

博士論文

項目反応間の局所依存性が受験者、項目および
テストの諸特性の推定に与える影響の検討

登藤直弥

項目反応間の局所依存性が受験者, 項目および
テストの諸特性の推定に与える影響の検討

登藤直弥

目次

第 1 章	序論	9
1.1	項目反応理論	9
1.2	項目特性関数と個人正答確率	10
1.3	局所独立性	11
1.4	局所依存性	14
1.5	局所依存性を考慮した項目反応モデル	15
1.6	本研究の目的	19
第 2 章	局所独立性を仮定した受験者特性の推定に与える影響	21
2.1	問題と目的	21
2.2	方法	23
2.2.1	項目反応データの発生モデル	23
2.2.2	シミュレーションの手続き	24
2.2.3	母数の推定法	27
2.3	結果と考察	30
2.3.1	発生モデルが 2PLM の場合	31
2.3.2	発生モデルが CCM の場合	31
2.3.3	発生モデルが 2PLCM の場合	33
2.3.4	先行研究との比較	34
2.3.5	局所依存性の発生プロセスと局所独立性を仮定したモデルの頑健性	35
第 3 章	局所独立性を仮定した項目特性の推定に与える影響	51
3.1	問題と目的	51
3.2	方法	53
3.2.1	仮想的な項目反応データの発生モデル	53
3.2.2	仮想的な項目反応データの分析モデル	54
3.2.3	2 値型の BTM に含まれる項目母数の変換方法	54

3.2.4	シミュレーション実験の手続き	55
3.2.5	MCMC による母数の推定	57
3.3	結果と考察	59
3.3.1	識別力の推定に対する局所依存性の影響	59
3.3.2	困難度の推定精度に関する結果	63
3.3.3	局所依存性が局所独立性を仮定した項目母数の推定に影響を与えるメカニズム	65
3.3.4	先行研究との関連	66
第 4 章	局所独立性を仮定したテスト情報量の推定に与える影響	96
4.1	問題と目的	96
4.2	方法	97
4.2.1	データの発生モデル	98
4.2.2	データの分析モデル	98
4.2.3	テスト情報量	99
4.2.4	シミュレーション実験の手続き	100
4.2.5	MCMC による母数の推定	102
4.3	結果と考察	103
4.3.1	真のテスト情報量の値	104
4.3.2	分析モデルが 2 値型の BTM の場合	104
4.3.3	分析モデルが 2PLM の場合	104
4.3.4	先行研究との関連	106
第 5 章	局所依存モデル間での受験者特性の推定精度の比較	118
5.1	問題と目的	118
5.2	方法	120
5.2.1	データの発生モデル	120
5.2.2	データの分析モデル	120
5.2.3	シミュレーションの手続き	121
5.2.4	母数の推定法	123
5.3	結果と考察	125
5.3.1	分析モデルが 2PLM の場合	125
5.3.2	分析モデルが GRM の場合	125
5.3.3	分析モデルが 2 値型の BTM の場合	127
5.3.4	分析モデルが CCM の場合	129

5.3.5	CCM の \overline{Bias} と局所依存度の強い項目群の数との関係	131
5.3.6	先行研究との関連	132
5.3.7	局所依存性を考慮するモデルのタイプと推定精度の改善度合いとの関係	132
第 6 章	総合考察	165
6.1	本論文で明らかになったこと	165
6.2	本論文の限界点と今後の課題	166

表目次

1.1	N 人の受験者に対し J 項目のテストを実施した結果	9
2.1	CCM を発生モデルとした場合に各項目に対して $(1, 1), (1, 0), (0, 1), (0, 0)$ と反応する受験者の割合 (100 人の受験者が 100 回受験した場合の平均値)	25
2.2	2PLCM を発生モデルとした場合に各項目に対して $(1, 1), (1, 0), (0, 1), (0, 0)$ と反応する受験者の割合 (100 人の受験者が 100 回受験者した場合の平均値)	25
2.3	発生モデルが 2PLM の場合の $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	36
2.4	発生モデルが CCM の場合の $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	37
2.5	発生モデルが CCM の場合の $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	37
2.6	発生モデルが 2PLCM の場合の $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	38
2.7	発生モデルが 2PLCM の場合の $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	38
3.1	分析モデルが 2 値型の BTM である時の識別力に関する $\overline{Bias}, \overline{RMSE}, cor(\mathbf{a}, \hat{\mathbf{a}})$ の値	68
3.2	分析モデルが 2PLM である時の識別力に関する $\overline{Bias}, \overline{RMSE}, cor(\mathbf{a}, \hat{\mathbf{a}})$ の値	69
3.3	識別力に関する $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値	70
3.4	識別力に関する $\Delta\overline{Bias}/\sigma_{a_j^*}$ と $\Delta\overline{RMSE}/\sigma_{a_j^*}$ の値	71
3.5	項目母数の比較可能性を確保しなかった場合の識別力に関する $\overline{Bias}, \overline{RMSE}, cor(\mathbf{a}, \hat{\mathbf{a}})$ の値	72
3.6	項目母数の比較可能性を確保しなかった場合の識別力に関する $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値	73
3.7	項目母数の比較可能性を確保しなかった場合の $\Delta\overline{Bias}/\sigma_{a_j^*}$ および $\Delta\overline{RMSE}/\sigma_{a_j^*}$ の値	74
3.8	分析モデルが 2 値型の BTM である時の困難度に関する $\overline{Bias}, \overline{RMSE}, cor(\mathbf{b}, \hat{\mathbf{b}})$ の値	75

3.9	分析モデルが 2PLM である時の困難度に関する $\overline{Bias}, \overline{RMSE}, cor(\mathbf{b}, \hat{\mathbf{b}})$ の値	76
3.10	困難度に関する $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\mathbf{b}, \hat{\mathbf{b}})$ の値	77
4.1	\bar{I} の値	107
4.2	分析モデルが 2 値型の BTM である場合の $\overline{Bias}, \overline{RMSE}$ の値	108
4.3	分析モデルが 2PLM である場合の $\overline{Bias}, \overline{RMSE}$ の値	109
4.4	分析モデルが 2PLM である場合の $\Delta\overline{Bias}, \Delta\overline{RMSE}$ の値	110
4.5	分析モデルが 2PLM である場合の $\Delta\overline{Bias}/\bar{I}, \Delta\overline{RMSE}/\bar{I}$ の値	111
5.1	分析モデルが 2PLM である時の $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	134
5.2	分析モデルが GRM である時の $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	135
5.3	分析モデルが GRM である時の $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	136
5.4	分析モデルが BTM である時の $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	137
5.5	分析モデルが BTM である時の $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	138
5.6	分析モデルが CCM である時の $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	139
5.7	分析モデルが CCM である時の $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値	139
5.8	$\sigma_{\gamma_{d(j)}}^2 = 1.4075$ とした BTM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 回分の平均)	140
5.9	$b_{jk} = -1.714$ とした CCM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 回分の平均)	140
5.10	$\sigma_{\gamma_{d(j)}}^2 = 0.155$ とした BTM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 回分の平均)	140
5.11	$b_{jk} = -0.612$ とした CCM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 回分の平均)	140

目次

2.1	発生モデルが CCM である場合の $\overline{\Delta Bias}$ の受験者数による変化	39
2.2	発生モデルが CCM である場合の $\overline{\Delta RMSE}$ の受験者数による変化	40
2.3	発生モデルが CCM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の受験者数による変化	41
2.4	発生モデルが CCM である場合の $\overline{\Delta Bias}$ の項目数による変化	42
2.5	発生モデルが CCM である場合の $\overline{\Delta RMSE}$ の項目数による変化	43
2.6	発生モデルが CCM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の項目数による変化	44
2.7	発生モデルが 2PLCM である場合の $\overline{\Delta Bias}$ の受験者数による変化	45
2.8	発生モデルが 2PLCM である場合の $\overline{\Delta RMSE}$ の受験者数による変化	46
2.9	発生モデルが 2PLCM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の受験者数による変化	47
2.10	発生モデルが 2PLCM である場合の $\overline{\Delta Bias}$ の項目数による変化	48
2.11	発生モデルが 2PLCM である場合の $\overline{\Delta RMSE}$ の項目数による変化	49
2.12	発生モデルが 2PLCM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の項目数による変化	50
3.1	$\overline{\Delta Bias}/\sigma_{a_j^*}$ と受験者数との関係	78
3.2	$\overline{\Delta RMSE}/\sigma_{a_j^*}$ と受験者数との関係	79
3.3	$\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ と受験者数との関係	80
3.4	$\overline{\Delta Bias}/\sigma_{a_j^*}$ と局所依存関係にある項目の数との関係	81
3.5	$\overline{\Delta RMSE}/\sigma_{a_j^*}$ と局所依存関係にある項目の数との関係	82
3.6	$\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ と局所依存関係にある項目の数との関係	83
3.7	$\overline{\Delta Bias}/\sigma_{a_j^*}$ と 4 つの項目群の局所依存度との関係	84
3.8	$\overline{\Delta RMSE}/\sigma_{a_j^*}$ と 4 つの項目群の局所依存度との関係	85
3.9	$\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ と 4 つの項目群の局所依存度との関係	86
3.10	困難度に関する $\overline{\Delta Bias}$ と受験者数との関係	87
3.11	困難度に関する $\overline{\Delta RMSE}$ と受験者数との関係	88
3.12	$\Delta cor(\mathbf{b}, \hat{\mathbf{b}})$ と受験者数との関係	89
3.13	困難度に関する $\overline{\Delta Bias}$ と局所依存関係にある項目の数との関係	90

3.14	困難度に関する $\overline{\Delta RMSE}$ と局所依存関係にある項目の数との関係	91
3.15	$\Delta cor(\mathbf{b}, \hat{\mathbf{b}})$ と局所依存関係にある項目の数との関係	92
3.16	困難度に関する $\overline{\Delta Bias}$ と4つの項目群の局所依存度との関係	93
3.17	困難度に関する $\overline{\Delta RMSE}$ と4つの項目群の局所依存度との関係	94
3.18	困難度に関する $\Delta cor(\mathbf{b}, \hat{\mathbf{b}})$ と4つの項目群の局所依存度との関係	95
4.1	$\overline{\Delta Bias}/\bar{I}$ と受験者数との関係	112
4.2	$\overline{\Delta RMSE}/\bar{I}$ と受験者数との関係	113
4.3	$\overline{\Delta Bias}/\bar{I}$ と局所依存関係にある項目の数との関係	114
4.4	$\overline{\Delta RMSE}/\bar{I}$ と局所依存関係にある項目の数との関係	115
4.5	$\overline{\Delta Bias}/\bar{I}$ と4つの項目群の局所依存度との関係	116
4.6	$\overline{\Delta RMSE}/\bar{I}$ と4つの項目群の局所依存度との関係	117
5.1	分析モデルが GRM である場合の $\overline{\Delta Bias}$ の受験者数による変化	141
5.2	分析モデルが GRM である場合の $\overline{\Delta RMSE}$ の受験者数による変化	142
5.3	分析モデルが GRM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の受験者数による変化	143
5.4	分析モデルが GRM である場合の $\overline{\Delta Bias}$ の局所依存関係にある項目の数による変化	144
5.5	分析モデルが GRM である場合の $\overline{\Delta RMSE}$ の局所依存関係にある項目の数による変化	145
5.6	分析モデルが GRM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の局所依存関係にある項目の数による変化	146
5.7	分析モデルが GRM である場合の $\overline{\Delta Bias}$ の4つの項目群の局所依存度による変化	147
5.8	分析モデルが GRM である場合の $\overline{\Delta RMSE}$ の4つの項目群の局所依存度による変化	148
5.9	分析モデルが GRM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の4つの項目群の局所依存度による変化	149
5.10	分析モデルが BTM である場合の $\overline{\Delta Bias}$ の受験者数による変化	150
5.11	分析モデルが BTM である場合の $\overline{\Delta RMSE}$ の受験者数による変化	151
5.12	分析モデルが BTM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の受験者数による変化	152
5.13	分析モデルが BTM である場合の $\overline{\Delta Bias}$ の局所依存関係にある項目の数による変化	153
5.14	分析モデルが BTM である場合の $\overline{\Delta RMSE}$ の局所依存関係にある項目の数による変化	154

5.15	分析モデルが BTM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の局所依存関係にある項目の数による変化	155
5.16	分析モデルが BTM である場合の $\Delta \overline{Bias}$ の 4 つの項目群の局所依存度による変化	156
5.17	分析モデルが BTM である場合の $\Delta \overline{RMSE}$ の 4 つの項目群の局所依存度による変化	157
5.18	分析モデルが BTM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の 4 つの項目群の局所依存度による変化	158
5.19	分析モデルが CCM である場合の $\Delta \overline{Bias}$ の受験者数による変化	159
5.20	分析モデルが CCM である場合の $\Delta \overline{RMSE}$ の受験者数による変化	160
5.21	分析モデルが CCM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の受験者数による変化	161
5.22	分析モデルが CCM である場合の $\Delta \overline{Bias}$ の 4 つの項目群の局所依存度による変化	162
5.23	分析モデルが CCM である場合の $\Delta \overline{RMSE}$ の 4 つの項目群の局所依存度による変化	163
5.24	分析モデルが CCM である場合の $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の 4 つの項目群の局所依存度による変化	164

第 1 章

序論

1.1 項目反応理論

ある能力を測定するために J 問から成るテスト (テスト A) を作成し, N 人の受験者に対してこのテストを実施した結果, 次のような結果が得られたとする (表 1.1).

表 1.1 N 人の受験者に対し J 項目のテストを実施した結果

	項目 1	項目 2	...	項目 J
受験者 1	1	1	...	0
受験者 2	0	1	...	1
⋮	⋮	⋮	⋮	⋮
受験者 N	1	1	...	1

なお, 表 1.1 中の 1 はその受験者が当該項目に正答していることを表しており, 0 は誤答していることを表している. 項目反応理論 (item response theory, IRT) では, このような項目反応データから各受験者の能力を評価する場合に, 正答数や配点で重みづけを行った和得点ではなく, 項目反応データそのものを用いて能力の評価が行われる.

仮にいま, その能力水準が既知である受験者集団が存在するものとし, この集団に対してテスト A を実施したとする. するとこのとき, 能力の水準毎にその正答率を算出すると, テスト A に含まれるどの項目においても, 能力の水準が高くなるほど各項目への正答率は高くなっていくと考えられる. 項目反応理論では, このような能力と各項目の正答率との関係を利用して, 各項目の能力水準毎の正答確率を項目反応モデルと呼ばれる関数で表現し, この項目反応モデルと項目反応データから, 項目反応モデルに含まれる項目の特性を表す母数と受験者の特性 (今の場合は能力) を表す母数を推定することになる.

項目反応理論を用いてテストの作成や運用を行うと、例えば、

- テスト情報関数と呼ばれる受験者特性の関数を用いて、受験者特性の推定精度をその水準毎に評価する
- 異なるテストを異なる集団に対して実施した場合に、受験者や項目の特性を共通の尺度上で評価する
- テスト特性関数と呼ばれる受験者特性の関数を利用して、テストを実施する前に、異なるテスト間でその得点を対応づける

といったことが可能となる。

このように、テストの作成、評価、運用に資する様々な利点を有するため、近年では多くのテストが項目反応理論に基づき作成され、運用されるようになってきている (e.g., Kan, van der Ven, Breteler, & Zitman, 2001; Simms, Goldberg, Roberts Watson, Welte, & Rotterman, 2011).

1.2 項目特性関数と個人正答確率

ある項目 j に対する正答確率と受験者特性 θ との関係を表す θ の関数を一般に項目 j の項目特性関数 (item characteristic function, ICF) と呼ぶ。1.1 節で言及した項目反応モデルとは、個々の項目に対し項目特性関数として設定される具体的な関数のことを表しており、項目反応モデルの代表的なものとしては、以下に示す 2 パラメタ・ロジスティックモデル (two-parameter logistic model, 2PLM) が挙げられる。

$$P_j(\theta_i) = \frac{1}{1 + \exp[-1.7a_j(\theta_i - b_j)]} \quad (1.1)$$

ここで、(1.1) 式中の $P_j(\theta_i)$ は θ の値が θ_i である受験者が項目 j に正答する確率を表しており、 a_j, b_j はそれぞれ項目 j の特性を表す母数となっている。 a_j は $\theta_i = b_j$ における (1.1) 式の傾きに比例する値となっており、この値が高いほど項目 j が受験者の特性を識別する力が強くなるため、一般に a_j は項目 j の識別力と呼ばれている。また、(1.1) 式において $\theta_i = b_j$ とすると $P_j(\theta_i) = 0.5$ となり、この値が高くなるほど項目 j に正答する確率が 0.5 となるのに必要な θ の値も高くなっていく。そのため、一般に b_j は項目 j の困難度と呼ばれている。

この他にも、(1.1) 式中の a_j が全ての項目間で等しいと仮定する 1 パラメタ・ロジスティックモデル (one-parameter logistic model, 1PLM) や (1.1) 式に新たな項目母数 c_j を加えて $P_j(\theta_i)$ を

$$P_j(\theta_i) = c_j + \frac{1 - c_j}{1 + \exp[-1.7a_j(\theta_i - b_j)]} \quad (1.2)$$

と表現する3パラメタ・ロジスティックモデル (three-parameter logistic model, 3PLM) などが、代表的な項目反応モデルとして挙げられる。なお、(1.2) 式中の θ_i を $\theta_i = -\infty$ であるとする、 $P_j(\theta_i) = c_j$ となる。このことから、 c_j は本来項目 j に正答することができない受験者が当て推量により正答する確率を表していると考えられており、一般に項目 j の当て推量母数と呼ばれている。

ところで、同じ国語力を有する受験者 A と B がある文学作品の題名を問う項目に回答する場面を考えてみると、A だけがこの作品の題名を知っている場合には、A がこの項目に正答する確率はほぼ 100% となるのに対し、B がこの項目に正答する確率はほぼ 0% になると考えられる。このように、特定の受験者 i が項目 j に正答する確率 (個人正答確率) を π_{ij} とすると、一般には、同じ受験者特性値 θ_i を有する受験者であっても同じ π_{ij} を有するとは限らないため、Lord (1980, pp.227-228) では、項目特性関数 $P_j(\theta_i)$ の解釈として、

1. 項目 j と同じ項目特性関数を有する項目の集団からランダムに抽出された項目に対し、受験者特性値が θ_i である特定の受験者が正答する確率
2. 受験者特性値が θ_i である受験者の集団からランダムに抽出された受験者が項目 j に正答する確率
3. 項目 j と同じ項目特性関数を有する項目の集団からランダムに抽出された項目に対し、受験者特性の値が θ_i である受験者集団からランダムに抽出された受験者が正答する確率

という3通りの解釈が提案されている。さらに、南風原 (1984) では、 $P_j(\theta_i)$ と π_{ij} との関係について更なる検討が加えられており、その結果、受験者特性値が θ_i である下位集団において π_{ij} の期待値をとると、その値が $P_j(\theta_i)$ に一致すること、つまり、

$$P_j(\theta_i) = E_{i|\theta_i} [\pi_{ij}] \quad (1.3)$$

なる関係が成立すること等が示されている。

1.3 局所独立性

項目反応理論における重要な仮定の一つに局所独立性 (local independence) の仮定と呼ばれるものがある (Lord & Novick, 1968, p.360)。例えば、2つの項目 j と k が局所独立の関係にあるとは、 θ の値を固定した場合に項目 j と k への反応 u_j, u_k が統計的に互いに独立になることを表しており、いま、 U_j を、項目 j に正答した場合には $U_j = 1$ となり、項目 j に誤答した場合には $U_j = 0$ となる確率変数であるとする、局所独立性の仮定と

は, 項目 j と k の間に

$$\begin{aligned} \text{Prob}(U_j = u_j, U_k = u_k | \theta_i) &= \text{Prob}(U_j = u_j | \theta_i) \text{Prob}(U_k = u_k | \theta_i) \\ &= P_j(\theta_i)^{u_j} (1 - P_j(\theta_i))^{1-u_j} P_k(\theta_i)^{u_k} (1 - P_k(\theta_i))^{1-u_k} \end{aligned} \quad (1.4)$$

なる関係を仮定することに他ならない. ここで, (1.4) 式中の $\text{Prob}(U_j = u_j, U_k = u_k | \theta_i)$ は θ のある水準 θ_i で条件付けた場合に U_j, U_k が u_j, u_k となる同時確率を表しており, $\text{Prob}(U_j = u_j | \theta_i)$ と $\text{Prob}(U_k = u_k | \theta_i)$ はそれぞれ θ_i で条件付けた場合に U_j, U_k が u_j, u_k となる周辺確率を表している.

項目 j と k が局所独立の関係にあるとすると, (1.4) 式より, θ_i と u_k で条件付けた場合の u_j の条件付き確率 $\text{Prob}(u_j | u_k, \theta_i)$ は

$$\begin{aligned} \text{Prob}(u_j | u_k, \theta_i) &= \frac{\text{Prob}(u_k, u_j | \theta_i)}{\text{Prob}(u_k | \theta_i)} \\ &= \frac{\text{Prob}(u_j | \theta_i) \text{Prob}(u_k | \theta_i)}{\text{Prob}(u_k | \theta_i)} \\ &= \text{Prob}(u_j | \theta_i) \end{aligned} \quad (1.5)$$

となる. このことから, 項目 j と k が局所独立であるとは, θ の値を固定した場合に u_j が u_k に依存しないことを意味していると言える.

ところで, 局所独立性と類似した概念に実験的独立性 (experimental independence) と呼ばれるものがある (Lord & Novick, 1968, p. 44). 実験的独立性とは, j, k という項目を受験者 i に与えて回答させるという測定手続き (実験) の結果得られる項目反応 u_{ij}, u_{ik} が統計的に互いに独立となることを表しており, u_{ij} と u_{ik} の同時確率を $\text{Prob}(u_{ij}, u_{ik})$ とし, その周辺確率をそれぞれ $\text{Prob}(u_{ij}), \text{Prob}(u_{ik})$ とすると, 項目 j, k が実験的に独立であるとは,

$$\begin{aligned} \text{Prob}(u_{ij}, u_{ik}) &= \text{Prob}(u_{ij}) \text{Prob}(u_{ik}) \\ &= \pi_{ij}^{u_{ij}} (1 - \pi_{ij})^{1-u_{ij}} \pi_{ik}^{u_{ik}} (1 - \pi_{ik})^{1-u_{ik}} \end{aligned} \quad (1.6)$$

なる関係が成立することを意味している. いま, 受験者 i の有する特性値が θ_i であり, 特性値が同じく θ_i である受験者全員の個人正答確率が等しかったとすると*1,

$$P_j(\theta_i) = \pi_{ij} \quad (1.7)$$

$$P_k(\theta_i) = \pi_{ik} \quad (1.8)$$

*1 Lord & Novick (1968, p. 539) における誤差の等質性が成立している状況に相当する.

となることから、受験者特性値が θ_i である受験者が項目 j, k に対し u_j, u_k と反応する確率 $Prob(u_j, u_k | \theta_i)$ は

$$Prob(u_j, u_k | \theta_i) = P_j(\theta_i)^{u_j} (1 - P_j(\theta_i))^{1-u_j} P_k(\theta_i)^{u_k} (1 - P_k(\theta_i))^{1-u_k} \quad (1.9)$$

となる。したがって、仮に項目 j と k が実験的に独立であり、かつ、同じ受験者特性値 θ_i を有する受験者全員が同じ個人正答確率を有するならば、項目 j と k は局所独立であると言える。

しかしながら、1.2 節でも述べたように、一般には同じ受験者特性値を有する受験者が必ずしも同じ個人正答確率を有するとは限らない。このような場合に、項目 j と k が局所独立となるための条件として、南風原 (2000) では、項目 j, k 間での実験的独立性と個人正答確率の局所独立性、2つの独立性の成立というものが挙げられている。ここで、個人正答確率の局所独立性とは、 π_{ij}, π_{ik} をその実現値とする確率変数をそれぞれ Π_j, Π_k とすると、受験者特性のある水準 θ_i で条件付けたときに Π_j, Π_k が π_j, π_k となる同時確率 $Prob(\Pi_j = \pi_j, \Pi_k = \pi_k | \theta_i)$ が θ_i で条件付けたときに Π_j, Π_k が π_j, π_k となる周辺確率 $Prob(\Pi_j = \pi_j | \theta_i), Prob(\Pi_k = \pi_k | \theta_i)$ の積に等しくなること、つまり、

$$Prob(\Pi_j = \pi_j, \Pi_k = \pi_k | \theta_i) = Prob(\Pi_j = \pi_j | \theta_i) Prob(\Pi_k = \pi_k | \theta_i) \quad (1.10)$$

なる関係が成立することを表している。この個人正答確率の局所独立性と項目 j, k 間の実験的独立性により局所独立性が成立する過程に関しては、南風原 (2000) において、以下のように説明されている (以下の引用においては、本論文との整合性を保つため、原文に記載されている式番号を本論文における式番号に置き換えている)。

最後にまとめとして、(1.4) 式によって局所独立性が実現する“プロセス”を確認しておく。最初に、特性値が θ である被験者 i が項目 1 に正答する確率を考える。その被験者の個人正答確率 π_{i1} は不明であるが、被験者 i を θ で規定される下位集団からランダムに選ばれたものと仮定すると、正答確率は、その下位集団における正答確率 $P_1(\theta)$ で与えられる。次に、その同じ被験者が項目 2 に正答する確率を考える。もしその被験者が項目 1 に正答していたら、その被験者の項目 1 における個人正答確率 π_{i1} は高いことが推測されるが、測定内容が一次元的で (1.10) 式が成り立っていれば、そのことはその被験者の項目 2 の個人正答確率 π_{i2} の高低について、何らヒントを与えるものではない。さらに、実験的独立性も満たされていれば、項目 1 に正答または誤答したことによって、項目 2 の正誤は何ら影響されないから、項目 2 の正答確率は項目 1 の場合と同様に $P_2(\theta)$ で与えられ、かつ、項目 1、項目 2 の両方に正答する確率は、積 $P_1(\theta)P_2(\theta)$ で与えられることになる。項目 3 以降もまったく同様である。

1.4 局所依存性

受験者や項目の特性を項目反応理論に基づいて推定する際には、通常対象となるデータに対して局所独立性が仮定される。しかしながら、分析の対象となるデータの中には局所独立性を仮定することができないと考えられるデータも多く存在する。Ferrara, Huynh, & Michaels (1999), Hoskens & De Boeck (1997), Kreiner & Christensen (2004), Yen (1993) では、このような項目反応間の局所依存性 (local dependence) を引き起こす要因について考察が加えられており、これらの研究で局所依存性の要因として取り上げられたものをその特性により分類していくと、局所依存性を引き起こすと考えられる状況を以下の3タイプに分類することができる。

- 項目反応間に直接的なつながりがある状況 (タイプ I)
- 項目反応間に間接的なつながりがある状況 (タイプ II)
- 複数の受験者特性の影響を受けている項目が複数存在する状況 (タイプ III)

タイプ I とは、ある項目への反応が他の項目への反応に影響を与える状況のことを指しており、このような状況の例としては、以下の項目群に回答する状況が挙げられる。

問 1. 変数 X の平均を求めよ.

問 2. 変数 X の分散を求めよ.

いま、 X が N 人の受験者のテストの得点 x_1, \dots, x_N を表しているとする、問 2 において求める X の分散 S_X^2 は X の平均 M_X を用いて

$$S_X^2 = \frac{1}{N} \sum_{i=1}^N (x_i - M_X)^2 \quad (1.11)$$

と表される。このため、これらの項目群においては、問 1 を誤答してしまった受験者は高い確率で問 2 にも誤答することになる。

このように、タイプ I の状況においては項目反応間の実験的独立性が満たされなくなるため、1.3 節の議論より、局所独立性が満たされなくなると考えられる。

タイプ II とは、1 つの成功 (失敗) が複数の項目への反応に影響を与える状況のことを指しており、国語や英語のテストに含まれる文章読解問題に回答する場面や社会や理科のテストで図表の読み取り問題に回答する場面がこのような状況の例として挙げられる。

例えばいま、項目 j と k がある課題文の内容理解について問う項目群であったとすると、この課題文をたまたま正しく読み取ることができた (できなかった) 受験者は項目 j と k の両方に正答 (誤答) する確率が高くなると考えられる。するとこのとき、特性値が θ_i で

ある受験者が項目 j に正答した上で項目 k に正答する確率は項目 j に誤答した上で項目 k に正答する確率よりも高くなると考えられる。このように、タイプ II の状況においては項目反応間において実験的独立性が成立しなくなるため、1.3 節の議論より、局所独立性が満たされなくなると考えられる。

タイプ III とは、テストに含まれる複数の項目への反応がテストが測定しようとしている受験者特性 θ 以外の特性 θ' から影響を受けている状況のことを指しており、このような状況の例としては、いくつかの文章題を含む数学のテストにおいて、各文章題に正答できるか否かが、このテストで測定しようとしている数学力のみならず、課題文を正確に読み取るために必要な文章読解能力にも依存している状況が挙げられる*2。

例えばいま、項目 j と k が θ ならびに θ' からの影響を受けており、同じ θ の値を有する受験者の間では θ' の値が高い受験者ほど項目 j, k に対する個人正答確率が高くなる傾向にあったとする。するとこのとき、特性値が θ_i である受験者集団において π_{ij} の値が高い受験者ほど π_{ik} の値が高くなり、 Π_j と Π_k の間に相関が生じることとなる。このように、タイプ III の状況においては個人正答確率の局所独立性が満たされなくなるため、1.3 節の議論より、局所独立性の仮定が満たされなくなると考えられる。

1.5 局所依存性を考慮した項目反応モデル

1.4 節でも述べたように、分析の対象となるデータの中には局所依存性が生じていると考えられるものも多く存在する。そこで、これまでに、局所依存性を考慮したモデルも先行研究において複数提案されてきており、これらのモデルをその特性により分類すると以下の3タイプに分類することができる。

- 局所依存関係にある項目群を1つの多値型項目と考えるモデル (タイプ a)
- 局所依存性を表現する受験者母数を加えるモデル (タイプ b)
- 局所依存性を表現する項目母数を加えるモデル (タイプ c)

タイプ a とは、局所依存関係にあると考えられる項目が J 個存在する場合にこれら J 個の項目 $d(j)$ における正答数をその分析対象とするモデルのことを指しており、このタイプ a のモデルとしては、例えば以下の段階反応モデル (graded response model, GRM) が挙

*2 このようなテストにおいては、同じ数学力を有する受験者であっても文章読解能力の低い受験者はそうでない受験者に比べて文章題に正答する確率が低くなってしまうと考えられる。このように、同じ受験者特性値 θ_i を有する受験者であってもテストが測定しようとしている特性以外の要因によりある項目への回答傾向が異なってしまう場合、この項目においては「特異項目機能 (Differential Item Functioning, DIF) が生じている」という。

げられる (Samejima, 1969).

$$P_{d(j)}(r|\theta_i) = P_{d(j)}^*(r|\theta_i) - P_{d(j)}^*(r+1|\theta_i) \quad (1.12)$$

ここで, $P_{d(j)}(r|\theta_i)$ は θ の値が θ_i である受験者が $d(j)$ において r 個の項目に正答する確率を表しており, $P_{d(j)}^*(r|\theta_i)$ は θ の値が θ_i である受験者が $d(j)$ において r 個以上の項目に正答する確率を表している. なお, 通常 $P_{d(j)}^*(r|\theta_i)$ を表す項目反応モデルとしては 2PLM ((1.1) 式) が用いられるため, (1.12) 式の $P_{d(j)}(r|\theta_i)$ は

$$P_{d(j)}(r|\theta_i) = \frac{1}{1 + \exp[-1.7a_{d(j)}^*(\theta_i - b_r^*)]} - \frac{1}{1 + \exp[-1.7a_{d(j)}^*(\theta_i - b_{r+1}^*)]} \quad (1.13)$$

と表されることが多い. ただし, (1.13) 式中の困難度母数 b_r^*, b_{r+1}^* に関しては, $b_r^* \leq b_{r+1}^*$ なる順序制約が課されることとなる.

このように, $d(j)$ における正答数を算出してこれに対し多値型の項目反応モデルを当てはめてデータの分析を行うことにより, 項目レベルでの独立, 依存の問題を回避することができるため, タイプ a のモデル群は局所依存性を考慮したモデルであると言える.

なお, タイプ a のモデルとしては, この他にも, 名義反応モデル (Bock, 1972) や一般化部分採点モデル (Muraki, 1992) などといったモデルを考えることができ, 例えば, Sireci, Thissen & Wainer (1991) においては, 項目反応間に局所依存性が生じていると考えられる文章読解問題の正答数に対して名義反応モデルを当てはめて分析を行っている.

タイプ b のモデルとは, 局所独立性を仮定して分析する際に用いられる 2PLM のようなモデルに対して何らかの受験者母数を導入し, この新たに導入された受験者母数によって局所依存性を表現するモデル群のことを指している. このタイプ b のモデルの例としては, 以下の 2 値型のベイジアンテストレットモデル (Baeyesian testlet model, BTM) が挙げられる (Bradlow, Wainer, & Wang, 1999).

$$P_{j \in d(j)}(\theta_i) = \frac{1}{1 + \exp[-1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)} - \gamma_{id(j)})]} \quad (1.14)$$

ここで, $P_{j \in d(j)}(\theta_i)$ は受験者特性値が θ_i であるような受験者が $d(j)$ に含まれる項目 j に正答する確率を表しており, $a_{j \in d(j)}, b_{j \in d(j)}$ は, 2PLM 同様, 項目 j の識別力と困難度を表す項目母数となっている. ただし, $a_{j \in d(j)}$ は $\gamma_{id(j)}$ が 0 であるときの $\theta_i = b_{j \in d(j)}$ における (1.14) 式の傾きに比例する値となっており, $b_{j \in d(j)}$ は $\gamma_{id(j)}$ が 0 であるときに $P_{j \in d(j)}(\theta_i)$ が 0.5 となる θ の値を表している. また, (1.14) 式の exp 内を

$$-1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)} - \gamma_{id(j)}) = -1.7a_{j \in d(j)}((\theta_i - \gamma_{id(j)}) - b_{j \in d(j)}) \quad (1.15)$$

のように変形すると, 同じ特性値 θ_i を有する受験者であっても $\gamma_{id(j)}$ の値が負となる受験者は正となる受験者に比べて項目 j に正答する確率が高くなることがわかる. このように,

(1.14) 式中の $\gamma_{id(j)}$ は、受験者特性値が θ_i である受験者 i と $d(j)$ との間の交互作用を表す受験者母数となっている。通常 $\gamma_{id(j)}$ の事前分布としては $N(0, \sigma_{\gamma_{d(j)}}^2)$ なる分布が仮定されることとなる*3。ここで、 $\sigma_{\gamma_{d(j)}}^2$ の値が 0 より大きい場合には、 $d(j)$ に含まれる項目の間で項目反応が局所依存となり、 $\sigma_{\gamma_{d(j)}}^2$ の値が 0 である場合には、 $d(j)$ に含まれる項目の間で反応が局所独立となるため、 $\sigma_{\gamma_{d(j)}}^2$ は $d(j)$ における局所依存度を表す母数となっていることがわかる。

このように、 $d(j)$ に含まれるそれぞれの項目に対し局所独立性を仮定する通常の項目反応モデルを仮定して、これらのモデルに対し、 θ 以外の受験者特性を表す $d(j)$ 内で共通の受験者母数を導入することにより、同じ受験者特性値 θ_i を有する受験者であっても、項目群単位で得意、不得意があることを表現することができるため、タイプ b のモデル群は局所依存性を考慮したモデルであると言える。

なお、タイプ b のモデルに関しても、タイプ a のモデル同様、2 値型の BTM 以外のモデルを考えることができ、受験者が当て推量により正答する確率を考慮した 3 パラメタ・ベイジアンテストレットモデル (three parameter Bayesian testlet model, 3PBTM) や多値型のベイジアンテストレットモデル (Wang, Bradlow, & Wainer, 2002)、多次元の項目反応モデル (multidimensional item response model, MIRM) などがこれにあたる。このうち、例えば、Nandakumar (1990) や Li, Bolt & Fu (2006) においては、MIRM および MIRM に含まれる項目母数に制約を課したモデルが、局所依存性が生じているデータの発生、分析モデルとして使用されている。

タイプ c のモデルとは、局所独立性を仮定して分析する際に用いられる 2PLM のようなモデルに対して何らかの項目母数を導入し、この新たに導入した項目母数によって局所依存性を表現するモデル群のことを指している。このタイプ c のモデルの例として、ここではまず、以下の固定組み合わせモデル (constant combination model, CCM) を紹介する (Hoskens & De Boeck, 1997)。

$$P(u_j, u_k | \theta_i) = \frac{\exp[u_j Z_j + u_k Z_k - u_j u_k b_{jk}]}{1 + \exp[Z_j] + \exp[Z_k] + \exp[Z_j + Z_k - b_{jk}]} \quad (1.16)$$

$$Z_j = 1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)}) \quad (1.17)$$

$$Z_k = 1.7a_{k \in d(j)}(\theta_i - b_{k \in d(j)}) \quad (1.18)$$

ここで、 $P(u_j, u_k | \theta_i)$ は受験者特性値が θ_i である受験者が局所依存関係にある項目 j と k に対し u_j, u_k と反応する確率を表しており、(1.17) 式中の $a_{j \in d(j)}, b_{j \in d(j)}$ は項目 j の識別力と困難度を、(1.18) 式中の $a_{k \in d(j)}, b_{k \in d(j)}$ は項目 k の識別力と困難度をそれぞれ表している。ただし、 $a_{j \in d(j)}$ および $a_{k \in d(j)}$ は、 b_{jk} が 0 であるときに限り、 $\theta_i = b_{j \in d(j)}$ および $\theta_i = b_{k \in d(j)}$ における各項目の項目特性関数の傾きに比例する値となり、 $b_{j \in d(j)}$ および

*3 事前分布に関しては、2.2.3 節を参照のこと。

$b_{k \in d(j)}$ は, b_{jk} が 0 であるときに限り, 各項目の項目特性関数の値が 0.5 となる θ の値と等しくなる. また (1.16) 式より, θ_i で条件付けた場合の項目 j と k に関するログオッズ比 ω_{jk} を求めると,

$$\begin{aligned}\omega_{jk} &= \ln \left(\frac{P(U_j = 1, U_k = 1 | \theta_i)}{P(U_j = 1, U_k = 0 | \theta_i)} \bigg/ \frac{P(U_j = 0, U_k = 1 | \theta_i)}{P(U_j = 0, U_k = 0 | \theta_i)} \right) \\ &= \ln \frac{P(U_j = 1, U_k = 1 | \theta_i) P(U_j = 0, U_k = 0 | \theta_i)}{P(U_j = 1, U_k = 0 | \theta_i) P(U_j = 0, U_k = 1 | \theta_i)} \\ &= -b_{jk}\end{aligned}\tag{1.19}$$

となる. 項目 j と k が局所独立な関係にある場合には ω_{jk} が 0 となることから, (1.16) 式中の b_{jk} は項目 j と k の局所依存度を表す母数となっており, b_{jk} が正の値を取る場合には項目 j, k に対して同一の反応 (正答あるいは誤答) をする受験者が異なる反応をする受験者に比べて少なくなり, 負の値を取る場合には項目 j, k に対して同一の反応をする受験者が異なる反応をする受験者に比べて多くなることがわかる. また, 本論文で取り上げられるタイプ c のモデルとして, ここではもう一つ, 2 パラメタ・ロジスティックコピュラモデル (two-parameter logistic copula model, 2PLCM) を紹介する. 2PLCM とは Braeken, Tuerlinckx, & DeBoeck (2007) にて提案されたラッシュコピュラモデル (Rasch copula model) を各項目の項目特性関数が 2PLM となるように拡張したものであり, 2PLCM においては, 局所依存関係にある 2 つの項目 j, k に対し u_j, u_k と反応する確率が以下の様に表される.

$$\begin{aligned}P(u_j, u_k | \theta_i) &= u_j u_k + (-1)^{2-u_j} u_k Q_j(\theta_i) + (-1)^{2-u_k} u_j Q_k(\theta_i) \\ &\quad + (-1)^{u_j+u_k} C(Q_j(\theta_i), Q_k(\theta_i))\end{aligned}\tag{1.20}$$

ここで, (1.20) 式中の $Q_j(\theta_i)$ は, 受験者特性値が θ_i である受験者が項目 j に誤答する確率を表しており, 今の場合,

$$Q_j(\theta_i) = 1 - \frac{1}{1 + \exp[-1.7a_j(\theta_i - b_j)]}\tag{1.21}$$

となる. また, $C(Q_j(\theta_i), Q_k(\theta_i))$ は $Q_j(\theta_i), Q_k(\theta_i)$ を引数とするコピュラ関数 (フランクコピュラ) を表しており, 今の場合,

$$C(Q_j(\theta_i), Q_k(\theta_i)) = -\frac{1}{\delta_{jk}} \log \left[1 - \frac{W(Q_j(\theta_j)) W(Q_j(\theta_k))}{W(1)} \right]\tag{1.22}$$

$$W(x) = 1 - \exp[-\delta_{jk} x]\tag{1.23}$$

となる (Frank, 1979). なお, (1.22) 式ならびに (1.23) 式中の δ_{jk} が正の値を取る場合には, 受験者特性値が θ_i である下位集団において項目 j と k に対し同一の反応をする受験

者が異なる反応をする受験者よりも多くなり、 δ_{jk} が負の値を取る場合には、受験者特性値が θ_i である下位集団において項目 j と k に対して異なる反応をする受験者が同一の反応をする受験者よりも多くなる。このように、 $\delta_{jk} \neq 0$ である場合には $\omega_{jk} \neq 0$ となることから、 δ_{jk} は項目 j と k の間の局所依存度を表す項目母数となっていることがわかる。

このように、 $d(j)$ に含まれる各項目に対しまず局所独立性を仮定する通常の項目反応モデルを仮定し、その上で、項目対毎の依存度を表す項目母数を導入することにより、 θ_i で条件づけたときの項目反応間の関連を表現することができるようになるため、タイプ c のモデル群は局所依存性を考慮したモデルであると言える。

なお、タイプ c のモデルとしては、この他にも、受験者が当て推量により正答する確率を考慮した 3 パラメタ・固定組み合わせモデル (three parameter constant combination model, 3PCCM, Chen & Wang, 2007) や hybrid kernel を利用した項目反応モデル (Ip, 2002)、局所依存線形ロジスティックテストモデル (locally dependent linear logistic test model, LDLLTM, Ip, Smits, & De Boeck, 2009)、連結型項目反応モデル (conjunctive item response model, CIRM, Jannarone, 1986) などといったモデルが存在し、例えば、豊田 (2005) においては、CIRM を用いて局所依存性が生じていると考えられるのデータの分析を行っている。

1.6 本研究の目的

日本で実施されるテストについて考えてみると、その出題形式の特徴として大問形式の出題が多いことが挙げられる (荒井・前川, 2005; 池田, 1992; 石塚・中畝・内田・前川, 2001)。これに加えて近年では、日本においても、項目反応理論に基づくテストの作成、運用が行われるようになってきており (e.g., 語彙・読解力検定)、これらのことから、今後日本において大問形式の問題を含むテストを項目反応理論に基づき作成、評価、運用する機会が増加していくであろうことが予想される。

通常大問形式の問題には、ある項目への回答を利用して他の項目への回答を導き出す項目やある文章や図表に対する理解を問う一連の項目群等が含まれており、1.4 節での議論と考え合わせると、大問形式の問題内においては、多くの場合、項目反応間に局所依存性が生じていると考えられる。1.5 節では、先行研究において局所依存性を考慮した様々なモデルが提案されていることを述べたのだが、これらのモデルに関しては、モデル自体に関する理解や普及が局所独立性を仮定するモデルほどは進んでおらず、現時点では、局所依存性が生じていると考えられるデータを項目反応理論に基づいて分析する際には、局所独立性を仮定したモデルを利用することが一般的となっている (e.g., Yang & Gao, 2008)。したがって、今後日本において項目反応理論に基づくテストの作成、評価、運用が増加していくであろうことを見据えると、どのような条件下であれば局所独立性を仮定した分析が

局所依存性に対して頑健であると言えるのか、詳細に検討することは重要であると考えられる。また、局所依存性を考慮したモデルを使用することにより、受験者特性や項目母数の推定精度がどの程度改善されるのか、それは局所独立性を仮定するモデルに代えてより複雑な局所依存性を考慮したモデルを使用するに値するのか、使用するに値するのであれば、どのような条件下のことであり、その条件下では、用いるモデルの種類によって改善の度合いが異なるのか、詳細に調べていくことも必要であると考えられる。

そこで、本論文では、これらの点に関して主にシミュレーションを通して検討を行い、大問形式の問題を含むテストの項目反応理論に基づく作成、評価、運用に資する知見の獲得を目指す。具体的には、まず第2章にて、局所依存性が局所独立性を仮定した受験者特性の推定に対しどのような影響を与えるのかシミュレーションを通して検討を行う。次に、第3章にて、局所依存性が局所独立性を仮定した項目特性の推定に与える影響についてシミュレーションによる検討を行い、解析的な観点からも考察を加える。さらに、第4章では、局所依存性が局所独立性を仮定したテスト情報量の推定に対しどのような影響を与えるのかシミュレーションを通して検討を加える。そして、第5章にて、局所依存関係を考慮することにより受験者特性の推定精度がどれほど改善されるのか、それは局所依存性を考慮したモデルを使用するに値するものなのか、値するのであればそれはどのような条件下のことであり、使用するモデルによって改善の度合いは異なるのか、といったことについてシミュレーションを通じた検討を加える。最後に、第6章にて、これらの研究成果に関する総合的な考察を行う。

第 2 章

局所独立性を仮定した受験者特性の推定に与える影響

2.1 問題と目的

本章では、局所依存性が局所独立性を仮定した受験者特性の推定にどのような影響を及ぼすのか、シミュレーションを通して検討を行う*1。このテーマに関しては既にいくつかの研究において検討が加えられており、例えば、Bradlow et al. (1999) では、発生モデルを 2 値型の BTM とし、分析モデルを 2PLM とするシミュレーションを実施した結果、

- 局所依存性を有するデータに対し局所独立性を仮定して受験者特性の推定を行うと、局所独立性が満たされているデータにおける値よりも、推定量のバイアスが大きくなる
- 局所依存性を有するデータに対し局所独立性を仮定して受験者特性の推定を行うと、局所独立性が満たされているデータにおける値よりも、事後分布の 95% 確信区間の幅 (mean 95% posterior interval width, M95%PIW) が大きくなる
- 局所依存性を有するデータに対し局所独立性を仮定して受験者特性の推定を行うと、局所独立性が満たされているデータにおける値よりも、M95%PIW が実際に真値を含む確率が小さくなる
- 局所依存性を有するデータに対し局所独立性を仮定して受験者特性の推定を行っても、局所独立性が満たされているデータにおける値に比べて、推定値と真値との順位相関係数はそれほど低下しない

*1 本章は、登藤 (2010) を加筆, 再構成したものである。

ことが報告されている。また, Junker (1991) では, 局所依存性を有するデータに対し局所独立性を仮定して受験者特性値の最尤推定を行うと, 最尤推定量の標準誤差が増加することが数学的に証明されており, 安井 (2013) においては, ある項目に誤答した場合には必ず以降の項目においても誤答するという構造を持ったデータと局所独立性を仮定できるデータを発生させ, これらのデータに対し 2PLM をあてはめて受験者特性の推定を行った結果,

- 局所依存性を有するデータに対し局所独立性を仮定して受験者特性の推定を行うと, 局所独立性が満たされているデータにおける値よりも, 推定量のバイアスが負の方向に大きくなる
- 局所依存性を有するデータに対し局所独立性を仮定して受験者特性の推定を行うと, 局所独立性が満たされているデータにおける値よりも, 推定量の平均二乗誤差平方根の値が小さくなる
- 局所依存性を有するデータに対し局所独立性を仮定して受験者特性の推定を行うと, 局所独立性が満たされているデータにおける値よりも, 推定値と真値の間の相関係数が大きくなる

ことが報告されている。

以上のように, 先行研究においては, 局所依存性が局所独立性を仮定した受験者特性の推定に及ぼす影響について, 主にシミュレーションを通じた検討が行われてきたのだが, これらの研究においては, 主に受験者数や項目数の多い状況が想定されてきた。例えば, Bradlow et al. (1999) の研究においては, 受験者数が 1000, 項目数が 60 である状況が想定されており, 安井 (2013) においては, 受験者数が 1000, 項目数が 12 である状況が想定されている。しかしながら, 実際にテストを作成, 評価, 運用する場面へ本研究の知見を応用するという観点からは, 受験者数や項目数を系統的に変化させてシミュレーションを行い, 受験者数が少なく項目数が多い状況等において局所依存性がどのような影響を与えるのか, 検討する必要があると考えられる。また, 先行研究においては, 局所依存性の発生メカニズム (モデル) として一種類の特定のモデルを取り上げて検討を行っている。しかしながら, 先述の通り, 局所依存性の発生メカニズムとしては様々なものが考えられ, これらのメカニズム間で受験者特性値の推定の頑健性が異なるかどうかについても, 複数の発生モデルを用いたシミュレーションを行い検討する必要があると考えられる。

そこで, 本章では, 発生メカニズムの異なる複数のモデルから項目反応データを発生させてシミュレーションを実施し, 局所独立性を仮定した受験者特性値の推定の頑健性が対象となるデータの発生メカニズムによって異なるのかどうか, 検討を行った。また, シミュレーションを実施する際にデータの受験者数や項目数を系統的に変化させ, 局所依存性の影響が受験者数や項目数とどのような関係にあるのか, 検討を行った。

2.2 方法

本章では、全ての項目間で局所独立性が満たされているテストと局所依存性が生じているテスト、2種類のテストに対する仮想的な項目反応データを発生させて、このデータに対し局所独立性を仮定する同一の項目反応モデルを当てはめて母数の推定を行った。そして、この手続きを多数回繰り返し、得られた推定値から受験者特性値の推定量のバイアス ($Bias(\hat{\theta}_i)$), 平均二乗誤差平方根 ($RMSE(\hat{\theta}_i)$), ならびに、真値と推定値との相関係数の平均 ($cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$) を算出し、2種類のテスト間でこれらの値の比較を行った。

2.2.1 項目反応データの発生モデル

全ての項目間で局所独立性が満たされているテストの場合

全ての項目間で局所独立性が満たされているテストを想定し、このテストに対する項目反応データを発生させる際には、(1.1) 式の 2PLM を使用した (以下にモデル式を再掲)。

- 2PLM のモデル式 (再掲)

$$P_j(\theta_i) = \frac{1}{1 + \exp[-1.7a_j(\theta_i - b_j)]} \quad (2.1)$$

全ての項目間で局所依存性が生じているテストの場合

本章では、異なる発生モデルを用いた場合に局所依存性の影響が異なるのかどうか検討するため、全ての項目間で局所依存性が生じているテストを想定し、このテストに対する項目反応データを発生させる際には、1.5 節で述べたタイプ c のモデルである、(1.16) 式の CCM と (1.20) 式の 2PLCM を使用した (CCM, 2PLCM のモデル式に関しては以下に再掲)。

- CCM のモデル式 (再掲)

$$P(U_j, U_k | \theta_i) = \frac{\exp[U_j Z_j + U_k Z_k - U_j U_k b_{jk}]}{1 + \exp[Z_j] + \exp[Z_k] + \exp[Z_j + Z_k - b_{jk}]} \quad (2.2)$$

$$Z_j = 1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)})$$

$$Z_k = 1.7a_{k \in d(j)}(\theta_i - b_{k \in d(j)})$$

- 2PLCM のモデル式 (再掲)

$$\begin{aligned}
 P(U_j, U_k | \theta_i) &= U_j U_k + (-1)^{2-U_j} U_k Q_j(\theta_i) + (-1)^{2-U_k} U_j Q_k(\theta_i) \\
 &\quad + (-1)^{U_j+U_k} C(Q_j(\theta_i), Q_k(\theta_i)) \tag{2.3} \\
 Q_j(\theta_i) &= 1 - \frac{1}{1 + \exp[-1.7a_j(\theta_i - b_j)]} \\
 C(Q_j(\theta_i), Q_k(\theta_i)) &= -\frac{1}{\delta_{jk}} \log \left[1 - \frac{W(Q_j(\theta_j)) W(Q_j(\theta_k))}{W(1)} \right] \\
 W(x) &= 1 - \exp[-\delta_{jk}x]
 \end{aligned}$$

2.2.2 シミュレーションの手続き

本研究では、受験者数や項目数についてより広範な状況を想定し、局所独立性を仮定した受験者特性値の推定の頑健性が受験者数や項目数によってどのような影響を受けるのか、系統的な検討を行うため、受験者数に関しては100人、300人、500人、1000人という4水準を設定し、項目数に関しては10項目、30項目、50項目という3水準を設定して、計12条件について以下に述べるシミュレーションを行なった。

なお、全ての項目間で局所依存性が生じているテストに対する項目反応データを発生させる場合には、 $2j - 1$ 番目の項目と $2j$ 番目の項目が局所依存関係にあるものと仮定した ($j = 1, 2, \dots, J/2$, J はテストに含まれる項目の数)。また、局所依存関係にある項目 j, k 間の局所依存度に関しては、発生モデルとして CCM を使用する場合には $b_{jk} = -2$ とし、2PLCM を使用する場合には $\delta_{jk} = 30$ とした。このように b_{jk} および δ_{jk} を定めると、事前に行ったシミュレーションより、どちらの発生モデルを用いた場合にも、同様の、具体的には、局所依存関係にある2つの項目 j, k に対し $(1, 1), (0, 0)$ という同一の反応をする受験者の割合がおおよそ等しくなるように、データを発生させることができることが分かっている。例えば、受験者数を100、項目数を10とした場合に、局所依存関係にある2項目1, 2に対し各受験者がどのように反応するのか、同じ100人の受験者が同じテストを100回受験する状況をシミュレートして調べてみると、表2.1および表2.2にあるように、CCMを発生モデルとした場合には項目1, 2に対し $(1, 1), (0, 0)$ と反応する受験者の割合(100回分の平均)は84%(73%+11%)となり、2PLCMを発生モデルとした場合には88%(52%+36%)となった。

表 2.1 CCM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 人の受験者が 100 回受験した場合の平均値)

	$U_2 = 1$	$U_2 = 0$
$U_1 = 1$	73%	10%
$U_1 = 0$	6%	11%

表 2.2 2PLCM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 人の受験者が 100 回受験者した場合の平均値)

	$U_2 = 1$	$U_2 = 0$
$U_1 = 1$	52%	12%
$U_1 = 0$	0%	36%

加えて, Hoskens & De Boeck (1997) では, 局所依存関係にある項目間で見られる b_{jk} の値としては 2 から -2 の範囲が現実的であるとの報告がなされており, これらのことから, b_{jk} の値として -2 を, δ_{jk} の値として 30 を仮定してシミュレーションを行うことにより, 実際に局所依存関係にあるデータを分析する際に生じるであろう局所依存性の影響について検討することが可能となる.

実際にシミュレーションを行う際には, 2.2.1 節で紹介した 3 種類の発生モデルから項目反応データを発生させて, これらのデータに対し局所独立性を仮定する 2PLM をあてはめて母数の推定を行った. なお, 母数の推定にはマルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo, MCMC) を利用し^{*2}, シミュレーション用のプログラムの記述および実行には統計解析向けのプログラミング言語およびその実行環境である R を利用した.

本研究で行ったシミュレーションの詳細は以下のようにまとめられる.

1. 各受験者の受験者特性値 θ_i ($i = 1, 2, \dots, N$) と項目 j ($j = 1, 2, \dots, J$) の識別力 a_j および困難度 b_j の真値を, それぞれ標準正規分布 $N(0, 1)$, 一様分布 $U(0.3, 1.5)$ および一様分布 $U(-2.0, 2.0)$ から発生させる^{*3}.
2. 全ての項目間で局所独立性が満たされているテストにおいて各受験者が各項目に正答する確率を, 手続き 1 で得られた受験者特性, 項目母数の真値と 2PLM の式

^{*2} MCMC の詳細に関しては 2.2.3 節を参照のこと.

^{*3} なお, 擬似乱数を発生させる際には R に実装されているメルセンヌツイスターを使用した.

- ((2.1) 式) から算出し, $N \times J$ の正答確率行列 \mathbf{A} を作成する.
3. 一様分布 $U(0, 1)$ から乱数を NJ 個発生させて, $N \times J$ の一様乱数行列 \mathbf{B} を作成する.
 4. 行列 \mathbf{A}, \mathbf{B} の各要素を比較して, 全ての項目間で局所独立性が満たされているテストに対する項目反応データの行列 \mathbf{U} を作成する. 具体的には, $a_{ij} \geq b_{ij}$ である場合には $u_{ij} = 1$ とし, $a_{ij} < b_{ij}$ である場合には $u_{ij} = 0$ とする.
 5. 全ての項目間で局所依存性が生じているテストにおいて各受験者が局所依存関係にある項目 j, k に対し $(0, 0), (1, 0), (0, 1), (1, 1)$ と反応する確率を, 手続き 1 で得られた受験者特性, 項目母数の真値と CCM の式 ((2.2) 式) を用いて算出し, 反応パターン毎に $N \times (J/2)$ の同時確率行列を作成してそれぞれ D, E, F, G とする.
 6. 一様分布 $U(0, 1)$ から乱数を $NJ/2$ 個発生させて, $N \times (J/2)$ の一様乱数行列 \mathbf{H} を作成する.
 7. 行列 D, E, F, G と \mathbf{H} の各要素を比較して, 全ての項目間で局所依存性が生じているテストに対する $N \times J$ の項目反応データの行列 \mathbf{U}' を作成する. 具体的には, $d_{ij} \geq h_{ij}$ であるならば $(u'_{i(2j-1)}, u'_{i(2j)}) = (0, 0)$ であるとし, $e_{ij} + d_{ij} \geq h_{ij} > d_{ij}$ であるならば $(u'_{i(2j-1)}, u'_{i(2j)}) = (1, 0)$ である, などとする.
 8. 発生モデルが 2PLCM である場合に関しても手続き 5 から 7 までを行い, 全ての項目間で局所依存性が生じているテストに対する $N \times J$ の項目反応データの行列 \mathbf{U}'' を作成する.
 9. $\mathbf{U}, \mathbf{U}', \mathbf{U}''$ に対し 2PLM をあてはめて, MCMC により母数の推定を行う.
 10. 手続き 1 から 9 までを 100 回繰り返す.
 11. 手続き 10 の結果得られた 100 組の推定値を用いて, 以下の統計量を算出する.

$$Bias(\hat{\theta}_i) = \frac{1}{100} \sum_{r=1}^{100} \hat{\theta}_{ir} - \theta_i \quad (2.4)$$

$$RMSE(\hat{\theta}_i) = \sqrt{\frac{1}{100} \sum_{r=1}^{100} (\hat{\theta}_{ir} - \theta_i)^2} \quad (2.5)$$

$$cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \frac{1}{100} \sum_{r=1}^{100} cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_r) \quad (2.6)$$

ここで, θ_i は手続き 1 で発生させた受験者特性の真値を表しており, $\hat{\theta}_i$ は θ_i の推定量を, $\hat{\theta}_{ir}$ は手続き 10 の結果得られた θ_i の r 個目の推定値をそれぞれ表わしている. また, $\boldsymbol{\theta}$ は N 人分の θ_i をその要素とする受験者特性の真値ベクトルを表しており, $\hat{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}$ の推定量を, $\hat{\boldsymbol{\theta}}_r$ は手続き 10 の結果得られた $\boldsymbol{\theta}$ の r 個目の推定値をそれぞれ表わしている. さらに, (2.6) 式中の $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_r)$ は $\boldsymbol{\theta}$ と $\hat{\boldsymbol{\theta}}_r$ の相関係数を表している. したがって, $Bias(\hat{\theta}_i)$

は $\hat{\theta}_i$ にどの程度の過大あるいは過小推定の傾向があるのかを表す統計量となっており、 $RMSE(\hat{\theta}_i)$ は $\hat{\theta}_i$ の推定の誤差の大きさを表す統計量となっている。また、 $cor(\theta, \hat{\theta})$ は $\hat{\theta}$ と θ との間で平均的にどの程度の線形性が見られるのかを表す統計量となっている。

2.2.3 母数の推定法

マルコフ連鎖

2.2.2 節でも述べたように、本研究では、2PLM をあてはめて母数の推定を行う際、MCMC を利用した。本節では、MCMC の基礎となるマルコフ連鎖について豊田 (2008) を参考に解説する。

いま、等間隔で離散的な時点の列を考える。このとき、ある時点 $t (t = 0, 1, 2, \dots)$ とし、時点 t における確率変数とその実現値をそれぞれ X^t, x^t とすると、マルコフ連鎖とは、任意の時点 t に対して

$$Prob(X^{t+1} = x^{t+1} | X^0 = x^0, X^1 = x^1, \dots, X^t = x^t) = Prob(X^{t+1} = x^{t+1} | X^t = x^t) \quad (2.7)$$

が常に成立する確率変数の系列のことを表している。

次に、 X^t が取りうる値の集合 Ω を状態空間と呼ぶこととし、 $X^t (t = 0, 1, 2, \dots)$ が同じ Ω を有するものとする。また、 $Prob(X^{t+1} = x^{t+1} | X^t = x^t)$ を推移確率と呼ぶこととし、推移確率を $p_{x^{t+1}|x^t}$ と表すことにする。ただし、推移確率は特定の時点 t には依存しないものとする。ここで、 $p_{x^{t+1}|x^t}$ を x^t 行 x^{t+1} 列の要素とする推移確率行列 \mathbf{P} を考える。すると、時点 t において X^t が x^t となる確率を $\pi_{x^t}^{t+1}$ 、この確率を Ω に含まれる全ての値について求め行ベクトルとしたものを π^{t+1} としたときに、 X^{t+1} が状態空間に含まれる各値をとる確率の分布は、

$$\pi^{t+1} = \pi^t \mathbf{P} \quad (2.8)$$

と表されることになる。さらに、上記の式から、

$$\pi^{t+1} = \pi^0 \mathbf{P}^{t+1} \quad (2.9)$$

なる関係が導き出されることから、時点 $t+1$ において X^{t+1} がある値をとる確率 π^{t+1} は $t=0$ における各状態の確率と推移確率行列によって決定されることがわかる。

あるマルコフ連鎖が、

1. 任意のどの状態からでも任意の他の状態に到達可能である
2. 確率変数がある値 $\omega (\omega \in \Omega)$ をとるという事象の起こる間隔については周期性が見られず、全くのランダムである

という2つの性質を満たすとき、任意の π^0 に対して、 $t \rightarrow \infty$ のときに π^{t+1} が不変分布と呼ばれる一意の π に収束することが知られている。このとき、先の (2.8) 式から、

$$\pi = \pi P \quad (2.10)$$

という関係が成り立つこととなり、不変分布に至った後のマルコフ連鎖の系列においては、各確率変数が状態空間の各値をとる確率は全て π で表されることとなる。

マルコフ連鎖モンテカルロ法

いま、推定対象の母数をその要素とする母数ベクトルを λ とすると、 λ は以下のように表わされる ((2.11) 式)。

$$\lambda = (\theta_1, \theta_2, \dots, \theta_N, a_1, a_2, \dots, a_J, b_1, b_2, \dots, b_J)' \quad (2.11)$$

この λ を得られたデータ U から推定する際、ベイズ統計学では、ベイズの定理*4 を用いて λ の事後分布 $Prob(\lambda|U)$ を以下のように算出し、事後分布の最頻値 (MAP 推定値) あるいは期待値 (EAP 推定値) をもって、 λ の推定値とする ((2.12) 式)。

$$\begin{aligned} Prob(\lambda|U) &= \frac{L(U|\lambda)Prob(\lambda)}{Prob(U)} \\ &= \frac{L(U|\lambda)Prob(\lambda)}{\int L(U|\lambda)Prob(\lambda)d\lambda} \end{aligned} \quad (2.12)$$

ここで、(2.12) 式中の $L(U|\lambda)$ は、仮に母数ベクトルの値が λ であった場合に、その条件の下で U というデータが得られる確率 (λ の尤度) を表しており、今の場合、

$$L(U|\lambda) = \prod_{i=1}^N \prod_{j=1}^J P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{1-u_{ij}} \quad (2.13)$$

となる。

このように、データ U から母数ベクトル λ をベイズ推定する際には、(2.12) 式に見られるような λ の事後分布を算出することになるのだが、一般に、この事後分布を解析的に求めることは難しいことが多い。これに対し、MCMC とは、データ U とマルコフ連鎖の特性を利用して λ の事後分布を不変分布とするマルコフ連鎖を構築し、このマ

*4 ベイズの定理とは、事象 A と事象 B に関して以下の等式が常に成り立つことをいう。

$$Prob(B|A) = \frac{Prob(A|B)Prob(B)}{Prob(A)}$$

ここで、上式中の $Prob(B|A)$ は、事象 A が生じた上で事象 B が生じる確率を表わしており、事象 B の事後確率と呼ばれる。また、 $Prob(B)$ は、事象 A が生じる前の事象 B の生起確率を表わしており、事象 B の事前確率と呼ばれる。

マルコフ連鎖から λ の事後分布に従う多数の標本を抽出して、MAP 推定値や EAP 推定値を数値的に求める方法のことをいう。 λ の事後分布を不変分布とするマルコフ連鎖を構築するための MCMC の具体的なアルゴリズムとしては、ギブスサンプラー (Gibbs sampler, Geman & Geman, 1984), データ拡大アルゴリズム (data augmentation and Gibbs sampling, Tanner & Wong, 1987), メトロポリス-ヘイスティングスアルゴリズム (Metropolis-Hastings algorithm, Hastings, 1970) といったものが存在するのだが、本研究では、Patz & Junker (1999) にて提案されたギブス内メトロポリス-ヘイスティングスアルゴリズム (Metropolis-Hastings within Gibbs algorithm) を利用することとした。そして、2.2.2 節中の手続き 9 においては、このアルゴリズムおよび $N \times J$ の項目反応行列 \mathbf{U} から母数ベクトル λ の事後分布を不変分布とするマルコフ連鎖を構築して、 λ に従う標本を多数抽出し、これらの平均値をもって λ の推定値 (EAP 推定値) とした (\mathbf{U}' , \mathbf{U}'' を利用した場合も同様)。なお、MCMC を利用して母数を推定するにあたっては、事前には各母数が統計的に互いに独立であったと仮定し、 θ_i, a_j, b_j の事前分布を、それぞれ標準正規分布 $N(0, 1)$, 正規分布 $N(1, 0.25)$, 標準正規分布 $N(0, 1)$ であると仮定した。

MCMC を用いた母数推定の詳細は以下のようにまとめられる。

1. \mathbf{U} より各受験者の標準得点 z_i と各項目の点双列相関係数 r_j , 正答率 p_j を算出し、 $z_i, r_j, 1 - p_j$ をそれぞれ θ_i, a_j, b_j の初期値 θ_i^0, a_j^0, b_j^0 として、 λ の初期値 λ^0 を

$$\begin{aligned}\lambda^0 &= (\lambda_1^0, \lambda_2^0, \dots, \lambda_{N+2J}^0)' \\ &= (\theta_1^0, \theta_2^0, \dots, \theta_N^0, a_1^0, a_2^0, \dots, a_J^0, b_1^0, b_2^0, \dots, b_J^0)' \\ &= (z_1, z_2, \dots, z_N, r_1, r_2, \dots, r_J, 1 - p_1, 1 - p_2, \dots, 1 - p_J)'\end{aligned}\quad (2.14)$$

とする。

2. λ_1^0 の候補となる λ_1^* を提案分布と呼ばれる以下の分布 ((2.15) 式) から抽出する。

$$h(\lambda_1^* | \lambda_1^0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\lambda_1^* - \lambda_1^0)^2}{2\sigma^2}\right]\quad (2.15)$$

3. λ_1^* と λ_1^0 から、採択確率と呼ばれる以下の確率 ((2.16) 式) を計算する。

$$\alpha(\lambda_1^* | \lambda_1^0) = \min\left(\frac{\text{Prob}(\lambda_1^* | \lambda_{-1}^0, \mathbf{U})h(\lambda_1^0 | \lambda_1^*)}{\text{Prob}(\lambda_1^0 | \lambda_{-1}^0, \mathbf{U})h(\lambda_1^* | \lambda_1^0)}, 1\right)\quad (2.16)$$

ただし、(2.16) 式中の λ_{-1}^0 は λ^0 より λ_1^0 を取り除いたベクトル

$$\lambda_{-1}^0 = (\lambda_2^0, \dots, \lambda_{N+2J}^0)'\quad (2.17)$$

を表している。

4. 一様分布 $U(0, 1)$ から発生させた乱数が (2.16) 式の値よりも小さければ $\lambda_1^1 = \lambda_1^*$ とし、大きければ $\lambda_1^1 = \lambda_1^0$ とする.
5. $\lambda_2^0, \dots, \lambda_{N+2J}^0$ に関しても 2 から 4 までの手続きを実行し、更新された母数ベクトルの実現値

$$\boldsymbol{\lambda}^1 = (\lambda_1^1, \lambda_2^1, \dots, \lambda_{N+2J}^1)' \quad (2.18)$$

を得る.

6. $\boldsymbol{\lambda}^1$ に対して 2 から 5 までの手続きを実行し、 $\boldsymbol{\lambda}^2$ を得る.
7. 手続き 6 を 20000 回繰り返す.
8. 手続き 7 の結果得られた 20000 組の母数ベクトルより $\boldsymbol{\lambda}$ の推定値を算出する.

なお, (2.15) 式中の σ^2 は, 手続き 4 における $\lambda_l^*(l = 1, \dots, N + 2J)$ の採択確率に影響を及ぼす母数となっており, 本研究では, 事前にシミュレーションを行い, 条件毎に λ_l^* の採択率が 25% ~ 50% になるように σ^2 の値を設定した. また, マルコフ連鎖の特性より, 手続き 7 の結果得られた 20000 組の母数ベクトルの実現値のうち冒頭の数千組の母数ベクトルに関しては, 手続き 1 で設定した初期値の影響を受けている (burn-in である) であると考えられた. そこで, 20000 組の母数ベクトルの時系列プロットを使用して初期値からの影響の有無を確認し, 事前に行ったシミュレーションの結果から, 全ての条件において, 最初の 3000 組の母数ベクトルを burn-in であるとして除外し, 残りの 17000 組の母数ベクトルの平均値をもって $\boldsymbol{\lambda}$ の推定値とすることにした.

2.3 結果と考察

本研究では, 局所依存性が局所独立性を仮定した受験者特性の推定に対しどのような影響を与えるのか平均的な傾向を把握するため, 条件毎に発生モデルが 2PLM である場合と CCM である場合, 2PLCM である場合との間で \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値に関する比較を行った. ただし, \overline{Bias} は $Bias(\hat{\theta}_i)$ の N 人分の平均を表しており, 2.2.2 節の説明より, 受験者特性値の推定量の平均的な過大あるいは過小推定の傾向を表す統計量となっている. また, \overline{RMSE} は $RMSE(\hat{\theta}_i)$ の N 人分の平均を表しており, 2.2.2 節の説明より, 受験者特性値の推定量の平均的な推定誤差の大きさを表す統計量となっている. したがって, これらの値を条件間で比較することにより, 局所独立性を仮定した受験者特性の推定に対し局所依存性が与える平均的な影響について検討することが可能となる. また, 2.2.2 節の説明より, $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値を条件間で比較することによって受験者特性値の推定量と真値との間に見られる線形関係に対し局所依存性が与える平均的な影響についても検討することが可能となる.

なお, 発生モデルが 2PLM である場合と CCM, 2PLCM である場合との間で \overline{Bias} や

\overline{RMSE} の値について比較を行う際には、その差異の大きさを母集団分布における θ_i の標準偏差に照らして評価することが必要となってくる。2.2.2 節のシミュレーションの手続きでも述べたように、本研究では、 θ_i の母集団分布として $N(0, 1)$ を仮定しており、母集団分布における θ_i の標準偏差は 1 となっている。したがって、 \overline{Bias} と \overline{RMSE} の発生モデル間での差異については、この値を、そのまま θ_i の母集団分布における標準偏差に対する相対的な大きさとして評価できることになる。

2.3.1 発生モデルが 2PLM の場合

まず、発生モデルが 2PLM である場合に $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値がどのようになるか、各条件において算出したところ、表 2.3 が得られた。

2.3.2 発生モデルが CCM の場合

次に、発生モデルが CCM である場合に $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値がどのようになるか、各条件において算出したところ、表 2.4 が得られた。また、発生モデルが CCM である場合に、発生モデルが 2PLM である場合との間で、 $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値がどの程度異なるのか、各条件において算出したところ、表 2.5 が得られた。ここで、 $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ はそれぞれ、 $\overline{Bias}, \overline{RMSE}, cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の発生モデルが 2PLM である時の値との差異を表している。さらに、発生モデルが CCM である場合に $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が受験者数によりどのように変化するか確認するためグラフを作成したところ、図 2.1, 図 2.2, 図 2.3 が得られた。加えて、発生モデルが CCM である場合に $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が項目数によりどのように変化するか確認するためグラフを作成したところ、図 2.4, 図 2.5, 図 2.6 が得られた。

これらの結果から、発生モデルが CCM である場合のバイアス、平均二乗誤差平方根、真値と推定値との相関係数について、以下のようなことが明らかとなった。

バイアスの観点からの評価

表 2.5 より、受験者数が $N = 100$ と少ない場合には、すべての項目数において、発生モデルが 2PLM である場合に比べて、発生モデルが CCM である場合の \overline{Bias} が 0.1 (θ_i の母集団分布における標準偏差の 10%) 以上正の方向に大きくなっており、局所依存性が無視できない程度の影響を与えていることがわかる。また、 $N = 300$ や $N = 500$ という条件においても、受験者数に比して項目数が多い場合 (項目数が受験者数の 10% 以上) には、発生モデルが CCM である場合の \overline{Bias} が 0.1 以上正の方向に大きくなっていった。さらに、図 2.1 や図 2.4 からは、すべての項目数において、受験者数の増加に伴い、 $\Delta\overline{Bias}$ の値が

減少するものの、すべての受験者数において、テストの項目数の増加に伴い、 $\overline{\Delta Bias}$ の値が正の方向に大きくなっていることが読み取れ、 \overline{Bias} の値に関しては、項目数を増やすことが局所依存性の影響の減少には結びつかないことが明らかとなった。

平均二乗誤差平方根の観点からの評価

表 2.5 より、受験者数が $N = 100$ で項目数が $J = 50$ である場合には、発生モデルが 2PLM である場合に比べて、 \overline{RMSE} の値が 0.15 (θ_i の母集団分布における標準偏差の 15%) 以上大きくなっており、局所依存性が大きな影響を与えていることがわかる。また、図 2.2 からは、受験者数が $N = 100$ と少ない場合には、発生モデルが 2PLM である場合との \overline{RMSE} の差が大きくなることが読み取れる。さらに、図 2.2 からは、 $N \geq 300$ の場合に、項目数が $J = 10$ から $J = 30$ と増加することによって $\overline{\Delta RMSE}$ の値が減少し、 $J = 30$ から $J = 50$ へと増加する際に $\overline{\Delta RMSE}$ の値が増加することが読み取れる。これは、 \overline{RMSE} がバイアスと標準誤差という 2 つの要因に作用される統計量であるため、項目数の増加によるバイアスの増加と標準誤差の減少のうち、項目数が $J = 10$ から $J = 30$ に増加する際には後者の影響が強く、 $J = 30$ から $J = 50$ に増加する際には前者の影響が強かったためであると考えられる。

真値と推定値との相関係数の観点からの評価

表 2.5 より、真値と推定値との相関係数に関しては、発生モデルが CCM である場合でも、発生モデルが 2PLM である場合に比べて極端に低くなることはなく、条件によっては、わずかに上回るケースもあることがわかる。したがって、真値と推定値との相関係数という観点からは、局所依存性の影響は大きくないであろうと判断することができる。また、図 2.6 からは、すべての受験者数において、項目数が $J = 10$ から $J = 30$ と増加する際に $\Delta cor(\theta, \hat{\theta})$ の値が増加し、 $J = 30$ から $J = 50$ へと増加する際に $\Delta cor(\theta, \hat{\theta})$ の値が減少することが読み取れる。

まとめ

様々あるテストの利用方法のうち、特定のテストの成績のみを用いて個人差の評価を行うという文脈においては、本章で検討対象となった 3 種類の統計量のうち、受験者特性値の真値と推定値との相関係数という観点からの評価が重要となってくる。そういった観点から上記の結果を見直してみると、真値と推定値との相関係数に対する局所依存性の影響は小さく、局所独立性を仮定したモデルは局所依存性に対して十分に頑健であると言える。

一方、1.1 節でも述べたように、項目反応理論を用いてテストの作成、評価、運用を行うことの利点として、特定のテストを超えて、各受験者の受験者特性値を共通の尺度上で比

較することができる、というものが挙げられ、そのような比較を行う文脈においては、バイアスの観点からの評価および平均二乗誤差平方根の観点からの評価も重要になってくる。そういった観点から上記の結果を見直してみると、特に受験者数が $N = 100$ のように少ない場合には、バイアス、平均二乗誤差平方根に対する局所依存性の影響が無視できないほど大きくなり、局所独立を仮定したモデルは局所依存性に対し頑健であるとは言えないであろう。

また、テストの項目数を増やすことにより、一般的には、種々の頑健性を増やすことができると考えられているのだが、上記の結果からは、局所依存関係にある項目を増やすことは局所依存性に対する頑健性を強めることにはつながらず、特にバイアスに関しては、項目数が多くなるほど局所独立性が満たされている場合とのバイアスの差が広がるという明確な逆の効果が見られることがわかった。

2.3.3 発生モデルが 2PLCM の場合

続いて、発生モデルが 2PLCM である場合に $\overline{Bias}, \overline{RMSE}, cor(\theta, \hat{\theta})$ の値がどのようになるか、各条件において算出したところ、表 2.6 が得られた。また、発生モデルが 2PLCM である場合に $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\theta, \hat{\theta})$ の値がどのようになるか、各条件において算出したところ、表 2.7 が得られた。さらに、発生モデルが 2PLCM である場合に $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\theta, \hat{\theta})$ の値が受験者数によりどのように変化するか確認するためグラフを作成したところ、図 2.7, 図 2.8, 図 2.9 が得られた。加えて、発生モデルが 2PLCM である場合に $\Delta\overline{Bias}, \Delta\overline{RMSE}, \Delta cor(\theta, \hat{\theta})$ の値が項目数によりどのように変化するか確認するためグラフを作成したところ、図 2.10, 図 2.11, 図 2.12 が得られた。

これらの結果から、発生モデルが 2PLCM である場合のバイアス、平均二乗誤差平方根、真値と推定値との相関係数について、以下のようなことが明らかとなった。

バイアスの観点からの評価

表 2.7 より、全ての受験者数および項目数において、 \overline{Bias} の値が発生モデルが 2PLM の場合とほとんど変わらないことがわかる。

平均二乗誤差平方根の観点からの評価

表 2.7 より、全ての受験者数および項目数において、発生モデルが 2PLM である場合に比べて、 \overline{RMSE} の値が大きくなっていた。特に、項目数が $J = 10$ であるという条件においては、 \overline{RMSE} の値が 0.10 (θ_i の母集団分布における標準偏差の 10%) 以上大きくなっており、局所依存性が大きな影響を与えていることが伺える。したがって、平均二乗誤差平方根という観点からは、局所独立性を仮定した受験者特性値の推定に対する局所依存性の

影響は無視できないものであると言える。

真値と推定値との相関係数からの評価

表 2.7 より、全ての受験者数および項目数において、発生モデルが 2PLM である場合に比べて、 $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が低下しており、特に、項目数が $J = 10$ と少ない条件においては、どの受験者数においても、0.10 前後 $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が低下している。また、図 2.12 からは、テストに含まれる項目の数を増やすことにより、 $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が発生モデルが 2PLM である場合の値に近づいていることが伺えるのだが、図 2.9 においては、受験者数の増加による $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値の改善は見られず、項目数が $J = 10$ である場合にはむしろ、受験者数が $N = 1000$ と最も多い場合に、発生モデルが 2PLM である場合との $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値の差が最大となっている。

まとめ

特定のテストにおける成績のみを用いて個人差の評価を行うという文脈から上記の結果を捉えなおしてみると、真値と推定値との相関係数に対する局所依存性の影響は無視できないものとなっており、特に項目数が少ない場合には、局所独立性を仮定したモデルは局所依存性に対して頑健ではなくなると言うことができる。

また、特定のテストを超えて各受験者の受験者特性値を共通の尺度上で比較するという文脈から上記の結果を捉えなおしてみると、バイアスに対する局所依存性の影響はほとんど見られなかったものの、平均二乗誤差平方根に対しては局所依存性が無視できないほどの影響を与えており、真値と推定値との相関係数に対する局所依存性の影響も考慮すると、局所独立を仮定したモデルは局所依存性に対し頑健であるとは言えないと考えられる。

また、上記の結果からは、受験者数を増やすことは局所依存性に対する頑健性の改善にはつながらず、特に真値と推定値との相関係数に関しては、項目数が少ない場合に、受験者数が多くなるほど局所独立性が満たされている場合の値との差が広がるという傾向が確認された。

2.3.4 先行研究との比較

2.2 節でも述べたように、本研究では、受験者数や項目数が少ない状況などより広範な状況を想定し、このような状況において、局所独立性を仮定した受験者特性値の推定が局所依存性に対してどの程度頑健であるのか検討を行うため、受験者数および項目数を系統的に変化させて局所依存性の影響との関連について検討し、また、Bradlow et al. (1999) とは異なる発生モデルを 2 種類用いて相互に局所依存性の影響に関して比較を行った。これらの点に関しては、2 値型の BTM を用いて局所依存性が生じているデータを発生させ

て検討を行った Bradlow et al. (1999) の研究との間で相違が見られるのだが、以下の点に関しては、本研究で CCM を発生モデルとした場合との間で、得られた結果が類似していた。

- 項目反応間に局所依存性が生じているデータにおいては、局所独立性が満たされている場合に比べて、 \overline{Bias} の値が増加する
- 項目反応間に局所依存性が生じているデータにおいては、局所独立性が満たされている場合に比べて、M95%PIW が受験者特性の真値を含む確率が低下する
- 項目反応間に局所依存性が生じているデータにおいて受験者特性の真値と推定値との順位相関係数の値を求めると、局所独立性が満たされている場合に比べて、その値はそれほど低下していない

なお、受験者特性値の推定誤差が大きくなればなるほど、M95%PIW が実際に真値を含む確率は低くなると考えられるので、M95%PIW という観点からの評価は平均二乗誤差平方根という観点からの評価に近いものになると考えられる。同様に、真値と推定値との順位相関係数という観点からの評価は、本研究で検討対象となった3種類の統計量のうち、真値と推定値との相関係数という観点からの評価に近いものになると考えられる。また、Junker (1991) の研究に関しては、本章で行ったシミュレーション研究とは異なるものの、本研究で CCM, 2PLCM を発生モデルとした場合との間で、以下の点に関して、得られた結果が類似していたと言える。

- 項目反応間に局所依存性が生じているデータに対し局所独立性を仮定して受験者特性の推定を行うと、推定量の標準誤差が増加する

2.3.5 局所依存性の発生プロセスと局所独立性を仮定したモデルの頑健性

本研究では、局所依存性が生じているデータの発生プロセス (モデル) 毎に、局所独立性を仮定した受験者特性値の推定に対する頑健性の評価結果に違いが見られた。また、Bradlow et al. (1999) では、本研究とは異なる発生プロセスが想定されていたのだが、その結果は、CCM を発生モデルとした場合に得られる結果と類似しており、2PLCM を発生モデルとした場合に得られる結果とは異なっていた。さらに、本研究の後に実施された安井 (2013) の研究においては、ある項目に誤答した場合にその項目と局所依存関係にある以降の項目にも必ず誤答するというやや特殊な発生プロセスを想定した結果、局所依存性が生じているデータにおいては、局所独立性が満たされている場合に比べて、 \overline{Bias} の値が負の方向に大きくなることが報告されており、一部の状況においては、局所独立性が満たされている場合に比べて、 \overline{RMSE} の値が小さくなる、 $cor(\theta, \hat{\theta})$ の値が大きくなるという

結果が得られている。これらの結果は、上述の3種類の発生プロセスのもとで得られた結果とはかなり異なるものとなっている。加えて、発生プロセスとして、CCMなどの特定のモデルを想定した場合であっても、本研究のように2項目ずつが局所依存の関係にあるという状況と、例えば、3項目ずつが局所依存の関係にあるという状況、4項目ずつが局所依存の関係にあるという状況とでは、局所独立性を仮定して受験者特性値を推定した場合に得られる結果が異なってくる可能性がある。したがって、究極的には、対象とするテストの構成 (e.g., テストに含まれる大問の数, 大問内の項目数) やデータの発生プロセス, 項目間の局所依存度等にできるだけ近いかたちでのシミュレーションを実施し、局所依存性に対する局所独立性を仮定した推定の頑健性を評価することが必要となる。

その一方で、本研究で得られた知見の応用という観点からは、多くの発生プロセスに共通する一般的な頑健性の評価というものも必要になってくると考えられる。そこで、その観点から本研究で得られた結果をまとめなおすと、以下のようなになる。

- $N = 100$ のように受験者数が少ない場合には、平均二乗誤差平方根に対する局所依存性の影響が無視できないほど大きくなる
- 一般に、項目数が増えることにより局所依存性に対する頑健性は増すことになると考えられるのだが、それらの項目が局所依存性を有している場合、項目数の増加に伴うバイアスに対する局所依存性の影響に関してはほとんど改善が見られない

表 2.3 発生モデルが 2PLM の場合の \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 100, J = 10$	0.134	0.638	0.768
$N = 100, J = 30$	0.083	0.438	0.900
$N = 100, J = 50$	0.017	0.356	0.936
$N = 300, J = 10$	0.059	0.625	0.779
$N = 300, J = 30$	0.042	0.431	0.905
$N = 300, J = 50$	0.028	0.349	0.940
$N = 500, J = 10$	0.055	0.618	0.773
$N = 500, J = 30$	0.044	0.428	0.902
$N = 500, J = 50$	0.056	0.342	0.941
$N = 1000, J = 10$	0.022	0.615	0.779
$N = 1000, J = 30$	0.018	0.426	0.904
$N = 1000, J = 50$	0.019	0.347	0.938

表 2.4 発生モデルが CCM の場合の \overline{Bias} , \overline{RMSE} , $cor(\theta, \hat{\theta})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\theta, \hat{\theta})$
$N = 100, J = 10$	0.237	0.683	0.764
$N = 100, J = 30$	0.339	0.530	0.912
$N = 100, J = 50$	0.388	0.512	0.943
$N = 300, J = 10$	0.096	0.649	0.770
$N = 300, J = 30$	0.145	0.446	0.909
$N = 300, J = 50$	0.196	0.440	0.940
$N = 500, J = 10$	0.078	0.646	0.759
$N = 500, J = 30$	0.110	0.436	0.904
$N = 500, J = 50$	0.156	0.382	0.936
$N = 1000, J = 10$	0.034	0.643	0.764
$N = 1000, J = 30$	0.052	0.421	0.907
$N = 1000, J = 50$	0.076	0.360	0.935

表 2.5 発生モデルが CCM の場合の $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\theta, \hat{\theta})$ の値

条件	$\Delta\overline{Bias}$	$\Delta\overline{RMSE}$	$\Delta cor(\theta, \hat{\theta})$
$N = 100, J = 10$	0.103	0.045	-0.004
$N = 100, J = 30$	0.256	0.092	0.012
$N = 100, J = 50$	0.372	0.156	0.008
$N = 300, J = 10$	0.037	0.023	-0.008
$N = 300, J = 30$	0.103	0.015	0.004
$N = 300, J = 50$	0.168	0.051	-0.001
$N = 500, J = 10$	0.024	0.029	-0.014
$N = 500, J = 30$	0.066	0.008	0.002
$N = 500, J = 50$	0.100	0.040	-0.005
$N = 1000, J = 10$	0.012	0.028	-0.015
$N = 1000, J = 30$	0.034	-0.005	0.003
$N = 1000, J = 50$	0.057	0.013	-0.003

表 2.6 発生モデルが 2PLCM の場合の \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 100, J = 10$	0.133	0.758	0.678
$N = 100, J = 30$	0.082	0.513	0.859
$N = 100, J = 50$	0.017	0.435	0.899
$N = 300, J = 10$	0.059	0.752	0.685
$N = 300, J = 30$	0.043	0.509	0.865
$N = 300, J = 50$	0.029	0.417	0.912
$N = 500, J = 10$	0.055	0.756	0.670
$N = 500, J = 30$	0.045	0.510	0.859
$N = 500, J = 50$	0.055	0.410	0.912
$N = 1000, J = 10$	0.022	0.781	0.653
$N = 1000, J = 30$	0.018	0.509	0.861
$N = 1000, J = 50$	0.019	0.414	0.911

表 2.7 発生モデルが 2PLCM の場合の $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	$\Delta\overline{Bias}$	$\Delta\overline{RMSE}$	$\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 100, J = 10$	-0.001	0.120	-0.089
$N = 100, J = 30$	-0.001	0.075	-0.041
$N = 100, J = 50$	0.000	0.079	-0.037
$N = 300, J = 10$	0.000	0.127	-0.093
$N = 300, J = 30$	0.001	0.077	-0.040
$N = 300, J = 50$	0.001	0.068	-0.028
$N = 500, J = 10$	0.000	0.138	-0.103
$N = 500, J = 30$	0.001	0.082	-0.043
$N = 500, J = 50$	-0.001	0.068	-0.028
$N = 1000, J = 10$	0.000	0.166	-0.125
$N = 1000, J = 30$	0.000	0.084	-0.043
$N = 1000, J = 50$	0.000	0.067	-0.027

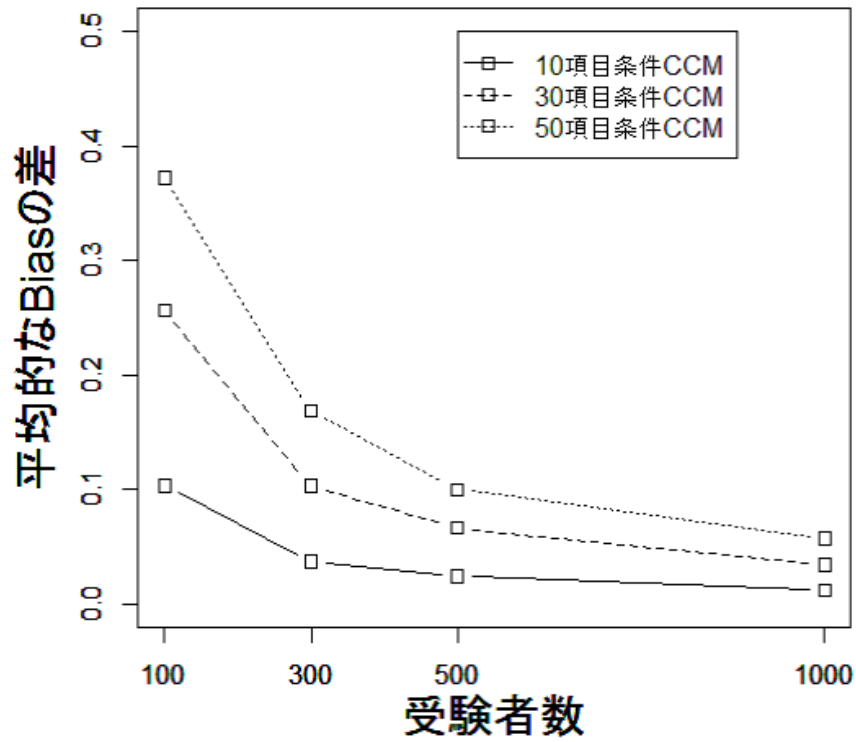


図 2.1 発生モデルが CCM である場合の $\Delta \overline{Bias}$ の受験者数による変化

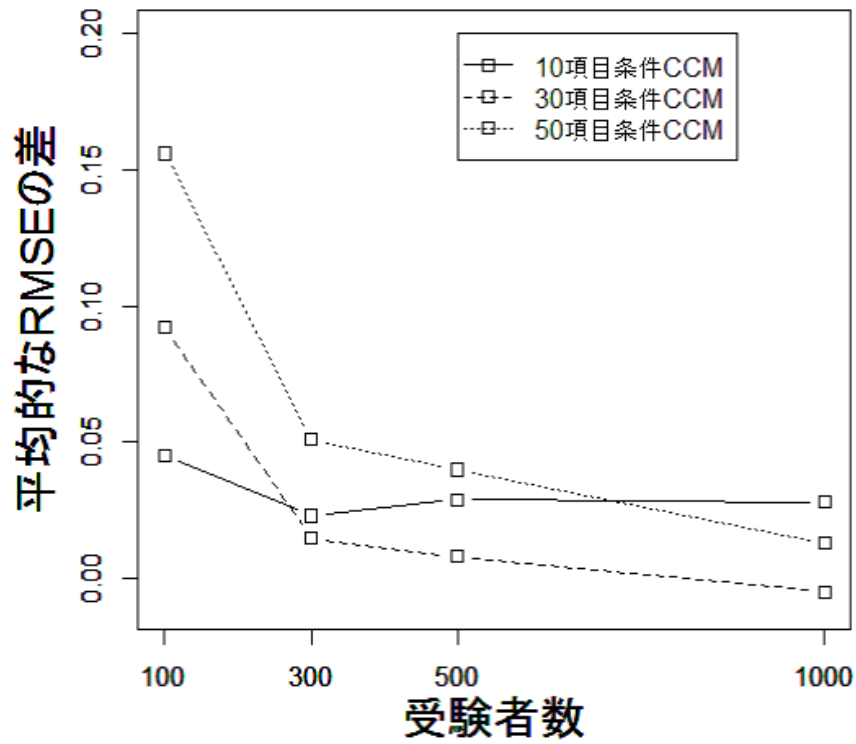


図 2.2 発生モデルが CCM である場合の $\overline{\Delta RMSE}$ の受験者数による変化

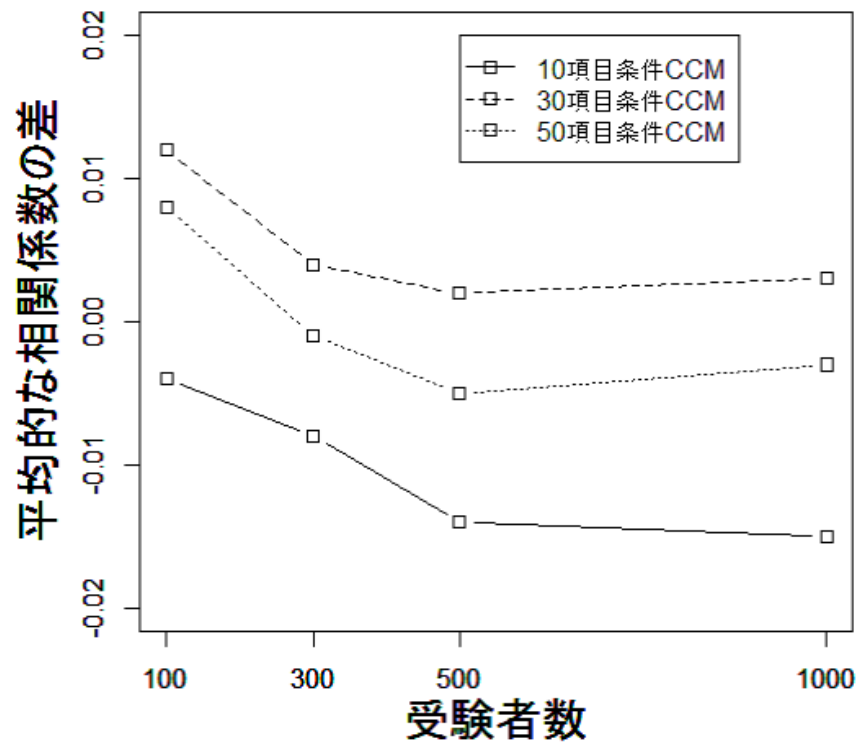


図 2.3 発生モデルが CCM である場合の $\Delta cor(\theta, \hat{\theta})$ の受験者数による変化

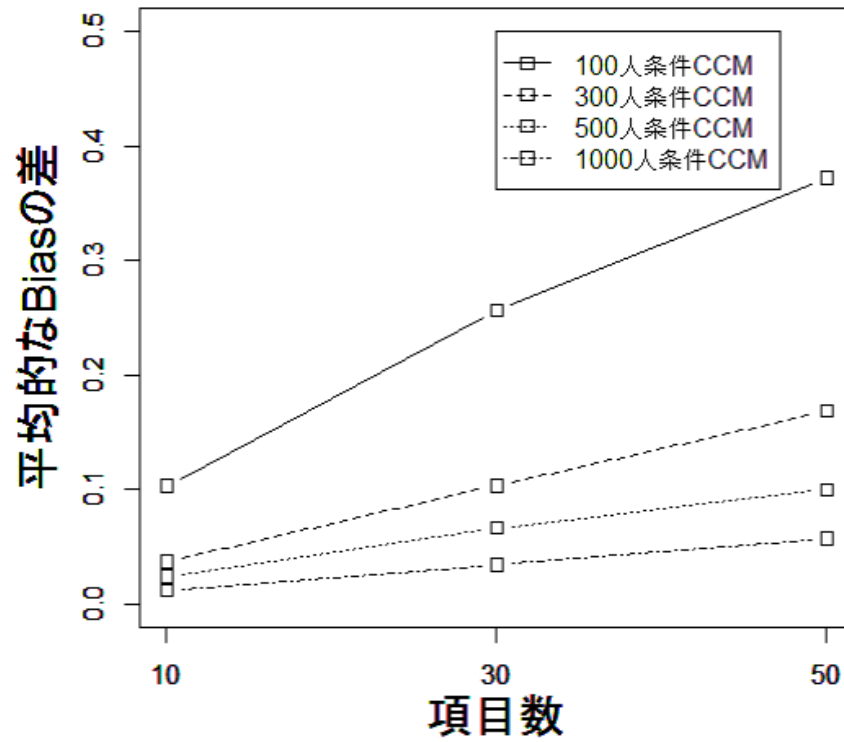


図 2.4 発生モデルが CCM である場合の $\Delta \overline{Bias}$ の項目数による変化

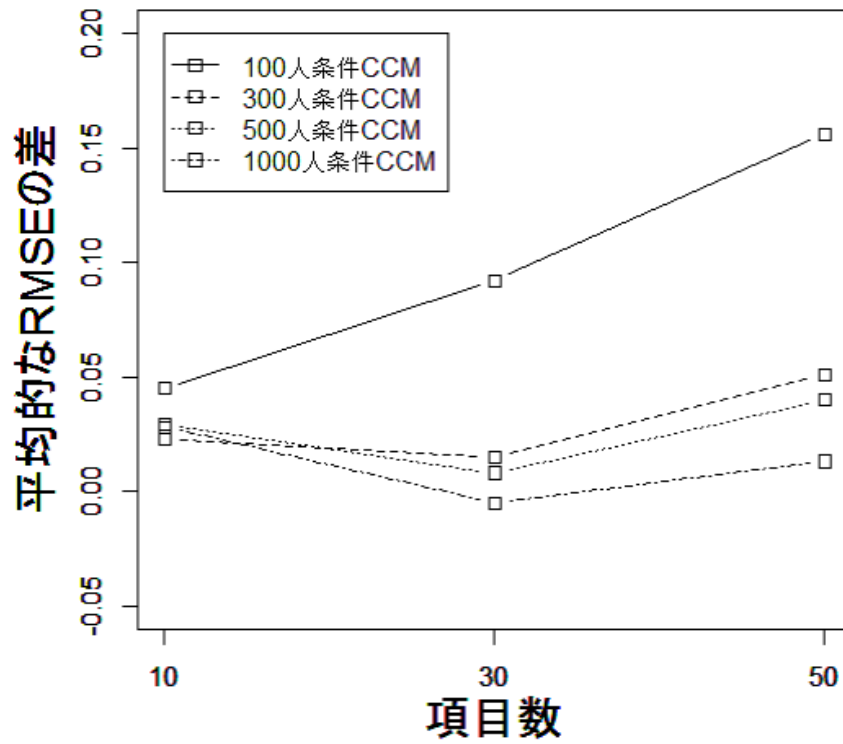


図 2.5 発生モデルが CCM である場合の $\Delta \overline{RMSE}$ の項目数による変化

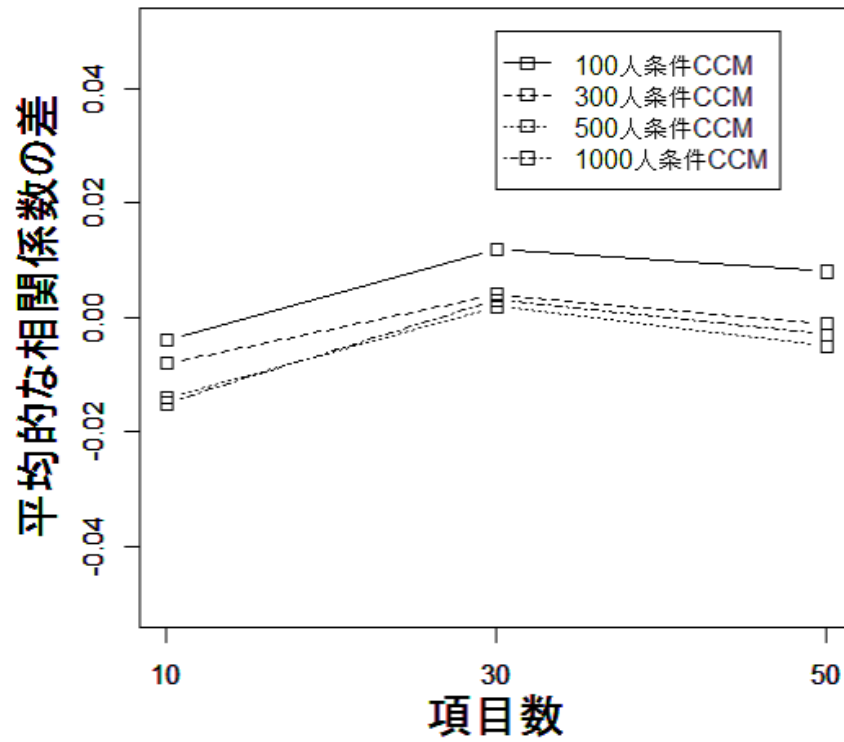


図 2.6 発生モデルが CCM である場合の $\Delta cor(\theta, \hat{\theta})$ の項目数による変化

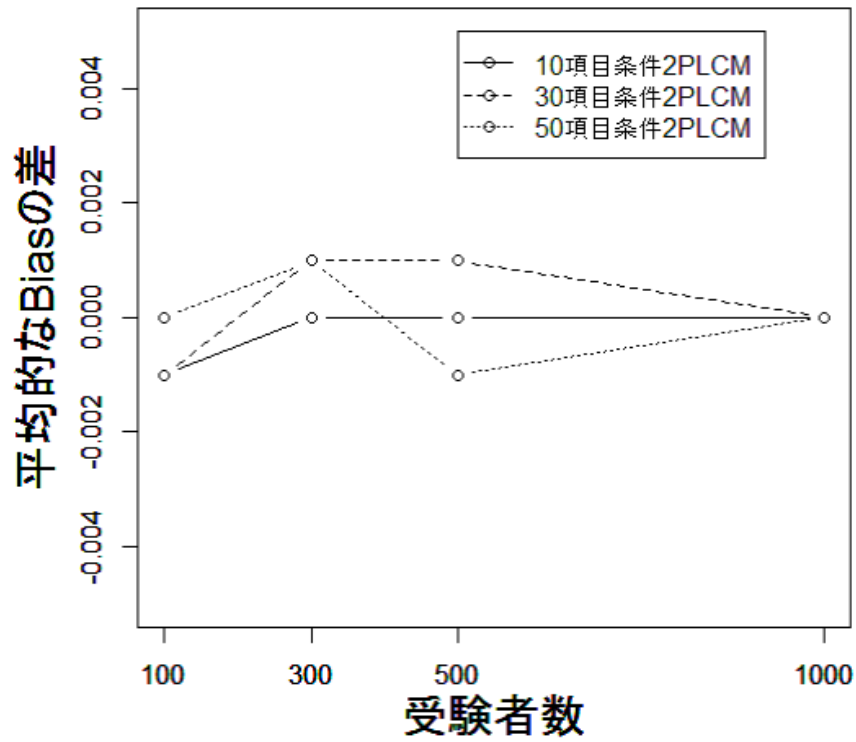


図 2.7 発生モデルが 2PLCM である場合の $\overline{\Delta Bias}$ の受験者数による変化

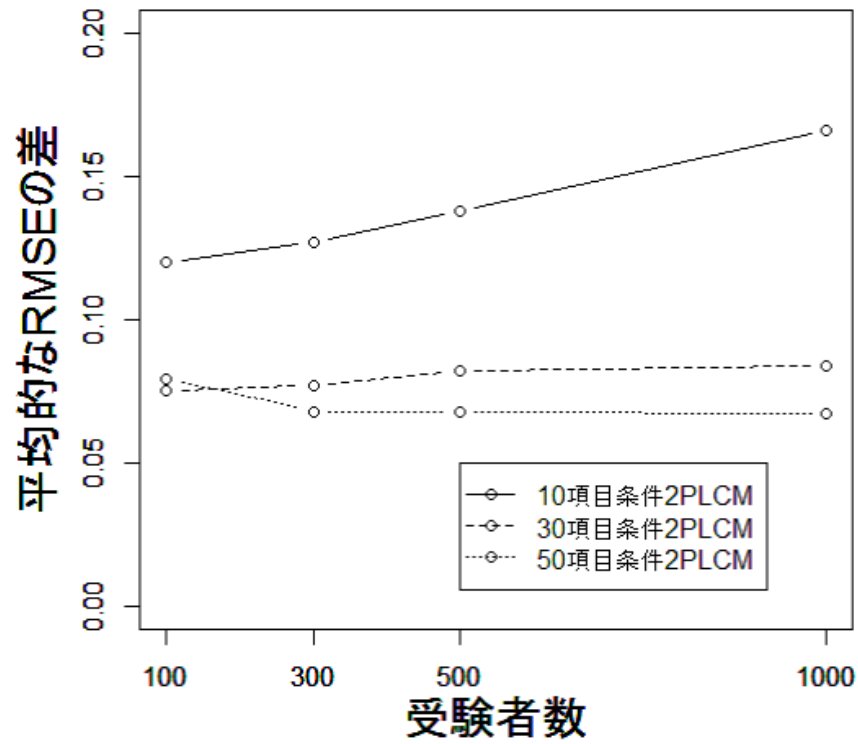


図 2.8 発生モデルが 2PLCM である場合の $\overline{\Delta RMSE}$ の受験者数による変化

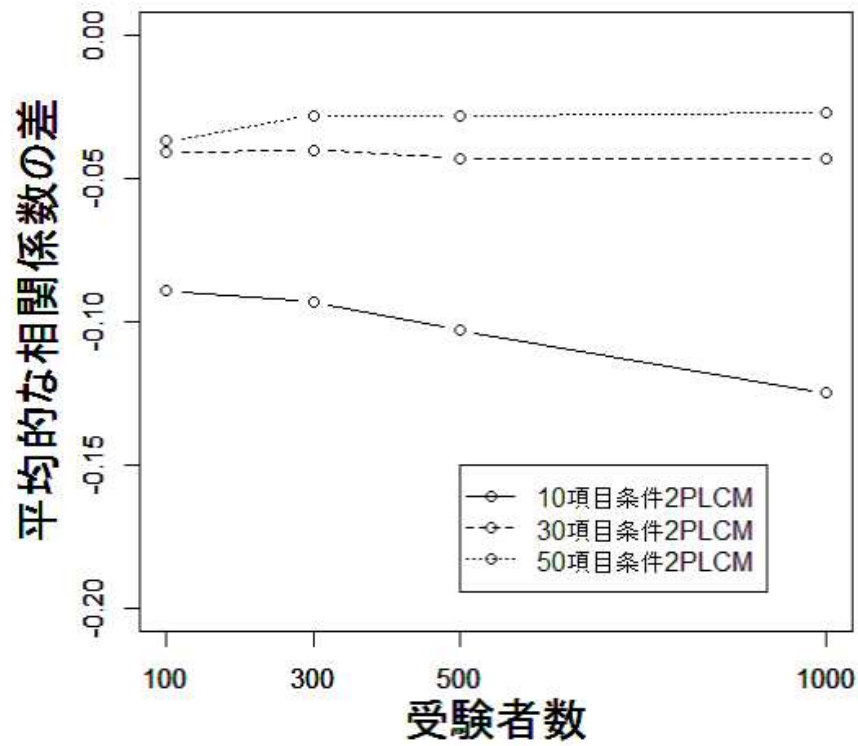


図 2.9 発生モデルが 2PLCM である場合の $\Delta cor(\theta, \hat{\theta})$ の受験者数による変化

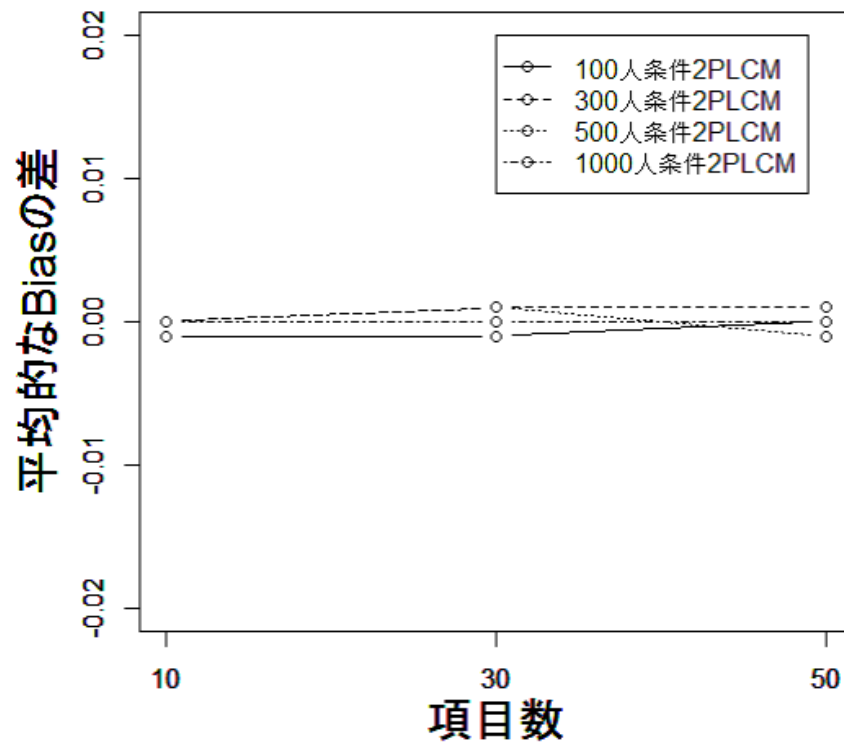


図 2.10 発生モデルが 2PLCM である場合の $\overline{\Delta Bias}$ の項目数による変化

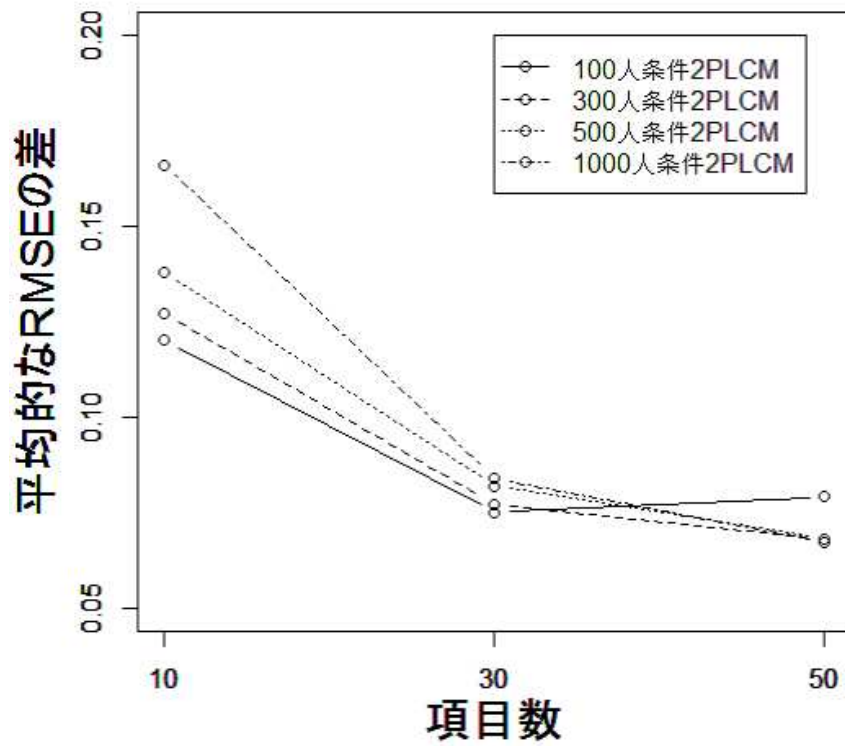


図 2.11 発生モデルが 2PLCM である場合の $\overline{\Delta RMSE}$ の項目数による変化

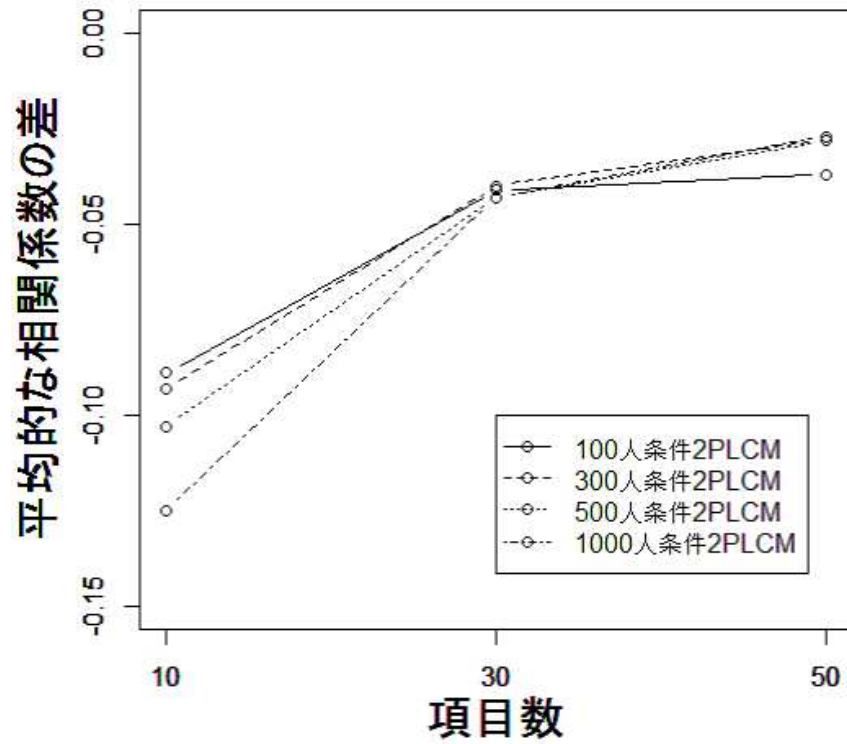


図 2.12 発生モデルが 2PLCM である場合の $\Delta cor(\theta, \hat{\theta})$ の項目数による変化

第3章

局所独立性を仮定した項目特性の推定に与える影響

3.1 問題と目的

本章では、局所依存性が局所独立性を仮定した項目特性の推定に対してどのような影響を与えるのか、シミュレーションを通して検討を行い、得られた結果に対して解析的な観点からも検討を加える*1。このテーマに関しても既にいくつかの先行研究において検討が加えられており、例えば、Bradlow et al. (1999) では、2 値型の BTM を発生モデルとし、2PLM と BTM を分析モデルとしてシミュレーションを行った結果、

- 識別力、困難度ともに、分析モデルが 2PLM である場合のバイアスの値が分析モデルが BTM である場合のバイアスよりも大きくなる
- 分析モデルが 2PLM である場合の方が BTM である場合よりも、識別力に関する M95%PIW の値が小さくなっている
- 識別力、困難度ともに、分析モデルが 2PLM である場合の方が、BTM である場合よりも、M95%PIW が実際に真値を含む確率が小さくなっている
- 識別力、困難度ともに、分析モデルが 2PLM である場合の方が、BTM である場合よりも、推定値と真値との順位相関係数が小さくなっている
- 局所依存関係にある項目に対し 2PLM をあてはめて識別力の推定を行うと、当該項目の識別力を過小推定する傾向にある

といったことが報告されている。また、Chen & Wang (2007) では、3PCCM を発生モデルとし、3PLM を分析モデルとしてシミュレーションを行った結果、項目反応間に正の依

*1 本章は、登藤 (2012b) を加筆、再構成したものである。

存関係がある場合、局所依存性を有するデータに対し局所独立性を仮定して項目特性の推定を行うと、

- 識別力が過大推定される
- 困難度が過小推定される
- 識別力、困難度ともに、推定量の平均二乗誤差平方根が大きくなる

ことが報告されている。さらに、Jiao Kamatani, Wang, & Jin (2012) においては、受験者間での局所依存性と項目反応間の局所依存性、2つの局所依存性を有するデータを発生させ、これに対し局所独立性を仮定するラッシュモデル (Rasch, 1960) をあてはめて困難度の推定を行ったところ、局所依存性を考慮した場合に比べて、推定量の平均二乗誤差平方根が増加したことが報告されている。また、Looney & Spray (1992) では、同じ項目を何度も繰り返すテストを想定したシミュレーションを実施したところ、局所独立性の仮定が満たされないデータに対し局所独立性を仮定して困難度の推定を行った場合に、困難度の推定量にバイアスが生じてしまうことが報告されており、Tuerlinckx & De Boeck (2001) においては、発生モデルを CCM とし、分析モデルを 2PLM としてシミュレーションを実施した結果、項目反応間に正の依存関係がある場合に、局所依存性を有するデータに対し局所独立性を仮定して識別力の推定を行うと、識別力が過大推定されることが報告されている。加えて、Wainer & Wang (2000) では、局所依存性を有すると考えられる TOEFL のデータに対して、3PLM と 3PBTM をあてはめて項目特性の推定を行ったところ、局所独立性を仮定した場合には、識別力の推定値にバイアスが生じることが報告されており、山野井・山田 (2012) では、実験的独立性の満たされないデータに対し 2PLM と GRM をあてはめて識別力の推定を行った結果、分析モデルとして 2PLM を使用した場合のバイアスの値が分析モデルが GRM である場合のバイアスに比べて正の方向に大きくなることが報告されている。

以上のように、先行研究においては、局所依存性が局所独立性を仮定した項目特性の推定に及ぼす影響について、主にシミュレーションを通じた検討が行われてきたのだが、これらの研究においては、局所依存性が生じているデータの発生モデルとこれを分析する際に用いるモデルとの間で、モデルに含まれる項目母数の意味するものが異なっており、その結果、これらの研究で得られた知見には本来比較不可能なものを比較した影響も含まれていると考えられる (Braeken, 2011; Braeken, et al., 2007; Ip, 2010; Ip, Smits, & De Boeck, 2009)。例えば、Bradlow et al. (1999) においては、発生モデルとして 2 値型の BTM が使用され、分析モデルとして 2PLM と BTM が使用されているのだが、1.5 節でも述べたように、BTM と 2PLM とではモデルに含まれる項目母数の意味するところが異なっており、この研究で得られた知見には、本来比較不可能なものを比較したことによる影響が含まれていると考えられる。同様に、Chen & Wang (2007) や Tuerlinckx & De

Boeck (2001) においても、発生モデルとして CCM 系のモデルを使用し、分析モデルとして 2PLM や 3PLM を使用しているのだが、1.2 節および 1.5 節より、これらのモデル間では、モデルに含まれる項目母数の意味するところが互いに異なっており、これらの研究で得られた知見に関して、本来比較不可能なものを比較したことによる影響が含まれていると考えられる。したがって、どのような条件の下であれば局所独立性を仮定した項目特性の推定が局所依存性に対して頑健であるのか、比較不可能なものを比較したことによる影響を取り除いて評価するためには、発生モデルと分析モデルの間で項目母数の比較可能性を確保した上でシミュレーションを実施する必要がある。これに関連して、Ip (2010) では、2 値型の BTM の項目母数を 2PLM の項目母数と比較可能なものに変換するための方法が提案されており、この方法を用いることにより、項目母数の比較可能性を確保した上でシミュレーションを実施することが可能となる。また、本研究で得られた知見を実際にテストを作成、評価、運用する場面へ応用するという観点からは、受験者数をはじめとする種々の要因を系統的に変化させてシミュレーションを行い、各要因が局所独立性を仮定した分析の頑健性にどのような影響を与えるのか、検討する必要があると考えられる。

そこで、本章では、発生モデルとして 2 値型の BTM を使用し、分析モデルとして 2PLM を使用して、Ip (2010) の方法により項目母数の比較可能性を確保した上でシミュレーションを行い、局所依存性が局所独立性を仮定した項目特性の推定に対してどのような影響を与えるのか、検討を行った。また、シミュレーションを実施する際に受験者数をはじめとする種々の要因を系統的に変化させ、各要因が局所独立性を仮定した分析の頑健性にどのような影響を与えるのか、検討を行った。

3.2 方法

本章で行うシミュレーション実験では、局所依存性が生じている仮想的な項目反応データに対し局所独立性を仮定するモデルと局所依存性を考慮したモデルをあてはめて母数の推定を行い、得られた項目母数の推定値と真値から各項目母数の推定量のバイアスと平均二乗誤差平方根、および、各項目母数を要素とするベクトルの推定量と真値の相関係数を算出して分析モデル間で比較を行った。

3.2.1 仮想的な項目反応データの発生モデル

局所依存性が生じている仮想的な項目反応データを発生させる際には、3.1 節でも述べたように、(1.14) 式の 2 値型の BTM を使用した (以下にモデル式を再掲)。

$$P_{j \in d(j)}(\theta_i) = \frac{1}{1 + \exp[-1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)} - \gamma_{ad(j)})]} \quad (3.1)$$

3.2.2 仮想的な項目反応データの分析モデル

局所依存性が生じている仮想的な項目反応データにあてはめるモデルとして以下の2つのモデルを使用することとした。

局所独立性を仮定する分析モデル

局所依存性が生じている仮想的な項目反応データにあてはめる局所独立性を仮定した分析モデルとしては、3.1節でも述べたように、(1.1)式の2PLMを使用した(以下にモデル式を再掲)。

$$P_j(\theta_i) = \frac{1}{1 + \exp[-1.7a_j(\theta_i - b_j)]} \quad (3.2)$$

局所依存性を考慮した分析モデル

局所依存性が生じている仮想的な項目反応データにあてはめる局所依存性を考慮した分析モデルとしては仮想的な項目反応データの発生モデルでもある2値型のBTMを使用することとした(モデル式については(3.1)式を参照のこと)。

3.2.3 2値型のBTMに含まれる項目母数の変換方法

1.5節でも述べたように、(3.2)式中の a_j は $\theta_i = b_j$ における(3.2)式の傾きに比例する値となっており、(3.2)式中の b_j は $P_j(\theta_i) = 0.5$ となる θ_i の値を表している。一方、(3.1)式中の $a_{j \in d(j)}$ は $\gamma_{ad(j)}$ が0であるときの $\theta_i = b_{j \in d(j)}$ における(3.1)式の傾きに比例する値となっており、 $b_{j \in d(j)}$ は $\gamma_{ad(j)}$ が0であるときに $P_{j \in d(j)}(\theta_i) = 0.5$ となる θ_i の値を表している。このように、2PLMに含まれる項目母数と2値型のBTMに含まれる項目母数とではその意味するところが互いに異なっており、このままではこれらの母数を互いに直接比較することはできない。

3.1節で紹介したIp(2010)では、(3.1)式から $\gamma_{ad(j)}$ を積分消去することによって2値型のBTMに含まれる項目母数と2PLMの項目母数が比較可能になると考え、(3.1)式より2値型のBTMを $\gamma_{ad(j)}$ について周辺化した周辺化テストレット項目反応関数(marginalized testlet item response function, MIRF)が導出されており、このMIRFが

らは、(3.1) 式中の $a_{j \in d(j)}$ を

$$a_j^* = \tau a_{j \in d(j)} \quad (3.3)$$

$$\tau = \frac{1}{\sqrt{\kappa^2 (1.7 a_{j \in d(j)})^2 \sigma_{\gamma_{d(j)}}^2 + 1}} \quad (3.4)$$

$$\kappa = \frac{16\sqrt{3}}{15\pi} \quad (3.5)$$

$$b_j^* = b_{j \in d(j)} \quad (3.6)$$

のように変換することで、2 値型の BTM の項目母数が 2PLM の項目母数と比較可能な項目母数 a_j^*, b_j^* となることが明らかとなっている*2。

本章で行うシミュレーション実験では、3.1 節でも述べたように、(3.3) 式～(3.6) 式を利用して $a_{j \in d(j)}, b_{j \in d(j)}$ を a_j^*, b_j^* に変換した上でバイアス、平均二乗誤差平方根、および、推定量と真値との相関係数を算出し、モデル間で比較を行った。

3.2.4 シミュレーション実験の手続き

本章では、データを発生させる際に、テストは局所依存関係にある 4 つの項目群から成るものであるとし、各項目群は 5 つの項目から構成されていると仮定した。また、3.1 節でも述べたように、本研究では、局所依存性の影響と各種要因との関係についても検討を加えるため、シミュレーションを実施する際に、受験者数、局所依存関係にある項目の数、4 つの項目群の局所依存度に関しても操作することとした。なお、受験者数に関しては、300 人、1000 人という 2 つの水準を設定して検討を行い、局所依存関係にある項目の数に関しては、各項目群内で局所依存関係にある項目の数に関して以下の 3 つの水準を設定し、検討を行った。

- 5 項目が局所依存関係にある ($J_{d(j)} = 5$)
- 3 項目が局所依存関係にある ($J_{d(j)} = 3$)
- 2 項目が局所依存関係にある ($J_{d(j)} = 2$)

また、4 つの項目群の局所依存度に関しては、4 つの項目群の局所依存度について以下の 5 つの水準を設定し、検討を行った。

- 4 つとも局所依存度が強い (強 $d(j) \times 4$)
- 4 つのうち 3 つの局所依存度が強い (強 $d(j) \times 3$)
- 4 つのうち 2 つの局所依存度が強い (強 $d(j) \times 2$)

*2 MIRF の導出過程など、詳細については Ip (2010) を参照のこと

- 4つのうち1つだけ局所依存度が強い (強 $d(j) \times 1$)
- 4つとも局所依存度が弱い (強 $d(j) \times 0$)

なお、項目間の局所依存度の強弱に関しては、データの発生モデルである2値型のBTMを提案したBradlow et al. (1999) およびこれを引用している論文のうち2値型のBTMを局所依存性を有すると考えられる実際の回答データに対して適用している3件の論文 (Bradlow et al., 1999; Li, Bolt, & Fu, 2005; Li, Bolt, & Fu, 2006) を参考にして決定した。具体的には、この3本の論文の中で報告されている17個の $\sigma_{\gamma_{d(j)}}^2$ の推定値 $\hat{\sigma}_{\gamma_{d(j)}}^2$ のうち上位4個の平均値 (1.4075) を局所依存度が強い場合の $\sigma_{\gamma_{d(j)}}^2$ の値とし、下位4個の平均値 (0.155) を局所依存度が弱い場合の $\sigma_{\gamma_{d(j)}}^2$ の値として、この基準に従い、局所依存性を有するデータを発生させることとした。

本研究では受験者数 (2水準) \times 局所依存関係にある項目の割合 (3水準) \times 項目間の局所依存度 (5水準) の計30条件の下でシミュレーション実験を行い、その詳細は以下のようまとめられる*3。

1. 各受験者の受験者特性値 θ_i ($i = 1, 2, \dots, N$) と $\gamma_{ad(j)}$ ($a = 1, 2, \dots, N; d(j) = 1, 2, 3, 4$) および各項目の識別力 $a_{j \in d(j)}$ と困難度 $b_{j \in d(j)}$ ($j = 1, 2, \dots, 20$) の真値をそれぞれ標準正規分布 $N(0, 1)$, 正規分布 $N(0, \sigma_{\gamma_{a(j)}}^2)$, 一様分布 $U(0.5, 1.5)$, 標準正規分布 $N(0, 1)$ から発生させる*4。
2. 各受験者の各項目に対する正答確率を手続き1で発生させた各母数の真値と (3.1) 式から算出し、 $N \times 20$ の正答確率行列 \mathbf{A} を作成する。
3. 一様分布 $U(0, 1)$ から乱数を $N \times 20$ 個発生させて $N \times 20$ の一様乱数行列 \mathbf{B} を作成する。
4. \mathbf{A}, \mathbf{B} の各要素を比較して局所依存性が生じている仮想的な項目反応データの行列 \mathbf{U} を作成する。具体的には、 $a_{ij} \geq b_{ij}$ である場合には $u_{ij} = 1$ とし、 $a_{ij} < b_{ij}$ である場合には $u_{ij} = 0$ とする。
5. \mathbf{U} に対して3.2.2節で紹介した2タイプの分析モデルを当てはめて母数の推定を行う。
6. 3から5までの手続きを R 回繰り返す。
7. 3.2.3節にて解説したIp (2010) の方法を利用し、発生モデルに含まれる項目母数と2値型のBTMをあてはめた場合に得られる項目母数の推定値を2PLMの項目母数と比較可能な形に変換する。
8. 7の結果得られた R 組の比較可能な推定値と真値から分析モデル毎に以下の指標

*3 実験用プログラムの記述および実行には統計解析向けプログラミング言語でありその実行環境でもあるRを利用した。

*4 なお、擬似乱数を発生させる際にはRに実装されているメルセンヌツイスターを使用した。

値を算出する.

$$Bias(\hat{\lambda}_j) = \frac{1}{R} \sum_{r=1}^R \hat{\lambda}_{jr} - \lambda_j \quad (3.7)$$

$$RMSE(\hat{\lambda}_j) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\lambda}_{jr} - \lambda_j)^2} \quad (3.8)$$

$$cor(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}) = \frac{1}{R} \sum_{r=1}^R cor(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}_r) \quad (3.9)$$

9. 8にて得られる20項目分の $Bias(\hat{\lambda}_j)$ と $RMSE(\hat{\lambda}_j)$ のうち, 実際に局所依存関係にある項目の $Bias(\hat{\lambda}_j)$ と $RMSE(\hat{\lambda}_j)$ の値を平均してそれぞれ \overline{Bias} , \overline{RMSE} とする.

ここで, λ_j は項目 j に関する項目母数の真値を表しており, $\hat{\lambda}_j$ は λ_j の推定量を, $\hat{\lambda}_{jr}$ は $\hat{\lambda}_j$ の r 回目の推定値をそれぞれ表わしている. また, $\boldsymbol{\lambda}$ はテストに含まれる20項目のうち実際に局所依存の関係にある項目の a_j^* あるいは b_j^* をその要素とする項目母数の真値ベクトルを表しており, $\hat{\boldsymbol{\lambda}}$ はその推定量を, $\hat{\boldsymbol{\lambda}}_r$ はその r 回目の推定値をそれぞれ表わしている.

なお, 手続き6における反復回数 R に関しては, 繰り返し毎の $\hat{\lambda}_{ir}$ の変動を考慮して, 本研究においては, 受験者数が1000人の場合には $R = 50$ とし受験者数が300人の場合には $R = 100$ とした.

3.2.5 MCMCによる母数の推定

本研究でもシミュレーション実験において母数の推定を行う際にMCMCを利用した(MCMCの詳細については2.2.3節を参照のこと). 実際にMCMCを用いて母数の推定を行う際にはWinBUGS 1.4 (Spiegelhalter, Thomas, & Best, 2003)を利用し, MCMCの具体的なアルゴリズムとしてはスライスサンプリング (slice sampling) を利用することとした (Neal, 1997). また, 実際に母数を推定するにあたっては各母数が互いに独立であると仮定し, 分析モデル毎に各母数の事前分布を以下のように設定した.

- 2PLMの場合
 - θ_i の事前分布: 標準正規分布 $N(0, 1)$
 - a_j の事前分布: 正規分布 $N(1, 0.25)$
 - b_j の事前分布: 標準正規分布 $N(0, 1)$
- 2値型のBTMの場合

- θ_i の事前分布：標準正規分布 $N(0, 1)$
- $a_{j \in d(j)}$ の事前分布：正規分布 $N(1, 0.25)$
- $b_{j \in d(j)}$ の事前分布：標準正規分布 $N(0, 1)$
- $\gamma_{ad(j)}$ の事前分布：正規分布 $N(0, \sigma_{\gamma_{d(j)}}^2)$
- $\sigma_{\gamma_{d(j)}}^{-2}$ ：ガンマ分布 $\Gamma(3, 1)$

また、本研究でも時系列プロットを利用してMCMCにより得られた標本の初期値からの影響の有無を確認し、事前に行ったシミュレーションの結果から、分析モデル毎に burn-in を以下のように定めた。

- 2PLM の場合
全ての条件において最初の 1000 組の実現値を burn-in とした。
- 2 値型の BTM の場合
全ての条件において最初の 1000 組の実現値を burn-in とした。

さらに、MCMC により実際に抽出する標本の数に関しては、WinBUGS User Manual (Spiegelhalter, Thomas, Best, & Lunn, 2003) に従い、各母数の標本平均の標準誤差が事後分布から抽出された標本の標準偏差の約 5% 以下になるよう以下のように定めた。

- 2PLM の場合
全ての条件において 4000 個の標本を抽出することとした。
- 2 値型の BTM の場合
全ての条件において 5000 個の標本を抽出することとした。

また、MCMC を行う際に使用する各母数の初期値に関しては、

- θ_i の初期値：0
- $a_j, a_{j \in d(j)}$ の初期値：1
- $b_j, b_{j \in d(j)}$ の初期値：0
- $\gamma_{ad(j)}$ の初期値：0
- $\sigma_{\gamma_{d(j)}}^{-2}$ の初期値：3

とした。

3.3 結果と考察

3.3.1 識別力の推定に対する局所依存性の影響

まず、局所依存性が局所独立性を仮定した識別力の推定に対しどのような影響を与えるのか検討を行うため、項目母数の比較可能性を確保した上で、分析モデル毎に、各条件における \overline{Bias} , \overline{RMSE} , $cor(\mathbf{a}, \hat{\mathbf{a}})$ の値を算出し、比較を行った。

なお、分析モデルが 2PLM である場合と 2 値型の BTM である場合との間で \overline{Bias} や \overline{RMSE} の値について比較を行う場合には、その差異の大きさを母集団分布における a_j^* の標準偏差に照らして評価することが必要となってくる。本研究では、3.2.4 節でも述べたように、 $a_{j \in d(j)}$ の母集団分として $U(0.5, 1.5)$ を仮定しているのだが、 a_j^* はこの $a_{j \in d(j)}$ を 3.2.3 節の (3.3) 式から (3.5) 式を用いて変換したものであるため、 a_j^* の母集団分布は条件毎に異なってくる。そこで、本研究においては、便宜上 a_j^* の母集団分布が $a_{j \in d(j)}$ の母集団分布 $U(0.5, 1.5)$ であるとし、母集団分布における a_j^* の標準偏差を 0.289 であると仮定する。したがって、 \overline{Bias} と \overline{RMSE} の分析モデル間での差異については、この値を 0.289 で除したものを、 a_j^* の母集団分布における標準偏差に対する相対的な大きさであるとして評価することができる。

分析モデルが 2 値型の BTM の場合

まず、分析モデルが 2 値型の BTM である場合に \overline{Bias} , \overline{RMSE} , $cor(\mathbf{a}, \hat{\mathbf{a}})$ の値がどのようになるのか、条件毎に算出したところ、表 3.1 が得られた。

分析モデルが 2PLM の場合

次に、分析モデルが 2PLM の場合に \overline{Bias} , \overline{RMSE} , $cor(\mathbf{a}, \hat{\mathbf{a}})$ の値がどのようになるのか、条件毎に算出したところ、表 3.2 が得られた。また、分析モデルが 2PLM である場合に、分析モデルが 2 値型の BTM である場合との間で \overline{Bias} , \overline{RMSE} , $cor(\mathbf{a}, \hat{\mathbf{a}})$ の値がどの程度異なるのか、条件毎に算出したところ、表 3.3 が得られた。ここで、 $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ はそれぞれ、分析モデルが 2PLM である場合の \overline{Bias} , \overline{RMSE} , $cor(\mathbf{a}, \hat{\mathbf{a}})$ から分析モデルが 2 値型の BTM である場合の値を引いたものとなっている。さらに、分析モデルが 2PLM である場合に、 $\Delta\overline{Bias}$ および $\Delta\overline{RMSE}$ の値が a_j^* の母標準偏差に比べてどの程度の大きさとなるのか、条件毎に算出したところ、表 3.4 が得られた。ここで、 $\Delta\overline{Bias}/\sigma_{a_j^*}$ および $\Delta\overline{RMSE}/\sigma_{a_j^*}$ は、それぞれ、 $\Delta\overline{Bias}$ および $\Delta\overline{RMSE}$ の値を a_j^* の母標準偏差で除した値を表している。加えて、分析モデルが 2PLM である場合に $\Delta\overline{Bias}/\sigma_{a_j^*}$, $\Delta\overline{RMSE}/\sigma_{a_j^*}$, $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が受験者数によりどのように

変化するのか確認するためグラフを作成したところ、図 3.1, 図 3.2, 図 3.3 が得られた。続いて、分析モデルが 2PLM である場合に $\overline{\Delta Bias}/\sigma_{a_j^*}$, $\overline{\Delta RMSE}/\sigma_{a_j^*}$, $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が局所依存関係にある項目の数によりどのように変化するのか確認するためグラフを作成したところ、図 3.4, 図 3.5, 図 3.6 が得られた。最後に、分析モデルが 2PLM である場合に $\overline{\Delta Bias}/\sigma_{a_j^*}$, $\overline{\Delta RMSE}/\sigma_{a_j^*}$, $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が 4 つの項目群の局所依存度によりどのように変化するのか確認するためグラフを作成したところ、図 3.7, 図 3.8, 図 3.9 が得られた。

これらの結果から、分析モデルが 2PLM である場合の識別力に関するバイアス、平均二乗誤差平方根、真値と推定値との相関係数について、以下のようなことが明らかとなった。

バイアスの観点からの評価

表 3.4 より、全ての条件において、分析モデルが 2PLM である場合の \overline{Bias} が $\sigma_{a_j^*}$ の 0.1 以上 ($a_{j^*}^l$ の母標準偏差の 10% 以上)、分析モデルが 2 値型の BTM である場合の \overline{Bias} の値よりも正の方向に大きくなっていった。また、図 3.4 からは、全ての項目群の局所依存度が弱い条件を除いて、局所依存関係にある項目の数が増加するにつれて、 $\overline{\Delta Bias}/\sigma_{a_j^*}$ の値が増加していく様子が見える。これは、局所依存関係にある項目の数が増加するにつれて、局所独立性を仮定した 2PLM の \overline{Bias} に関してはその値が増加していく一方、局所依存性を考慮した 2 値型の BTM においては、そのような増加傾向が見られなかったことに起因する。

平均二乗誤差平方根の観点からの評価

表 3.4 より、全ての条件において、分析モデルが 2PLM である場合の \overline{RMSE} の値が分析モデルが BTM である場合の \overline{RMSE} の値よりも大きくなっており、特に、 $N = 300$ と受験者数が少ない場合には、その差が $\sigma_{a_j^*}$ の 0.1 以上 (a_j^* の母標準偏差の 10% 以上) となっていた。また、図 3.2 からは、受験者数が増加するのに伴って、 $\overline{\Delta RMSE}/\sigma_{a_j^*}$ の値が減少していく様子が見える。これは、受験者数の増加に伴って 2PLM の頑健性が向上し、2PLM の \overline{RMSE} の値が減少する一方、局所依存性を考慮する 2 値型の BTM においては、2PLM ほど、受験者数の増加に伴う減少が見られなかったためだと考えられる。さらに、図 3.5 からは、全ての項目群の局所依存度が弱い条件を除いて、局所依存関係にある項目の数が増加するのに伴い、 $\overline{\Delta RMSE}/\sigma_{a_j^*}$ の値が増加する様子が確認できる。これは、局所依存関係にある項目の数が増加するにつれて 2PLM の \overline{RMSE} の値が増加する一方、BTM においてはそのような変化が見られなかったためだと考えられる。

真値と推定値との相関係数の観点からの評価

表 3.3 より, 全ての条件において, 分析モデルが 2PLM である場合の $cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が分析モデルが BTM である場合の値以下となっており, 特に, 局所依存関係にある項目の数が 5 項目で, かつ, 各項目群の局所依存度が異なる場合には, $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が -0.15 よりも小さくなっていった. また, 図 3.6 からは, 各項目群の局所依存度が異なる場合に, 局所依存関係にある項目の数が増加するにつれて, $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が減少していく様子が見える. これは, 各項目群の局所依存度が異なる場合には, 項目群毎に項目母数の推定値にかかるバイアスが異なってくるため, 局所依存関係にある項目の数が増加するほど, 2PLM の $cor(\mathbf{a}, \hat{\mathbf{a}})$ が減少する一方, BTM においては, そのような傾向は見られなかったためであると考えられる. さらに, 図 3.9 からは, 局所依存関係にある項目の数が 5 項目である場合に, 全ての項目群の局所依存度が等しい場合には, $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が 0 に近くなる一方, 各項目群の局所依存度が異なる場合には, $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が大きく低下している様子が確認できる. これは, 4 つの項目群の局所依存度が等しい場合には, 全ての項目母数の推定値に同程度のバイアスが生じるため, 2PLM を使用した場合にも, 識別力母数の線形性が BTM と同程度に保たれるものの, 局所依存度が 4 つの項目群の間で異なる場合には, 項目群毎に項目母数にかかるバイアスの程度が異なってくるため, 2PLM を使用した場合に識別力母数の線形性が保たれなくなり, 局所依存関係にある項目が 5 項目である場合には, その程度が他の条件に比べて高くなるためだと考えられる.

まとめ

特定のテスト内において項目特性の評価を行うという文脈においては, 識別力の真値と推定値との相関係数という観点から評価を行うことが重要となってくる. そういった観点から上記の結果を見直してみると, 特に, 局所依存関係にある項目のテスト全体に占める割合が高く, 各項目群の局所依存度が異なるような場合には, 局所依存性の影響が無視できないほど大きくなり, 局所独立性を仮定したモデルは頑健であるとはいえない.

また, 特定のテストを超えて, 各項目の項目特性を共通の尺度上で比較するという文脈においては, バイアスおよび平均二乗誤差平方根という観点からの評価も重要になってくる. そういった観点から上記の結果を見直してみると, バイアス, 平均二乗誤差平方根, どちらの統計量に関しても, 局所依存性の影響は無視できないほど大きくなっており, 局所独立性を仮定した項目特性の推定は, 局所依存性に対し頑健ではないといえる.

項目母数の比較可能性を確保しなかった場合

最後に, 分析モデルが 2PLM である場合に, 項目母数の比較可能性を確保せず発生モデルの $a_{j \in d(j)}$ と推定された a_j の値を直接比較すると, $\overline{Bias}, \overline{RMSE}, cor(\mathbf{a}, \hat{\mathbf{a}})$ の値はど

のようになるのか、各条件において算出したところ、表 3.5 が得られた。また、項目母数の比較可能性を確保しなかった場合に、 $\Delta \overline{Bias}$, $\Delta \overline{RMSE}$, $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値がどのようになるのか、各条件において算出したところ、表 3.6 が得られた。さらに、項目母数の比較可能性を確保しなかった場合に $\Delta \overline{Bias}/\sigma_{a_j^*}$ と $\Delta \overline{RMSE}/\sigma_{a_j^*}$ の値がどのようになるのか、各条件において算出したところ、表 3.7 が得られた。

これらの結果から、項目母数の比較可能性を確保しなかった場合の識別力に関するバイアス、平均二乗誤差平方根、真値と推定値との相関係数について、以下のようなことが明らかとなった。

バイアスの観点からの評価

表 3.7 より、項目母数の比較可能性を確保せずに局所依存性の影響を評価すると、局所依存性の影響が無視できないと考えられる条件においては、分析モデルとして 2PLM を使用したときの \overline{Bias} の値が分析モデルが BTM の場合の \overline{Bias} よりも小さくなっていると判断することになる。このことから、項目母数の比較可能性を確保せずに局所依存性の影響について評価を行うと、項目母数の比較可能性を確保した上での評価とは逆の方向に推定量のバイアスを評価してしまうことになると考えられる。

平均二乗誤差平方根の観点からの評価

表 3.7 より、項目母数の比較可能性を確保せずに局所依存性の影響を評価すると、全ての条件において、分析モデルが 2PLM であるときの \overline{RMSE} の値が分析モデルが BTM であるときの \overline{RMSE} よりも大きくなっており、特に、 $N = 300$ と受験者数が少ない場合には、その差が $\sigma_{a_j^*}$ の 0.1 以上 (a_j^* の母標準偏差の 10% 以上) となっていると判断することになる。このことから、項目母数の比較可能性を確保せずに局所依存性の影響を評価したとしても、平均二乗誤差平方根に関しては、比較可能性を確保した場合と同様の評価を下すことができると考えられる。

真値と推定値との相関係数の観点からの評価

表 3.6 より、項目母数の比較可能性を確保せずに局所依存性の影響を評価すると、局所依存性の影響が無視できないほど大きい条件においては、 $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が負になっており、特に、局所依存度の強い項目群の数が 3 や 2 である場合には、 $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ の値が 0.1 以上小さくなっていると判断することになる。このことから、項目母数の比較可能性を確保せずに局所依存性に関する評価を行うと、局所依存性の影響が強くなっている条件について、項目母数の比較可能性を確保した場合とは異なる評価を下すことになると考えられる。

まとめ

項目母数の比較可能性を確保しなかった場合には、バイアスや真値と推定値との相関係数という観点からの評価が、項目母数の比較可能性を確保した場合のものと異なるものとなっていた。したがって、特定のテスト内において項目特性の比較を行うにしても、特定のテストを超えて、項目特性を共通の尺度上で評価するにしても、先行研究で得られた知見を利用するためには見直しが必要であろうと考えられる。

3.3.2 困難度の推定精度に関する結果

続いて、局所依存性が局所独立性を仮定した困難度の推定に対しどのような影響を与えるのか検討を行うため、項目母数の比較可能性を確保した上で、分析モデル毎に、各条件における \overline{Bias} , \overline{RMSE} , $cor(\mathbf{b}, \hat{\mathbf{b}})$ の値を算出し、比較を行った。

なお、分析モデルが 2PLM である場合と 2 値型の BTM である場合との間で \overline{Bias} や \overline{RMSE} の値に関して比較を行う際には、その差異の大きさを母集団分布における b_j^* の標準偏差に照らして評価することが必要となってくる。本研究では、3.2.4 節でも述べたように、 $b_{j \in d(j)}$ の母集団分布として $N(0, 1)$ を仮定しており、3.2.3 節の (3.6) 式より、 b_j^* の母集団分布も $N(0, 1)$ となる。したがって、 \overline{Bias} , \overline{RMSE} の分析モデル間での差異に関しては、この値を、そのまま b_j^* の母標準偏差に対する相対的な大きさとして評価できることになる。

分析モデルが 2 値型の BTM の場合

まず、分析モデルが 2 値型の BTM である場合に \overline{Bias} , \overline{RMSE} , $cor(\mathbf{b}, \hat{\mathbf{b}})$ の値がどのようになるのか、条件毎に算出したところ、表 3.8 が得られた。

分析モデルが 2PLM の場合

次に、分析モデルが 2PLM である場合に \overline{Bias} , \overline{RMSE} , $cor(\mathbf{b}, \hat{\mathbf{b}})$ の値がどのようになるのか、条件毎に算出したところ、表 3.9 が得られた。また、分析モデルが 2PLM である場合に、 $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\mathbf{b}, \hat{\mathbf{b}})$ の値がどのようになるのか、条件毎に算出したところ、表 3.10 が得られた。加えて、分析モデルが 2PLM である場合に、 $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\mathbf{b}, \hat{\mathbf{b}})$ の値が受験者数によりどのように変化するか確認するためグラフを作成したところ、図 3.10, 図 3.11, 図 3.12 が得られた。続いて、分析モデルが 2PLM である場合に、 $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\mathbf{b}, \hat{\mathbf{b}})$ の値が局所依存関係にある項目の数によりどのように変化するか確認するためグラフを作成したところ、図 3.13, 図 3.14, 図 3.15 が得られた。最後に、分析モデルが 2PLM である場合に、

$\Delta \overline{Bias}$, $\Delta \overline{RMSE}$, $\Delta \text{cor}(\mathbf{b}, \hat{\mathbf{b}})$ の値が4つの項目群の局所依存度によりどのように変化するのか確認するためグラフを作成したところ、図 3.16, 図 3.17, 図 3.18 が得られた。

これらの結果から、分析モデルが 2PLM である場合の困難度に関するバイアス、平均二乗誤差平方根、真値と推定値との相関係数について、以下のようなことが明らかとなった。

バイアスの観点からの評価

表 3.10 より、全ての条件において、 \overline{Bias} の値に関して、分析モデルが 2PLM である場合と 2 値型の BTM である場合との間にほとんど差は見られなかった。

平均二乗誤差平方根の観点からの評価

表 3.10 より、ほとんどの条件において、分析モデルが 2PLM と 2 値型の BTM である場合との間で、 \overline{RMSE} の値に差は見られなかったが、 $J_{d(j)} = 5$, 強 $d(j) \times 4, 3, 2$ である条件に限り、分析モデルが 2PLM である場合と 2 値型の BTM である場合との間で、 \overline{RMSE} の値に 0.05 程度 (b_j^* の母標準偏差の 5% 程度) の差が見られた。また、図 3.14 からは、局所依存関係にある項目の数が増えるほど、 $\Delta \overline{RMSE}$ の値が増加する傾向がうかがえる。これは、局所依存関係にある項目の数が増えることにより、2PLM の \overline{RMSE} の値が増加する一方、2 値型の BTM においてはそのような傾向が見られなかったためだと考えられる。さらに、図 3.17 からは、局所依存度の強い項目群の増加に伴って、 $\Delta \overline{RMSE}$ の値が増加していく傾向がみられる。これは、局所依存度の強い項目群が増加することにより 2PLM の \overline{RMSE} の値が増加する一方で、BTM においてはそのような傾向がみられなかったためだと考えられる。

真値と推定値との相関係数の観点からの評価

表 3.10 より、全ての条件において、 $\text{cor}(\mathbf{b}, \hat{\mathbf{b}})$ の値に関し、分析モデルが 2PLM である場合と BTM である場合との間でほとんど差は見られなかった。また、図 3.18 からは、局所依存関係にある項目が 5 項目である場合に、全ての項目群の局所依存度が等しい場合には $\Delta \text{cor}(\mathbf{b}, \hat{\mathbf{b}})$ の値が 0 に近くなる一方、各項目群の局所依存度が異なる場合には、 $\Delta \text{cor}(\mathbf{b}, \hat{\mathbf{b}})$ の値が小さくなる様子がうかがえた。これは、4つの項目群の局所依存度が全て等しい場合には、全ての項目特性の推定量に対し同程度のバイアスが生じるため、2PLM を使用した場合にも、困難度母数の線形性が BTM と同程度に保たれるものの、局所依存度が項目群の間で異なる場合には、項目群毎に推定量にかかるバイアスの程度が異なってくるため、困難度母数の線形性が保たれなくなり、局所依存関係にある項目が 5 項目である場合には、その程度が他の条件に比べて高くなるためだと考えられる。

まとめ

特定のテスト内で項目特性の評価を行うという文脈においては、困難度の真値と推定値との相関係数という観点から評価を行うことが重要となってくる。そういった観点から上記の結果を見直してみると、局所独立性を仮定した困難度の推定に対し局所依存性の影響はほとんど見られず、局所独立性を仮定するモデルは局所依存性に対して頑健であると言える。

また、特定のテストを超えて、各項目の困難度を共通の尺度上で比較するという文脈においては、バイアスおよび平均二乗誤差平方根という観点からの評価も重要になってくる。そういった観点から上記の結果を見直してみると、 $J_{d(j)} = 5$ 、強 $d(j) \times 4, 3, 2$ のように、局所依存関係にある項目のテスト全体に占める割合が高く、局所依存度の高い項目の占める割合も多い場合には、局所依存性の影響が無視できないものとなっており、このような条件においては、局所独立性を仮定するモデルは頑健であるとは言えなかった。

項目母数の比較可能性を確保しなかった場合

分析モデルが 2PLM の場合、3.2.3 節の (3.6) 式より、発生モデルの $b_{j \in d(j)}$ と推定された b_j の値を直接比較した結果は b_j^* と b_j の値を比較した結果と等しくなる。したがって、本研究では、困難度推定に対する局所依存性の影響と項目母数の比較可能性との関係に関する検討は行わない。

3.3.3 局所依存性が局所独立性を仮定した項目母数の推定に影響を与えるメカニズム

本研究においては、特に局所独立性を仮定して識別力を推定した際に、一貫した過大推定の傾向と推定誤差の増大が確認された。Lord & Novick (1968) 等に示されているように、2PLM に含まれる a_j は、一定の仮定の下、項目 j への項目反応 u_j と和得点 s の間の双列相関係数 r_{bj} の関数として

$$a_j \cong \frac{r_{bj}}{\sqrt{1 - r_{bj}^2}} \quad (3.10)$$

のように表すことができる。また、データの発生モデルである 2 値型の BTM には θ_i に加えて $d(j)$ における局所依存性を表現するために受験者母数 $\gamma_{id(j)}$ が含まれており、この $\gamma_{id(j)}$ の影響で局所依存関係にある項目間の四分相関係数は θ_i のみに基づくものよりも強いものとなっている。したがって、2PLM をあてはめて母数の推定を行った場合には、局所依存関係にある項目間の四分相関係数は θ_i のみに基づくものとして過大推定されることがとなり、同様に、他の項目と局所依存関係にある項目の双列相関係数 r_{bj} も過大推定さ

れることとなる。ここで、(3.10) 式より、 a_j が r_{bj} の単調増加関数となっていることを考慮すると、2 値型の BTM から発生させたデータに対し 2PLM をあてはめて識別力の推定を行うと、推定量 \hat{a}_j は真の識別力を過大推定することになると考えられる。本章で行ったシミュレーションにおいて、分析モデルが 2PLM であるときの識別力に関する \overline{Bias} の値が正の方向に大きくなったのはこのような理由に因るものだと考えられた。また、平均事情誤差平方根はこのバイアスと推定誤差の 2 つの要因の影響を受ける統計量であるため、本章で行ったシミュレーションにおいて分析モデルが 2PLM であるときの識別力に関する \overline{RMSE} の値が大きくなったのは、上述のように、 \hat{a}_j のバイアスが正の方向に大きくなったためであると考えられた。

3.3.4 先行研究との関連

3.1 節でも述べたように、本研究では、局所依存性が局所独立性を仮定した項目特性の推定に対してどのような影響を与えるのか、項目母数の比較可能性を確保した上で検討を行うため、発生モデルとして 2 値型の BTM を使用し、分析モデルとして 2PLM, BTM を使用して、Ip (2010) の方法を用いて項目母数を比較可能な形に変換した上で、局所依存性の影響に関する検討を行った。この点に関しては、2 値型の BTM に含まれる項目母数を 2PLM の項目母数と直接比較した Bradlow et al. (1999) の研究との間で相違が見られる。しかしながら、本研究では、比較可能性の有無と局所依存性の影響との関連を調べるため、比較可能性を確保しなかった場合の推定精度についても検討を加えており、これらの結果を Bradlow et al. (1999) で得られた結果と比較すると、以下の点に関しては、得られた結果が類似していた。

- 局所依存性を有するデータに対して局所独立性を仮定して識別力の推定を行った場合、局所依存関係にある項目の識別力を過小推定する
- 局所依存性を有するデータに対して局所独立性を仮定して識別力、困難度の推定を行うと、局所依存性を考慮した場合に比べて、M95%PIW が真値を含む確率が低下する
- 局所依存性を有するデータに対して局所独立性を仮定して識別力の推定を行うと、局所依存性を考慮した場合に比べて、真値と推定値との順位相関係数の値が低下する

また、Chen & Wang (2007) や Jiao et al. (2012) の研究においては、発生モデルや分析モデルとして、本研究で使用されたモデルとは異なるものが用いられているのだが、以下の点に関しては、本研究で比較可能性を確保しなかった場合の結果と得られた結果が類似していた。

- 局所依存性を有するデータに対して局所独立性を仮定して困難度の推定を行うと、困難度が過小推定される傾向にある
- 局所依存性を有するデータに対して局所独立性を仮定して困難度の推定を行うと、局所依存性を考慮した場合に比べて、平均二乗誤差平方根が増加する

表 3.1 分析モデルが 2 値型の BTM である時の識別力に関する \overline{Bias} , \overline{RMSE} , $cor(\mathbf{a}, \hat{\mathbf{a}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\mathbf{a}, \hat{\mathbf{a}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.094	0.118	0.715
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.072	0.117	0.877
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.060	0.119	0.893
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.048	0.119	0.872
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.026	0.119	0.848
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.112	0.134	0.787
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.085	0.135	0.890
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.066	0.131	0.901
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.050	0.129	0.884
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.024	0.122	0.880
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.142	0.162	0.626
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.096	0.155	0.872
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.075	0.149	0.882
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.048	0.138	0.853
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.015	0.125	0.837
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.033	0.060	0.827
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.020	0.056	0.963
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.017	0.062	0.964
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.017	0.069	0.953
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.007	0.071	0.948
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.034	0.060	0.874
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.022	0.065	0.965
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.018	0.069	0.968
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.019	0.076	0.958
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.007	0.075	0.959
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.039	0.065	0.805
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.024	0.072	0.962
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.021	0.078	0.962
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.019	0.085	0.945
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.002	0.083	0.940

表 3.2 分析モデルが 2PLM である時の識別力に関する \overline{Bias} , \overline{RMSE} , $cor(\mathbf{a}, \hat{\mathbf{a}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\mathbf{a}, \hat{\mathbf{a}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.227	0.257	0.647
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.191	0.248	0.566
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.163	0.223	0.682
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.144	0.207	0.708
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.118	0.181	0.837
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.174	0.206	0.735
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.148	0.194	0.849
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.135	0.188	0.873
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.122	0.182	0.877
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.107	0.175	0.876
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.173	0.204	0.607
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.141	0.189	0.872
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.135	0.192	0.881
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.123	0.189	0.861
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.111	0.186	0.823
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.155	0.169	0.797
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.116	0.175	0.658
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.091	0.153	0.723
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.082	0.136	0.726
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.060	0.099	0.943
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.096	0.114	0.843
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.076	0.107	0.949
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.068	0.104	0.951
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.067	0.108	0.942
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.055	0.101	0.953
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.074	0.096	0.767
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.063	0.098	0.964
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.062	0.104	0.959
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.065	0.111	0.939
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.057	0.111	0.930

表 3.3 識別力に関する $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta\overline{cor}(\mathbf{a}, \hat{\mathbf{a}})$ の値

条件	$\Delta\overline{Bias}$	$\Delta\overline{RMSE}$	$\Delta\overline{cor}(\mathbf{a}, \hat{\mathbf{a}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.132	0.139	-0.068
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.119	0.131	-0.311
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.103	0.104	-0.210
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.096	0.088	-0.164
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.092	0.062	-0.011
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.062	0.071	-0.052
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.063	0.059	-0.041
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.068	0.057	-0.028
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.072	0.053	-0.008
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.084	0.053	-0.004
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.031	0.042	-0.018
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.045	0.034	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.060	0.043	-0.001
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.075	0.051	0.009
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.095	0.061	-0.015
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.122	0.109	-0.030
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.096	0.119	-0.305
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.074	0.091	-0.241
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.065	0.068	-0.228
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.054	0.028	-0.004
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.062	0.054	-0.031
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.054	0.042	-0.015
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.049	0.035	-0.016
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.048	0.032	-0.017
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.049	0.026	-0.006
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.035	0.032	-0.038
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.039	0.026	0.002
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.042	0.026	-0.003
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.047	0.026	-0.006
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.054	0.029	-0.009

表 3.4 識別力に関する $\Delta\overline{Bias}/\sigma_{a_j^*}$ と $\Delta\overline{RMSE}/\sigma_{a_j^*}$ の値

条件	$\Delta\overline{Bias}/\sigma_{a_j^*}$	$\Delta\overline{RMSE}/\sigma_{a_j^*}$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.459	0.481
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.412	0.454
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.359	0.361
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.334	0.305
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.317	0.214
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.214	0.247
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.217	0.204
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.237	0.198
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.250	0.184
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.290	0.183
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.108	0.146
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.154	0.119
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.207	0.148
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.259	0.176
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.331	0.210
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.423	0.378
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.333	0.412
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.256	0.314
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.224	0.235
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.186	0.099
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.214	0.186
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.187	0.145
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.171	0.122
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.167	0.110
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.169	0.092
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.121	0.109
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.134	0.090
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.145	0.088
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.162	0.091
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.187	0.100

表 3.5 項目母数の比較可能性を確保しなかった場合の識別力に関する \overline{Bias} , \overline{RMSE} , $cor(\mathbf{a}, \hat{\mathbf{a}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\mathbf{a}, \hat{\mathbf{a}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.135	0.216	0.642
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.091	0.199	0.657
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.045	0.175	0.729
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.002	0.163	0.767
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.041	0.154	0.833
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.196	0.251	0.741
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.132	0.219	0.713
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.073	0.188	0.753
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.021	0.176	0.761
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.026	0.155	0.872
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.291	0.328	0.628
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.213	0.273	0.616
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.123	0.218	0.642
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.064	0.206	0.616
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.003	0.166	0.821
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.207	0.230	0.797
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.166	0.198	0.825
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.117	0.159	0.856
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.065	0.120	0.907
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.017	0.087	0.945
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.274	0.287	0.873
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.203	0.234	0.735
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.140	0.179	0.786
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.075	0.129	0.840
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	-0.026	0.097	0.954
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.389	0.396	0.825
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.291	0.317	0.636
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.195	0.231	0.679
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.121	0.170	0.713
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.051	0.116	0.934

表 3.6 項目母数の比較可能性を確保しなかった場合の識別力に関する $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta\overline{cor}(\mathbf{a}, \hat{\mathbf{a}})$ の値

条件	$\Delta\overline{Bias}$	$\Delta\overline{RMSE}$	$\Delta\overline{cor}(\mathbf{a}, \hat{\mathbf{a}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.229	0.098	-0.073
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.164	0.082	-0.220
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.105	0.057	-0.163
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.050	0.045	-0.105
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.014	0.035	-0.015
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.308	0.117	-0.046
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.217	0.084	-0.177
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.140	0.057	-0.148
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.071	0.047	-0.123
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.002	0.033	-0.007
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.432	0.166	0.002
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.310	0.117	-0.256
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.197	0.069	-0.240
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.112	0.069	-0.237
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.012	0.041	-0.016
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.240	0.170	-0.030
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.187	0.142	-0.138
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.134	0.097	-0.108
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.082	0.051	-0.046
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.024	0.016	-0.003
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.308	0.227	-0.001
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.226	0.169	-0.230
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.159	0.110	-0.182
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.094	0.052	-0.118
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	-0.033	0.022	-0.005
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.428	0.331	0.020
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.315	0.246	-0.326
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.215	0.152	-0.283
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.140	0.085	-0.232
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.053	0.033	-0.006

表 3.7 項目母数の比較可能性を確保しなかった場合の $\Delta \overline{Bias}/\sigma_{a_j^*}$ および $\Delta \overline{RMSE}/\sigma_{a_j^*}$ の値

条件	$\Delta \overline{Bias}/\sigma_{a_j^*}$	$\Delta \overline{RMSE}/\sigma_{a_j^*}$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.794	0.339
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.567	0.283
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.362	0.197
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.173	0.155
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.049	0.120
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-1.066	0.404
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.751	0.292
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.484	0.199
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.244	0.164
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.006	0.113
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-1.498	0.575
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-1.072	0.407
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.684	0.238
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.387	0.238
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.042	0.141
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.830	0.588
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.646	0.492
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.465	0.336
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.282	0.178
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.083	0.056
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-1.066	0.786
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.782	0.584
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.549	0.381
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.327	0.181
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	-0.115	0.076
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-1.484	1.147
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-1.093	0.852
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.746	0.528
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.484	0.294
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.185	0.115

表 3.8 分析モデルが 2 値型の BTM である時の困難度に関する \overline{Bias} , \overline{RMSE} , $cor(\mathbf{b}, \hat{\mathbf{b}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\mathbf{b}, \hat{\mathbf{b}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.022	0.173	0.986
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.004	0.148	0.989
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.002	0.146	0.990
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.009	0.134	0.991
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.003	0.127	0.992
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.018	0.166	0.989
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.008	0.152	0.991
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.002	0.148	0.991
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.006	0.141	0.992
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	-0.004	0.134	0.993
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.028	0.142	0.994
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.046	0.142	0.993
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.038	0.138	0.993
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.012	0.132	0.994
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.014	0.126	0.995
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.005	0.090	0.996
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.006	0.078	0.997
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.006	0.074	0.997
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.013	0.073	0.997
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.013	0.073	0.998
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.002	0.089	0.997
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.005	0.084	0.997
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.004	0.078	0.998
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.009	0.078	0.998
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.010	0.076	0.998
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.009	0.086	0.997
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.002	0.086	0.997
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.002	0.082	0.998
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.004	0.081	0.998
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.011	0.076	0.998

表 3.9 分析モデルが 2PLM である時の困難度に関する \overline{Bias} , \overline{RMSE} , $cor(\mathbf{b}, \hat{\mathbf{b}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\mathbf{b}, \hat{\mathbf{b}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.055	0.227	0.987
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.039	0.193	0.984
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.022	0.188	0.984
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.005	0.158	0.989
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.019	0.140	0.992
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.023	0.205	0.990
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.024	0.175	0.990
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.016	0.168	0.991
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.003	0.150	0.992
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	-0.013	0.141	0.993
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.020	0.178	0.993
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.040	0.156	0.993
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.033	0.148	0.994
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.011	0.135	0.994
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.018	0.127	0.995
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.045	0.144	0.996
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.030	0.131	0.991
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.012	0.125	0.992
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.006	0.101	0.994
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.003	0.076	0.998
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.014	0.124	0.997
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.018	0.101	0.997
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.012	0.096	0.997
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.002	0.089	0.997
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	-0.001	0.079	0.998
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.007	0.110	0.997
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.012	0.094	0.997
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.008	0.089	0.998
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.002	0.085	0.998
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.003	0.075	0.998

表 3.10 困難度に関する $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta\text{cor}(\mathbf{b}, \hat{\mathbf{b}})$ の値

条件	$\Delta\overline{Bias}$	$\Delta\overline{RMSE}$	$\Delta\text{cor}(\mathbf{b}, \hat{\mathbf{b}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.033	0.054	0.001
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.035	0.045	-0.005
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.024	0.043	-0.005
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.014	0.023	-0.002
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.016	0.013	0.000
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.004	0.039	0.001
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.016	0.023	0.000
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.014	0.020	0.000
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.009	0.010	0.000
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	-0.008	0.007	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.008	0.035	-0.001
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.005	0.013	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.005	0.010	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.002	0.004	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.004	0.001	0.000
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.040	0.054	0.000
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.037	0.053	-0.006
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.018	0.051	-0.006
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.007	0.027	-0.003
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.016	0.003	0.000
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.016	0.035	0.000
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.024	0.017	0.000
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.017	0.018	-0.001
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.011	0.011	0.000
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	-0.011	0.003	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.002	0.024	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.015	0.008	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.010	0.006	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.005	0.004	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.008	-0.001	0.000

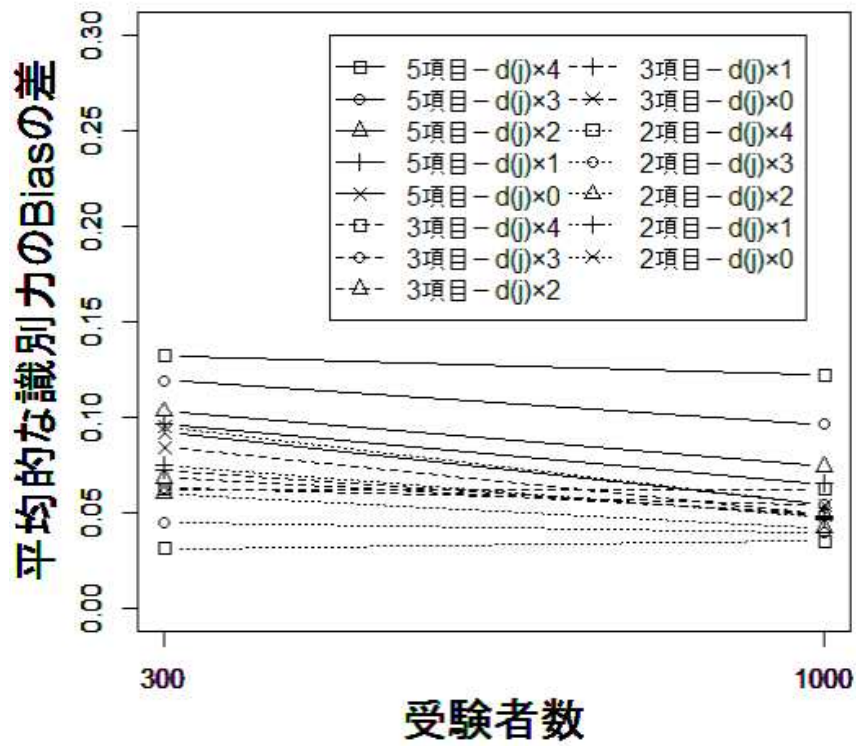


図 3.1 $\Delta \overline{Bias} / \sigma_{a_j^*}$ と受験者数との関係

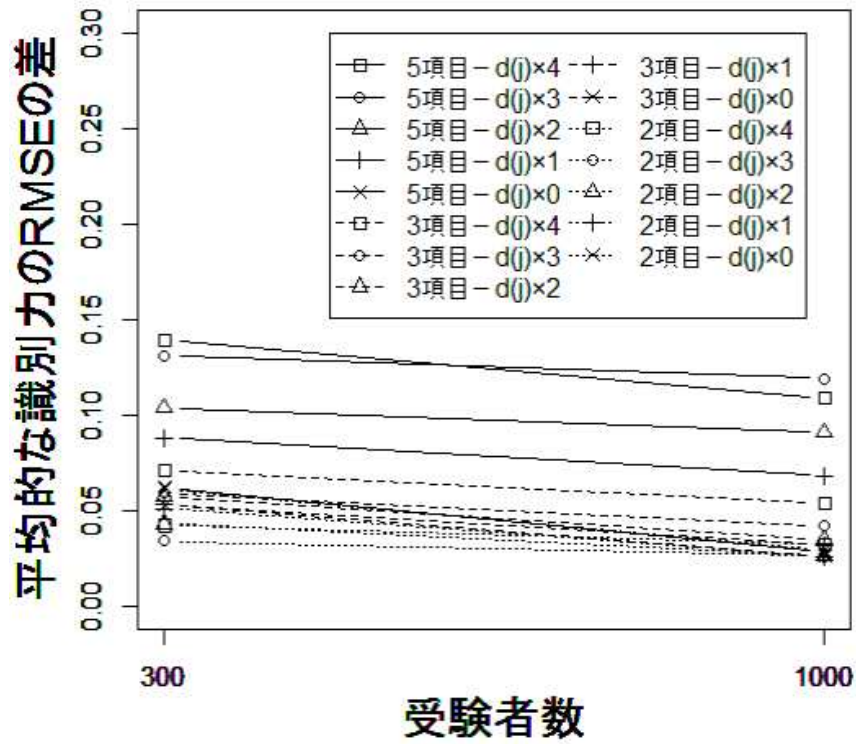


図 3.2 $\overline{\Delta RMSE}/\sigma_{a_j^*}$ と受験者数との関係

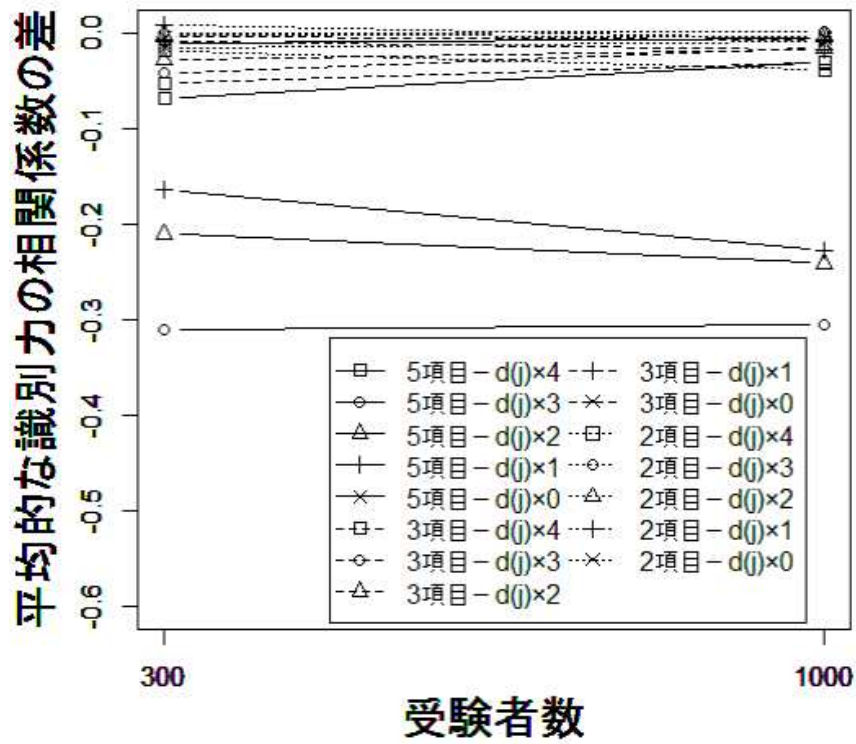


図 3.3 $\Delta cor(\mathbf{a}, \hat{\mathbf{a}})$ と受験者数との関係

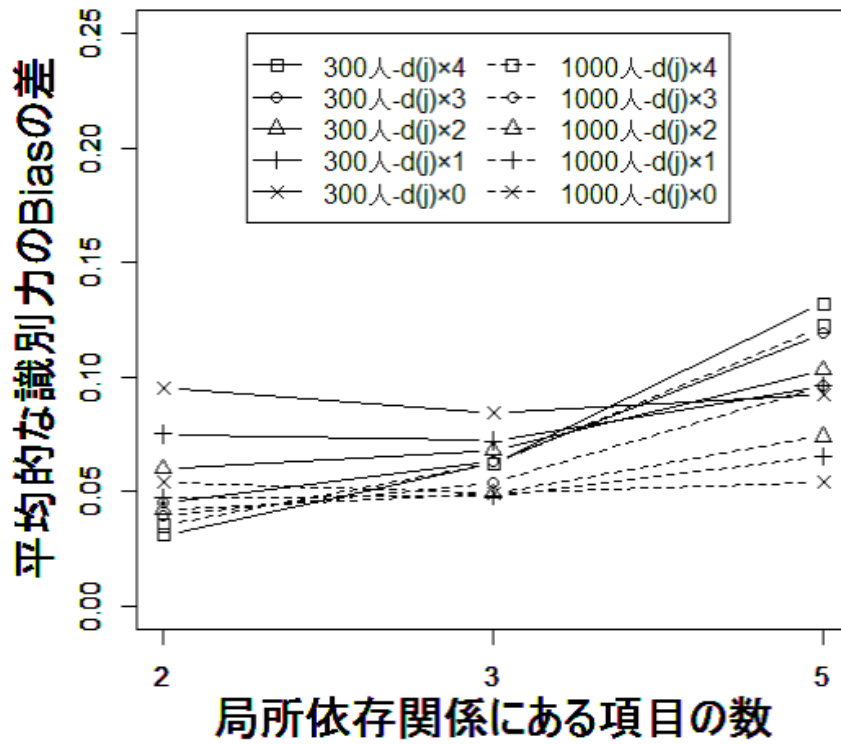


図 3.4 $\Delta \overline{Bias} / \sigma_{a_j^*}$ と局所依存関係にある項目の数との関係

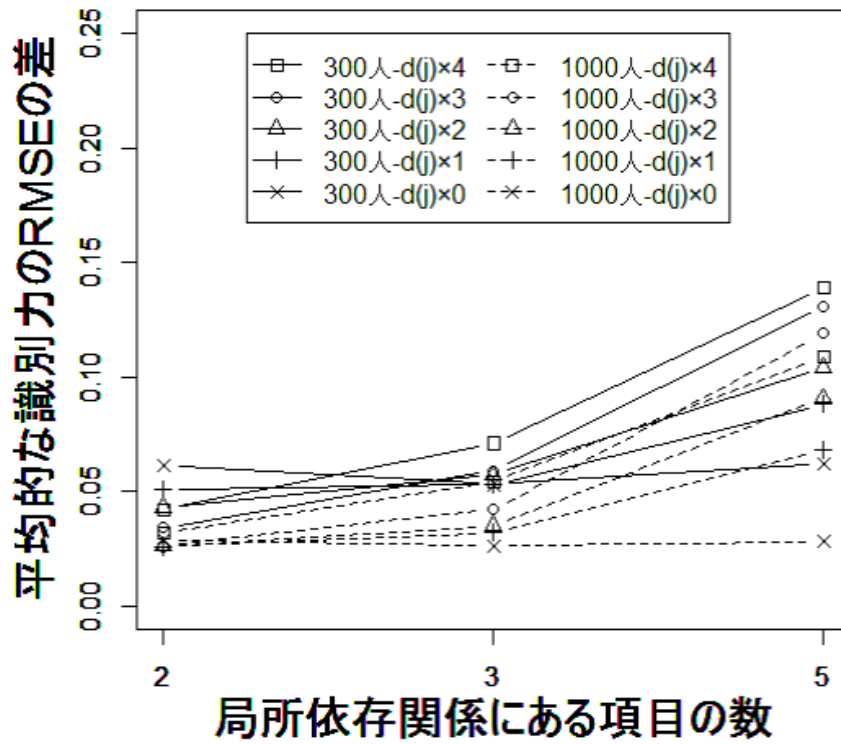


図 3.5 $\overline{\Delta RMSE}/\sigma_{a_j^*}$ と局所依存関係にある項目の数との関係

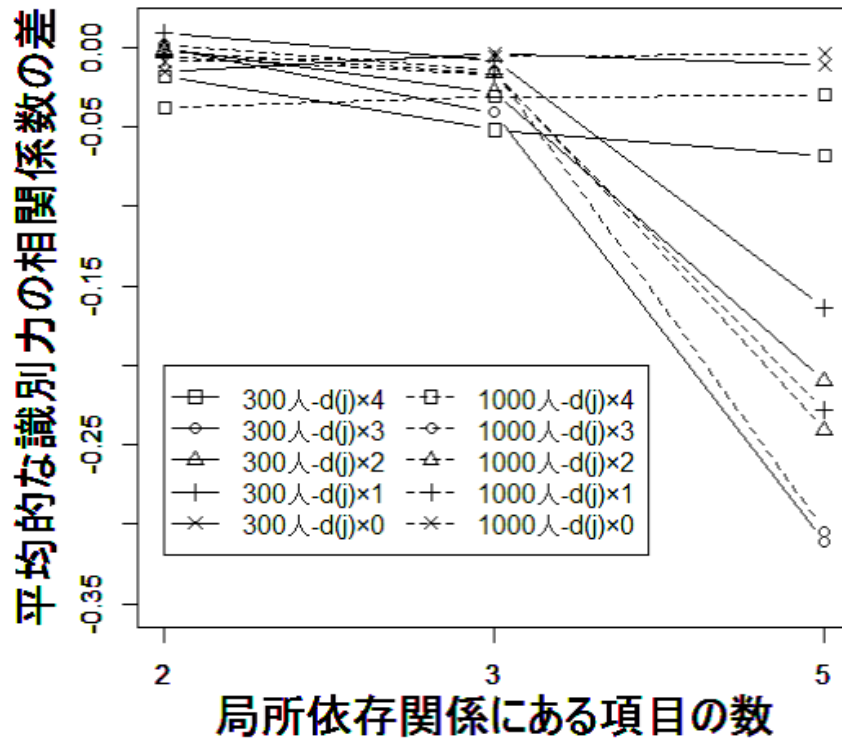


図 3.6 $\Delta cor(a, \hat{a})$ と局所依存関係にある項目の数との関係

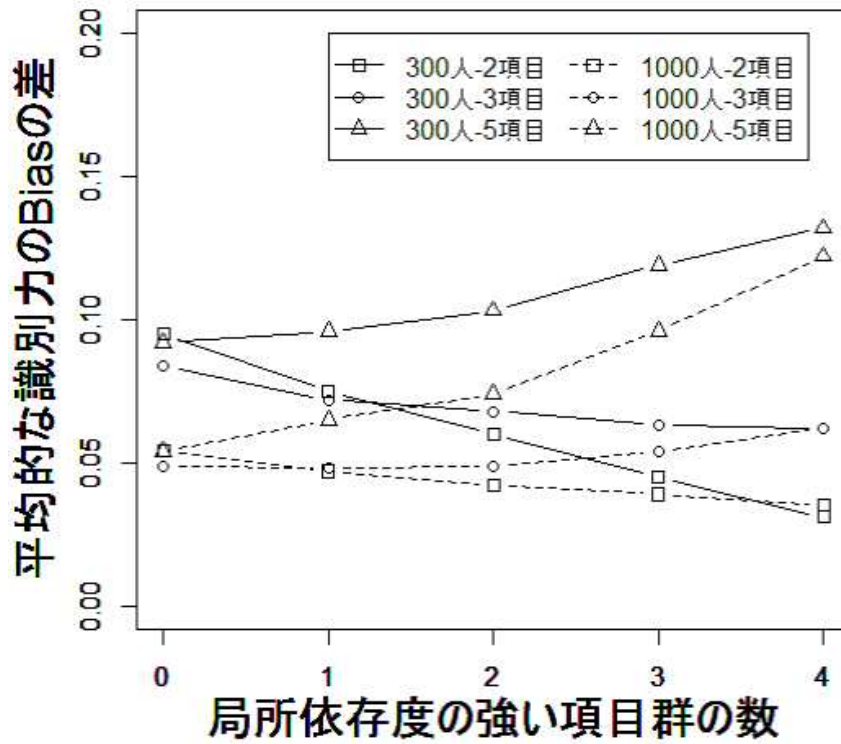


図 3.7 $\Delta \overline{Bias} / \sigma_{a_j^*}$ と 4 つの項目群の局所依存度との関係

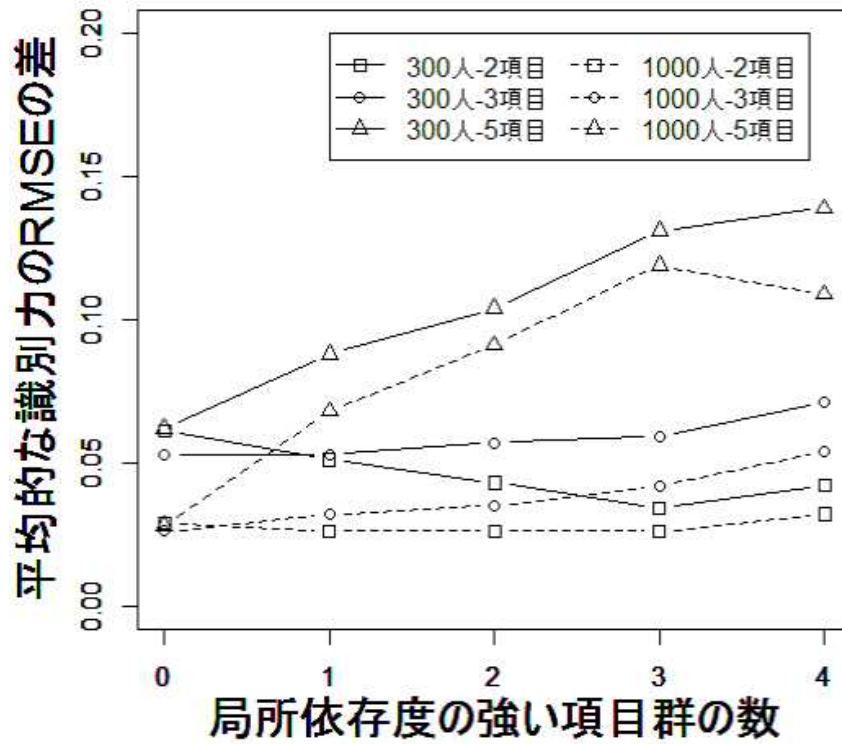


図 3.8 $\overline{\Delta RMSE}/\sigma_{a_j^*}$ と 4 つの項目群の局所依存度との関係

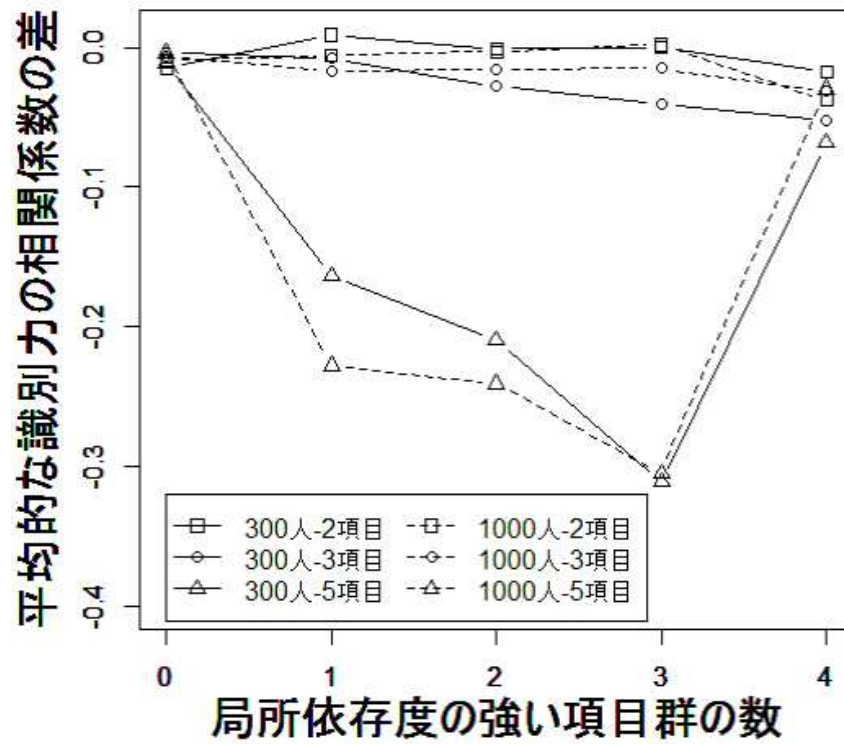


図 3.9 $\Delta cor(a, \hat{a})$ と 4 つの項目群の局所依存度との関係

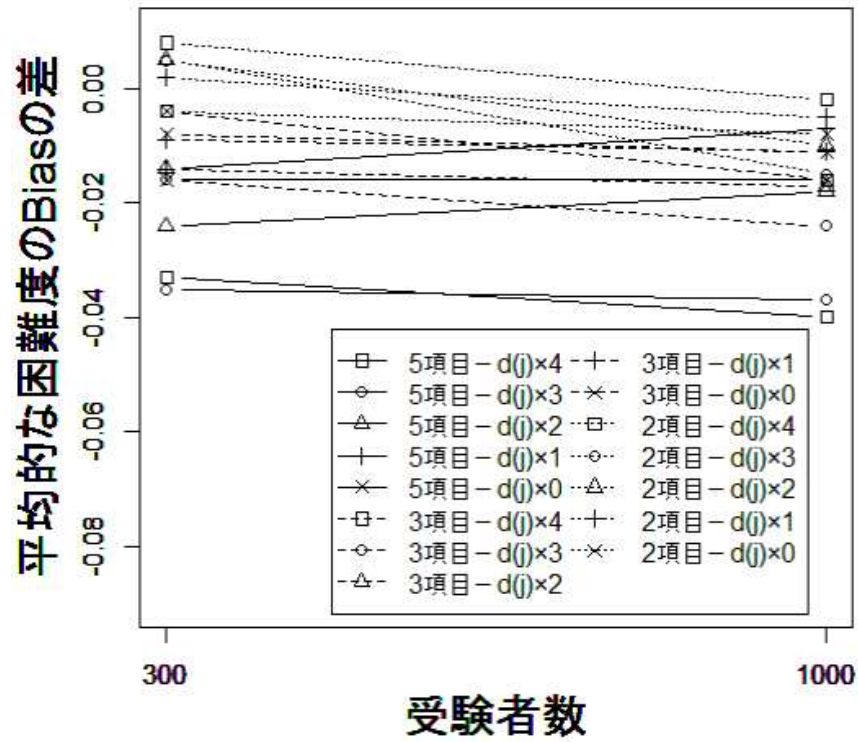


図 3.10 困難度に関する $\overline{\Delta Bias}$ と受験者数との関係

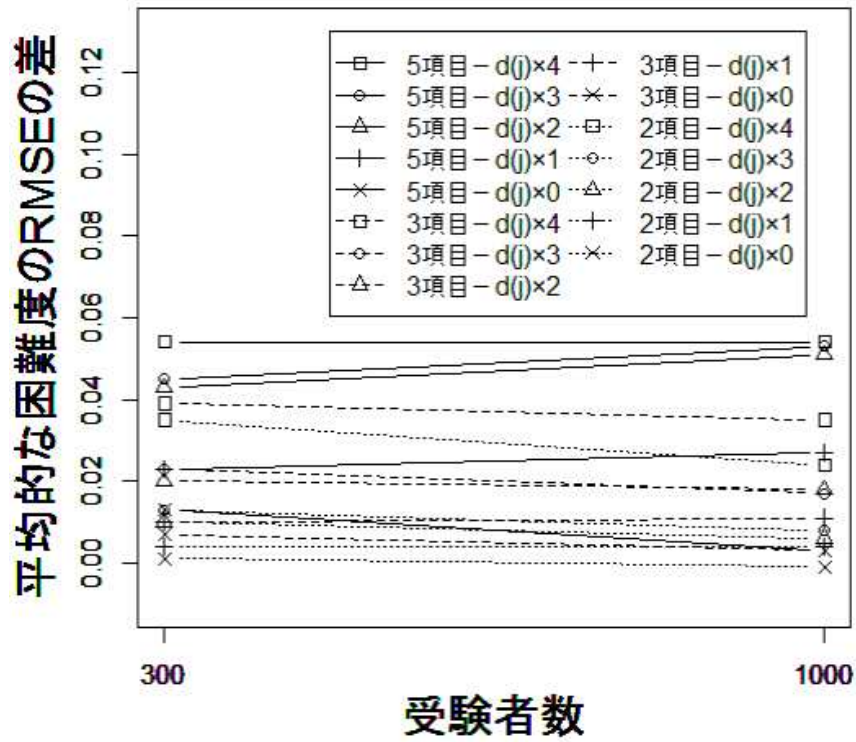


図 3.11 困難度に関する $\overline{\Delta RMSE}$ と受験者数との関係

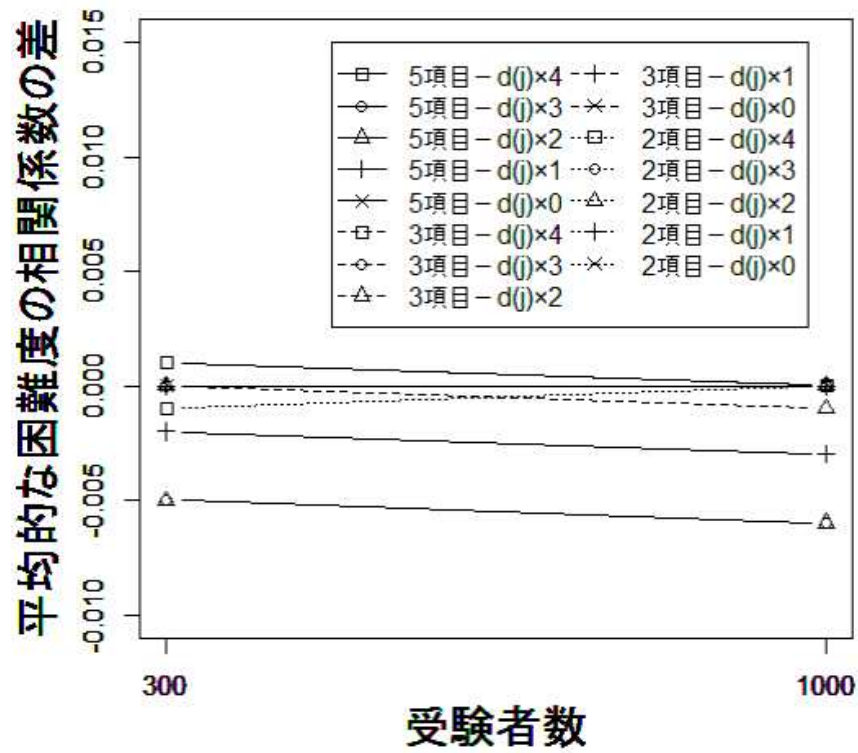


図 3.12 $\Delta cor(b, \hat{b})$ と受験者数との関係

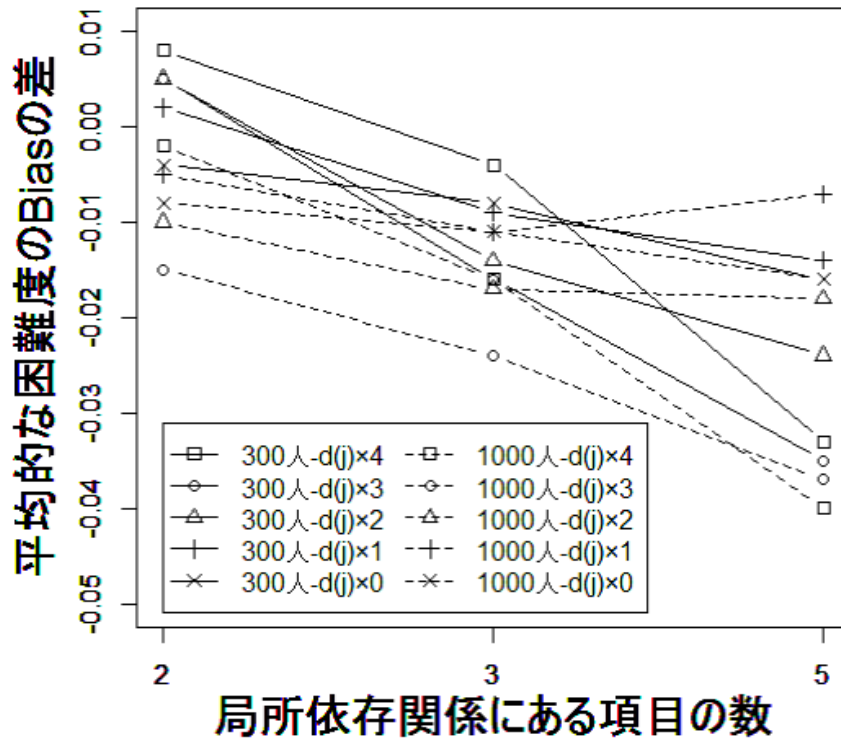


図 3.13 困難度に関する $\overline{\Delta Bias}$ と局所依存関係にある項目の数との関係

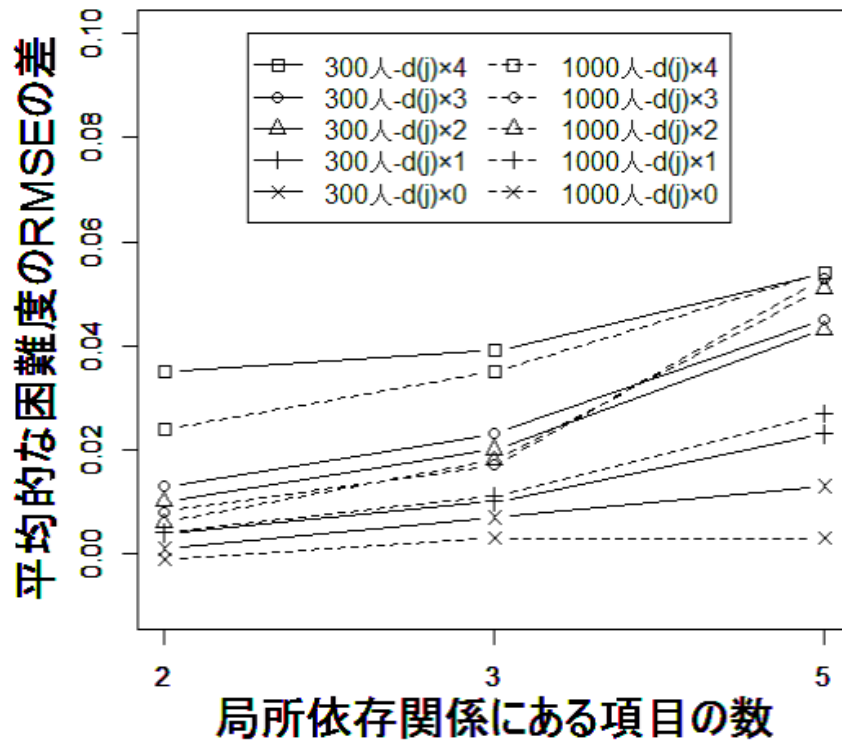


図 3.14 困難度に関する $\Delta \overline{RMSE}$ と局所依存関係にある項目の数との関係

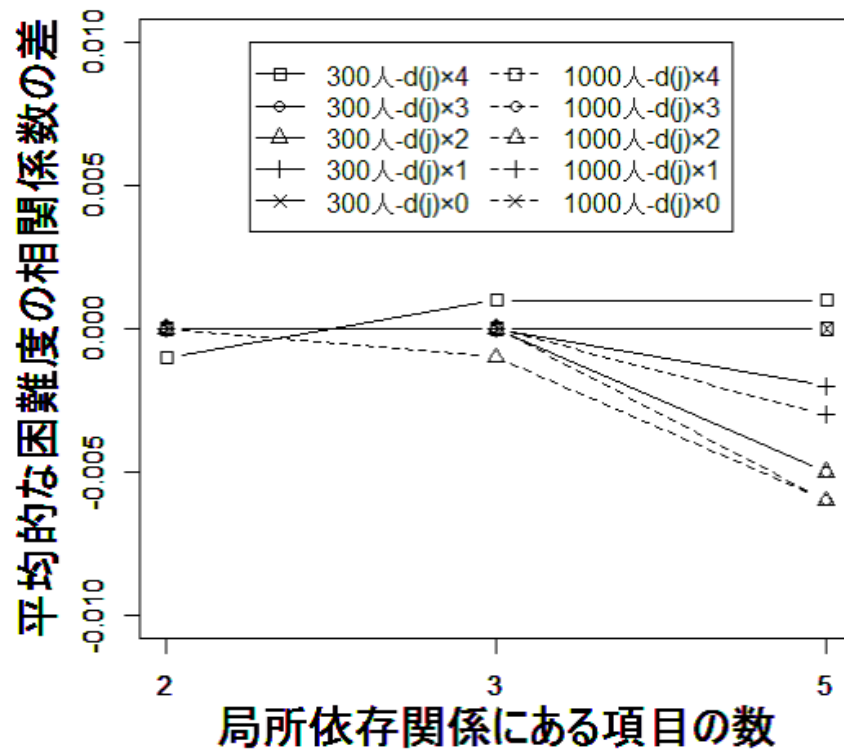


図 3.15 $\Delta cor(\mathbf{b}, \hat{\mathbf{b}})$ と局所依存関係にある項目の数との関係

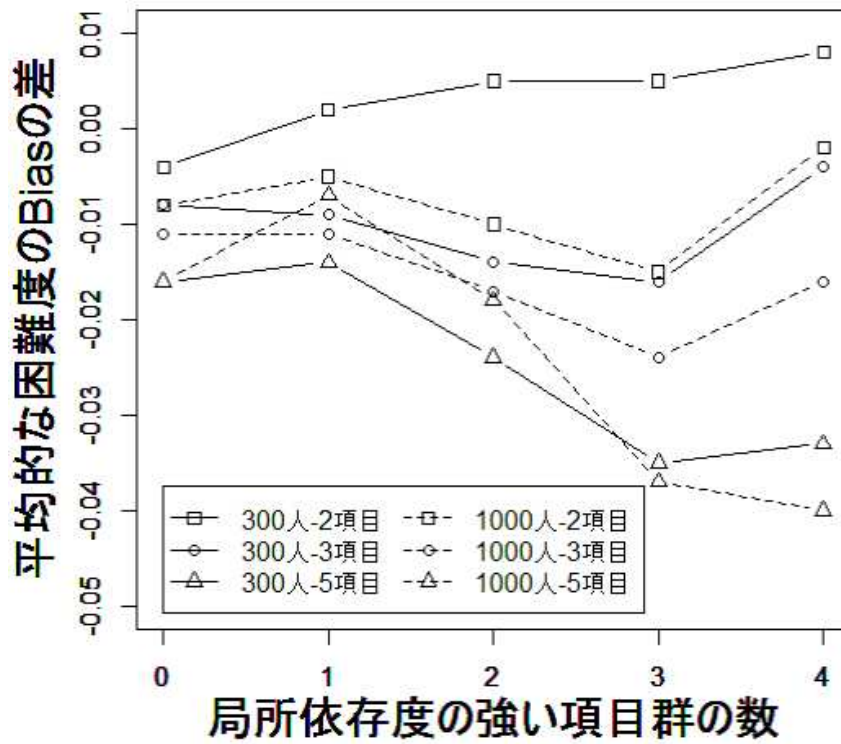


図 3.16 困難度に関する $\Delta \overline{Bias}$ と4つの項目群の局所依存度との関係

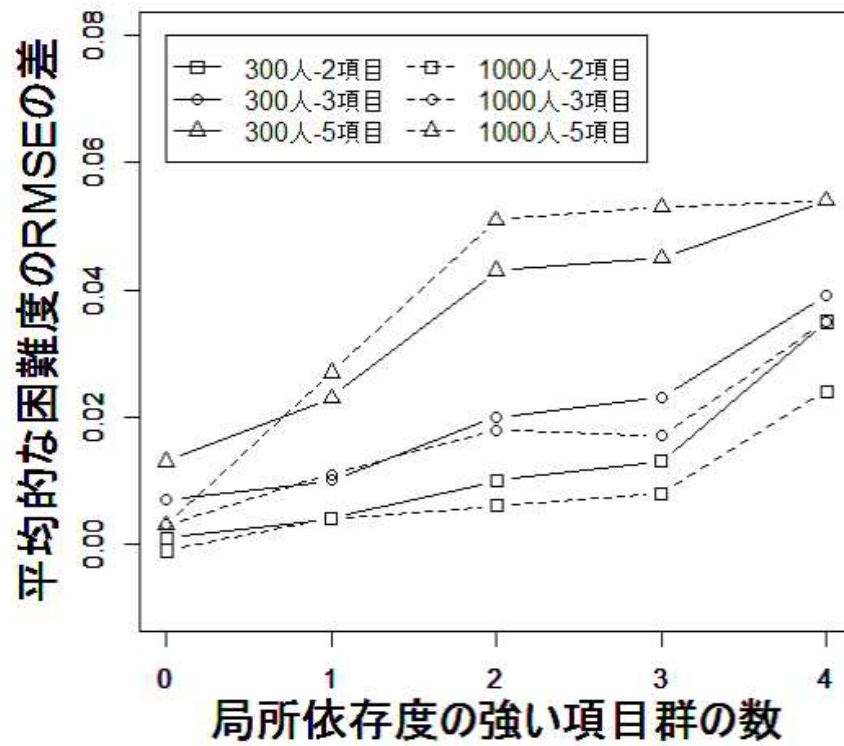


図 3.17 困難度に関する $\overline{\Delta RMSE}$ と 4 つの項目群の局所依存度との関係

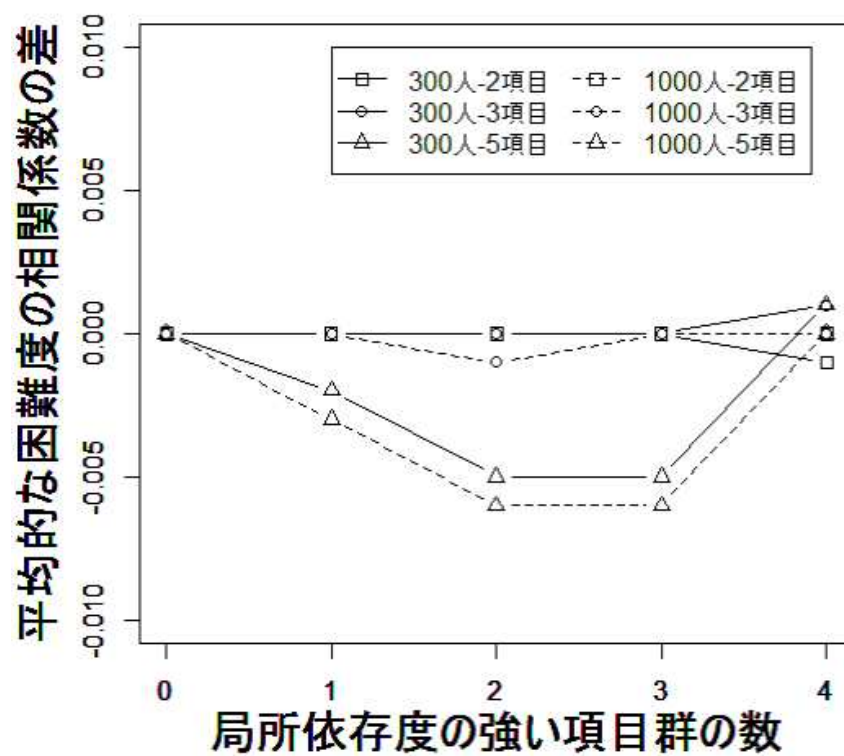


図 3.18 困難度に関する $\Delta cor(\mathbf{b}, \hat{\mathbf{b}})$ と 4 つの項目群の局所依存度との関係

第4章

局所独立性を仮定したテスト情報量の推定に与える影響

4.1 問題と目的

本章では、局所依存性が局所独立性を仮定したテスト情報量の推定に対してどのような影響を与えるのか、シミュレーションを通して検討を加える。このテーマに関しては、既にいくつかの先行研究において検討が加えられており、例えば、Ip (2010) においては、局所依存性を有すると考えられる National Assessment of Educational Progress (NAEP) のデータに対して局所独立性を仮定して情報量の推定を行った結果、局所依存性を考慮した場合に比べて、情報量が過大推定される傾向にあったことが報告されている。また、石塚他 (2001) において、2000 年度に実施された大学入試センター試験英語の回答データに対して GRM と 2PLM をあてはめてテスト情報量の算出を行ったところ、2PLM のテスト情報量が GRM のテスト情報量を概ね上回る傾向にあったことが報告されている。さらに、Keller, Swaminathan, & Sireci (2003) の研究でも、実データを用いた検討を行ったところ、上記の先行研究と同様の結果が報告されており、Wainer & Wang (2000) においても、局所依存性を有すると考えられる TOEFL のデータに対し局所独立性を仮定して情報量の推定を行った結果、局所依存性を考慮した場合に比べて、情報量が過大推定される傾向にあることが報告されている。このようなことから、例えば、Ip (2000) においては、局所独立性を仮定した場合に生じるテスト情報量の過大推定を修正するための方法が提案されている。また、これと関連して、Lee (2000) においては、局所依存性を有するデータに対し局所独立性を仮定してテスト得点の条件付き標準誤差を求めると、局所依存性を考慮した場合に比べて、条件付き標準誤差を過小推定する傾向にあることが報告されており、Wainer et al. (2007, p. 182) では、局所依存性を有すると考えられる乳がん患者への調査データに対し局所独立性を仮定して受験者特性の推定を行ったところ、局所依存性

を考慮した場合に比べて、特性値の事後分布の幅が狭まったことが報告されている。さらに、テスト情報量と関連するテストの信頼性という観点からは、Anastasi (1961, p. 121) や Thorndike (1951, p. 585), Guilford (1936, p.417), Kelly (1924) において、項目反応間の局所依存性を考慮せずに信頼性の推定を行うと、当該テストの信頼性を過大推定することになることが指摘されており、Wainer (1995) においては、局所依存性を有すると考えられる Low School Admission Test (LSAT) のデータに対し局所独立性を仮定して信頼性の算出を行った結果、局所依存性を考慮した場合に比べてテストの信頼性が過大推定されたことが報告されている。加えて、Wang & Wilson (2005) においても、台湾で発行されている17の心理学のジャーナルを154名の評定者が8つの観点から評定したデータに対して局所独立性を仮定して分析を行ったところ、局所依存性を考慮した場合に比べて、テストの信頼性が過大推定されたことが報告されている。

以上のように、先行研究においては、局所依存性が局所独立性を仮定したテスト情報量の推定に対しどのような影響を及ぼすのか、様々な角度から検討がなされてきたのだが、これらの先行研究においては、主に実データの分析を通して、局所依存性の影響に関し検討が行われてきた。しかしながら、本研究を含め関連する研究において得られた知見を実際にテストを作成、評価、運用する場面に応用するという観点からは、シミュレーションを実施し、局所独立性を仮定してテスト情報量を推定した場合に推定量のバイアス、平均二乗誤差平方根がどの程度となるのか、真のテスト情報量との比較し定量的に把握する必要があると考えられる。また、実際のテスト運用や評価が、様々な受験者数のテスト、様々な構成のテストに対して行われ得ることを考慮すると、シミュレーションを実施する際に受験者数等の要因を操作し、推定量のバイアス、平均二乗誤差平方根と受験者数等の要因がどのような関係にあるのか、検討する必要があると考えられる。

そこで、本章では、局所依存性が局所独立性を仮定したテスト情報量の推定に対してどのような影響を及ぼすのか、真のテスト情報量と比較し検討を行うためシミュレーションを実施した。また、シミュレーションを行う際に受験者数等の要因を系統的に変化させ、推定量のバイアス、平均二乗誤差平方根と受験者数等の要因がどのような関係にあるのか、検討を行った。

4.2 方法

本研究では、局所依存性を有するデータに対して局所独立性を仮定するモデルと局所依存性を考慮したモデルをあてはめて母数の推定を行い、得られた推定値を用いて、分析モデル毎に、テスト情報量を算出した。そして、この手続きを多数回繰り返し、多数のテスト情報量の推定値から、分析モデル毎に、推定量のバイアス、平均二乗誤差を算出して、これらの統計量に関して比較を行った。

4.2.1 データの発生モデル

局所依存性を有するデータを発生させる際には、(1.14) 式の 2 値型の BTM を使用した (以下にモデル式を再掲).

$$P_{j \in d(j)}(\theta_i) = \frac{1}{1 + \exp[-1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)} - \gamma_{id(j)})]} \quad (4.1)$$

1.5 節でも述べたように、局所依存性を考慮したモデルには 2 値型の BTM が属するタイプ b のモデルの他にもタイプ a のモデルとタイプ c のモデルが存在する. 本研究では、シミュレーションを行う際に各受験者の各項目に対する項目反応パターンが必要となってくるのだが、タイプ a に属するモデルは局所依存関係にある項目群 $d(j)$ における正答数を分析対象としているため、基本的には、各受験者の各項目に対する項目反応パターンを表現することができない. また、タイプ c のモデルに関しては、3 項目以上が局所依存関係にある場合であっても、その状況に即した項目反応パターンを表現することが可能ではあるが、先行研究での使用例を見る限り、基本的には 2 つの項目が局所依存関係にある場合を想定したモデルであると考えられる (Braeken et al., 2007; Hoskens & De Boeck, 1997; Ip, 2002; Ip et al., 2009).

そこで、本研究では、局所依存性を有するデータを発生させる際に、タイプ b に属する 2 値型の BTM を使用することとした.

4.2.2 データの分析モデル

本研究では、局所依存性を有するデータを分析する際に、以下の 2 つのモデルを使用した.

局所独立性を仮定した場合の分析モデル

本研究では、局所依存性を有するデータに対し局所独立性を仮定して分析を行う際、(1.1) 式の 2PLM を使用した (以下にモデル式を再掲).

$$P_j(\theta_i) = \frac{1}{1 + \exp[-1.7a_j(\theta_i - b_j)]} \quad (4.2)$$

局所依存性を考慮した場合の分析モデル

本研究では、局所依存性を有するデータの局所依存性を考慮して分析を行う際、データの発生モデルでもある 2 値型の BTM を使用した (モデル式については (4.1) 式を参照のこと).

4.2.3 テスト情報量

4.2節の冒頭でも述べたように、本研究では、局所依存性を有するデータに対して局所独立性を仮定するモデルと局所依存性を考慮するモデルをあてはめて母数の推定を行い、得られた推定値から、分析モデル毎に、テスト情報量の推定を行った。ここで、テスト情報量とは、データ行列 \mathbf{U} の i 行目 \mathbf{u}_i (受験者 i の項目反応パターン) が受験者 i の特性値 θ_i に関して有する情報の量を表しており、いま、 θ_i に関するテスト情報量を $I(\theta_i)$ と表記することになると、 $I(\theta_i)$ は以下の式で定義される ((4.3) 式)。

$$I(\theta_i) = E \left[\left(\frac{\partial}{\partial \theta_i} \log L(\mathbf{u}_i | \theta_i) \right)^2 \middle| \theta_i \right] \quad (4.3)$$

また、最尤推定量 $\hat{\theta}_i$ の性質より、項目数が十分大きい場合には、漸近的に、 θ_i が与えられたときの $\hat{\theta}_i$ の分散 $\sigma_{\hat{\theta}_i | \theta_i}^2$ とテスト情報量 $I(\theta_i)$ との間に (4.4) 式の関係が成立する。

$$\sigma_{\hat{\theta}_i | \theta_i}^2 = \frac{1}{I(\theta_i)} \quad (4.4)$$

したがって、これらのことから、テスト情報量 $I(\theta_i)$ が大きくなるほど受験者特性値の推定量 $\hat{\theta}_i$ の推定誤差が小さくなるため、一般的には、受験者特性の全域にわたってテスト情報量の大きいテストほど、良いテストであると考えられる。

局所独立性を仮定した場合のテスト情報量

本研究では、局所独立性を仮定した場合の分析モデルとして 2PLM を使用した。分析モデルとして 2PLM を使用した場合には、テスト情報量の具体的な関数形が (4.5) 式となる (詳細については Lord & Novick (1968) や豊田 (2002) を参照のこと)。

$$I(\theta_i) = (1.7)^2 \sum_{j=1}^J a_j^2 P_j(\theta_i) Q_j(\theta_i) \quad (4.5)$$

ここで、(4.5) 式中の $P_j(\theta_i)$ は 2PLM の式 ((4.2) 式) を表しており、 $Q_j(\theta_i)$ は $1 - P_j(\theta_i)$ を表している。

本研究では、推定された 2PLM の項目母数と (4.5) 式を用いて、局所独立性を仮定した場合のテスト情報量を算出した。

局所依存性を考慮した場合のテスト情報量

本研究では、局所依存性を考慮した場合の分析モデルとして 2 値型の BTM を使用した。分析モデルとして 2 値型の BTM を使用した場合には、テスト情報量の具体的な関数形が

(4.6) 式となる (詳細については Wainer, Bradlow, & Du (2000) や Ip (2010) を参照のこと).

$$I(\theta_i) = \sum_{j=1}^J \int_{-\infty}^{\infty} (1.7a_{j \in d(j)})^2 \left[\frac{\exp(1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)} - \gamma_{id(j)}))}{(1 + \exp(1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)} - \gamma_{id(j)})))^2} \right] d\gamma_{id(j)} \quad (4.6)$$

本研究では, 推定された 2 値型の BTM の項目母数と (4.6) 式を用いて, 局所依存性を考慮した場合のテスト情報量を算出した.

4.2.4 シミュレーション実験の手続き

本研究では, データを発生させる際, テストは局所依存関係にある 4 つの項目群から構成されているとし, 各項目群は 5 つの項目から構成されていると仮定した. また, 4.1 節でも述べたように, 本研究では, 推定量のバイアス, 平均二乗誤差と種々の要因がどのような関係にあるのか検討を行うため, シミュレーションを行う際に, 受験者数, 局所依存関係にある項目の数, 4 つの項目群の局所依存度を系統的に変化させることとした. 受験者数に関しては, 300 人, 1000 人という 2 つの水準を設定し, 局所依存関係にある項目の数に関しては, 各項目群内で実際に局所依存関係にある項目の数について以下の 3 つの水準を設定した.

- 5 項目が局所依存関係にある ($J_{d(j)} = 5$)
- 3 項目が局所依存関係にある ($J_{d(j)} = 3$)
- 2 項目が局所依存関係にある ($J_{d(j)} = 2$)

また, 項目間の局所依存度に関しては, テストに含まれる 4 つの項目群の局所依存度について以下の 5 つの水準を設定した.

- 4 つとも局所依存度が強い (強 $d(j) \times 4$)
- 4 つのうち 3 つの局所依存度が強い (強 $d(j) \times 3$)
- 4 つのうち 2 つの局所依存度が強い (強 $d(j) \times 2$)
- 4 つのうち 1 つだけ局所依存度が強い (強 $d(j) \times 1$)
- 4 つとも局所依存度が弱い (強 $d(j) \times 0$)

なお、項目間の局所依存度の強弱に関しては、3章と同様の基準を用いて操作を行った。

本研究で行ったシミュレーションの詳細は、以下のようにまとめられる*1。

1. 各受験者の受験者特性値 θ_i ($i = 1, 2, \dots, N$) と $\gamma_{id(j)}$ ($a = 1, 2, \dots, N; d(j) = 1, 2, 3, 4$) および各項目の識別力 $a_{j \in d(j)}$ と困難度 $b_{j \in d(j)}$ ($j = 1, 2, \dots, 20$) の真値をそれぞれ標準正規分布 $N(0, 1)$, 正規分布 $N(0, \sigma_{\gamma_{d(j)}}^2)$, 一様分布 $U(0.5, 1.5)$, 標準正規分布 $N(0, 1)$ から発生させる*2。
2. 各受験者の各項目に対する正答確率を手続き 1 で発生させた各母数の真値と 2 値型の BTM の式 ((4.1) 式) から算出し, $N \times 20$ の正答確率行列 \mathbf{A} を作成する。
3. 一様分布 $U(0, 1)$ から乱数を $N \times 20$ 個発生させて $N \times 20$ の一様乱数行列 \mathbf{B} を作成する。
4. \mathbf{A}, \mathbf{B} の各要素を比較して局所依存性が生じている仮想的な項目反応データの行列 \mathbf{U} を作成する。具体的には, $a_{ij} \geq b_{ij}$ である場合には $u_{ij} = 1$ とし, $a_{ij} < b_{ij}$ である場合には $u_{ij} = 0$ とする。
5. \mathbf{U} に対して 4.2.2 節で紹介した 2 つの分析モデルを当てはめて母数の推定を行う。
6. 4.2.3 節で紹介したテスト情報量の式を利用して, 手続き 5 で得られた母数の推定値から分析モデル毎にテスト情報量を算出する。
7. 手続き 3 から 6 までを R 回繰り返す。
8. 手続き 7 の結果得られた R 組のテスト情報量より, 分析モデル毎に以下の統計量を算出する。

$$Bias(\hat{I}(\theta_i)) = \frac{1}{R} \sum_{r=1}^R \hat{I}_r(\theta_i) - I(\theta_i) \quad (4.7)$$

$$RMSE(\hat{I}(\theta_i)) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{I}_r(\theta_i) - I(\theta_i))^2} \quad (4.8)$$

ここで, (4.7), (4.8) 式中の $\hat{I}(\theta_i)$ は $I(\theta_i)$ の推定量を表しており, $\hat{I}_r(\theta_i)$ は $I(\theta_i)$ の r 回目の推定値を表わしている。また, 手続き 7 における反復回数 R に関しては, 繰り返し毎の母数の推定値の変動を考慮して, 受験者数が 1000 人の場合には $R = 50$ とし, 受験者数が 300 人の場合には $R = 100$ とした。

*1 シミュレーション用プログラムの記述および実行には, 統計解析向けプログラミング言語であり実行環境でもある R を利用した。

*2 なお, 擬似乱数を発生させる際には R に実装されているメルセンヌツイスターを使用した。

4.2.5 MCMC による母数の推定

本研究においても、手続き 5 における母数の推定は MCMC により行った。MCMC の具体的なアルゴリズムとしてはスライスサンプリング (Neal, 1997) を利用し、WinBUGS 1.4 (Spiegelhalter et al., 2003) を利用して MCMC を実行した。また、事前には各母数が統計的に互いに独立であったと仮定し、分析モデル毎に、各母数の事前分布を以下のように設定した。

- 2PLM を使用した場合
 - θ_i の事前分布：標準正規分布 $N(0, 1)$
 - a_j の事前分布：正規分布 $N(1, 0.25)$
 - b_j の事前分布：標準正規分布 $N(0, 1)$
- 2 値型の BTM を使用した場合
 - θ_i の事前分布：標準正規分布 $N(0, 1)$
 - $a_{j \in d(j)}$ の事前分布：正規分布 $N(1, 0.25)$
 - $b_{j \in d(j)}$ の事前分布：標準正規分布 $N(0, 1)$
 - $\gamma_{id(j)}$ の事前分布：正規分布 $N(0, \sigma_{\gamma_{d(j)}}^2)$
 - $\sigma_{\gamma_{d(j)}}^{-2}$: ガンマ分布 $\Gamma(3, 1)$

さらに、各母数の初期値に関しては、分析モデル毎に、以下の様に設定した。

- 2PLM を使用した場合
 - θ_i の初期値：0
 - a_j の初期値：1
 - b_j の初期値：0
- 2 値型の BTM を使用した場合
 - θ_i の初期値：0
 - $a_{j \in d(j)}$ の初期値：1
 - $b_{j \in d(j)}$ の初期値：0
 - $\gamma_{id(j)}$ の初期値：0
 - $\sigma_{\gamma_{d(j)}}^{-2}$ の初期値：3

加えて、本研究においても、抽出された母数ベクトル λ の時系列プロットにより初期値からの影響の有無を確認し、事前に行ったシミュレーションの結果から、分析モデル毎に burn-in を以下のように定めた。

- 2PLM を使用した場合
全ての条件において最初の 1000 組の標本を burn-in とした.
- 2 値型の BTM を使用した場合
全ての条件において最初の 1000 組の標本を burn-in とした.

最後に、抽出する標本の数に関しては、WinBUGS User Manual (Spiegelhalter et al., 2003) に従い、各母数の標本平均の標準誤差が事後分布から抽出された標本の標準偏差の約 5% 以下になるよう、分析モデル毎に以下のように定めた.

- 2PLM を使用した場合
全ての条件において 4000 個の標本を抽出することとした.
- 2 値型の BTM を使用した場合
全ての条件において 5000 個の標本を抽出することとした.

4.3 結果と考察

本研究では、局所依存性が局所独立性を仮定したテスト情報量の推定に対しどのような影響を与えるのか平均的な傾向を把握するため、条件毎に、分析モデルが 2PLM である場合と 2 値型の BTM である場合の \overline{Bias} , \overline{RMSE} を算出し、分析モデル間でこれらの値に関する比較を行った。ただし、 \overline{Bias} は、 $\theta_i = -3.00, -2.75, \dots, 2.75, 3.00$ の計 25 点における $Bias(\hat{I}(\theta_i))$ の値を適当な値で重みづけし、和をとったものとなっている。具体的には、4.2.5 節でも述べたように、どちらの分析モデルにおいても、 θ_i の母集団分布として標準正規分布 $N(0, 1)$ を仮定しているため、 $N(0, 1)$ を各 θ_i の前後 0.125 の区間で (-3.00 と 3.00 に関しては、それぞれ $-\infty$ から -2.875 と 2.875 から ∞ という区間で) 積分したもので重みづけを行っている。したがって、 \overline{Bias} というのは、テスト情報量の推定量が平均的にどの程度の過大あるいは過小推定の傾向を有するのかを表す統計量となっている。また、 \overline{RMSE} は、 \overline{Bias} 同様、 $\theta_i = -3.00, -2.75, \dots, 2.75, 3.00$ における $RMSE(\hat{I}(\theta_i))$ の値を適当な値で重みづけし、和をとったものとなっており、したがって、推定量の平均的な推定誤差を表す統計量となっている。これらの統計量の値を分析モデル間で比較することにより、局所独立性を仮定したテスト情報量の推定に対し局所依存性が与える平均的な影響について検討することが可能となる。

なお、分析モデルが 2PLM である場合と 2 値型の BTM である場合との間で \overline{Bias} や \overline{RMSE} の値について比較を行う際には、その差異の大きさを真のテスト情報量の値 (発生モデルを用いて算出されるテスト情報量の値) に照らして評価することが必要となってくる。そこで、本研究では、条件毎に、 $\theta_i = -3.00, -2.75, \dots, 2.75, 3.00$ における真のテ

スト情報量の値を \overline{Bias} , \overline{RMSE} と同様の重みで重みづけし、この和をとったものを \bar{I} とした。したがって、 \overline{Bias} , \overline{RMSE} の分析モデル間での差異に関しては、この値を \bar{I} で除したものを、真のテスト情報量の値に対する相対的な大きさとして評価できることになる。

4.3.1 真のテスト情報量の値

まず、条件毎に、発生モデルを用いて算出されるテスト情報量の値から \bar{I} の値を算出したところ、表 4.1 が得られた。

4.3.2 分析モデルが 2 値型の BTM の場合

次に、分析モデルが 2 値型の BTM である場合に \overline{Bias} , \overline{RMSE} の値がどのようになるのか、条件毎に算出したところ、表 4.2 が得られた。

4.3.3 分析モデルが 2PLM の場合

続いて、分析モデルが 2PLM である場合に \overline{Bias} , \overline{RMSE} の値がどのようになるのか、条件毎に算出したところ、表 4.3 が得られた。また、分析モデルが 2PLM である場合に、分析モデルが 2 値型の BTM である場合との間で \overline{Bias} , \overline{RMSE} の値がどの程度異なるのか、条件毎に算出したところ、表 4.4 が得られた。ここで、 $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$ はそれぞれ、分析モデルが 2PLM である場合の \overline{Bias} , \overline{RMSE} の値から分析モデルが 2 値型の BTM である場合の値を引いたものとなっている。さらに、分析モデルが 2PLM である場合に、 $\Delta\overline{Bias}$ および $\Delta\overline{RMSE}$ の値が \bar{I} のに比べてどの程度の大きさとなるのか、条件毎に算出したところ、表 4.5 が得られた。ここで、 $\Delta\overline{Bias}/\bar{I}$ および $\Delta\overline{RMSE}/\bar{I}$ はそれぞれ、 $\Delta\overline{Bias}$ および $\Delta\overline{RMSE}$ の値を \bar{I} の値で除したものを表している。加えて、分析モデルが 2PLM である場合に $\Delta\overline{Bias}/\bar{I}$, $\Delta\overline{RMSE}/\bar{I}$ の値が受験者数によりどのように変化するか確認するためグラフを作成したところ、図 4.1, 図 4.2 が得られた。また、分析モデルが 2PLM である場合に $\Delta\overline{Bias}/\bar{I}$, $\Delta\overline{RMSE}/\bar{I}$ の値が局所依存関係にある項目の数によりどのように変化するか確認するためグラフを作成したところ、図 4.3, 図 4.4 が得られた。最後に、分析モデルが 2PLM である場合に $\Delta\overline{Bias}/\bar{I}$, $\Delta\overline{RMSE}/\bar{I}$ の値が 4 つの項目群の局所依存度によりどのように変化するか確認するためグラフを作成したところ、図 4.5, 図 4.6 が得られた。

これらの結果から、分析モデルが 2PLM である場合のバイアス、平均二乗誤差平方根について、以下のようなことが明らかとなった。

バイアスの観点からの評価

表 4.5 より、全ての条件において、分析モデルが 2PLM である場合には、2 値型の BTM である場合に比べて、 \overline{Bias} の値が正の方向に大きくなっており、特に、局所依存度の強い項目群の数が $d(j) \times 2, 1, 0$ と半分以下である場合には、受験者数や局所依存関係にある項目の数に関わらず、全ての条件において、 \overline{Bias} の値が \bar{I} の 0.1 以上大きくなっていった。また、図 4.3 からは、全ての条件において、局所依存関係にある項目の数が増加するにつれて $\Delta\overline{Bias}/\bar{I}$ の値が増加していくことが読み取れるが、図 4.5 からは、受験者数や局所依存関係にある項目の数に関わらず、どの条件においても、局所依存度の強い項目群の数が増加するにつれて $\Delta\overline{Bias}/\bar{I}$ の値が減少していくことが読み取れ、 \overline{Bias} の値に関しては、項目間の局所依存度を減らすことが局所依存性の影響の減少に結びつかないことが明らかとなった。さらに、図 4.1 より、どの条件においても、 $\Delta\overline{Bias}/\bar{I}$ の値に関して、受験者数の増加に伴う大きな変動は確認されず、受験者数を増やすことも、 \overline{Bias} に対する局所依存性の影響を低下させることには結びつかないことが明らかとなった。

平均二乗誤差平方根の観点からの評価

表 4.5 より、全ての条件において、分析モデルが 2PLM である場合には、2 値型の BTM である場合に比べて、 \overline{RMSE} の値が正の方向に大きくなっており、特に、局所依存度の強い項目群の数が $d(j) \times 2, 1, 0$ とテスト全体の半分以下である場合には、受験者数や局所依存関係にある項目の数に関わらず、全ての条件において、 \overline{RMSE} の値が \bar{I} の 0.1 以上大きくなっていった。また、図 4.4 からは、全ての条件において、局所依存関係にある項目の数が増加するにつれて $\Delta\overline{RMSE}/\bar{I}$ の値が増加していくことが読み取れるが、図 4.6 からは、受験者数や局所依存関係にある項目の数に関わらず、どの条件においても、局所依存度の強い項目群の数が増加するのに伴い $\Delta\overline{RMSE}/\bar{I}$ の値が減少していくことが読み取れ、 \overline{RMSE} の値に関しては、項目間の局所依存度の減少が局所依存性の影響の減少には結びつかないことが明らかとなった。さらに、図 4.2 より、どの条件においても、 $\Delta\overline{RMSE}/\bar{I}$ の値は受験者数の増加に伴いほとんど変動しておらず、受験者数を増やすことも、 \overline{RMSE} に対する局所依存性の影響を減少させることには結びつかないということが明らかとなった。

まとめ

4.2.3 節でも述べたように、項目反応理論を用いてテストを作成、評価する場合には、一般的に、受験者特性の全域にわたってテスト情報量の値が高くなるようにテストが構成されるため、このような文脈においては、バイアスや平均二乗誤差平方根の観点からの評価が重要となってくる。そのような観点から上記の結果を見直してみると、全ての条件に

において、局所依存性の影響は無視できないものとなっており、特に、 $d(j) \times 2, 1, 0$ のように局所依存度の強い項目群がテスト全体に占める割合が半分以下である場合には、局所依存性に対する局所独立性を仮定したテスト情報量の推定は頑健であるとは言い難い。

また、一般的には、受験者数を増やしたり項目間の局所依存度を減少させることにより、局所依存性に対する頑健性を増すことができると考えられるのだが、上記の結果からは、受験者数を増やすことは局所依存性に対する頑健性を増すことにはつながらず、項目間の局所依存度を低下させた場合には、かえって、局所依存性の影響が強くなるということが明らかとなった。

4.3.4 先行研究との関連

4.1 節でも述べたように、本研究では、局所独立性を仮定したテスト情報量の推定に対する局所依存性の影響について、真のテスト情報量との比較という観点から検討を行うため、シミュレーションを実施した。また、推定量のバイアス、平均二乗誤差平方根と受験者数等の要因との関係についても検討を加えるため、シミュレーションを実施する際に、受験者数、局所依存関係にある項目の数、4つの項目群の局所依存度を系統的に変化させ、局所依存性の影響とこれらの要因との関連について検討を行った。これらの点に関しては、実データの解析を通して局所依存性の影響について検討した Ip (2010)、石塚他 (2001)、Keller et al. (2003)、Wainer & Wang (2000) の研究との間で相違が見られるのだが、以下の点に関しては、本研究との間で、類似した結果が得られている。

- 局所依存性を有するデータに対し局所独立性を仮定してテスト情報量の推定を行うと、局所依存性を考慮した場合に比べて、テスト情報量が過大推定される

表 4.1 \bar{I} の値

条件	\bar{I}
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	6.645
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	6.322
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	5.752
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	5.140
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	4.519
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	8.035
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	7.836
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	7.519
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	7.155
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	6.834
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	8.833
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	8.658
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	8.389
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	8.217
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	8.010
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	6.645
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	6.322
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	5.752
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	5.140
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	4.519
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	8.035
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	7.836
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	7.519
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	7.155
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	6.834
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	8.833
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	8.658
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	8.389
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	8.217
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	8.010

表 4.2 分析モデルが 2 値型の BTM である場合の \overline{Bias} , \overline{RMSE} の値

条件	\overline{Bias}	\overline{RMSE}
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	1.252	1.291
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	1.066	1.047
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.852	0.889
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.546	0.592
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.311	0.366
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.977	1.053
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.855	0.925
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.670	0.741
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.525	0.599
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.416	0.492
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.712	0.830
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.691	0.803
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.533	0.648
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.427	0.552
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.436	0.530
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.458	0.501
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.267	0.325
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.143	0.215
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.216	0.274
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.005	0.160
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.424	0.487
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.382	0.437
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.295	0.360
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.374	0.428
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.273	0.329
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.312	0.394
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.310	0.380
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.244	0.327
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.248	0.325
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.175	0.258

表 4.3 分析モデルが 2PLM である場合の \overline{Bias} , \overline{RMSE} の値

条件	\overline{Bias}	\overline{RMSE}
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	3.140	3.148
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	3.562	3.571
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	4.406	4.414
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	5.021	5.028
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	5.767	5.773
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	1.833	1.850
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	2.043	2.060
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	2.507	2.523
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	2.907	2.921
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	3.347	3.359
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	1.168	1.195
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	1.364	1.389
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	1.764	1.767
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	1.955	1.977
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	2.322	2.342
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	2.425	2.428
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	2.844	2.847
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	3.716	3.718
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	4.629	4.632
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	5.488	5.490
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	1.158	1.165
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	1.464	1.471
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	2.071	2.076
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	2.699	2.703
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	3.236	3.240
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.490	0.515
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.781	0.799
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	1.323	1.331
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	1.665	1.673
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	2.073	2.080

表 4.4 分析モデルが 2PLM である場合の $\overline{\Delta Bias}$, $\overline{\Delta RMSE}$ の値

条件	$\overline{\Delta Bias}$	$\overline{\Delta RMSE}$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	1.888	1.858
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	2.556	2.524
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	3.553	3.525
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	4.475	4.436
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	5.456	5.407
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.856	0.798
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	1.888	1.135
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	1.837	1.781
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	2.383	2.321
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	2.932	2.867
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.456	0.365
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.673	0.586
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	1.212	1.119
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	1.528	1.425
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	1.866	1.812
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	1.967	1.927
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	2.577	2.522
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	3.572	3.503
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	4.413	4.358
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	5.483	5.330
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.733	0.679
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	1.082	1.034
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	1.776	1.715
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	2.326	2.275
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	2.932	2.911
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.177	0.121
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.470	0.419
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	1.079	1.004
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	1.417	1.348
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	1.899	1.822

表 4.5 分析モデルが 2PLM である場合の $\Delta\overline{Bias}/\bar{I}$, $\Delta\overline{RMSE}/\bar{I}$ の値

条件	$\Delta\overline{Bias}/\bar{I}$	$\Delta\overline{RMSE}/\bar{I}$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.284	0.280
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.404	0.399
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.618	0.613
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.870	0.863
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	1.207	1.196
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.106	0.099
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.152	0.145
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.244	0.237
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.333	0.324
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.429	0.420
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.052	0.041
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.078	0.068
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.145	0.133
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.186	0.173
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.235	0.226
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.296	0.290
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.408	0.399
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.621	0.609
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.859	0.848
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	1.213	1.179
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.091	0.084
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.138	0.132
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.236	0.228
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.325	0.318
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.434	0.426
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.020	0.014
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.054	0.048
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.129	0.120
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.172	0.164
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.237	0.227

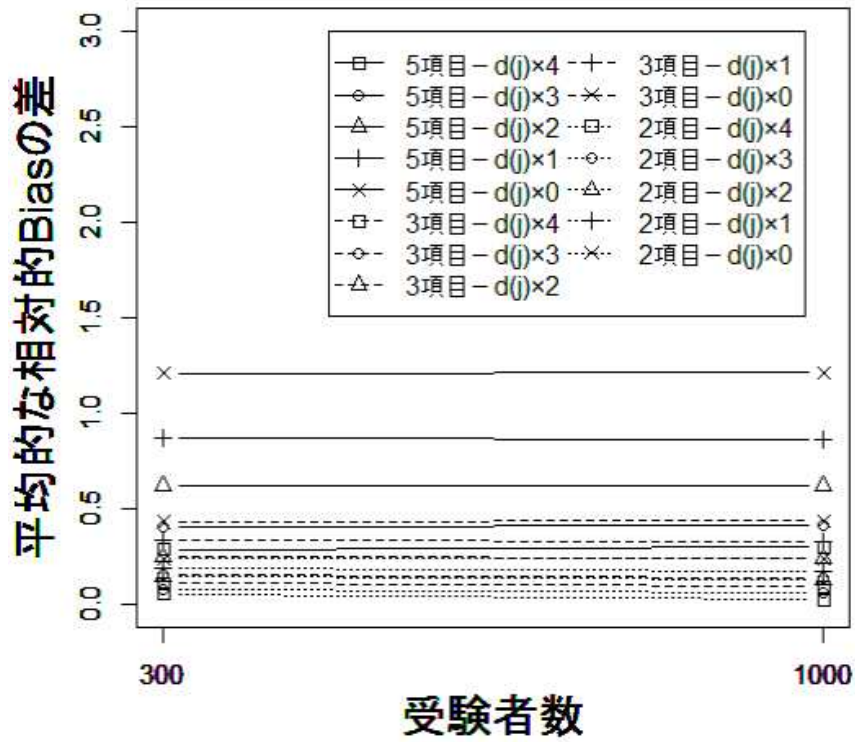


図 4.1 $\overline{\Delta Bias/I}$ と受験者数との関係

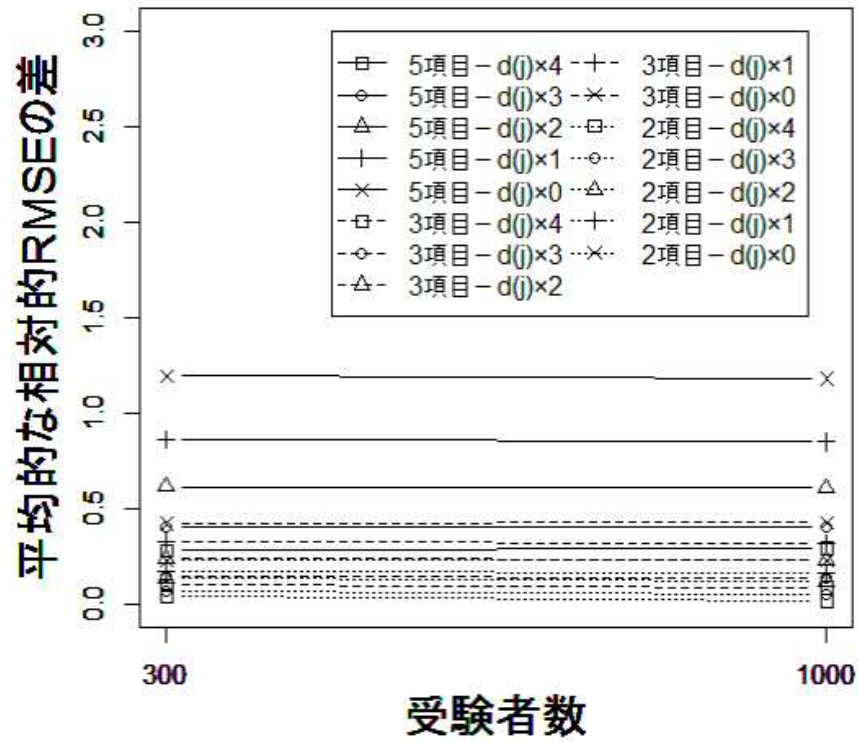


図 4.2 $\overline{\Delta RMSE/I}$ と受験者数との関係

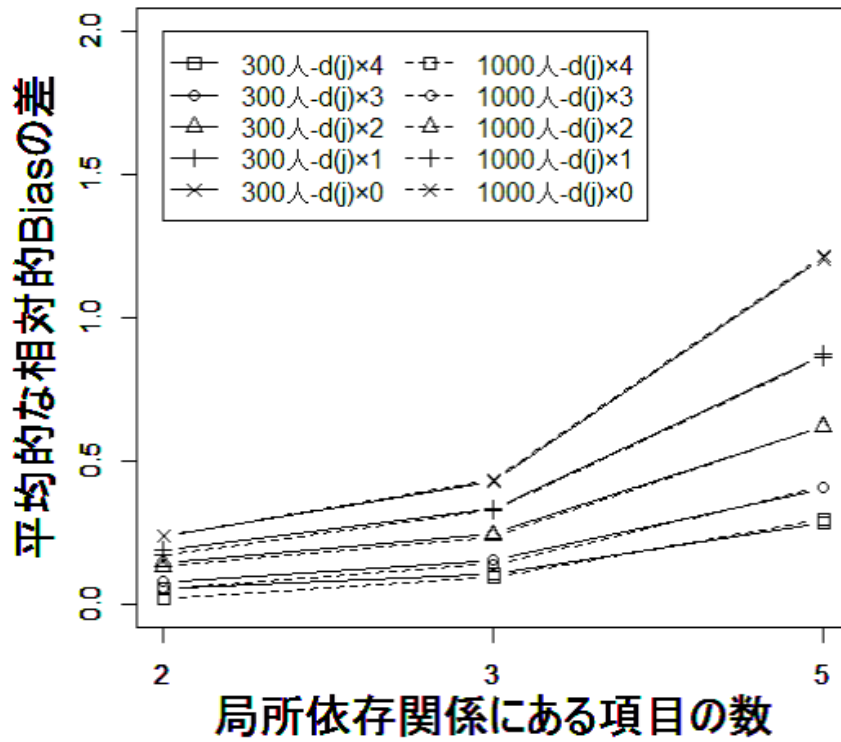


図 4.3 $\overline{\Delta Bias}/\bar{I}$ と局所依存関係にある項目の数との関係

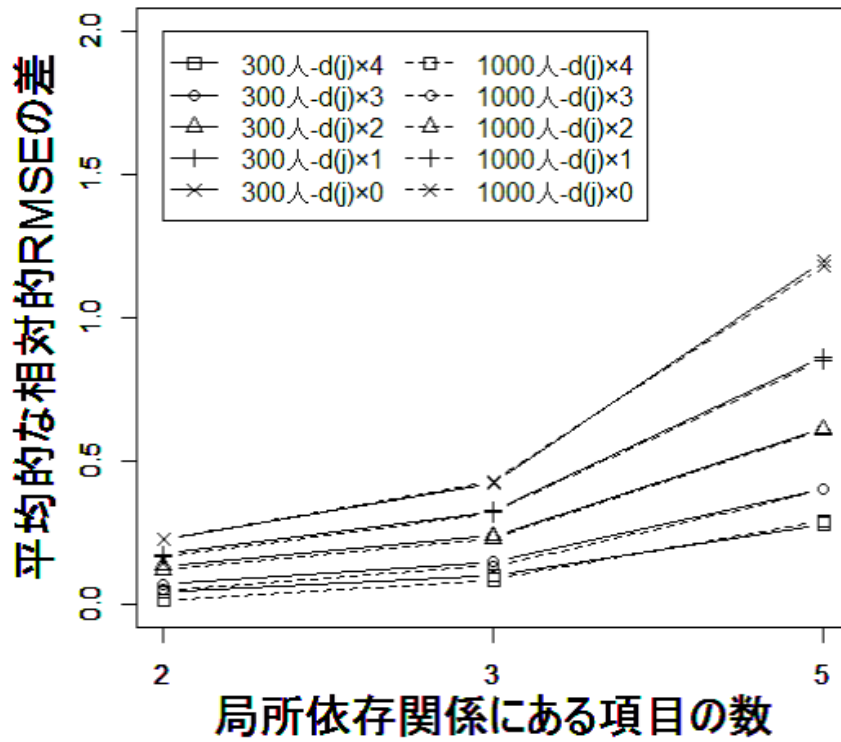


図 4.4 $\overline{\Delta RMSE}/\bar{I}$ と局所依存関係にある項目の数との関係

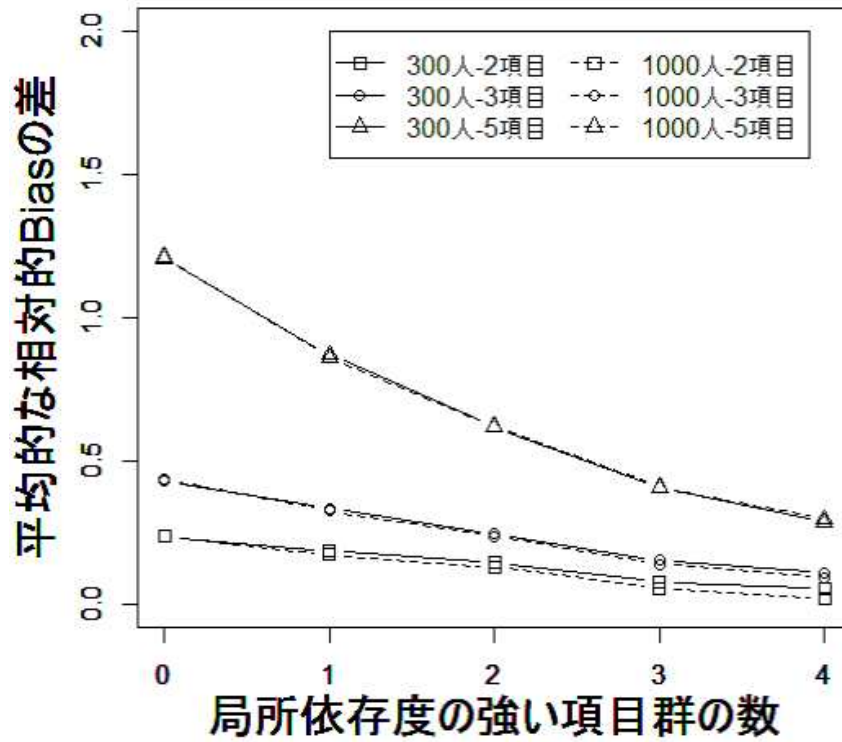


図 4.5 $\overline{\Delta Bias/I}$ と 4 つの項目群の局所依存度との関係

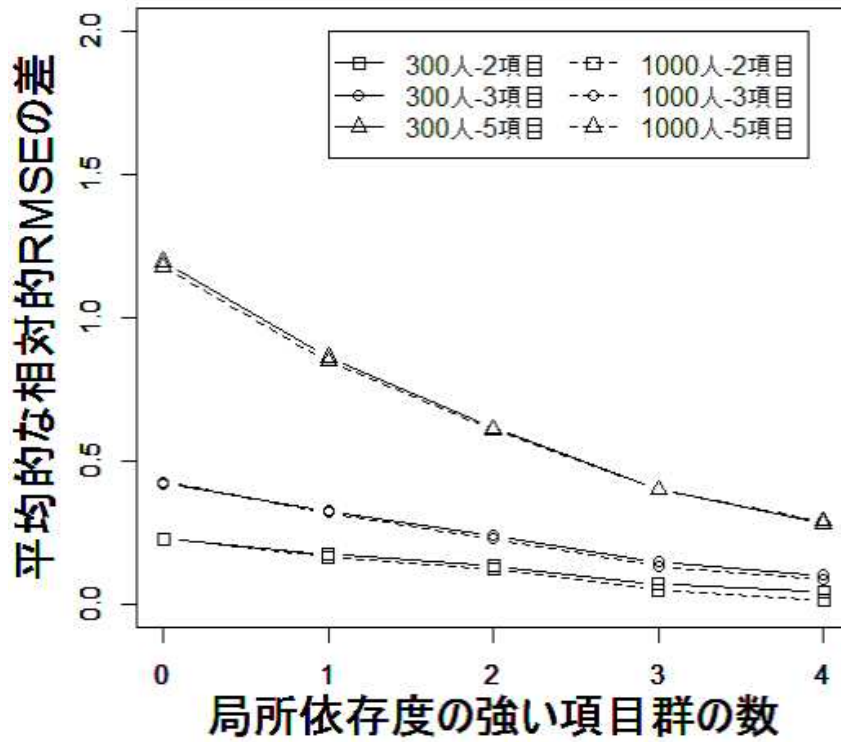


図 4.6 $\overline{\Delta RMSE/I}$ と 4 つの項目群の局所依存度との関係

第 5 章

局所依存モデル間での受験者特性の推定精度の比較

5.1 問題と目的

本章では、局所依存性を考慮することにより受験者特性の推定精度がどの程度改善されるのか、もし改善されるのであれば、それは局所依存性を考慮したモデルを使用するに値するほどのものであるのか、シミュレーションを通じた検討を行う。また、局所依存性を考慮したモデルを使用したほうが良いという結果が得られた場合には、それがどのような条件の下でのことであり、その改善度合いは使用するモデルによっても異なるのかどうか、検討を行うこととする*¹。このテーマに関連して、局所依存性を考慮したモデルや局所独立性を仮定したモデルとの間で比較を行った研究としては、Keller et al. (2003), Lee (2000), Lee, Kolen, Frisbie, & Ankenmann (2001), Reise, Horan, & blanchard (2011), Tuerlinckx & De Boeck (1999), Wainer & Wang (2000), Wang, Cheng, & Wilson (2005), Zhang (2010) といったものが挙げられるのだが、特に、受験者特性の推定精度の比較を行った研究としては、Braedlow et al. (1999), DeMars (2006), 石塚他 (2001), 泉・山野井・山田・白川・対馬 (2013), Wainer et al. (2007, pp. 137-140) といったものが挙げられる。例えば、Bradlow et al. (1999) では、2 値型の BTM を発生モデルとし、2PLM と BTM を分析モデルとしてシミュレーションを実施した結果、

- 分析モデルが BTM である場合の方が、2PLM である場合よりも、推定量にかかるバイアスが小さくなる
- 分析モデルが BTM である場合の方が、2PLM である場合よりも、受験者特性に関する M95%PIW の幅が大きくなる

*¹ 本章は、登藤 (2012a) を加筆、再構成したものである。

- 分析モデルが BTM である場合の方が, 2PLM である場合よりも, M95%PIW が実際に受験者特性の真値を含む確率が大きくなる
- 分析モデルが BTM である場合と 2PLM である場合との間で, 受験者特性の推定値と真値との順位相関係数ほとんど変わらない

ことが報告されている。また, DeMars (2006) においては, 局所依存性を考慮するタイプ b のモデルから局所依存性を有するデータを発生させて, タイプ a のモデル, タイプ b のモデル, 局所独立性を仮定するモデルの間で受験者特性の推定精度を比較したところ, 分析モデル間で, 推定量のバイアス, 平均二乗誤差平方根に関しほとんど差は見られなかったことが報告されている。さらに, 石塚他 (2001) においては, 2000 年度に実施された大学入試センター試験英語の長文読解問題に対して GRM と 2PLM をあてはめ受験者特性値の推定を行ったところ, 推定値間の相関係数が 0.987 となったことが報告されており, 泉他 (2013) においては, 局所依存性が生じていると考えられる複数のテストデータに対して 2PLM と GRM をあてはめて受験者特性の推定を行った結果, 推定値間の相関係数が 0.973 から 0.995 となったことが報告されている。加えて, Wainer et al. (2007, pp. 137-140) においては, タイプ b のモデルから局所依存性を有するデータを発生させ, タイプ b のモデルと局所独立性を仮定するモデルとの間で受験者特性の推定値と真値との相関係数に関して比較を行ったところ, 分析モデル間で, 相関係数の値にほとんど差が見られなかったことが報告されている。

以上のように, 先行研究においては, 局所依存性を考慮することにより受験者特性の推定精度がどの程度改善されるのかが, シミュレーションや実データの分析を通して検討されてきた。1.5 節でも述べたように, 局所依存性を考慮したモデルにはタイプ a, タイプ b, タイプ c という計 3 タイプのモデルが存在し, これらのモデルは互いに局所依存性への対処法が異なっているため, 局所依存性を有するデータに対しこれら 3 タイプのモデルをあてはめて受験者特性の推定を行った場合, 得られる推定量の性質もタイプ間で異なるものになると考えられる。したがって, 今後大問形式の問題を含むテストの項目反応理論に基づく作成, 評価, 運用が増加していくであろうことを考慮すると, これら 3 タイプのモデルと局所独立性を仮定したモデル (以下「タイプ d」と呼ぶ) との間で, 受験者特性の推定精度に関して比較を行う必要があると考えられる。しかしながら, 現在までのところ, これら 4 タイプのモデル間で受験者特性の推定精度に関し比較, 検討を行った研究は見当たらない。さらに, 本研究で得られた知見を実際にテストを作成, 評価, 運用する場面へ応用することを見据えると, 受験者数等の要因を系統的に変化させてシミュレーションを行い, 局所依存性を考慮することによる推定精度の改善がこれらの要因とどのような関係にあるのか, 検討する必要もあると考えられる。

そこで, 本章では, 局所依存性の生じているデータに対して局所依存性を考慮したモデ

ル (タイプ a, タイプ b, タイプ c) と局所独立性を仮定したモデル (タイプ d) をあてはめて受験者特性の推定を行うシミュレーションを実施し, 局所依存性を考慮することにより受験者特性の推定精度がどの程度改善されるのか, 検討を行った. また, シミュレーションを実施する際に受験者数等の要因を系統的に変化させ, 受験者特性の推定精度の改善がこれらの要因とどのような関係にあるのか, 検討を行った.

5.2 方法

本章では, 局所依存性を有する仮想的な項目反応データを発生させ, これに対しタイプ a からタイプ d までの項目反応モデルをあてはめて母数の推定を行った. そして, この手続きを多数回繰り返し, 得られた多数の $\hat{\theta}_{ir}$ から $\hat{\theta}_i$ の $Bias(\hat{\theta}_i)$, $RMSE(\hat{\theta}_i)$ と $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ を算出して, モデル間でこれらの統計量に関する比較を行った.

5.2.1 データの発生モデル

本研究においても, 4.2.1 節で述べたのと同様の理由から, 局所依存性を有するデータの発生モデルとして (1.14) 式の 2 値型の BTM を使用した (以下にモデル式を再掲).

$$P_{j \in d(j)}(\theta_i) = \frac{1}{1 + \exp[-1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)} - \gamma_{id(j)})]} \quad (5.1)$$

5.2.2 データの分析モデル

局所依存性を有するデータを分析する際には, タイプ a からタイプ d のモデルとして, それぞれ以下のモデルを使用した.

タイプ a の分析モデル

局所依存性を有するデータを分析する際に用いるタイプ a のモデルとしては, (1.12) 式の GRM を使用した (以下にモデル式を再掲).

$$P_{d(j)}(r|\theta_i) = \frac{1}{1 + \exp[-1.7a_{d(j)}^*(\theta_i - b_r^*)]} - \frac{1}{1 + \exp[-1.7a_{d(j)}^*(\theta_i - b_{r+1}^*)]} \quad (5.2)$$

タイプ b の分析モデル

局所依存性を有するデータを分析する際に用いるタイプ b のモデルとしては, データの発生モデルでもある 2 値型の BTM を使用した (モデル式については (5.1) 式を参照のこと).

タイプ c の分析モデル

局所依存性を有するデータを分析する際に用いるタイプ c のモデルとしては, (1.16) 式の CCM を使用した (以下にモデル式を再掲).

$$P(U_j, U_k | \theta_i) = \frac{\exp [U_j Z_j + U_k Z_k - U_j U_k b_{jk}]}{1 + \exp [Z_j] + \exp [Z_k] + \exp [Z_j + Z_k - b_{jk}]} \quad (5.3)$$

$$Z_j = 1.7a_{j \in d(j)}(\theta_i - b_{j \in d(j)}) \quad (5.4)$$

$$Z_k = 1.7a_{k \in d(j)}(\theta_i - b_{k \in d(j)}) \quad (5.5)$$

タイプ d の分析モデル

局所依存性を有するデータを分析する際に用いるタイプ d のモデルとしては, (1.1) 式の 2PLM を使用した (以下にモデル式を再掲).

$$P_j(\theta_i) = \frac{1}{1 + \exp [-1.7a_j(\theta_i - b_j)]} \quad (5.6)$$

5.2.3 シミュレーションの手続き

本研究では, データを発生させる際, テストは局所依存関係にある 4 つの項目群から構成されており, 各項目群は 5 つの項目から構成されていると仮定した. また, 5.1 節でも述べたように, 局所依存性を考慮することによる推定精度の改善が種々の要因とどのような関係にあるのか調べるため, 本研究では, 受験者数, 局所依存関係にある項目の数, 4 つの項目群の局所依存度を系統的に変化させ, シミュレーションを行った. 受験者数に関しては 300 人, 1000 人という 2 水準を設定し, 局所依存関係にある項目の数に関しては, 項目群内で実際に局所依存関係にある項目の数に関して以下の 3 つの水準を設定した.

- 5 項目が局所依存関係にある ($J_{d(j)} = 5$)
- 3 項目が局所依存関係にある ($J_{d(j)} = 3$)
- 2 項目が局所依存関係にある ($J_{d(j)} = 2$)

また, 項目間の局所依存度に関しては, テストに含まれる 4 つの項目群の局所依存度に関して以下の 5 つの水準を設定した.

- 4 つとも局所依存度が強い (強 $d(j) \times 4$)
- 4 つのうち 3 つの局所依存度が強い (強 $d(j) \times 3$)
- 4 つのうち 2 つの局所依存度が強い (強 $d(j) \times 2$)
- 4 つのうち 1 つだけ局所依存度が強い (強 $d(j) \times 1$)

- 4つとも局所依存度が弱い (強 $d(j) \times 0$)

ここで、項目間の局所依存度の強弱に関しては、3章と同様の基準を用いて操作を行った。本研究で行ったシミュレーションの詳細は以下のようにまとめられる*2。

1. 各受験者の θ_i ($i = 1, 2, \dots, N$) と $\gamma_{id(j)}$ ($i = 1, 2, \dots, N; d(j) = 1, 2, 3, 4$) および各項目の $a_{j \in d(j)}$, $b_{j \in d(j)}$ ($j = 1, 2, \dots, 20$) の真値をそれぞれ標準正規分布 $N(0, 1)$, 正規分布 $N(0, \sigma_{\gamma_{d(j)}}^2)$, 一様分布 $U(0.5, 1.5)$, 標準正規分布 $N(0, 1)$ から発生させる*3。
2. 各受験者の各項目に対する正答確率を手続き 5.3 で得られた各母数の真値と 2 値型の BTM の式 ((5.1) 式) から算出し、 $N \times 20$ の正答確率行列 \mathbf{A} を作成する。
3. 一様分布 $U(0, 1)$ から $N \times 20$ 個の乱数を発生させ、 $N \times 20$ の一様乱数行列 \mathbf{B} を作成する。
4. \mathbf{A} , \mathbf{B} の各要素を比較して局所依存性を有するデータ \mathbf{U} を作成する。具体的には、 $a_{ij} \geq b_{ij}$ である場合には $u_{ij} = 1$ とし、 $a_{ij} < b_{ij}$ である場合には $u_{ij} = 0$ とする。
5. \mathbf{U} に対して 5.2.2 節で紹介した 4 つの分析モデルを当てはめて母数の推定を行う。
6. 手続き 3 から 5 までを R 回繰り返す。
7. 手続き 6 の結果得られた R 組の $\hat{\theta}_{ir}$ を用いて、分析モデル毎に、 $Bias(\hat{\theta}_i)$, $RMSE(\hat{\theta}_i)$, $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値を算出する (各統計量の計算式に関しては以下に再掲)。

$$Bias(\hat{\theta}_i) = \frac{1}{R} \sum_{r=1}^R \hat{\theta}_{ir} - \theta_i \quad (5.7)$$

$$RMSE(\hat{\theta}_i) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_{ir} - \theta_i)^2} \quad (5.8)$$

$$cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \frac{1}{R} \sum_{r=1}^R cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_r) \quad (5.9)$$

なお、手続き 6 における反復回数 R に関しては、繰り返し毎の $\hat{\theta}_{ir}$ の変動を考慮して、受験者数が 1000 人の場合には $R = 50$ とし、受験者数が 300 人である場合には $R = 100$ とした。また、タイプ c のモデルは、4.2.1 節でも述べたように、基本的に 2 つの項目が局所依存関係にある場合を想定したモデルであると考えられる。そのため、タイプ c の分析モデルに関しては、実際に使用され得る文脈等を考慮して、 $J_{d(j)} = 2$ である条件に限り、分析モデルとして使用することとした。

*2 シミュレーションプログラムの記述および実行には統計解析向けのプログラミング言語でありその実行環境でもある、R を利用した。

*3 なお、擬似乱数を発生させる際には R に実装されているメルセンヌツイスターを使用した。

5.2.4 母数の推定法

本研究においても、手続き5における母数の推定はMCMCにより行った。MCMCの具体的なアルゴリズムとしてはスライスサンプリング (Neal, 1997) を利用し、WinBUGS 1.4 (Spiegelhalter et al., 2003) を利用してMCMCを実行した。また、事前には各母数が統計的に互いに独立であったと仮定し、分析モデル毎に、各母数の事前分布を以下のように設定した。

- タイプ a の場合
 - θ_i の事前分布：標準正規分布 $N(0, 1)$
 - $a_{d(j)}$ の事前分布：正規分布 $N(1, 0.25)$
 - b_r^* の事前分布：標準正規分布 $N(0, 1)$
- タイプ b の場合
 - θ_i の事前分布：標準正規分布 $N(0, 1)$
 - $a_{j \in d(j)}$ の事前分布：正規分布 $N(1, 0.25)$
 - $b_{j \in d(j)}$ の事前分布：標準正規分布 $N(0, 1)$
 - $\gamma_{d(j)}$ の事前分布：正規分布 $N(0, \sigma_{\gamma_{d(j)}}^2)$
 - $\sigma_{\gamma_{d(j)}}^{-2}$ ：ガンマ分布 $\Gamma(3, 1)$
- タイプ c の場合
 - θ_i の事前分布：標準正規分布 $N(0, 1)$
 - $a_{j \in d(j)}$ の事前分布：正規分布 $N(1, 0.25)$
 - $b_{j \in d(j)}$ の事前分布：標準正規分布 $N(0, 1)$
 - b_{jk} の事前分布：正規分布 $N(-2, 1)$
- タイプ d の場合
 - θ_i の事前分布：標準正規分布 $N(0, 1)$
 - a_j の事前分布：正規分布 $N(1, 0.25)$
 - b_j の事前分布：標準正規分布 $N(0, 1)$

さらに、各母数の初期値に関しては、分析モデル毎に、以下の様に設定した。

- タイプ a の場合
 - θ_i の初期値：0
 - $a_{d(j)}$ の初期値：1
 - b_r^* の初期値： $r = 1, 2, 3, 4, 5$ に対して、それぞれ 0, 0.1, 0.2, 0.3, 0.4
- タイプ b の場合

- θ_i の初期値 : 0
- $a_{j \in d(j)}$ の初期値 : 1
- $b_{j \in d(j)}$ の初期値 : 0
- $\gamma_{id(j)}$ の初期値 : 0
- $\sigma_{\gamma_{d(j)}}^{-2}$ の初期値 : 3
- タイプ c の場合
 - θ_i の初期値 : 0
 - $a_{j \in d(j)}$ の初期値 : 1
 - $b_{j \in d(j)}$ の初期値 : 0
 - b_{jk} の初期値 : -2
- タイプ d の場合
 - θ_i の初期値 : 0
 - a_j の初期値 : 1
 - b_j の初期値 : 0

加えて、本研究においても、抽出された母数ベクトル λ の時系列プロットにより初期値からの影響の有無を確認し、事前に行ったシミュレーションの結果から、分析モデル毎に burn-in を以下のように定めた。

- タイプ a の場合
 - 全ての条件において最初の 1000 組の標本を burn-in とした。
- タイプ b の場合
 - 全ての条件において最初の 1000 組の標本を burn-in とした。
- タイプ c の場合適用された全ての条件において最初の 1000 組の標本を burn-in とした。
- タイプ d の場合
 - 全ての条件において最初の 1000 組の標本を burn-in とした。

最後に、抽出する標本の数に関しては、WinBUGS User Manual (Spiegelhalter et al., 2003) に従い、各母数の標本平均の標準誤差が事後分布から抽出された標本の標準偏差の約 5% 以下になるよう、分析モデル毎に以下のように定めた。

- タイプ a の場合
 - $J_{d(j)} = 5$ である場合には 5000 個の標本を抽出し、それ以外の条件では 4000 個の標本を抽出することとした。
- タイプ b の場合
 - 全ての条件において 5000 個の標本を抽出することとした。

- タイプ c の場合
適用された全ての条件において 4000 個の標本を抽出することとした。
- タイプ d の場合
全ての条件において 4000 個の標本を抽出することとした。

5.3 結果と考察

本研究では、局所依存性を考慮することにより受験者特性の推定精度がどれほど改善されるのか検討するため、条件毎に、分析モデルが 2PLM である場合と GRM, BTM, CCM である場合との間で \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ に関する比較を行った。

なお、分析モデルが 2PLM である場合と GRM, BTM, CCM である場合との間で \overline{Bias} や \overline{RMSE} の値について比較を行う際には、その差異の大きさを母集団分布における θ_i の標準偏差に照らして評価することが必要となってくる。5.2.3 節の手続きでも述べたように、本研究では、 θ_i の母集団分布として $N(0, 1)$ を仮定しており、母集団分布における θ_i の標準偏差は 1 となっている。したがって、 \overline{Bias} と \overline{RMSE} の分析モデル間での差異については、この値を、そのまま θ_i の母標準偏差に対する相対的な大きさとして評価できることになる。

5.3.1 分析モデルが 2PLM の場合

まず、分析モデルが 2PLM である場合に \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値がどのようになるのか、条件毎に算出したところ、表 5.1 が得られた。

5.3.2 分析モデルが GRM の場合

次に、分析モデルが GRM である場合に \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値がどのようになるのか、条件毎に算出したところ、表 5.2 が得られた。また、分析モデルが GRM である場合に $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値がどのようになるのか、条件毎に算出したところ、表 5.3 が得られた。ここで、 $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ はそれぞれ、分析モデルが GRM である場合の \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値から分析モデルが 2PLM である場合の \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値を減じたものとなっている。さらに、分析モデルが GRM である場合に $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が受験者数によりどのように変化するか確認するためグラフを作成したところ、図 5.1, 図 5.2, 図 5.3 が得られた。加えて、分析モデルが GRM である場合に $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が局所依存関係にある項目の数によりどのように変化するか確認するためグラフを作成した

ところ、図 5.4, 図 5.5, 図 5.6 が得られた。最後に、分析モデルが GRM である場合に $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta\text{cor}(\theta, \hat{\theta})$ の値が 4 つの項目群の局所依存度によりどのように変化するのかが確認するためグラフを作成したところ、図 5.7, 図 5.8, 図 5.9 が得られた。

これらの結果から、分析モデルが GRM である場合のバイアス、平均二乗誤差平方根、真値と推定値との相関係数について、以下のようなことが明らかとなった。

バイアスの観点からの評価

表 5.7 より、条件によっては、分析モデルが 2PLM である時のほうが推定量にかかるバイアスの値が小さくなることもあるが、全ての条件において、分析モデルが 2PLM である場合と GRM である場合との間で、受験者特性値の推定量にかかるバイアスの値にほとんど差は見られなかった。また、図 5.1 からは、受験者数の増加に伴い、 $\Delta\overline{Bias}$ の値が増加する傾向が確認された。これは、受験者数の増加に伴い局所独立性を仮定した 2PLM の頑健性が向上し、バイアスの値が減少する一方で、局所依存性を考慮した GRM においては、バイアスの値に関して、それほどの減少は見られなかったためである。

平均二乗誤差平方根の観点からの評価

表 5.7 より、条件によっては、分析モデルが 2PLM である時のほうが推定量の推定誤差の大きさが小さくなることもあるのだが、全ての条件において、分析モデルが 2PLM である場合と GRM である場合との間で、推定量の推定誤差の大きさにはほとんど差が見られなかった。また、図 5.8 からは、項目群内で局所依存関係にある項目の数が 2 項目や 3 項目である場合、局所依存度の強い項目群が増加するにつれて $\Delta\overline{RMSE}$ の値が減少し、項目群内で局所依存関係にある項目の数が 5 項目である場合には、項目群間で局所依存度が異なる場合には、等しい場合に比べて、 $\Delta\overline{RMSE}$ の値が小さくなっている様子が見える。局所依存関係にある項目が 2 項目や 3 項目である場合には、局所依存度の強い項目群の増加に伴う 2PLM の \overline{RMSE} の値の上昇が、GRM における値の上昇に比べて強かったため、 $\Delta\overline{RMSE}$ の減少が生じたと考えられ、局所依存関係にある項目が 5 項目である場合には、表 5.7 にもあるように、4 つの項目群の局所依存度が異なる時の 2PLM の推定値と真値との相関係数が、GRM のものに比べて、他の条件における場合よりも小さくなるため、2PLM を使用した場合の推定誤差が GRM を使用した場合に比べてより大きくなったためであると考えられた。

真値と推定値との相関係数の観点からの評価

表 5.7 より、条件によっては、分析モデルが 2PLM である時のほうが推定値と真値との相関係数が大きくなっているのだが、全ての条件において、分析モデルが 2PLM である場合と GRM である場合との間で、受験者特性の推定値と真値との相関係数に関してほとん

ど差は見られなかった。また、図 5.9 からは、局所依存関係にある項目の数が 2 項目や 3 項目である場合には、局所依存度の強い項目群の数が増加するにつれて、 $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が上昇していく様子が見える。これは、局所依存度の強い項目群の増加に伴う $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の減少が、GRM に比べて 2PLM においてより強く見られるためである。さらに、図 5.9 からは、局所依存関係にある項目の数が 5 項目である場合に、4 つの項目群の局所依存度が等しい時に比べて 4 つの項目群の局所依存度が異なる時、 $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が大きくなっていることが見える。これは、4 つの項目群の局所依存度が全て等しい場合には、全ての項目特性の推定値に同程度のバイアスが生じるため、2PLM を使用した場合のほうが、項目反応パタンの情報を利用しない GRM に比べて相関係数の値が高くなるものの、局所依存度が 4 つの項目群の間で異なる場合には、項目群毎に項目特性にかかるバイアスの程度が異なってくるため、項目母数と共に推定される受験者特性の線形性が保たれなくなり、局所依存関係にある項目が 5 項目である場合には、その程度が他の条件に比べて高くなるためだと考えられる。

まとめ

特定のテストの成績のみを用いて個人差の評価を行うという文脈においては、受験者特性値の真値と推定値との相関係数という観点から評価を行うことが重要となってくる。そういった観点から上記の結果を見直してみると、真値と推定値との相関係数に関して分析モデル間ではほとんど差が見られず、本研究においては、局所独立性を仮定したモデルに代えてより複雑な局所依存性を考慮したタイプ a のモデルの利用を積極的に肯定するような結果は得られなかったといえる。

また、特定のテストを超えて、各受験者の受験者特性値を共通の尺度上で比較するという文脈においては、バイアスおよび平均二乗誤差平方根という観点からの評価も重要になってくる。そういった観点から上記の結果を見直してみると、分析モデル間でこれらの統計量の値にほとんど差は見られず、局所独立性を仮定したモデルに代えてタイプ a のモデルを利用することを積極的に推奨するような結果は得られなかった。

5.3.3 分析モデルが 2 値型の BTM の場合

続いて、分析モデルが BTM である場合に \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値がどのようになるのか、条件毎に算出したところ、表 5.4 が得られた。また、分析モデルが BTM である場合に $\Delta \overline{Bias}$, $\Delta \overline{RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値がどのようになるのか、条件毎に算出したところ、表 5.5 が得られた。さらに、分析モデルが BTM である場合に $\Delta \overline{Bias}$, $\Delta \overline{RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が受験者数によりどのように変化するか確認するためグラフを作成したところ、図 5.10、図 5.11、図 5.12 が得られた。加えて、

分析モデルが BTM である場合に $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta\text{cor}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が局所依存関係にある項目の数によりどのように変化するのかが確認するためグラフを作成したところ、図 5.13, 図 5.14, 図 5.15 が得られた。最後に、分析モデルが BTM である場合に $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta\text{cor}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値が4つの項目群の局所依存度によりどのように変化するのかが確認するためグラフを作成したところ、図 5.16, 図 5.17, 図 5.18 が得られた。

これらの結果から、分析モデルが BTM である場合のバイアス、平均二乗誤差平方根、真値と推定値との相関係数について、以下のようなことが明らかとなった。

バイアスの観点からの評価

表 5.9 より、受験者数が $N = 1000$ と多い場合には、2PLM を使用した場合のバイアスが2値型の BTM を使用した場合のバイアスよりも小さくなっているものの、全ての条件において、分析モデルが 2PLM である場合と2値型の BTM である場合との間で、推定量にかかるバイアスの値にほとんど差は見られなかった。また、図 5.10 からは、受験者数の増加に伴い、 $\Delta\overline{Bias}$ の値が増加している様子がうかがえる。これは、受験者数の増加に伴って、局所独立性を仮定する 2PLM の頑健性が上昇し、バイアスの値が減少する一方、局所依存性を考慮する2値型の BTM においてはそのような変化が見られなかったためである。

平均二乗誤差平方根の観点からの評価

表 5.9 より、条件によっては、2PLM を使用した場合の推定誤差の大きさが2値型の BTM を使用した場合の値よりも小さくなっていたが、全ての条件において、分析モデルが 2PLM である場合と2値型の BTM である場合との間で、推定量の推定誤差の大きさにほとんど差は見られなかった。また、図 5.14 からは、全ての項目群の局所依存度が揃っていない場合に、局所依存関係にある項目の数が増加すると $\Delta\overline{RMSE}$ の値が減少することが確認できる。これは、局所依存関係にある項目数の増加に伴う \overline{RMSE} の増加が、2値型の BTM よりも局所独立性を仮定する 2PLM において、より強く見られたためだと考えられる。さらに、図 5.17 からは、各項目群内で局所依存関係にある項目が5項目である場合に、4つの項目群の局所依存度が等しい時に比べて、4つの項目群の局所依存度が異なる場合には $\Delta\overline{RMSE}$ の値が小さくなっていることがうかがえる。これは、局所依存関係にある項目が5項目である場合には、表 5.9 にもあるように、4つの項目群の局所依存度が異なる時の 2PLM の推定値と真値との相関係数が2値型の BTM に比べて他の条件におけるものよりも小さくなるため、2PLM を使用した場合の推定誤差が BTM を使用した場合に比べてより大きくなったためであると考えられる。

真値と推定値との相関係数の観点からの評価

表 5.9 より、条件によっては、2PLM を使用した場合の値が 2 値型の BTM を使用した場合の値よりも大きくなることがあったものの、全ての条件において、分析モデルが 2PLM である場合と 2 値型の BTM である場合との間で、推定値と真値との相関係数にほとんど差は見られなかった。また、図 5.18 からは、局所依存関係にある項目の数が 5 項目である場合に、4 つの項目群の局所依存度が等しい時に比べて 4 つの項目群の局所依存度が異なる時には、 $\Delta cor(\theta, \hat{\theta})$ の値が大きくなっていることがうかがえる。これは、4 つの項目群の局所依存度が全て等しい場合には、全ての項目特性の推定値に同程度のバイアスが生じるため、2PLM を使用した場合にも、受験者特性推定値の線形性は BTM と同程度に保たれるものの、局所依存度が 4 つの項目群の間で異なる場合には、項目群毎に項目特性にかかるバイアスの程度が異なってくるため、項目母数と共に推定される受験者特性の線形性が保たれなくなり、局所依存関係にある項目が 5 項目である場合には、その程度が他の条件に比べて高くなるためだと考えられる。

まとめ

特定のテストの成績のみを用いて個人差の評価を行うという文脈から上記の結果を見直してみると、真値と推定値との相関係数に関する分析モデル間での差はほとんど見られなかった。そのため、本研究においては、局所独立性を仮定したモデルに代えてタイプ b のモデルを利用することを積極的に推奨するような結果は得られなかったといえる。

また、特定のテストを超えて、各受験者の受験者特性値を共通の尺度上で比較するという文脈から上記の結果を見直してみると、バイアス、平均二乗誤差平方根ともに、分析モデル間でその値にほとんど差は見られず、このような文脈においても、タイプ b のモデルの利用を積極的に推奨するような結果は、本研究では得られなかった。

5.3.4 分析モデルが CCM の場合

最後に、分析モデルが CCM である場合に $\overline{Bias}, \overline{RMSE}, cor(\theta, \hat{\theta})$ の値がどのようになるのか、条件毎に算出したところ、表 5.6 が得られた。また、分析モデルが CCM である場合に $\Delta \overline{Bias}, \Delta \overline{RMSE}, \Delta cor(\theta, \hat{\theta})$ の値がどのようになるのか、条件毎に算出したところ、表 5.7 が得られた。さらに、分析モデルが CCM である場合に $\Delta \overline{Bias}, \Delta \overline{RMSE}, \Delta cor(\theta, \hat{\theta})$ の値が受験者数によりどのように変化するか確認するためグラフを作成したところ、図 5.19, 図 5.20, 図 5.21 が得られた。最後に、分析モデルが CCM である場合に $\Delta \overline{Bias}, \Delta \overline{RMSE}, \Delta cor(\theta, \hat{\theta})$ の値が 4 つの項目群の局所依存度によりどのように変化するか確認するためグラフを作成したところ、図 5.22, 図 5.23, 図

5.24 が得られた。

これらの結果から、分析モデルが CCM である場合のバイアス、平均二乗誤差平方根、真値と推定値との相関係数について、以下のようなことが明らかとなった。

バイアスの観点からの評価

表 5.11 より、条件によっては、2PLM を使用した場合の方が CCM を使用した時よりもバイアスの値が小さくなっていたものの、全ての条件において、分析モデルが 2PLM である場合と CCM である場合との間で、推定量にかかるバイアスにほとんど差は見られなかった。また、図 5.19 より、 $\Delta \overline{Bias}$ の値に関しては、受験者数の増加に伴って、その値が増加する傾向が見られた。これは、受験者数の増加に伴って、2PLM においては \overline{Bias} の値が減少する一方、CCM においてはそのような傾向が見られなかったことに起因する。さらに、図 5.22 からは、局所依存度の強い項目群が増加するにつれて、 $\Delta \overline{Bias}$ の値が減少していく様子が見える。これは、局所依存度の強い項目群が増えるにつれて、CCM においては \overline{Bias} の値が減少する一方、2PLM においては特に増加、減少等の傾向が確認されなかったためである。

平均二乗誤差平方根の観点からの評価

表 5.11 より、条件によっては、2PLM を使用した場合の推定誤差の大きさが CCM を使用した時のものよりも小さくなることがあったが、全ての条件において、分析モデルが 2PLM である場合と CCM である場合との間で、推定量の推定誤差の大きさにはほとんど差は見られなかった。また、図 5.20 からは、受験者数の増加に伴い、 $\Delta \overline{RMSE}$ の値が減少している様子が見える。これは、局所依存性を考慮する CCM に比べて局所独立性を仮定する 2PLM において、受験者数の増加に伴う \overline{RMSE} の値の減少がより大きかったためだと考えられる。

真値と推定値との相関係数の観点からの評価

表 5.11 より、条件によっては、2PLM を使用した場合の方が CCM を使用した場合よりも推定値と真値の相関係数が大きくなることもあったが、全ての条件において、分析モデルが 2PLM である場合と CCM である場合との間で、推定値と真値との相関係数にほとんど差は見られなかった。

まとめ

特定のテストの成績のみを用いて個人差の評価を行うという観点から上記の結果を見直してみると、真値と推定値との相関係数に関しては、分析モデル間でほとんど差は見られず、本研究では、局所独立性を仮定したモデルに代えてより複雑なタイプ c のモデルを利

用することを強く推奨するような結果は、得られなかった。

また、特定のテストを超えて、各受験者の受験者特性値を共通の尺度上で比較するという観点から上記の結果を見直してみると、バイアス、平均二乗誤差平方根の値に関しては、分析モデル間でほとんど差は見られず、タイプcのモデルの利用を積極的に肯定するような結果は、本研究では得られなかった。

さらに、一般的には、局所依存度の強い項目群の数が増えることによりバイアスの値は増加すると考えられているが、CCMに関しては、それとは逆の関係が確認された。この点に関しては、5.3.5節において、改めて考察を加える。

5.3.5 CCMの \overline{Bias} と局所依存度の強い項目群の数との関係

本研究では、局所依存度の強い項目群に対しCCMをあてはめ b_{jk} の値を推定したところ、その推定値の平均が-1.714となり、局所依存度の弱い項目群に対する b_{jk} の推定値の平均は-0.612となった。そこで、2値型のBTMとCCMをそれぞれ発生モデルとし、局所依存度の強い2つの項目 j, k に対して100人の受験者がどのように反応するのかシミュレーションを行って調べてみたところ、 $\sigma_{\gamma_{a(j)}}^2 = 1.4075$ とした2値型のBTMを発生モデルとした場合には j, k に対し(1,1), (1,0), (0,1), (0,0)と反応する受験者の割合(100回分の平均)が表5.8のようになり、 $b_{jk} = -1.714$ としたCCMを発生モデルとした場合には表5.9のようになった。また、2値型のBTMとCCMをそれぞれ発生モデルとし、局所依存度の弱い2つの項目 j, k に対し各受験者($N = 100$)がどのように反応するのかシミュレーションを行って調べてみたところ、 $\sigma_{\gamma_{a(j)}}^2 = 0.155$ とした2値型のBTMを発生モデルとした場合には表5.10のようになり、 $b_{jk} = -0.612$ としたCCMを発生モデルとした場合には表5.11のようになった。表5.8, および、表5.9, 表5.10, 表5.11からは、

- 2値型のBTMを発生モデルとした場合、局所依存度の強弱に関わらず、項目 j, k に同一の反応をする受験者が異なる反応をする受験者に比べて多くなり、項目 j, k の両方に正答する受験者が誤答する受験者に比べて多くなっている
- CCMを発生モデルとした場合にも、局所依存度の強弱に関わらず、項目 j, k に同一の反応をする受験者が異なる反応をする受験者に比べて多くなっているが、項目反応間の局所依存度が弱い場合には、項目 j, k の両方に正答する受験者が誤答する受験者に比べて少なくなっている

ことが読み取れ、このように、項目反応間の局所依存度が弱い項目群においては、2値型のBTMが表現している局所依存関係とCCMが表現している局所依存関係とが異なるものとなっていたため、本研究では、局所依存度の強い項目群の数が増加するにつれて、CCMの \overline{Bias} が減少したのではないかと考えられる。

5.3.6 先行研究との関連

5.1 節でも述べたように、本研究では、局所依存性を考慮することにより受験者特性の推定精度がどの程度改善されるのか、局所依存性を考慮する3タイプのモデル間での比較を行うため、分析モデルとして GRM, BTM, CCM, 2PLM を用い、シミュレーションを実施した。この点に関しては、分析モデルとして BTM と 2PLM を使用し、受験者特性の推定精度に関する比較を行った Bradlow et al. (1999) の研究との間で相違が見られるのだが、本研究において分析モデルが BTM である場合の結果に着目してみると、以下の点に関しては、類似した結果が得られていると言える。

- 分析モデルが BTM である場合の方が、2PLM である場合よりも、M95%PIW が実際に真値を含む確率が大きくなる
- 分析モデルが BTM である場合と 2PLM である場合との間で、受験者特性の真値と推定値との順位相関係数がほとんど変わらない

また、DeMars (2006) の研究においては、タイプ a, タイプ b, タイプ d のモデル間での比較が行われており、この点に関しては、本研究との間で相違が見られるが、以下の点に関しては、本研究との間で類似した結果が得られている。

- 分析モデル間で、推定量のバイアス、平均二乗誤差平方根の値に関してほとんど差が見られない

さらに、Wainer et al. (2007, pp. 137-140) の研究においても、タイプ b のモデルとタイプ d のモデルとの間で比較が行われており、この点に関しては、本研究との間で相違が見られるのだが、分析モデルが BTM である場合の結果に着目すると、以下の点に関しては、本研究との間で類似した結果が得られている。

- 推定値と真値との相関係数に関して、分析モデル間でほとんど差は見られない

5.3.7 局所依存性を考慮するモデルのタイプと推定精度の改善度合いとの関係

本研究では、局所独立性を仮定したモデルとの間で受験者特性の推定精度を比較した場合に、改善度合いの評価に関して、局所依存性を考慮する3タイプのモデル間に違いは見られなかった。このことから、多くの分析モデルに共通する一般的な評価として、局所依存性を考慮することによる推定精度の改善に関し、以下のようなことが言えると考えら

れる。

- 局所依存性を有するデータに対し、局所依存性を考慮するモデルをあてはめて受験者特性の推定を行った場合、推定量のバイアス、平均二乗誤差平方根、推定値と真値との相関係数に関しては、局所独立性を仮定するモデルを使用したときのものとほとんど差が見られない

一方、推定精度の改善度合いと受験者数、局所依存関係にある項目の数、4つの項目群の局所依存度との関係に関しては、局所依存性を考慮するモデルの間で違いが見られたのだが、本研究で得られた知見を応用するという観点からは、多くの分析モデルに共通する一般的な関係性を抽出する必要があると考えられる。そのような観点から本研究で得られた結果をまとめなおすと、以下のようなようになる。

- 受験者数の増加に伴って、局所独立性を仮定したモデルの頑健性が上昇し、推定量にかかるバイアスの値が減少するため、局所依存性を考慮したモデルを使用した場合と比較した際に推定精度の差異が減少する

表 5.1 分析モデルが 2PLM である時の \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.023	0.530	0.822
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.023	0.501	0.848
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.021	0.459	0.880
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.020	0.429	0.899
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.021	0.385	0.923
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.021	0.423	0.902
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.021	0.409	0.910
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.021	0.393	0.920
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.021	0.381	0.925
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.021	0.366	0.932
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.021	0.398	0.916
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.022	0.388	0.922
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.022	0.377	0.928
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.021	0.371	0.930
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.021	0.361	0.934
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.016	0.546	0.804
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.016	0.517	0.831
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.015	0.471	0.867
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.015	0.435	0.891
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.015	0.385	0.919
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.015	0.424	0.897
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.016	0.409	0.907
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.015	0.391	0.916
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.015	0.377	0.923
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.015	0.363	0.929
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.016	0.397	0.913
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.015	0.386	0.919
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.015	0.374	0.925
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.015	0.367	0.928
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.015	0.358	0.932

表 5.2 分析モデルが GRM である時の \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.021	0.533	0.817
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.020	0.499	0.858
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.020	0.454	0.888
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.020	0.426	0.904
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.020	0.397	0.918
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.021	0.422	0.906
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.021	0.409	0.913
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.021	0.396	0.920
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.021	0.386	0.924
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.022	0.375	0.929
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.022	0.399	0.918
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.022	0.389	0.923
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.021	0.378	0.928
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.022	0.373	0.930
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.022	0.365	0.933
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.019	0.541	0.800
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.019	0.506	0.848
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.018	0.457	0.881
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.018	0.426	0.899
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.008	0.397	0.914
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.018	0.416	0.902
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.019	0.403	0.910
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.018	0.387	0.917
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.018	0.374	0.922
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.019	0.363	0.926
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.018	0.396	0.914
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.019	0.386	0.920
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.018	0.375	0.925
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.019	0.368	0.927
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.019	0.361	0.930

表 5.3 分析モデルが GRM である時の $\overline{\Delta Bias}$, $\overline{\Delta RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	$\overline{\Delta Bias}$	$\overline{\Delta RMSE}$	$\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.002	0.003	-0.005
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.003	-0.002	0.010
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.001	-0.005	0.008
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.000	-0.003	0.005
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.001	0.012	-0.005
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.000	-0.001	0.004
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.000	0.000	0.003
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.000	0.003	0.000
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.000	0.005	-0.001
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.001	0.009	-0.003
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.001	0.001	0.002
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.000	0.001	0.001
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.001	0.001	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.001	0.002	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.001	0.004	-0.001
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.003	-0.005	-0.004
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.003	-0.011	0.017
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.003	-0.014	0.014
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.003	-0.009	0.008
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.003	0.012	-0.005
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.003	-0.004	0.005
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.003	-0.003	0.003
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.003	0.001	0.001
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.003	0.003	-0.001
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.004	0.007	-0.003
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.002	-0.001	0.001
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.004	0.000	0.001
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.003	0.001	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.004	0.001	-0.001
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.004	0.003	-0.002

表 5.4 分析モデルが BTM である時の \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.018	0.526	0.820
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.017	0.486	0.863
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.016	0.442	0.893
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.016	0.415	0.908
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.017	0.386	0.923
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.018	0.418	0.907
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.018	0.405	0.915
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.017	0.390	0.922
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.017	0.379	0.927
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.017	0.366	0.932
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.018	0.397	0.918
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.018	0.388	0.923
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.019	0.377	0.928
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.017	0.371	0.930
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.017	0.362	0.934
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.019	0.535	0.803
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.018	0.494	0.852
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.018	0.446	0.886
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.018	0.416	0.903
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.018	0.385	0.919
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.018	0.416	0.903
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.018	0.403	0.912
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.018	0.387	0.919
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.019	0.374	0.925
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.018	0.363	0.930
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.018	0.395	0.915
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.018	0.385	0.920
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.018	0.373	0.925
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.018	0.367	0.928
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.018	0.358	0.932

表 5.5 分析モデルが BTM である時の $\overline{\Delta Bias}$, $\overline{\Delta RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	$\overline{\Delta Bias}$	$\overline{\Delta RMSE}$	$\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 4$	-0.005	-0.004	-0.002
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 3$	-0.006	-0.015	0.015
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 2$	-0.005	-0.017	0.013
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 1$	-0.004	-0.014	0.009
$N = 300, J_{d(j)} = 5, \text{強 } d(j) \times 0$	-0.004	0.001	0.000
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 4$	-0.003	-0.005	0.005
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 3$	-0.003	-0.004	0.005
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 2$	-0.004	-0.003	0.002
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 1$	-0.004	-0.002	0.002
$N = 300, J_{d(j)} = 3, \text{強 } d(j) \times 0$	-0.004	0.000	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.003	-0.001	0.002
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.004	0.000	0.001
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.003	0.000	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.004	0.000	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.004	0.001	0.000
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 4$	0.003	-0.011	-0.001
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 3$	0.002	-0.023	0.021
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 2$	0.003	-0.025	0.019
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 1$	0.003	-0.019	0.012
$N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0$	0.003	0.000	0.000
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 4$	0.003	-0.008	0.006
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 3$	0.002	-0.006	0.005
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 2$	0.003	-0.004	0.003
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 1$	0.004	-0.003	0.002
$N = 1000, J_{d(j)} = 3, \text{強 } d(j) \times 0$	0.003	0.000	0.001
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.002	-0.002	0.002
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.003	-0.001	0.001
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.003	-0.001	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.003	0.000	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.003	0.000	0.000

表 5.6 分析モデルが CCM である時の \overline{Bias} , \overline{RMSE} , $cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	\overline{Bias}	\overline{RMSE}	$cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.001	0.400	0.917
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.005	0.389	0.923
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.008	0.378	0.927
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.012	0.372	0.930
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.015	0.363	0.934
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	0.010	0.395	0.914
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	0.011	0.384	0.920
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	0.014	0.373	0.925
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.015	0.366	0.928
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.017	0.358	0.932

表 5.7 分析モデルが CCM である時の $\Delta\overline{Bias}$, $\Delta\overline{RMSE}$, $\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ の値

条件	$\Delta\overline{Bias}$	$\Delta\overline{RMSE}$	$\Delta cor(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.020	0.002	0.001
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.017	0.001	0.001
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.014	0.001	-0.001
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 1$	-0.009	0.001	0.000
$N = 300, J_{d(j)} = 2, \text{強 } d(j) \times 0$	-0.006	0.002	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 4$	-0.006	-0.002	0.001
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 3$	-0.004	-0.002	0.001
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 2$	-0.001	-0.001	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 1$	0.000	-0.001	0.000
$N = 1000, J_{d(j)} = 2, \text{強 } d(j) \times 0$	0.002	0.000	0.000

表 5.8 $\sigma_{\gamma_{d(j)}}^2 = 1.4075$ とした BTM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 回分の平均)

	$U_k = 1$	$U_k = 0$
$U_j = 1$	38%	20%
$U_j = 0$	13%	29%

表 5.9 $b_{jk} = -1.714$ とした CCM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 回分の平均)

	$k = 1$	$k = 0$
$j = 1$	46%	5%
$j = 0$	6%	43%

表 5.10 $\sigma_{\gamma_{d(j)}}^2 = 0.155$ とした BTM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 回分の平均)

	$U_k = 1$	$U_k = 0$
$U_j = 1$	38%	23%
$U_j = 0$	15%	24%

表 5.11 $b_{jk} = -0.612$ とした CCM を発生モデルとした場合に各項目に対して (1, 1), (1, 0), (0, 1), (0, 0) と反応する受験者の割合 (100 回分の平均)

	$k = 1$	$k = 0$
$j = 1$	32%	8%
$j = 0$	8%	52%

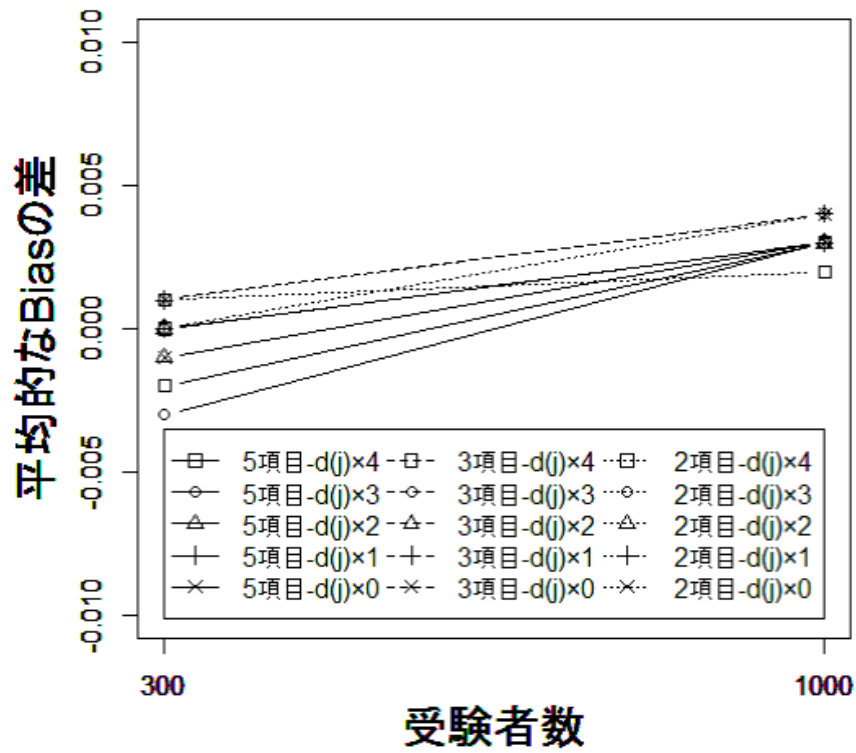


図 5.1 分析モデルが GRM である場合の $\Delta \overline{Bias}$ の受験者数による変化

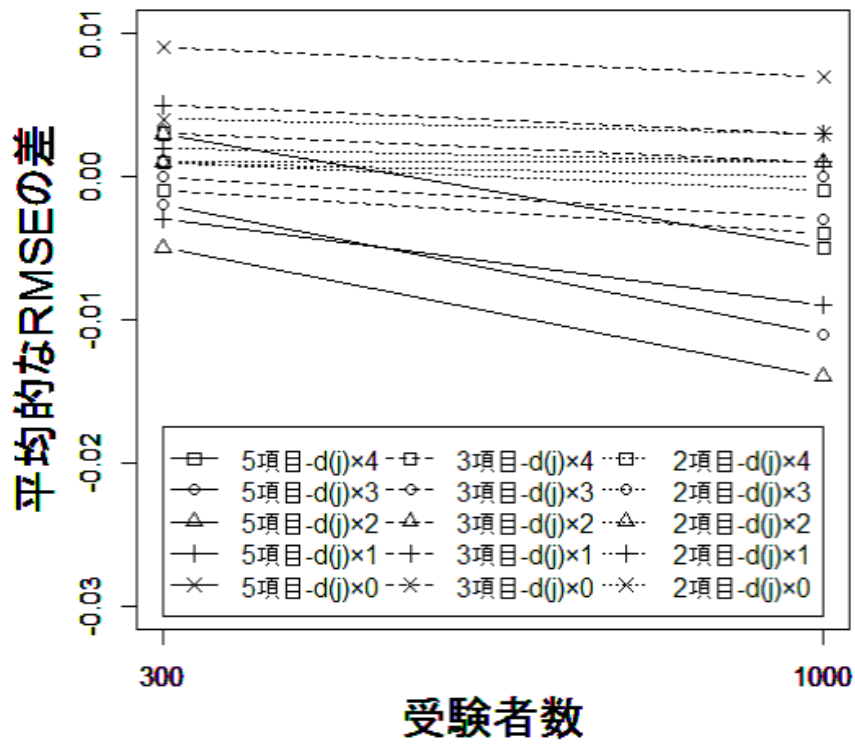


図 5.2 分析モデルが GRM である場合の $\overline{\Delta RMSE}$ の受験者数による変化

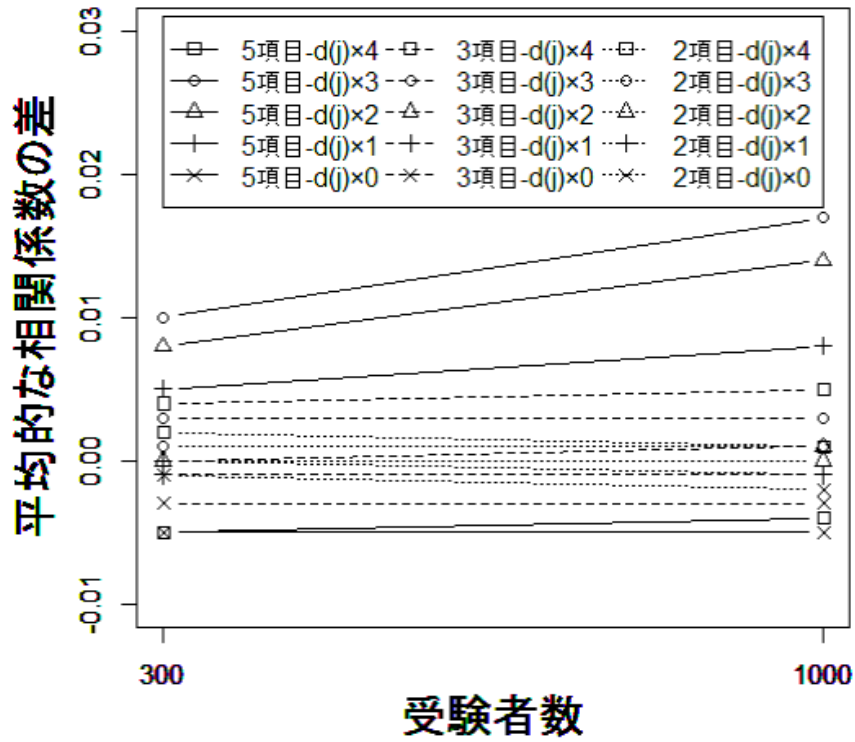


図 5.3 分析モデルが GRM である場合の $\Delta cor(\theta, \hat{\theta})$ の受験者数による変化

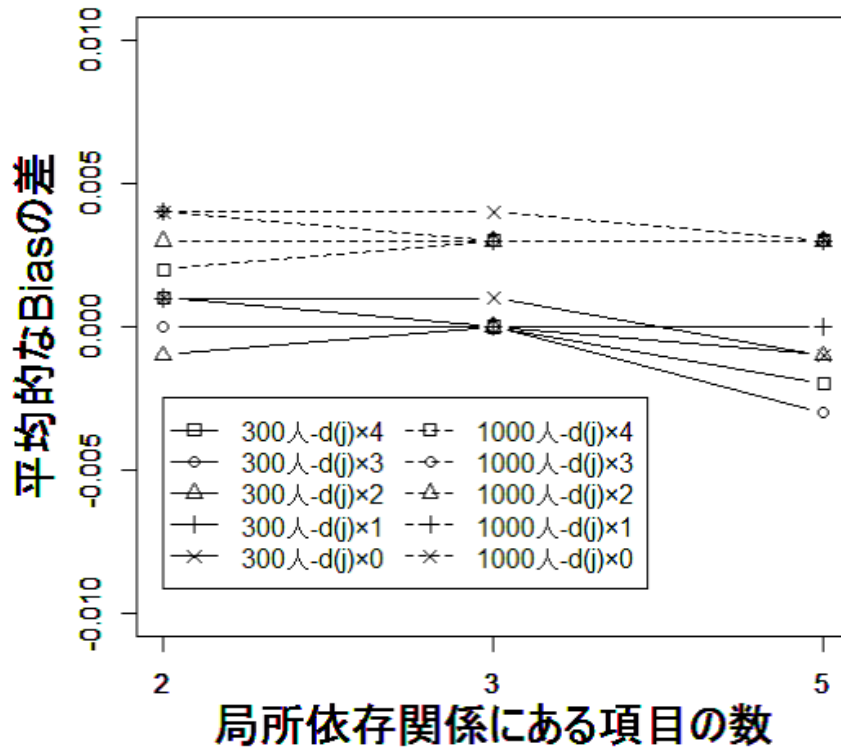


図 5.4 分析モデルが GRM である場合の $\overline{\Delta Bias}$ の局所依存関係にある項目の数による変化

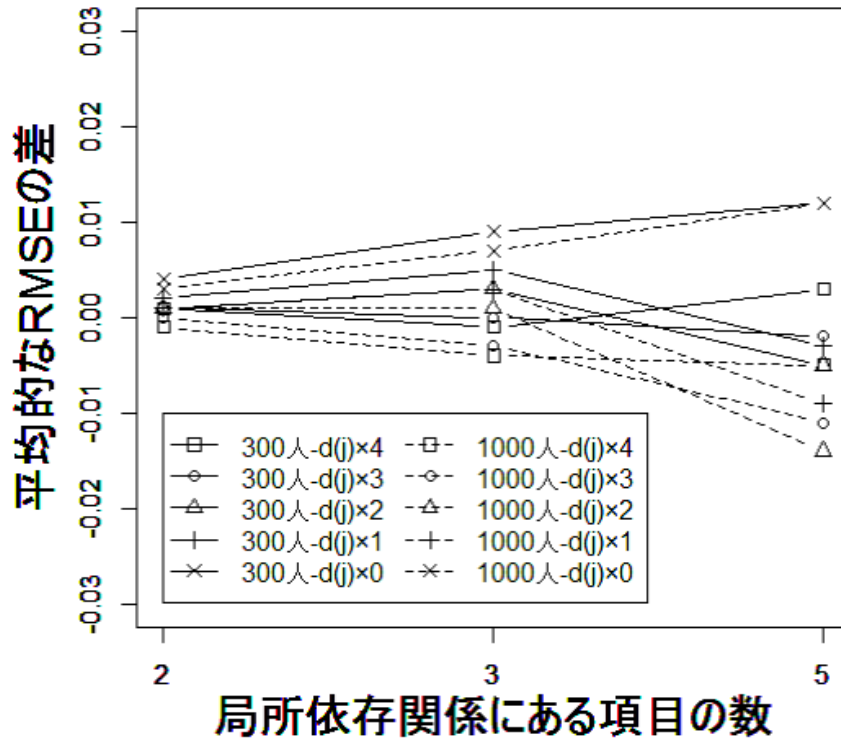


図 5.5 分析モデルが GRM である場合の $\overline{\Delta RMSE}$ の局所依存関係にある項目の数による変化

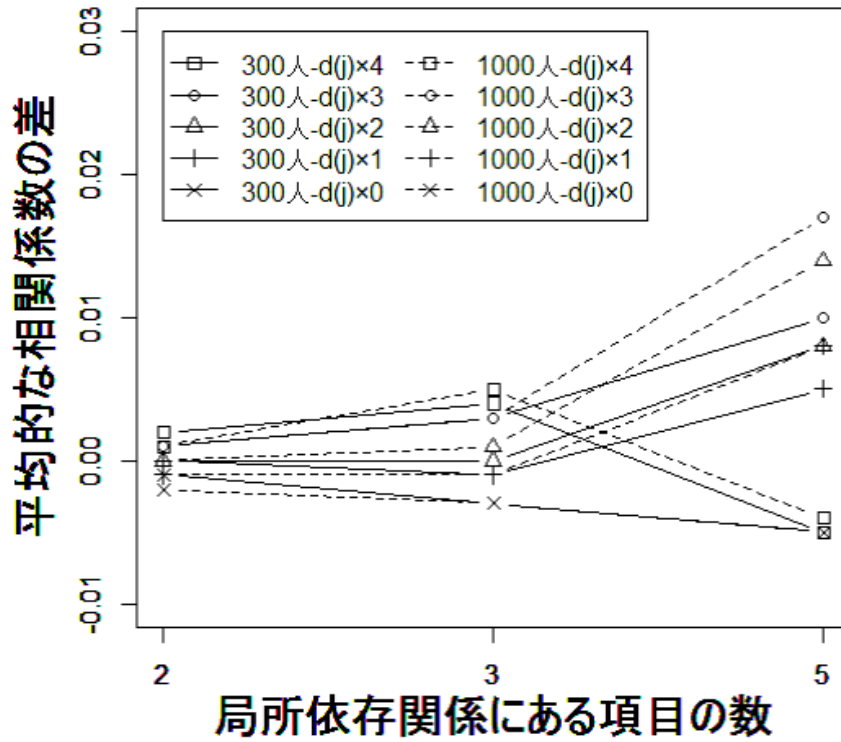


図 5.6 分析モデルが GRM である場合の $\Delta cor(\theta, \hat{\theta})$ の局所依存関係にある項目の数による変化

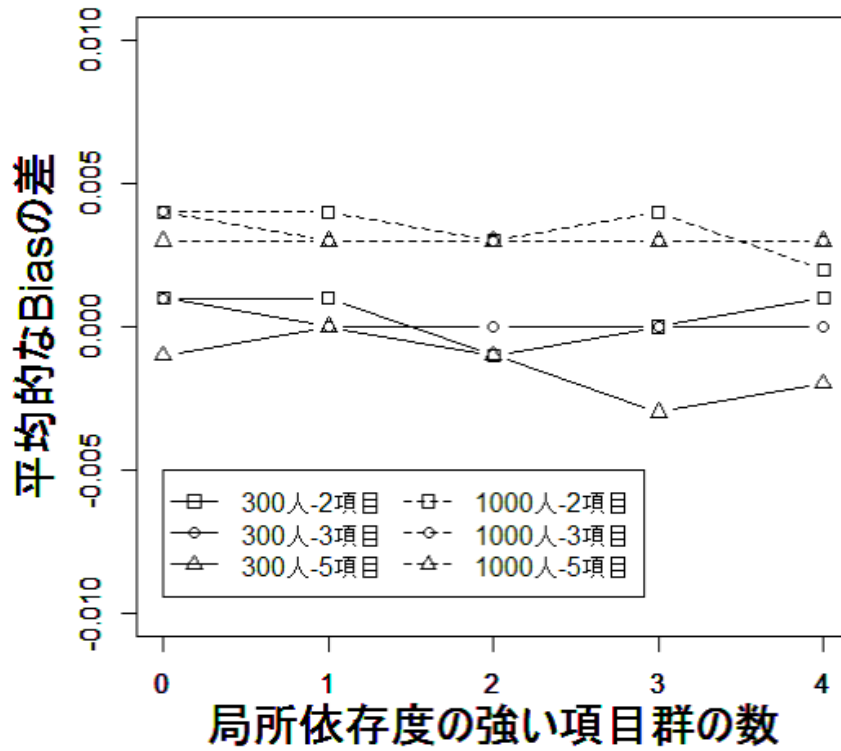


図 5.7 分析モデルが GRM である場合の $\overline{\Delta Bias}$ の 4 つの項目群の局所依存度による変化

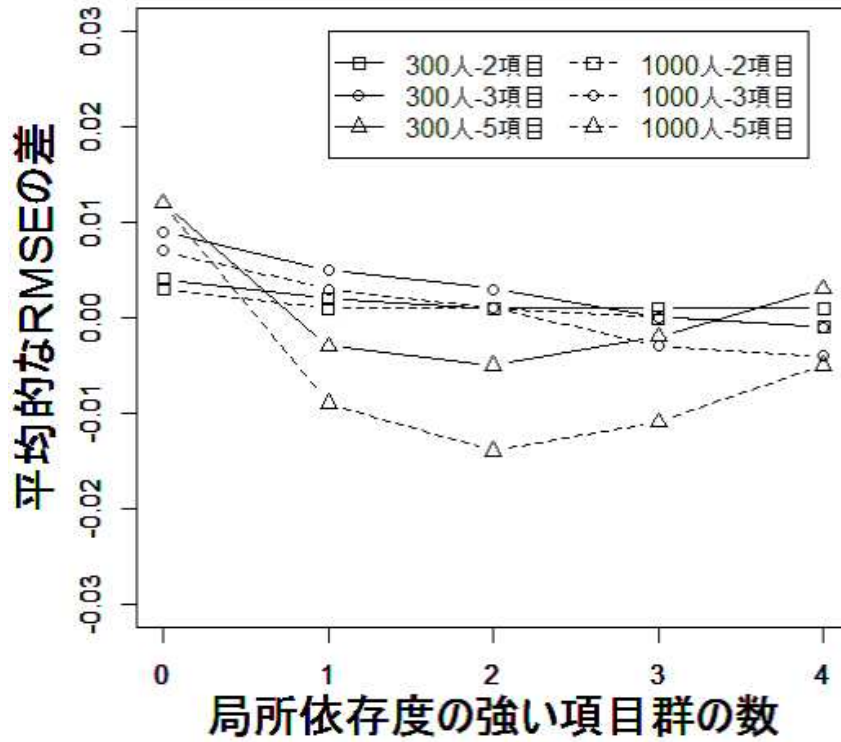


図 5.8 分析モデルが GRM である場合の $\overline{\Delta RMSE}$ の 4 つの項目群の局所依存度による変化

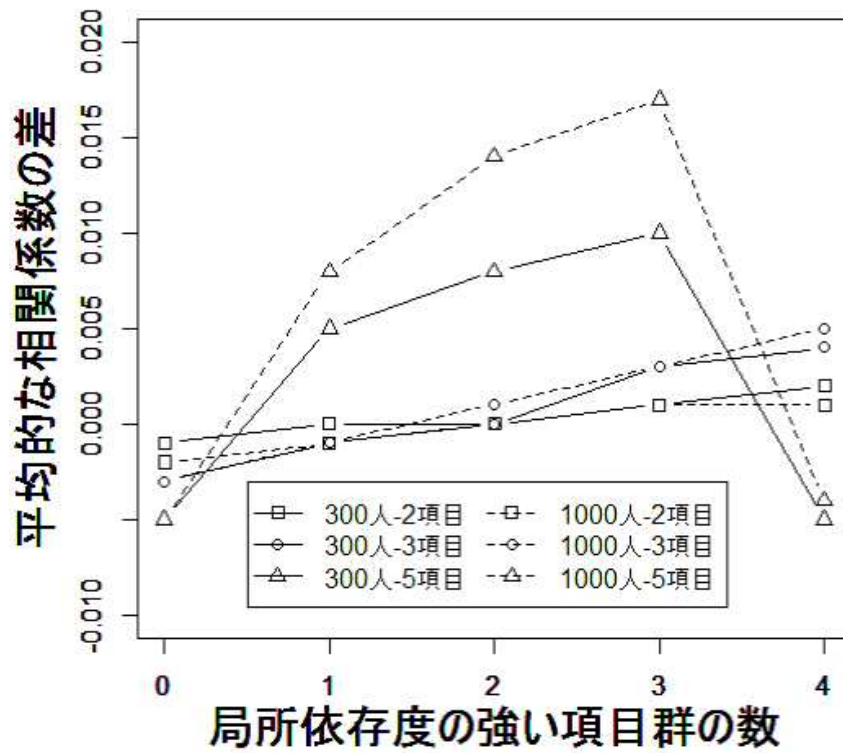


図 5.9 分析モデルが GRM である場合の $\Delta cor(\theta, \hat{\theta})$ の 4 つの項目群の局所依存度による変化

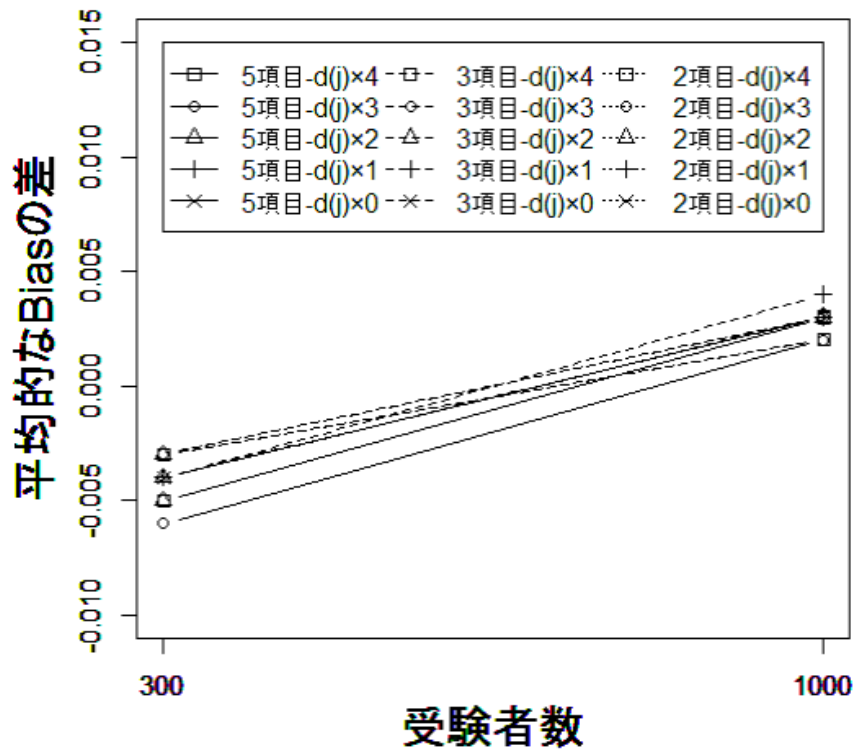


図 5.10 分析モデルが BTM である場合の $\Delta \overline{Bias}$ の受験者数による変化

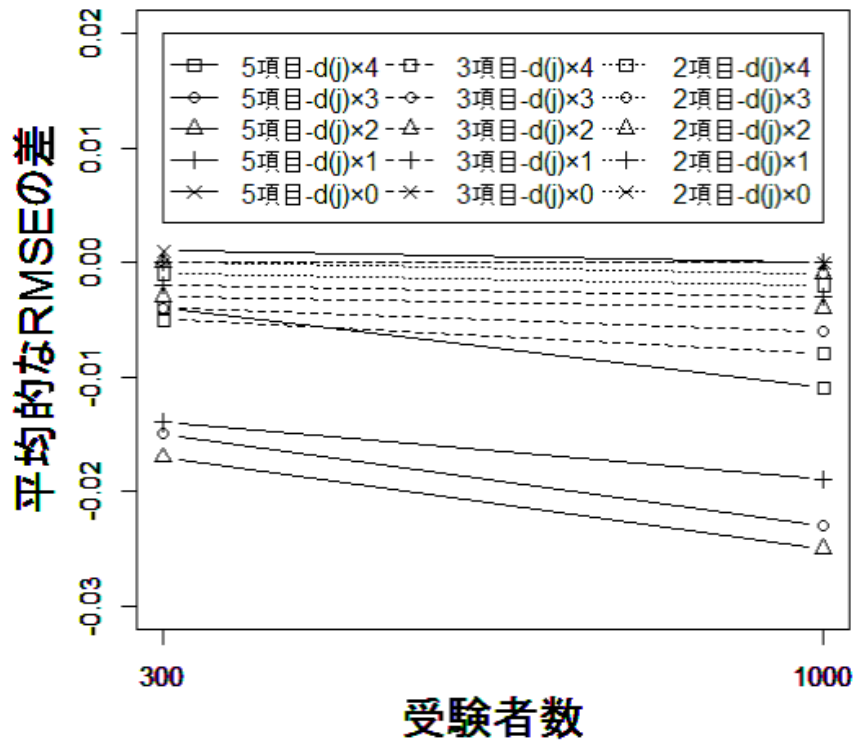


図 5.11 分析モデルが BTM である場合の $\overline{\Delta RMSE}$ の受験者数による変化

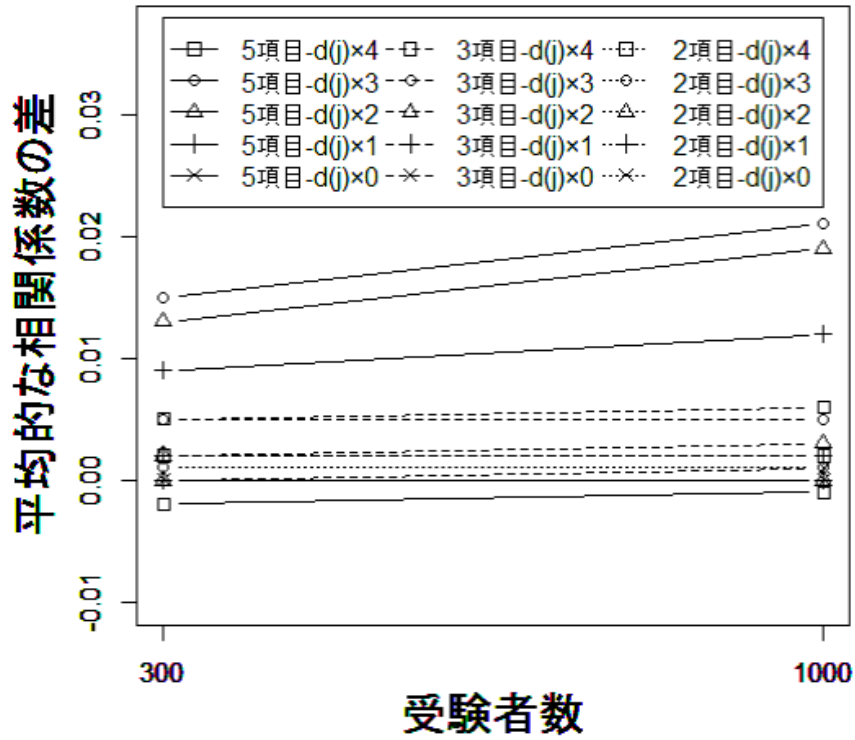


図 5.12 分析モデルが BTM である場合の $\Delta cor(\theta, \hat{\theta})$ の受験者数による変化

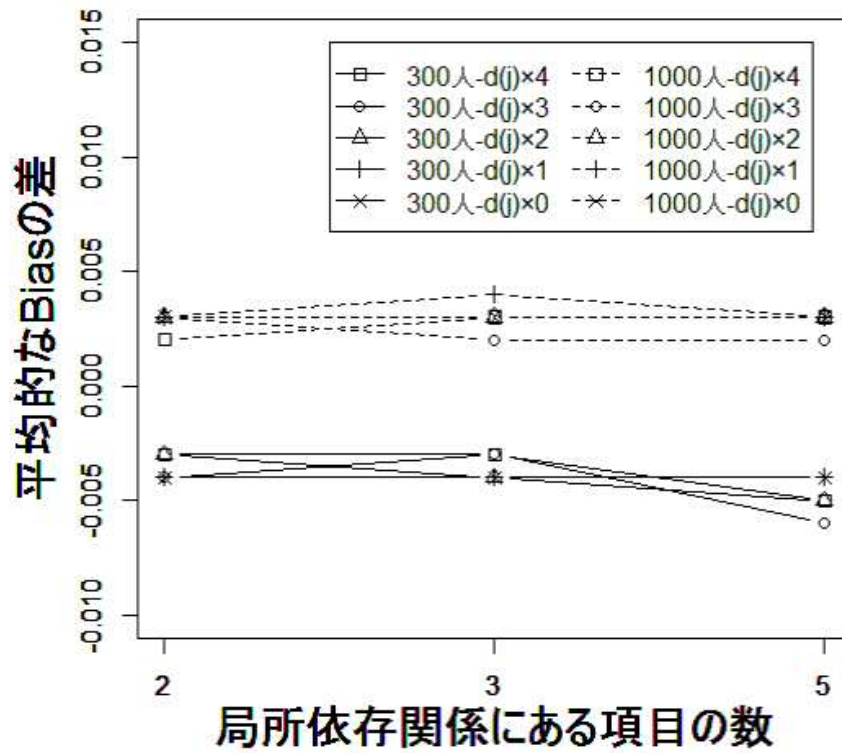


図 5.13 分析モデルが BTM である場合の $\overline{\Delta Bias}$ の局所依存関係にある項目の数による変化

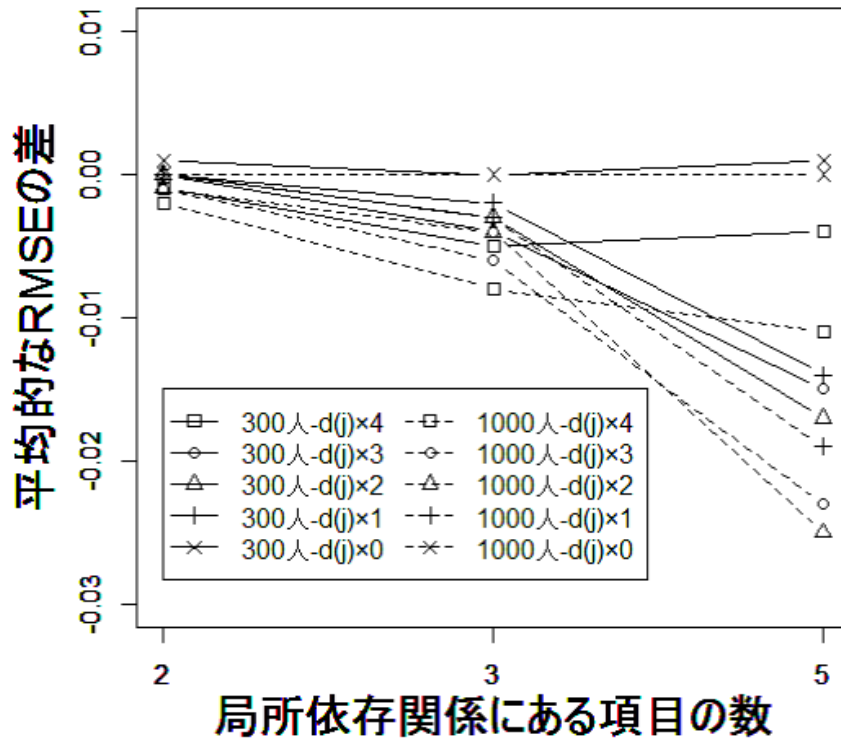


図 5.14 分析モデルが BTM である場合の $\Delta \overline{RMSE}$ の局所依存関係にある項目の数による変化

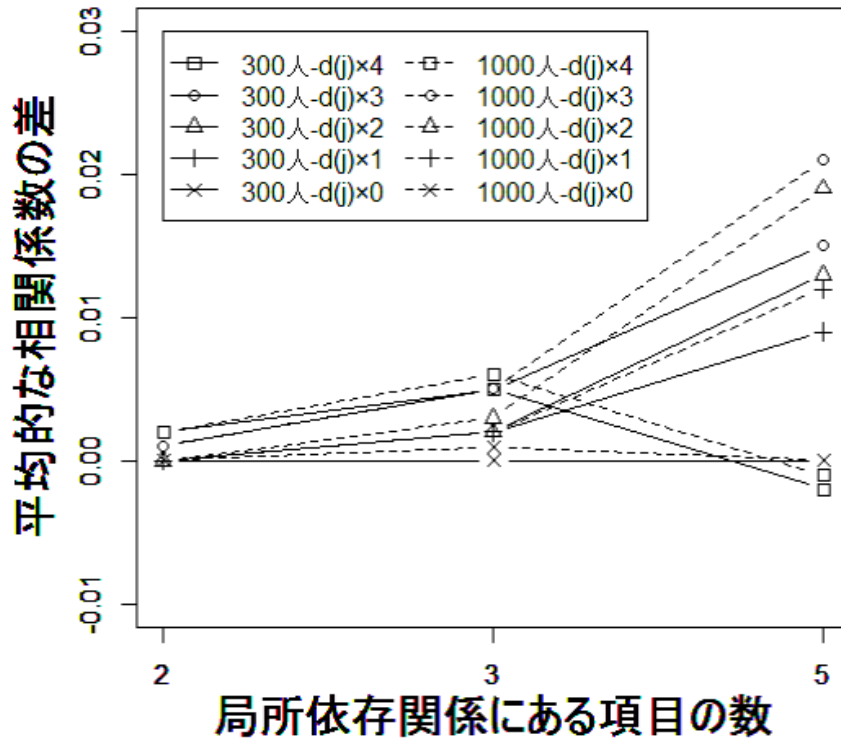


図 5.15 分析モデルが BTM である場合の $\Delta cor(\theta, \hat{\theta})$ の局所依存関係にある項目の数による変化

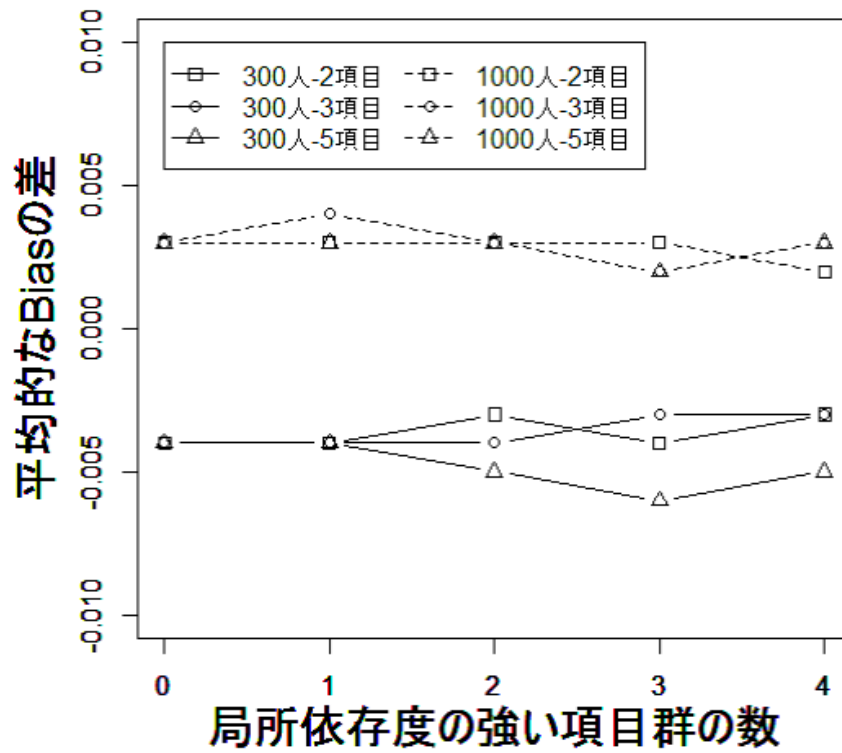


図 5.16 分析モデルが BTM である場合の $\Delta \overline{Bias}$ の 4 つの項目群の局所依存度による変化

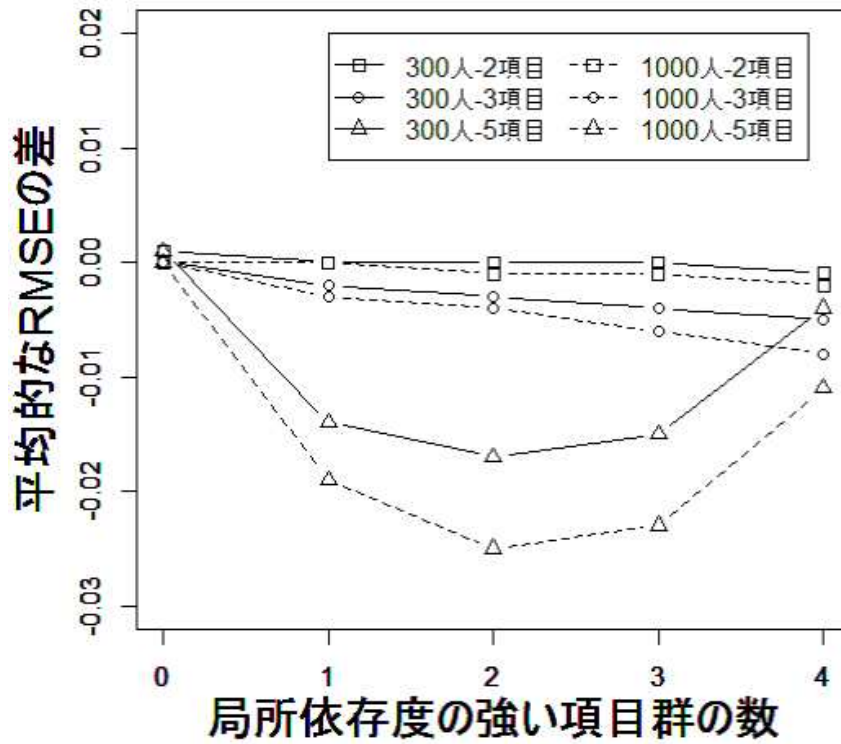


図 5.17 分析モデルが BTM である場合の $\overline{\Delta RMSE}$ の 4 つの項目群の局所依存度による変化

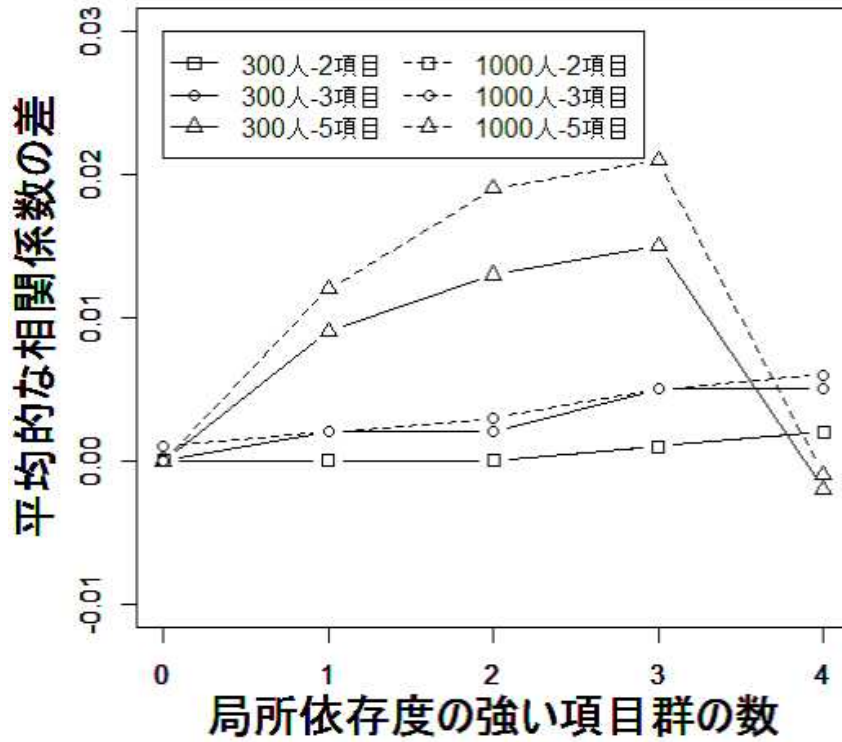


図 5.18 分析モデルが BTM である場合の $\Delta cor(\theta, \hat{\theta})$ の 4 つの項目群の局所依存度による変化

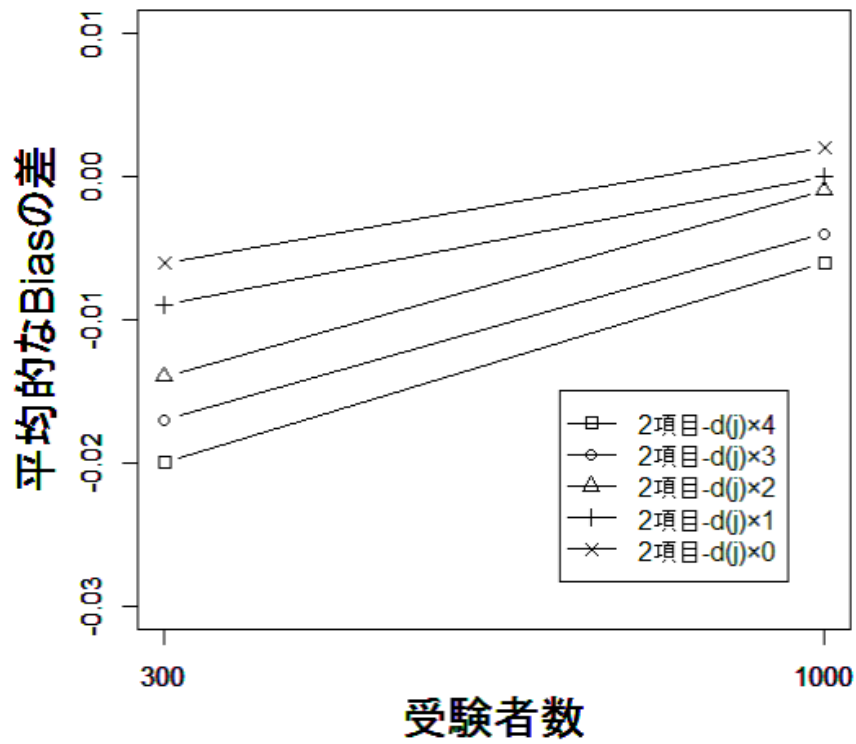


図 5.19 分析モデルが CCM である場合の $\Delta \overline{Bias}$ の受験者数による変化

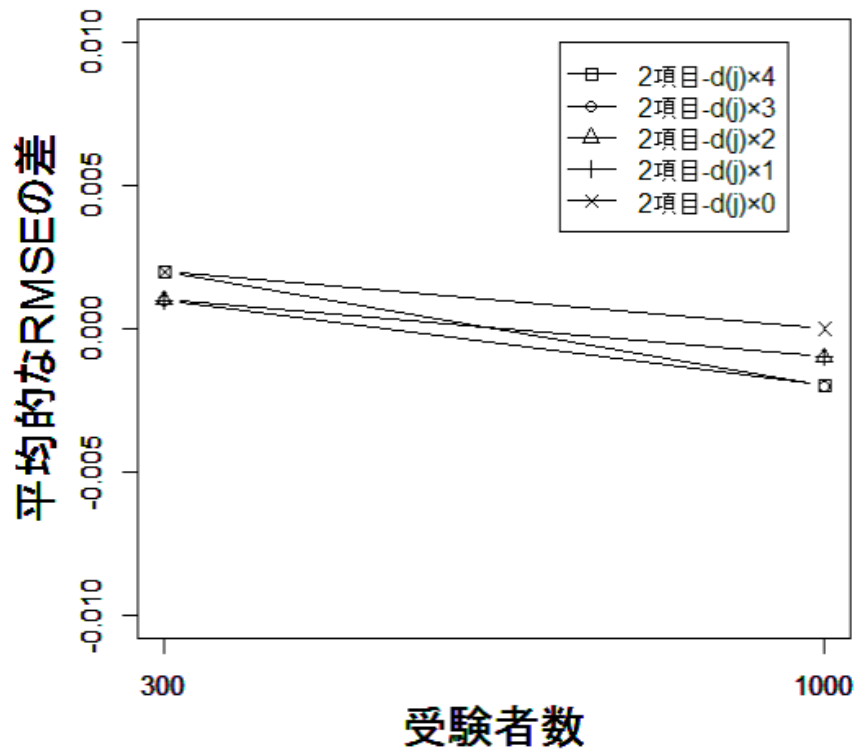


図 5.20 分析モデルが CCM である場合の $\overline{\Delta RMSE}$ の受験者数による変化

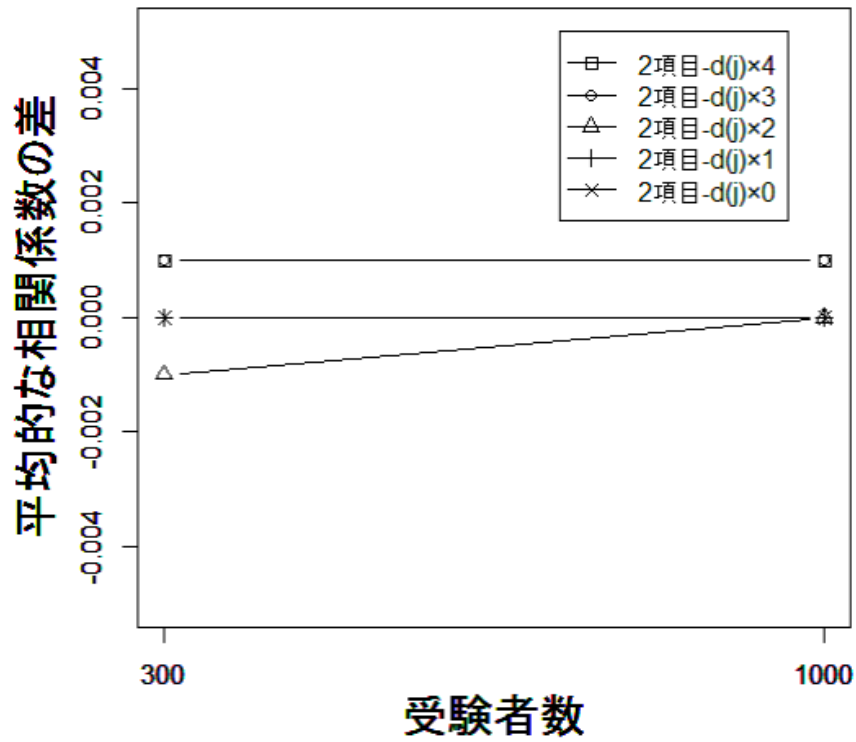


図 5.21 分析モデルが CCM である場合の $\Delta cor(\theta, \hat{\theta})$ の受験者数による変化

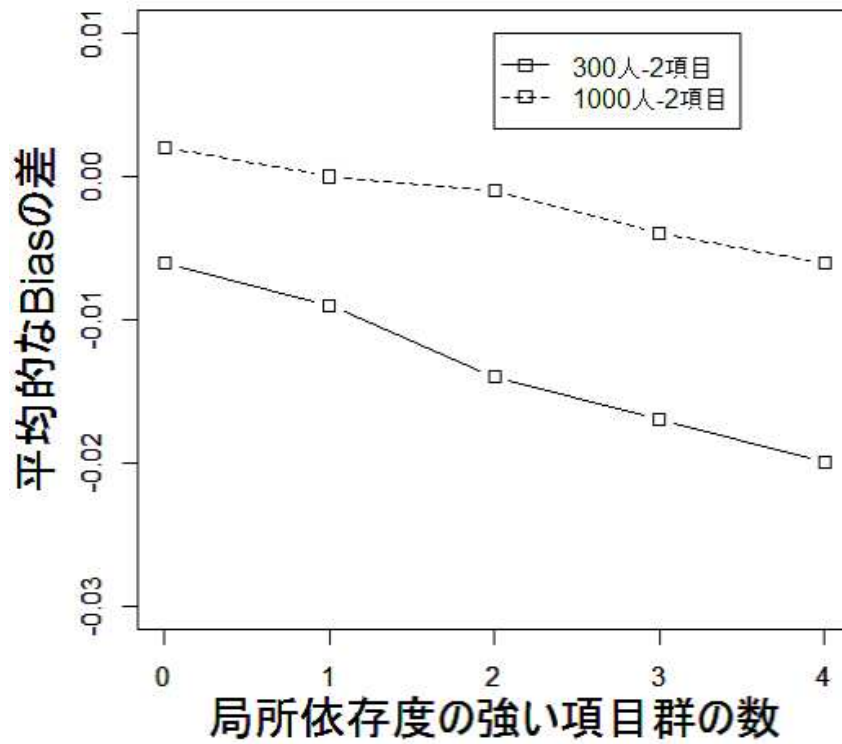


図 5.22 分析モデルが CCM である場合の $\overline{\Delta Bias}$ の4つの項目群の局所依存度による変化

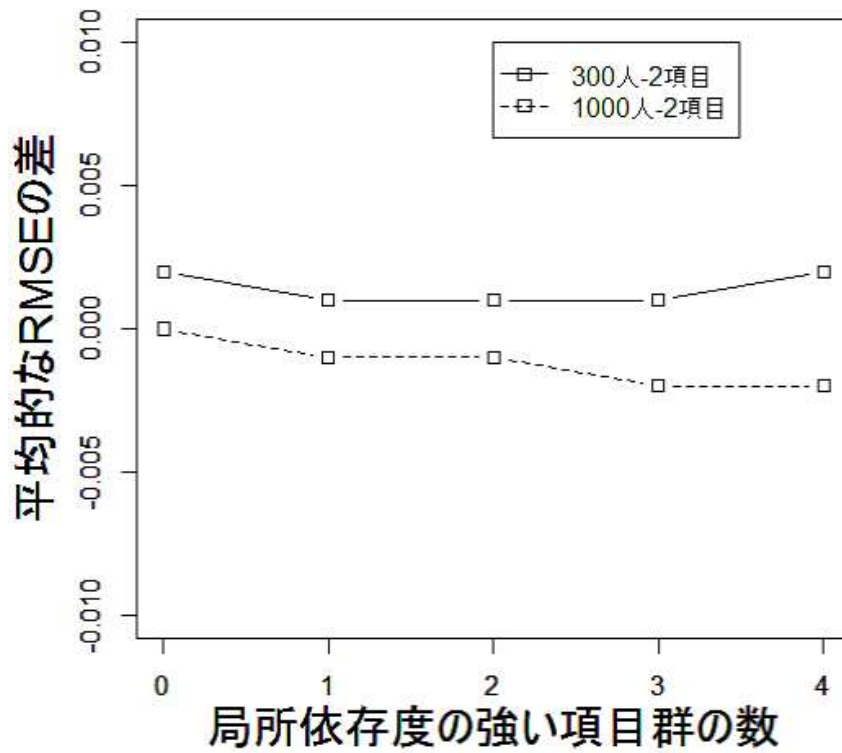


図 5.23 分析モデルが CCM である場合の $\overline{\Delta RMSE}$ の4つの項目群の局所依存度による変化

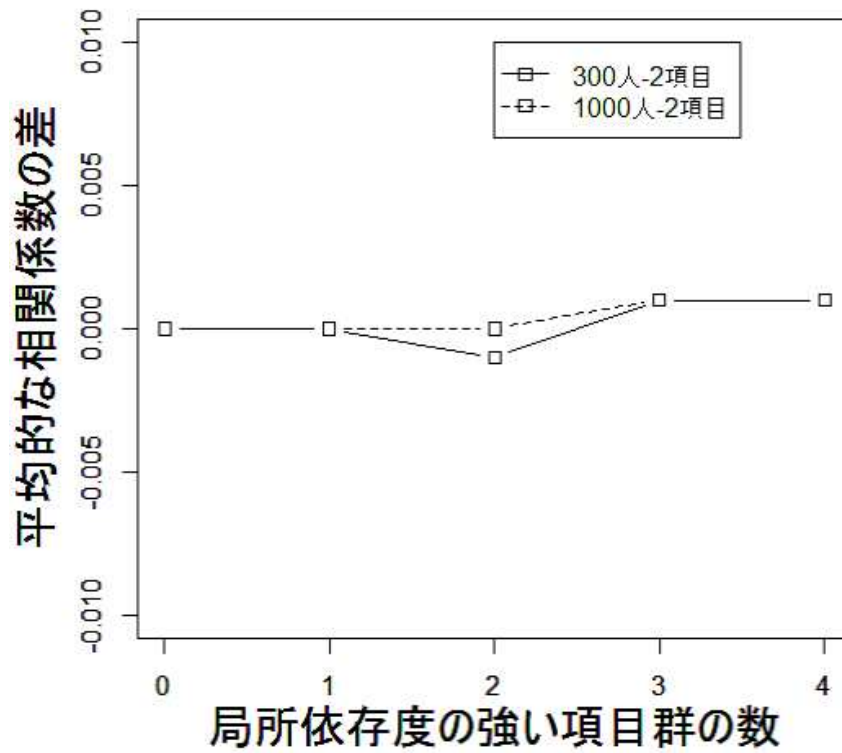


図 5.24 分析モデルが CCM である場合の $\Delta cor(\theta, \hat{\theta})$ の 4 つの項目群の局所依存度による変化

第6章

総合考察

6.1 本論文で明らかになったこと

第1章で局所独立性の概念についての整理と局所依存性を考慮した項目反応モデルのレビューを行った後、まず第2章において、局所依存性が局所独立性を仮定した受験者特性の推定に対してどのような影響を与えるのか、異なる発生メカニズムを表現する複数のモデルからデータを発生させてシミュレーションを実施し、検討を行った。その結果、発生モデル毎に局所依存性に対する頑健性の評価結果が異なった一方、多くの発生メカニズムに共通する一般的な知見として、受験者数が少ない場合に局所依存性の影響が無視できないほど大きくなる、という知見が得られた。特に、特定のテストを超えて各受験者の受験者特性値を共通の尺度上で評価するという文脈からは、受験者数が少ない場合、局所独立性を仮定したモデルは頑健ではないといえる。また、一般的な知見として、推定量のバイアスに対する局所依存性の影響は、項目が局所依存性を有している場合、項目数の増加に伴う改善が見られないことが明らかとなった。

次に第3章において、局所依存性を有するデータに対して局所独立性を仮定するモデルと局所依存性を考慮するモデルをあてはめてシミュレーションを実施し、局所依存性が局所独立性を仮定した項目特性の推定に対してどのような影響を与えるのか、項目母数の比較可能性を確保した上で検討を行った。その結果、特定のテスト内において項目の特性を評価するという文脈においては、局所依存関係にある項目のテスト全体に占める割合が高く、項目間の局所依存度が項目群間で異なる場合にのみ、局所依存性の影響が無視できないほど大きくなっていった。一方、特定のテストを超えて各項目の特性値を共通の尺度上で比較するという文脈においては、局所依存性の影響が無視できないほど大きくなっており、局所独立性を仮定したモデルは頑健であるとはいえないことがわかった。また、受験者数や局所依存関係にある項目数との関係については、対象とする項目特性毎に異なる関係が得られたものの、共通する一般的な知見として、局所依存関係にある項目の割合が増加す

ると、局所独立性を仮定した場合の推定誤差が局所依存性を考慮した場合に比べて大きくなる、という関係が見られた。加えて、局所依存関係にある項目のテスト全体に占める割合が高い場合、項目間の局所依存度が項目群により異なると、局所独立性を仮定した場合の真値と推定値との相関係数が局所依存性を考慮した場合に比べて低下する、という関係が見られた。さらに、項目母数の比較可能性が頑健性の評価にどのような影響を与えるのか検討したところ、比較可能性を確保した場合としない場合とで識別力の推定量のバイアスの符号が逆転するなど、比較可能性を確保するか否かにより頑健性の評価が異なる結果となった。したがって、項目母数の比較可能性を確保していなかった先行研究における結果は、本研究における結果によって見直される必要がある。

さらに第4章において、局所依存性が局所独立性を仮定したテスト情報量の推定に対しどのような影響を与えるのか、シミュレーションを通して検討を行った。その結果、誤って局所独立性を仮定したモデルをあてはめた場合、テスト情報量が過大推定されることがわかった。しかし、局所依存性の程度とこの過大推定の程度との関係は単純ではなく、局所依存度の強い項目群が多い場合には、過大推定の程度はむしろ小さくなっていった。

最後に第5章において、局所依存性を有するデータに対して局所依存性を考慮する複数のモデルと局所独立性を仮定するモデルをあてはめてシミュレーションを実施し、局所依存性を考慮することにより受験者特性の推定精度がどれほど改善されるのか、検討を行った。その結果、受験者特性の推定精度に関しては分析モデル間でほとんど差が見られず、特定のテストの成績のみを用いて個人差の評価を行うという文脈においても、特定のテストを超えて各受験者の受験者特性値を共通の尺度上で比較するという文脈においても、局所独立性を仮定するモデルに代えてより複雑な局所依存性を考慮するモデルを使用することを積極的に推奨するような結果は得られなかった。また、受験者数や局所依存関係にある項目の数等との関係については、分析モデル毎に異なる関係が得られたものの、共通する一般的な知見として、受験者数を増加させることにより、局所独立性を仮定した場合のバイアスが局所依存性を考慮した場合のバイアスと同等かそれよりも値が小さくなる、という関係が得られた。

6.2 本論文の限界点と今後の課題

本論文では、局所依存性が局所独立性を仮定した受験者特性および項目特性の推定に対しどのような影響を与えるのか、局所依存性を考慮したモデルは局所独立性を仮定したモデルに代えて使用するに値するものなのか、といったことについて、シミュレーションを通して検討を行った。しかしながら、今後、実際の回答データを用いて受験者特性や項目母数の推定を行う場面に対し本研究の知見を応用していくという観点からは、局所依存性があると想定される大問形式の実際のテストについて、実データに基づいて局所依存度を推

定したうえで、本研究のシミュレーションで得られた結果と同様の結果が得られるかの確認を行う必要があると考えられる。また、本論文では、局所依存性を考慮した現存のモデルが、受験者特性の推定において必ずしも有効に機能していないことがわかったことから、現実の局所依存のあり方をよりの確に反映した新たなモデルの開発に取り組むことも重要な課題であると考えられる。

引用文献

1. Anastasi, A. (1961). *Psychological testing* (2nd ed.). New York: Macmillan.
2. 荒井清佳・前川眞一 (2005). 日本の公的な大規模試験に見られる特徴—標準化の観点から— 日本テスト学会誌, **1**, 81-92.
3. Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more latent categories. *Psychometrika*, **37**, 29-51.
4. Bradlow, E. T., Wainer, H., & Wang, X. H. (1999). A Bayesian random effects model for testlets. *Psychometrika*, **64**, 153-168.
5. Braeken, J. (2011). A boundary mixture approach to violation of conditional independence. *Psychometrika*, **64**, 153-168.
6. Braeken, J., Tuerlinckx, F., & De Boeck, P. (2007). Copula functions for residual dependency. *Psychometrika*, **72**, 393-411.
7. Chen, C., & Wang, W. (2007). Effect of ignoring item interaction on item parameter estimation and detection of interacting items. *Applied Psychological Measurement*, **31**, 388-411.
8. DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement*, **43**, 145-168.
9. Ferrara, S., Huynh, H., & Michaels, H. (1999). Contextual explanation of local dependence in item clusters in a large scale hands-on science performance assessment. *Journal of Educational Measurement*, **6**, 119-140.
10. Frank, M. J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematicae*, **19**, 194-226.
11. Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *Institute of Electrical and Electronics Engineers, Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

12. Guilford, J. P. (1936). *Psychometric methods*. New York: McGraw-Hill.
13. 南風原朝和 (1984). テスト理論への個人正答確率に基づくアプローチ 新潟大学教育学部紀要, **26**, 21-28.
14. 南風原朝和 (2000). 個人正答確率に基づく局所独立性の概念の明確化 —— 実験的独立性および一次元性との関係を中心に —— <http://www.p.u-tokyo.ac.jp/haebara/local_ind/> (2012年5月10日)
15. Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97-109.
16. Hoskens, M., & De Boeck, P. (1997). A parametric model for local dependence among test items. *Psychological Methods*, **2**, 261-277.
17. 池田 央 (1992). テストの科学—試験に関わる全ての人に 日本文化化学社.
18. Ip E. H. (2000). Adjusting for information inflation due to local dependency in moderately large item clusters. *Psychometrika*, **65**, 73-91.
19. Ip, E. H. (2002). Locally dependent latent trait model and the Dutch identity revisited. *Psychometrika*, **67**, 367-386.
20. Ip, E. H. (2010). Interpretation of the three parameter testlet response model and information function. *Applied Psychological Measurement*, **34**, 467-482.
21. Ip, E. H., Smits, D. J. M., & De Boeck, P. (2009). Locally dependent linear logistic test model with person covariates. *Applied Psychological Measurement*, **33**, 555-569.
22. 石塚智一・中畝菜穂子・内田照久・前川眞一 (2001). テストレットモデルによる英語試験問題の分析 大学入試センター研究紀要, **30**, 21-38.
23. 泉 毅・山野井真児・山田剛史・白川隆朋・対馬英樹 (2013). 局所独立性を満たさないデータに対する段階反応モデルの適用 ~2PLMによる分析との比較検討~ 日本テスト学会誌, **9**, 37-55.
24. Jannarone, R. J. (1986). Conjunctive item response theory kernels. *Psychometrika*, **51**, 357-373.
25. Jiao, H., Kamata, A., Wang, S., & Jin, Ying. (2012). A multilevel testlet model for dual local dependence. *Journal of Educational Measurement*, **49** 82-100.
26. Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika*, **56**, 255-278.
27. Kan, C. C., van der Ven, A. H. G. S., Breteler, M. H. M., & Zitman, F. G. (2001). Latent trait standardization of the Benzodiazepine dependence self-report questionnaire using the Rasch scaling model. *Comprehensive Psychiatry*, **42**, 424-432.

28. Keller, L. A., Swaminathan, H., & Sireci, S. G. (2003). Evaluating scoring procedures for context-dependent item sets. *Applied Measurement in Education*, **16**, 207-222.
29. Kelly, T. L. (1924). Note on the reliability of a test: A reply to Dr. Crumm's criticism. *Journal of Educational Psychology*, **15**, 193-204.
30. Kreiner, S., & Christensen, K. B. (2004). Analysis of local dependence and multidimensionality in graphical loglinear Rasch models. *Communications in Statistics: Theory and Methods*, **33**, 1239-1276.
31. Lee, G. (2000). A comparison of methods of estimating conditional standard errors of measurement for testlet-based test scores using simulation techniques. *Journal of Educational Measurement*, **36**, 91-112.
32. Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets. *Applied Psychological Measurement*, **25**, 357-372.
33. Li, Y., Bolt, D. M., & Fu, J. (2005). A test characteristic curve linking method for the testlet model. *Applied Psychological Measurement*, **29**, 340-356.
34. Li, Y., Bolt, D. M., & Fu, J. (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement*, **30**, 3-21.
35. Looney, M. A., & Spray, J. A. (1992). Effects of violating local independence on IRT parameter estimation for the binomial trials model. *Research Quarterly for Exercise and Sport*, **63**, 356-359.
36. Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associate.
37. Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
38. Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, **16**, 159-176.
39. Nandakumar, R. (1990). Traditional dimensionality versus essential dimensionality. *Journal of Educational Measurement*, **28**, 99-117.
40. Neal, R. (1997). Markov chain Monte Carlo methods based on 'slicing' the density function. *Technical Report 9722*, Department of Statistics, University of Toronto, Canada.
41. Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, **24**, 146-178.

42. Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research.
43. Reise, S. P., Horan, W. P., & Blanchard, J. J. (2011). The challenges of fitting an item response theory model to the social anhedonia scale. *Journal of Personality Assessment*, **93**, 213-224.
44. Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, **34**, 100-114.
45. Simms, L. J., Goldberg, L. R., Roberts, J. E., Watson, D., Welte, J., & Rotterman, J. H. (2011). Computerized adaptive assessment of personality disorder: Introducing the CAT-PD project. *Journal of Personality Assessment*, **93**, 380-389.
46. Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, **28**, 237-247.
47. Spiegelhalter, D. J., Thomas, A., & Best, N. (2003). WinBUGS 1.4 [Computer Program]. Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health.
48. Spiegelhalter, D. J., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS User Manual Version 1.4. Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health.
49. Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528-550.
50. Thorndike, R. L. (1951). Reliability. In E. F. Lindquist (Ed.), *Educational Measurement*. Washington, DC: American Council on Education.
51. 登藤直弥 (2010). 局所独立性の仮定が満たされない場合の潜在特性推定への影響 日本テスト学会誌, **6**, 17-28.
52. 登藤直弥 (2012a). 大問形式の問題の項目群への項目反応に対する確率モデルの比較 日本テスト学会誌, **8**, 85-100.
53. 登藤直弥 (2012b). 項目反応間の局所依存性が項目母数の推定に与える影響——項目母数の比較可能性を確保した上での検討—— 行動計量学, **39**, 81-91.
54. 豊田秀樹 (2002). 項目反応理論 [入門編] ——テストと測定の科学—— 朝倉書店.
55. 豊田秀樹 (編著) (2005). 項目反応理論 [理論編] ——テストの数理—— 朝倉書店.
56. 豊田秀樹 (編著) (2008). マルコフ連鎖モンテカルロ法 朝倉書店.
57. Tuerlinckx, F., & De Boeck, P. (1999). Distinguishing constant and dimension-dependent interaction: A simulation study. *Applied Psychological Measurement*, **23**, 299-307.

58. Tuerlinckx, F., & De Boeck, P. (2001). The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological Methods*, **6**, 181-195.
59. Wainer, H. (1995). Precision and differential item functioning on a testlet-based test: The 1991 Law School Admission Test as an example. *Applied Measurement in Education*, **8**, 157-186.
60. Wainer, H., Bradlow, E. T., & Du, Z. (2000). Testlet response theory: An analog for the 3-PL useful in testlet-based adaptive testing. In W. J. van der Linden, & C. A. W. Glas (Eds.), *Computerized adaptive testing, theory and practice*. Boston, MA: Kluwer-Nijhoff. pp. 246-270.
61. Wainer, H., Bradlow E. T., & Wang X. H. (2007). *Testlet response theory and its applications*. USA: Cambridge University Press.
62. Wainer, H., & Wang, X. H. (2001). Using a new statistical model for testlets to score TOEFL. *TOEFL Technical Report No. 16*. Princeton, NJ: Educational Testing Service.
63. Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied Psychological Measurement*, **26**, 109-128.
64. Wang, C. W., Cheng, Y. Y., & Wilson, M. (2005). Local item dependence for items across tests connected by common stimuli. *Educational and Psychological Measurement*, **65**, 5-27.
65. Wang, C. W., & Wilson, M. (2005). Exploring local item dependence using a random-effect facet model. *Applied Psychological Measurement*, **29**, 296- 318.
66. 山野井真児・山田剛史 (2012). 局所依存性のあるデータ分析における項目反応モデルの比較——項目識別力の過大推定に着目して—— 日本教育心理学会第 54 回総会発表論文集, 536.
67. Yang, W. L., & Gao, R. (2008). Invariance of score linkings across gender groups for forms of a testlet-based college-level examination program examination. *Applied Psychological Measurement*, **32**, 45-61.
68. 安井彩乃 (2013). 項目の局所依存性が潜在特性推定値に及ぼす影響について 日本テスト学会第 11 回大会発表論文抄録集, 84-87.
69. Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, **30**, 187-213.
70. Zhang, B. (2010). Assessing the accuracy and consistency of language proficiency classification under competing measurement models. *Language Testing*,

27, 119-140.

付録

第 2 章で使用したプログラム

発生モデルが 2PLM の場合 ($(N = 1000, J = 50)$ の場合の例)

```
#乱数の種を設定する
set.seed(1010)

#被験者を 100 人発生させる
 $\theta$  t1 <- rnorm(n = 100)

#各テスト項目の困難度を発生させる
bt1 <- runif(n = 30, min = -2.0, max = 2.0)

#各テスト項目の識別力を発生させる
at1 <- runif(n = 30, min = 0.3, max = 1.5)

#残りの被験者 900 人を発生させる
 $\theta$  t2 <- rnorm(n = 900)
 $\theta$  t <- c( $\theta$  t1,  $\theta$  t2)

#残りの 20 個の項目を発生させる
bt2 <- runif(n = 20, min = -2.0, max = 2.0)
bt <- c(bt1, bt2)
at2 <- runif(n = 20, min = 0.3, max = 1.5)
at <- c(at1, at2)

#50 個の項目に対する 1000 人の正答確率・誤答確率 (2 パラメタロジスティック)
を計算する
 $\theta$  tvec <- rep( $\theta$  t, 50)
 $\theta$  tmat <- matrix( $\theta$  tvec, nrow = 50, ncol = 1000, byrow = TRUE)
ft <-  $\theta$  tmat - bt
et <- exp(-1.0 * at * ft)
Pt <- 1 / (1 + et)
Qt <- 1 - Pt

#推定値の入れ物を用意する
```



```
theta_m <- matrix(nrow = 1000, ncol = 100)
am <- matrix(nrow = 50, ncol = 100)
bm <- matrix(nrow = 50, ncol = 100)
Xe <- matrix(nrow = 1000, ncol = 100)
DPe <- matrix(nrow = 50, ncol = 100)
PRe <- matrix(nrow = 50, ncol = 100)
pc <- matrix(nrow = 3, ncol = 100)
kc <- matrix(nrow = 3, ncol = 100)
sc <- matrix(nrow = 3, ncol = 100)

#MCMCによる推定を100回行なう
for(r in 1:100){

#項目反応行列を作成する
一様乱数 <- runif(n = 1000 * 50, min = 0, max = 1)
一様乱数行列 <- matrix(一様乱数, nrow = 50, ncol = 1000)
U <- ifelse(Pt >= 一様乱数行列, 1, 0)

#メモリを解放する
gc()
gc()

#項目母数、ならびに、被験者母数の推定を行なう
#推定材料の入れ物を用意する
theta_e <- matrix(nrow = 1000, ncol = 20000)
be <- matrix(nrow = 50, ncol = 20000)
ae <- matrix(nrow = 50, ncol = 20000)

#各パラメタの初期値を決める
X <- colSums(U)
Xe[, r] <- X
mu <- mean(X)
Sigma <- mean((X - mu) ^ 2)
sigma <- sqrt(Sigma)
theta_e[, 1] <- (X - mu) / sigma
PR <- rowMeans(U)
PRe[, r] <- PR
be[, 1] <- 1 - PR
DP <- apply(U, 1, cor, X)
DPe[, r] <- DP
ae[, 1] <- DP
for(i in 1:19999){

#各項目への正答確率と誤答確率(2パラメタロジスティック)を計算する
theta_vec <- rep(theta_e[, i], 50)
theta_mat <- matrix(theta_vec, nrow = 50, ncol = 1000, byrow = TRUE)
```

```

f <-  $\theta$  mat - be[, i]
e <- exp(-1.0 * ae[, i] * f)
P <- 1 / (1 + e)
Q <- 1 - P

#続いて、次の時点における $\theta$ の候補を発生させる(ランダムウォーク連鎖)
t <- rnorm(n = 1000, mean = 0, sd = 1.7)
 $\theta$  c <-  $\theta$  e[, i] + t

#潜在特性値が $\theta$  cであった場合の、各項目に対する正答確率・誤答確率(2パラメタロジスティック)を求める
 $\theta$  cvec <- rep( $\theta$  c, 50)
 $\theta$  cmat <- matrix( $\theta$  cvec, nrow = 50, ncol = 1000, byrow = TRUE)
fc <-  $\theta$  cmat - be[, i]
ec <- exp(-1.0 * ae[, i] * fc)
Pc <- 1 / (1 + ec)
Qc <- 1 - Pc

# $\theta$  cを採択するかどうか、判断する
P  $\theta$  <- dnorm( $\theta$  e[, i], mean = 0, sd = 1)
P  $\theta$  c <- dnorm( $\theta$  c, mean = 0, sd = 1)
L  $\theta$  <- (P ^ U) * (Q ^ (1 - U))
L  $\theta$  c <- (Pc ^ U) * (Qc ^ (1 - U))
LP  $\theta$  <- log(P  $\theta$ )
LP  $\theta$  c <- log(P  $\theta$  c)
LL  $\theta$  <- log(L  $\theta$ )
LL  $\theta$  c <- log(L  $\theta$  c)
LL  $\Theta$  <- colSums(LL  $\theta$ )
LL  $\Theta$  c <- colSums(LL  $\theta$  c)
L  $\alpha$  <- (LL  $\Theta$  c + LP  $\theta$  c) - (LL  $\Theta$  + LP  $\theta$ )
 $\alpha$  c <- exp(L  $\alpha$ )
dim( $\alpha$  c) <- c(1, 1000)
 $\alpha$  <- apply( $\alpha$  c, 2, min, 1)
urn <- runif(n = 1000, min = 0, max = 1)
 $\alpha$  logic <-  $\alpha$  >= urn
 $\theta$  e[which( $\alpha$  logic), i+1] <-  $\theta$  c[which( $\alpha$  logic)]
 $\theta$  e[which(! $\alpha$  logic), i+1] <-  $\theta$  e[which(! $\alpha$  logic), i]

#次に、潜在特性値が推移した $\theta$ である場合の正答確率・誤答確率(2パラメタロジスティック)を求める
 $\theta$  vec <- rep( $\theta$  e[, i+1], 50)
 $\theta$  mat <- matrix( $\theta$  vec, nrow = 50, ncol = 1000, byrow = TRUE)
f <-  $\theta$  mat - be[, i]
e <- exp(-1.0 * ae[, i] * f)
P <- 1 / (1 + e)
Q <- 1 - P

```

```

#次の時点における a の推移先の候補を発生させる (ランダムウォーク連鎖)
ta <- rnorm(n = 50, mean = 0, sd = 0.25)
ac <- ae[, i] + ta

#識別力が ac であった場合の、各項目に対する正答確率・誤答確率 (2 パラメタロ
ジスティック) を求める
ec <- exp(-1.0 * ac * f)
Pc <- 1 / (1 + ec)
Qc <- 1 - Pc

#ac を採択するかどうか決める
Pa <- dnorm(ae[, i], mean = 1, sd = 0.5)
Pac <- dnorm(ac, mean = 1, sd = 0.5)
La <- (P ^ U) * (Q ^ (1 - U))
Lac <- (Pc ^ U) * (Qc ^ (1 - U))
LPa <- log(Pa)
LPac <- log(Pac)
LLa <- log(La)
LLac <- log(Lac)
LLA <- rowSums(LLa)
LLAc <- rowSums(LLac)
L $\alpha$  <- (LLAc + LPac) - (LLA + LPa)
 $\alpha$  c <- exp(L $\alpha$ )
dim( $\alpha$  c) <- c(1, 50)
 $\alpha$  <- apply( $\alpha$  c, 2, min, 1)
urn <- runif(n = 50, min = 0, max = 1)
 $\alpha$  logic <-  $\alpha$  >= urn
ae[which( $\alpha$  logic), i+1] <- ac[which( $\alpha$  logic)]
ae[which(! $\alpha$  logic), i+1] <- ae[which(! $\alpha$  logic), i]

#最後に、潜在特性値が推移した  $\theta$  で、識別力が推移した a である場合の正答確率・
誤答確率 (2 パラメタロジスティック) を求める
 $\theta$  vec <- rep( $\theta$  e[, i+1], 50)
 $\theta$  mat <- matrix( $\theta$  vec, nrow = 50, ncol = 1000, byrow = TRUE)
f <-  $\theta$  mat - be[, i]
e <- exp(-1.0 * ae[, i+1] * f)
P <- 1 / (1 + e)
Q <- 1 - P

#次の時点における b の推移先の候補を発生させる (ランダムウォーク連鎖)
tb <- rnorm(n = 50, mean = 0, sd = 0.3)
bc <- be[, i] + tb

#困難度が bc であった場合の、各項目に対する正答確率・誤答確率 (2 パラメタロ
ジスティック) を求める

```

```

fc <-  $\theta$  mat - bc
ec <- exp(-1.0 * ae[, i+1] * fc)
Pc <- 1 / (1 + ec)
Qc <- 1 - Pc

#bc を採択するかどうか決める
Pb <- dnorm(be[, i], mean = 0, sd = 1)
Pbc <- dnorm(bc, mean = 0, sd = 1)
Lb <- (P ^ U) * (Q ^ (1 - U))
Lbc <- (Pc ^ U) * (Qc ^ (1 - U))
LPb <- log(Pb)
LPbc <- log(Pbc)
LLb <- log(Lb)
LLbc <- log(Lbc)
LLB <- rowSums(LLb)
LLBc <- rowSums(LLbc)
L  $\alpha$  <- (LLBc + LPbc) - (LLB + LPb)
 $\alpha$  c <- exp(L  $\alpha$ )
dim( $\alpha$  c) <- c(1, 50)
 $\alpha$  <- apply( $\alpha$  c, 2, min, 1)
urn <- runif(n = 50, min = 0, max = 1)
 $\alpha$  logic <-  $\alpha$  >= urn
be[which( $\alpha$  logic), i+1] <- bc[which( $\alpha$  logic)]
be[which(! $\alpha$  logic), i+1] <- be[which(! $\alpha$  logic), i]
}

#メモリを解放する
gc()
gc()

#得られた標本から  $\theta$ 、a、b の推定値を計算する
 $\theta$  e <-  $\theta$  e[, -c(1:3000)]
ae <- ae[, -c(1:3000)]
be <- be[, -c(1:3000)]
 $\theta$  <- rowMeans( $\theta$  e)
a <- rowMeans(ae)
b <- rowMeans(be)
 $\theta$  m[, r] <-  $\theta$ 
am[, r] <- a
bm[, r] <- b
pc[1, r] <- cor( $\theta$  t,  $\theta$ , method = c("pearson"))
pc[2, r] <- cor(at, a, method = c("pearson"))
pc[3, r] <- cor(bt, b, method = c("pearson"))
kc[1, r] <- cor( $\theta$  t,  $\theta$ , method = c("kendall"))
kc[2, r] <- cor(at, a, method = c("kendall"))
kc[3, r] <- cor(bt, b, method = c("kendall"))

```

```
sc[1, r] <- cor( $\theta$  t,  $\theta$ , method = c("spearman"))
sc[2, r] <- cor(at, a, method = c("spearman"))
sc[3, r] <- cor(bt, b, method = c("spearman"))

#メモリを解放する
gc()
gc()
}

# $\theta \cdot a \cdot b$ について、Bias・RMSEの計算を行なう
Bias  $\theta$  <- rowMeans( $\theta$  m) -  $\theta$  t
Biasa <- rowMeans(am) - at
Biasb <- rowMeans(bm) - bt
RMSE  $\theta$  1 <- ( $\theta$  m -  $\theta$  t) ^ 2
RMSE  $\theta$  2 <- rowMeans(RMSE  $\theta$  1)
RMSE  $\theta$  <- sqrt(RMSE  $\theta$  2)
RMSEa1 <- (am - at) ^ 2
RMSEa2 <- rowMeans(RMSEa1)
RMSEa <- sqrt(RMSEa2)
RMSEb1 <- (bm - bt) ^ 2
RMSEb2 <- rowMeans(RMSEb1)
RMSEb <- sqrt(RMSEb2)

#Bias、RMSEの平均と標準偏差を求める
 $\mu$  B  $\theta$  <- mean(Bias  $\theta$ )
 $\mu$  Ba <- mean(Biasa)
 $\mu$  Bb <- mean(Biasb)
 $\mu$  R  $\theta$  <- mean(RMSE  $\theta$ )
 $\mu$  Ra <- mean(RMSEa)
 $\mu$  Rb <- mean(RMSEb)
 $\sigma$  B  $\theta$  <- sqrt(mean((Bias  $\theta$  -  $\mu$  B  $\theta$ ) ^ 2))
 $\sigma$  Ba <- sqrt(mean((Biasa -  $\mu$  Ba) ^ 2))
 $\sigma$  Bb <- sqrt(mean((Biasb -  $\mu$  Bb) ^ 2))
 $\sigma$  R  $\theta$  <- sqrt(mean((RMSE  $\theta$  -  $\mu$  R  $\theta$ ) ^ 2))
 $\sigma$  Ra <- sqrt(mean((RMSEa -  $\mu$  Ra) ^ 2))
 $\sigma$  Rb <- sqrt(mean((RMSEb -  $\mu$  Rb) ^ 2))

#ファイルに出力する

write.table( $\theta$  m, file = "独立  $\theta$  (1.0) 個別 1000 人 50 問固定せず.xls", sep = "\t")
write.table(am, file = "独立 a(1.0) 個別 1000 人 50 問固定せず.xls", sep = "\t")
write.table(bm, file = "独立 b(1.0) 個別 1000 人 50 問固定せず.xls", sep = "\t")
write.table(Xe, file = "独立 X(1.0) 個別 1000 人 50 問固定せず.xls", sep = "\t")
write.table(DPe, file = "独立 DP(1.0) 個別 1000 人 50 問固定せず.xls", sep = "\t")
```

```
write.table(PRe, file = "独立 PR(1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
write.table(pc, file = "独立 pc(1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
write.table(kc, file = "独立 kc(1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
write.table(sc, file = "独立 sc(1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
write.table(Bias  $\theta$ , file = "独立 Bias  $\theta$  (1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
write.table(Biasa, file = "独立 Biasa(1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
write.table(Biasb, file = "独立 Biasb(1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
write.table(RMSE  $\theta$ , file = "独立 RMSE  $\theta$  (1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
write.table(RMSEa, file = "独立 RMSEa(1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
write.table(RMSEb, file = "独立 RMSEb(1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")  
sv <- cbind( $\mu$  B  $\theta$ ,  $\mu$  Ba,  $\mu$  Bb,  $\mu$  R  $\theta$ ,  $\mu$  Ra,  $\mu$  Rb,  $\sigma$  B  $\theta$ ,  $\sigma$   
Ba,  $\sigma$  Bb,  $\sigma$  R  $\theta$ ,  $\sigma$  Ra,  $\sigma$  Rb)  
write.table(sv, file = "独立 sv(1.0) 個別 1000 人 50 問固定せ  
ず.xls", sep = "\t")
```

発生モデルが CCM の場合 ($(N = 500, J = 30)$ の場合の例)

```
#乱数の種を設定する  
set.seed(1010)  
  
#被験者を 100 人発生させる  
 $\theta$  t1 <- rnorm(n = 100)  
  
#各テスト項目の困難度を発生させる  
bt <- runif(n = 30, min = -2.0, max = 2.0)  
  
#各テスト項目の識別力を発生させる  
at <- runif(n = 30, min = 0.3, max = 1.5)  
  
#残りの被験者 400 人を発生させる  
 $\theta$  t2 <- rnorm(n = 400)  
 $\theta$  t <- c( $\theta$  t1,  $\theta$  t2)  
  
#30 個の項目に対する 500 人の正答確率・誤答確率 (2 パラメタロジスティック)  
を計算する  
 $\theta$  tvec <- rep( $\theta$  t, 30)
```

```
theta tmat <- matrix(theta tvec, nrow = 30, ncol = 500, byrow = TRUE)
ft <- theta tmat - bt
et <- exp(-1.0 * at * ft)
Pt <- 1 / (1 + et)
Qt <- 1 - Pt
```

#局所依存の関係にある項目 (item1 ~ item30) に、特定の反応を見せる確率を計算する

```
b12 <- (-2)
ec <- at * ft
eec <- exp(ec)
d1 <- eec[seq(1, 29, by = 2), ]
d2 <- eec[seq(2, 30, by = 2), ]
d3 <- exp(ec[seq(1, 29, by = 2), ] + ec[seq(2, 30, by = 2), ] - b12)
d <- 1 + d1 + d2 + d3
P00 <- 1 / d
P10 <- d1 / d
P01 <- d2 / d
P11 <- d3 / d
```

以下は発生モデルが 2PLM のときと同様なので割愛

発生モデルが 2PLCM の場合 (($N = 100, J = 10$) の場合の例)

```
#乱数の種を設定する
set.seed(1010)
```

```
#被験者を 100 人発生させる
theta t <- rnorm(n = 100)
```

```
#各テスト項目の困難度を発生させる
bt <- runif(n = 30, min = -2.0, max = 2.0)
```

```
#各テスト項目の識別力を発生させる
at <- runif(n = 30, min = 0.3, max = 1.5)
```

```
#テスト項目を 10 問にする
bt <- bt[c(1:10)]
at <- at[c(1:10)]
```

#10 個の項目に対する 100 人の正答確率・誤答確率 (2 パラメタロジスティック) を計算する

```
theta tvec <- rep(theta t, 10)
theta tmat <- matrix(theta tvec, nrow = 10, ncol = 100, byrow = TRUE)
ft <- theta tmat - bt
et <- exp(-1.0 * at * ft)
```

```

Pt <- 1 / (1 + et)
Qt <- 1 - Pt

#局所依存の関係にある項目 (item1 ~ item10) に、特定の反応を見せる確率を計算する
δ <- 30.0
C1 <- -δ * Qt
C2 <- 1 - exp(C1)
C3 <- log(C2)
C4 <- matrix(nrow = 5, ncol = 100)
for(i in 1:5){
C4[i, ] <- colSums(C3[c((2 * i - 1), (2 * i)), ])
}
C5 <- C4 - log(1 - exp(-δ))
C6 <- 1 - exp(C5)
C7 <- log(C6)
Cop <- (-1 / δ) * C7
P00 <- Cop
P10 <- Qt[seq(2, 10, by = 2), ] - Cop
P01 <- Qt[seq(1, 9, by = 2), ] - Cop
P11 <- 1 - Qt[seq(1, 9, by = 2), ] - Qt[seq(2, 10, by = 2), ] + Cop

```

以下は発生モデルが 2PLM のときと同様なので割愛

第 3, 4, 5 章で使用したプログラム

データの発生用プログラム ($(N = 1000, J_{d(j)} = 5, \text{強 } d(j) \times 0)$ の場合)

```

#乱数の種を設定
set.seed(1010)

#真の母数値を発生 (上位:1.186、下位:0.394)
tt <- rnorm(n = 1000, mean = 0, sd = 1)
bt <- rnorm(n = 20, mean = 0, sd = 1)
at <- runif(n = 20, min = 0.5, max = 1.5)
te1t <- rnorm(n = 1000, mean = 0, sd = 0.394)
te2t <- rnorm(n = 1000, mean = 0, sd = 0.394)
te3t <- rnorm(n = 1000, mean = 0, sd = 0.394)
te4t <- rnorm(n = 1000, mean = 0, sd = 0.394)
tet <- cbind(te1t, te2t, te3t, te4t)

#関数を定義
subtract <- function(x, y){x - y}

#20 個の項目に対する 1000 人の正答確率 (5 項目 ver.)

```



```

c1 <- apply(as.matrix(tt), 1, subtract, bt)
c2 <- numeric()
for(l in 1:4){
zp <- t(c1)[, (5*l-4):(5*l)] - tet[, l]
c2 <- cbind(c2, zp)
} #l
Pt <- 1/(1 + exp((-1.7)*t(t(c2)*at)))

#項目反応行列 (row & sum) の作成
U <- vector(mode = "list", length = 50)
S <- vector(mode = "list", length = 50)
for(i in 1:50){
u <- matrix(runif(n = 20000, min = 0, max = 1), nrow = 1000, ncol = 20)
U[[i]] <- ifelse(Pt >= u, 1, 0)
z <- numeric()
for(l in 1:4){
zp <- rowSums(U[[i]][, (5*l-4):(5*l)])
z <- cbind(z, zp)
} #l
S[[i]] <- z
} #i

#ファイルに出力する
for(l in 1:50){
write.table(t(as.vector(t(U[[l]]))), paste("U_",
formatC(l, format = "d", width = 2, flag = "0"), ".txt", sep = ""),
sep = ",", row.names = FALSE, col.names = FALSE)
write.table(t(as.vector(t(S[[l]]))), paste("S_",
formatC(l, format = "d", width = 2, flag = "0"), ".txt", sep = ""),
sep = ",", row.names = FALSE, col.names = FALSE)
} #l

```

WinBUGS で使用する data file (分析モデルが GRM で $J_{d(j)} = 5$ の場合)

```

#model file (GRM)
#5item ver.
model{
for(i in 1:N){
for(j in 1:J){
Ps[i, j, 6] <- 0
for(k in 1:5){
Ps[i, j, k] <- 1/(1 + exp(-1.7*a[j]*(t[i]-b[j, k])))
P[i, j, k+1] <- Ps[i, j, k] - Ps[i, j, k+1]
} #k
P[i, j, 1] <- 1 - Ps[i, j, 1]
U[i, j] <- S[i, j] + 1

```

```

        U[i, j] ~ dcat(P[i, j, ])
      } #j
      t[i] ~ dnorm(0, 1)
    } #i
    for(j in 1:J){
      a[j] ~ dnorm(1, 4)
      b[j, 1] ~ dnorm(0, 1)
      for(k in 2:5){
        b[j, k] ~ dnorm(0, 1)I(b[j, k-1], )
      } #k
    } #j
  }

```

WinBUGS で使用する model file (分析モデルが BTM で $J_{d(j)} = 3$ の場合)

```

#model file (TEM)
#3item ver.
model{
  for(i in 1:N){
    for(l in 1:4){
      for(j in (5*l-4):(5*l-2)){
        P[i, j] <- 1/(1+exp(-1.7*a[j]*(t[i]-b[j]-g[i, l])))
        U[i, j] ~ dbern(P[i, j])
      } #j
      for(j in (5*l-1):(1*5)){
        P[i, j] <- 1/(1+exp(-1.7*a[j]*(t[i]-b[j])))
        U[i, j] ~ dbern(P[i, j])
      } #j
      g[i, l] ~ dnorm(0, tau[l])
    } #l
    t[i] ~ dnorm(0, 1)
  } #i
  for(l in 1:4){
    tau[l] ~ dgamma(3, 1)
  } #l
  for(j in 1:J){
    a[j] ~ dnorm(1, 4)
    b[j] ~ dnorm(0, 1)
  } #j
} #model

```

WinBUGS で使用する model file (分析モデルが CCM で $J_{d(j)} = 2$ の場合)

```

#model file (CCM)
#2item ver.

```

```

model{
  for(i in 1:N){
    for(j in 1:10){
      z1[i, j] <- (1.7)*a[2*j-1]*(t[i]-b[2*j-1])
      z2[i, j] <- (1.7)*a[2*j]*(t[i]-b[2*j])
      dP[i, j] <- 1 + exp(z1[i, j]) + exp(z2[i, j])
        + exp(z1[i, j]+z2[i, j]-g[j])
      P[i, j, 1] <- 1/dP[i, j]
      P[i, j, 2] <- exp(z1[i, j])/dP[i, j]
      P[i, j, 3] <- exp(z2[i, j])/dP[i, j]
      P[i, j, 4] <- exp(z1[i, j]+z2[i, j]-g[j])/dP[i, j]
    } #j
    t[i] ~ dnorm(0, 1)
  } #i
  for(l in 1:L){
    a[l] ~ dnorm(1, 4)
    b[l] ~ dnorm(0, 1)
  } #j
  for(j in 1:4){
    g[j] ~ dnorm(-2, 1)
  } #j
  for(j in 5:10){
    g[j] <- 0
  } #j
  for(i in 1:N){
    for(j in 1:10){
      U[i, j] <- V[i, j] + 1
      U[i, j] ~ dcat(P[i, j, ])
    } #j
  } #i
} #model #ok

```

WinBUGS で使用する model file (分析モデルが 2PLM の場合)

```

model{
  for(i in 1:N){
    for(j in 1:J){
      P[i, j] <- 1/(1+exp(-1.7*a[j]*(t[i]-b[j])))
      U[i, j] ~ dbern(P[i, j])
    } #j
    t[i] ~ dnorm(0, 1)
  } #i
  for(j in 1:J){
    a[j] ~ dnorm(1, 4)
    b[j] ~ dnorm(0, 1)
  } #j

```

} #model

この他, WinBUGS にて MCMC を実行する際には, data file と inits file が必要となるが, これらの作成方法に関しては WinBUGS User Manual Version 1.4. (Spiegelhalter et al., 2003) を参照のこと.

謝辞

東京大学大学院教育学研究科の南風原朝和先生には、本論文を執筆するにあたり、大変貴重なご意見、ご指導を賜りました。心より感謝申し上げます。また、本日まで私を見守ってくれた私の両親に、心より感謝申し上げます。