデルのパラメタ推定に Pólya-gamma 分布を用いた方法（Zhang et al., 2020）や，Pólya-gamma 分布を用いてモデルパラメタと Q 行列の推定も同時に実行する方法も提案されている（Balamuta and Culpepper, 2022; Jimenez et al., 2023）．

4.3 変分ベイズ推定法

MCMC 法は乱数を生成することにより，パラメタと潜在変数の事後分布を近似する方法であった．MCMC 法の欠点とそれを補う変分ベイズ推定法の利点については以下のようにまとめられる（Bishop, 2006）．まず，MCMC 法はサンプリングの回数を増やせば事後分布の近似精度を高めることができるが，計算量が多くなり推定に時間がかかってしまうことがある．また，MCMC サンプリングが定常分布からのサンプリングになっているかどうかを判断するための手続きが必要であるが，その判断は必ずしも容易ではない．一方，変分ベイズ推定法はあるクラスの確率分布の中から事後分布を最もよく近似する分布を見つける方法であり，計算量を押さえつつアルゴリズムの収束判定も行いやすく，ベイズ的な推定量を得ることができる．また，変分ベイズ推定法は，対数尤度の下界を最適化する方法として定式化されることもある（e.g., Jeon et al., 2017）．上記のように，変分ベイズ推定法は最適化問題を解くことで，MCMC よりも高速な推定を可能にしている．ただし，事後分布を近似する分布（変分事後分布）はデータを得た元での真の事後分布ではないため，変分事後分布における分散が真の事後分布における分散よりも小さくなるなどの問題も有している．

変分ベイズ推定法は，項目反応理論モデルを含む潜在変数モデルにおいても近年適用例が増加してきている方法である．例えば，Jeon et al.（2017）は一般化線形混合モデルに対しての変分ベイズアルゴリズムを提案した．その後，多次元項目反応理論モデルへの適用例（Cho et al., 2021, 2024）もあり，ますますの発展が見られる．変分ベイズ推定法についての理論的な解説は Nakajima et al.（2019）においてなされている．また，入門的レビューについては Blei et al.（2017）や Grimmer（2011）が参考になる．

DCM においては，Yamaguchi and Okada（2020c）において変分ベイズ推定法に基づく DINA モデルの推定アルゴリズムが提案されている．さらに，Yamaguchi（2020）は多枝選択式のデータに対する DINA モデル（multiple choice DINA model）について，Yamaguchi and Okada（2020b）は前述の MCMC 法で用いたモデルと事前分布を仮定した変分ベイズ推定法を提案した．この他，Yamaguchi and Martinez（2024）は隠れマルコフモデルにもとづく縦断的 DCM に対して変分ベイズ推定法を提案し，Yamaguchi（2023a）は外部情報を含めた DCM のうちの一つである二段階 DCM に対する変分ベイズ推定法を提案している．本節では，Yamaguchi and Okada（2020b）にもとづいた DCM における変分ベイズ推定法のアルゴリズムを示す．アルゴ

リズムの具体的な導出方法については，Yamaguchi and Okada (2020b) に詳細な記載がある．

変分ベイズ推定法では，DCM における潜在変数とパラメタの事後分布 $P(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}|\mathrm{X})$ を適当な分布 $q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})$ によって近似することを考える．$q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})$ と $P(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}|\mathrm{X})$ の間の Kullback-Leibler divergence を

$$KL(q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})||P(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}|\mathrm{X})) = \int \sum_{\boldsymbol{\mathcal{A}}} q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}) \log \frac{q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})}{P(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}|\mathrm{X})} d\boldsymbol{\Theta} d\boldsymbol{\pi} \tag{4.10}$$

とする．DCM の変分ベイズ推定法においては，$KL(q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})||P(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}|\mathrm{X}))$ を潜在変数とパラメタの間の独立性の仮定

$$q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}) = q(\boldsymbol{\mathcal{A}}) q(\boldsymbol{\Theta}, \boldsymbol{\pi}) = \left(\prod_{i=1}^{I} q(\boldsymbol{\alpha}_i) \right) \left(\prod_{j=1}^{J} \prod_{h=1}^{H_j} q(\theta_j h) \right) q(\boldsymbol{\pi}) \tag{4.11}$$

のもと最小化する分布 q^* を変分事後分布とよび，変分事後分布を事後分布の近似分布とする．ここでは，表記の簡略化のため，引数に対応して異なった分布を仮定している．上式の $q(\boldsymbol{\mathcal{A}})q(\boldsymbol{\Theta}, \boldsymbol{\pi})$ が最小限の独立性の仮定であり，2 つ目の等号はモデルの仮定から導かれる独立性である．これを用いて，変分ベイズ推定法における最適化問題は

$$q^* = \underset{q}{\operatorname{argmin}}\, KL(q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})||P(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}|\mathrm{X})), \text{s.t. } q(\boldsymbol{\mathcal{A}})q(\boldsymbol{\Theta}, \boldsymbol{\pi}) \tag{4.12}$$

となる．近似する分布に独立性を仮定し，変分分布のクラスを限定して事後分布を近似する方法は平均場近似と呼ばれる．変分ベイズ推定法では，真の事後分布を近似するための変分分布の独立性を仮定した最適化問題を解くため，MCMC 法よりも高速な推定が可能となる．

また，等価な定式化として，周辺対数尤度の下界 $\underline{l}(q)$ を同様の制約条件のもと最適化する分布を変分布とすることもある．すなわち，

$$\begin{aligned} q^* &= \underset{q}{\operatorname{argmax}} \int \sum_{\boldsymbol{\mathcal{A}}} q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}) \log \frac{P(\mathrm{X}, \boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})}{q(\boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})}, \text{s.t. } q(\boldsymbol{\mathcal{A}})q(\boldsymbol{\Theta}, \boldsymbol{\pi}) d\boldsymbol{\Theta} d\boldsymbol{\pi} \\ &= \underset{q}{\operatorname{argmax}} \{\underline{l}(q)\}, \text{s.t. } q(\boldsymbol{\mathcal{A}})q(\boldsymbol{\Theta}, \boldsymbol{\pi}) \end{aligned} \tag{4.13}$$

により変分事後分布を求める問題を定式することもある．ただし，$P(\mathrm{X}, \boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})$ はデータとパラメタの同時確率であり，式 (3.1) とパラメタの事前分布から，

$$P(\mathrm{X}, \boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi}) = P(\mathrm{X}, \boldsymbol{\mathcal{A}}|\boldsymbol{\Theta}, \boldsymbol{\pi}) P(\boldsymbol{\Theta}) P(\boldsymbol{\pi}) \tag{4.14}$$

と表される．対数周辺尤度の下界 $\underline{l}(q)$ は ELBO (evidence lower bound) とも呼ばれ，本論文の DCM の設定においては解析的な表現が与えられている (Yamaguchi, 2020, Equation 32)．

上記の設定のもと，

$$q(\boldsymbol{\mathcal{A}}) \propto \exp\{\mathbb{E}_{q(\boldsymbol{\Theta}, \boldsymbol{\pi})}[\log P(\mathrm{X}, \boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})]\}, \tag{4.15}$$

$$q(\boldsymbol{\Theta}, \boldsymbol{\pi}) \propto \exp\{\mathbb{E}_{q(\boldsymbol{\mathcal{A}})}[\log P(\mathrm{X}, \boldsymbol{\mathcal{A}}, \boldsymbol{\Theta}, \boldsymbol{\pi})]\} \tag{4.16}$$

により，最適な変分事後分布を計算できる．しかし，変分事後分布は相互に関係しているので，最適な変分事後分布を得るために反復計算を行う必要がある．期待値の計算 $\mathbb{E}_{q(\boldsymbol{\Theta}, \boldsymbol{\pi})}[\cdot], \mathbb{E}_{q(\boldsymbol{\mathcal{A}})}[\cdot]$ はそれぞれ変分事後分布 $q(\boldsymbol{\Theta}, \boldsymbol{\pi})$ と $q(\boldsymbol{\mathcal{A}})$ について取ることを意味している．前述の事前分布の設定のもと，この変分事後分布はよく知られた分布になることが示されている．まず，$\boldsymbol{\alpha}_i$ の変分事後分布は

$$q^*(\boldsymbol{\alpha}_i) \propto \prod_{l=1}^{L} r_{il}^{\mathcal{I}(\boldsymbol{\alpha}_i=\boldsymbol{\alpha}_l)} \tag{4.17}$$

であり，変分パラメタが $r_{il} = \frac{\rho_{il}}{\sum_{l'=1}^{L}\rho_{il'}}$ であるカテゴリカル分布である．ここで ρ_{il} は

$$\begin{aligned}\log\rho_{il} &= \mathbb{E}_{q(\boldsymbol{\pi})}[\log\pi_l] \\ &\quad + \sum_{j=1}^{J}\sum_{h=1}^{H_j}\mathcal{I}(\boldsymbol{\alpha}_{jh}^* = \boldsymbol{\alpha}_l)(x_{ij}\mathbb{E}_{q(\theta_{jh})}[\log\theta_{jh}] + (1-x_{ij})\mathbb{E}_{q(\theta_{jh})}[\log(1-\theta_{jh})])\end{aligned} \tag{4.18}$$

である．上式中に現れる期待値の具体的な値は後述する．つぎに，混合比率パラメタの変分事後分布は，

$$q^*(\boldsymbol{\pi}) \propto \prod_{l=1}^{L}\pi_l^{\delta_l^*} \tag{4.19}$$

と表されることからディリクレ分布であることがわかり，変分パラメタ δ_l^* は

$$\delta_l^* = \sum_{i=1}^{I}\mathbb{E}_{q(\boldsymbol{\alpha}_i)}[\mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)] + \delta_l^0 \tag{4.20}$$

である．最後に，項目パラメタ θ_{jh} の変分事後分布は

$$q^*(\theta_{jh}) \propto \theta_{jh}^{a_{jh}^*-1}(1-\theta_{jh})^{b_{jh}^*-1} \tag{4.21}$$

となるベータ分布であり，変分パラメタ a_{jh}^*, b_{jh}^* は，

$$\begin{cases} a_{jh}^* &= \sum_{i=1}^{I}\sum_{l=1}^{L}\mathcal{I}(\boldsymbol{a}_{jh}^* = \boldsymbol{a}_l)\mathbb{E}_{q(\boldsymbol{\alpha}_i)}[\mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)]x_{ij} + a_{jh}^0 \\ b_{jh}^* &= \sum_{i=1}^{I}\sum_{l=1}^{L}\mathcal{I}(\boldsymbol{a}_{jh}^* = \boldsymbol{a}_l)\mathbb{E}_{q(\boldsymbol{\alpha}_i)}[\mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)](1-x_{ij}) + b_{jh}^0 \end{cases} \tag{4.22}$$

と計算される．上記の変分パラメタの更新に現れる期待値は

$$\mathbb{E}_{q(\boldsymbol{\alpha}_i)}[\mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)] = r_{il}, \tag{4.23}$$

$$\mathbb{E}_{q(\theta_{jh})}[\log\theta_{jh}] = \psi(a_{jh}^*) - \psi(a_{jh}^* + b_{jh}^*), \tag{4.24}$$

$$\mathbb{E}_{q(\theta_{jh})}[\log(1-\theta_{jh})] = \psi(b_{jh}^*) - \psi(a_{jh}^* + b_{jh}^*), \tag{4.25}$$

$$\mathbb{E}_{q(\boldsymbol{\pi})}[\log\pi_l] = \psi(\delta_l^*) - \psi\left(\sum_{l=1}^{L}\delta_l^*\right) \tag{4.26}$$

と計算される．ここで，$\psi(\cdot)$ はディガンマ関数であり $\frac{d}{dx}\log\Gamma(x) = \int_0^\infty y^{(x-1)}\exp(-y)dy$ と定義される．

DCM に対する変分ベイズアルゴリズムは以下のようになる．まず，変分パラメタ $A^* = \{a_{jh}^*\}_{j,h}^{J,H_j}, B^* = \{B_{jh}^*\}_{j,h}^{J,H_j}, \boldsymbol{\delta}^* = (\delta_1^*,\ldots,\delta_L^*)^\top$ を初期化し，$q(\boldsymbol{\alpha}_i), \forall i,$ の変分パラメタを式(4.18)により更新する(変分 E-step)．混合比率パラメタと項目パラメタの変分事後分布 $q(\boldsymbol{\pi})$ と $q(\theta_j h), \forall j, h,$ の変分パラメタを式(4.20)と(4.22)により更新する(変分 M-step)．変分 E-step および変分 M-step を収束基準(例えば，$\underline{l}(q)$ の変化量が $1 > \epsilon > 0$ 以下になる，など)を満足するまで，変分 E-step と変分 M-step を反復する．上記の更新式により $\underline{l}(q)$ が単調に増加することが保証されている．最後に，得られた変分事後分布をもとにその期待値・中央値や分散などの要約統計量をパラメタの点推定値や事後分散の推定値とすればよい．

ここで示した変分ベイズアルゴリズムは，単調性を完全に満足した事後分布の構成はできて

いない点に注意が必要である．Yamaguchi and Okada (2020c) では，事前分布の平均値が単調性を満たすようにハイパーパラメタを調整しておく方法を提案し，シミュレーションや実データ分析ではフルベイズ推定の場合と類似したの推定値を得ることができることを報告している．しかし，当該の方法の理論的な性質は検討されていない．

5. (一般化) ノンパラメトリック推定法

最尤推定法とベイズ推定法はいずれもパラメトリックな DCM の推定法である．すなわち，これらの方法では DCM の最終的なアウトプットであるアトリビュート習得パタンとモデルのパラメタの両方を仮定していた．これらの方法はサンプルサイズが小さい場合にはアトリビュート習得パタンの推定が安定しないといった問題を持つとされている．これらの推定法に対して，正答確率パラメタや混合比率パラメタといったモデルパラメタを仮定せずにアトリビュート習得パタンを推定する方法として，(一般化) ノンパラメトリック推定法 (generalized non-parametric method) も提案されている (Chiu and Douglas, 2013; Chiu et al., 2018)．加えて，Ma et al. (2023a) はノンパラメトリック推定法とパラメトリック法の関連を議論し，これらを統一的に表現できる損失関数に基づく推定法を提案した．本章では，基本的なノンパラメトリック推定法を示し，さらに一般化ノンパラメトリック推定法についても示す．

後述するように，(一般化) ノンパラメトリック推定法で用いられる損失関数では DINA モデルや DINO (deterministic input noisy “OR”-gate) モデル (Templin and Henson, 2006) といった具体的なデータ生成モデルを仮定しているように見える．しかし，(一般化) ノンパラメトリック推定法によるアトリビュート習得パタンの推定値は，データ生成モデルが DINA モデルや DINO モデルでなくとも一致性があることが示されている．

まず，Wang and Douglas (2015) はノンパラメトリック推定法によるアトリビュート習得パタンの推定値の一致性を証明した．より具体的には，DINA モデルを含む結合的 (非補償的) な DCM をデータ生成モデルし，既知の Q 行列のもとで，ノンパラメトリック推定法によるアトリビュート習得パタンが一致的であるための項目反応関数について必要な条件は以下の 2 つであることが示されている (Wang and Douglas, 2015, p. 99)．第 1 の条件は，それぞれの項目に対して項目の正答に必要なアトリビュートを習得している場合の正答確率が 0.5 以上であることで，第 2 の条件は項目の正答に必要なアトリビュートを習得していない場合の正答確率が 0.5 未満であることである．データ生成モデルがこれらの条件を満足している場合，ノンパラメトリック推定法によるアトリビュート習得パタンの推定値は，個人ごとに，項目数の増加に伴って一致的であることが示されている．

さらに，Chiu and Köhn (2019) は一般化ノンパラメトリック推定法によるアトリビュート習得パタンの推定値の一致性を証明した．この一致性は，データ生成モデルが LCDM と同等の一般的な DCM である generalized DINA モデル (de la Torre, 2011) で表現できるモデルに対して，一定の仮定のもと成り立つものである (Chiu and Köhn, 2019, p. 841, Theorem 2)．これはすなわち，データ生成モデルがよく利用される DCM の範囲では，一般化ノンパラメトリック推定法のよるアトリビュート習得パタンの推定値が統計的に望ましい性質を有しており，この推定法を用いる根拠があることを意味している．この意味で，一般化ノンパラメトリック推定法は，パラメトリックなモデルを仮定した推定法と同様に一般に利用できる方法であるといえる．

5.1 ノンパラメトリック推定法

ここから，具体的なノンパラメトリック推定法について説明する．ノンパラメトリック推定

法で用いる距離関数を定義するために理想反応(ideal response)を定義する．DINA モデルにおける理想反応は

$$\eta_j^{DINA}(\boldsymbol{\alpha}_l) = \prod_{k=1}^{K} \alpha_{lk}^{q_{jk}} \tag{5.1}$$

で定義される．ここで，0^0 は 1 と定義する．理想反応 $\eta_j(\boldsymbol{\alpha}_l)$ は，l 番目のアトリビュート習得パタンが項目 j で測定されているアトリビュートをすべて習得しているものであれば 1，そうでなければ 0 を取る変数であり，誤差がない場合の l 番目のアトリビュート習得パタンの項目 j への正答・誤答を表している．さて，この理想反応と実際の項目反応とのハミング距離(Hamming distance) $d_h(\cdot)$ を

$$d_h(\boldsymbol{x}_i, \boldsymbol{\eta}^{DINA}(\boldsymbol{\alpha}_l)) = \sum_{j=1}^{J} |x_{ij} - \eta_j^{DINA}(\boldsymbol{\alpha}_l)| = \sum_{j=1}^{J} \left| x_{ij} - \prod_{k=1}^{K} \alpha_{lk}^{q_{jk}} \right| \tag{5.2}$$

とする．ただし，$\boldsymbol{\eta}^{DINA}(\boldsymbol{\alpha}_l) = (\eta_1^{DINA}(\boldsymbol{\alpha}_l), \ldots, \eta_J^{DINA}(\boldsymbol{\alpha}_l))^\top$ である．ノンパラメトリック推定法は，$\boldsymbol{x}_i, \boldsymbol{\eta}_i$ 間の距離を適当に定め，

$$\hat{\boldsymbol{\alpha}}_i = \underset{\boldsymbol{\alpha}_l}{\mathrm{argmin}}\{d_h(\boldsymbol{x}_i, \boldsymbol{\eta}^{DINA}(\boldsymbol{\alpha}_l))\} \tag{5.3}$$

と距離関数を最小化するアトリビュート習得パタンを推定値とする方法である．距離関数の定義として，重み付きハミング距離(weighted Hamming distance)を用いることもでき，その場合の距離関数 $d_{wh}(\cdot)$ は

$$d_{wh}(\boldsymbol{x}_i, \boldsymbol{\eta}^{DINA}(\boldsymbol{\alpha}_l)) = \sum_{j=1}^{J} \frac{1}{\bar{p}_j(1-\bar{p}_j)} |x_{ij} - \boldsymbol{\eta}_j(\boldsymbol{\alpha}_l)| = \sum_{j=1}^{J} \frac{1}{\bar{p}_j(1-\bar{p}_j)} \left| x_{ij} - \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}} \right| \tag{5.4}$$

と定義される．ここで $\bar{p}_j = \frac{1}{I}\sum_{i=1}^{I} x_{ij}$ で定義される項目 j の正答確率である．なお，距離関数を最小化するアトリビュート習得パタンが複数ある場合には，それらのなかから一様にランダムで一つのパタンを選択するという操作が必要になる．この手続きは一般化ノンパラメトリック推定法においても必要である．

5.2 一般化ノンパラメトリック推定法

一般化ノンパラメトリック推定法では DINA モデルの理想反応に加えて，DINO モデルの理想反応

$$\eta_j^{DINO}(\boldsymbol{\alpha}_l) = 1 - \prod_{k=1}^{K} (1-\alpha_{lk})^{q_{jk}} \tag{5.5}$$

を用いる．$\eta_j^{DINO}(\boldsymbol{\alpha}_l)$ は項目 j で必要なアトリビュートのうち少なくとも 1 つ習得しているパタンであれば 1，そうでなければ 0 となる．すなわち，$\eta_j^{DINO}(\boldsymbol{\alpha}_l) = 0$ となるのは項目 j で必要なアトリビュートを全く習得していない場合とも考えられる．さらに，重み w_{lj} (ただし，$0 \le w_{lj} \le 1$)を導入し，重み付き理想反応(weighted ideal response) $\eta_j^{(w)}(\boldsymbol{\alpha}_l)$ を

$$\eta_j^{(w)}(\boldsymbol{\alpha}_l) = w_{lj}\eta_j^{DINA}(\boldsymbol{\alpha}_l) + (1-w_{lj})\eta_j^{DINO}(\boldsymbol{\alpha}_l) \tag{5.6}$$

と定義する．重み w_{lj} は項目 j の反応が DINA 的か DINO 的かどうかをあらわしている．

この重み付き理想反応と項目反応の距離関数を

$$d_g(\boldsymbol{x}_j, \eta_j^{(w)}(\boldsymbol{\alpha}_l)) = \sum_{i=1}^{I} \mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)(x_{ij} - \eta_j^{(w)}(\boldsymbol{\alpha}_l))^2 \tag{5.7}$$

とする．このとき，上記の距離関数を最小化する重みは

$$\hat{w}_{lj} = \frac{\sum_{i=1}^{I} \mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l)(x_{ij} - \eta_j^{DINO}(\boldsymbol{\alpha}_l))^2}{(\sum_{i=1}^{I} \mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l))(\eta_j^{DINA}(\boldsymbol{\alpha}_l) - \eta_j^{DINO}(\boldsymbol{\alpha}_l))} \tag{5.8}$$

となる．この推定された重みを用いた重み付き理想反応を $\hat{\eta}_j^{(w)}(\boldsymbol{\alpha}_l)$ とし，アトリビュート習得パタンを推定するための損失関数を

$$d_g(\boldsymbol{x}_i, \hat{\boldsymbol{\eta}}^{(w)}(\boldsymbol{\alpha}_l)) = \sum_{j=1}^{J} (x_{ij} - \hat{\eta}_j^{(w)}(\boldsymbol{\alpha}_l))^2 \tag{5.9}$$

とする．ただし，$\hat{\boldsymbol{\eta}}^{(w)}(\boldsymbol{\alpha}_l) = (\hat{\eta}_1^{(w)}(\boldsymbol{\alpha}_l), \ldots, \hat{\eta}_J^{(w)}(\boldsymbol{\alpha}_l))^\top$ である．上記の重みを用いて，一般化ノンパラメトリック推定法によるアトリビュート習得パタンは

$$\hat{\boldsymbol{\alpha}}_i = \underset{\boldsymbol{\alpha}_l}{\operatorname{argmin}} \{d_g(\boldsymbol{x}_i, \hat{\boldsymbol{\eta}}^{(w)}(\boldsymbol{\alpha}_l))\} \tag{5.10}$$

と推定される．

以上で見たように，一般化ノンパラメトリック推定法においては，まずアトリビュート習得パタンの初期値を決め，次に重みパラメタを式(5.8)により更新する．この更新された重みを用いて式(5.10)によりアトリビュート習得パタンを更新し，再度重み w_{lj} の更新を行う．このように，重みとアトリビュート習得パタンの更新を，アトリビュート習得パタンの変化が十分に小さくなるまで反復するのが一般化ノンパラメトリック推定法の手続きである．アトリビュート習得パタンの初期値としては通常のノンパラメトリック推定法による推定値を用いる．

このように一般化ノンパラメトリック推定法においては，補償的な(あるいは disjunctive) DCM の理想反応と非補償的な(あるいは conjunctive な)ものを組み合わせる方法を用いている．パラメトリック法においても，Yamaguchi and Okada (2020a) は DINA モデルと DINO モデルの項目反応関数の混合モデルとしての Hybrid DCM を提案しており，一般化ノンパラメトリック推定法とも類似した考え方をしている．

6. 推定法の選択

ここまで，最尤推定法，ベイズ推定法，ノンパラメトリック推定法の観点から DCM のパラメタ推定についてレビューを行った．本節では，これらのパラメタ推定法について，どのような場合にどのような方法を選択するのが望ましいのか，パラメタ推定法の特徴を踏まえて議論を行う．特に，サンプルサイズの大小とテスト項目数(多い・少ない)の組み合わせの観点から整理を行う．

まず，小サンプル・小項目数の状況は，例えばクラスルームの簡単な確認テストのような状況とみなすことができる．DCM が学習改善のための情報をテストから引き出すための統計モデルであることに鑑みると，小サンプル・小項目数の状況における DCM のパラメタ推定を考察することに意義はあるだろう．ただし，この状況はデータが多くないため，そもそも複雑な統計モデルの推定が容易ではない状況であるという点に留意する必要がある．この状況では，サンプルサイズが小さく，また項目数も多くないため最尤推定ベースの推定法の適用は難しいだろう．この場合には，計算量が多くなったとしても，事前情報を含めたフルベイズ推定を行うのが望ましい可能性がある．ただし，ベイズ推定法の利用のためにはパラメトリックモデル

の特定が必要である点に注意が必要である．また，データが少ないため，事前情報が強すぎるとデータからアトリビュート推定を行う意味も薄くなってしまう点にも注意しなければならない．ノンパラメトリック推定法は，データが少ない場合でもアトリビュート習得パタンの推定がある程度正確であることが数値実験で示されているが，Q 行列の特定が正しい必要がある．項目数が少ない場合には，ノンパラメトリック推定法であっても，アトリビュート習得パタンの推定が安定しにくい可能性もあるので，やはり注意が必要だろう．小サンプル・小項目数の状況でパラメトリックモデルを用いたい場合には，事前に項目パラメタを推定済の項目プールを用意しおくことなど，工夫が必要になると思われる．

次に，小サンプル・大項目数の状況は，例えばクラスルームの定期テストや 1 年の学習成果を振り返るようなまとめテストや定期テストなどの状況が考えられる．多くの項目数が利用できるという点では，先のノンパラメトリック推定法の利用が第 1 の選択肢に挙がるだろう．特に，制約がより少ない一般化ノンパラメトリック推定法が利用しやすいと考えられる．また，小サンプル・小項目数の状況と同様に，ベイズ推定法も候補に挙がる．一つ一つの項目パラメタの推定には不確実性がありながらも，ベイズ推測においてはそれらを考慮して，アトリビュート習得パタンの推定ができると考えられるからである．ただし，MAP 推定は，アトリビュート習得パタンの推定の際に，最尤推定のように項目パラメタの点推定値を使うため，項目パラメタの推定の不確実性は考慮しにくいと考えられる．もし，アトリビュートの数が多い場合や推定時間を短縮したい場合には，変分ベイズ推定法も利用できるだろう．最尤推定法は，やはり項目パラメタの推定の際にサンプルサイズが小さいことが影響し，正確な項目パラメタの推定が難しい可能性がある．また，アトリビュート習得パタンの推定には，項目パラメタ推定値を固定する必要があるが，項目パラメタの不確実性を考慮できないため，最尤推定法の利用には慎重になる必要がある．

さらに，大サンプル・小項目数の状況は，やや想定しにくいが，行政が行う標本調査の一部に，簡単な学力調査が含まれているような状況などが考えられる．この場合，個々人のアトリビュート習得パタンの推定を行うよりも，出題した項目の項目パラメタの特性や混合比率パラメタの推定に興味があると思われる．こうした状況であれば，最尤推定ベースの方法は，項目パラメタの点推定値だけではなく，漸近論に依拠した標準誤差も利用できると考えられるため，周辺最尤推定法を用いた方法が良い可能性もある．また，やや探索的にデータ分析を行いたい場合には，正則化推定法を用いてもよいだろう．MCMC を用いたベイズ的な方法は，サンプルサイズが大きい状況では計算時間が長くなる傾向があるものの，最尤推定法とあまりかわらない推定値が得られる可能性もあるので，積極的に用いる必要はないかもしれない．ただし，事前情報を積極的に活用した場合はその限りではない．この場合には，変分ベイズ推定法や MAP 推定法は，計算速度が MCMC より早いため，利用してもよいだろう．ノンパラメトリック推定法は，その目的に鑑みて，大サンプル・小項目数の推定には適さないのではないかと考えられる．ただし，他の推定法の初期値に用いるアトリビュート習得パタンをノンパラメトリック推定法により推定するという方略は，他の推定法の収束を早めることに繋がると期待できる．

最後に，大サンプル・大項目数の状況は，PISA（Programme for International Student Assessment）調査や国際数学・理科教育動向調査（Trends in International Mathematics and Science Study [TIMSS]）といった国際調査のような状況が思いつく．この場合，先の大サンプル・小項目数の状況と類似して，個々人のアトリビュート習得パタンの推定は必ずしも最終的な目的ではないと考えられる．また，この場合アトリビュート数が多くなるなど，モデルの規模も大きくなると考えられる．こうした場合には，まず許容できる待ち時間で推定が実行できることが望ましいので，変分ベイズ推定法や同時最尤推定法が候補に挙がる．ただし，同時最尤推定法

は点推定値しか与えない方法であるので，標準誤差の算出は別途行う必要がある．周辺最尤推定法や MCMC 法による推定法は，計算コストが高くなりやすいで，大サンプル・大項目数かつモデルが複雑な状況は使いにくいかもしれない．ノンパラメトリック推定法は大サンプル・小項目数の状況と同様に初期値を推定する方法としての活用が考えられる．

以上のように，テストの分析状況や目的によって使いやすい推定法が変わることがわかる．上記の議論から，大雑把に言って，サンプルサイズが小さい場合には，ノンパラメトリック推定法や事前情報を加味できるベイズ推定法が使いやすく，サンプルサイズが大きい場合には計算効率も考慮して，変分ベイズ推定法や同時・周辺最尤推定法が使いやすいと考えられる．

7. 総合考察

本稿では DCM のモデルパラメタ推定法について，現在提案されている方法を概観した．推定法としては，主として最尤推定法，ベイズ推定法，およびノンパラメトリック推定法の 3 つの方法を軸に発展がなされていた．最尤推定法においては，正則化推定法の活用がみられ，ベイズ推定法においては変分ベイズ推定法などの発展がみられた．正則化推定法や変分ベイズ推定法など機械学習でも用いられている方法も積極的に利用されており，活発に研究が進んでいることがうかがえる．本研究でレビューを行った方法により，DCM においては基本的なパラメタ推定についての研究は概ねしつくされていると考えられる．

以下では，推定法についての残された課題について議論を行う．特に，発展的なモデルのパラメタ推定，効率的な推定アルゴリズムの開発，より現実的な状況に即した推定法の開発の必要性といった観点から議論する．

本研究でレビューした方法は，多値反応や多値アトリビュート習得を想定した場合に拡張されている場合や，アトリビュート階層構造を含めた推定法につながる方法も多い．特に，診断情報をより精緻にするために，反応時間やアンケート調査の状況などのテストへの解答パタン以外の外部情報を取り入れたモデルの推定法も必要となる．こうしたモデルは一般にパラメタの数が増え，複雑化されていくと考えられる．そうした場合にも対応できるような効率的な推定法の開発は今後重要な観点である．特に，ある時点までの診断情報を捉えつつ，その後の診断を行ううえでは，最尤推定法よりもベイズ推定によるパラメタ更新を行えるような枠組みが望ましいだろう．特にベイズ推定の中でも，計算量の少ない変分ベイズ推定法などは今後より活用が期待できる推定法と考えられる．

こうした点に関連して，効率的な推定アルゴリズムの開発も必要となる．教育現場への情報機器の普及に伴い，タブレット端末などを用いた学習も一般的になりつつある．こうした現状においては，学習者のテストの反応パタン以外の普段の学習状況などのログデータも大量に取得できると考えられる．また，学習の進展に伴い，アトリビュートの数も次第に増加することが想定できる．こうした複雑なモデルかつ大量のデータを用いて診断情報を即時的に提供するためには，効率的な推定アルゴリズムの開発が不可欠になると考えられる．この意味で，ノンパラメトリック推定法などは高速に機能する方法として，有用であると考えられる．また，パラメトリックな方法においても，変分ベイズ推定法のように，比較的計算量が少ない方法なども提案されているものの，大規模なデータ状況下でも効率的に機能する方法を追求する意義は大きいだろう．

さらに，DCM はその目的から，教室場面など比較的サンプルサイズが小さい状況での利用も未だに一般的である．先の大量のデータにも対応できる推定法に加えて，こうした小サンプルでも，アトリビュート習得パタンを安定して推定できる方法の開発も必要であろう．ノンパラメトリック推定法は，こうした状況でも機能することがシミュレーションにより示されてい

るが，パラメトリックな方法においてもこうした方法が望まれる．

本研究の限界として，Q 行列の推定法についてはレビューを行わなかった点が挙げられる．Q 行列を含めたモデルの推定法も近年活発に研究がなされており，Q 行列推定は DCM の理論・応用の両側面から重要な観点と考えられる．Q 行列推定はモデルパラメタの識別性とも関係し，理論的にも高度な話題であり，DCM の応用研究者がそれらの原典を参照することは必ずしも容易ではない．また，DCM 研究者にとっても，先端的な課題として Q 行列推定の現状を日本語で参照できる包括的なレビューの価値は高いと考えられる．今後は，Q 行列推定の方法の展開の包括的なレビューが望まれる．

加えて，DCM は制約付き潜在クラスモデルとして，識別性など理論的な研究もなされている．後述の Q 行列の推定の話題とも関連して，DCM を現実的な場面で利用するためにも，DCM の理論的な性質の検討は重要である．こうした理論的な発展の現状の解説も必要となるだろう．

さらに，関連して，アトリビュート階層構造の推定法についても重要な研究関心である．本研究では，構造化されていないアトリビュート習得パタンを想定したが，実際にはアトリビュートには習得の順序関係などが想定される．こうしたアトリビュートの習得の順序関係について，理論的な想定のみならずデータから推定を行う方法も，DCM の応用を促す上で必要な観点である．こうしたアトリビュート階層構造の推定法や関連する話題についても包括的なレビューが必要と考えられる．

謝　辞

本研究は JSPS 科研費基盤研究（A）23H00065，基盤研究（B）20H01720，21H00936，23H00985，若手研究 22K13810 の助成を受けたものです．

参 考 文 献

Balamuta, J. J. and Culpepper, S. A. (2022). Exploratory restricted latent class models with monotonicity requirements under POLYA–GAMMA data augmentation, *Psychometrika*, **87**(3), 903–945, https://doi.org/10.1007/s11336-021-09815-9.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, **4**, Springer, New York.

Blei, D. M., Kucukelbir, A. and McAuliffe, J. D. (2017). Variational inference: A review for statisticians, *Journal of the American statistical Association*, **112**(518), 859–877, https://doi.org/10.1080/01621459.2017.1285773.

Cai, L. (2008). SEM of another flavour: Two new applications of the supplemented EM algorithm, *British Journal of Mathematical and Statistical Psychology*, **61**(2), 309–329, https://doi.org/10.1348/000711007X249603.

Chen, Y., Liu, J., Xu, G. and 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.

Chen, Y., Li, X., Liu, J. and Ying, Z. (2017). Regularized latent class analysis with application in cognitive diagnosis, *Psychometrika*, **82**, 660–692, https://doi.org/10.1007/s11336-016-9545-6.

Chen, Y., Culpepper, S. A., Chen, Y. and Douglas, J. (2018). Bayesian estimation of the DINA Q-matrix, *Psychometrika*, **83**, 89–108, https://doi.org/10.1007/s11336-017-9579-4.

Chen, Y., Culpepper, S. A. and Liang, F. (2020). A sparse latent class model for cognitive diagnosis, *Psychometrika*, **85**, 121–153, https://doi.org/10.1007/s11336-019-09693-2.

Chiu, C.-Y. and Douglas, J. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns, *Journal of Classification*, **30**, 225–250, https://doi.org/10.1007/s00357-013-9132-9.

Chiu, C.-Y. and Köhn, H.-F. (2019). Consistency theory for the general nonparametric classification method, *Psychometrika*, **84**, 830–845, https://doi.org/10.1007/s11336-019-09660-x.

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

Chiu, C.-Y., Sun, Y. and Bian, Y. (2018). Cognitive diagnosis for small educational programs: The general nonparametric classification method, *Psychometrika*, **83**, 355–375, https://doi.org/10.1007/s11336-017-9595-4.

Cho, A. E., Wang, C., Zhang, X. and Xu, G. (2021). Gaussian variational estimation for multidimensional item response theory, *British Journal of Mathematical and Statistical Psychology*, **74**, 52–85, https://doi.org/10.1111/bmsp.12219.

Cho, A. E., Xiao, J., Wang, C. and Xu, G. (2024). Regularized variational estimation for exploratory item factor analysis, *Psychometrika*, **89**, 347–375, https://doi.org/10.1007/s11336-022-09874-6.

Culpepper, S. A. (2019). An exploratory diagnostic model for ordinal responses with binary attributes: Identifiability and estimation, *Psychometrika*, **84**(4), 921–940, https://doi.org/10.1007/s11336-019-09683-4.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic, *Journal of Educational and Behavioral Statistics*, **34**(1), 115–130, https://doi.org/10.3102/1076998607309474.

de la Torre, J. (2011). The generalized DINA model framework, *Psychometrika*, **76**, 179–199, https://doi.org/10.1007/s11336-011-9207-7.

Grimmer, J. (2011). An introduction to Bayesian inference via variational approximations, *Political Analysis*, **19**(1), 32–47, https://doi.org/10.1093/pan/mpq027.

Henson, R. A., Templin, J. L. and Willse, J. T. (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.

Hoijtink, H. (1998). Constrained latent class analysis using the Gibbs sampler and posterior predictive p-values: Applications to educational testing, *Statistica Sinica*, 691–711, https://www.jstor.org/stable/24306458.

Hong, C.-Y., Chang, Y.-W. and Tsai, R.-C. (2016). Estimation of generalized DINA model with order restrictions, *Journal of Classification*, **33**, 460–484, https://doi.org/10.1007/s00357-016-9215-5.

Jamshidian, M. and Jennrich, R. I. (2000). Standard errors for EM estimation, *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), **62**(2), 257–270, https://doi.org/10.1111/1467-9868.00230.

Jeon, M., Rijmen, F. and Rabe-Hesketh, S. (2017). A variational maximization–maximization algorithm for generalized linear mixed models with crossed random effects, *Psychometrika*, **82**, 693–716, https://doi.org/10.1007/s11336-017-9555-z.

Jimenez, A., Balamuta, J. J. and Culpepper, S. A. (2023). A sequential exploratory diagnostic model using a Pólya-gamma data augmentation strategy, *British Journal of Mathematical and Statistical Psychology*, **76**(3), 513–538, https://doi.org/10.1111/bmsp.12307.

Junker, B. W. and 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.

Laudy, O., Boom, J. and Hoijtink, H. (2004). Bayesian computational methods for inequality constrained latent class analysis, *New Developments in Categorical Data Analysis for the Social and Behavioral Sciences*, 63–82, Psychology Press, New York, https://doi.org/10.4324/9781410612021.

Leighton, J. P., Gierl, M. J. and 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, https://doi.org/10.1111/j.1745-3984.2004.tb01163.x.

Liu, X. and Johnson, M. S. (2019). Estimating CDMs using MCMC, *Handbook of Diagnostic Classification Models*: *Models and Model Extensions, Applications, Software Packages* (eds. M. von Davier and Y.-S. Lee), Chapter 31, 629–646, Springer, New York.

Liu, Y., Xin, T., Andersson, B. and Tian, W. (2019). Information matrix estimation procedures for cognitive diagnostic models, *British Journal of Mathematical and Statistical Psychology*, **72**(1), 18–37, https://doi.org/10.1111/bmsp.12134.

Ma, C., de la Torre, J. and Xu, G. (2023a). Bridging parametric and nonparametric methods in cognitive diagnosis, *Psychometrika*, **88**(1), 51–75, https://doi.org/10.1007/s11336-022-09878-2.

Ma, C., Ouyang, J. and Xu, G. (2023b). Learning latent and hierarchical structures in cognitive diagnosis models, *Psychometrika*, **88**(1), 175–207, https://doi.org/10.1007/s11336-022-09867-5.

Ma, W. and Jiang, Z. (2021). Estimating cognitive diagnosis models in small samples: Bayes modal estimation and monotonic constraints, *Applied Psychological Measurement*, **45**(2), 95–111, https://doi.org/10.1177/0146621620977681.

Meng, X.-L. and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm, *Journal of the American Statistical Association*, **86**(416), 899–909.

Nakajima, S., Watanabe, K. and Sugiyama, M. (2019). *Variational Bayesian Learning Theory*, Cambridge University Press, Cambridge.

Philipp, M., Strobl, C., de la Torre, J. and Zeileis, A. (2018). On the estimation of standard errors in cognitive diagnosis models, *Journal of Educational and Behavioral Statistics*, **43**(1), 88–115, https://doi.org/10.3102/1076998617719728.

Rupp, A. A. and Templin, J. (2008a). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model, *Educational and Psychological Measurement*, **68**(1), 78–96, https://doi.org/10.1177/0013164407301545.

Rupp, A. A. and Templin, J. (2008b). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art, *Measurement*, **6**(4), 219–262, https://doi.org/10.1080/15366360802490866.

Rupp, A. A., Templin, J. and Henson, R. A. (2010). *Diagnostic Measurement*: *Theory, Methods, and Applications*, Guilford Press, New York.

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, 日本テスト学会誌, **11**(1), 81–97, https://doi.org/10.24690/jart.11.1_81.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach, *Journal of Educational Statistics*, **10**(1), 55–73, https://doi.org/10.3102/10769986010001055.

Templin, J. and Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus, *Educational Measurement*: *Issues and Practice*, **32**(2), 37–50, https://doi.org/10.1111/emip.12010.

Templin, J. L. and Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models, *Psychological Methods*, **11**(3), 287–305, https://doi.org/10.1037/1082-989X.11.3.287.

von Davier, M. and Lee, Y.-S. (2019). *Handbook of Diagnostic Classification Models*, Springer, New York.

Wang, C. (2021). Using penalized EM algorithm to infer learning trajectories in latent transition CDM, *Psychometrika*, **86**(1), 167–189, https://doi.org/10.1007/s11336-020-09742-1.

Wang, C. and Lu, J. (2021). Learning attribute hierarchies from data: Two exploratory approaches, *Journal of Educational and Behavioral Statistics*, **46**(1), 58–84, https://doi.org/10.3102/1076998620931094.

Wang, S. and Douglas, J. (2015). Consistency of nonparametric classification in cognitive diagnosis, *Psychometrika*, **80**, 85–100, https://doi.org/10.1007/s11336-013-9372-y.

Xu, G. (2017). Identifiability of restricted latent class models with binary responses, *The Annals of Statistics*, **45**(2), 675–707, https://doi.org/10.1214/16-AOS1464.

Xu, G. and Shang, Z. (2018). Identifying latent structures in restricted latent class models, *Journal of* *the American Statistical Association*, **113**(523), 1284–1295, https://doi.org/10.1080/01621459.2017.1340889.

Xu, X., De la Torre, J., Zhang, J., Guo, J. and Shi, N. (2020). Estimating CDMs using the slice-within-gibbs sampler, *Frontiers in Psychology*, **11**, 2260, https://doi.org/10.3389/fpsyg.2020.02260.

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), 21–44, https://doi.org/10.24690/jart.15.1_21.

Yamaguchi, K. (2020). Variational Bayesian inference for the multiple-choice DINA model, *Behaviormetrika*, **47**(1), 159–187, https://doi.org/10.1007/s41237-020-00104-w.

Yamaguchi, K. (2023a). Bayesian analysis methods for two-level diagnosis classification models, *Journal of Educational and Behavioral Statistics*, **48**(6), 773–809, https://doi.org/10.3102/10769986231173594.

Yamaguchi, K. (2023b). On the boundary problems in diagnostic classification models, *Behaviormetrika*, **50**(1), 399–429, https://doi.org/10.1007/s41237-022-00187-7.

Yamaguchi, K. and Martinez, A. J. (2024). Variational Bayes inference for hidden Markov diagnostic classification models, *British Journal of Mathematical and Statistical Psychology*, **77**(1), 55–79, https://doi.org/10.1111/bmsp.12308.

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.

Yamaguchi, K. and Okada, K. (2020a). Hybrid cognitive diagnostic model, *Behaviormetrika*, **47**(2), 497–518, https://doi.org/10.1007/s41237-020-00111-x.

Yamaguchi, K. and Okada, K. (2020b). Variational Bayes inference algorithm for the saturated diagnostic classification model, *Psychometrika*, **85**(4), 973–995, https://doi.org/10.1007/s11336-020-09739-w.

Yamaguchi, K. and Okada, K. (2020c). Variational Bayes inference for the DINA model, *Journal of Educational and Behavioral Statistics*, **45**(5), 569–597, https://doi.org/10.3102/1076998620911934.

Yamaguchi, K. and Templin, J. (2022a). Direct estimation of diagnostic classification model attribute mastery profiles via a collapsed Gibbs sampling algorithm, *Psychometrika*, **87**(4), 1390–1421, https://doi.org/10.1007/s11336-022-09857-7.

Yamaguchi, K. and Templin, J. (2022b). A Gibbs sampling algorithm with monotonicity constraints for diagnostic classification models, *Journal of Classification*, **39**(1), 24–54, https://doi.org/10.1007/s00357-021-09392-7.

Zhang, Z., Zhang, J., Lu, J. and Tao, J. (2020). Bayesian estimation of the DINA model with Pólya-gamma Gibbs sampling, *Frontiers in Psychology*, **11**, 384, https://doi.org/10.3389/fpsyg.2020.00384.

Recent Development of Parameter Estimation Methods in Diagnostic Classification Models

Kazuhiro Yamaguchi

Division of Psychology, Institute of Human Sciences, University of Tsukuba

Diagnostic classification models (DCMs) or cognitive diagnostic models (CDMs) are a family of models that consider elements of cognitive abilities to solve test items named "attributes" and classify test takers into attribute mastery patterns. DCMs are a useful tool to analyze educational tests and have been applied to various tests. However, estimation methods of DCMs have not been assessed in Japan. This study aimed to review the current state of development of estimation methods of DCMs and to contribute to improving the application and theoretical development of DCMs. As a result, estimation methods were developed according to the maximum likelihood estimation, Bayesian estimation, and non-parametric estimation methods. In particular, regularization methods in the maximum likelihood estimation and variational Bayes, which are relatively new methods, were employed. Finally, the remaining estimation problems and future research topics in this area were discussed.

Key words: Diagnostic classification models, cognitive diagnostic models, parameter estimation methods.