

項目反応理論における
能力潜在変数と困難度パラメタの変分ベイズ推定

2024年1月30日

情報数理工学プログラム

学籍番号 2010041

石川 文弥

指導教員 植野 真臣

令和5年度 情報数理工学プログラム卒業論文概要

令和2年度 入学	学籍番号 2010041
指導教員 植野 真臣	氏名 石川 文弥
題目	項目反応理論における 能力潜在変数と困難度パラメタの変分ベイズ推定

概要

e テスティングとは、異なる問題項目を解答したにもかかわらず、受検者の能力を同一精度で評価できるコンピュータテストである。e テスティングには、項目反応理論 (IRT) と呼ばれる確率モデルが用いられる。このIRTによる予測精度向上のためには、モデルパラメタの推定精度が重要となる。変分ベイズ法による項目反応理論のパラメタ推定は、受検者数や項目数が多いデータを高速に推定することができる手法として知られている。

本論では、IRT の能力値パラメタと困難度パラメタを変分ベイズ法により正規分布に近似してパラメタ推定を行う手法を提案する。

Cho らによる GVEM では、能力値パラメタを変分ベイズ法により正規分布に近似することで項目パラメタの推定を大幅に高速化した。しかし、GVEM ではデータサイズが小さい場合に推定できない場合がある。また、GVEM では項目パラメタを互いに所与として推定するため、各パラメタの誤差が相互に影響し合い、推定精度が低下している可能性がある。

そこで本論では、能力値パラメタと困難度パラメタを変分ベイズ法によりベイズ推定し、能力値パラメタと困難度パラメタを周辺化して識別力パラメタを推定する。

提案手法には以下の利点がある。(1) 推定時間が短く、困難度パラメタもベイズ推定することで、GVEM に比べてデータサイズが小さい場合でも推定可能である。(2) 能力値パラメタと困難度パラメタを周辺化して識別力パラメタを推定することで、推定精度を向上できる。

シミュレーションデータによる比較実験で、提案手法がデータサイズが小さい場合に従来手法よりも推定精度を向上できることを示す。また、GVEM よりも収束までに必要なアルゴリズムの反復数が少ないことを比較実験により示す。

1 まえがき

近年, e テスティングと呼ばれるコンピュータ上で行われるテスト形式が普及している. e テスティングを用いることで, 異なる問題項目を解答したにもかかわらず, 受検者の能力を同一精度で評価することができる. これにより, 受検者は同一精度のテストを複数回受検可能となる [1, 2]. e テスティングは, 医療系大学間共用試験や情報処理技術者試験などで既に実用化されている [3, 4].

e テスティングでは, 一般的に項目反応理論 (Item Response Theory; IRT)[5, 6] と呼ばれる確率モデルを用いる. IRT では, 2 パラメタロジスティックモデル (2PLM) が一般的に使用される. 2PLM は, 受検者の能力を表す潜在変数 (能力値パラメタ) と, 問題項目の困難度と識別力を表す 2 つのパラメタ (項目パラメタ) により, 受検者の各問題項目への正答確率を予測できる. IRT による予測精度を向上させるためには, 各受検者の問題項目への解答データによる 2PLM の各パラメタの推定精度が重要である.

IRT のパラメタ推定には, ベイズ的な手法であるマルコフ連鎖モンテカルロ法 (Markov Chain Monte Carlo; MCMC)[7] が一般的に使用される. MCMC 法は, マルコフ連鎖の特性を利用して事後分布に従う乱数を生成することで, 事後分布の近似を行う. MCMC 法の利点は, サンプルングを十分に繰り返すことで真の事後分布に近い分布が得られることであるが, 計算コストが非常に大きくなる可能性がある. また, IRT のパラメタ推定においては受検者数や項目数が増加すると未知パラメタ数が増加し, 一回のサンプルングに必要な計算コストが大きくなる. そのため, 受検者数や項目数が多い解答データに対しては, 莫大な計算時間がかかり MCMC 法による推定が困難となる場合がある.

この問題を解決するために, 変分ベイズ法 (Variational Bayesian method; VB method) が提案されている [8]. VB 法は事後分布に制約を与え, 事後分布の近似分布 (変分事後分布) を最適化問題により導出することで計算速度の向上を図る手法であり, 自然言語処理や音声認識, 遺伝学など多様な分野に適用されている [9, 10, 11]. VB 法は事後分布を単純な関数形で近似することが可能であり, 解答データの受検者数や項目数が多い場合に, 計算速度を MCMC よりも著しく向上できる [12]. 近年, この VB 法を利用した IRT のパラメタ推定手法が数多く提案されている [13, 14, 15, 16, 17]. これらの VB 法による IRT のパラメタ推定は, 多くの場合で MCMC 法に匹敵する精度により推定可能であることが知られている [12, 15, 18].

Cho ら [14] は, 変分事後分布を正規分布に近似して推定に利用する GVEM 法 (Gaussian Variational Expectation-Maximization method) を提案した. GVEM 法では, 能力潜在変数の変分事後分布を正規分布に近似し, その分布を EM アルゴリズム [19] に用いて項目パラメタを推定する. GVEM 法では, EM アルゴリズムにおける能力潜在変数の条件付き分布に, 正規分布に近似した変分事後分布を用いる. これにより, GVEM 法は勾配法などの数値最適化を使用せず, 区分求積法による積分計算を含まない更新式の反復による推定を可能とし, 計算速度を向上させた. さらに, Cho らはシミュレーション実験により, GVEM 法は CJMLE[20] よりもバイアスが少なくことや MHRM アルゴリズム [21] より小さい平均二乗誤差を示すことなどを報告した.[14]. しかし, GVEM 法は解答データの受検者数や項目数が少ない場合に, 収束せず推定が不安定になる場合がある. これは, GVEM 法の EM ア

ルゴリズムによる推定値は、データ数が少ない場合にデータのばらつきの影響を大きく受けるためである。一般に、このような場合はベイズ推定が有効であることが知られている [22, 23].

この課題を解決するために、本研究では GVEM 法における能力潜在変数と困難度パラメタの変分事後分布を導出し、正規分布に近似する手法を提案する。提案手法は、能力潜在変数と困難度パラメタを VB 法によりベイズ推定することで、解答データの受検者数や項目数が少ない場合でも安定した推定が可能となる。また、困難度パラメタは事前分布に正規分布を仮定することが一般的であり、事後分布も正規分布の形状をとることが多い [24]. 提案手法では、能力潜在変数と困難度パラメタの変分事後分布を正規分布に正確に近似することで、アルゴリズムに含まれる期待値計算を高精度に近似できる。そのため、提案手法は GVEM 法と同様にアルゴリズムに数値最適化を含まず、積分計算を含まない更新式による推定を可能にする。さらに、提案手法における識別力パラメタは能力潜在変数に加えて困難度パラメタを周辺化した周辺尤度の最大化によって推定されるため、困難度パラメタの誤差の影響が軽減され、推定精度を向上できる。

本研究では、シミュレーション実験により提案手法を従来手法と比較した。その結果、提案手法は多くの条件で従来手法とほぼ同程度のパラメタ推定精度を保ちつつ、解答データの受検者数が少ない場合に識別力の推定精度を従来手法よりも向上させることができた。さらに、提案手法は従来手法よりも推定時間を削減可能し、GVEM 法よりも少ない EM アルゴリズムの反復による推定が可能であることを示した。

2 項目反応理論

項目反応理論 (Item Response Theory; IRT)[5, 6] は確率モデルを用いたテスト理論のひとつであり, e テスティングに用いられている [1, 2]. IRT は, 受検者の能力や項目の特性を表すパラメータを用いた確率モデルにより受検者の各問題項目への正答確率を予測できる.

項目反応理論の一般的なモデルの一つとして, 2 母数ロジスティックモデル (2-Parameter Logistic Model; 2PLM) がある. 2PLM は, 能力値が $\theta \in [-\infty, \infty]$ のある受検者 $i \in \{1, 2, \dots, N\}$ が項目 $j \in \{1, 2, \dots, J\}$ に正答する確率を以下のように表す.

$$p(y_{ij} = 1|\theta) = \frac{1}{1 + \exp(-a_j(\theta_i - b_j))} \quad (1)$$

ここで, y_{ij} は受検者 i が項目 j に正答すれば 1, それ以外の場合は 0 をとる変数である. また, $a_j \in [0, \infty], b_j \in [-\infty, \infty]$ はそれぞれ項目 j の識別力, 困難度を示すパラメータである. このモデルにおいて識別力 $a_{ij} = 1$ としたものは 1 母数ロジスティックモデル (1-Parameter Logistic Model; 1PLM), あるいは Rasch モデルと呼ばれる.

IRT のパラメータ推定には, 次に示す尤度関数を一般的に使用する. 受検者を N 人, 項目数を J 個とした場合, 受検者 i の項目 j への反応の尤度は以下の式で表される.

$$L(y_{ij}|\theta_i, a_j, b_j) = p(y_{ij} = 1|\theta)^{y_{ij}} (1 - p(y_{ij} = 1|\theta_i))^{1-y_{ij}} \quad (2)$$

さらに, 局所独立と個人の交換可能性を仮定することで, 項目パラメータと能力値の同時尤度は以下の式で表される [25].

$$L(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}) = \prod_{i=1}^N \prod_{j=1}^J p(y_{ij} = 1|\theta)^{y_{ij}} (1 - p(y_{ij} = 1|\theta_i))^{1-y_{ij}} \quad (3)$$

ここで, $\mathbf{a} = (a_1, \dots, a_j, \dots, a_J)$, $\mathbf{b} = (b_1, \dots, b_j, \dots, b_J)$ は, 項目ごとの識別力と困難度を表す J 次元ベクトルである. また, \mathbf{Y} は受検者ごとの各項目反応を表す $N \times J$ 行列である ((4) 式).

$$\mathbf{Y} = \begin{pmatrix} y_{11} & \dots & y_{1j} & \dots & y_{1J} \\ \vdots & & \vdots & & \vdots \\ y_{i1} & \dots & y_{ij} & \dots & y_{iJ} \\ \vdots & & \vdots & & \vdots \\ y_{N1} & \dots & y_{Nj} & \dots & y_{NJ} \end{pmatrix} \quad (4)$$

IRT のパラメータ推定には, 尤度 ((3) 式) のパラメータを周辺化した周辺尤度を用いる場合がある. 能力潜在変数 $\boldsymbol{\theta}$ を事前分布 $\pi(\cdot)$ で周辺化した場合の周辺尤度は以下の通りである.

$$p(\mathbf{Y}|\mathbf{a}, \mathbf{b}) = \int_{\boldsymbol{\theta}} L(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{a}, \mathbf{b})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \quad (5)$$

3 IRT のパラメタ推定法

IRT のパラメタ推定には, マルコフ連鎖モンテカルロ法 (Markov Chain Monte Carlo method; MCMC method)[7] が一般的に使用される. MCMC は, ベイズ推定の枠組みの中で, 数値的な解法により未知パラメタを推定する手法である.

3.1 ベイズ推定法

ベイズ推定は, 観測データ \mathbf{Y} と未知パラメタベクトル γ の事前分布 $\pi(\gamma)$ を用いて, 以下のベイズの定理により γ の事後分布を求める.

$$\begin{aligned} p(\gamma|\mathbf{Y}) &= \frac{p(\mathbf{Y}, \gamma)}{p(\mathbf{Y})} \\ &= \frac{L(\mathbf{Y}|\gamma)\pi(\gamma)}{p(\mathbf{Y})} \end{aligned} \quad (6)$$

さらに, ベイズ推定では得られた事後分布を用いて様々な推定量を計算して評価を行う. 例えば, 事後期待値による推定値 (EAP 推定量) は以下のように定義される [26].

$$\hat{\gamma}_{eap} = \int \gamma p(\gamma|\mathbf{Y}) d\gamma \quad (7)$$

しかし, 周辺尤度 $p(\mathbf{Y}) = \int L(\mathbf{Y}|\gamma)\pi(\gamma)d\gamma$ により (6) 式や (7) 式には多重積分が含まれ, これらを解析的に解くことはほぼ不可能である.

3.2 MCMC 法

MCMC 法は, マルコフ連鎖の特性を利用したサンプリングにより事後分布を近似する. マルコフ連鎖は, 次の状態が現在の状態にのみ依存し過去の状態には依存しない確率過程であり, 適切に設計したマルコフ連鎖により導かれた確率分布は, 定常分布に収束する. MCMC 法は, 定常分布が事後分布と一致するようにマルコフ連鎖を設計し, 事後分布に従う乱数を順次生成することで事後分布を近似する. マルコフ連鎖の設計方法については様々な手法が存在するが, MCMC 法のアルゴリズムを効率的に行うハミルトニアンモンテカルロ法 (Hamiltonian Monte Carlo; HMC) が一般的に使用される [27].

HMC 法は, ハミルトニアンと呼ばれる物理量を利用して効率的なサンプリングを行う手法である. ハミルトニアンは, 物体の有する力学的エネルギーの総和を表し, 物体の位置と運動量を直行軸とする位相空間で表現される. また, 力学的エネルギー保存の法則によりハミルトニアンは一定の値をとる. HMC 法では, S 次元の未知パラメタベクトル $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_S)$ を物体の位置と見なし, 運動量を独立な S 個の標準正規分布からの乱数 $\rho = (\rho_1, \rho_2, \dots, \rho_S)$ とすることで, パラメタ空間の探索を位相空間上の物体の運動になぞらえて行う. HMC 法におけるハミルトニアンの定義は以下の通りであ

る [26].

$$H(\boldsymbol{\gamma}, \boldsymbol{\rho}) = h(\boldsymbol{\gamma}) + \frac{1}{2} \sum_{s=1}^S \rho_s^2$$

ここで, $h(\boldsymbol{\gamma})$ は位置 $\boldsymbol{\gamma}$ に対応する高さを表す関数である. HMC では $h(\boldsymbol{\gamma}) = -\log p(\boldsymbol{\gamma}|\mathbf{Y})$ とおくことで, 事後分布 $p(\boldsymbol{\gamma})$ とそれとは独立な標準正規分布 $p(\boldsymbol{\rho})$ の同時分布 $p(\boldsymbol{\gamma}, \boldsymbol{\rho}|\mathbf{Y})$ をハミルトニアンを用いて表すことができる.

$$p(\boldsymbol{\gamma}, \boldsymbol{\rho}|\mathbf{Y}) \propto \exp(-H(\boldsymbol{\gamma}, \boldsymbol{\rho}))$$

HMC 法は, $p(\boldsymbol{\gamma}, \boldsymbol{\rho}|\mathbf{Y})$ から乱数を生成する. $p(\boldsymbol{\gamma})$ と $p(\boldsymbol{\rho})$ は互いに独立であるため, $p(\boldsymbol{\gamma}, \boldsymbol{\rho}|\mathbf{Y})$ からの乱数 $\boldsymbol{\gamma}$ は $p(\boldsymbol{\gamma})$ からの乱数と等価である. HMC 法における乱数は, 位置 $\boldsymbol{\gamma}$ の物体に対して運動量 $\boldsymbol{\rho}$ を与えた場合の, 遷移先の位置をリープフロッグ法 [27] により予測することで生成される. HMC 法において, $h(\boldsymbol{\gamma})$ の定義より対数事後確率が高さを物理空間はの低さと見なすことができる. 物理空間上では, 運動する物体は低い所へと遷移するため, 乱数を生成するごとに, 対数事後確率が高くなる方向へパラメタを遷移することができる. リープフロッグ法では数値計算上の誤差が生じるため, Metropolis-Hastings 法 [28] と同様に以下の補正係数を用いて確率補正を行う.

$$r = \frac{p(\boldsymbol{\gamma}', \boldsymbol{\rho}')}{p(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\rho}^{(t)})} = \exp(H(\boldsymbol{\gamma}^{(t)}, \boldsymbol{\rho}^{(t)}) - H(\boldsymbol{\gamma}', \boldsymbol{\rho}')), \quad (8)$$

HMC のアルゴリズムは以下の通りである [26]. なお, ϵ, L はリープフロッグ法で用いられるチューニングパラメタであり, T は繰り返し回数を表す.

1. 初期値 $\boldsymbol{\gamma}^{(t)}, \epsilon, L, T$, バーンイン期間を定める. $t = 1$ とする.
2. 独立な S 個の標準正規乱数 $\boldsymbol{\rho}^{(t)}$ を発生させる.
3. リープフロッグ法で変移させ, 更新先の候補値 $\boldsymbol{\gamma}', \boldsymbol{\rho}'$ を生成する.
4. 確率 $\alpha = \min\{1, r\}$ に従って候補値 $\boldsymbol{\gamma}'$ を受容し, $\boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}'$ とする. 受容されなかった場合は, $\boldsymbol{\gamma}^{(t+1)} = \boldsymbol{\gamma}^{(t)}$ とする.
5. $T = t$ ならば終了. そうでなければ $t = t + 1$ としてステップ 2 に戻る.

理論的にはハミルトニアンは保存されるが, リープフロッグ法による数値計算上の誤差により, $\boldsymbol{\gamma}', \boldsymbol{\rho}'$ のハミルトニアンは変化する. そのため, (8) 式右辺の指数部分はほぼ 0 となり, r は 1 に近い値を取る. これにより, HMC 法は, 高い受容率による効率的なサンプリングが可能となる.

HMC 法は適切なチューニングを行う必要があり, 実装が難しい場合がある. 近年では, チューニングパラメタを自動調整するアルゴリズム (No-U-Trun Sampler; NUTS) が提案され, HMC 法の実用性が向上している [29].

MCMC 法は, サンプリングを十分に繰り返すことで真の事後分布に近い分布を推定可能だが, 計算コストが非常に大きくなる可能性がある. また, IRT のパラメタ推定においては受検者数や項目数が増加すると未知パラメタ数が増加し, 一回のサンプリングに必要な計算コストが大きくなる. そのため, 大規模なデータセットや複雑なモデルを扱う際には莫大な計算時間が必要となり, MCMC 法による推定が困難な場合がある.

4 変分ベイズ法を利用した IRT のパラメタ推定法

MCMC 法の課題を解決するために、変分ベイズ法 (Variational Bayes method; VB method, あるいは Variational Inference; VI) が提案されている。VB 法は事後分布に制約を与え、事後分布の近似分布 (変分事後分布) を最適化問題として導出することで計算速度の向上を図る手法であり、自然言語処理や音声認識、遺伝学など多様な分野に適応されている [9, 10, 11]。

4.1 変分ベイズ法

変分ベイズ法では、周辺尤度の下界の最大化を目的関数とした最適化問題を解くことで変分事後分布を定式化できる。まず、推定するパラメタ γ の事後分布 $p(\gamma|\mathbf{Y})$ を別の確率分布 $q(\gamma)$ で近似するために、KL ダイバージェンスの最小化問題を考える。KL ダイバージェンス (Kullback-Leibler divergence) とは、2 つの確率分布間の類似度を表す指標であり、以下のように定義される。

$$\begin{aligned} KL(q(\gamma)||p(\gamma|\mathbf{Y})) &= \int q(\gamma) \left[\log \frac{q(\gamma)}{p(\gamma|\mathbf{Y})} \right] d\gamma \\ &= \mathbb{E}_{q(\gamma)} \left[\log \frac{q(\gamma)}{p(\gamma|\mathbf{Y})} \right] \end{aligned} \quad (9)$$

KL ダイバージェンスは非負の値を取り、0 に近い値であるほど分布間の類似度は高いことを示す。次に、ベイズの定理 ((6) 式) により、(9) 式を変形すると以下ようになる。

$$\begin{aligned} KL(q(\gamma)||p(\gamma|\mathbf{Y})) &= \mathbb{E}_{q(\gamma)} \left[\log q(\gamma) \frac{p(\mathbf{Y})}{L(\mathbf{Y}, \gamma)} \right] \\ &= \log p(\mathbf{Y}) - (\mathbb{E}_{q(\gamma)} [\log p(\mathbf{Y}, \gamma)] - \mathbb{E}_{q(\gamma)} [\log q(\gamma)]) \end{aligned} \quad (10)$$

KL ダイバージェンスは非負の値をとるため、(10) 式から対数周辺尤度の下界 $\underline{l}(\mathbf{Y}; q(\gamma))$ が以下のように導出される。

$$\begin{aligned} \log p(\mathbf{Y}) &= \mathbb{E}_{q(\gamma)} [\log p(\mathbf{Y}, \gamma)] - \mathbb{E}_{q(\gamma)} [\log q(\gamma)] + KL(q(\gamma)||p(\gamma|\mathbf{Y})) \\ &\geq \mathbb{E}_{q(\gamma)} [\log p(\mathbf{Y}, \gamma)] - \mathbb{E}_{q(\gamma)} [\log q(\gamma)] \\ &=: \underline{l}(\mathbf{Y}; q(\gamma)) \end{aligned} \quad (11)$$

$\log p(\mathbf{Y})$ は γ に依存しないので定数とすれば、KL ダイバージェンスの最小化問題は $\underline{l}(\mathbf{Y}; q(\gamma))$ の最大化問題と等価である。そのため、VB 法では $\underline{l}(\mathbf{Y}; q(\gamma))$ の最大化を目的関数とした最適化問題を解くことで、事後分布の近似分布 $q(\gamma)$ を導出する。

$\underline{l}(\mathbf{Y}; q(\gamma))$ の最大化問題を解くために、VB 法では以下に示した独立性の仮定を目的関数の制約として導入する。

$$q(\boldsymbol{\gamma}) = \prod_{s=1}^S q(\gamma_s) \quad (12)$$

また, 事前分布 $\pi(\boldsymbol{\gamma})$ の独立性を (12) 式と同様に仮定する. $\underline{l}(\mathbf{Y}; q(\boldsymbol{\gamma}))$ の式は独立性の仮定により以下のように変形できる.

$$\begin{aligned} \underline{l}(\mathbf{Y}; q(\boldsymbol{\gamma})) &= \mathbb{E}_{q(\boldsymbol{\gamma})} [\log p(\mathbf{Y}, \boldsymbol{\gamma})] - \mathbb{E}_{q(\boldsymbol{\gamma})} [\log q(\boldsymbol{\gamma})] \\ &= \int_{\boldsymbol{\gamma}} q(\boldsymbol{\gamma}) (\log \pi(\boldsymbol{\gamma}) + \log p(\mathbf{Y}|\boldsymbol{\gamma}) - \log q(\boldsymbol{\gamma})) d\boldsymbol{\gamma} \\ &= \int_{\boldsymbol{\gamma}} \left(\prod_{s=1}^S q_s(\gamma_s) \right) \left(\log \prod_{s=1}^S \pi(\gamma_s) + \log p(\mathbf{Y}|\boldsymbol{\gamma}) - \log \prod_{s=1}^S q_s(\gamma_s) \right) d\boldsymbol{\gamma} \\ &= \int_{\boldsymbol{\gamma}} \left(q_s(\gamma_s) \prod_{s' \neq s} q_{s'}(\gamma_{s'}) \right) \left(\sum_{s=1}^S \log \pi(\gamma_s) + \log p(\mathbf{Y}|\boldsymbol{\gamma}) - \sum_{s=1}^S \log q_s(\gamma_s) \right) d\boldsymbol{\gamma} \quad (13) \end{aligned}$$

(13) 式を関数 $q_s(\gamma_s)$ で微分し, その微分値を 0 と置くことで $\underline{l}(\mathbf{Y}; q(\boldsymbol{\gamma}))$ を最大化する最適な変分事後分布が導出される. $\underline{l}(\mathbf{Y}; q(\boldsymbol{\gamma}))$ を関数 $q_s(\gamma_s)$ で微分すると以下のようになる.

$$\begin{aligned} \frac{\partial \underline{l}(\mathbf{Y}; q(\boldsymbol{\gamma}))}{\partial q_s(\gamma_s)} &= \int_{\boldsymbol{\gamma}} \left(\prod_{s' \neq s} q_{s'}(\gamma_{s'}) \right) \left(\sum_{s=1}^S \log \pi(\gamma_s) + \log p(\mathbf{Y}|\boldsymbol{\gamma}) - \sum_{s=1}^S \log q_s(\gamma_s) \right) \\ &\quad + \left(q_s(\gamma_s) \prod_{s' \neq s} q_{s'}(\gamma_{s'}) \right) \left(-\frac{1}{q_s(\gamma_s)} \right) d\boldsymbol{\gamma} \\ &= \int_{\boldsymbol{\gamma}} \left(\prod_{s' \neq s} q_{s'}(\gamma_{s'}) \right) \left(\sum_{s=1}^S \log \pi(\gamma_s) + \log p(\mathbf{Y}|\boldsymbol{\gamma}) - \sum_{s=1}^S \log q_s(\gamma_s) - 1 \right) d\boldsymbol{\gamma} \\ &= \mathbb{E}_{\prod_{s' \neq s} q_{s'}(\gamma_{s'})} \left[\sum_{s=1}^S \log \pi(\gamma_s) + \log p(\mathbf{Y}|\boldsymbol{\gamma}) - \sum_{s=1}^S \log q_s(\gamma_s) \right] - 1 \\ &= \mathbb{E}_{\prod_{s' \neq s} q_{s'}(\gamma_{s'})} \left[\sum_{s' \neq s} \log \pi(\gamma_{s'}) + \log p(\mathbf{Y}|\boldsymbol{\gamma}) - \sum_{s' \neq s} \log q_{s'}(\gamma_{s'}) \right] + \log \frac{\pi(\gamma_s)}{q_s(\gamma_s)} - 1 \\ &= \mathbb{E}_{\prod_{s' \neq s} q_{s'}(\gamma_{s'})} [\log p(\mathbf{Y}|\boldsymbol{\gamma})] + \log \frac{\pi(\gamma_s)}{q_s(\gamma_s)} + \text{const} \quad (14) \end{aligned}$$

(14) 式の値を 0 とおくことで, 最適な変分事後分布は以下の通りとなる.

$$q_s(\gamma_s) \propto \pi(\gamma_s) \exp \left(\mathbb{E}_{\prod_{s' \neq s} q_{s'}(\gamma_{s'})} [\log L(\mathbf{Y}|\boldsymbol{\gamma})] \right) \quad (15)$$

なお, (15) 式は変分事後分布の関数形を示す. $q_s(\gamma_s)$ を確率分布として扱うためには, (15) 式を規格化する必要がある.

VB法は、(15)式により各パラメタの変分事後分布を逐次更新していくことで推定できる。変分事後分布は単純な関数形で近似されるため、VB法はMCMC法より計算速度を向上させることができる。特に解答データの受検者数や項目数が多い場合に、VB法はMCMC法よりも大幅に計算速度を高速化することが可能である[12]。近年、VB法を利用したIRTのパラメタ推定法が提案されている[13, 14, 15, 16, 17]。これらのVB法によるIRTのパラメタ推定は、多くの場合でMCMC法に匹敵する精度で推定可能であることが知られている[12, 15, 18]。

4.2 能力潜在変数の変分事後分布を用いたEMアルゴリズム

VB法を利用した最新のIRTのパラメタ推定法として、ChoらはGVEMアルゴリズム(Gaussian Variational Expectation-Maximization)を提案した[14]。GVEMは、能力潜在変数の変分事後分布を正規分布に近似し、その分布をEMアルゴリズム[19]に用いて項目パラメタを推定する。

能力潜在変数の変分事後分布は4.1節と同様に導出できる。受検者の独立性の仮定と(11)式より、対数周辺尤度の下界 $\underline{l}(\mathbf{Y}_i; q_i(\theta_i))$ は以下のように表される。

$$\underline{l}(\mathbf{Y}_i; q(\theta_i)) =: \sum_{i=1}^N \mathbb{E}_{q_i(\theta_i)} [\log p(\mathbf{Y}_i, \theta_i | \mathbf{a}, \mathbf{b})] - \sum_{i=1}^N \mathbb{E}_{q_i(\theta_i)} [\log q_i(\theta_i)] \quad (16)$$

GVEMでは、 $\underline{l}(\mathbf{Y}_i; q(\theta_i))$ の導出に以下のBishop, Jaakkolaら[8, 30]によるシグモイド関数の下界を利用する。

$$\begin{aligned} \frac{e^{x_{ij}}}{1 + e^{x_{ij}}} &= \max_{\xi_{ij}} \frac{e^{\xi_{ij}}}{1 + e^{\xi_{ij}}} \exp \left\{ \frac{x_{ij} - \xi_{ij}}{2} - \eta(\xi_{ij}) (x_{ij}^2 - \xi_{ij}^2) \right\} \\ &\geq \frac{e^{\xi_{ij}}}{1 + e^{\xi_{ij}}} \exp \left\{ \frac{x_{ij} - \xi_{ij}}{2} - \eta(\xi_{ij}) (x_{ij}^2 - \xi_{ij}^2) \right\} \end{aligned} \quad (17)$$

ここで、 ξ_{ij} は目的関数 $e^{x_{ij}}/1 + e^{x_{ij}}$ を近似するためのパラメタであり、Mステップで項目パラメタと共に推定される。また、 $\eta(\xi_{ij}) = (2\xi_{ij})^{-1} [e^{\xi_{ij}}/(1 + e^{\xi_{ij}}) - 1/2]$ である。

$x_{ij} = -a_j(\theta_i - b_j)$ として(17)を利用することで、受検者*i*における対数尤度 $\log L(\mathbf{Y}_i | \theta_i, \mathbf{a}, \mathbf{b})$ の下界が以下のように求められる。

$$\begin{aligned} \log L(\mathbf{Y}_i | \theta_i, \mathbf{a}, \mathbf{b}) &= \sum_{j=1}^J \left\{ y_{ij} \log \frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))} + (1 - y_{ij}) \log \frac{1}{1 + \exp(a_j(\theta_i - b_j))} \right\} \\ &= \sum_{j=1}^J \left\{ y_{ij} (a_j(\theta_i - b_j)) + \log \frac{\exp(-a_j(\theta_i - b_j))}{1 + \exp(-a_j(\theta_i - b_j))} \right\} \\ &\geq \sum_{j=1}^J \log \frac{e^{\xi_{ij}}}{1 + e^{\xi_{ij}}} + \sum_{j=1}^J y_{ij} (a_j(\theta_i - b_j)) + \sum_{j=1}^J \frac{((b_j - \theta_i)a_j - \xi_{ij})}{2} \\ &\quad - \sum_{j=1}^J \eta(\xi_{ij}) \left(((b_j - \theta_i)a_j)^2 - \xi_{ij}^2 \right) \end{aligned} \quad (18)$$

したがって、 θ_i と \mathbf{Y}_i の同時分布の対数 $\log p(\mathbf{Y}_i, \theta_i | \mathbf{a}, \mathbf{b})$ の下界は以下の通りとなる。

$$\begin{aligned}
\log p(\mathbf{Y}_i, \theta_i | \mathbf{a}, \mathbf{b}) &= \log L(\mathbf{Y}_i | \theta_i, \mathbf{a}, \mathbf{b}) + \log \pi(\theta_i) \\
&\geq \sum_{j=1}^J \log \frac{e^{\xi_{ij}}}{1 + e^{\xi_{ij}}} + \sum_{j=1}^J y_{ij} (a_j (\theta_i - b_j)) + \sum_{j=1}^J \frac{((b_j - \theta_i) a_j - \xi_{ij})}{2} \\
&\quad - \sum_{j=1}^J \eta(\xi_{ij}) \left(((b_j - \theta_i) a_j)^2 - \xi_{ij}^2 \right) + \log \pi(\theta_i) \\
&=: l(\mathbf{Y}_i, \theta_i, \boldsymbol{\xi}_i | \mathbf{a}, \mathbf{b})
\end{aligned} \tag{19}$$

(19) 式より $l(\mathbf{Y}_i; q(\theta_i))$ を最大化する能力潜在変数の最適な変分事後分布 $q_i(\theta_i)$ を定式化することができる。受検者 i の能力潜在変数の事前分布を $\pi(\theta_i) \sim N(0, 1)$ とすると、 θ_i の変分事後分布の対数 $\log q_i(\theta_i)$ は以下のように求められる。

$$\log q_i(\theta_i) \propto \sum_{j=1}^J \left(y_{ij} - \frac{1}{2} \right) a_j \theta_i - \sum_{j=1}^J \eta(\xi_{ij}) a_j^2 (b_j - \theta_i)^2 - \frac{\theta_i^2}{2} \tag{20}$$

Bishop の文献 [8] より、項目数 J が適度に大きくなると事後分布は正規分布に近似できることが知られている。得られた (20) 式を正規分布の形に変形することで、近似した正規分布のパラメータ $q_i(\theta_i) \sim N(\mu_i, \sigma_i^2)$ が以下のように得られる。

$$\mu_i = \sigma_i^2 \times \sum_{j=1}^J \left\{ 2\eta(\xi_{ij}) a_j b_j + y_{ij} - \frac{1}{2} \right\} a_j \tag{21}$$

$$\sigma_i^2 = \frac{1}{2 \sum_{j=1}^J \eta(\xi_{ij}) a_j^2 + 1} \tag{22}$$

GVEM では、導出した変分事後分布を利用して、EM アルゴリズム [19] により項目パラメタを推定する。

E ステップでは、(21), (22) 式により能力潜在変数の変分事後分布を更新し、更新した変分事後分布を用いて $l(\mathbf{Y}_i; \theta_i, \boldsymbol{\xi}_i | \mathbf{a}, \mathbf{b})$ の期待値 $E(\mathbf{a}, \mathbf{b}, \boldsymbol{\xi})$ を求める。 $E(\mathbf{a}, \mathbf{b}, \boldsymbol{\xi})$ は以下のように求められる。

$$\begin{aligned}
E(\mathbf{a}, \mathbf{b}, \boldsymbol{\xi}) &= \sum_{i=1}^N \int_{\theta_i} l(\mathbf{Y}_i, \theta_i, \boldsymbol{\xi}_i | \mathbf{a}, \mathbf{b}) \times q_i(\theta_i) d\theta_i \\
&= \sum_{i=1}^N \sum_{j=1}^J \left(\log \frac{e^{\xi_{ij}}}{1 + e^{\xi_{ij}}} + \left(y_{ij} - \frac{1}{2} \right) \frac{a_j (\mu_i - b_j)}{2} - \frac{\xi_{ij}}{2} \right) \\
&\quad - \sum_{i=1}^N \sum_{j=1}^J \left(\eta(\xi_{ij}) \left((b_j^2 - 2b_j \mu_i + \mu_i^2 + \sigma_i^2) a_j^2 - \xi_{ij}^2 \right) \right) - \sum_{i=1}^N \left(\frac{1}{2} \log 2\pi + \frac{\mu_i^2 + \sigma_i^2}{2} \right)
\end{aligned} \tag{23}$$

M ステップでは、 $E(\mathbf{a}, \mathbf{b}, \boldsymbol{\xi})$ を最大化する項目パラメタと $\boldsymbol{\xi}$ を更新する。各パラメタの更新式は、 $E(\mathbf{a}, \mathbf{b}, \boldsymbol{\xi})$ の導関数を 0 と置くことで得られる。各パラメタの更新式は以下の通りである。

$$a_j = \frac{1}{2} \left[\sum_{i=1}^N \eta(\xi_{ij}) (b_j^2 - 2b_j\mu_i + \mu_i^2 + \sigma_i^2) \right]^{-1} \left(\sum_{i=1}^N \left(y_{ij} - \frac{1}{2} \right) (\mu_i - b_j) \right) \quad (24)$$

$$b_j = \frac{\sum_{i=1}^N \left[\left(\frac{1}{2} - y_{ij} \right) + 2\eta(\xi_{ij}) a_j \mu_i \right]}{\sum_{i=1}^N 2\eta(\xi_{ij})} \quad (25)$$

$$\xi_{ij}^2 = a_j^2 (b_j^2 - 2b_j\mu_i + \mu_i^2 + \sigma_i^2) \quad (26)$$

2PLM のパラメタ推定のための、GVEM アルゴリズムは以下のようになる。

1. 項目パラメタ $\mathbf{M}^{(t)} = \{\mathbf{a}^{(t)}, \mathbf{b}^{(t)}\}$ とパラメタ $\xi_{ij}^{(t)}$ の初期値を設定. $t = 0$ とする.
2. E ステップ: (21), (22) 式を用いて $\mu_i^{(t+1)}, \sigma_i^{(t+1)}$ を更新.
3. M ステップ: (24), (25), (26) 式を用いて $\mathbf{M}^{(t+1)}, \xi_{ij}^{(t+1)}$ を更新.
4. 収束条件を満たしていれば推定終了. そうでなければ, $t = t + 1$ としてステップ 2 に戻る.

GVEM は, E・M ステップで行われる更新に積分計算を含まず数値最適化を使用しないアルゴリズムであるため, 計算速度を向上できる. Cho らは, シミュレーション実験により CJMLE よりもバイアスが少なかったことや, MHRM アルゴリズムより小さい平均二乗誤差を示すことなどを報告した. さらに, 一定の条件のもと, サンプルサイズと項目数が大きくなるにつれて, 推定値と真値の間の誤差の期待値が 0 に近づくことも示した [14].

しかし, GVEM は, 受検者数や項目数が少ない解答データの推定時に, 推定が不安定になる場合がある. これは, GVEM の EM アルゴリズムによる推定値は, データ数が少ない場合にデータのばらつきの影響を大きく受けるためである. 一般に, このような場合はベイズ推定が有効であることが知られている [22, 23].

5 提案手法

この課題を解決するために, 本研究では GVEM における能力潜在変数と困難度パラメタの事後分布を VB 法により導出し, 正規分布に近似する手法を提案する. 提案手法は, 能力潜在変数と困難度パラメタを VB 法によりベイズ推定することで, 受検者数や項目数が少ない解答データでも安定した推定が可能となる. また, 困難度パラメタは事前分布に正規分布を仮定することが一般的であり, 事後分布は正規分布に近い分布となることが多い [24]. 提案手法における EM アルゴリズムでは, 正規分布に近似した変分事後分布により, 能力潜在変数と困難度パラメタによる対数尤度の期待値が計算される. そのため, 識別力パラメタの推定において困難度パラメタの推定誤差の影響が軽減され, 推定精度を向上できる.

5.1 変分事後分布の導出

能力値と困難度の変分事後分布を導出するために、以下の独立性を仮定する.

$$q(\boldsymbol{\theta}, \mathbf{b}) = \prod_{i=1}^N q_i(\theta_i) \prod_{j=1}^J q_j(b_j) \quad (27)$$

独立性の仮定と VB 法により、双方の変分事後分布は以下の通りになる.

$$\log q_i(\theta_i) \propto \log \pi(\theta_i) + \sum_{j=1}^J \int_{b_j} q_j(b_j) \log p(y_{ij} | \theta_i, a_j, b_j) db_j \quad (28)$$

$$\log q_j(b_j) \propto \log \pi(b_j) + \sum_{i=1}^N \int_{\theta_i} q_i(\theta_i) \log p(y_{ij} | \theta_i, a_j, b_j) d\theta_i \quad (29)$$

ここで、 $\pi(\theta_i) \sim N(0, 1)$, $\pi(b_j) \sim N(0, 1)$ とする. また、尤度は GVEM と同様にシグモイド関数を近似した (18) 式を用いる. ここで、 $q_i(\theta_i)$, $q_i(b_j)$ は正規分布 $q_i(\theta_i) \sim N(\mu_{\theta_i}, \sigma_{\theta_i}^2)$, $q_j(b_j) \sim N(\mu_{b_j}, \sigma_{b_j}^2)$ に近似可能であるため、(28) 式と (29) 式は以下のように変形される.

$$\log q_i(\theta_i) \propto \sum_{j=1}^J \left(y_{ij} - \frac{1}{2} \right) a_j \theta_i - \sum_{j=1}^J \eta(\xi_{ij}) a_j^2 (\mu_{b_j}^2 + \sigma_{b_j}^2 - 2\mu_{b_j} \theta_i + \theta_i^2) - \frac{\theta_i^2}{2} \quad (30)$$

$$\log q_j(b_j) \propto \sum_{i=1}^N \left(\frac{1}{2} - y_{ij} \right) a_j b_i - \sum_{i=1}^N \eta(\xi_{ij}) a_j^2 (b_j^2 - 2b_j \mu_{\theta_i} + \mu_{\theta_i}^2 + \sigma_{\theta_i}^2) - \frac{b_j^2}{2} \quad (31)$$

(30), (31) 式を変形し、それぞれ正規分布に近似することで、以下の正規分布のパラメタの更新値を得る.

$$\mu_{\theta_i} = \sigma_{\theta_i}^2 \times \sum_{j=1}^J \left\{ 2\eta(\xi_{ij}) a_j \mu_{b_j} + y_{ij} - \frac{1}{2} \right\} a_j \quad (32)$$

$$\sigma_{\theta_i}^2 = \frac{1}{2 \sum_{j=1}^J \eta(\xi_{ij}) a_j^2 + 1} \quad (33)$$

$$\mu_{b_j} = \sigma_{b_j}^2 \times \sum_{i=1}^N \left\{ 2\eta(\xi_{ij}) a_j \mu_{\theta_i} - y_{ij} + \frac{1}{2} \right\} a_j \quad (34)$$

$$\sigma_{b_j}^2 = \frac{1}{2 \sum_{i=1}^N \eta(\xi_{ij}) a_j^2 + 1} \quad (35)$$

5.2 識別力パラメタの推定

識別力パラメタ a_j とパラメタ ξ_{ij} は, GVEM と同様に EM アルゴリズムにより推定される. 提案手法は, E ステップで θ_i と \mathbf{Y}_i の同時分布の対数の下界 ((19) 式) を求め, 変分事後分布 $q(\boldsymbol{\theta}, \mathbf{b})$ を用いて周辺化する. $E(\mathbf{a}, \boldsymbol{\xi})$ は以下の通りである.

$$\begin{aligned}
 E(\mathbf{a}, \boldsymbol{\xi}) &= \int l(\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\xi} | \mathbf{a}, \mathbf{b}) \times q(\boldsymbol{\theta}, \mathbf{b}) d\boldsymbol{\theta} d\mathbf{b} \\
 &= \sum_{i=1}^N \sum_{j=1}^J \int_{\theta_i} \int_{b_j} l(y_{ij}, \theta_i, \xi_{ij} | a_j, b_j) \times q_{b_j}(b_j) q_{\theta_i}(\theta_i) db_j d\theta_i \\
 &= \sum_{i=1}^N \sum_{j=1}^J \left(\log \frac{e^{\xi_{ij}}}{1 + e^{\xi_{ij}}} + \left(y_{ij} - \frac{1}{2} \right) \frac{a_j(\mu_{\theta_i} - \mu_{b_j})}{2} - \frac{\xi_{ij}}{2} \right) \\
 &\quad - \sum_{i=1}^N \sum_{j=1}^J \left(\eta(\xi_{ij}) \left((\mu_{b_j}^2 + \sigma_{b_j}^2 - 2\mu_{b_j}\mu_{\theta_i} + \mu_{\theta_i}^2 + \sigma_{\theta_i}^2) a_j^2 \right) - \xi_{ij}^2 \right) - \sum_{i=1}^N (\log 2\pi + 1) \quad (36)
 \end{aligned}$$

M ステップの $E(\mathbf{a}, \boldsymbol{\xi})$ の最大化により, 識別力パラメタとパラメタ ξ_{ij} の更新式は以下が得られる.

$$a_j = \frac{1}{2} \left[\sum_{i=1}^N \eta(\xi_{ij}) (\mu_{b_j}^2 + \sigma_{b_j}^2 - 2\mu_{b_j}\mu_{\theta_i} + \mu_{\theta_i}^2 + \sigma_{\theta_i}^2) \right]^{-1} \left(\sum_{i=1}^N \left(y_{ij} - \frac{1}{2} \right) (\mu_{\theta_i} - \mu_{b_j}) \right) \quad (37)$$

$$\xi_{ij}^2 = a_j^2 (\mu_{b_j}^2 + \sigma_{b_j}^2 - 2\mu_{b_j}\mu_{\theta_i} + \mu_{\theta_i}^2 + \sigma_{\theta_i}^2) \quad (38)$$

5.3 提案手法のアルゴリズム

2PLM のパラメタ推定のための, 提案手法のアルゴリズムは以下のようになる.

1. パラメタ $a_j^{(t)}, \xi_{ij}^{(t)}$ と困難度の変分事後分布のパラメタ $\mu_{b_j}^{(t)}, \sigma_{b_j}^{(t)}$ の初期値を設定. $t = 0$ とする.
2. E ステップ: (32), (33), (34), (35) 式を用いて $\mu_{\theta_i}^{(t+1)}, \sigma_{\theta_i}^{(t+1)}, \mu_{b_j}^{(t+1)}, \sigma_{b_j}^{(t+1)}$ を順に更新.
3. M ステップ: (37), (38) 式を用いて $a_j^{(t+1)}, \xi_{ij}^{(t+1)}$ を更新.
4. 収束条件を満たしていれば推定終了. そうでなければ, $t = t + 1$ としてステップ 2 に戻る.

提案手法は, GVEM と同様にアルゴリズムに数値最適化を含まず, 積分計算を含まない更新式による推定を可能にする. さらに, 正規分布に近似された能力潜在変数と困難度パラメタの変分事後分布により対数尤度の期待値は計算される. これにより, 識別力パラメタの推定において困難度パラメタの推定誤差の影響が軽減され, 推定精度を向上できる. 提案手法は GVEM の利点を保ちつつ識別力パラメタの推定精度を向上する.

6 評価実験

提案手法の有効性を示すために、シミュレーション実験により提案手法を従来手法と比較する。なお、本研究の実行環境は Windows11 を OS とする計算機 (CPU: 13th Gen Intel(R) Core(TM) i9-13900KF 3.00 GHz, RAM: 128GB) である。また、推定は python で実装したプログラムを用いて行う。

6.1 実験設定

データセットは以下の設定のもと生成したものを使用する。

- データサイズ
 1. 受検者数を 100, 1000, 10000, 100000 に固定し、項目数を 20, 40, ..., 200 まで 20 刻みに設定したデータ (データセット 1)
 2. 項目数を 10, 100, 1000 に固定し、受検者数を 100, 200, ..., 1000 まで 100 刻みに設定したデータ (データセット 2)
- パラメタの真値 (宮澤ら [31] の実験設定により生成)
 - $\theta_i \sim N(0, 1), \log a_j \sim N(-0.5, 0.2), b_j \sim N(0, 1)$
- 正答率の極端な項目 (正答率が 0.9 より大きい, または 0.1 未満) と素点との相関が低い項目 (0.3 未満) を除く。

データセット 1, 2 に対して各手法の推定を 10 回ずつ行い、各手法の推定精度、推定時間を測定した。推定精度の評価には、各パラメタごとに RMSE と Bias の平均値を利用した。例として、能力潜在変数における RMSE と Bias の平均値を求める計算式を以下に示す。 i 番目の受検者について、 θ_i が能力潜在変数の真値、 $\hat{\theta}_i^{(r)}$ が r 回目の能力潜在変数の推定値を表している。また、推定結果が事後分布として得られているものは、その分布の平均値を推定値として計算した。

$$\text{Avg.RMSE} = \frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{10} \sum_{r=1}^{10} (\theta_i - \hat{\theta}_i^{(r)})^2} \quad (39)$$

$$\text{Avg.Bias} = \frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{10} \sum_{r=1}^{10} (\theta_i - \hat{\theta}_i^{(r)})} \quad (40)$$

提案手法と GVEM の推定時には、EM アルゴリズムの反復回数も測定した。さらに、一回の推定において反復ごとの推定値の RMSE を計算した。

MCMC 法は、Wu らの実験と同様に NUTS(No-U-Turn Sampler) を使用し、繰り返し回数を 200, ウォームアップステップを 100, 並列処理を行わない設定で推定を行った [18]。また、計算コストの課題により MCMC 法による推定は、データセット 1 の受検者数を 100000 に固定した場合を除いて

行った.

提案手法と GVEM の収束条件は, Cho らが GVEM による推定に使用していた収束条件 [14] と同じ条件に設定した. 収束条件は以下の通りである. なお, $a^{(t)}, b^{(t)}$ は t 回目の反復による項目パラメタの推定値を表す.

$$M_p < 0.0001, M_p = \sqrt{(a^{(t+1)} - a^{(t)})^2 + (b^{(t+1)} - b^{(t)})^2} \quad (41)$$

また, GVEM と提案手法の最大反復回数は 10000 回とした.

6.2 実験結果

次の表 1, 2 と表 3, 4 は, それぞれデータセット 1, 2 について推定して得られた RMSE と Bias の平均値を示す. 表 1, 2, 3, 4 より, 受検者数が 1000 以下のデータに対して, 提案手法は多くの条件で識別力パラメタの RMSE と Bias を従来手法よりも大きく改善することができた. さらに, 受検者数が 100, 項目数が 20 と小規模なデータに対して, 提案手法は GVEM と比較して困難度パラメタの RMSE と Bias を向上できた. 提案手法における項目パラメタの RMSE と Bias は, 受検者数や項目数が多い場合でも多くの場合でほぼ同程度の値を示した. 提案手法における能力潜在変数の RMSE と Bias は, 多くの条件で従来手法とほぼ同程度であったが, 表 1 より項目数が 1000 になると従来手法よりも RMSE が大きくなった. これは, 提案手法による能力潜在変数の更新式で, 困難度パラメタの正規分布による期待値の総和を算出しているためであると考えられる ((32) 式). 期待値の総和を求める際に困難度パラメタの正規分布の近似により生じた誤差が蓄積されるため, 項目数の増加に伴い提案手法による能力潜在変数の推定精度が低下した可能性がある.

表 1: 各パラメタにおける RMSE の平均値 (データセット 1)

受検者数	項目数	識別力			困難度			能力値		
		MCMC	GVEM	propose	MCMC	GVEM	propose	MCMC	GVEM	propose
100	20	0.16	0.18	0.13	0.70	0.83	0.69	0.53	0.56	0.53
	40	0.19	0.25	0.15	0.32	0.31	0.32	0.36	0.35	0.35
	60	0.21	0.24	0.18	0.67	0.67	0.67	0.28	0.28	0.30
	80	0.22	0.25	0.19	0.60	0.60	0.60	0.28	0.30	0.28
	100	0.23	0.26	0.19	0.91	0.93	0.92	0.26	0.26	0.28
	120	0.19	0.24	0.14	0.41	0.41	0.40	0.25	0.26	0.25
	140	0.22	0.30	0.17	0.54	0.54	0.54	0.25	0.27	0.28
	160	0.21	0.29	0.16	0.30	0.32	0.29	0.20	0.24	0.22
	180	0.22	0.27	0.17	0.81	0.81	0.83	0.17	0.18	0.21
	200	0.19	0.25	0.14	0.41	0.40	0.40	0.21	0.19	0.26
1000	20	0.09	0.09	0.08	0.78	0.83	0.80	0.45	0.46	0.45
	40	0.12	0.10	0.10	0.57	0.57	0.57	0.36	0.36	0.36
	60	0.07	0.06	0.06	0.26	0.25	0.25	0.29	0.29	0.29
	80	0.08	0.08	0.07	0.48	0.48	0.48	0.25	0.25	0.25
	100	0.10	0.08	0.07	0.36	0.21	0.21	0.24	0.23	0.23
	120	0.08	0.08	0.08	0.46	0.46	0.46	0.20	0.20	0.20
	140	0.07	0.07	0.06	0.10	0.10	0.10	0.20	0.20	0.20
	160	0.09	0.09	0.08	0.54	0.54	0.54	0.18	0.18	0.19
	180	0.06	0.06	0.05	0.10	0.09	0.10	0.18	0.18	0.18
	200	0.09	0.09	0.08	0.52	0.52	0.53	0.16	0.16	0.16
10000	20	0.04	0.06	0.06	0.25	0.28	0.27	0.46	0.45	0.45
	40	0.12	0.11	0.11	0.66	0.67	0.67	0.33	0.33	0.33
	60	0.05	0.05	0.05	0.45	0.46	0.46	0.29	0.29	0.29
	80	0.04	0.04	0.04	0.28	0.28	0.28	0.25	0.25	0.25
	100	0.07	0.06	0.06	0.54	0.55	0.55	0.23	0.22	0.22
	120	0.03	0.03	0.03	0.06	0.06	0.06	0.21	0.21	0.21
	140	0.07	0.07	0.07	0.54	0.54	0.54	0.19	0.19	0.19
	160	0.05	0.05	0.05	0.44	0.44	0.44	0.18	0.18	0.18
	180	0.05	0.05	0.05	0.27	0.28	0.28	0.17	0.17	0.17
	200	0.08	0.08	0.08	0.65	0.65	0.65	0.16	0.16	0.16
100000	20		0.04	0.04		0.03	0.03		0.43	0.43
	40		0.03	0.03		0.06	0.06		0.34	0.34
	60		0.06	0.06		0.52	0.52		0.28	0.28
	80		0.08	0.08		0.53	0.54		0.25	0.25
	100		0.06	0.06		0.54	0.54		0.22	0.22
	120		0.07	0.07		0.40	0.40		0.20	0.20
	140		0.03	0.03		0.30	0.30		0.19	0.19
	160		0.01	0.01		0.02	0.02		0.18	0.18
	180		0.07	0.07		0.64	0.64		0.17	0.17
	200		0.05	0.05		0.50	0.50		0.16	0.16

表 2: 各パラメタにおける Bias の平均値 (データセット 1)

受検者数	項目数	識別力			困難度			能力値		
		MCMC	GVEM	propose	MCMC	GVEM	propose	MCMC	GVEM	propose
100	20	-0.12	-0.15	-0.06	0.23	0.44	0.23	0.07	0.23	0.07
	40	-0.17	-0.24	-0.12	0.17	0.13	0.17	0.06	0.01	0.06
	60	-0.15	-0.20	-0.09	-0.04	0.02	-0.04	-0.07	-0.01	-0.07
	80	-0.13	-0.21	-0.06	-0.05	-0.12	-0.05	-0.07	-0.16	-0.07
	100	-0.17	-0.23	-0.10	0.06	0.11	0.06	-0.03	0.03	-0.03
	120	-0.16	-0.23	-0.06	-0.03	-0.04	-0.02	-0.06	-0.08	-0.05
	140	-0.17	-0.28	-0.08	0.03	-0.02	0.03	0.04	-0.02	0.04
	160	-0.17	-0.28	-0.08	0.07	0.11	0.07	0.08	0.13	0.08
	180	-0.16	-0.23	-0.06	-0.03	-0.08	-0.03	0.00	-0.06	0.01
	200	-0.15	-0.23	-0.05	0.10	0.05	0.09	0.09	0.04	0.09
1000	20	-0.03	-0.01	0.01	-0.09	-0.14	-0.08	-0.01	-0.06	-0.01
	40	-0.06	-0.03	-0.03	0.05	0.05	0.05	0.06	0.05	0.06
	60	-0.05	-0.04	-0.03	-0.01	0.00	-0.01	-0.02	-0.02	-0.02
	80	-0.04	-0.03	-0.01	-0.04	-0.03	-0.04	-0.03	-0.02	-0.03
	100	-0.05	-0.05	-0.03	-0.09	-0.03	-0.03	-0.11	-0.02	-0.02
	120	-0.03	-0.03	-0.01	-0.04	-0.03	-0.04	0.00	0.02	0.00
	140	-0.04	-0.04	-0.02	-0.04	-0.05	-0.04	-0.05	-0.07	-0.06
	160	-0.03	-0.03	-0.01	-0.01	-0.02	-0.01	-0.02	-0.03	-0.02
	180	-0.05	-0.05	-0.02	0.07	0.06	0.07	0.07	0.06	0.07
	200	-0.04	-0.04	-0.01	0.01	0.02	0.01	0.01	0.02	0.01
10000	20	0.00	0.04	0.04	0.10	0.10	0.10	0.01	0.01	0.01
	40	-0.02	0.00	0.00	0.04	0.04	0.04	-0.01	-0.01	-0.01
	60	0.00	0.01	0.02	-0.03	-0.03	-0.03	0.01	0.01	0.01
	80	-0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01
	100	-0.02	-0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
	120	-0.02	-0.01	-0.01	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01
	140	-0.01	-0.01	0.00	0.00	0.00	0.00	-0.02	-0.02	-0.02
	160	-0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01
	180	0.00	0.00	0.01	-0.01	-0.01	-0.01	0.00	0.00	0.00
	200	0.00	0.00	0.01	0.00	0.00	0.00	0.00	-0.01	0.00
100000	20		0.04	0.04		-0.01	-0.01		0.00	0.00
	40		0.02	0.02		-0.03	-0.02		0.00	0.00
	60		0.01	0.01		-0.02	-0.02		0.00	0.00
	80		0.02	0.02		-0.01	-0.01		0.00	0.00
	100		0.01	0.01		-0.01	-0.01		0.01	0.00
	120		0.01	0.01		0.04	0.04		0.00	0.00
	140		0.01	0.01		0.00	0.00		0.00	0.00
	160		0.00	0.01		0.00	0.00		0.00	0.00
	180		0.00	0.00		-0.02	-0.02		0.00	0.00
	200		0.05	0.01		-0.02	-0.02		0.02	0.00

表 3: 各パラメタにおける RMSE の平均値 (データセット 2)

受検者数	項目数	識別力			困難度			能力値		
		MCMC	GVEM	propose	MCMC	GVEM	propose	MCMC	GVEM	propose
100	10	0.26	0.23	0.22	0.81	0.81	0.81	0.60	0.68	0.60
200		0.18	0.15	0.14	0.43	0.45	0.43	0.63	0.59	0.62
300		0.13	0.13	0.08	0.63	0.61	0.63	0.57	0.52	0.57
400		0.29	0.19	0.21	0.95	0.87	0.95	0.59	0.60	0.58
500		0.16	0.18	0.15	1.01	1.13	1.04	0.57	0.57	0.56
600		0.19	0.18	0.13	0.56	0.88	0.53	0.63	0.66	0.62
700		0.11	0.11	0.11	0.20	0.20	0.19	0.56	0.56	0.56
800		0.1	0.15	0.13	0.52	0.73	0.52	0.61	0.63	0.60
900		0.1	0.12	0.11	0.39	0.48	0.39	0.61	0.61	0.61
1000		0.14	0.14	0.14	0.73	0.98	0.75	0.55	0.56	0.55
100	100	0.23	0.26	0.19	0.91	0.93	0.92	0.26	0.26	0.28
200		0.14	0.16	0.11	0.19	0.18	0.18	0.25	0.25	0.26
300		0.15	0.15	0.13	0.37	0.37	0.37	0.24	0.24	0.23
400		0.11	0.11	0.10	0.33	0.33	0.33	0.23	0.23	0.24
500		0.10	0.10	0.08	0.14	0.14	0.13	0.24	0.24	0.24
600		0.11	0.12	0.10	0.36	0.36	0.36	0.23	0.23	0.23
700		0.11	0.11	0.09	0.50	0.50	0.50	0.22	0.22	0.22
800		0.09	0.09	0.08	0.33	0.33	0.33	0.24	0.23	0.24
900		0.08	0.08	0.07	0.22	0.22	0.22	0.24	0.24	0.23
1000		0.10	0.08	0.07	0.36	0.21	0.21	0.24	0.23	0.23
100	1000	0.18	0.26	0.14	0.68	0.68	0.73	0.08	0.11	0.31
200		0.15	0.18	0.13	0.69	0.69	0.74	0.08	0.09	0.25
300		0.15	0.18	0.11	0.63	0.63	0.67	0.10	0.12	0.17
400		0.10	0.12	0.09	0.15	0.15	0.15	0.08	0.08	0.19
500		0.12	0.14	0.11	0.67	0.66	0.72	0.13	0.10	0.20
600		0.10	0.12	0.09	0.46	0.46	0.48	0.08	0.10	0.15
700		0.08	0.09	0.08	0.15	0.16	0.16	0.08	0.08	0.15
800		0.10	0.11	0.10	0.65	0.65	0.68	0.08	0.09	0.13
900		0.07	0.08	0.07	0.10	0.10	0.10	0.08	0.08	0.13
1000		0.07	0.08	0.06	0.10	0.09	0.10	0.08	0.08	0.12

表 4: 各パラメタにおける Bias の平均値 (データセット 2)

受検者数	項目数	識別力			困難度			能力値		
		MCMC	GVEM	propose	MCMC	GVEM	propose	MCMC	GVEM	propose
100	10	-0.10	-0.06	0.01	-0.51	-0.02	-0.56	-0.04	-0.21	-0.04
200		-0.06	0.04	0.03	0.16	-0.34	0.16	-0.04	-0.16	-0.04
300		-0.07	0.01	0.01	-0.37	-0.21	-0.36	0.02	0.01	0.02
400		-0.10	-0.01	-0.02	0.34	-0.34	0.36	0.01	0.05	0.01
500		-0.06	-0.01	0.01	0.04	0.21	0.07	0.02	0.12	0.02
600		-0.04	-0.04	0.04	0.08	-0.20	0.05	-0.02	-0.16	-0.02
700		0.03	0.07	0.09	-0.13	-0.12	-0.12	-0.04	-0.04	-0.04
800		0.05	0.05	0.10	0.14	0.25	0.13	0.01	0.13	0.01
900		0.05	0.07	0.09	0.17	0.18	0.19	0.01	0.01	0.01
1000		-0.01	-0.02	0.05	-0.02	0.04	-0.02	0.02	0.06	0.02
100	100	-0.17	-0.23	-0.10	0.06	0.11	0.06	-0.03	0.03	-0.03
200		-0.11	-0.14	-0.05	-0.02	0.03	-0.02	-0.03	0.02	-0.03
300		-0.09	-0.10	-0.04	0.00	0.03	0.01	0.02	0.05	0.02
400		-0.06	-0.07	-0.02	0.08	0.09	0.08	0.04	0.04	0.04
500		-0.07	-0.07	-0.03	0.04	0.03	0.04	0.02	0.02	0.02
600		-0.07	-0.08	-0.04	0.03	0.04	0.03	0.03	0.03	0.02
700		-0.05	-0.06	-0.02	0.03	0.02	0.03	0.05	0.04	0.05
800		-0.06	-0.06	-0.04	-0.03	-0.02	-0.03	-0.01	-0.01	-0.02
900		-0.05	-0.05	-0.02	-0.02	0.00	-0.02	0.00	0.02	0.00
1000		-0.05	-0.05	-0.03	-0.09	-0.03	-0.03	-0.11	-0.02	-0.02
100	1000	-0.13	-0.24	0.04	0.00	-0.07	-0.01	0.00	-0.08	0.02
200		-0.09	-0.15	0.06	0.01	0.06	0.02	0.01	0.06	0.00
300		-0.10	-0.16	0.03	0.04	0.04	0.04	0.05	0.05	0.05
400		-0.07	-0.11	0.05	-0.02	0.00	-0.02	-0.03	-0.01	-0.03
500		-0.06	-0.10	0.05	-0.12	-0.08	-0.12	-0.11	-0.07	-0.13
600		-0.06	-0.09	0.05	0.01	-0.07	0.00	0.01	-0.07	0.02
700		-0.04	-0.07	0.06	0.01	0.03	0.01	0.01	0.03	0.01
800		-0.04	-0.06	0.05	-0.02	-0.05	-0.02	-0.02	-0.06	-0.02
900		-0.04	-0.06	0.05	-0.02	-0.03	-0.02	-0.02	-0.03	-0.02
1000		0.04	-0.06	0.04	0.05	0.00	0.04	0.04	-0.01	0.05

次に, GVEM と提案手法の推定時間と反復回数の比較を行う. 以下の表 5, 6 は, データセット 1, 2 の推定に要した時間と反復回数の平均値, 1 反復あたりに要した平均推定時間を示している.

表 5, 6 より, 提案手法は全てのデータセットにおいて従来手法よりも推定時間を著しく短縮できた. さらに, 提案手法と GVEM の反復回数の結果を比較すると, 同じ収束条件にもかかわらず, 反復ごとの平均時間はほぼ同程度で平均反復回数を大幅に削減することができた. 特に, 項目数が小さいデータセットに対しては, GVEM による推定は収束せず最大反復回数かそれに近い回数の反復を行っているが, 提案手法では少ない反復回数で推定できていることが確認できる.

以下の図 1, 2 は, それぞれ受検者数が 100, 項目数が 20 のデータ (小規模データ) と受検者数が 100000, 項目数が 200 のデータ (大規模データ) を推定した結果を示す. 図 1, 2 における 3 つの図は, 横軸を反復回数, 縦軸を反復ごとの推定値により求めた RMSE とし, 項目パラメタの実験結果を図示したものである. なお, 図示する範囲は提案手法の平均反復回数までとした.

図 1 より, 受検者数や項目数が少ない場合でも提案手法は反復回数を増すごとに RMSE が収束していることを確認できる. 一方, GVEM による推定ではどのパラメタ (表 5) も推定値が収束せず, 困難度パラメタが特に大きく発散している様子が確認できる. また, 図 2 からは提案手法における各パラメタの RMSE が GVEM よりも少ない反復回数で収束していることが確認できる. そのため, 提案手法は GVEM よりも少ない反復回数でパラメタを推定できることが分かった.

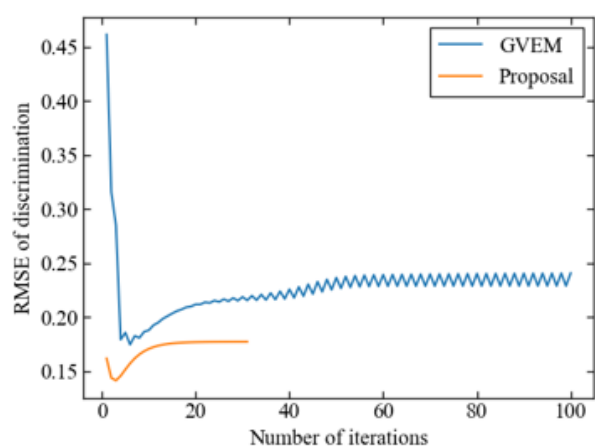
提案手法が GVEM よりも収束までにかかる反復回数が少ない原因は, 各パラメタの更新式に他のパラメタの更新値を含むためであると考えられる. 識別力パラメタの推定精度向上により, 全てのパラメタの更新において更新値がより収束値に速く近づくため, 提案手法は GVEM よりも反復回数を削減できたと考えられる.

表 5: 各手法の推定時間と反復回数 (データセット 1)

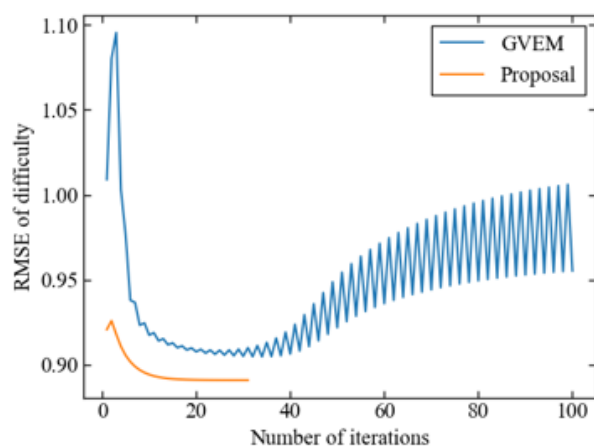
受検者数	項目数	平均推定時間			平均反復回数		反復ごとの推定時間	
		MCMC	GVEM	propose	GVEM	propose	GVEM	propose
100	20	9.79	89.88	0.17	10000.00	28.50	0.01	0.01
	40	16.51	5.56	0.41	219.61	35.90	0.03	0.01
	60	22.57	7.44	0.79	462.60	47.50	0.02	0.02
	80	29.14	8.30	1.06	388.30	48.50	0.02	0.02
	100	36.05	11.40	1.54	427.80	56.40	0.03	0.03
	120	43.61	12.50	1.62	391.00	48.80	0.03	0.03
	140	52.22	19.83	2.38	534.30	61.30	0.04	0.04
	160	58.25	25.38	2.93	597.20	66.50	0.04	0.04
	180	70.36	30.48	3.27	635.90	66.20	0.05	0.05
	200	87.23	36.72	3.77	692.00	69.20	0.05	0.05
1000	20	163.30	630.91	1.90	10000.00	33.80	0.06	0.06
	40	321.33	34.46	5.48	272.30	49.70	0.13	0.11
	60	445.36	85.51	10.64	451.20	65.40	0.19	0.16
	80	550.52	88.99	14.02	351.80	64.10	0.25	0.22
	100	699.52	101.04	20.51	378.00	75.90	0.27	0.27
	120	799.22	166.76	29.32	445.10	89.90	0.37	0.33
	140	966.45	241.82	41.61	538.50	108.00	0.45	0.39
	160	1029.28	262.08	46.30	511.70	104.60	0.51	0.44
	180	1195.38	343.82	58.79	594.90	119.30	0.58	0.49
	200	1249.21	372.03	67.64	583.90	122.90	0.64	0.55
10000	20	9225.03	1446.95	19.14	2251.30	33.40	0.64	0.57
	40	19267.56	369.23	53.48	286.30	47.90	1.29	1.12
	60	23294.31	703.84	106.38	364.50	63.70	1.93	1.67
	80	19413.54	839.30	155.12	324.50	70.00	2.59	2.22
	100	24213.71	1216.56	225.82	376.20	81.50	3.23	2.77
	120	29566.70	1763.88	335.95	461.90	101.50	3.82	3.31
	140	32234.00	1476.13	368.42	339.10	95.40	4.35	3.86
	160	38880.77	2162.40	511.65	454.70	115.70	4.76	4.42
	180	42635.93	3052.19	640.31	529.20	128.50	5.77	4.98
	200	41905.97	2987.95	736.74	532.20	133.60	5.61	5.51
100000	20		1056.08	209.53	187.00	36.70	5.65	5.71
	40		3415.12	582.57	305.60	51.70	11.18	11.27
	60		7779.79	1344.01	454.90	80.70	17.10	16.65
	80		6459.60	1498.59	280.30	67.10	23.05	22.33
	100		9804.73	2247.51	346.20	80.90	28.32	27.78
	120		14460.48	3341.23	426.70	99.10	33.89	33.72
	140		20288.98	5462.47	533.60	122.50	38.02	44.59
	160		22191.01	6287.99	511.20	120.60	43.41	52.14
	180		15696.87	7834.84	266.00	139.40	59.01	56.20
	200		14445.40	8023.51	267.4	141.10	54.02	56.86

表 6: 各手法の推定時間と反復回数 (データセット 2)

受検者数	項目数	平均推定時間			平均反復回数		反復ごとの推定時間	
		MCMC	GVEM	propose	GVEM	propose	GVEM	propose
100	10	6.35	63.85	0.10	10000.00	34.20	0.01	0.00
200		11.62	55.40	0.17	10000.00	28.50	0.01	0.01
300		23.96	8.15	0.26	978.30	29.80	0.01	0.01
400		31.97	4.13	0.33	373.10	29.00	0.01	0.01
500		45.18	137.01	0.48	10000.00	33.80	0.01	0.01
600		46.19	164.02	0.57	10000.00	33.10	0.02	0.02
700		57.94	3.95	0.52	203.70	26.10	0.02	0.02
800		67.14	217.38	0.78	10000.00	34.70	0.02	0.02
900		97.49	244.70	0.76	10000.00	29.80	0.02	0.03
1000		100.22	275.54	0.95	10000.00	33.40	0.03	0.03
100	100	36.05	11.40	1.54	427.80	56.40	0.03	0.03
200		73.33	19.23	3.19	362.10	58.80	0.05	0.05
300		130.58	35.79	5.85	450.50	72.10	0.08	0.08
400		194.27	49.31	7.65	465.00	70.30	0.11	0.11
500		304.76	44.52	9.32	335.40	68.60	0.13	0.14
600		360.21	43.98	10.53	275.80	64.90	0.16	0.16
700		445.12	57.54	13.03	308.00	69.20	0.19	0.19
800		518.10	75.19	15.93	350.90	73.40	0.21	0.22
900		620.37	72.88	16.27	301.20	66.80	0.24	0.24
1000		699.52	101.04	20.51	378.00	75.90	0.27	0.27
100	1000	657.07	307.79	31.13	1017.40	100.90	0.30	0.31
200		1432.36	650.29	75.13	1061.30	120.30	0.61	0.62
300		1925.19	844.84	139.52	922.50	149.50	0.92	0.93
400		2751.92	1289.49	205.74	1050.80	165.20	1.23	1.25
500		3198.28	1579.45	332.16	1030.90	213.00	1.53	1.56
600		3971.08	1893.71	463.84	1027.90	249.30	1.84	1.86
700		4245.52	2158.87	570.24	1007.70	262.40	2.14	2.17
800		5213.04	2538.02	741.28	1035.30	294.40	2.45	2.52
900		5332.49	2902.69	889.20	1046.10	314.60	2.77	2.83
1000		6436.25	3080.67	1089.98	989.00	339.40	3.11	3.21

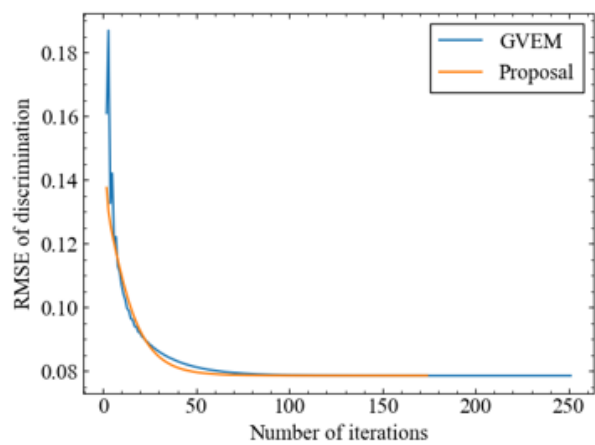


(a) 識別力

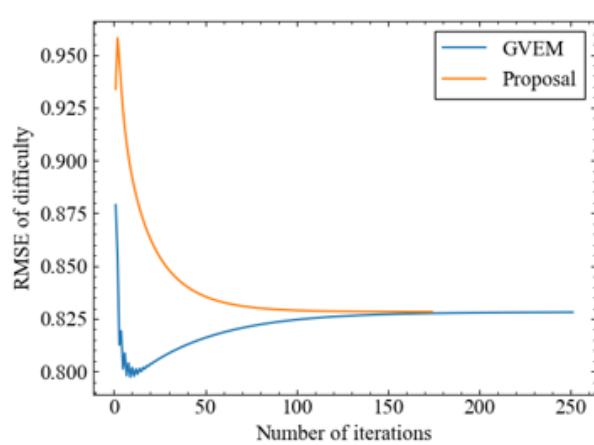


(b) 困難度

図 1: 小規模データセットにおける反復回数と項目パラメタの RMSE の関係



(a) 識別力



(b) 困難度

図 2: 大規模データセットにおける反復回数と項目パラメタの RMSE の関係

7 むすび

本論では、項目反応理論のパラメタ推定の課題を解決するため、能力潜在変数と困難度パラメタの事後分布を VB 法により導出し、正規分布に近似してパラメタ推定する手法を提案した。提案手法は能力潜在変数と困難度パラメタをベイズ推定することで、解答データの受検者数や項目数が少ない場合でも安定した推定を可能にした。また、事後分布を正規分布に近似することで、識別力パラメタの推定に必要な期待値計算を短時間で行いつつ、周辺化により困難度パラメタの推定誤差の影響を軽減させた。シミュレーションデータを用いた比較実験により、提案手法は、多くの条件で従来手法とほぼ同程度のパラメタ推定精度を保ちつつ、解答データの受検者数が少ない場合に識別力の RMSE と Bias を従来手法よりも向上させることができた。さらに、提案手法は従来手法よりも計算時間を削減し、Chorらの GVEM[14] よりも少ない反復による推定を可能にした。

提案手法の課題は 2 つ挙げられる。1 つ目は、項目数の多い解答データに対する能力潜在変数の推定精度が低いことである。これは、項目数が多いと能力潜在変数の更新式に含まれる困難度パラメタの期待値の和を計算する際に、正規分布の近似により生じた誤差が蓄積されるため、能力潜在変数の更新式に誤差が生じて推定精度が低下すると考えられる。したがって、困難度パラメタの正規分布をより正確に近似できれば、項目数が多いデータに対しても能力潜在変数の推定精度を向上できる可能性がある。2 つ目は、識別力パラメタをベイズ推定することである。識別力パラメタを期待値計算可能な単純な関数形に正確に近似することで、提案手法のパラメタ推定精度を保ちつつ、受検者数や項目数の少ない解答データをより安定して推定できる可能性がある。

参考文献

- [1] Maomi Ueno, Kazuma Fuchimoto, and Emiko Tsutsumi. e-testing from artificial intelligence approach. *Behaviormetrika*, 48(2):409–424, July 2021.
- [2] Maomi Ueno. AI based e-testing as a common yardstick for measuring human abilities. In *2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 1–5. IEEE, June 2021.
- [3] 公益社団法人医療系大学間共用試験実施評価機構 (2023). 共用試験の概要. <https://www.cato.or.jp/cbt/establish/index.html>.
- [4] 独立行政法人情報処理推進機構 (2023). 【it パスポート試験】情報処理推進機構. <https://www3.jitec.ipa.go.jp/JitesCbt/index.html>.
- [5] Frederic M Lord and Melvin R Novick. *Statistical theories of mental test scores*. IAP, 2008.
- [6] Frank B Baker and Seock-Ho Kim. *Item response theory: Parameter estimation techniques*. CRC press, 2004.
- [7] Dirk P Kroese and Reuven Y Rubinstein. Monte carlo methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(1):48–58, 2012.
- [8] Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer New York, 2006.
- [9] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *J. Am. Stat. Assoc.*, 112(518):859–877, April 2017.
- [10] Shinji Watanabe, Yasuhiro Minami, Atsushi Nakamura, and Naonori Ueda. Variational bayesian estimation and clustering for speech recognition. *IEEE Transactions on Speech and Audio Processing*, 12(4):365–381, 2004.
- [11] David M Blei and Michael I Jordan. Variational methods for the dirichlet process. In *Proceedings of the twenty-first international conference on Machine learning, ICML '04*, page 12, New York, NY, USA, July 2004. Association for Computing Machinery.
- [12] J T Ormerod and M P Wand. Gaussian variational approximate inference for generalized linear mixed models. *J. Comput. Graph. Stat.*, 21(1):2–17, January 2012.
- [13] Minjeong Jeon, Frank Rijmen, and Sophia Rabe-Hesketh. A variational Maximization–Maximization algorithm for generalized linear mixed models with crossed random effects. *Psychometrika*, 82(3):693–716, September 2017.
- [14] April E Cho, Chun Wang, Xue Zhang, and Gongjun Xu. Gaussian variational estimation for multidimensional item response theory. *Br. J. Math. Stat. Psychol.*, 74 Suppl 1:52–85, July 2021.
- [15] Prathiba Natesan, Ratna Nandakumar, Tom Minka, and Jonathan D Rubright. Bayesian prior choice in IRT estimation using MCMC and variational bayes. *Front. Psychol.*, 7:1422,

September 2016.

- [16] Frank Rijmen and Minjeong Jeon. Fitting an item response theory model with random item effects across groups by a variational approximation method. *Ann. Oper. Res.*, 206(1):647–662, July 2013.
- [17] Arto Klami. Poly-gamma augmentations for factor models. In Dinh Phung and Hang Li, editors, *Proceedings of the Sixth Asian Conference on Machine Learning*, volume 39 of *Proceedings of Machine Learning Research*, pages 112–128, Nha Trang City, Vietnam, 2015. PMLR.
- [18] Mike Wu, Richard L Davis, Benjamin W Domingue, Chris Piech, and Noah Goodman. Variational item response theory: Fast, accurate, and expressive. February 2020.
- [19] R Darrell Bock and Murray Aitkin. Marginal maximum likelihood estimation of item parameters: Application of an em algorithm. *Psychometrika*, 46(4):443–459, 1981.
- [20] Yunxiao Chen, Xiaoou Li, and Siliang Zhang. Joint maximum likelihood estimation for High-Dimensional exploratory item factor analysis. *Psychometrika*, 84(1):124–146, March 2019.
- [21] A metropolis -hastings robbins -monro algorithm for maximum likelihood nonlinear latent structure analysis with a comprehensive measurement model, 2008.
- [22] Frederic M Lord. Maximum likelihood and bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, 23(2):157–162, 1986.
- [23] Frank B Baker and Seock-Ho Kim. *Item Response Theory: Parameter Estimation Techniques, Second Edition*. CRC Press, July 2004.
- [24] 豊田秀樹. 項目反応理論 「理論編」. 朝倉書店, 2005.
- [25] 山口 一大. 項目反応理論モデルのパラメタ推定法の展開. *日本テスト学会誌*, 18(1):103–131, 2022.
- [26] 豊田秀樹. 基礎からのベイズ統計学 ハミルトニアンモンテカルロ法による実践的入門. 朝倉書店, 2015.
- [27] Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.
- [28] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.
- [29] Matthew D Hoffman, Andrew Gelman, et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.
- [30] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37:183–233, 1999.
- [31] 宮澤芳光 and 植野真臣. 高精度能力推定を保証する 2 段階等質適応型テスト. *電子情報通信学会論文誌 D*, 106(1):34–46, 2023.