

独立行政法人大学入試センター
入学者選抜研究機構入試評価部門報告書

大学入試の標準化、 多様化、および精密化

平成 23 年（2011 年）8 月

独立行政法人大学入試センター入学者選抜研究機構

独立行政法人大学入試センター
入学者選抜研究機構入試評価部門報告書

大学入試の標準化、 多様化、および精密化

目次

はじめに	11
第1章 大学入試へのベイズ的アプローチ	15
1.1 視点と目的	15
1.2 入試の結果の分割表表示	16
1.3 期待効用最大化とは	17
1.4 入試の結果の望ましさと損失	18
1.5 選抜方式の最適化	19
1.6 合否判定方法の最適化	20
1.7 合否判定資料の最適化	21
1.8 合否判定システムの事後評価	21
1.9 アドミッションズオフィスの仕事	22
1.10 具体的提言	24
1.11 入学者選抜研究機構発足記念セミナーにおけるスライド	25
第2章 セクターと難易度による分類の下での近年の大学入学者選抜の分析	35
2.1 はじめに	35
2.2 調査の概要	38
2.3 国立, 公立, 及び私立大学ごとの推移	38
2.3.1 センター試験	38
2.3.2 個別学力試験	41

2.3.3	非学力型選抜	44
2.4	難易度によってグループ分けした場合の推移	46
2.4.1	センター試験	48
2.4.2	個別学力試験	52
2.4.3	非学力型選抜	57
2.5	まとめ	62
第3章	小標本における選抜効果を補正する相関係数の推定について	67
3.1	導入	67
3.2	母数推定	71
3.2.1	最尤推定法	71
3.2.2	ベイズ推定法	73
3.3	シミュレーション	75
3.3.1	設定	75
3.3.2	結果	76
3.4	まとめと議論	77
3.5	Appendix	82
3.5.1	最尤推定のための R 関数	82
3.5.2	ベイズ推定のための R 関数	83
第4章	相関構造を仮定したテストレットモデルのベイズ推論	85
4.1	はじめに	85
4.2	テストレットに対する IRT モデルの提案	88
4.3	シミュレーション研究	92
4.4	事例分析	95
4.4.1	北アイルランドの標準テストデータ	95
4.4.2	日本における国立大学法人等の教育研究評価データ	100

4.4.3	事例分析の考察	106
4.5	まとめと今後の展開	107
4.6	補足: マルコフ連鎖モンテカルロ法による母数推定について	108
第5章	セミパラメトリックベイズの枠組みによる因果効果の推定	111
	引用文献	129

目次

2.1	センター試験の科目数の推移 (国立大学)	39
2.2	センター試験の科目数の推移 (公立大学)	40
2.3	センター試験の科目数の推移 (私立大学)	41
2.4	個別学力試験の科目数の推移 (国立大学)	42
2.5	個別学力試験の科目数の推移 (公立大学)	43
2.6	個別学力試験の科目数の推移 (私立大学)	44
2.7	非学力型選抜の動向 (国立大学)	45
2.8	非学力型選抜の動向 (公立大学)	46
2.9	非学力型選抜の動向 (私立大学)	47
2.10	グループ分けした場合のセンター試験の科目数の推移 (国立大学)	50
2.11	グループ分けした場合のセンター試験の科目数の推移 (公立大学)	51
2.12	グループ分けした場合のセンター試験の科目数の推移 (私立大学)	53
2.13	グループ分けした場合の個別学力試験の科目数の推移 (国立大学)	54
2.14	グループ分けした場合の個別学力試験の科目数の推移 (公立大学)	56
2.15	グループ分けした場合の個別学力試験の科目数の推移 (私立大学)	58
2.16	グループ分けした場合の非学力型選抜の動向 (国立大学)	60
2.17	グループ分けした場合の非学力型選抜の動向 (公立大学)	61
2.18	グループ分けした場合の非学力型選抜の動向 (私立大学)	63
3.1	人工データによる選抜効果の例	68

表目次

1.1	入試の4つの結果	17
2.1	選抜方式のデータの整理の形式	37
2.2	入学者選抜方式の推移の特徴	65
3.1	(x_i, y_i) が共に観測された標本のみから計算した, 希薄化を修正していない 相関係数の各条件における平均および標準偏差	78
3.2	最尤推定法とベイズ推定法による 10,000 回分の ρ の推定値の平均	79
3.3	最尤推定法とベイズ推定法による 10,000 回分の ρ の推定値の平均二乗誤 差 (MSE)	80
4.1	シミュレーションによる推定精度の評価	94
4.2	シミュレーションによる能力母数 θ に関する真値と推定値の相関行列	94
4.3	シミュレーションによる識別力母数 α に関する真値と推定値の相関行列	94
4.4	シミュレーションによる困難度母数 β に関する真値と推定値の相関行列	95
4.5	標準テストデータにおける単元項目間の相関	97
4.6	標準テストデータにおける識別力母数 α に関する各モデルによる推定値	98
4.7	標準テストデータにおける困難度母数 β に関する各モデルによる推定値	99
4.8	標準テストデータにおける能力母数 θ に関する各モデル推定値の相関行列	100
4.9	標準テストデータにおける識別力母数 α に関する各モデル推定値の相関行列	100
4.10	標準テストデータにおける困難度母数 β に関する各モデル推定値の相関行列	101
4.11	国立大学法人等の教育研究評価データにおける項目間の相関	102

4.12 国立大学法人等の教育研究評価データにおける識別力母数 α に関する各モデルによる推定値	103
4.13 国立大学法人等の教育研究評価データにおける困難度母数 β に関する各モデルによる推定値	104
4.14 国立大学法人等の教育研究評価データにおける能力母数 θ に関する各モデル推定値の相関行列	104
4.15 国立大学法人等の教育研究評価データにおける識別力母数 α に関する各モデル推定値の相関行列	105
4.16 国立大学法人等の教育研究評価データにおける困難度母数 β に関する各モデル推定値の相関行列	105

はじめに

大学入試センターの入学者選抜研究機構には3つの部門があり、それぞれに2つのプロジェクトを持つ。よって、合計6つのプロジェクトがある。本報告書は、それらのうちの1つである「大学入試の標準化、多様化、および精密化」プロジェクトの最初の活動報告である。

本プロジェクトの目的は、大学入試における理想を念頭に置きながら、その理想を実現するための現実的な方策を考えることである。この目的を達成するための当面の目標として、次の2つを考えている。1つは、大学入試において最小限やるべきこと、またやってはならないこと(DO's and DON'Ts)をまとめることであり、大学入試のスタンダードを作ることである。もう1つの目標は、アドミッションズオフィスの仕事を明確化し、その仕事のそれぞれについて、実践的な方策を提言することである。アドミッションズオフィスは、日本では、AO入試を実施する組織のようなイメージがあるが、本来のアドミッションズオフィスに期待される業務は多様であり、広範に渡る。

入試に関しては、当然のことながら、凡百の議論がある。どちらかといえば、その弊害についての指摘が多く、入試がいかにかに若者の生活を捻じ曲げているかがマスコミでもよく取り上げられる。それらは実際のことであり、入試に関わるものが心しておくべきことである。しかし一方で、入試が受験生に対して大学での学習に備えるための勉強の動機付けになるというようなプラスの面があることも忘れてはならない。入試の功罪を的確に把握し、入試システムを決めなければならない。この決定は、方程式の解のような正解があるわけではない。むしろ、各大学が独自の個性を発揮する入試システムを採用し、結果的に多様な入試システムが存在するほうが望ましいと考える。入試システムの設計と実施に大いに力

をふるうのが、アドミッションズオフィスである。アドミッションズオフィスという組織を実際にたちあげても、その仕事をこなす人材がないという意見がある。しかし、私達は、アドミッションズオフィスが実際にでき、その仕事の重要性についての認識が広まれば、志望する人も多くなるであろうと考えている。ある程度の統計学的な知識を持っていることが望ましいが、アドミッションズオフィスが必要とする人材の専門分野を制限する必要はなく、多くの分野の人材が協力して、アドミッションズオフィスを作っていく機運が望ましい。

本報告書は、意思決定として入試を見るという基本姿勢を示す論考や、大学入試の時間的展望、また最近行った入試に関わる方法論的な試みなどを収録している。我々のプロジェクトは始まったばかりであり、未熟な点も多く、まだ十分に具体的ではないかも知れない。多様なバックグラウンドを持つ多くの方から、ご意見・ご叱正をいただき、軌道修正をするつもりである。

大学入試センター入学者選抜研究機構
「大学入試の標準化、多様化、および精密化」プロジェクト
繁梲 算男, 小谷野 仁

研究組織

研究代表者

繁柵 算男 大学入試センター入学者選抜研究機構 客員教授

研究分担者

小谷野 仁 大学入試センター入学者選抜研究機構 特任助教

研究協力者

林 篤裕 九州大学高等教育開発推進センター 教授

大森 拓哉 多摩大学経営情報学部 准教授

星野 崇宏 名古屋大学大学院経済学研究科 准教授

岡田 謙介 専修大学人間科学部 講師

森 一将 東京大学教養学部附属教養教育高度化機構 特任講師

第1章 大学入試へのベイズ的アプローチ

繁柵 算男

1.1 視点と目的

この小論では、各大学が行う入試は、それぞれの大学が“望ましい”学生を選抜するための意思決定システムであるということを経験的な視点とする。現在は、大学全入時代(ユニバーサル化)といわれ、相対的な人数合わせという観点からいえば、進学希望者のほぼ全員が大学に合格できるようになっている。しかし、だからといって、大学入試の選抜性がなくなっていると世の中を観察するのは間違っている。有力な大学では相変わらずの激しい受験競争があり、また受験生の全員を合格とせざるをえない大学でも、高校生に大学の魅力を伝え、志願者を増やす努力をすることによって、ある種の選抜性を入試に求めるべきである。自分の長所を見つけ、大学教育によって実力を伸ばしてくれそうな入試をする大学は、学生にとって魅力的である。入試はそれぞれの大学がどのような学生に入学してほしいかを具体的に示すメッセージであり、個々の大学が入試をより良くしようと努力するならば、日本の教育力の底上げにつながるであろう。

本論考の目的は、大学入試を意思決定システムと見なし、入試という合否決定システムを数式で表現することを通して、入試を改善するためには何が問題となるか、またその問題の解決のためにはどうすればいいかを明らかにすることである。(意思決定については、数多くの書物がある。たとえば、ベイズ統計学の最近の発展について、多くの成果を網羅しているのは、Dey and Rao [8] である。) 以下の論考でいくつか数式を使うが、それは計算のためではなく、また証明のためでもなく、ただ内容を正確に表現し、具体的方策の指針を得

1. 大学入試へのベイズ的アプローチ

るためである。数式による表現はその意味が明確であるため、表現の違いや個人的な経験に伴う情緒的価値観から比較的自由に、大学入試において何をなすべきかという課題を明らかにでき、具体的にその課題に対する対処法を考えることができるという利点がある。

1.2 入試の結果の分割表表示

大学入試において、志願者の一人一人に対して大学がとりうる行動は、その志願者を合格させる (a_1 とする) か不合格とする (a_2 とする) かだけである。一方、その決定を受ける志願者は多様であり、受験生は一人一人別々の結果を受ける。しかし、ここでは、最大限の単純化をし、受験生の受験時点での状態を、受験した大学の教育を享受する資質を持っているという状態と、その大学での教育を享受するには準備性が不十分であるという状態の2つに分けよう。前者を(大学にとって)望ましい受験生 (s_1)、後者を(大学にとって)望ましくない受験生 (s_2) とする。その結果は、 2×2 の分割表にまとめられる。なお、合格者を決めることを選抜ということもあるが、選抜という言葉には多くの中から「少数の」良いもの、優れたものを選ぶというニュアンスがあり、ここでは合否判定ということにする。(大辞林によれば、「」に入れた「少数の」という意味は本来の意味ではないようであるが)。

この表では、結果は4つに分けられる。それぞれについて、短い記述をしておく。この図は極度に単純化してある。各受験生についての望ましさが異なることなどを考慮して、全体的な評価をする必要があるが、少なくとも、この単純化した表は入試に絡む議論の共通の土俵になるものである。これら4つの結果はお互いにトレードオフの関係にあり、どれか1つだけを考慮して、全体の望ましさ(あるいは損失)を最大(あるいは最小)にすることは不可能である。たとえば、受験生になったとして、あるいはその父兄になったとして、望ましい受験生が不合格になるような結果はあってはならないとし、そのほかの損失よりもはるかに大きく「比較にならない」とすると、残された方式は全員入学しかありえないが、それは個々の大学の独自性を損ない、現実的ではない。

表 1.1. 入試の 4 つの結果

	望ましい受験生 (s_1)	望ましくない受験生 (s_2)
合格 (a_1)	GOOD (結果 1)	大学資源の非効率運用 (結果 2)
不合格 (a_2)	残念 (結果 3)	やむをえず (結果 4)

1.3 期待効用最大化とは

入試を表 1.1 のようにまとめると、大学入試の最適化が意味するところは単純である。大学の立場からすれば、大学入試は、この選抜から大学が最も益するところが大きいように決定をすればよいということになる。あるいは同じことであるが、大学入試によってもたらされる損失を最小化すればよい。しかし、単純に表現することはできても、これを実行することは簡単ではない。2 つの課題が存在する。

- (1) 常に、合格させるべき特徴を持っている受験生を合格とし、合格に値しない受験生を不合格とする決定を行えるならば、理想的であるが、現実にはこれは難しい。したがって、この理想になるべく近くなるように、合格させるべき受験生を合格させる確率を最大化し、合格に値しない受験生を合格とする確率を最小化することが、1 つ目の課題である。
- (2) 2 つ目の課題は、入試の後に、不合格とすべき受験生を合格させた損失と、合格させるべき受験生を不合格とした損失を評価することである。

前者の課題 (1) は、言い換えれば、どのような入試資料と合否決定方法を用いればよいかという問題である。この課題は、第 1.5 節と第 1.6 節で論じる。後者の課題 (2) は、合格させるべきではない受験生を合格させた場合の損失と、合格させるべき受験生を不合格とした損失を評価することと考えてもよい。この言い方は、厳密には、ベイズ的アプローチから見て正確ではないし、このように表現したからといって、実際の評価の困難さが消えるわけではないが、考えるべき課題の本質を伝えることはできる。各大学は、予想される損失を最も

1. 大学入試へのベイズ的アプローチ

小さくするように、合否の決定を行えばよい。

実際には、受験生はそれぞれ異なっており、評価の観点も多様である。受験生を合格に値するものとそうではないものに二分することはできない。次の節では、各受験生に大学に入ってもらいたい程度を評価する方法について論じる。この望ましさの程度をベイズ的アプローチの用語にならって効用と呼ぶ。(違和感のある用語の使い方であろうが、統計理論との橋渡しのためである)。

1.4 入試の結果の望ましさと損失

意思決定問題として大学入試を考えると、表 1.1 はわかりやすい図式であるが、実際には、各受験生の (大学にとっての) 望ましさは、合格か不合格かに二分されるものではなく、連続的なものである。また、理論的な展開のためにもそのほうが都合がよい。その望ましさを各受験生の効用とよぶ。今、その効用は、大学において発揮される各人の特質 (基準 y) の関数であるとしよう。大学における特質とは、大学でのパフォーマンスに関連するもの (例: 授業の成績), 社会に対する貢献 (例: ボランティア活動), 課外活動 (例: サークル活動), 学習意欲に関連するもの (例: 出席率), 適応状態 (例: 留年, 退学) などである。すなわち、 $u_i = u(y_i)$ である。入試を合否判定の意思決定問題とすると、それは、基準 (y) を選抜資料 (x) によって予測し、その予測に基づき、各受検者の効用を評価し、予想される全体の効用を最大化するように合格か不合格かを定める問題であるといえる。学生の望ましさを構成概念 (学力, 学士力, やる気など) の関数として表現したほうがよい場合には、その構成概念の指標 (潜在変数) を推定し、決定はこの推定された分布に基づいて行う。(構成概念と観測される変数 y の関係を統計モデルとして設定する必要があり、議論がますます複雑になるので、この小論では潜在変数についてはこれ以降とり上げない)

この効用の評価には、各大学の方針がはっきりと正確に反映されるべきである。各大学のアドミッションポリシーは具体的でなければならないといわれることが多いが、意思決定の図式でいえば、この効用の評価が具体的にできるように、アドミッションポリシーは

書かれていなければならない。その意味では、構成概念よりも、具体的な指標で各受験生の将来の望ましさを評価したほうがよいように思われる。以下の議論は、それを具体的にイメージできるように、(それでなくても複雑な議論であるから、) 各受験生の望ましさは1つの基準 y_i によって評価されると考えて読んでいただいてもよい。この基準は、たとえば大学における成績 (g.p.a.) である。

1.5 選抜方式の最適化

当然のことながら、大学での成績は、入試の時には得られていない情報である。入試の時に意思決定のために利用できる情報は入試資料 x である。入試資料は複数得られるので、 x はベクトルである。議論に必要な概念を整理し、次のように記号化する。

$y = {}^t(y_1, \dots, y_q)$ 基準 (例: g.p.a., 出席率),

$x = {}^t(x_1, \dots, x_p)$ 選抜資料 (例: 入学試験, 面接評価, 内申書),

a_1 代替案 1 (合格という決定),

a_2 代替案 2 (不合格という決定),

$u(a_1, y) = u_1(y)$ 基準 y を持つ受験生を合格とした場合の効用,

$u(a_2, y) = u_2(y)$ 基準 y を持つ受験生を不合格とした場合の効用.

基準 y は受験時には未知であり、個々の受験者の効用は推測するしかない。個々の受験者の効用は、この不確実性を考慮して次のように計算できる。

$$\begin{aligned}\bar{u}_1(i|x_i) &= \int u_1(y_i) p(y_i|x_i) dy_i, \\ \bar{u}_2(i|x_i) &= \int u_2(y_i) p(y_i|x_i) dy_i.\end{aligned}$$

ここで使われている $p(y|x)$ は予測分布とよばれる分布であり、選抜資料 x_i が与えられた場合に、それを所与として基準 y を予測する分布である。この予測では1つの数値で表されるようなものは適切ではなく、不確実性を考慮して分布による予測が行われる。

1. 大学入試へのベイズ的アプローチ

得られたデータを活用するには、選抜資料 x_i と基準 y_i をつなぐモデルが必要である。データの情報によって、モデルの未知の部分、すなわち未知パラメータに関する推測を行うのである。たとえば、モデルとして、単純に次のような多変量回帰モデルを考える。

$$p(y_i|x_i) = N_{q \times (p+1)}(Bx_i, \Sigma).$$

ここで、 B は回帰係数の行列であり、 Σ は誤差の分散共分散行列である。 x_i と y_i についての複数の観測値の組を Data とおくと、このモデルのもとでは、

$$p(y|x, \text{Data}) = \int \int p(y|x, B, \Sigma) p(B, \Sigma | \text{Data}) dB d\Sigma$$

となる。パラメータに関する知識は、ベイズの定理によってアップデートされる。

1.6 合否判定方法の最適化

合否判定は、受験者の数が合格可能数を上回る場合には、 $u_1(i|x_i)$ の大きい受験者から定員に達するまで合格とするという方式で行えばよい。定員をある程度柔軟に設定できるならば、 $u_1(i|x_i)$ と $u_2(i|x_i)$ を比較し、相対的に決めるのがよい。問題は、受験生の評価を1次元で行えない場合である。受験生を、大学におけるパフォーマンスやまじめさ、将来希望する進路に進む熱意など、多様な観点から評価したいとき、合否決定方式は複雑になりうる。合格者の集団が多様性をもつようにしたい場合には、合否判定方式も複数用意するべきである。すなわち、合格者は複数の経路 (パス) によって決められるべきである。

判定方式として、多段階の決定方式を用いるかという問題もある。ただし、この問題においては、コストを問題にしない限り、多段階の決定方式を用いる理由はない。選抜資料の収集にコストがかかるからこそ、たとえば、スクリーニングとしての1段階目と、丁寧に資料を収集して判定する2段階目が必要なのである。判定方式 d のコストを c_i とするとき、

$$u(d) = \sum_i \{ \bar{u}_1(i|x_i) - \bar{u}_2(i|x_i) - c_i \}$$

がその決定方式の予想される効用であり、この $u(d)$ を最大化するものが最適な合否判定方法である。(ただし、効用とコスト c_i は、同じ原点と同じ目盛を持つ間隔尺度ではないので、現実にこれを実行するのは難しい。このような現実的困難については、第 1.9 節でまとめて論じる。)

1.7 合否判定資料の最適化

合否判定資料として最も良いのは、合格とした場合の効用と不合格とした場合の効用の差を最も大きくする資料の組み合わせである。理想的には、受験生一人一人に対して最も適切な組み合わせを選ぶべきであるが、総体として最も良い資料は、下記の情報量 (I) を最大化することによって得られる。

$$I = \sum_i \int \{u_{1i}(y_i)p(y_i|x_i, \text{Data}) - u_{2i}(y_i)p(y_i|x_i, \text{Data})\} dy_i - \sum_i \int \{u_{1i}(y_i)p(y_i|x_i) - u_{2i}(y_i)p(y_i|x_i)\} dy_i. \quad (1.1)$$

たとえば、大学における学力習得に重きを置くような基準を用いるならば、入学試験の結果が有力な判定資料になるであろう。将来の就職のための技量や態度がどの程度準備されているかに重きを置くならば、判定資料として面接の印象を重視するかもしれない。

以上で、ベイズ的アプローチの枠組みを用いて、意思決定としての合否判定システムの構図を示した。この後、第 1.8 節では、このシステムがうまく機能したかどうかを評価する方法について論じ、第 1.9 節では、この判定方式の具体化について述べる。

1.8 合否判定システムの事後評価

実際に上記のベイズ的合否判定システムを用いた後、これが目論見どおりに目的を果たしているかどうかを点検することは必須である。このシステムは、いくつかの仮定と、入試の時点では観測されていない基準 (y , たとえば g.p.a.) に基づいて設計されているので、実

1. 大学入試へのベイズ的アプローチ

際に基準が観測された後、判定システムが良い結果をもたらしたかどうかを点検する必要がある。この点検には多様な観点がありうる。その多様な観点からの評価をどのように総合するかが最も重要である。そのことを踏まえたうえで、統計学的な点検について説明する。上記のように、どのような入試資料に基づき、どのような合否判的方式をとるか、そしてどの受験生を合格とするかという決定のカギは、入試資料 x を所与としたときの基準 y の予測分布である。この予測分布から、十分な精度の予測が得られるかを見極める必要がある。予測分布に関する点検は、統計的には、予測分布に含まれるパラメータの推定から始まる。その後、推定されたパラメータによる予測の成績を、推定に用いていないデータ (検証用データ) を使って検証する。この2つの過程で特に気をつけなければならないのは、基準 y が合格者に対してのみしか得られていない点である。すべての受験生に対してある分布を仮定するのだが、実際には、一部の受験生からの観測値しか得られないから、これは1つの欠損値問題である。この問題については、岡田と繁樹 [31] を参照していただきたい。この論文における結果は、現在では、より一般的な状況に拡張されていることも指摘しておく。

1.9 アドミッションズオフィスの仕事

ここまでの議論によって、ベイズ的アプローチを用いた入試システムの講義調の説明は終わる。入試に関する議論は、政治家によるもの、教育学者によるもの、大学生、受験生、あるいはその家族によるものなどがあり、トーンが様々である。ここまで読まれた方 (少数だと思われるが) の中には、以上の議論がどのように現実の入試の問題に関わるのか疑問に思われた方も多いであろう (少数の中の多数派である)。本節では、大学入試に関する組織 (アドミッションズオフィスと書くことにする) がどのようなことをなすべきかについて、ここまで述べてきた観点からまとめる。まず、アドミッションズオフィスがやるべきことは、次のような4つのステップに分けられる。

- (1) どのような受験生に大学に入学してほしいのかを具体的に述べたアドミッションポリシーを策定する援助をする。

- (2) それぞれのアドミSSIONポリシーに則り, 各受験生の望ましさを入試資料から数値化する.
- (3) 望ましい受験生を合格させ, そうでない受験生を不合格とする合否判定システムを設計し, 各受験生に適用する.
- (4) 入試判定システムがうまく機能しているかどうかを, 入試の後で評価する.

以上の4つのステップのそれぞれについて, 前節までの数式を交えた説明は, 具体的指針を提示している.

- (1) アドミSSIONポリシーを明確に設定する必要があることは, いたるところで主張されている. この小論で強調したいことは, 受験生の望ましさが大学生活において観察される指標の関数として表現できるかどうかを意識して, 机上の空論でないアドミSSIONポリシーを作成すべきであるということである. たとえば, 学士力や就職力を高めることのできる学生を入学させたいという目標は, アドミSSIONポリシーとしてはあいまいに過ぎる. 観測されない概念 (構成概念) をアドミSSIONポリシーに加える場合には, その概念について丁寧に議論を行ない, 間接的にでも測定できるように概念を精錬させておかなければならない.
- (2) アドミSSIONポリシーに基づいて, 大学生活で観測される変数 y や議論によって明確にした潜在変数 z によって受験生の望ましさが表現されたならば, それをもとに, 具体的に各受験生の評価を数値化する必要がある. ここで, 受験生は数値化できないという議論は通用しないと考える. 合格か不合格かを決めている限り, 順序を付けているのである. 数値化に反対する論者は, (a) 望ましさを実数として表現できるはずがない, (b) それぞれに個性を持つ受験生の望ましさを1次元の数値に集約するのは, いたずらに競争意識を掻き立てるだけである, などの主張を論拠にするのかもしれない. 理由 (a) に対するこの小論の立場は以下のようなものである. 実際, 受験生の望ましさを実数で表現するか, 順序で表現するか, 表 1.1 のような, 合格させるべき, 不合格やむなしという単純な二分法のカテゴリーで表現するかは, 測定論的問題とし

1. 大学入試へのベイズ的アプローチ

て残る. ここでは「大は小を兼ねる」という考え方をとり, より精緻な測定を心がけ, その測定値を使う際に適当な粗さ (精緻さ) を選べばよい. 理由 (b) に対しては, 1次元ではなく, 多次元で測定することもできることを指摘しておく. このように受験生の望ましさが数値化されるならば, これを最も良く予測する入試資料の組み合わせを選ぶことが次のステップである. (より正確に言えば, 合格させた場合の効用と, 不合格とした場合の効用の差を最大化するものとして定義される. 式 (1.1) を参照.)

- (3) 合否の判定システムは, 入試資料を (a) 総合点方式で統合する, (b) それぞれの資料で基準点を設け, すべての資料において基準点以上であることを合格の条件とする, (c) それぞれの資料で基準点を設けるが, 少なくとも1つの資料において (あるいは2つ以上, etc.) 基準点を超えていることを合格の条件とする, などが考えられる. さらに, 上記 (a), (b), (c) の混合もありうる. 実際には入試資料の取得にはコストがかかるから, 1次のスクリーニングを行う必要があることもあるだろう. 一般に, 何段階で合否の決定を行うかは, コストの問題がカギとなる.
- (4) アドミッションポリシーに基づいて受験生の望ましさを指標化し, それを最も良く予測する入試資料を確定し, 効率的な合否決定システムを設計した後は, 妥当性検証のための追試研究を行って, そのパフォーマンスを検証する必要がある. そのためには, 受験生と入学した学生についてデータベースを作る必要がある.

1.10 具体的提言

最後に, アドミッションズオフィスの業務において, この小論で説明した統計的意思決定の専門家に何ができるのか, 具体的な提言を行いたい. アドミッションポリシーは, アドミッションズオフィスではなく, 教授会や理事会などで決定されるが, 統計や心理測定の専門家は, 先に述べたように, 観測できる事象とそれを結び付けるようにアドバイスをすることができる. また, 受験生の望ましさの評価を, 1つの指標 (たとえば, g.p.a.) によってではなく, 複数の指標を統合して行うほうが, 大学の構成員を多様にするために望まし

い。この統合について、心理測定の専門家は方針を示すことができる。専門用語でいえば、大学生の複数の属性を同時に考慮して、その望ましさを評価する問題は、多属性の効用の評価の問題であり、たとえば AHP などの手法が使える。入試資料や合否決定方式の最適化は、統計理論的には上記のとおりであるが、実際には、この最適化は機械的に容易に行えるわけではなく、ノウハウが必要である。専門家はそれらをアドバイスすることもできる。

1.11 入学者選抜研究機構発足記念セミナーにおけるスライド

平成22年9月10日

大学入学者選抜の必要条件

繁樹 算男
 入学者選抜研究機構(入試評価部門)

本発表の視点

- 入学試験は、大学が“望ましい”学生を選抜するための意思決定システムである。
- 大学全入時代(希望者ほぼ全員入学、ユニバーサル化)においても、入試をよくする努力は大切である。
- 個々の大学の努力が、日本の教育力の増進にも役に立つことを期待する。

日本の入試のユニバーサル化

入試の多様化の現状(1)

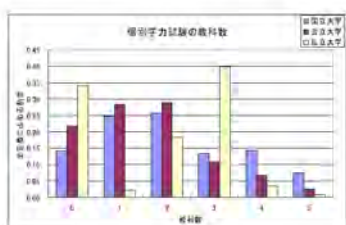
2011年度入試

(定員数による、小谷野、繁樹、準備中)

1. 大学入試へのベイズ的アプローチ

5

入試の多様化の現状(2)



6

アメリカのユニヴァーサル化

- 2007年 大学進学率48.4% しかし、多様
– (Shulenburger, 2010)
- **Less Selective** Colleges and Universities
– 高校の成績のみ、共通テストはplacementのため
- **More Selective** Colleges and Universities
– 科目修得表、成績、推薦状、共通テスト、論文
– しかし、大学の授業に準備ができていると判断されれば合格

7

アメリカの多様な入試

- **Highly Selective** Colleges and Universities
– 受検者は学業成績には優れていることが保障されているので他の要因(推薦状、小論文、面接、課外活動、階層的バラつきなど)が選抜において相対的に重要になる。

	合格者 10%以内	合格者 20%以内	計
公立大学	1	7	653
私立大学	8	29	2022

8

OUTLINE

- 入試は大学を主体とする意思決定システムである。
- その意思決定を効果的なものにするためには何が必要か? ⇒ 大学入試の必要条件
- その必要条件を踏まえて、具体的な提案をする。

9

意思決定としての大学入試

10

意思決定

選択肢	望ましい学生	望ましくない学生
合格(a_1)	GOOD 結果1	大学資源の非効率運用 結果2
不合格(a_2)	残念 結果3	やむを得ず 結果4

11

結果の望ましさの評価

- アドミッションポリシー⇒
4つの結果のそれぞれの望ましさの評価＝
どのような学生が入学することを望むか
- Q1:この望ましさは多属性評価である。たとえば次の属性の望ましさを相対的にどう評価するか？
- 第1属性(達成度、学力、結晶性)
- 第2属性(潜在的可能性、知能、流動性)
- 第3属性(非認知的次元、社会的かしこさ、やる気、Emotional Intelligence)

12

流動性vs結晶性

- 高校まで鋭意努力した結果としての知識の集成の程度を問うのか、将来の大学生活や職業生活におけるパフォーマンスを予測する潜在的な能力を重視するか？
- Q2:どの程度の勉強をすることによって解決できる問題をテストとするか。(e.g.週に何時間くらい勉強する受験生をよしとするか?)

13

社会的スキル

- 社会的かしこさ(木下,2006)
(実際に社会を生きていくうえでの知恵)
5つの要素:大局的発想、状況適合性、
相対化、内的世界の拡がり、感情の制御)
- 感情的知性(Emotional Intelligence)
- やる気
- Q3:面接はこれらを測れるか?面接をどのように構造化するか?A君に対して、3人の面接官が、3点、2点、1点を付けた。B君に対しては、全員2点であった。どちらを合格とさせるか?

14

どの属性を重視するか?



15

4人の受験生類型(1)



A君 受験優等生:受験のために最大限の努力をする。予備校にも通っている。その代わりに、スポーツや読書などはしない。まして、TVを見ることなどは時間の浪費であると思っている。孤立しているわけではないが、友人を作る努力もしない。学力テストの成績は、標準得点で75。

B君 バランス型:受験の勉強もするが、3年の夏までは生徒会活動でがんばっていた。友人が多く、適当に遊んでいる。学力テスト得点65。



16

4人の受験生類型(2)



C君 スポーツマン:勉強がきらいだが、特に興味のある理科については時々非常に良い得点を取る。ラグビー部員でレギュラー。遊びにエネルギーを費やしており、友人の数は多くないが非常に仲のよいグループがある。学力テスト得点50。

D君 芸術家:絵を描くのが得意で、市の展覧会に入選したこともある。孤立を好み、ユニークな考え方で時々人を驚かせる。学力テスト得点 55。



1. 大学入試へのベイズ的アプローチ

17

4人の受験生の評価



18

必要条件1

アドミッションポリシーの具体化

- 大学入試の目標を具体的かつ明快に定めること
- 入試はその目標を達成するための方便であることを意識すること
- できれば受験生に対し、大学においてどのような特性(パフォーマンス)を期待するかを具体的に述べること

19

選抜システムの設計

- 意思決定に携わる意見を集約し、アドミッションポリシーを決定する。(提案:入試関係者からの意見収集(アンケートの実施)⇒クラスター化(クラスター分析, KJ法など)による情報集約とフィードバック⇒意見の収束)
- アドミッションポリシーに基づき、各々の受験生の望ましさを評価する方式を決定する。たとえば、受験生の各属性における評価値と各属性の重要度を総合する。(提案:多属性効用の推定の手法を用いる。)

20

必要条件2

適切な選抜資料の選択

- アドミッションポリシーで目標とする重要なパフォーマンスを明確にしたうえで、それを最もよく予測する資料を選択する。(提案:想定される合格者の望ましさとその他の受験生のそれとの差を最大化するような資料を選択する。)
- 別の言い方では、求めるパフォーマンスを予測する測定用具として、信頼性や妥当性の高い資料を集めること
- Q4:選抜資料の収集に公平性や透明性、客観性を求めるか?

21

必要条件3

合否決定方法が目的に合うこと

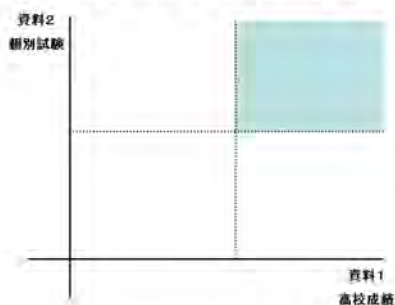
- 目標とする特性を最もよく実現するという意味で合否決定方法は合理的でなければならない。

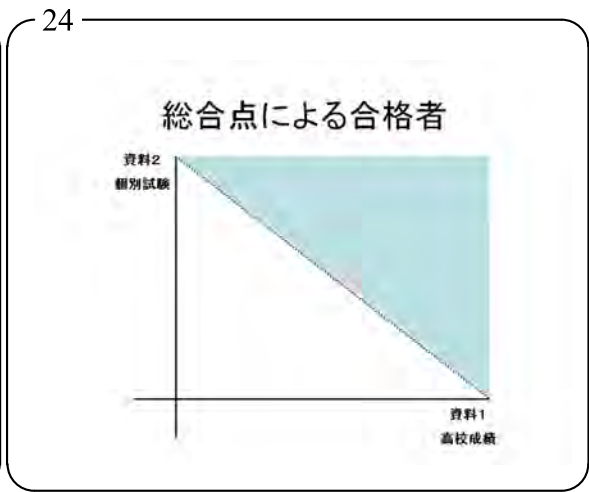
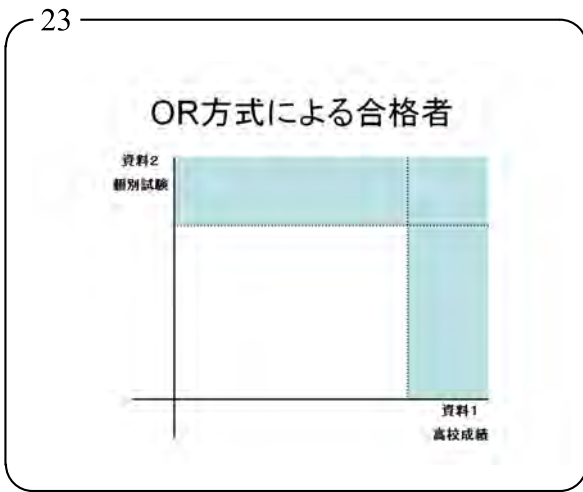
合否決定方法の例

- 総合点方式(例:傾斜配点)
- OR方式(並列的) どれか一つ(あるいはふたつ)良いものがあればよい。
- AND方式(直列的) すべてに最小の基準を満たしている。

22

AND方式による合格者





25

必要条件4 妥当性の確保と検証

- 入試選抜システムが有効に機能したかどうかをデータによって検証する(追跡調査など)。
- 妥当性検証のためには、妥当性を評価する基準が必要である。たとえば、初年時の成績(gpa、獲得された学力の具体的指標)、出席率(大学への適応度の指標)。
- 検証の結果を大学教員やアドミッションズオフィスにフィードバックし、選抜システムを改善する。

26

必要条件まとめ(ここまで)

1. アドミッションポリシーの具体化
2. 選抜資料の選択
3. 合否決定方法の確定
4. 妥当性の検証

Remark! 意思決定の枠組みから言えば、入試とは、選抜資料データから多属性の各受験者の望ましさ(効用という)を評価し、もっともよく合格者と不合格者の望ましさの評価値の差をひろげる資料や、決定方式を探す問題である。この決定システムがうまくいっているかどうかは不断に検証しなければならない
[参考資料参照]。

27

トピック1 学力偏差値の功罪

- 長所: 客観的であり、受験生にとって、あるいは、高校教師にとって、目標が明快。大学で学習に直結。進路指導が容易。
- 短所: 偏差値の高い大学を除き、学生のやる気を阻害する。
- 偏差値の高いところは、学力を重視する場合にも、学力のどの側面を見るか、たとえば、知識を問うか、解答プロセスを見るかなどの点で多様化する。
- 偏差値の低いところは、アウトカム(学士力)を高めるためにどのように入試を工夫するか、どうすれば、受験生にアピールできるかという点で多様化する。

28

提案: 多様な選抜システム

- Highly Selectiveな大学では、多様化する余裕がある。その余裕を生かして、偏差値的1次元性をこわす。
- Less Selectiveな大学では、たとえば、将来の可能性に重きを置く選抜をすることによって、受験する誘因となり、その大学の教育力を高める可能性がある。そのことによって、偏差値的1次元性を克服する。

1. 大学入試へのベイズ的アプローチ

29

トピック2 自由競争で十分か？

- “ある大学では、公表される偏差値を下げず、受験生をふやすためにAO入試を利用しています。そのため、入学者の一部が、たとえば、英語の基礎力がないなど、大学での教育に支障が出ています。”
- Q5:ある地方の、工学系大学は数学と理科のみ受験科目になっています。学内ではコミュニケーション力のために、英語を受験科目にするべきだという意見がありますが、そうすると、同じ地方の偏差値の高い総合大学へ流れてしまうという理由で英語はいつも受験科目に入りません。

30

入試のメッセージ性

- 大学入試は、各大学が将来の学生に対し、準備すべきことは何かということ具体的に、かつ、ある種の強制力を持って発するメッセージである。(例:センター試験にリスニングが導入されたことによる高校の英語教育の変化)
- 大学教員が入試に託したいメッセージは、しばしば、受験生の数を増やすという経営的目的と矛盾する。
- なんらかの公的なガイドラインが必要ではないか。まったくの自由競争ではなく、いわば、良い制約の中での競争が望ましい。

31

必要条件5 (desiderata)

大学入試のメッセージ性

- 大学入試は、学生にいかにかに学ぶかを直接的に訴えるメッセージ性を持つ。
- それぞれの大学が近視眼的に、それぞれの大学の効用を追求するとき、日本全体の入試システムとして不都合をきたす可能性がある。
- ほとんどの受験生を合格させるケースであっても、受験のために勉強させることは大学における学習のために良い効果をもたらす。
- 大学入試が大学教育に対する準備性を高めるため役立てるには、全国的なガイドラインが必要である。⇒大学評価に期待する。

32

大学(認証)評価

- 大学評価・学位授与機構、大学基準協会、日本高等教育機構のいずれかに、7年(以内)ごとに評価を受ける。
- いずれも、学生の受け入れに関する基準がある。

33

大学評価・学位授与機構の基準

(基準4)

- 教育の目的に沿って、求める学生像及び入学者選抜の基本方針などの入学者受入方針(アドミッション・ポリシー)が明確に定められ、公表、周知されていること。
- 入学者受入方針(アドミッション・ポリシー)に沿って適切な学生を受入が実施され、機能していること。
- 実入学者数が、入学定員と比較して適正な数となっていること。

34

まとめと提言

- 入試を意思決定システムとして、必要な五つの条件をまとめた。
- それぞれにおいて、具体的に決定すべき事柄を述べ、改善のための提案をした。
- それぞれの決定事項を最適化するための準備をし、入試システムが機能しているかどうかを調べるためにはアドミッションズオフィスが充実されなければならないと考える。(AO入試を本来のAdmissions Officeが行うべき入試とする。)

35

アドミッションズオフィスに期待すること

- アドミッションポリシーの決定の手助け(情報提供、受験生や関係教員へのアンケート調査など)
- 選抜資料の管理
- 合否決定のシステム化
- 広報活動(案内冊子、キャッチコピー)
- 追跡調査の設計とフィードバック

36

シンポジウム参考資料

統計的決定問題としての定式化

37

統計的決定問題としての入試

- 大学におけるパフォーマンス(基準 y)を選抜資料(x)によって予測し、その予測にもとづき、各受験者の望ましさ(以下、効用という)を評価し、合格か不合格かを定めること。
- 学生の望ましさ、構成概念(知能ややる気など)で説明されている場合には、その構成概念の指標(潜在変数、 ξ)を推定し、決定はこの推定された分布に基づく。

38

記号

$y = (y_1, y_2, \dots, y_p)$, 基準 例: GPA, 出席率
 $x = (x_1, x_2, \dots, x_p)$, 選抜資料 例: 大学入学試験, 面接評価, 内申書
 a_1 : 代替案1(合格という決定)
 a_2 : 代替案2(不合格という決定)
 $u(a_1, y) = u(y)$: 基準 y を持つ受験者を合格とした場合の効用
 $u(a_2, y) = u(y)$: 基準 y を持つ受験者を不合格とした場合の効用

39

受験者 i の期待効用

基準 y
 は、受験時には未知数である。
 故に、個々の受験者の効用は推測するしか術がない。
 個々の受験者の効用は期待効用である。

$$\begin{aligned} \bar{u}_1(i) &= \int u_1(y_i) p(y_i) dy_i, \\ \bar{u}_2(i) &= \int u_2(y_i) p(y_i) dy_i, \\ \bar{u}_1(i | x_i, Data) &= \int u_1(y_i) p(y_i | x_i, Data) dy_i, \\ \bar{u}_2(i | x_i, Data) &= \int u_2(y_i) p(y_i | x_i, Data) dy_i. \end{aligned}$$

40

予測分布

追跡調査などであるデータがあれば、必要な予測分布を導くことができる。
 そのためには、分布仮定が必要である。
 例えば、ひとつの可能なモデルは多変量正規分布であり、 x の観測値を
 所与とした場合の分布として、たとえば、多変量回帰モデルを採用する。

$$\begin{aligned} p(y) &= \int p(y | \theta_y) p(\theta_y) d\theta_y, \\ \text{e.g. } y &\sim N_q(\mu, \Sigma), \theta_y = (\mu', \text{vec}(\Sigma)')', \\ p(y | x, Data) &= \int p(y | x, \theta_{yx}) p(\theta_{yx} | Data) d\theta_y, \\ \text{e.g. } \mu &= B'x, \theta_{yx} = (\text{vec}(B)', \text{vec}(\Sigma)')' \end{aligned}$$

1. 大学入試へのベイズ的アプローチ

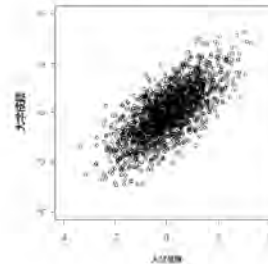
41

予測分布に関するコメント

1. 受験者の望ましさが、適性ややる気などの構成概念、すなわち、潜在変数(ξ) γ 表現される場合には、効用は、 ξ の関数となり、 ξ の事後分布をデータから求める必要が生じる。
2. 統計モデルの要であるところのパラメータの事後分布を求める際のDataには大学入試の場合に際立った特徴がある(次の図参照)すなわち、入試選抜資料と基準の両者のデータのサンプリングが変則的である。基準のデータは合格者のみに対して得られる。パラメータは、この変則的サンプリングに影響されにくい回帰係数パラメータと条件付共分散行列を用い、ベイズ的な推論方法によって元の母集団における予測分布を導出する。

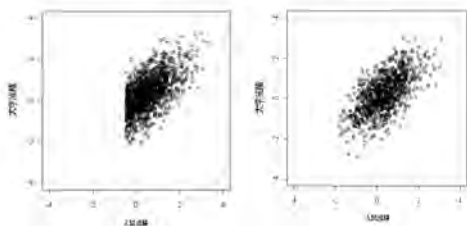
42

受験者集団



43

合格者集団



明確に上位が合格する場合

確率的に入試上位者が合格する場合

44

合格者の決定方式のいろいろ

- 単純加算
- 合成得点
- AND方式(MaxMin方式、最低点の最大化)
- OR方式(MaxMax方式、最高点の最大化)
- 制約条件の元での最大化

たとえば、 (γ, y) が大きい者を合格とする。
一般的には、合格者は、 $y \in Y_a$ として定義される。

45

合否の決定方法

n 人のうち、 n_0 人合格させる。
 $P(\gamma' y \geq c)$ が高い受験者から、 n_0 人を合格とする。
(いくつかの単純化のための仮定をおけば、この決定方式は、 $\gamma' B' x$ すなわち、選抜資料の合成得点の高い順に合格とする決定となる。
一般的には $P(y \in Y_a)$ の確率が高い受験者から n_0 人を合格とすることになる。

46

選抜資料の価値

n 人の受験者に対して、
受験者が合格としたとき、 $k(i) = 1$ 、不合格としたとき、 $k(i) = 2$ とする。
同様に、テスト情報元にも、過去のデータによってシステム構築後に、
受験者が合格としたとき、 $k'(i) = 1$ 、不合格としたとき、 $k'(i) = 2$ とする。
このとき、受験者集団全体に対するテスト情報の価値は、
$$I = \sum_{i=1}^n \int u_{\gamma(i)}(y_i) P(y_i | x_i, Data) dy_i - \sum_{i=1}^n \int u_{\gamma(0)}(y_i) P(y_i) dy_i$$

$$= \sum_{i=1}^n \tilde{u}_{\gamma(i)}(y_i | x_i, Data) - \sum_{i=1}^n \tilde{u}_{\gamma(0)}(y_i)$$

47

選抜資料の情報量＝ 統計的情報価値

- このようにテストがもたらす利益は、テストを得た後に行う決定によってもたらされる期待効用と、テスト情報なしで行う決定によってもたらされる期待効用の差である。これをテストの情報価値と呼ぶことにする。
- 妥当性のあるテストはこの情報価値の大きいものである。
- 妥当性のあるテストの組は、情報価値を最適化することによって得られる。

48

選抜資料の妥当性指標

特定の目的における、特定の受験者集団に対する妥当性指標は、次のように定義できる。
この妥当性指標は、あくまで特定の状況に適用できる性質のものであるが、妥当性とは本来そのようなものであろう。

$$Index = \frac{\sum_{i=1}^n \bar{u}_{i(0)}(y_i | x_i, Data) - \sum_{i=1}^n \bar{u}_{i(0)}(y_i)}{\sum_{i=1}^n \bar{u}_{i(0)}(y_i | x_i, Data)}$$

49

まとめ

- 入試選抜資料の妥当性は、その資料によってもたらされる期待効用によって測られる。
- 良い選抜資料のセットを見出す作業は、統計的には変数選択の問題であり、より具体的には、妥当性を基準として、その基準を最適化する変数の組を見出す統計的推論の問題として定式化できる。
- このような考え方から入試研究の改善のための知見を得ることができる。

第2章 セクターと難易度による分類の下 での近年の大学入学者選抜の分析

小谷野 仁, 繁樹 算男

本章では, 1991 年度から 2011 年度までの 21 年間の全国大学の入学者選抜方式の推移を把握する目的で, 1991, 1996, 2001, 2006, 及び 2011 年度についての全数調査からのデータに基づき, 公表された定員をベースに, センター試験と個別学力試験の科目数の推移, 小論文と面接の利用状況の推移, 及び非学力型選抜の動向を, 大学を国立, 公立, 及び私立に分け, またこれらを難易度によって更にいくつかのグループに分けた上でまとめ, それぞれの特徴を示した.

2.1 はじめに

以下では, 1991 年度から 2011 年度までの 21 年間に全国大学の入学者選抜方式がどのように推移して来たのか, また現状はどのようなのかを定量的に調べる. 本調査は, 大学入学者選抜における一連の手続きを意思決定システムと見なすという視点から, 入学者選抜で必要とされることを明らかにし, 入試をより良くするために望まれることをスタンダードとしてまとめることを目指した, 大学入試センター入学者選抜研究機構における入試スタンダード・プロジェクトの基礎研究として行われた. その基礎研究は, 1991 年度から 2011 年度までの 21 年間を 5 年間隔に分け, 1991, 1996, 2001, 2006, 及び 2011 年度について全数調査を行って, (i) センター試験の科目数の推移, (ii) 個別学力試験の科目数の推移, (iii) 小論文と面接の利用状況の推移, 及び (iv) 非学力型選抜の動向を, 大学を (a) 国立大学法人

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

(以下国立), 公立大学法人 (以下公立), 及び私立に分けて, また (b) 選抜の性質で更にいくつかのグループに分けて定量的に明らかにすることを目的としている。

大学の入学者選抜方式の歴史的な推移や変容を主題とした研究は, これまで主に教育社会学の領域で様々な視点からなされてきた。経時的なデータに基づいた近年の実証的な研究としては, 例えば, 推薦入学の拡大に注目して, マス選抜の部分で選抜方式がより多様化している傾向があることを実証した中村 [43], 国公立大学を中心に, 大学または学部を単位として, 入学者選抜方式のいくつかの側面についてのデータをまとめた野家 [45], 1994年度から 2005 年度までの総合試験と適性試験の利用状況を分析した伊藤 [28] などがある。著者等の知る限り, 定員をベースとしたデータに基づく研究は少ないが, 私立大学の推薦入学の割合の推移を考察した西井 [44] がある。1 時点のデータに基づくものでは, 例えば, 滝 [40] が多様化を軸に 1999 年度の入試を簡潔に要約している。より最近のものでは, 国公立大学の一般入試, 推薦入試, 及び AO 入試の全般に渡って多様化の実態を分析した山村 [49] が挙げられる。

これらの研究と比較して, 本調査の特徴は, 国公立大学の全てに対して, 大学が公表した入学者選抜で選抜方式と定員が明確なものについて, 原則として全数調査を行って得たデータを用い, 大学や学部を単位とはせずに, 定員をベースとして分析したこと, より数量的な接近法をとって経時的な推移を調べたことである。但し, 本研究は公表された定員に基づく調査であり, 実際の入学者数とは異なることに注意する必要がある。

表 2.1. 選抜方式のデータの整理の形式

学部	選抜方式名	定員	割合	セ科目	個別科目	総合問題	小論文	面接	書類	その他
法学部	前期	100	0.333	0	3	0	0	0	0	0
法学部	後期	70	0.023	0	3	0	0	0	0	0
法学部	セ利用	80	0.267	3	0	0	0	0	0	0
法学部	推薦	50	0.167	0	0	0	1	1	1	0
経済学部	I期	120	0.480	0	3	0	0	0	0	0
経済学部	II期	50	0.200	0	2	0	0	0	0	時事教養試験
経済学部	III期	30	0.120	0	2	0	1	0	0	0
経済学部	セ利用	30	0.120	3	0	0	0	0	0	0
経済学部	AO	20	0.080	0	0	0	0	1	1	応用力試験
理工学部	前期	100	0.500	0	3	0	0	0	0	0
理工学部	後期	50	0.250	0	3	0	0	0	0	0
理工学部	推薦	30	0.150	0	0	0	0	1	1	基礎学力調査
理工学部	AO	20	0.100	0	0	0	0	1	1	課題レポート

2.2 調査の概要

全国大学の入学者選抜方式の推移と現状を把握するために、晶文社出版編集部 [36–39] と晶文社学校案内編集部 [35] から、次の (i) から (ix) のデータを収集した: (i) 選抜方式名, (ii) 定員, (iii) センター試験の科目数, (iv) 個別学力試験の科目数, (v) 総合問題の有無, (vi) 小論文の有無, (vii) 面接の有無, (viii) 書類の有無, 及び (ix) その他. 前節で述べたように, 対象は, 1991 年度から 2011 年度のうち, 1991, 1996, 2001, 2006, 及び 2011 年度である. データの様式は, 表 2.1 に模式的に示されている. 表の中の「セ科目」と「個別科目」は, それぞれセンター試験と個別学力試験の科目数を表している. また, 総合問題, 小論文, 面接, 及び書類の欄の「1」はそれらが課されることを, 「0」は課されないことを示している.

2.3 国立, 公立, 及び私立大学ごとの推移

本節では, 大学を国立, 公立, 及び私立に分けた上で, (1) センター試験の科目数が 0 科目, …, 4 科目と 5 科目以上によって選抜される定員の全定員の中の割合, (2) 個別学力試験の科目数が 0 科目, …, 4 科目と 5 科目以上によって選抜される定員の全定員の中の割合, 及び (3) 小論文または面接が課される選抜方式の定員と, センター試験と個別学力試験が共に 0 科目である選抜方式の定員の全定員の中の割合を計算し, 最近 21 年間のそれらの推移を調べる.

2.3.1 センター試験

国立大学. 国立大学のセンター試験の科目数の推移が, 図 2.1 に示されている. まず, センター試験が 5 科目以上課される方式によって選抜される定員の国立大学の全定員に占める割合は, 1991 年度が 85.51% で最も高く, 1996 年度に 70.44% と大幅に減少した後, 2006 年度まで 70% 前後の水準で推移し, 2011 年度に 83.20% まで回復した. 一方で, センター試験が課されない選抜が, 1991 年度から 2011 年度まで, わずかずつではあるが一貫して上

昇している。1991年度には3.08%であったが、2001年度に10.45%まで大幅に上昇し、2011年度には13.01%にまでなっている。残りの科目数のうち、1科目と2科目はほとんどの年度において1%以下の水準で、3科目は5%前後の水準で安定的に推移しているが、4科目は1991年度の7.33%から1996年度の13.15%まで上昇した後、低下し続け、2011年度には1.41%にまでなっている。

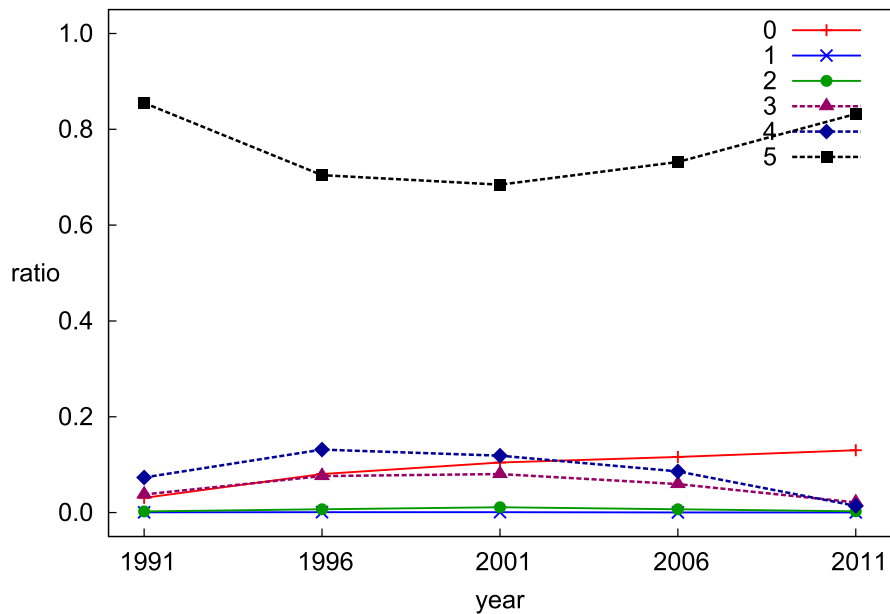


図 2.1. センター試験の科目数の推移 (国立大学)

公立大学. 図 2.2 を見ると、1991年度から2011年度までの公立大学のセンター試験の科目数の推移は、国立大学と似た特徴を示していることが分かる。5科目以上課される選抜は、1991年度が55.12%で最も高く、1996年度に28.08%と大幅に減少した後、2006年度まで30%前後の水準で推移し、2011年度に49.12%まで回復した。また、1991年度から2011年度まで、センター試験が課されない選抜が、わずかずつではあるが一貫して上昇している。1991年度には9.64%であったが、2001年度に20.06%まで大幅に上昇し、2011年度には21.29%にまでなっている。残りの科目数のうち、1科目と2科目はほとんどの年度において5%以下の水準で、4科目は15%前後の水準で安定的に推移しているが、3科目

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

は 1991 年度から 2001 年度までの 10 年間に 20.26% から 28.01% に上昇して、2011 年度に再び 14.99% まで低下するという、比較的大幅な変動を示している。

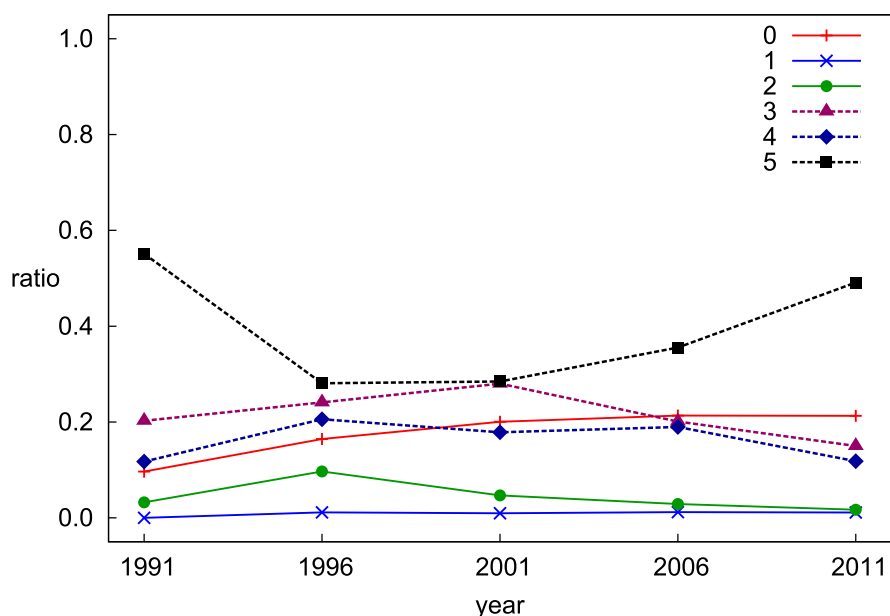


図 2.2. センター試験の科目数の推移 (公立大学)

私立大学. 図 2.3 は、私立大学のセンター試験の科目数の推移を示している。私立大学の定員の総数のうち、センター試験が課されない方式で選抜される定員の割合は、1991 年度には 99.88% であったが、2001 年度には 94.85% に、更に 2011 年度には 88.03% にまで減少している。2001 年度においては、私立大学の定員の総数のうち、2 科目が課される選抜の定員が 1.69%、3 科目が 2.64% であって、5 科目以上は 0.13% であるが、10 年後の 2011 年度においては、2 科目が 3.40%、3 科目が 4.65%、5 科目以上は 1.39% となっており、微増傾向にある。

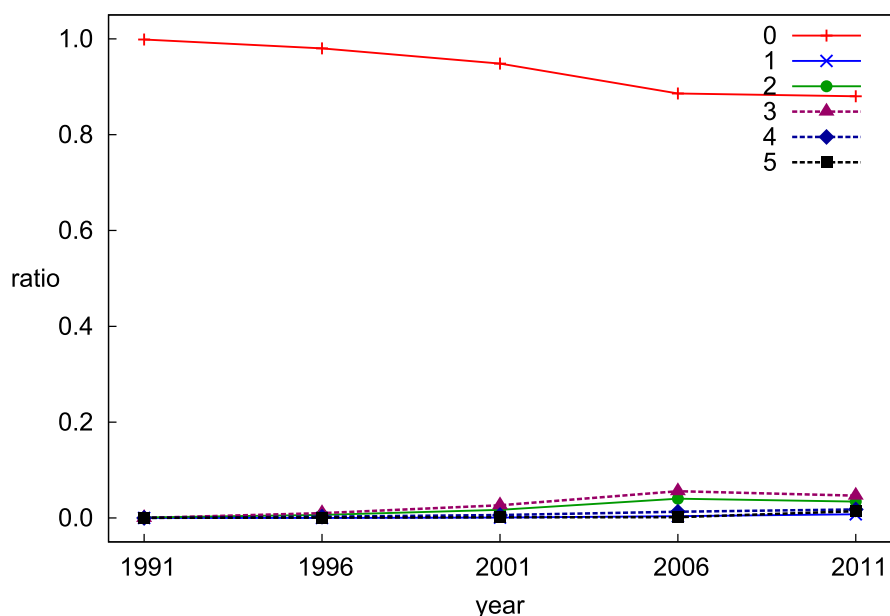


図 2.3. センター試験の科目数の推移 (私立大学)

2.3.2 個別学力試験

国立大学. 国立大学の個別学力試験の科目数の推移が、図 2.4 に示されている。まず、個別学力試験が 0 科目である選抜の定員の国立大学の全定員に占める割合は、1991 年度には 16.96% であったが、1996 年度の 28.55% を経て、2001 年度には 34.20% まで大幅に上昇し、その後 2006 年度までほぼ同じ水準で推移した。しかし、2011 年度に 14.14% まで大幅に低下して、1991 年度の水準も下回っている。次に、1 科目と 2 科目は、1991 年度から 2006 年度まで緩やかに減少した後、横ばいあるいは増加の傾向に転じている。3 科目は、1991 年度から 2006 年度までは 20% 前後の水準で推移していたが、2011 年度に 13.24% となり、目立つ減少を示した。4 科目と 5 科目以上の選抜の割合は、特に 2006 年度以降微増傾向にある。4 科目の選抜は、1991 年度から 2006 年度までの 16 年間は 5% 前後の水準で安定的に推移して来たが、2011 年度に 14.32% と大きく上昇した。また、5 科目以上が課される選抜は、2011 年度には全体の約 7% となっている。図には明示されていないが、個別に調べてみ

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

ると、医学部が多いことが分かる。

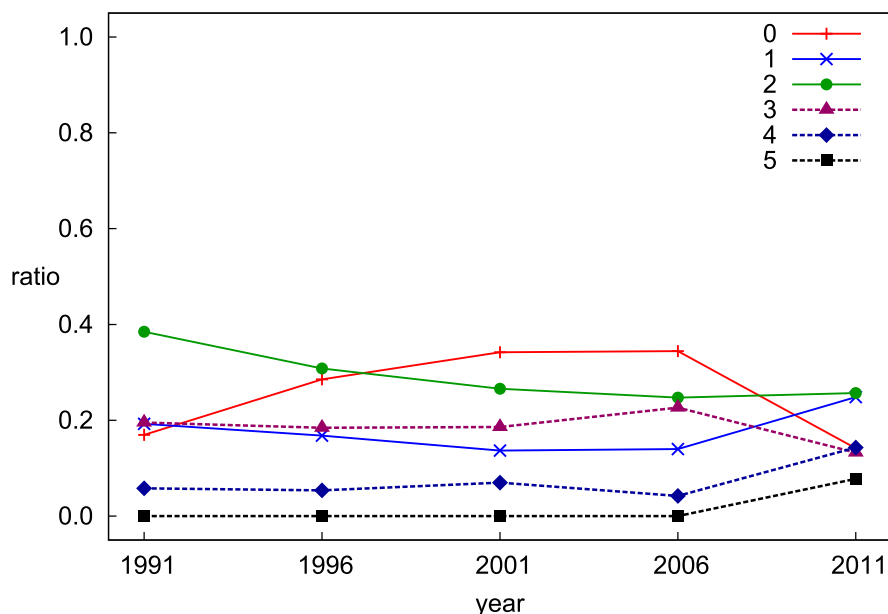


図 2.4. 個別学力試験の科目数の推移 (国立大学)

公立大学. 図 2.5 より、センター試験の科目数と同様に、公立大学の個別学力試験の科目数の推移も、国立大学と似た特徴を示していること読み取れる。個別学力試験が 0 科目である選抜は、1991 年度には 20.39% であったが、1996 年度の 32.06%、2001 年度の 48.35% を経て、2006 年度には 53.45% の水準にまで上昇した。しかし、2011 年度に 21.68% まで大幅に低下し、1991 年度に近い水準に戻っている。1 科目と 2 科目は、1991 年度から 2006 年度までは緩やかに低下する挙動を示していたが、2006 年度から 2011 年度までの 5 年間に、大きく上昇し、これらが 0 科目の大幅な低下分を吸収している。また、4 科目と 5 科目以上の選抜も、国立大学と同様、特に 2006 年度以降微増傾向にあり、2011 年度において 4 科目は 6.73%、5 科目以上は 3.46% になっている。5 科目以上を課しているのは、やはり医学系の学部が主である。

私立大学. 最後に、私立大学の個別学力試験の科目数の推移を示した図 2.6 を見る。1991 年度には、3 科目の選抜の定員が私立大学の定員の総数の 69.40% を占めており、圧倒的で

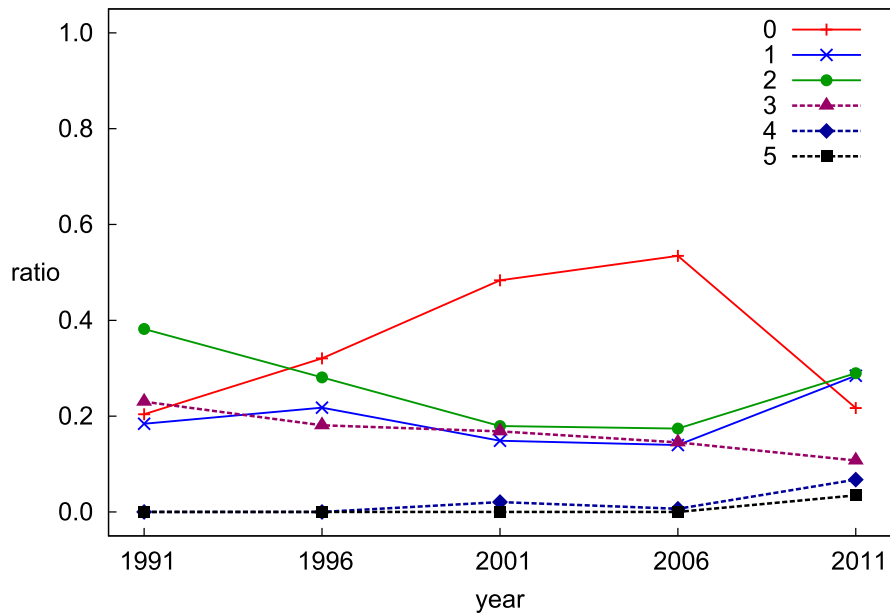


図 2.5. 個別学力試験の科目数の推移 (公立大学)

あったが、1996 年度に 52.44%、2001 年度に 45.44%、更に 2006 年度には 36.64% と大幅に低下し続けた。その後、2011 年度にやや持ち直して 39.94% になっている。1, 4, 及び 5 科目は、1991 年度から 2011 年度まで低い水準で安定的に推移している。また、2 科目は、1991 年度の 17.85% から 1996 年度に 28.02% まで大きく上昇し、2001 年度までこれとほぼ同じ水準で推移したが、その後低下して、2011 年度には 1991 年度とほぼ同じ水準の 8.18% に戻っている。最後に、0 科目は、1991 年度の 10.64% から上昇し続けて、2006 年度には 38.21% にまでなった。特に 2001 年度の段階で 23.85% であり、ここから 2006 年度までの 5 年間におよそ 15% も上昇している。その後やや回復の兆しを見せており、2011 年度には 34.10% となっている。従って、大雑把には、この 21 年間、私立大学の個別学力試験の科目数は、0 科目と 3 科目がお互いの変化を吸収する形で推移して来たことになる。

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

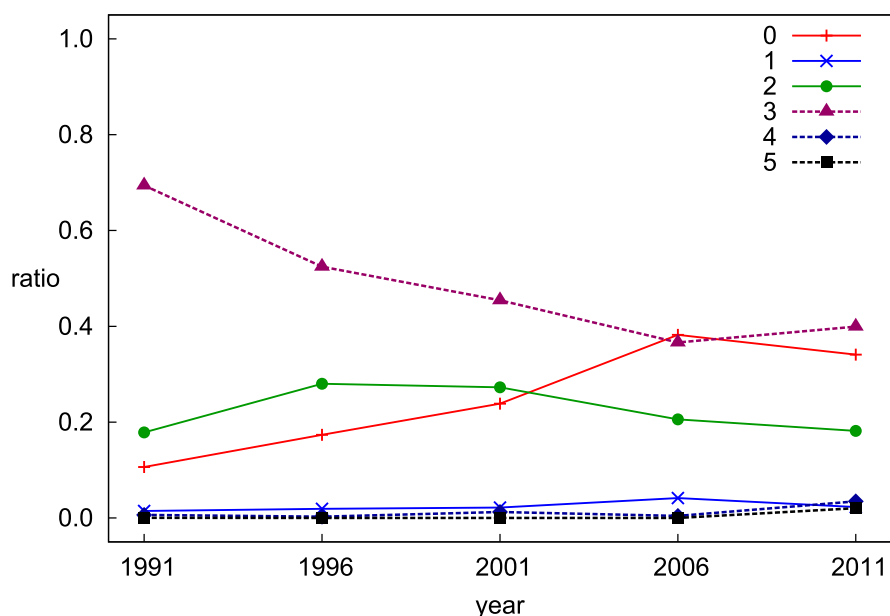


図 2.6. 個別学力試験の科目数の推移 (私立大学)

2.3.3 非学力型選抜

国立大学. 図 2.7 は、国立大学の小論文または面接が課される選抜と、学力試験が課されない選抜の動向を示している。国立大学の定員の総数のうち、小論文が課される選抜の定員の割合は、1991 年度から 2006 年度までの 16 年間、20% 弱の水準でほぼ一定であった。例えば、1991 年度は 19.94%、2006 年度は 18.15% である。しかし、2011 年度に 7.28% まで急激に低下している。次に、面接が課される選抜は、1991 年度の 9.25% から 2006 年度の 21.99% まで 16 年間で 2 倍以上の水準に上昇した。しかし、2011 年度に 11.75% まで大きく低下し、5 年間で 1991 年度に近い水準に戻っている。最後に、学力試験が課されない方式で選抜される定員の割合は、1991 年度から 2011 年度までの 21 年間一貫して上昇している。1991 年度には 3.07% に過ぎなかったが、2001 年度に 10.45% まで大きく上昇し、2011 年度には 12.95% になっている。5 年毎に上昇分が少なくなっているものの、この変数がこのような振る舞いを示したのは国立大学だけであった。

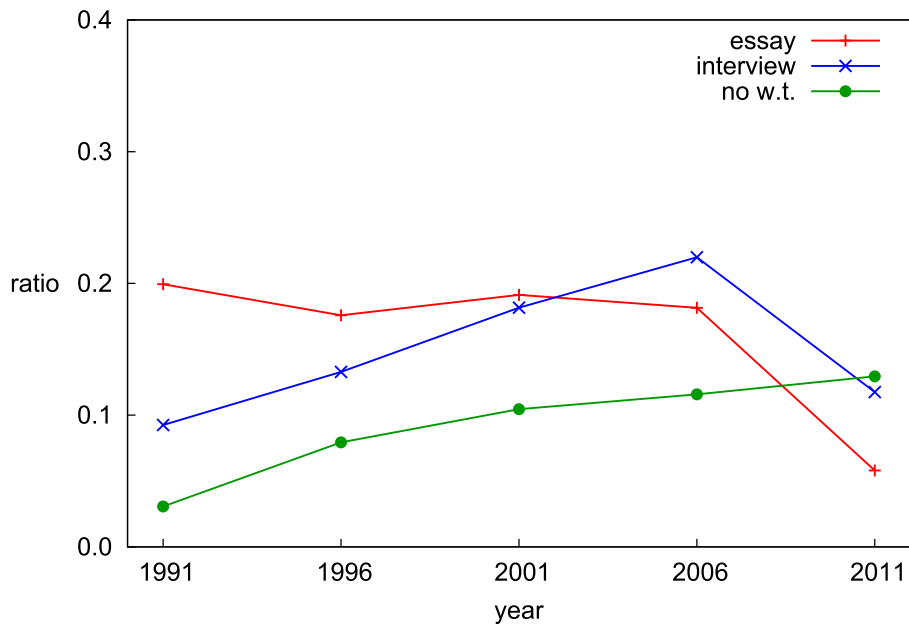


図 2.7. 非学力型選抜の動向 (国立大学)

公立大学. 次に, 公立大学についての結果が示された図 2.8 を見る. 小論文が課される選抜の定員の割合は, 1991 年度の 19.90% から 2006 年度の 38.37% まで, 16 年間で 2 倍近い水準にまで上昇した. その後, 2011 年度に 1991 年度の水準をはるかに下回る 13.85% まで急激に低下している. 公立大学の変数は, 国立大学と似た振る舞いをするものが多いが, 小論文は全く異なる振る舞いを示したもののうちの 1 つである. 次に, 面接が課される選抜の定員の割合は, 1991 年度から 1996 年度まで約 13% で一定であったが, その後 2006 年度までの 11 年間で 30.85% と 2 倍以上の水準に上昇した. しかし, 2011 年度に 16.02% まで大きく低下し, 1991 年度から 1996 年度に近い水準になっている. 学力試験が課されない選抜の定員の割合は, 1991 年度の 7.28% から 1996 年度の 12.64%, 2001 年度の 16.98% を経て, 2006 年度の 20.88% まで上昇し続けたが, 2011 年度に 19.75% となり, 下降に転じている.

私立大学. 私立大学についての結果が, 図 2.9 に示されている. まず, 小論文は, 国公立大学と異なり, この 21 年間に大きな上昇や低下を示すことなく, 10% 前後の水準で比較的安

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

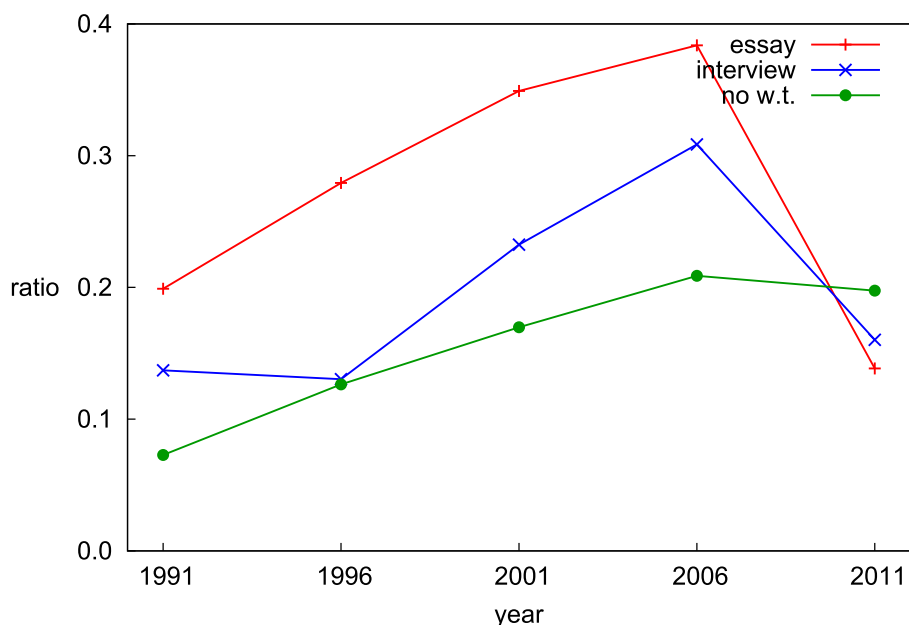


図 2.8. 非学力型選抜の動向 (公立大学)

定的に推移して来た。1991 年度の 10.54% から 2001 年度の 11.47% を経て、2011 年度は 7.28% である。面接は、やや振り幅が大きいものの、小論文と似た変化を示しており、20% 前後の水準で推移している。最も割合が低かったのは 1996 年度の 14.39% であり、高かったのは 2006 年度の 23.16% である。学力試験が課されない選抜の定員の割合は、1991 年度の 10.54% から 1996 年度の 15.40%、2001 年度の 18.72% を経て、2006 年度の 26.98% まで 16 年間上昇し続けたが、2011 年度に下降に転じて、23.28% になっている。

2.4 難易度によってグループ分けした場合の推移

前節では、大学を国立、公立、及び私立に分類したが、本節では、難易度によって、これらを更にいくつかのグループに分けて、各グループの選抜方式の推移を見る。具体的には、(1) センター試験が 0 科目、1 または 2 科目、3 科目、4 科目以上によって選抜される定員の全定員の中の割合、(2) 個別学力試験が 0 科目、1 または 2 科目、3 科目、4 科目以上によって

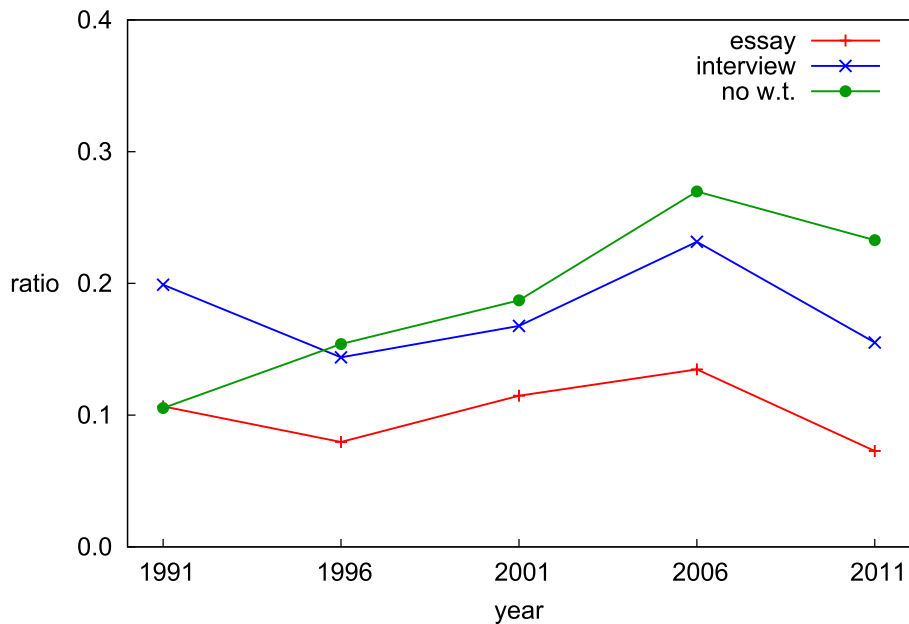


図 2.9. 非学力型選抜の動向 (私立大学)

選抜される定員の全定員の中の割合、及び (3) 小論文または面接が課される選抜方式の定員と、センター試験と個別学力試験が共に 0 科目である選抜方式の定員の全定員の中の割合を計算し、最近 21 年間のそれらの推移を調べる。

1つの大学の難易度の指標として、ここでは、その大学が実施した入学者選抜のうち、偏差値が公表されているものについて、偏差値の単純平均を用いた。全ての入学者選抜に対する偏差値が得られるならば、1つの大学の各選抜方式の定員の全定員に占める割合を重みとして平均をとるのが望ましいと考えられるが、実際に偏差値が公表されるのは一部の選抜方式に対してのみであるため、上の方法で計算した値を用いることとした。そうして、2011年度の偏差値のデータを用いて、各大学の上の意味での難易度を計算し、国立大学、公立大学、及び私立大学を難易度 65 以上、50 以上 65 未満、50 未満の 3つのグループに分けた。以下では、難易度 65 以上のグループをグループ a、50 以上 65 未満のグループをグループ b、50 未満のグループをグループ c と呼ぶことにする。ただし、公立大学については、グループ a に属する大学は存在しなかった。グループ分けは、1991、1996、2001、2006、及び

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

2011 年度の各年度について個別には行わずに、2011 年度のデータに基づくグループ分けを 1991, 1996, 2001, 及び 2006 年度の 4 年度にも適用した。従って、以下で見ているのは、2011 年度に各グループに属している大学が、最近 21 年間にどのような入学者選抜を行ってきたかである。

2.4.1 センター試験

国立大学. 図 2.12 は、国立大学の各グループにおけるセンター試験の科目数の推移を示している。グループ a においては、最近 21 年間、4 科目以上が課される選抜が圧倒的多数を占め続けており、状況にほとんど変化は見られない。その割合は、1991 年度においては 94.95%、2001 年度においては 94.16%、2011 年度においては 97.22% となっている。3 科目の選抜の割合は、最も高い 2006 年度においても 5.97% であり、2 科目以下の選抜はほとんどない。

グループ a と比較した時、グループ b においては、センター試験が課されない選抜の割合が、少しずつではあるが、一貫して増加している点が 1 つの特徴となっている。1991 年度においては 2.84% であったが、5 年後の 1996 年度に 8.18% まで大きく増加し、2011 年度には 13.63% となっている。4 科目以上が課される選抜は、1991 年度から 2001 年度まで減少傾向にあり、93.46% から 79.66% まで低下したが、2001 年度を境に増加傾向に転じ、2011 年度には 83.98% となっている。3 科目の選抜は、1991 年度の 3.35% から増加して、2001 年度には 7.95% になったが、その後減少し、2011 年度には、1991 年度の水準を下回る 2.09% になった。

グループ c においては、グループ a 及び b と比較して、科目数の変動が目につく。4 科目以上が課される選抜は、1991 年度の 85.96% から 1996 年度の 72.68% まで減少し、その後増加に転じて、2006 年度には 83.32% になったが、2011 年度には 78.87% まで再び減少している。3 科目の選抜は、1991 年度から 1996 年度まで増加して 8.02% から 14.21% になったが、その後減少して 2006 年度に 1.94% となり、2011 年度までおよそ 2% の水準が維持

2.4. 難易度によってグループ分けした場合の推移

されている。センター試験が課されない選抜の割合は、2001年度から2006年度までの間はほぼ横ばいであったが、この期間を除いて増加している。1991年度においては6.02%であったのが、2001年度には15.67%となり、2011年度には18.90%にまで上昇している。

公立大学. 次に、公立大学の各グループにおけるセンター試験の科目数の推移が示された図 2.12 を見る。グループ b において、4 科目以上が課される選抜は、1991 年度から 2001 年度まで減少して、78.48% から 49.66% になったが、その後増加に転じ、2006 年度には 57.12%、2011 年度には 63.70% になっている。3 科目が課される選抜は、1991 年度の 15.24% から増加して、2001 年度においては 27.63% と比較的高い割合を占めた。しかし、2001 年度を境に減少し始めて、2006 年度には増加し続けている 0 科目の選抜よりも少なくなり、2011 年度においては 15.25% の割合に戻っている。1 または 2 科目の選抜は、最近 21 年間非常に低い水準で推移して来た。0 科目の選抜は、1991 年度と 1996 年度の間に 6.27% から 14.86% まで大幅に増加した後も 2006 年度まで増加を続け 20.72% にまでなったが、その後、その横ばい傾向に転じ、2011 年度においては 19.69% となっている。

次にグループ b についてであるが、4 科目以上の選抜は 1991 年度の 4.96% から増加を続け、2011 年度には 40.29% の水準にまでなった。特に、1996 年度と 2001 年度の間に 10.28% から 32.86% に大きく増加している。3 科目の選抜は、1991 年度から 1996 年度までに 36.17% から 7.23% に大幅に減少したが、2001 年度に 18.94% まで回復し、その後は 10% 台後半の水準で安定的に推移している。1 または 2 科目の選抜は、1991 年度から 1996 年度までに、0 科目と 3 科目の選抜の減少分を吸収する形で、23.40% から 61.20% まで大幅に増加したが、2001 年度には 1991 年度とほぼ同じ 23.39% の水準に戻り、その後も減少して、2011 年度には 13.80% になっている。0 科目の選抜は、1991 年度から 1996 年度までに 35.46% から 21.29% まで減少したが、その後増加し続け、2011 年度には 31.41% の割合に戻っている。

私立大学. 最後に、私立大学の各グループにおけるセンター試験の科目数の推移が、図 2.12 に示されている。私立大学においては、グループの間でセンター試験の科目数の推移は非常に似ており、全てのグループにおいて 0 科目が少しずつ減少し、他の科目が少しずつ増

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

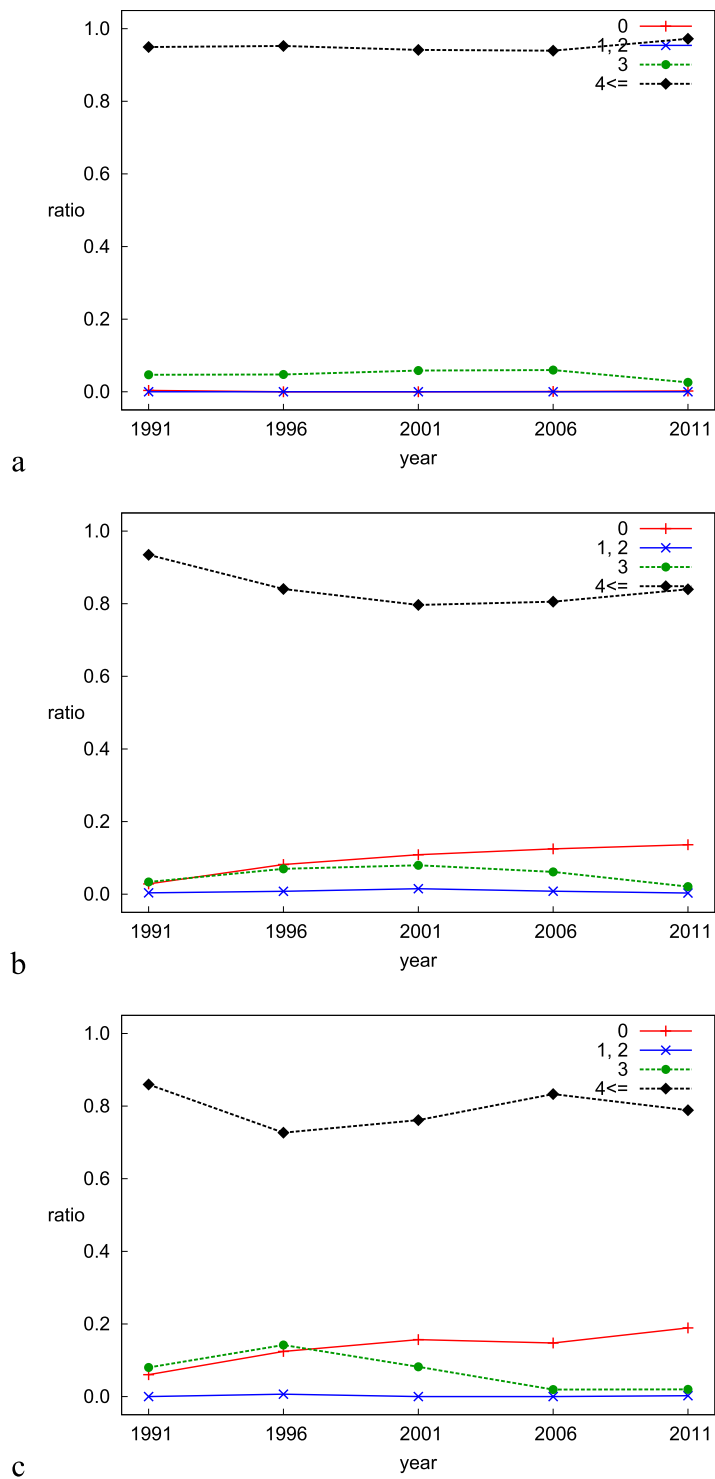
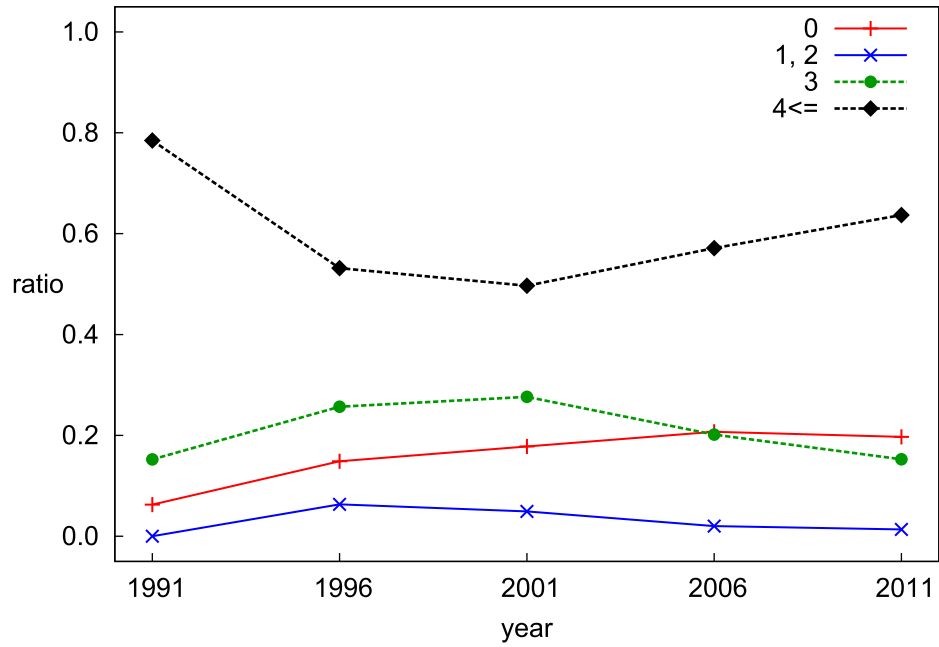
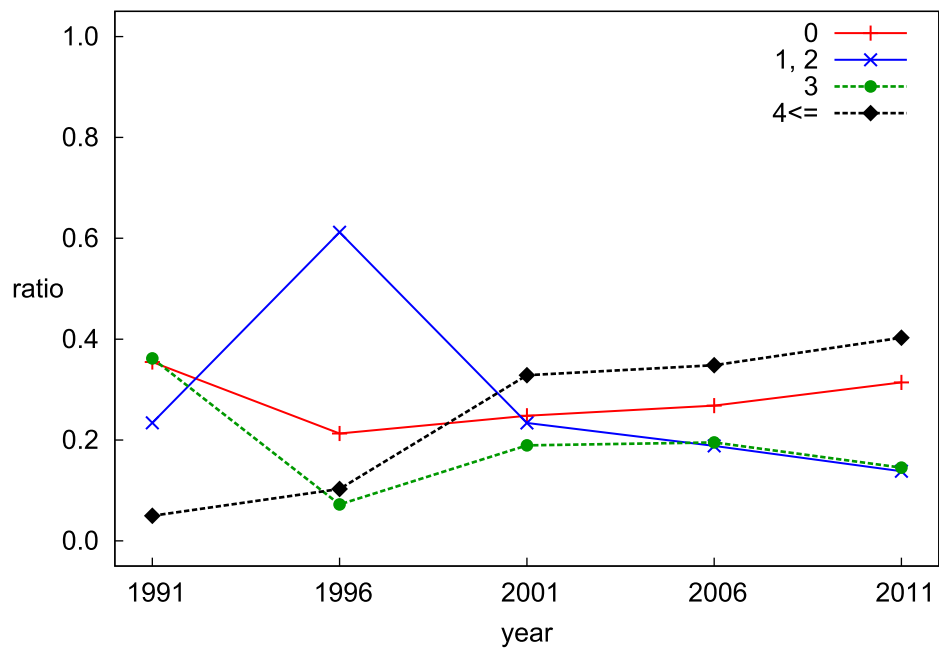


図 2.10. グループ分けした場合のセンター試験の科目数の推移 (国立大学)

2.4. 難易度によってグループ分けした場合の推移



b



c

図 2.11. グループ分けした場合のセンター試験の科目数の推移 (公立大学)

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

加する傾向にあり、私立大学全体で見た時と同様の推移を示している。

2.4.2 個別学力試験

国立大学. 図 2.13 に、国立大学の各グループの個別学力試験の科目数の推移が示されている。グループ a において、4 科目以上が課される選抜は、1991 年度から 2001 年度までおよそ 70% の水準で安定的に推移していたが、2006 年度に 52.33% まで低下し、その後 2011 年度に 87.37% まで大きく上昇している。この大幅な上昇は、3 科目の選抜の急速な低下に反映されており、その割合は、2011 年度にはおよそ 2% の水準にまで落ち込んでいる。1 または 2 科目の選抜と学力試験が課されない選抜の割合は、最近 21 年間、それぞれ 15% 前後の水準と 5% 以下の水準で安定的に推移している。

グループ b については、グループ a と比較して、4 科目以上の選抜の割合が低く、2 科目以下の選抜の割合が高いことが、1 つの特徴となっている。ただ、2006 年度を境に、学力試験が課されない選抜の割合は大きく減少し、代わって 1 または 2 科目の選抜の割合と 4 科目以上の選抜の割合が増加している。

グループ c においては、3 科目以上が課される選抜は、ほとんど行われていない。これは、グループ c の大きな特徴となっている。2 科目以下の選抜が大多数を占めているが、1991 年度から 1996 年度までの期間に、学力試験が課されない選抜の割合の大幅な増加と、1 または 2 科目が課される選抜の割合の大幅な減少が見られた。また、2006 年度から 2011 年度までの期間に、学力試験が課されない選抜の割合の大幅な減少と、1 または 2 科目が課される選抜の割合の大幅な増加が見られる。

公立大学. 図 2.14 は、公立大学の各グループの個別学力試験の科目数の推移を示している。4 科目以上の選抜は、2006 年度まで極めて低い水準で推移していたが、2006 年度と 2011 年度の間に 11.44% の水準まで大幅に上昇した。3 科目の選抜は少しずつ減少しており、1991 年度には 24.20% であったが、2001 年度には 16.53%、更にその 10 年後の 2011 年度には 11.48% となっている。1 または 2 科目の選抜は、1991 年度から 2006 年度まで一貫

2.4. 難易度によってグループ分けした場合の推移

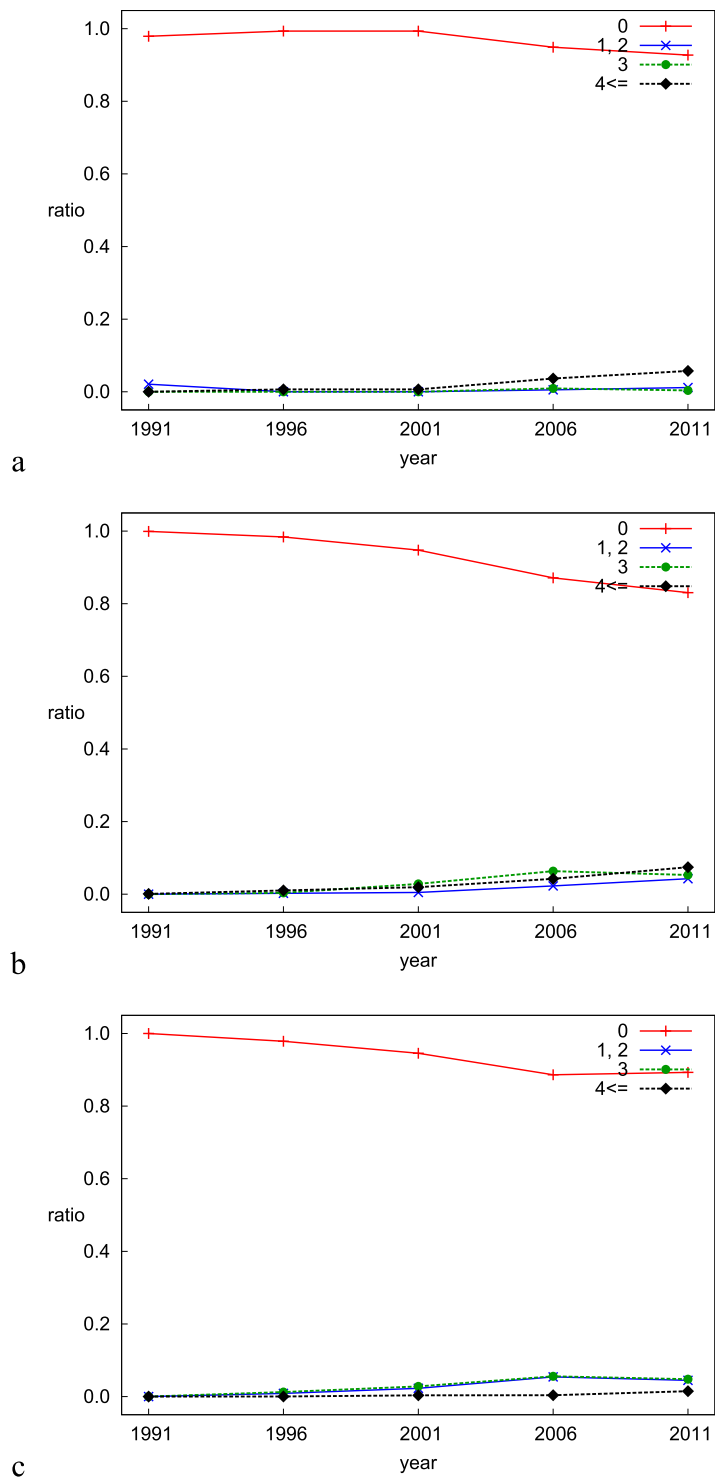


図 2.12. グループ分けした場合のセンター試験の科目数の推移 (私立大学)

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

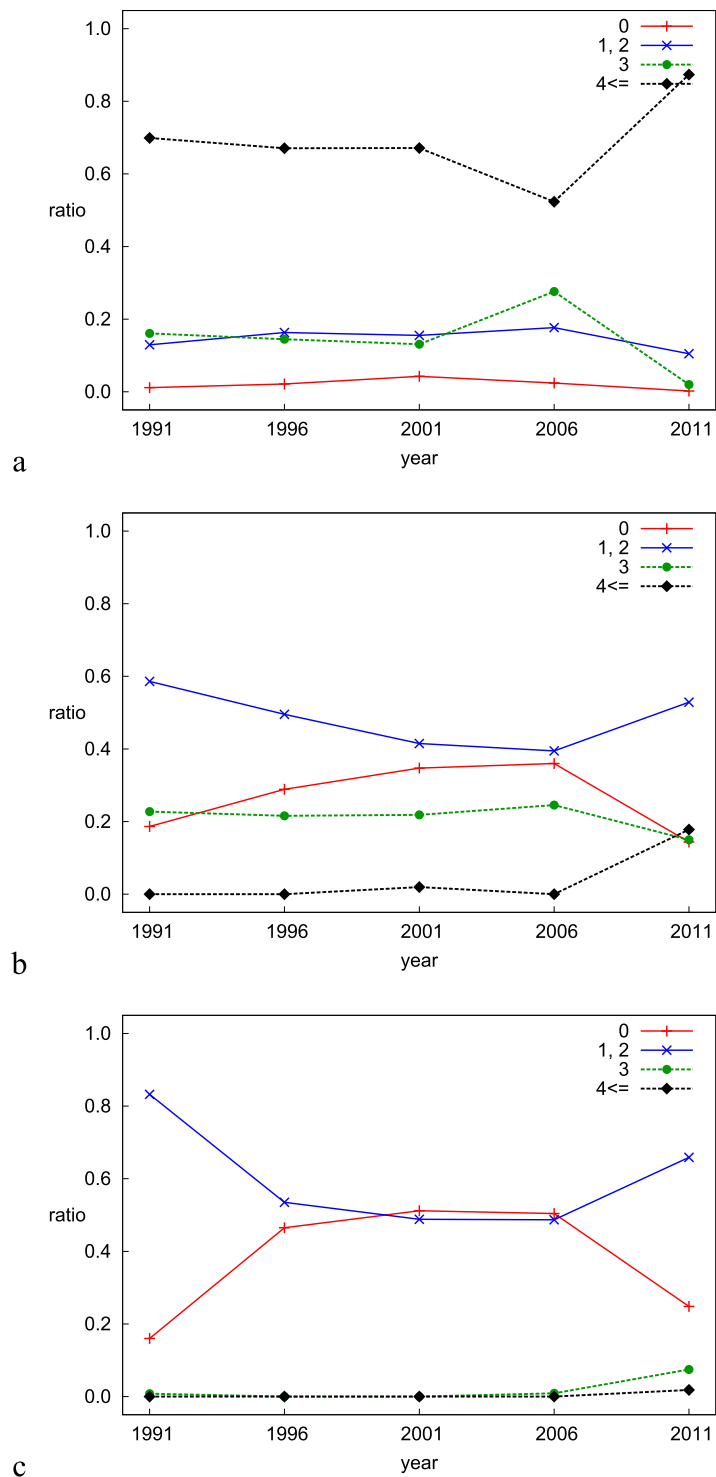


図 2.13. グループ分けした場合の個別学力試験の科目数の推移 (国立大学)

2.4. 難易度によってグループ分けした場合の推移

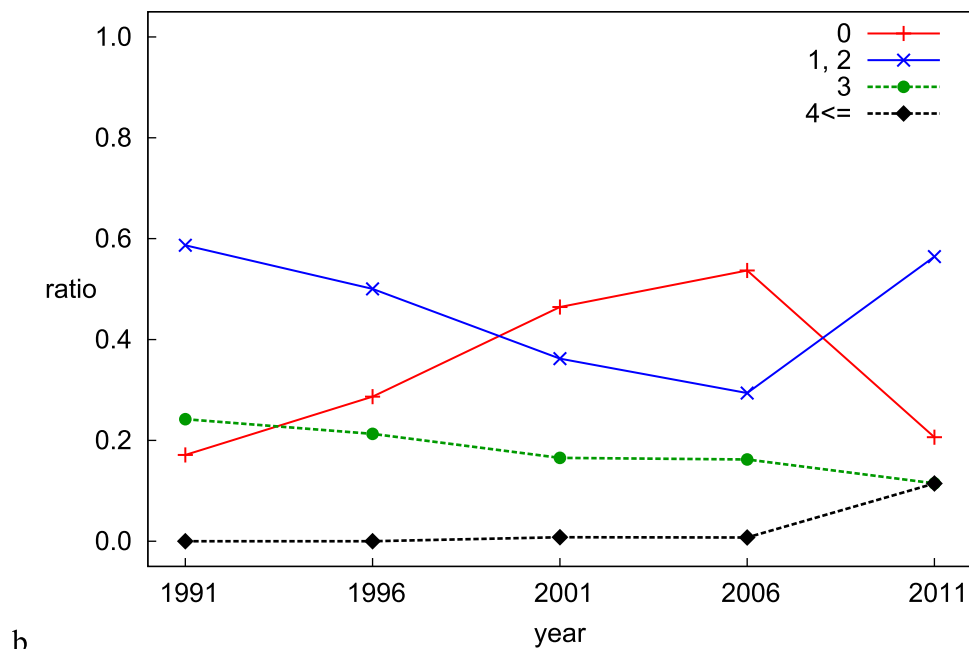
して減少し、58.68%からおよそ半分の29.38%にまでなったが、2006年度を境に増加に転じ、2011年度には56.44%の水準に戻っている。0科目の選抜は、これとほぼ反対の推移を示しており、1991年度の17.12%から2006年度の53.68%まで増加を続けた後、減少に転じ、2011年度には20.63%になっている。

このように、グループbにおいては2006年度が推移の仕方の変化点となっているが、グループcにおいては、4科目以上の選抜が最近21年間一貫して極めて低い水準にとどまっているものの、2001年度が推移の仕方の変化点となっているように見える。1または2科目の選抜は、1991年度の58.15%から2001年度の24.41%まで10年間で大幅に減少した後、急激な増加に転じ、2011年度には、1991年度の水準を上回る68.26%の水準にまでなっている。この変動は、2001年度における3科目の選抜の増加と、0科目の選抜の急速な減少に反映されている。0科目の選抜は、1991年度においても41.84%と高い水準であったが、その後の10年間増加を続け、2001年度においては60.49%の水準にまでなった。しかし、その後、減少に転じ、2011年度には、1991年度の水準を下回る29.02%となっている。

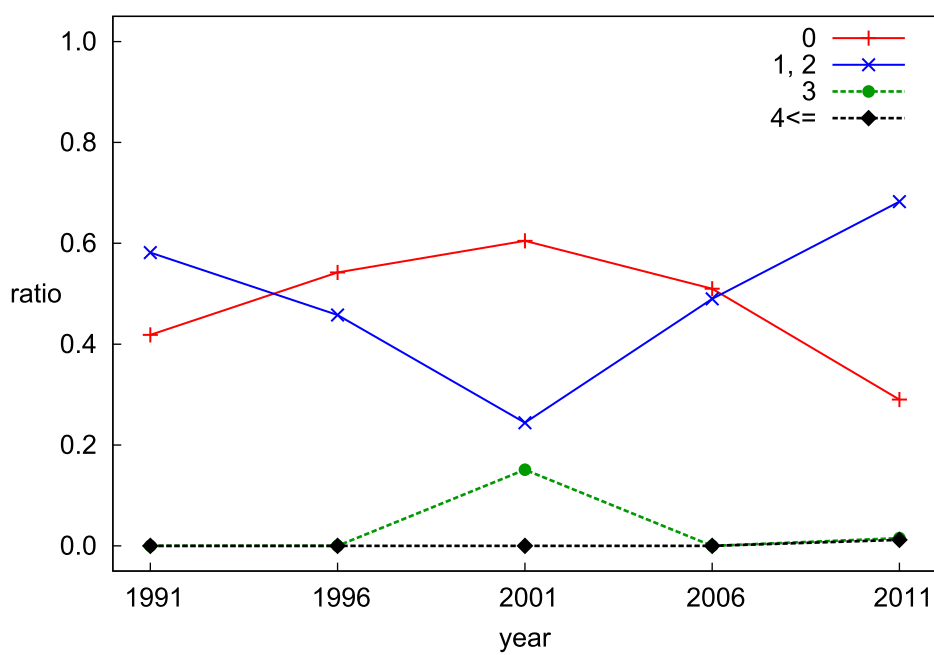
私立大学.最後に、私立大学の各グループにおける個別学力試験の科目数の推移が、図2.15に示されている。グループaにおいては、最近21年間を通して3科目の選抜が支配的であり、1991年度から1996年度までの間に68.94%から57.45%まで大きく減少した後、更に2006年度まで少しずつ減少して54.82%になったが、その後増加傾向に転じ、2011年度には59.82%になっている。0科目の選抜は、1991年度の5.46%から2006年度の18.17%まで上昇を続けたが、2011年度において18.51%であり、2006年度を境に安定的な推移に転じたように見える。1または2科目の選抜は、1991年度から2006年度まで20%台後半の水準で安定的に推移して来たが、2011年度に大幅に減少した。この減少分は概ね4科目以上の選抜の増加分に反映されており、2006年度までは低い水準で安定的に推移して来たが、2006年度から2011年度の間急激に上昇している。

グループbの個別学力試験の科目数の推移は、2006年度を境とした1または2科目の選抜の減少と4科目以上の選抜の増加がそれほど大幅なものではないものの、グループaのそれと似た特徴を示している。3科目の選抜は、1991年度の83.66%から2006年度の62.13%

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析



b



c

図 2.14. グループ分けした場合の個別学力試験の科目数の推移 (公立大学)

まで少しずつ減少して来たが、この水準で概ね安定し、2011年度においては61.80%である。また、0科目の選抜は、1991年度から2006年度まで増加を続け、3.54%から21.01%にまでなったが、その後安定に転じ、2011年度における水準は22.04%である。

グループaとbにおいては、最近21年間を通して3科目の選抜が最も大きな割合を占めていたが、グループcにおいてのみ状況が異なる。1991年度には3科目の選抜の割合が最も高かったが、1991年度の64.59%から2006年度の27.66%まで低下し続けた。一方で、0科目の選抜が1991の14.29%から2006年度の44.60%まで増加し続け、2001年度と2006年度の間、1または2科目の選抜と3科目の選抜を上回って、2011年度まで最も高い割合を占めている。

2.4.3 非学力型選抜

国立大学. 図2.16に、国立大学の各グループの非学力型選抜の動向が示されている。グループaにおいて、小論文が課される選抜の割合は、1991年度には14.39%であったが、この水準から低下し続け、小論文は2011年度にはほとんど課されなくなっている。また、面接が課される選抜の割合は、最も高い2006年度においても2.90%であり、3%以下の水準である。センター試験と個別学力試験が共に課されない選抜は、事実上行われていない。

グループbにおいて、小論文が課される選抜の割合は、1991年度から2006年度まで20%弱の水準で安定的に推移していたが、2011年度に大幅に低下し、6.52%となっている。面接が課される選抜の割合は、1991年度の8.86%から2006年度の21.36%まで上昇を続けたが、2006年度を境に低下に転じ、2011年度には12.37%の水準である。グループaと大きく異なり、学力試験が課されない選抜の割合は、1991年度から2011年度まで増加し続けており、1991年度の2.83%から2001年度の10.88%を経て、2011年度には13.57%にまでなっている。

グループcにおいては、グループa及びbと比較して大きな変動が見られる。小論文が課される選抜の割合は、1991年度の30.09%から2001年度の18.41%まで低下した後、2006

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

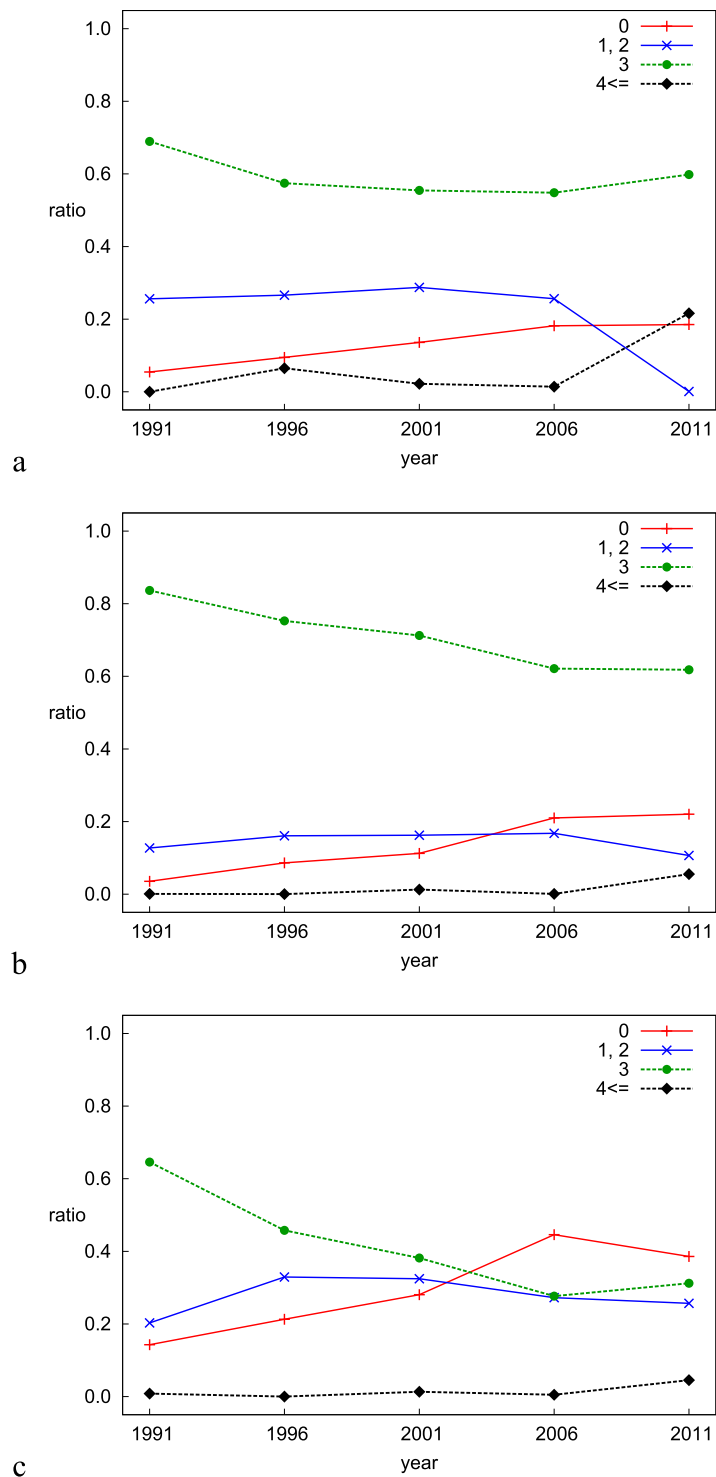


図 2.15. グループ分けした場合の個別学力試験の科目数の推移 (私立大学)

2.4. 難易度によってグループ分けした場合の推移

年度に 22.16% まで上昇したが、その後大幅に低下して、2011 年度には 4.28% となっている。面接が課される選抜の割合は、1991 年度と 2001 年度の間に 12.67% から 28.49% まで大幅に増加したが、2001 年度を境に減少に転じ、2011 年度における水準は 16.69% である。学力試験が課されない選抜は、ほぼ横ばいであった 2001 年度から 2006 年度までの期間を除いて増加しており、2011 年度には 20% に近い 18.90% の水準となっている。

公立大学. 図 2.17 は、公立大学の各グループの非学力型選抜の動向を示している。グループ b においては、2006 年度が推移の仕方の変化点のように見える。面接が課される選抜の割合が、1991 年度の 13.74% から 1996 年度の 13.31% までほぼ横ばいであったことを除き、2006 年度までが増加、その後が減少あるいは横ばいとなっている。小論文については、1991 年度から 2006 年度までの間に 20.35% から 39.67% まで増加したが、2011 年度に 13.21% まで急激に減少した。面接が課される選抜の割合は、ピークの 2006 年度は 31.11% であったが、2011 年度は 15.23% となっている。学力試験が課されない選抜は、1991 年度から 2006 年度までの間に 6.27% から 20.23% まで 3 倍以上の割合に増加したが、2006 年度を境に横ばいとなり、2011 年度は 18.64% となっている。

グループ c において、小論文と面接が課される選抜は、共に 1991 年度と 1996 年度の間に大幅に減少し、1996 年度と 2001 年度の間に大幅に増加した後、減少あるいは横ばいという挙動を示している。小論文は 1991 年度においては 22.70%、2011 年度においては 18.80% であって、面接は 1991 年度においては 22.70%、2011 年度においては 18.80% となっており、間で大きく変動したが、20 年前と比較して、これらが課される割合は減少している。学力試験が課されない選抜は、1991 年度が 5.67%、2001 年度が 24.82%、2011 年度が 27.30% となっており、グループ b よりやや早く、2001 年度の頃から横ばいの状態になっているように見える。

私立大学. 最後に、私立大学の各グループにおける非学力型選抜の動向が、図 2.18 に示されている。私立大学の非学力型選抜は、グループによってかなり大きく異なる変動を示している。グループ a においては、小論文が課される選抜の割合が、2006 年度を境に急激に低下していることが著しい特徴となっている。1991 年度から 2001 年度までは、30% 台

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

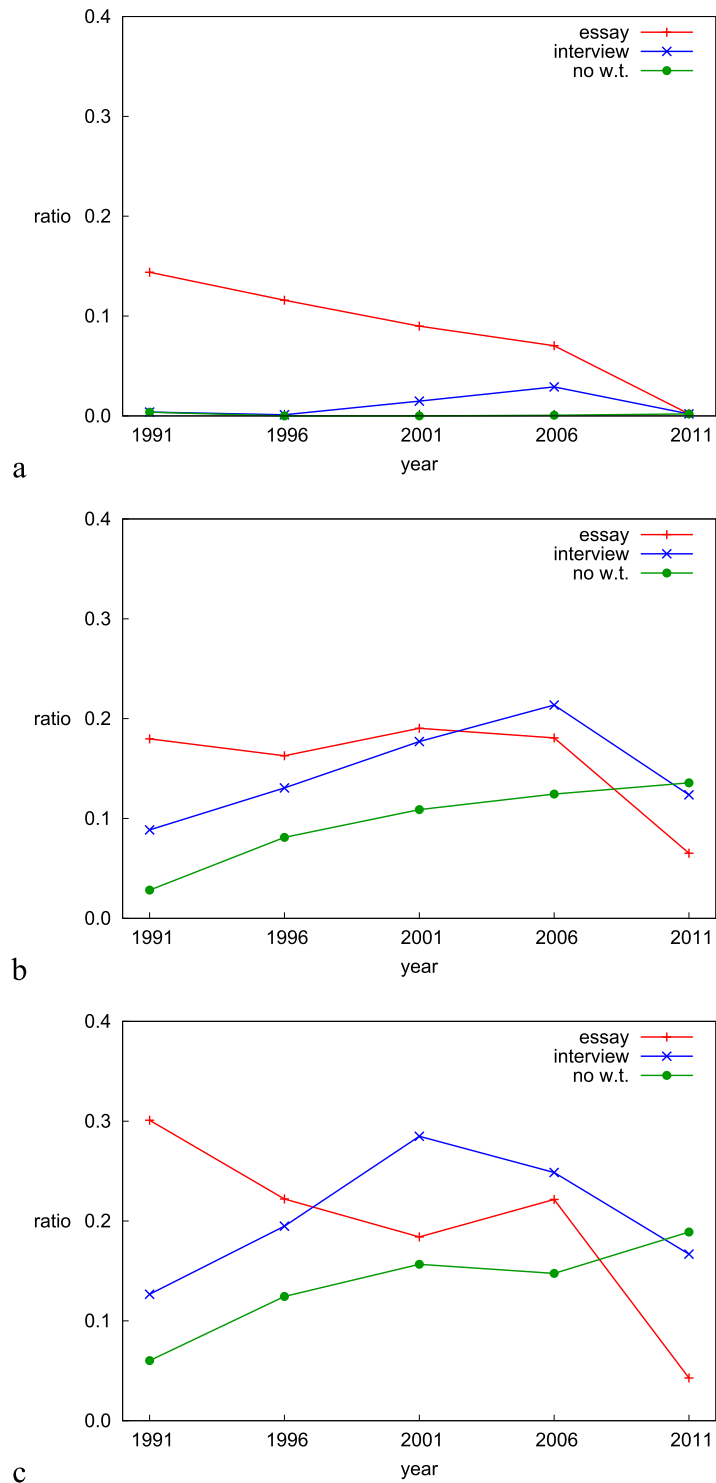
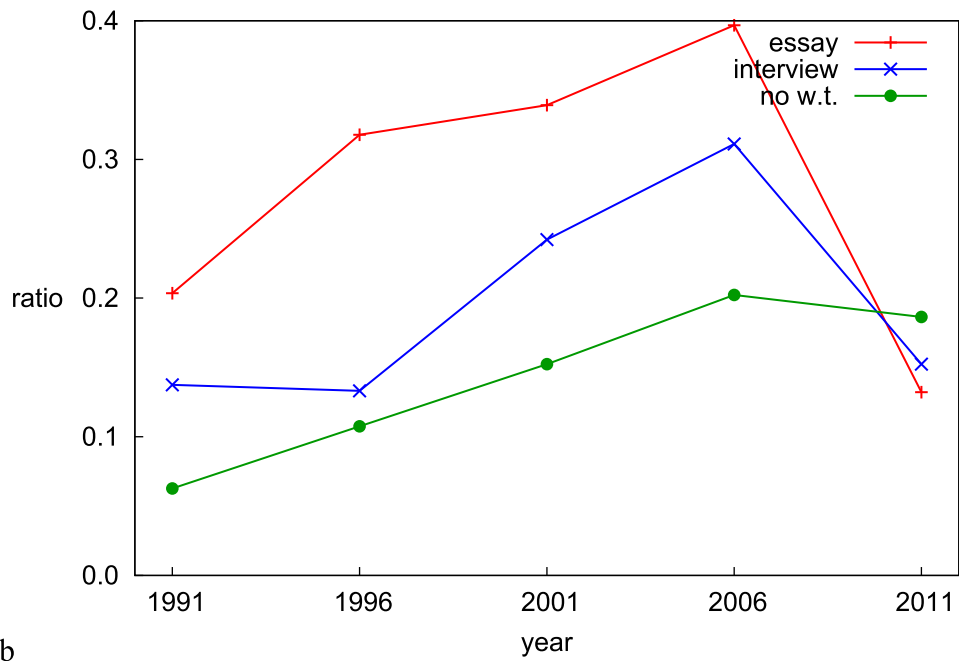
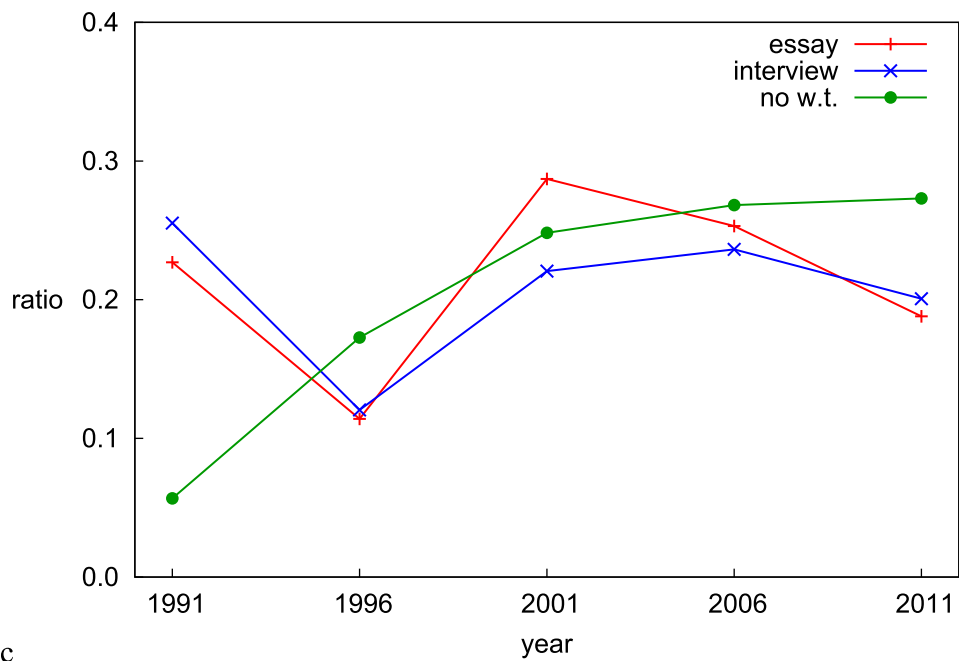


図 2.16. グループ分けした場合の非学力型選抜の動向 (国立大学)

2.4. 難易度によってグループ分けした場合の推移



b



c

図 2.17. グループ分けした場合の非学力型選抜の動向 (公立大学)

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

の前半という高い水準で安定的に推移していたが、2006年度に27.35%になり、2011年度には4.09%まで大幅に低下した。面接が課される選抜の割合は、最近21年間10%前後の水準で推移している。学力試験が課されない選抜は、1991年度の3.36%から2006年度の14.00%まで一貫して上昇してきたが、2011年度には下降に転じて11.24%となっている。

グループa及びcと比較して、グループbにおいては、小論文または面接が課される選抜と学力試験が課されない選抜の割合のどちらも、10%以下の低い水準で安定的に推移してきた。

最後に、グループcにおける非学力型選抜の動向であるが、まず、小論文が課される選抜の割合は、多少の変動はあるものの、10%前後の水準で比較的安定的に推移している。最も割合が高かったのが2006年度で15.20%、最も低かったのが1996年度で7.55%である。面接については20%前後の水準で推移しており、最も割合が高かったのが2006年度で27.76%、最も低かったのが1996年度で16.50%である。学力試験が課されない選抜の割合は、1991年度から2006年度まで上昇し続け、14.29%から33.24%になったが、その後下降に転じ、2011年度には28.35%になっている。

2.5 まとめ

表2.5に学力を中心とした選抜方式の推移の特徴が抽出されている。まず国立大学について見ると、センター試験5科目以上の結果と個別学力試験の結果からは、2001年度あるいは2006年度の頃までは科目数の減少傾向、それ以降には増加傾向が見られるのに対し、センター試験0科目の結果、及びセンター試験と個別学力試験が共に0科目の結果はこれと反するものになっており、2極化の傾向が見られる。次に公立大学については、センター試験0科目の結果、個別学力試験の結果、及びセンター試験と個別学力試験が共に0科目の結果から、2006年度の頃までは科目数の減少傾向、それ以降には増加傾向が見られる。最後に私立大学についてであるが、センター試験の結果から、徐々に利用は拡大しているものの、依然として、センター試験を利用した選抜方式の定員の全定員の中の割合は、それほ

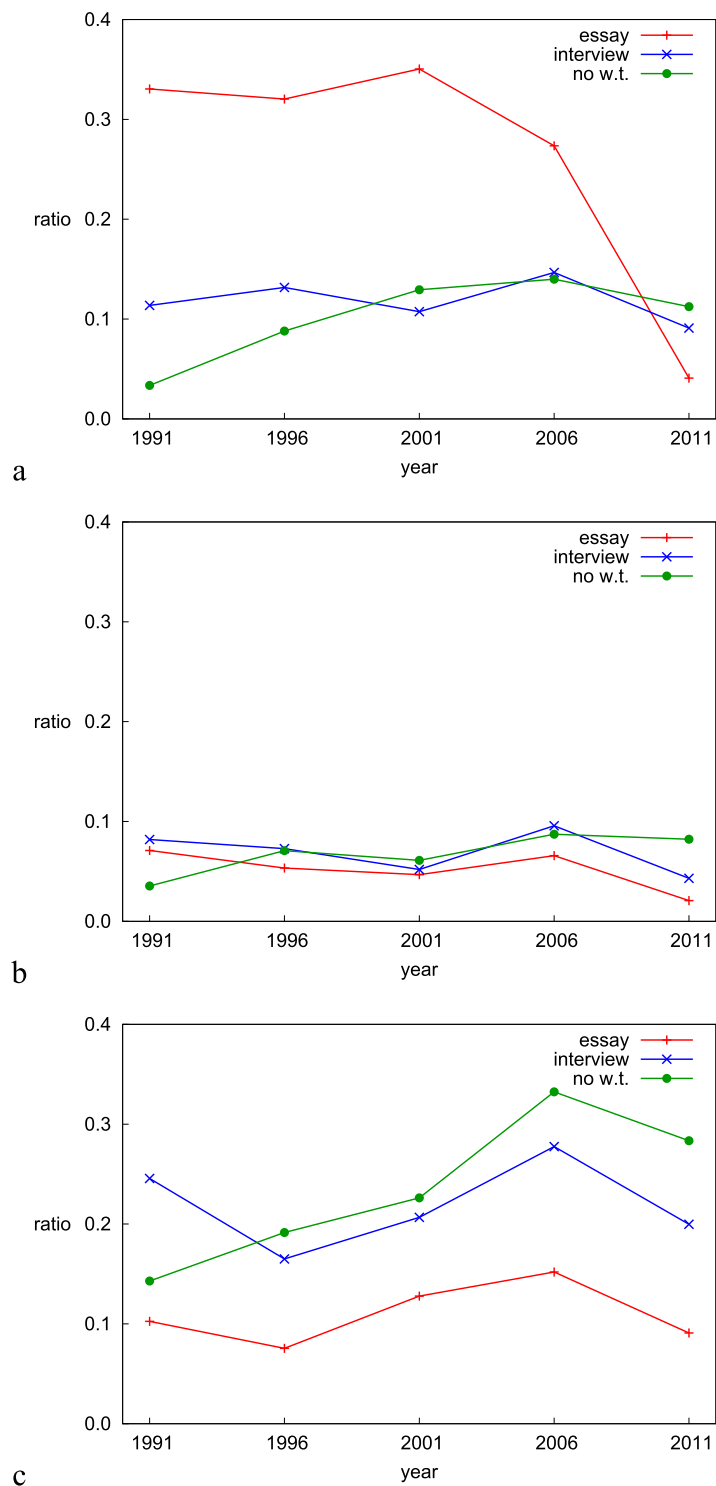


図 2.18. グループ分けした場合の非学力型選抜の動向 (私立大学)

2. セクターと難易度による分類の下での近年の大学入学者選抜の分析

ど大きくないことが分かる。また、個別学力試験の結果、及びセンター試験と個別学力試験が共に 0 科目の結果から、公立大学と同様に、2006 年度の頃までは科目数の減少傾向、それ以降には増加傾向が見られる。次に、小論文と面接に関する結果を、これらの選抜方式がこの 21 年間で普及したかという点からまとめる。国公立大学については、2011 年度の小論文または面接が課される選抜の定員の割合は、1991 年度とほぼ同じ水準か、またはそれを下回る水準となっている。私立大学については、この 21 年間、多少の変動を経験しながらも、小論文が 10% 前後、面接が 20% 前後で推移して来た。以上をまとめると、小論文と面接は広く使われた時期があったものの、これらの利用はあまり拡大していないと言える。

一般に、日本では、1980 年代の後半から大学入試の多様化が進展してきたと考えられている。入試の多様化は、1990 年代には、例えばそれが過熱した受験競争を冷却しているのかといった文脈で(天野 [27])、現在では、例えば全入時代の入学者選抜の課題(先崎 [33]、近藤 [32])といった文脈で論じられ、議論される際の文脈は変化しているが、長期間に渡り大学入試研究における 1 つのキーワードであり続けている。「入学者選抜の多様化をより一層推進することが大切」(文部省高等教育局大学課大学入試室 [48])であるという主張が 1990 年代から今日までしばしばなされてきたが、山村 [49] などによって、果たして入試の多様化は成功しているのかが、近年改めて問われている。本章では、1991 年度から 2011 年度までの 21 年間に全国大学の入学者選抜方式がどのように推移して来たのか、また現状はどのようなかを定量的に調べたが、第 2.1 節で述べたように、この調査は、入学者選抜で必要とされることを明らかにし、入試をより良くするために望まれることをスタンダードとしてまとめることを目指した入試スタンダード・プロジェクトの基礎研究として行われたものである。今後は、本章の調査の結果を活かして、大学入試のガイドラインの作成に取り組んでいきたいと考えている。

表 2.2. 入学者選抜方式の推移の特徴

センター試験	
国立	0 科目: 1991–2011 一貫して微増 5 科目以上: 1991–2001 減少, 2001–2011 増加
公立	0 科目: 1991–2006 微増 5 科目以上: 1991–1996 大幅減少, 2001–2011 増加
私立	0 科目: 1991–2011 一貫して微減 1–4 科目と 5 科目以上: 全体的に微増
個別学力試験	
国立	0 科目: 1991–2001 増加, 2006–2011 大幅減少 4 科目と 5 科目以上: 2006–2011 増加
公立	0 科目: 1991–2006 増加, 2006–2011 大幅減少 4 科目と 5 科目以上: 2006–2011 増加
私立	0 科目: 1991–2006 増加, 2006–2011 減少 3 科目: 1991–2006 減少, 2006–2011 増加
センター = 0 かつ個別 = 0	
国立	1991–2011 一貫して増加
公立	1991–2006 増加, 2006–2011 微減
私立	1991–2006 増加, 2006–2011 減少

第3章 小標本における選抜効果を補正する相関係数の推定について

岡田 謙介, 繁樹 算男

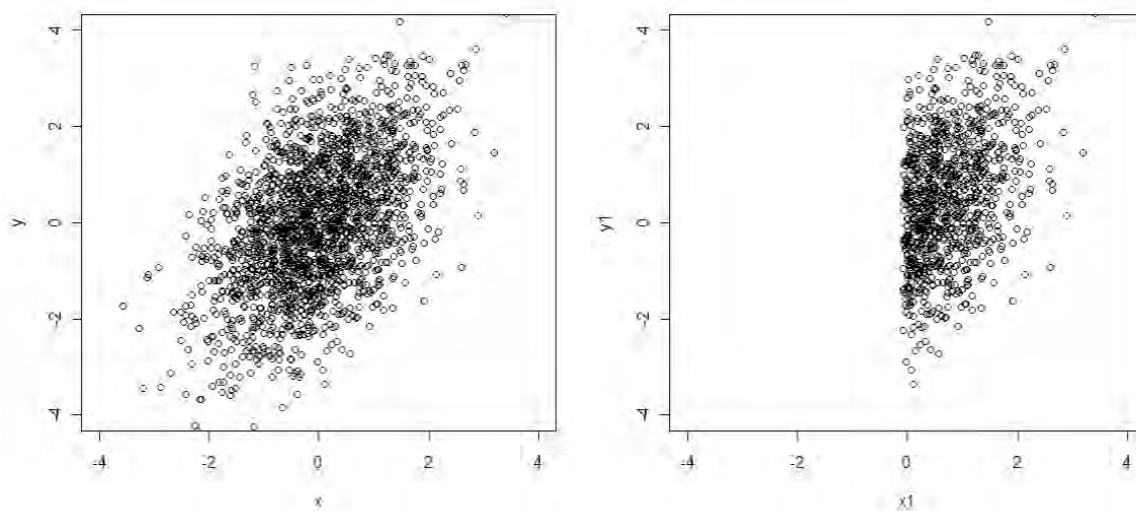
入試得点と入学後の成績との間で相関を計算すると、予想したよりも小さい値が算出されることが多い。これは入学しなかった標本において後者が欠測となるために生じる現象であり、選抜効果と呼ばれる。選抜効果の影響を除いて真の相関係数を推定する方法には最尤推定法とベイズ推定法が提案されているが、その小標本における挙動はこれまで十分検討されていない。そこで本研究では、小標本下において真の標本数・真の相関係数・欠測標本の割合を操作し、両手法を比較する数値実験を行った。結果の概要は以下の3点である。(1) 標本数が十分大きければ、どちらの手法を用いても適切な推定ができた。(2) 推定値の期待値が真値に近いかどうかという観点からは、ベイズ推定がより推奨される傾向にあった。(3) 推定値と真値の二乗誤差の観点からは、最尤推定がより推奨される傾向にあった。本研究で用いた推定のための R コードは Appendix に掲載されている。

3.1 導入

入学試験の得点と入学後の成績 (たとえば GPA) との間で相関を計算すると、当初の予想よりも小さな相関係数の値が得られることが多い。これは前者の試験を突破した入学者のみからしか後者が観測されないためである。たとえば図 3.1 に、2 変量正規分布より生成した模擬データの例を示す。図 3.1 (a) は、変数 (x, y) の組がすべての標本 (この例では 500 標本) について観測された場合の散布図である。この完全データについて標本相関係数を

3. 小標本における選抜効果を補正する相関係数の推定について

計算すると 0.70 となる. しかし, いま変数 x の上位 3 割についてしか変数 y が観測されないとする. その場合, 変数 x の値が図 3.1 (b) の点線以下である標本は欠測となってしまう, 観測される (x, y) の組のみで描いた散布図は図 3.1 (b) のように, 元の完全データと比べて変数のばらつきが小さくなってしまう. これが選抜効果の生じた典型的な例である. 実際, この図 3.1 (b) における観測データについて標本相関係数を計算すると 0.45 となっており, 元の 0.70 から大幅に小さな値が算出されてしまうことがわかる. こうした選抜効果が生じているデータから通常の相関係数を計算すると, 本来の 2 変数間の関係の大きさを小さく見積もってしまい, 極端な場合には本来関係があるにもかかわらず関係がないと結論づけてしまうことすら生じうる. 先ほどの図 3.1 で変数 x を入試の成績, 変数 y を入学後の成績と置き換えて考えてみると, 観測される標本のみから計算した相関係数では入試と入学後の成績との相関を過小に評価することになり, したがって入試の予測的妥当性が過小評価されてしまうことになる. このような事態は, 真の変数間の関係を解釈するうえで都合が悪い. したがって, 選抜効果の影響を受けない真の相関係数を統計学的に推定することが望まれる.



(a) 欠測がない場合

(b) x の値に依存して欠測がある場合

図 3.1. 人工データによる選抜効果の例

統計学では、無視可能 (ignorable) の仮定のもと、選抜効果の影響を除去した真の相関係数の値を推定する問題に取り組んできた。ここで無視可能とは、欠測がランダムな欠測 (missing at random, MAR) であり、かつ母数が特異的 (distinct) である場合をいう。順に説明する。

いま、データを $\mathbf{y} = \{y_i\}$ で、欠測データ指標 (missing data indicator) を $\mathbf{m} = \{m_i\}$ であらわすことにする。欠測データ指標は、データの i 番目の要素が観測されていれば 1、そうでなければ 0 の値をとる。また、未知母数を θ で書くことにする。このとき、欠測がデータ \mathbf{y} の値に依存しないとき、すなわちすべての \mathbf{y}, θ について

$$f(\mathbf{m}|\mathbf{y}, \theta) = f(\mathbf{m}|\theta) \quad (3.1)$$

であるとき、 \mathbf{y} は MAR であると言う。

また、簡単のため 2 母数 $\theta = (\theta_1, \theta_2)$ のモデルを考える。いま (θ_1, θ_2) の母数空間が θ_1 の母数空間と θ_2 の母数空間の直積空間となっているとき、母数は特異的であると言う。とくにベイズ推定では、これは事前分布が独立であるという仮定

$$p(\theta_1, \theta_2) = p(\theta_1)p(\theta_2) \quad (3.2)$$

となる。多母数の場合も同様である。

テストを含む社会科学分野における欠損データの解析法についての詳しい議論は岩崎 [30] や星野 [47] を参照してほしい。

統計的手法を実際のデータに適用するにあたっては、漸近論に基づく手法では言うまでもないが、そうでない場合においても小標本では推定値が期待通りのふるまいをしないことがしばしばある (レビューは、たとえば Hoyle [13])。そのため、推定量の小標本におけるふるまいは数値実験により確認することが重要である。これまでも、テストに関連する統計的手法における小標本時の統計量のふるまいを調べるシミュレーション研究は多く行われている。たとえば、Duncan and Layard [9] は相関係数の検定における漸近頑健 (asymptotically robust) 手法を比較するモンテカルロ・シミュレーションを行い、ジャックナイフ法に基づく手法が小標本下でも頑健であることを見出した。Swaminathan *et al.* [21]

3. 小標本における選抜効果を補正する相関係数の推定について

は小標本下での項目反応理論 (item response theory, IRT) モデルの項目母数推定についてのシミュレーション研究を行い, 事前情報を活用することで推定精度が向上すること, その効果はとくに 2 母数モデル・3 母数モデルで大きいことを見出した. 星野ほか [46] は既存の共分散構造分析における適合度指標についてのシミュレーション研究をレビューするとともに, 尤度比検定統計量の標本数依存性を確認する数値実験を行い, 有意水準を 5% に設定しても小標本下では棄却率が非常に大きくなる場合があることが示された.

しかし, 片方の変数の一部が欠測の場合の相関係数の推定を扱ったシミュレーション研究はわずかである. Gross and Torres-Quevedo [12] はこの問題に対するベイズ推定法を提案するとともに, 相関係数の区間推定についてのシミュレーション結果を示している. しかし, 実際のデータ解析で第一義的に考慮されるのは区間推定よりも点推定の結果であるのに対し, 彼らは点推定についての結果を一切示していない. また, 同論文では最尤法の近似信頼区間とベイズ法の事後最高密度区間という概念的に異なる区間が比較されている. とくに, この近似信頼区間は z 変換による近似を利用して導出されているものであるが, 同近似は漸近的にしか成立しない. したがって, ベイズ推定の方が区間推定幅が小さいという彼らの得た結果は, 推定法自体の性質というよりも, 筆者ら自身も指摘するように (前掲書, p.351) 区間推定に際しての近似の悪さが原因である可能性が高い. また Little and Rubin [16] でも本問題に対する最尤推定値とベイズ推定区間の比較を行っているが, これも区間推定についての結果である上に, 彼らは実データの分析結果を示しているにすぎないため真値との乖離について議論することが不可能である.

冒頭に述べたような, 選抜効果の影響を補正して真の相関係数を推定したいという場面は, 入試をはじめとして現実に多々存在すると考えられる. また, 社会科学分野の研究の常としてデータの入手にはコストがかかることが多く, とりわけ小標本下での推定量のふるまいが実際的には重要であると考えられる. このような関心に基づき, 本研究では小標本における片方の変数の一部が無視可能な欠測である場合の, 最尤推定法とベイズ推定法の点推定値の挙動をモンテカルロ・シミュレーションにより研究する. ベイズ推定における点推定値として, 本研究では事後モードを用いる. その理由は大別して 3 点ある. 第一に,

相関係数は $[-1, 1]$ の値をとる統計量であり, その真値がちょうど 0 でない場合には歪んだ標本分布をす

ベイズ推定における点推定値として, 本研究では事後モードを用いる. その理由は大別して 3 点ある. 第一に, 相関係数は $[-1, 1]$ の値をとる統計量であり, その真値がちょうど 0 でない場合には歪んだ標本分布をする. とくにその絶対値が大きくなるほど, 分布の歪みは大きくなる. こうした歪んだ分布では平均値による分布の代表性は低くなると考えられており, モードがより適切な代表値と考えられる. 第二に, 事後モードをベイズ推定量とすることは事後確率が最大となる母数の値を点推定値とすることを意味し, 点推定値の意味が理解しやすい. この枠組みは MAP (Maximum A Posteriori) 推定として広く知られ, 使われているものである. 第三に, ベイズ推定の事後モードは標本数が大きくなれば事前分布によらず最尤推定値と近づくことがわかっているが, 小標本での両者の挙動は自明ではない. そのため, 小標本において両者を比較することが有意義と考えられる. また大標本において両推定値が近づくことの確認により, シミュレーションが適切に実施されている傍証を与えることができる.

本論文の構成は以下のとおりである. 4.2 節では, 選抜効果の下で真の相関係数を推定するための方法として, 最尤推定法とベイズ推定法についてそれぞれ述べる. 4.3 節では, 数値実験の設定について具体的に述べる. 結果は 4.4 4 節で与えられ, 4.5 節でこれについての議論を行う. また, 本研究で作成された最尤推定法とベイズ推定法による真の相関係数推定のための R プログラムは Appendix に掲載されている. 読者はこれを用いて実際のデータ分析および本研究の追試を行うことができる.

3.2 母数推定

3.2.1 最尤推定法

ここでは Little and Rubin [16] の表記にならい, 選抜効果の影響を除去して相関係数の真値を最尤推定する方法について述べる. いま, それぞれ n 標本の同一な分布にしたがう

3. 小標本における選抜効果を補正する相関係数の推定について

データ (\mathbf{x}, \mathbf{y}) を考える. ここで $\mathbf{x} = \{x_i\}$ は入試得点, $\mathbf{y} = \{y_i\}$ は入学後の成績を想定している. この2変数について, 次の2変量正規分布を仮定する:

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \right). \quad (3.3)$$

ここで, この平均ベクトルを $\boldsymbol{\mu}$ で, 分散共分散行列を $\boldsymbol{\Sigma}$ であらわす. また, 変数 \mathbf{y} について, 実際には欠測のためその一部しか観測できないとする. 一般性を失わずに, $i = 1, \dots, r$ の r 標本については y_i が観測されるとし, $i = (r+1), \dots, n$ の $(n-r)$ 標本については y_i が欠測とする.

Anderson [3] は尤度分割 (factored likelihood) の手法により, この問題の最尤推定法を提案した. 彼のアイディアの骨子は, (x_i, y_i) の同時分布 $f(x_i, y_i | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ が, x_i の周辺分布 $f(x_i | \mu_x, \sigma_x^2)$ と, x_i を所与としたときの y_i の条件付き分布 $f(y_i | x_i, \beta_1, \beta_0, \sigma_1^2)$ との積で書けるということである. つまり,

$$f(x_i, y_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = f(x_i | \mu_x, \sigma_x^2) f(y_i | x_i, \beta_1, \beta_0, \sigma_1^2) \quad (3.4)$$

である. ただし,

$$\beta_1 = \frac{\sigma_{xy}}{\sigma_x^2}, \quad \beta_0 = \mu_y - \frac{\beta_1}{\mu_x}, \quad \sigma_1^2 = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2} \quad (3.5)$$

とする. ここで, 母数ベクトル

$$\boldsymbol{\Phi} = (\mu_x, \sigma_x^2, \beta_1, \beta_0, \sigma_1^2)' \quad (3.6)$$

は元の母数ベクトル

$$\boldsymbol{\theta} = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy})' \quad (3.7)$$

の1対1変換となっている.

ここで (3.4) 式の分割ができることに注意すると, 同式前半の項を最大にする最尤推定量は

$$\hat{\mu}_x = \frac{\sum_{i=1}^n x_i}{n}, \quad \hat{\sigma}_x^2 = \frac{\sum_{i=1}^n (x_i - \mu_x)^2}{n} \quad (3.8)$$

であり, 後半の項を最大化する最尤推定量は標準的な線形回帰の結果と同様に

$$\hat{\beta}_1 = \frac{s_{xy}}{s_x^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\sigma}_1^2 = s_y^2 - \frac{s_{xy}^2}{s_x^2} \quad (3.9)$$

となる. ただし, \bar{y} は (\mathbf{x}, \mathbf{y}) の両変数が観測された r 個の観測データの標本平均であり, s_x^2, s_y^2, s_{xy} は同じく r 個の観測データの標本分散および共分散である. このとき, σ_y^2 の最尤推定量は

$$\hat{\sigma}_y^2 = \hat{\sigma}_1^2 + \hat{\beta}_1^2 \hat{\sigma}_x^2 = s_y^2 + \hat{\beta}_1^2 (\hat{\sigma}_x^2 - s_x^2) \quad (3.10)$$

となり, これらを用いると, 真の相関係数の最尤推定量は

$$\hat{\rho} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} \sqrt{\frac{\hat{\sigma}_x^2}{s_x^2}} \sqrt{\frac{s_y^2}{\hat{\sigma}_y^2}} \quad (3.11)$$

によって与えられることがわかる.

なお, この値は Pearson [18] によって異なる観点から導出された推定量と一致するものである.

3.2.2 ベイズ推定法

Gross and Torres-Quevedo [12] はベイズ推定によって真の相関係数を推定することを提案した. ベイズ統計学の枠組みを利用した推論を行うためには事前分布の設定が必要である. 彼らは,

$$f(\Phi) \propto \sigma_x^{-2a} \sigma_y^{-2c} \quad (3.12)$$

の形の事前分布を用いることを提案した. しかし, 彼らの手法を用いるためには, 何らかのソフトウェア等を用いて事後分布の重積分を数値的に行う必要があった.

Little and Rubin [16] は同じ事前分布の設定のもと, 分割尤度の考え方を用いて, ρ の周辺事後分布からの標本を直接生成する方法を提案した. この事後標本を十分多く生成することにより, 容易に事後分布を評価することが可能となる. 彼らの方法は, t 回目の繰り返しにおいて以下の 3 ステップからなる.

3. 小標本における選抜効果を補正する相関係数の推定について

(1) カイ二乗分布から次の 2 乱数を生成する:

$$w_1^{(t)} \sim \chi^2(n+2a-3), \quad w_2^{(t)} \sim \chi^2(r+2c-4). \quad (3.13)$$

また, 正規分布から次の乱数を生成する:

$$z_1^{(t)} \sim N(0,1). \quad (3.14)$$

(2) $\sigma_x^2, \beta_1, \sigma_1^2$ の事後標本を以下のように生成する:

$$\sigma_x^{2(t)} = \frac{n\hat{\sigma}_x^2}{w_1^{(t)}}, \quad \beta_1^{(t)} = \hat{\beta}_1 + z_1^{(t)} \sqrt{\frac{\sigma_x^{2(t)}}{rs_x^2}}, \quad \sigma_1^{2(t)} = \frac{r\hat{\sigma}_1}{w_2^{(t)}}. \quad (3.15)$$

ただし $\hat{\sigma}_x^2, \hat{\beta}_1, \hat{\sigma}_1$ はそれぞれ (3.8) 式および (3.15) 式で与えた最尤推定量である.

(3) 次に相関係数の事後標本を生成する:

$$\rho^{(t)} = \frac{\beta_1^{(t)} \sqrt{\sigma_x^{2(t)}}}{\sigma_1^{2(t)} + \beta_1^{(t)2} \sigma_x^{2(t)}}. \quad (3.16)$$

(3.16) 式は $\rho = \sigma_{xy}/(\sigma_x^2\sigma_y^2)^{1/2}$ の定義より導かれるものである. 1 から 3 のステップを十分多く繰り返すことにより, ρ の事後分布を直接生成し, 評価することができる.

(3.12) 式の事前分布の a, c は分析に先だって与えなければならない. 今回考えている 2 変量正規分布のような多母数の位置・尺度母数族のモデルにおいては, $a = c = 1$ の場合, すなわち

$$f(\Phi) \propto \sigma_x^{-2} \sigma_y^{-2} \quad (3.17)$$

とすることによって, この事前分布が Jeffreys の事前分布となることが知られている (Box and Tiao [7]). Jeffreys の事前分布は母数の変換にたいして常にフィッシャー情報量の平方根に対応するという不変性を持ち, 代表的な無情報事前分布の一つである. Gross and Torres-Quevedo [12] および Little and Rubin [16] もこの事前分布を利用しており, 本研究でもこの設定を採用する.

3.3 シミュレーション

3.3.1 設定

前節で述べた最尤推定法とベイズ推定法の小標本下での挙動を調べるため、シミュレーションデータによる数値実験を行った。本シミュレーションにおいては、以下の3つの変数を操作した。

全標本数. 欠測がないとした場合の全標本数 n を操作した。 $n = 20, 30, 40, 50, 100, 200, 1000$ の7条件が設定された。小標本とは $n = 150$ 程度よりも小さい場合を指すことが多い (Hoyle [13]) が、 $n = 1000$ の条件は大標本の場合についての結果を押さえるために設定した。

欠測標本の割合. n のうち何割が欠測であったか、すなわち $(n-r)/n$ の比の値を操作した。 $(n-r)/n = 0.2, 0.4, 0.6, 0.8$ の4条件が設定された。

相関係数の真値. 真値 ρ の値を操作した。 $\rho = 0.2, 0.4, 0.6, 0.8$ の4条件が設定された。

以上の合計 $7 \times 4 \times 4 = 112$ 条件が設定された。1回の推定では、以下の2変量正規分布から (\mathbf{x}, \mathbf{y}) それぞれ n 個の乱数データを生成した:

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \quad (3.18)$$

その後、 $\{x_i\}$ を大きさ順に並べたときにその値が下位 $100 \times (n-r)/n \%$ に入った i について、対応する y_i の値を欠測とした。このような操作を経た (\mathbf{x}, \mathbf{y}) を入力データとした。なお、データの概要を理解しやすくするため、 (x_i, y_i) が共に観測された標本のみから計算した、希薄化を修正しない通常の相関係数の各条件における平均および標準偏差 (SD) の値を表1に示した。このように生成されたデータに対し、前節で述べた最尤推定法、ベイズ推定法のそれぞれによって真の相関係数の推定を行った。

最尤推定法は、(3.11) 式の $\hat{\rho}$ を推定値とした。ベイズ推定法における事前分布には Jeffreys の無情報事前分布を利用した。そして (3.16) 式の $\rho^{(t)}$ を $t = 1, \dots, 100,000$ の10万個生成

3. 小標本における選抜効果を補正する相関係数の推定について

することによって事後分布を構成した。これに対して、ガウスカーネルを用いたカーネル密度推定を行い、その分布のモードをベイズ推定値とした。

以上の推定を、1条件につき最尤法・ベイズ法の各々において 10,000 回反復した。今回の研究で利用した最尤推定およびベイズ推定を行うための、統計環境 R で用いることのできる関数は Appendix に掲載されている。

3.3.2 結果

表 3.2 には各条件における 10,000 回の推定値の平均値を、最尤推定およびベイズ推定の推定法別にまとめた。また、表 3.3 には真値からの二乗誤差 (mean squared error, MSE) を同じく最尤推定およびベイズ推定に分けてまとめた。いずれの表も、列方向が標本数の条件をあらわしており、右へ行くにつれて標本数が大きな条件となっている。また行方向は真の相関係数 ρ の値の 4 条件にネストした形で、欠測標本の割合の 4 条件が示されている。

まず表 3.2 において $n = 1000$ の場合を見ると、最尤推定法・ベイズ推定法のいずれにおいても推定値の平均が、一番左の列に示されている相関係数の真値とほぼ一致していることがわかる。推定値の期待値が真の値に等しくなるという性質はすなわち不偏性 (unbiasedness) である。この結果からは、標本数が大きくなれば、どちらの方法においても平均値の意味では真値を適切に推定できていると言える。なお $n = 1000$ の条件において表 3.3 の真値との二乗誤差を見ると、相関係数の真値が小さくかつ欠損標本の割合が大きい場合にやや MSE が大きくなっているものの、全体的に MSE も十分小さくなっていることがわかる。以上より、標本数が十分大きい場合には、どちらの推定法を用いてもおよそ適切な推定ができていると言える。

次に、表 3.2 より全体的な平均についての傾向を考察する。表 3.2 (a) を見ると、最尤推定法ではどの真値・欠測標本割合の条件においても、標本数が小さい場合には相関係数を過小推定し、標本数が大きくなるにしたがって平均値が下側から真値に近づいていくことがわかる。それに対して表 3.2 (b) のベイズ推定では、小標本でかつ欠損割合が大きい場合

合にはむしろ相関係数を過大推定し、標本数が大きくなるにしたがってその平均値が上側から真値に近づいていることがわかる。ただし、欠測割合が大きく標本数が小さい場合には最尤法と同様に過小推定が生じている。

いずれの手法においても、とくに平均推定値の真値からの逸脱が大きいのは、標本数が小さく、欠測の割合が大きな場合である。このような場合には、表 3.2 の (a) と (b) を見比べると、ベイズ推定の方が最尤推定よりも真値に近い平均推定値を与えていることがわかる。

次に、表 3.3 より MSE についての結果を見る。全体的な傾向としては、予想されるようにどちらの推定法においても標本数が大きくなれば、MSE は小さくなっていることが確認できる。表 3.3 の (a) と (b) を見比べると、真の相関係数が 0.2 および 0.4 の場合には、MSE は多くの条件において最尤推定の方が小さくなっていることがわかる。また、真の相関係数の値が大きくとも欠測標本の割合が大きい場合には、やはり最尤推定法が小さな MSE を与える傾向にある。しかし、真の相関係数の値が大きく、かつ欠損標本の割合が大きい場合には、ベイズ推定の方が小さな MSE を与えている。

なお、ベイズ推定では事後モードではなく事後平均を点推定値として用いる場合も少なからずある。しかし、4.1 節で述べたように相関係数の分布は歪んだ分布となるため、事後平均による分布の代表性には疑問がある。事実、本シミュレーションにおいて事後平均の平均値を計算した場合、一般に事後モードよりも真値から逸脱した値であった。

また、1 回の推定に要する時間は最尤推定のみならずベイズ推定でも手元の計算機 (Core 2 Duo 3GHz) で 0.5 秒以下であり、計算量については両手法とも実用面で問題ないと言える。

3.4 まとめと議論

前節で述べた結果を端的にまとめると次のようになる。推定値の平均が真の値に近いかどうか、すなわち不偏性と同じ考え方による評価の観点からは、とくに標本数・相関係数の真値が小さく欠測割合が大きい場合に、ベイズ推定のほうがよい傾向にあった。一方、

3. 小標本における選抜効果を補正する相関係数の推定について

表 3.1. (x_i, y_i) が共に観測された標本のみから計算した、希薄化を修正していない相関係数の各条件における平均および標準偏差

True Correlation Coefficient (ρ)	Missing Data Proportion ($(n-r)/n$)		Number of Samples (n)						
			20	30	40	50	100	200	1000
0.2	0.2	Mean	.1479	.1499	.1524	.1540	.1540	.1536	.1537
		(SD)	(.2545)	(.2031)	(.1767)	(.1574)	(.1104)	(.0775)	(.0350)
	0.4	Mean	.1270	.1247	.1297	.1288	.1310	.1299	.1313
		(SD)	(.2972)	(.2402)	(.2064)	(.1824)	(.1274)	(.0898)	(.0405)
	0.6	Mean	.1037	.1049	.1094	.1087	.1115	.1116	.1131
		(SD)	(.3770)	(.2989)	(.2562)	(.2275)	(.1603)	(.1113)	(.0495)
	0.8	Mean	.0748	.0829	.0860	.0868	.0918	.0917	.0957
		(SD)	(.5788)	(.4416)	(.3759)	(.3292)	(.2267)	(.1577)	(.0704)
0.4	0.2	Mean	.3085	.3110	.3135	.3147	.3154	.3153	.3161
		(SD)	(.2377)	(.1879)	(.1644)	(.1463)	(.1029)	(.0718)	(.0322)
	0.4	Mean	.2639	.2622	.2683	.2691	.2708	.2710	.2728
		(SD)	(.2876)	(.2303)	(.1972)	(.1748)	(.1217)	(.0858)	(.0382)
	0.6	Mean	.2171	.2251	.2306	.2290	.2340	.2338	.2366
		(SD)	(.3691)	(.2901)	(.2469)	(.2211)	(.1536)	(.1078)	(.0477)
	0.8	Mean	.1573	.1809	.1836	.1863	.1948	.1937	.1996
		(SD)	(.5772)	(.4323)	(.3684)	(.3250)	(.2229)	(.1564)	(.0689)
0.6	0.2	Mean	.4877	.4911	.4930	.4950	.4956	.4958	.4967
		(SD)	(.2027)	(.1613)	(.1405)	(.1245)	(.0869)	(.0608)	(.0273)
	0.4	Mean	.4229	.4268	.4307	.4311	.4353	.4364	.4380
		(SD)	(.2607)	(.2046)	(.1771)	(.1568)	(.1092)	(.0766)	(.0340)
	0.6	Mean	.3585	.3681	.3758	.3756	.3818	.3826	.3861
		(SD)	(.3439)	(.2702)	(.2288)	(.2032)	(.1407)	(.0991)	(.0438)
	0.8	Mean	.2725	.2965	.3099	.3079	.3225	.3225	.3307
		(SD)	(.5563)	(.4164)	(.3456)	(.3126)	(.2126)	(.1476)	(.0653)
0.8	0.2	Mean	.7039	.7078	.7097	.7109	.7120	.7126	.7131
		(SD)	(.1402)	(.1100)	(.0946)	(.0833)	(.0582)	(.0407)	(.0181)
	0.4	Mean	.6357	.6427	.6454	.6479	.6513	.6530	.6546
		(SD)	(.1986)	(.1539)	(.1323)	(.1163)	(.0797)	(.0557)	(.0246)
	0.6	Mean	.5645	.5743	.5809	.5837	.5907	.5934	.5967
		(SD)	(.2849)	(.2198)	(.1856)	(.1641)	(.1117)	(.0781)	(.0343)
	0.8	Mean	.4432	.4785	.4944	.4956	.5144	.5196	.5275
		(SD)	(.5122)	(.3704)	(.3104)	(.2720)	(.1829)	(.1254)	(.0552)

表 3.2. 最尤推定法とベイズ推定法による 10,000 回分の ρ の推定値の平均

(a) 最尤推定法

True Correlation Coefficient (ρ)	Missing Data Proportion $((n-r)/n)$	Number of Samples (n)						
		20	30	40	50	100	200	1000
0.2	0.2	.1804	.1864	.1918	.1949	.1973	.1982	.1994
	0.4	.1714	.1724	.1845	.1855	.1939	.1950	.1991
	0.6	.1440	.1535	.1676	.1717	.1851	.1909	.1985
	0.8	.0945	.1161	.1266	.1386	.1628	.1752	.1975
0.4	0.2	.3717	.3809	.3867	.3900	.3950	.3972	.3995
	0.4	.3497	.3585	.3734	.3783	.3890	.3936	.3992
	0.6	.2968	.3275	.3479	.3532	.3773	.3861	.3979
	0.8	.1969	.2515	.2709	.2924	.3395	.3592	.3934
0.6	0.2	.5726	.5824	.5869	.5906	.5952	.5974	.5994
	0.4	.5438	.5634	.5732	.5776	.5901	.5951	.5993
	0.6	.4813	.5217	.5460	.5540	.5792	.5882	.5984
	0.8	.3420	.4090	.4585	.4687	.5361	.5622	.5949
0.8	0.2	.7850	.7911	.7936	.7953	.7977	.7990	.7997
	0.4	.7688	.7832	.7870	.7908	.7960	.7983	.7997
	0.6	.7217	.7570	.7708	.7781	.7915	.7960	.7994
	0.8	.5437	.6407	.6865	.7079	.7651	.7850	.7978

(b) ベイズ推定法

True Correlation Coefficient (ρ)	Missing Data Proportion $((n-r)/n)$	Number of Samples (n)						
		20	30	40	50	100	200	1000
0.2	0.2	.2132	.2147	.2115	.2101	.2048	.2027	.2001
	0.4	.2002	.2154	.2153	.2047	.2085	.2053	.2015
	0.6	.1751	.1970	.2061	.2207	.2184	.2088	.2018
	0.8	.1276	.1445	.1603	.1812	.2065	.2151	.2044
0.4	0.2	.4250	.4243	.4181	.4172	.4078	.4040	.4007
	0.4	.4120	.4276	.4285	.4183	.4143	.4075	.4018
	0.6	.3712	.4047	.4214	.4340	.4278	.4166	.4033
	0.8	.2511	.2813	.3225	.3684	.4195	.4267	.4082
0.6	0.2	.6293	.6266	.6191	.6170	.6083	.6040	.6007
	0.4	.6245	.6334	.6312	.6226	.6147	.6071	.6017
	0.6	.5742	.6106	.6326	.6393	.6278	.6153	.6031
	0.8	.3810	.4621	.5122	.5656	.6293	.6310	.6081
0.8	0.2	.8215	.8170	.8125	.8108	.8053	.8024	.8004
	0.4	.8298	.8243	.8195	.8152	.8084	.8042	.8011
	0.6	.7997	.8244	.8290	.8279	.8171	.8085	.8018
	0.8	.5950	.6951	.7456	.7873	.8289	.8199	.8042

3. 小標本における選抜効果を補正する相関係数の推定について

表 3.3. 最尤推定法とベイズ推定法による 10,000 回分の ρ の推定値の平均二乗誤差 (MSE)

(a) 最尤推定法

True Correlation Coefficient (ρ)	Missing Data Proportion ($(n-r)/n$)	Number of Samples (n)						
		20	30	40	50	100	200	1000
0.2	0.2	.0988	.0642	.0491	.0391	.0195	.0097	.0020
	0.4	.1705	.1164	.0879	.0695	.0347	.0176	.0036
	0.6	.3111	.2171	.1661	.1349	.0709	.0353	.0071
	0.8	.6221	.4614	.3791	.3119	.1752	.0934	.0201
0.4	0.2	.0794	.0494	.0377	.0298	.0146	.0071	.0014
	0.4	.1482	.0967	.0702	.0550	.0262	.0129	.0025
	0.6	.2918	.1918	.1405	.1145	.0543	.0264	.0051
	0.8	.6419	.4506	.3603	.2955	.1521	.0773	.0144
0.6	0.2	.0476	.0289	.0215	.0166	.0078	.0037	.0007
	0.4	.1025	.0585	.0421	.0318	.0142	.0066	.0013
	0.6	.2336	.1391	.0926	.0722	.0299	.0135	.0025
	0.8	.6161	.4092	.2912	.2463	.1059	.0450	.0070
0.8	0.2	.0151	.0084	.0060	.0046	.0021	.0010	.0002
	0.4	.0360	.0171	.0118	.0082	.0034	.0016	.0003
	0.6	.1171	.0525	.0313	.0216	.0071	.0030	.0005
	0.8	.5097	.2721	.1807	.1271	.0368	.0108	.0015

(b) ベイズ推定法

True Correlation Coefficient (ρ)	Missing Data Proportion ($(n-r)/n$)	Number of Samples (n)						
		20	30	40	50	100	200	1000
0.2	0.2	.1338	.0801	.0588	.0456	.0211	.0100	.0019
	0.4	.2915	.1725	.1184	.0941	.0407	.0189	.0035
	0.6	.5457	.3876	.2917	.2190	.0935	.0406	.0075
	0.8	.8188	.7272	.6441	.5746	.3257	.1430	.0223
0.4	0.2	.0976	.0566	.0414	.0318	.0153	.0072	.0014
	0.4	.2277	.1259	.0840	.0647	.0280	.0133	.0025
	0.6	.4712	.3109	.2214	.1596	.0640	.0274	.0051
	0.8	.7958	.6911	.5897	.4979	.2506	.0982	.0151
0.6	0.2	.0505	.0280	.0209	.0161	.0079	.0037	.0007
	0.4	.1307	.0622	.0390	.0303	.0134	.0065	.0012
	0.6	.3359	.1975	.1209	.0803	.0289	.0126	.0024
	0.8	.7521	.5850	.4691	.3637	.1388	.0437	.0068
0.8	0.2	.0122	.0070	.0051	.0040	.0020	.0010	.0002
	0.4	.0284	.0133	.0084	.0063	.0031	.0015	.0003
	0.6	.1383	.0534	.0264	.0155	.0051	.0026	.0005
	0.8	.5641	.3607	.2489	.1599	.0294	.0072	.0013

個々の推定値と真値との二乗誤差の観点からは、とくに標本数・相関係数の真値が小さく欠測割合が大きい場合に、最尤推定法がよい傾向にあった。平均推定値の全体的な傾向としては、おおむねベイズ推定は(標本数と観測標本の割合が共に小さい場合を除いては)過大推定、最尤推定は過小推定の傾向があった。いずれの推定法においても、標本数が大きい場合には真値は適切に推定されており、推定法間の差はあまり見られなかった。

表 3.2 に示した推定値の平均と真値との近さと、表 3.3 に示した MSE の小ささとは、最尤推定法とベイズ推定法の間で片方がよければ片方がよくないという風に、相反するような結果が得られた。これには、相関係数が 0 から 1 の間の値をとるため、歪んだ分布をする統計量であることが原因の一つとして考えられる。これらの結果より、筆者らが推奨する方法は以下のとおりである。まず、当然のことであるが標本数を多く集めることが可能であるならば、そうすることが望ましい。シミュレーションの結果からは、1000 程度の標本数があれば、いずれの推定法を用いてもおよそ適切な推定が行えるようである。やむをえず標本数の大きくない分析をしなければならない場合には、最尤推定法・ベイズ推定法の双方を適用してみることが考えられる。計算コストはベイズ推定法の方が大きいとはいえ、現代の計算機を用いれば計算はどちらも瞬時に終わる計算である。実際の計算には本論文の末尾に掲載される R 関数が利用できる。その上で、結果の解釈にあたっては、最尤推定には全体的に平均的な意味で過小推定、ベイズ推定には同じく過大推定の傾向があることを念頭に入れつつ、また本研究で示された MSE の大きさ(不確実性)を考慮して解釈を行うことが望まれる。本研究で用いたベイズ推定法について、理論的観点から言及すべきことがある。とくに小標本でみられた過大推定の傾向には、(3.17) 式で設定した事前分布の影響があらわれていることが考えられる。この事前分布は Jeffreys の無情報事前分布として設定されたものであり、先行研究でも事前情報がない場合に利用されたものである。しかし、この分布は尺度母数 (σ_x^2, σ_y^2) のみならず、その他の母数も分散の逆数に比例させているものであり、こうした母数について必ずしも無情報と言えない形になっていると考えられる。(3.17) 式の設定は、先行研究で第一義的には事後分布の評価を容易にするために行われたものである。しかし、現在では MCMC 法の発展により、より一般的な形の事前分布も利

3. 小標本における選抜効果を補正する相関係数の推定について

用することが可能である。したがって、他の母数には正規分布など別の事前分布をおくことが検討できるが、これは今後の研究課題と言える。また、本研究で扱った y の欠測パターンは、いずれも x の値の小さい側 $100 \times (n-r)/n\%$ の標本を欠測としたものである。これは入試データに用いる場合には現実的な設定と考えられるが、欠測が x の値に依存する何らかの関数である場合などのより一般的な場合においても同様の傾向がみられるかどうかは今後の研究課題である。

3.5 Appendix

3.5.1 最尤推定のための R 関数

まず以下の 2 関数 (標本分散, 標本共分散) を定義しておく。

```
svar <- function(x) {  
  var(x) * (length(x) - 1) / length(x)  
}  
  
scov <- function(x, y) {  
  cov(x, y) * (length(x) - 1) / length(x)  
}
```

本論文で考える設定下での、最尤推定により相関係数を推定するための関数は次のとおりである。

```
selecorml <- function(x, y) {  
  y1 <- y[!is.na(y)]  
  x1 <- x[!is.na(y)]  
  sxy <- scov(x1, y1)  
  sx2 <- scov(x1, x1)  
  sy2 <- scov(y1, y1)  
  sigx2h <- svar(x)  
  beta1h <- sxy / sx2  
  sigy2h <- sy2 + beta1h^2 * (sigx2h - sx2)  
  rho.ml <- (sxy / sqrt(sx2 * sy2)) *
```

```

    sqrt(sigx2h / sx2) * sqrt(sy2 / sigy2h)
  rho.ml
}

```

この関数を用いると、たとえば x に入試得点、 y に入学後の成績のデータが入っている場合には、次によって最尤法による相関係数の点推定値が計算できる。

```
selecor.ml(x, y)
```

3.5.2 ベイズ推定のための R 関数

前節で定義した `svar()`、`scov()` の 2 関数が定義されているとする。本論文で考える設定下での、ベイズ推定により相関係数を推定するための関数は次のとおりである。

```

selecor.bayes <- function(x, y, niter = 100000){
  y1 <- y[!is.na(y)]
  x1 <- x[!is.na(y)]
  n <- length(y)
  r <- n - sum(is.na(y))
  sxy <- scov(x1, y1)
  sx2 <- svar(x1)
  sy2 <- svar(y1)
  sigx2h <- svar(x)
  beta1h <- sxy / sx2
  sig12h <- sy2 - (sxy^2) / sx2
  a <- 1
  c <- 1
  w1 <- rchisq(niter, df = n + 2 * a - 3)
  w2 <- rchisq(niter, df = r + 2 * c - 4)
  z1 <- rnorm(niter, mean = 0, sd = 1)
  sigx2 <- n * sigx2h / w1
  sig12 <- r * sig12h / w2
  beta1 <- beta1h + z1 * sqrt(sig12 / (r * sx2))
  rho <- beta1 * sqrt(sigx2) / sqrt(sig12 + beta1^2 * sigx2)
  dx <- density(rho, bw = "nrd")
  rhomode <- dx$x[which.max(dx$y)]
}

```

3. 小標本における選抜効果を補正する相関係数の推定について

```
    rhomode  
}
```

この関数を用いると、たとえば x に入試得点、 y に入学後の成績のデータが入っており、10 万個 (デフォルト) のモンテカルロ標本より事後分布を構成したい場合には、次によってベイズ法による相関係数の点推定値が計算できる。

```
selecor.bayes(x, y)
```

第4章 相関構造を仮定したテストレット モデルのベイズ推論

森 一将, 大森 拓哉, 繁樹 算男

大問形式のテストに代表されるテストレットに対して新しい IRT (項目反応理論) モデルを提案した. 提案モデルはテストレット群内の関係性をより詳細に反映させるように能力母数の分散行列に相関構造を導入したものである. 提案モデルはシミュレーション研究により推定精度が評価され既存モデルよりも全体として適切な推定を行うことが検証された. 事例分析として, モデルは学力テストデータと国立大学法人等の教育研究評価データに適用され, 既存モデルと比較した母数推定の特徴が明らかにされた. 同時に国立大学法人等の教育研究評価データがテストレットとみなせることを明らかにし, より合理的・統合的な評価手法として IRT の適用可能性を示した. 提案モデルは学力テストの結果に限らず, 授業評価や人事評価など様々なデータへの適用が期待される.

4.1 はじめに

特定の知識領域についての 1 つの文脈に関連した項目群のことをテストレットという (Wainer and Kiely [24]). テストレットは, 大問形式からなるテストがその典型例であり, 近年 Computer Based Test における適応型テストの発達と共に日本においても注目されるようになってきた.

このテストレットにおける大きな問題の 1 つが局所独立性の仮定の問題である. 局所独立性は, Rosenbaum [20] に基づくと以下のような仮定として定義される. あるテストのあ

4. 相関構造を仮定したテストレットモデルのベイズ推論

る項目への反応について、誤答を 0, 正答を 1 とし, このような 2 種類の値を取る確率変数を $\mathbf{X} = \{X_1, \dots, X_M\}$ とする. この確率変数 \mathbf{X} はその背後に存在する潜在変数 Θ と共に構造化されるとする. このとき, 局所独立性は以下のように定義される.

$$X_1 \perp X_2 \perp \dots \perp X_M \mid \Theta \quad (4.1)$$

つまり, 潜在変数 Θ を所与とした場合におけるにおける項目反応の発生確率は独立であるという仮定である. これは, ある被験者が同質の問題を複数回答したとき, それぞれの正答・誤答の反応は, 被験者の能力や問題の内容に依存せず, ランダムに発生するという状況に対応する. テストレットからなる項目反応データに項目反応理論 (Item Response Theory; IRT) モデルを適用すると, 推定が過剰に行われることは様々な研究で指摘されている. 例えば, Thissen *et al.* [22] においては読解力テストの事例を用い, (4.1) の局所独立性に関する評価を行っている. この研究では読解力テストをいくつかのテストレットから構成されるものとみなし, 局所独立性を仮定するモデルと仮定しないモデルによる情報関数曲線の比較を行った. 同時にこれらのモデルより算出された母数の推定値と, より妥当性の高い言語テストの相関を計算した. その結果, テストレットにおいては局所独立性を仮定しないモデルは言語テストとの相関が比較的低い一方で, 情報関数曲線の変化がより大きく推定されてしまうという過剰推定の状態にあることを指摘した. より以前の研究における同様の指摘については, Wainer *et al.* [23] に詳しく見られる.

この局所独立の問題に対して, Rosenbaum [20] は以下のようなテストレットに適した独立性に関する仮定を提案した. 項目反応 X_1, \dots, X_M に対して互いに要素が重複しない $L \leq M$ 個の部分集合 V_1, \dots, V_L を考え, 次のような独立性を定義する.

$$V_1 \perp V_2 \perp \dots \perp V_L \mid \Theta \quad (4.2)$$

部分集合 V_1, \dots, V_L をテストレットとして扱うと, この仮定は潜在変数 Θ 所与の下において, テストレット群間の独立のみを仮定し, 群内における項目については (4.1) のような局所独立の仮定を行わない制約条件が緩和されたモデルであることがわかる. これは, ある被験者が同質の問題を複数回答したとき, 問題が同じテストレットに属さない場合には場

合には局所独立の場合と同様に正答・誤答の反応がランダムに発生する一方で、同じテストレットに属する場合にはこれらの反応に、相関関係が発生してしまうという状況に対応する。この仮定はテストレットにおける状況をより適切にあらわしており、その後の多くのIRTモデルに用いられている。

一方で、従来のIRTテストレットモデルにおいてはテストレット群内の項目間の関係については比較的注意が払われてこられず、多くの研究において群内における相関が一定であるとみなされてきた。ところが、この設定は大問形式テストなど、ある分野の学力テストにおいては実際的なものとはいえない。加えて、学力テスト以外の評価データにおいてもこの設定が実際的でない場合がある。例えば、大学の教育研究活動評価においては、各評価項目の背後に特定の知識領域に関する文脈(例えば「教育」、「研究」)を仮定したテストレットとみなすことが妥当な場合がある。加えて、実際的には、このテストレット群内においても「教育実施体制」などの基礎的な評価項目と「学業の成果」などの応用的な評価項目に分類できることがある。この場合において、基礎的な評価項目間の相関と、基礎的な評価項目と応用的な評価項目の間の相関は必ずしも等しくないことが容易に考えられる。このように、一般的な状況においてはテストレット群内の相関は一定でないと考えるのが妥当である。

これらの問題に対して、本研究ではテストレット群内における能力母数に関して潜在構造を設定し、より適切な母数推定を行えるモデルの提案を行う。加えて事例分析を行い、大学評価に代表される学力テストを含むより広範な範囲の実例への適用について検討を行う。

本論は4節から成る。4.2節において従来モデルの検討と新しいIRTの提案を行う。提案されたモデルは4.3節においてシミュレーション研究によりその推定精度が評価・考察される。4.4節では事例分析が示される。対象は、従来のIRTの主な適用範囲である学力テストデータと、拡張的事例としての大学評価データの2つである。全体的なまとめと今後の課題は4.5節で示される。

4.2 テストレットに対する IRT モデルの提案

まず, IRT モデルに対しての定式化を行う. i 番目の被験者の j 番目のテストレットの k 番目の項目に対する反応を $x_{ijk} = \{0, 1\}$ ($i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_j$) とおき, 以下のようなモデルを設定する.

$$P(x_{ijk} = 1 | \Theta) = \Phi(u_{ijk}) \quad (4.3)$$

ここで, P は確率, Φ は正規累積関数, u_{ijk} は被験者 i の項目 (j, k) に対する潜在得点, Θ は推定の対象となる母数全体の集合を示す. このモデルを IRT における正規累積モデル (Lord and Novick [17]) という. (4.3) において, 項目反応 x_{ijk} が発生する確率は, 他の項目への反応や他の要因は所与とされておらず, Θ のみが所与とされている. つまり, 母数集合 Θ が所与の下で, 項目反応 x_{ijk} が独立と仮定されていることがわかる. したがって, このモデルは, (4.1) で定義された局所独立の仮定を満たすモデルであることがわかる. このモデルは, (4.3) の潜在得点 u_{ijk} と被験者 i の能力を示す潜在変数 θ_i の間に線形関係を設定し利用される. 例えば一般的によく利用されるものの 1 つである 2 母数モデルでは u_{ijk} に以下のような仮定をおく.

$$u_{ijk} = \alpha_{jk}(\theta_i - \beta_{jk}), \quad \theta_i \sim N(0, 1) \quad (4.4)$$

ここで, N は正規分布を示し, θ_i は被験者 i の真の能力を示す母数を示す. また, α_{jk}, β_{jk} は項目 jk の反応傾向を制御する母数でそれぞれ識別力母数, 困難度母数という. また, この 2 母数正規累積モデルの応用として, 正答反応がなされる確率 (4.3) を以下のロジスティック関数に置き換えたモデルも提案されている (Birnbaum [5]).

$$P(x_{ijk} = 1 | \Theta) = \frac{1}{\exp\{-\alpha_{jk}(\theta_i - \beta_{jk})\} + 1} \quad (4.5)$$

一方, 2 母数正規累積モデルの拡張として, 次のような誤差を仮定した変量効果モデルも提案されている (Albert [2] など).

$$u_{ijk} = \alpha_{jk}(\theta_i - \beta_{jk}) + \varepsilon_{ijk}, \quad \theta_i \sim N(0, 1), \quad \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2) \quad (4.6)$$

ここで, ε_{ijk} は誤差項を示す.

テストレットにおいては, 同一テストレット群内に属する項目の反応傾向に関して相関関係が観察される場合が多い. したがって, この様な場合, 潜在的に仮定する潜在得点 u_{ijk} についても (4.1), (4.3) のような局所独立の仮定が成立しないと考えられ, 従来から様々なモデルが提案されてきた. 代表的なテストレット IRT モデルに, Bladlow *et al.* [6] が (4.6) の変量効果モデルを元にして提案した次のようなモデルがある.

$$u_{ijk} = \alpha_{jk}(\theta_i - \beta_{jk} - \gamma_{ik}) + \varepsilon_{ijk}, \quad \theta_i \sim N(0, 1), \quad \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2) \quad (4.7)$$

このモデルでは, 2 母数モデル (4.7) に対して新たに γ_{ik} という被験者 i , テストレット群 j に依存する母数が追加される. これは, 特定のテストレットが被験者に及ぼす効果を困難度に反映したものとも考えられ, 特定被験者に対するテストレット効果と呼ばれる. Bladlow *et al.* [6] は, (4.7) の設定の下で, 被験者 i に対する 2 つの異なる項目の潜在得点 u_{ijk} の相関を解析的に計算し, 同一テストレット群 j に属する場合には高く, 属さない場合には低くなることを示した. この場合にはテストレット群内における各項目の相関はすべて一定の値をとる. ところが, 大問形式のテストや大学教育における専門科目のように, 一般的な状況においては, 同一カテゴリ (テストレット) 内においても相関の高低に差が生じてしまうことがしばしば発生する. これは同一カテゴリにおいても, 近接する問題や科目については類似する特有知識や能力を利用することにより相関が高くなり, 近接しないものに対してはあまり類似する知識や能力を利用しないため相関が低くなるというごく妥当な状況により生じていると考えられる. この場合 (4.7) のような同一テストレット内の相関を一定にするモデル設定は必ずしも適切であるとは言えず, あらたなモデルの枠組みが必要となってくる.

そこで, 本研究では (4.3) の正規累積モデルの潜在得点 u_{ijk} に対し, 2 段階モデルを提案する.

$$\mathbf{u}_{ij}^* \sim N(\theta_i \mathbf{1}_{n_j}, \Sigma_j), \quad \theta_i \sim N(0, 1) \quad (4.8)$$

$$u_{ijk} = \alpha_{jk}(u_{ijk}^* - \beta_{jk}) + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2) \quad (4.9)$$

4. 相関構造を仮定したテストレットモデルのベイズ推論

ここで, $\mathbf{1}_{n_j}$ は全要素が 1 である $n_j \times 1$ の列ベクトル, u_{ijk} は被験者の真の能力を補正する母数である. これは, 先ほどの例における特有知識や技能を考慮した, テストレットで測定する「部分的な能力」に相当すると考える. また, (4.9) において, 任意の 2 つの潜在得点 u_{ijk}^* と $u_{i'j'k'}^*$ は $i \neq i'$ かつ $j \neq j'$ であるときに独立であることが明らかであるので, (4.2) のテストレットにおける独立性の仮定を満たしていることがわかる. 加えて, (4.8) の分散共分散行列に以下のような制約を入れる.

$$\Sigma_j = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1n_j} \\ \sigma_{21} & \cdot & \cdot & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{n_j 1} & \cdots & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n_j-1} & \rho_{n_j} \\ \rho_1 & 1 & \rho_1 & \cdots & \cdot & \rho_{n_j-1} \\ \rho_2 & \rho_1 & \vdots & \vdots & \cdot & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{n_j-1} & \cdot & \vdots & \cdot & \cdot & \rho_1 \\ \rho_{n_j} & \rho_{n_j-1} & \cdots & \rho_2 & \rho_1 & 1 \end{bmatrix} = \sigma^2 \mathbf{R}_j \quad (4.10)$$

$$\rho_1 \geq \cdots \geq \rho_{n_j}$$

ただし, \mathbf{R}_j は $n_j \times n_j$ の行列であり, 以下のように定義される相関行列である.

$$\begin{cases} r_{k'k} = 1 & k' = k \\ r_{k'k} = \rho_{|k'-k|} & k' \neq k \end{cases} \quad (4.11)$$

(4.10), (4.11) で示されたような特殊な分散共分散行列の構造を自己回帰共分散構造 (Autoregressive Covariance Structure) という. 自己回帰共分散構造は成長曲線分析を中心とした時系列解析モデルでしばしば利用される仮定であるが (例えば, Lee [14], Fujikoshi *et al.* [10], Lee and Chang [15] など), 本研究ではこの構造をテストレットに適用し, 同一群内における部分的な能力 u_{ijk}^* の相関関係のモデル化を行う.

本モデルに類似するものとしてとして多次元 IRT (Multidimensional IRT; MIRT) がある. しかし, MIRT は, 能力母数 θ_i を多次元化してしまう設定をとったため (例えば, Beguin and

Glas [4] や Zheng [26]), 推定された能力母数の評価が比較的難しくなるという側面を持っている. また, このような MIRT のテストレットへの適用においても多次元化された θ_i の分散共分散行列を対角行列として扱い, 必ずしも本研究で指摘されているテストレット群内における反応傾向の相関関係を適切にモデル化していないと思われるものも存在する (Wang and Wilson [25]). 本研究においては, 自己回帰共分散構造の導入により, 近接する項目への反応の類似性の高さを適切にモデル化し, かつこのような類似性の影響を真の能力に反映させない, より柔軟な分析を可能にした.

また, 本研究ではベイジアンアプローチを導入する. ベイジアンアプローチでは, まず推定対象の母数に対して観測データ (テストデータ) を得る前の知識を事前分布によって表現する. この事前分布と母数の尤度を表すモデル関数を用いて, 観測データ所与の下での母数の確率分布 (事後分布) を求め, これを基にして母数の推定を行う (繁樹 [34]). ベイジアンアプローチは, テストデータを得る前に分析者が持つ知識を事前分布という形で取り入れることができるため, 柔軟にモデル構築・推定が行えるという利点を持つ. ベイジアンアプローチの導入のために, (4.8), (4.10), (4.11) をモデル関数とみなし, 推定対象の母数 $\alpha_{jk}, \beta_{jk}, \rho_k$ に対して以下の事前分布を設定する.

$$\alpha_{jk} \sim N(\mu_\alpha, \sigma_\alpha^2), \beta_{jk} \sim N(\mu_\beta, \sigma_\beta^2), \rho_{jk} \sim \text{TN}_{a,b}(\mu_{\rho,a,b}, \sigma_\rho^2) \quad (4.12)$$

ここで, $\mu_\alpha, \sigma_\alpha^2, \mu_\beta, \sigma_\beta^2, \mu_{\rho,a,b}, \sigma_\rho^2$ は事前分布の母数, $\text{TN}_{a,b}$ は区間 $[a, b]$ における切断正規分布を示す.

また, 一般的に (4.10) のような複雑な構造を持つ母数の推定は計算が複雑になるため困難である場合が多い. 本研究では, このような困難性を軽減するために, ベイジアンアプローチにおける数値計算アルゴリズムの一種である Gibbs Sampler (Geman and Geman [11]) をはじめとしたマルコフ連鎖モンテカルロ法を導入する. マルコフ連鎖モンテカルロ法を用いた場合における推定方法については本論末の補足に記す.

4.3 シミュレーション研究

前節で提案されたモデルに対してシミュレーション研究によりその母数推定精度を評価した。まずは、シミュレーション研究による評価のためにデータセットを用意した。このデータセットは $\sum_{j=1}^J n_j$ 個の項目より構成されている。そのうち、50% のデータを (4.3), (4.4) により、残りの 50% のデータを (4.3), (4.8), (4.9), (4.10) より発生させた。このとき、後者の 50% のデータには局所独立は仮定できず、一種のテストレットとみなすことが出来るため、通常の単問テストとテストレットの混合という一般的な状況を仮定したものといえる。また、データにおいては、単問テストとテストレットが等比率で含まれているため、モデルの構造による有利・不利が極力存在しないような設定となっている。このようなデータセットを 100 セット生成した。なお、各データセットについて $I = 100, J = 8, n_j = 8$ (すべてのテストレットで一定) とした。また、テストレットにおける各項目の相関は以下のように設定した。

$$\mathbf{R}_j = \begin{bmatrix} 1 & & & & & & & & \\ 0.6 & 1 & & & & & & & \\ 0.5 & 0.6 & 1 & & & & & & \\ 0.4 & 0.5 & 0.6 & 1 & & & & & \\ 0.3 & 0.4 & 0.5 & 0.6 & 1 & & & & \\ 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 1 & & & \\ 0.15 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 1 & & \\ 0.1 & 0.15 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 1 & \end{bmatrix}$$

これらのデータセットに対して、先行研究 (Bladlow *et al.* [6]) に基づき 4 つのモデルを適用し、能力母数 θ , 識別力母数 α , 困難度母数 β に対する推定精度の評価を行った。適用したモデルは以下の通り。

モデル 1 (4.3), (4.9), (4.10) を用いた本研究で提案された新しいテストレットモデル。

モデル 2 (4.3), (4.8) を用いた Bladlow *et al.* [6] におけるテストレットモデル。

モデル3 (4.3), (4.7) を用いた 2 母数正規累積 IRT における変量効果モデル.

モデル4 (4.3), (4.5) を用いた 2 母数ロジスティックモデル.

なお, モデル 1, 2, 3 についてはソフトウェア Open BUGS (Spiegelhalter *et al.* ??) を用いて Gibbs Sampler による推定を行い, モデル 4 については統計ソフトウェア R のパッケージ ltm (Rizopoulos [19]) を用いて最尤法による推定を行った.

推定精度評価は平均絶対誤差 (Mean Absolute Error; MAE¹) と相関係数により行った. 推定精度の評価結果を表 4.1 に示す. IRT における主な目的は被験者の能力母数 θ を正しく測定することであるが, 提案モデルはテストレットが混在する状況下において他のモデルよりも正しく θ を推定していることがわかる. また同様に困難度母数 β についても推定精度も他のモデルに比べて高い. 識別力母数 α については, Bladlow *et al.* [6] によるテストレットモデル (モデル 2), 及び 2 母数正規累積 IRT における変量効果モデル (モデル 3) の 2 つのベイズ推定よりは精度が低くなっているが, よく用いられる最尤法による 2 母数ロジスティックモデル (モデル 4) よりは精度が高くなった.

また, 各モデルの推定値と真値の相関行列を表 4.2 から表 4.4 に示す. 表 4.2 より, 能力母数 θ に関しては, 各モデルとも真値 - 推定値の相関が非常に高いことが分かる. その一方で, 識別力母数 α はどの推定値も真値との相関は高いとはいえない (表 4.3). その中でも提案モデル (モデル 1) は比較となる他のモデルに比べて, 比較的高い相関を示している. このことより, 提案モデルは識別力母数 α の推定においても限定的ではあるがある程度の推定力を持っているとも考えられる. 困難度母数 β についてはテストレットを前提としたモデル (モデル 1, 2) は 0.6 以上と高い相関を示しているが, 一般的なモデル (モデル 3, 4) は相関が 0.5 前後と比較的低くなっている (表 4.4).

以上より, 提案したモデルは群内に相関関係を想定したテストレットが存在するような通常のテストにおいて全体的により適切な推定を行うということがいえる.

¹MAE は $|(\text{母数の推定値}) - (\text{母数の真値})|$ の平均として示され, その値が小さい方がより良い推定を行っているともみなされる指標である.

4. 相関構造を仮定したテストレットモデルのベイズ推論

表 4.1. シミュレーションによる推定精度の評価

推定値対象の母数	推定値の MAE			
	モデル 1	モデル 2	モデル 3	モデル 4
能力母数 θ	109.48	124.80	110.03	115.66
識別力母数 α	15.99	7.34	13.71	20.72
困難度母数 β	16.97	19.38	17.04	29.68

表 4.2. シミュレーションによる能力母数 θ に関する真値と推定値の相関行列

	真値	モデル 1	モデル 2	モデル 3	モデル 4
真値	1.00	0.97	0.97	0.97	0.96
モデル 1	0.97	1.00	0.99	0.99	0.98
モデル 2	0.97	0.99	1.00	0.99	0.98
モデル 3	0.97	0.99	0.99	1.00	0.98
モデル 4	0.96	0.98	0.98	0.98	1.00

表 4.3. シミュレーションによる識別力母数 α に関する真値と推定値の相関行列

	真値	モデル 1	モデル 2	モデル 3	モデル 4
真値	1.00	0.29	0.24	0.23	0.13
モデル 1	0.29	1.00	0.44	0.46	0.22
モデル 2	0.24	0.44	1.00	0.29	0.16
モデル 3	0.23	0.46	0.29	1.00	0.15
モデル 4	0.13	0.22	0.16	0.15	1.00

表 4.4. シミュレーションによる困難度母数 β に関する真値と推定値の相関行列

	真値	モデル 1	モデル 2	モデル 3	モデル 4
真値	1.00	0.72	0.60	0.50	0.49
モデル 1	0.72	1.00	0.52	0.44	0.38
モデル 2	0.60	0.52	1.00	0.31	0.28
モデル 3	0.50	0.44	0.31	1.00	0.20
モデル 4	0.49	0.38	0.28	0.20	1.00

4.4 事例分析

提案モデル及び既存のモデルをテストレット型の事例データに適用し、その母数推定の特長について考察する。適用される事例は、北アイルランドの標準テストデータと日本の国立大学法人等の教育研究評価データである。

4.4.1 北アイルランドの標準テストデータ

1つ目の事例分析は、北アイルランドの小学校における標準テスト導入時の得点データ(標準テストデータ)を用いる (Airasian *et al.* [1]). このデータは、1973年から1976年に渡り、北アイルランドにおいて標準テストを導入した際の教師、生徒、父兄などへの影響を評価するために取得され、175の小学校ののべ40,000人以上の学生が参加した。本節ではそのうち、調査4年目(1976年)における4年生の評価テストデータを用いる。評価テストは、算数、アイルランド語、英語の3科目であり、それぞれは以下のような複数の単元項目から構成される。

- 算数: 計算, 概念テスト, 総合問題.
- アイルランド語: 読解 (語彙), 読解 (理解力), 用法, スペリング.

4. 相関構造を仮定したテストレットモデルのベイズ推論

- 英語: 読解 (単語分析), 読解 (理解力), 文法 (大文字小文字の区別, 句読点の用法), 文法 (用法, 会話文の選択), スペリング

この複数の単元項目は算数, アイルランド語, 英語という各科目それぞれごとについて関連性を持っている. したがって, 各科目を単元項目から構成される一種のテストレットとのみならずことが可能である.

データセットでは, 上記のそれぞれの単元項目について各受験者 (学生) の得点データが与えられている. 分析ではこれを上位 20% 以内の得点 (1) とそれ以下の得点 (0) と変換し, 単元項目ごとの 2 値データを作成した. 対象となったデータは 3,850 (人) である.

変換後の 2 値データに対して, 単元項目ごとの相関係数 (ϕ 係数) を計算した結果を表 4.5 に示す. この表に示された相関係数について代表値を計算すると, テストレット群内の相関係数の中央値は 0.53 である一方, テストレット群内の相関係数は 0.41 と差が生じていることがわかる. 項目反応の相関係数が高いということは, (4.3), (4.8), (4.9) で導かれるように, その背後に存在する潜在得点 u_{ijk} の相関が高いことを意味する. したがって, 群内の相関係数が高いことはテストレットにおいて背後に存在する潜在得点 u_{ijk} に相関関係が存在することを示唆している. 加えて, テストレット群内の各項目についても, ϕ 係数が一定ではなく, 適切な並べ替えを行うことで (4.10) のような構造に近くなることが分かる. したがって, 前述の議論と同様に, テストレット群内において事例データ (4.10) の構造を仮定することは一定の妥当性を持つ.

表 4.5. 標準テストデータにおける単元項目間の相関

評価項目		算数			アイルランド語					英語				
		1	2	3	1	2	3	4	1	2	3	4	5	
算数	1. 総合問題	1.00												
	2. 計算	0.57	1.00											
	3. 概念テスト	0.55	0.45	1.00										
アイルランド語	1. 読解 (語彙)	0.43	0.41	0.41	1.00									
	2. 読解 (理解力)	0.44	0.42	0.41	0.65	1.00								
	3. スペリング	0.37	0.40	0.36	0.61	0.53	1.00							
	4. 用法	0.35	0.35	0.34	0.58	0.54	0.52	1.00						
外国語	1. 読解 (理解力)	0.49	0.40	0.43	0.47	0.49	0.42	0.41	1.00					
	2. 読解 (単語分析)	0.49	0.38	0.42	0.48	0.47	0.42	0.42	0.63	1.00				
	3. スペリング	0.41	0.38	0.36	0.50	0.48	0.56	0.43	0.53	0.53	1.00			
	4. 文法 (用法・会話)	0.37	0.33	0.36	0.41	0.40	0.38	0.38	0.49	0.52	0.49	1.00		
	5. 文法 (大文字小文字・句読点)	0.48	0.42	0.44	0.45	0.46	0.41	0.38	0.49	0.49	0.46	0.43	1.00	

²網掛け部分はテストレット群内をしめし、それ以外の部分はテストレット群間を示す。また、テストレット群内の相関構造を分かりやすく見せるために、次の手順で一部項目の並べ替えを行った。

1. テストレット群内の各単元項目ごとの ϕ 係数の合計値を計算する。
2. ϕ 係数の合計値が一番大きい各単元項目を一番右に並べ替える。
3. その他の項目をこの項目との ϕ 係数が降順になるように並べ替える。

4. 相関構造を仮定したテストレットモデルのベイズ推論

この2値データに3節で適用した4つのモデルを適用した。識別力母数 α と困難度母数 β の推定結果を表4.6と表4.7に示す。困難度母数 β の推定値が最尤法による2母数ロジスティックモデル(モデル4)で他よりも低く推定されている。これはテストレット構造を仮定しないモデル構造であること、比較的外れ値の影響を受けやすい固定効果モデルであること、推定に最尤法を用いていることなどの理由から推定が不安定になったものと考ええる。

表4.6. 標準テストデータにおける識別力母数 α に関する各モデルによる推定値

		モデル1	モデル2	モデル3	モデル4
算数	計算	2.28	1.37	1.57	1.84
	概念テスト	2.16	1.31	1.55	1.87
	総合問題	3.78	2.74	1.88	2.26
アイルランド語	読解(語彙)	3.85	3.50	2.72	3.18
	読解(理解力)	2.23	1.43	1.80	2.10
	用法	3.33	1.98	2.54	2.94
	スペリング	2.74	1.55	2.04	2.42
英語	読解(単語分析)	3.18	2.44	2.21	2.80
	読解(理解力)	3.25	1.97	2.26	2.83
	文法(用法・会話)	1.91	1.20	1.60	1.94
	文法(大小文字・句読点)	1.85	1.20	1.88	2.26
	スペリング	2.52	1.42	2.16	2.59

また、能力母数 θ 、識別力母数 α 、困難度母数 β に対する各モデルの推定値の相関を表4.8から表4.10に示す。能力母数 θ 、困難度母数 β についてはシミュレーションと同様に各モデルとも非常に強い相関を示している。識別力母数 α については、提案モデルを含むテストレットを前提としたモデル(モデル1, 2)が非常に強い相関を示していることが見て取

表 4.7. 標準テストデータにおける困難度母数 β に関する各モデルによる推定値

		モデル 1	モデル 2	モデル 3	モデル 4
算数	計算	-0.09	-0.11	-0.07	-0.16
	概念テスト	-0.27	-0.36	-0.26	-0.50
	総合問題	-0.03	-0.03	-0.03	-0.08
アイルランド語	読解 (語彙)	-0.16	-0.18	-0.12	-0.41
	用法	-0.05	-0.05	-0.04	-0.09
	読解 (理解力)	-0.12	-0.14	-0.10	-0.30
	スペリング	-0.16	-0.19	-0.13	-0.32
英語	読解 (単語分析)	-0.19	-0.24	-0.17	-0.52
	読解 (理解力)	-0.14	-0.17	-0.13	-0.39
	文法 (用法・会話)	-0.33	-0.44	-0.30	-0.62
	文法 (大小文字・句読点)	-0.19	-0.25	-0.17	-0.39
	スペリング	-0.09	-0.11	-0.09	-0.23

4. 相関構造を仮定したテストレットモデルのベイズ推論

れる. このことから提案モデルは既存のテストレットモデルと整合的な推定を行っていることが標準テストデータの分析からも示された.

表 4.8. 標準テストデータにおける能力母数 θ に関する各モデル推定値の相関行列

	モデル 1	モデル 2	モデル 3	モデル 4
モデル 1	1.00	0.97	0.98	0.98
モデル 2	0.97	1.00	0.97	0.98
モデル 3	0.98	0.97	1.00	0.99
モデル 4	0.98	0.98	0.99	1.00

表 4.9. 標準テストデータにおける識別力母数 α に関する各モデル推定値の相関行列

	モデル 1	モデル 2	モデル 3	モデル 4
モデル 1	1.00	0.92	0.75	0.75
モデル 2	0.92	1.00	0.70	0.69
モデル 3	0.75	0.70	1.00	0.99
モデル 4	0.75	0.69	0.99	1.00

4.4.2 日本における国立大学法人等の教育研究評価データ

2 つ目の事例分析は独立行政法人大学評価・学位授与機構が 2008 年に発表した「国立大学法人等の中期目標期間に係る教育研究の状況 (平成 16 年度から平成 19 年度)」に関するデータ (国立大学法人等の教育研究評価データ) である (大学評価・学位授与機構 [50]). この評価データは, 国立大学法人等の学部・研究科を対象にして, 教育水準, 研究水準の状況についての評価を第三者機関である大学評価・学位授与機構が行ったものであり, 「中

表 4.10. 標準テストデータにおける困難度母数 β に関する各モデル推定値の相関行列

	モデル 1	モデル 2	モデル 3	モデル 4
モデル 1	1.00	1.00	0.99	0.94
モデル 2	1.00	1.00	1.00	0.93
モデル 3	0.99	1.00	1.00	0.93
モデル 4	0.94	0.93	0.93	1.00

期目標の達成状況」と「学部・研究科等の現況分析」の2つから構成される。事例分析では、学部・研究科等の現況分析に含まれる「教育の分析項目ごとの水準」と「研究の分析項目ごとの水準」を抜き出し、データ化したものを用いる。

学部・研究科等の現況分析において利用した評価項目は以下の7項目である。

- 教育水準: 教育実施体制, 教育内容, 教育方法, 学業の成果, 進路・就職の状況.
- 研究水準: 研究活動の状況, 研究成果の状況.

この複数の評価項目は、教育水準, 研究水準というより統合的な評価項目ごとに関連性を持っており、テストレットの一種とみなすことが可能である。

データセットにおいて、各評価項目は、「期待される水準を大きく上回る」、「期待される水準を上回る」、「期待される水準にある」、「期待される水準を下回る」の4段階評価を行っている。本研究における分析ではこの評価水準を以下のように加工し2値データを作成した³。

- 「期待される水準を大きく上回る」、「期待される水準を上回る」⇒ 期待される水準より優れている (1)

³実際のデータは、大学評価・学位授与機構のホームページ (<http://www.niad.ac.jp/n.hyouka/kokuritsu/hyoukakekka/index.html>) より、各法人の評価データを入手し、集計・加工を行った。また、2値データの分割方法については、元のデータが4段階であること、変換後のデータによる比率に著しい偏りが生じないことを考慮し、本文の通りに決定した。

4. 相関構造を仮定したテストレットモデルのベイズ推論

- 「期待される水準にある」, 「期待される水準を下回る」 ⇒ 期待される水準以下である (0)

教育水準の評価結果 (5 項目) 及び研究水準の評価結果 (2 項目) の両方に欠損値がない組織のみをその分析対象とした. 対象となったデータは 225 (学科・研究科) である.

標準テストの事例と同様に変換後の 2 値データに対して, 評価項目ごとの相関係数 (ϕ 係数) を計算した結果を表 4.11 に示す⁴. テストレット群間の相関に一部高い値が見られるが, この表に示された相関係数について代表値を計算すると, テストレット群内の相関係数の中央値は 0.34 である一方, テストレット群内の相関係数は 0.25 と差が生じていることがわかる. したがって, 前述の議論と同様にテストレット群内において事例データ (4.10) の構造を仮定することは一定の妥当性を持つ.

表 4.11. 国立大学法人等の教育研究評価データにおける項目間の相関

評価項目		教育水準					研究水準	
		1	2	3	4	5	1	2
教育水準	1. 教育実施体制	1.00						
	2. 教育内容	0.58	1.00					
	3. 教育方法	0.37	0.39	1.00				
	4. 学業の成果	0.36	0.34	0.25	1.00			
	5. 進路・就職の状況	0.23	0.22	0.21	0.28	1.00		
研究水準	1. 研究活動の状況	0.31	0.44	0.21	0.18	0.23	1.00	
	2. 研究成果の状況	0.28	0.37	0.18	0.20	0.29	0.66	1.00

この 2 値データに 3 節で適用した 4 つのモデルを適用した. 識別力母数 α と困難度母数 β の推定結果を表 4.12 と表 4.13 に示す. 困難度母数 β の推定値が最尤法による 2 母数

⁴網掛け部分はテストレット群内を示し, それ以外の部分はテストレット群間を示す. また, テストレット群内において, 標準テストデータの分析事例と同様の手順により項目の並べ替えを行った.

ロジスティックモデル (モデル 4) で他よりも高く推定されている。これは標準テストデータの分析結果と同様に推定の不安定さから来るものと考えられる。加えて、 β の推定において教育水準に関するテストレットの方が研究水準に関するテストレットよりも値が大きくなっている。これは、国立大学法人の真の能力 (教育・研究などを行うための潜在的能力) が第三者から評価されやすい項目は教育分野よりも研究分野であると解釈される。ただし、この傾向が国立大学法人以外のすべての大学組織に共通する一般的傾向であるかどうかは、今後より広範な対象による検討を行う必要がある。

表 4.12. 国立大学法人等の教育研究評価データにおける識別力母数 α に関する各モデルによる推定値

		モデル 1	モデル 2	モデル 3	モデル 4
教育水準	教育実施体制	1.86	1.77	2.28	1.95
	教育内容	2.71	1.84	3.13	2.71
	教育方法	1.10	1.04	1.09	1.16
	学業の成果	1.72	1.03	1.16	1.26
	進路・就職の状況	1.26	0.74	1.10	1.27
研究水準	研究活動の状況	4.00	2.33	1.99	3.70
	研究成果の状況	4.25	2.84	1.66	3.01

また、能力母数 θ 、識別力母数 α 、困難度母数 β に対する各モデルの推定値の相関を表 4.14 から表 4.16 に示す。能力母数 θ についてはシミュレーションや標準テストデータの事例と同様に各モデルとも非常に強い相関を示している (表 4.14)。ところが、識別力母数 α においては、2 母数正規累積 IRT における変量効果モデル (モデル 3) が他の 3 つのモデルに比べて低い相関を示している (表 4.15)。同様に、困難度母数 β においては 2 母数ロジスティックモデル (モデル 4) が他の 3 つに比べてやや低い相関を示している (表 4.16)。以上より、3 つの母数 θ 、 α 、 β すべてにおいて整合的な推定を行っているモデルは、提案モデル (モデル 1) と既存のテストレットモデル (モデル 2) であることがわかる。

4. 相関構造を仮定したテストレットモデルのベイズ推論

表 4.13. 国立大学法人等の教育研究評価データにおける困難度母数 β に関する各モデルによる推定値

		モデル 1	モデル 2	モデル 3	モデル 4
教育水準	教育実施体制	1.26	1.33	1.03	2.32
	教育内容	1.00	1.11	0.80	2.45
	教育方法	1.43	1.53	1.35	1.71
	学業の成果	1.57	2.11	1.74	2.39
	進路・就職の状況	1.85	2.55	1.84	2.50
研究水準	研究活動の状況	0.54	0.54	0.39	1.42
	研究成果の状況	0.49	0.49	0.42	1.17

表 4.14. 国立大学法人等の教育研究評価データにおける能力母数 θ に関する各モデル推定値の相関行列

	モデル 1	モデル 2	モデル 3	モデル 4
モデル 1	1.00	0.94	0.94	0.96
モデル 2	0.94	1.00	0.95	0.97
モデル 3	0.94	0.95	1.00	0.97
モデル 4	0.96	0.97	0.97	1.00

表 4.15. 国立大学法人等の教育研究評価データにおける識別力母数 α に関する各モデル推定値の相関行列

	モデル 1	モデル 2	モデル 3	モデル 4
モデル 1	1.00	0.94	0.40	0.94
モデル 2	0.94	1.00	0.50	0.89
モデル 3	0.40	0.50	1.00	0.59
モデル 4	0.94	0.89	0.59	1.00

表 4.16. 国立大学法人等の教育研究評価データにおける困難度母数 β に関する各モデル推定値の相関行列

	モデル 1	モデル 2	モデル 3	モデル 4
モデル 1	1.00	0.98	0.98	0.77
モデル 2	0.98	1.00	0.99	0.76
モデル 3	0.98	0.99	1.00	0.70
モデル 4	0.77	0.76	0.70	1.00

4.4.3 事例分析の考察

本節においては、テストレット型の事例データとして北アイルランドの標準テストデータと日本の国立大学法人等の教育研究評価データを適用・分析した。2つの事例における推定では、能力母数 θ の推定値が共通して高い相関を持つ結果を示した。これは前節で示されたシミュレーション研究の結果と整合的である。

識別力母数 α 、困難度母数 β については、テストレットを仮定しないモデル (2 母数正規累積 IRT における変量効果モデル, 2 母数ロジスティックモデル) についてはその相関の度合いにばらつきが生じた。これについては、対象データがテストレットの形式を持ち、局所独立性の仮定が成立しないため、推定値が安定していないことにより生じているとも考えられる。その一方でテストレットを仮定して構築されたモデル (提案モデル, 既存のテストレットモデル) は一貫して相関の高い推定を行っている。このことより、提案モデルは実際のテストレットデータにおいて、先行研究 (Bladlow *et al.* [6]) におけるモデルが類似した推定を行っていることがわかる。

また、本節においては、IRT モデルの適応範囲の拡張として国立大学法人等の教育研究評価データを分析した。結果は前述の通り、シミュレーション研究や学力テストの分析結果と整合的であり、テストレットデータとしての扱いが可能であることが明らかになった。IRT に関する代表的なテキストである豊田 [41, 42] で示されているように従来の IRT の適用は人間 (個人) を主な対象としており、組織や組織が提供するサービスに対する適用はまだ少ない。特に、大学評価は近年多面化かつ詳細化してきており、多くの項目に基づく複雑性の高い評価が行なわれており、項目を合理的に統合するための分析手法の要請も高い。これに対し、国立大学法人等の教育研究評価データをテストレットとみなし、IRT を用いて分析することにより、大学自体の総合的な評価を能力母数 θ を用いて推定する方法は単純で分かりやすく、学内で行われる学力テストの分析結果などの既存のほかの IRT による分析結果との統合も行いやすいという利点を持つ。以上より、事例分析で適用可能性が明らかになったテストレット IRT の国立大学法人等の教育研究評価データへの適用は評価の統合的理解に対しより合理的かつ有効な手段であると考えられる。

4.5 まとめと今後の展開

本論文ではテストレット群内の各項目間の背後に一定でない相関が存在するという大問形式のテストにおけるデータ構造を想定し、これに対する母数推定を適切に行うモデルを提案した。このモデルにはベイジアンアプローチを導入し、複雑な潜在構造をより柔軟に分析することを可能とした。またこのモデルは、シミュレーション研究により既存モデルと比較してより全体としてより適切な推定を行うことが検証された。

事例分析では北アイルランドにおける標準テストのデータ及び日本の国立大学法人の評価データを分析した。結果としてシミュレーション研究と同様の結果が出たことに加え、提案モデルと先行研究 (Bladlow *et al.* [6]) におけるモデルが類似した推定を行っていることが示唆された。加えて、従来個人の学力テストをその対象としていた IRT の適用範囲を大学評価という組織に対する評価データにまで拡張することについて検討を行い、提案モデルに代表されるテストレット IRT が有効な分析方法の 1 つになる可能性を示した。

今後の展開としては、まず提案モデルの事例適用が挙げられる。テストレットは実際上の適用が多い一方で、本論文で検討されているような群内の相関関係まで考慮に入れられた IRT モデルはまだ考えられていない。このような状況は、事例で挙げられた学力テストや大学評価のほか、大学カリキュラムにおける単位評価、企業における人材育成の評価測定など応用の分野は非常に多いと考える。今後、事例適用を進めていくことで、モデルの有効性の検証や応用モデルの構築を行うことが必要になってくると考えられる。

また、本研究で提案されたモデルは (4.12) で示されたように相関係数 ρ の事前分布に区間を指定した切断正規分布を用いることにより、より数値計算時の負荷を減らしている。その一方、切断正規分布の区間を指定することにより、相関係数の推定範囲が限定されてしまい、テストレット内の項目数 n_j が大きくなると柔軟性に欠ける推定を行う可能性がある。この問題への対処に関して、Lee and Chang [15] では、成長曲線モデルを対象として (4.11) で設定されたような順序性の制約のみを用いた母数推定を行っている。今後はこの研究や関連の研究を参考にしながら、テストレット IRT においてもより制約の少ないモデルの構

築法について検討を行っていくことが必要である。

4.6 補足: マルコフ連鎖モンテカルロ法による母数推定について

マルコフ連鎖モンテカルロ法では, 事前分布とモデル関数によって求められた事後分布について, 完全条件付事後分布, そして提案分布を定義する. 完全条件付事後分布は推定対象となる母数の1つのみに注目し, それ以外の母数と観測データを所与にした場合の条件付確率である. また, 提案分布は完全条件付事後分布が既知の確率分布に従わないときに用いられる代替的な役割を果たす分布である. マルコフ連鎖モンテカルロ法についてのより詳細な理論的説明については伊庭ほか [29] の第 I 部及び第 III 部を参照のこと. 本論文で提案されたモデルにおける推定手順の概略は以下の通りである.

- (1) 能力母数 θ , 識別力母数 α , 困難度母数 β , 相関係数 ρ の初期値を設定し, それぞれ $\theta^{(0)}, \alpha^{(0)}, \beta^{(0)}, \rho^{(0)}$ とする.
- (2) 以下の計算を適切な回数 (I 回) 繰り返す ($i = 1, \dots, I$).
 - (a) 完全条件付事後分布 $\pi(\theta|X, \alpha^{(i-1)}, \beta^{(i-1)}, \rho^{(i-1)})$ により i 番目のサンプル $\theta^{(i)}$ を発生させる.
 - (b) 完全条件付事後分布 $\pi(\alpha|X, \theta^{(i-1)}, \beta^{(i-1)}, \rho^{(i-1)})$ により i 番目のサンプル $\alpha^{(i)}$ を発生させる.
 - (c) 完全条件付事後分布 $\pi(\beta|X, \theta^{(i-1)}, \alpha^{(i-1)}, \rho^{(i-1)})$ により i 番目のサンプル $\beta^{(i)}$ を発生させる.
 - (d) i 番目のサンプル $\rho^{(i)}$ を発生させる. 発生は単変数 ρ_{jk} ごとに以下の手順を繰り返す.

i. i 番目のサンプル $\rho^{(i)}$ の候補 ρ' を提案分布 $q(\rho'|\mathbf{X}, \boldsymbol{\theta}^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\beta}^{(i-1)})$ より発生させる.

ii. 採択率

$$a = \frac{\pi(\rho'|\mathbf{X}, \boldsymbol{\theta}^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\beta}^{(i-1)})q(\rho^{(i-1)}|\mathbf{X}, \boldsymbol{\theta}^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\beta}^{(i-1)})}{\pi(\rho^{(i-1)}|\mathbf{X}, \boldsymbol{\theta}^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\beta}^{(i-1)})q(\rho'|\mathbf{X}, \boldsymbol{\theta}^{(i-1)}, \boldsymbol{\alpha}^{(i-1)}, \boldsymbol{\beta}^{(i-1)})}$$

を計算する.

iii. 計算された値を基に, 確率 a で $\rho^{(i)} = \rho'$ とし, 確率 $1 - a$ で $\rho^{(i)} = \rho^{(i-1)}$ とする.

(3) 適切な回数の繰り返し後, それぞれの母数に対して得られたサンプルに対して初期値に依存する最初の部分を取り除き, 残りの部分の平均をとる. これを推定値とする.

ここで手順 2 の (a) から (c) までを Gibbs Sampler といい, (d) を Metropolis-Hastings アルゴリズムという. 特に (d) では相関係数 ρ に関する推定を行っているが, (4.12) で示されているとおり, ρ に関する事前分布は切断正規分布であり, 値域が制限されているため, Metropolis-Hastings アルゴリズムにより発生した値やこれを基に計算される推定値もその取り得る値が制限を受けることになる.

第5章 セミパラメトリックベイズの枠組みによる因果効果の推定

星野 崇宏

以下は、研究会における星野崇宏氏の講演の資料である。この講演は、実際のデータから因果関係を推論しようとするときに生ずる問題について、Rubin の因果モデルという基本から始め、具体的に因果を浮き彫りにするための精緻なモデルに基づく方法について、自身の研究業績を含めた広範なトピックをカバーしている。因果を論ずるとき、原因および結果とする変数以外に関与する変数について、因果と関連しない統計的変動と見なすことができるのは、実験を事前に慎重に計画して、実施した場合のみである。実際には、無作為割り付けはもともと計画できないことが多いし、計画しても、計画通りに実施できない場合も多い。この問題に対処するために、マッチング法や共分散分析などの手法を用いるというのが、教科書的な知識である。しかし、マッチング法には実施上の困難があるし、共分散分析（とその類縁の方法）には、強い仮定を持つモデルに依存するという欠点がある。仮定がそれほど制約的でない傾向性スコアを用い、かつモデルの制約が少ないセミパラメトリック法を使うという方向が、以下の講演の骨子である。これまで、入試データの解析は、多変量正規分布のようなスタンダードな分布に頼って行われてきたが、入試において得られるデータは、母集団をきれいに反映しないものがほとんどである。入試データから入試業務に役立つ知見を得ようとするとき、以下の星野氏の講演で解説されている方法論は、非常に役に立つと考えられる。 繁榊 算男(文責)

5. セミパラメトリックベイズの枠組みによる因果効果の推定

1



名古屋大学 大学院経済学研究科
星野崇宏
(JSTさきがけ「知の創生と情報社会」研究代表者兼務)
Mail: bayesian@jasmine.ocn.ne.jp

2

内容

- ❖ 反実仮想データとRubinの因果モデル
- ❖ 欠測データモデルの広がり
- ❖ 様々な共変量調整法とその問題点
- ❖ ノンパラメトリックベイズ推定の近年の発展
特にディリクレ過程混合モデルについて
- ❖ Probit Stick-breaking Process Mixture (PSBPM)によるセミパラメトリックベイズ解析と共変量選択

3



反実仮想データと
Rubinの因果モデル

4

問題意識: 因果効果の推定

社会科学や医学・疫学などの研究の関心対象
研究者による実験的操作が困難な要因

疫学: 喫煙・睡眠・病気そのもの
医学: 手術と放射線治療・既存薬と新薬
経済学: 失業給付・職業訓練
マーケティング: 広告接触・クーポン受け取り
経営学: 人事処遇・合併行動
教育学/労働経済: 教育プログラム・親による教育投資

→ 無作為割り当てによる実験研究が困難なので、
共変量の影響を除去した、関心のある要因単独の効果
= **因果効果**を推定する必要がある
「共変量調整の必要性」と「Rubinの因果モデルの利用」

5

Rubinの因果モデル(1974)とは?

- ❖ 反実仮想(counterfactual)の仮定
ある被験者*i*が「条件1だった時の結果」 y_{1i}
と「条件2だった時の結果」 y_{2i}
(潜在的結果変数: **Potential Outcome**と呼ぶ)
の差の期待値 $E(y_1 - y_2)$ が**因果効果**
- ❖ 実際は各被験者でどちらか一方しか得られない
- ❖ よって推定値 $\frac{1}{N} \sum_{i=1}^N (y_{1i} - y_{2i})$ は得られない
- ❖ もともとのアイデアはNeyman(1923)に遡る

6

無作為割り当ての意義

ある教育プログラムの効果が知りたい。
(例) 子供の学習援助プログラム

第1群: 援助を受けさせる(実験群)
第2群: 援助を受けない(=通常教育、対照群)
(割り当ては無作為ではない)

一年後、プログラムの効果を評価するには?
~~たとえば援助期間後のテスト得点の平均の差で
プログラムの有効性を評価してよいのか?~~

7

割り当てと欠測データ

所属群	実験群			対照群		
	1	1	1	2	2	2
被験者番号	1	2	N-1	N
y_1	y_{11}	y_{12}	y_{1N-1}	y_{1N}
y_2	y_{21}	y_{22}	y_{2N-1}	y_{2N}

網掛けは実際には得られないデータの部分

- y_1 : 援助を受けた場合の1年後のテスト得点
- y_2 : 援助を受けない場合の1年後のテスト得点

プログラムの評価の基準(Rubinの因果効果)は

- 「もし全員が援助を受けた場合」のテスト得点の平均
 - 「もし全員が援助を受けない場合」のテスト得点の平均
- の差 = 個人差 (共変量) に影響されない援助のみの効果

8

無作為割り当てなら

所属群	実験群			対照群		
	1	1	1	2	2	2
被験者番号	1	2	N-1	N
y_1	y_{11}	y_{12}	y_{1N-1}	y_{1N}
y_2	y_{21}	y_{22}	y_{2N-1}	y_{2N}

2群は条件以外は等質

「もし全員が援助を受けた場合」のテスト得点の平均 $E(y_1)$ (計算できない)

|| 被験者数が多いと

「実験群に割り当てられた被験者」が援助を受けた場合の得点の平均 = 普通の「実験群の平均」

9

無作為割り当てでない場合

所属群	実験群			対照群		
	1	1	1	2	2	2
被験者番号	1	2	N-1	N
y_1	y_{11}	y_{12}	y_{1N-1}	y_{1N}
y_2	y_{21}	y_{22}	y_{2N-1}	y_{2N}

低い (例) これまでの教育水準
高い
「実験群の平均」 - 「対照群の平均」

- 「もし全員が援助を受けた場合」の検査得点の平均
 - 「もし全員が援助を受けない場合」の検査得点の平均
- の差 (因果効果)

10

無作為割り当ての効用

無作為割り当てならば
共変量の影響を無視できる

⇒ 群間差から (Rubinの) 因果効果を推定できる

無作為割り当てで無い場合はどうするか?

⇒ 従属変数や割り当てに関連する共変量の影響を除去する **共変量調整** を行う

11



12

非実験研究での因果効果の推定

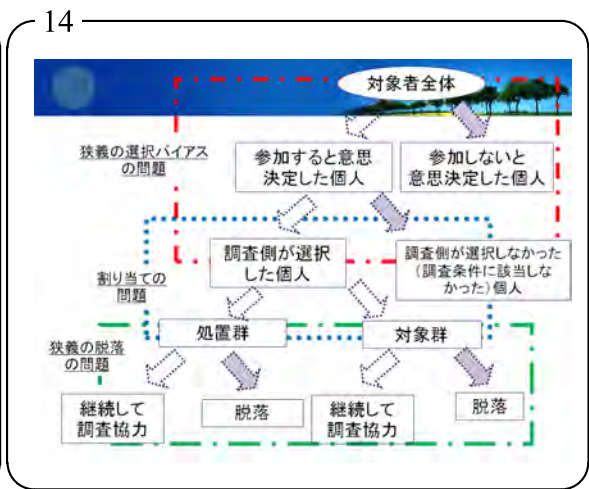
	処置群	対照群
介入を受けた場合の結果 y_1	処置群のデータ	欠測
介入を受けない場合の結果 y_0	欠測	対照群のデータ
共変量項目	全対象者に共通して得られている変数	

5. セミパラメトリックベイズの枠組みによる因果効果の推定

13

選択バイアス

	調査対象者(就労女性)	非対象者(非就労女性)
従属変数 (労働所得)	対象者のデータ	非対象者のデータ
共変量項目 (学歴等)	全対象者に共通して得られている変数	



15

疑似パネルデータ

	時点aのデータ	時点bのデータ
変数群A 2005年での就業形態	データA	欠測
変数群B 2010年での就業形態	欠測	データB
共変量項目	全対象者に共通して得られている変数	

Browning '85, Econometrica; Moffitt, '93, JofEconometrics

16

データ融合

	購買履歴データ	市場調査データ
変数群A (購買履歴) 例)IDカードからのデータ	購買履歴調査のデータ	欠測
変数群B (広告接触)	欠測	市場調査の回答データ
共変量項目	全対象者に共通して得られている変数	



- 18
- ### 既存の共変量調整法
- 1: マッチング(均衡化)
関連する方法として層別解析など
 - 2: 回帰分析的モデルの利用
 - 3: ノンパラメトリック回帰
カーネル回帰など
 - 4: 傾向スコア
 - 5: Heckmanモデルの利用

19

既存の共変量調整法の前提条件

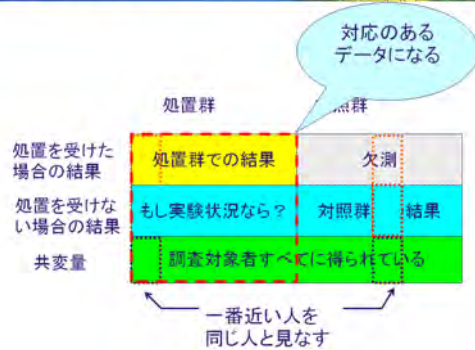
5のHeckmanモデルを除く共変量調整法ではランダムな欠測(Missing at random) または統計的因果推論ではStrong Ignorabilityを仮定
 具体的には

$$p(z | y_1, y_2, x) = p(z | x)$$

どちらの群になるかは「潜在的な結果変数」には依存しない
 ⇒ 共変量によってのみ決定される
 = どんな共変量を選ぶかが重要になる

20

マッチングによる因果効果の推定



21

マッチング(均衡化)

処置群(実験群)と対照群で共変量の値が同じになる対象者のペアを作る
 例) 年齢と性別、病気の進行度が同じ「手術群」と「投薬群」の患者をペアにして比較
 → 「共変量が同じ人は(治療後の)経過がまったく同じはず」ならばこれでよい

方法

- 完全マッチング 共変量がすべて同じペアを作成
- 最近傍マッチング(Nearest Neighbor Matching) 距離が近いものをマッチング

例) マハラノビスマッチング $\{(x_i - x_j)' S^{-1} (x_i - x_j)\}^{1/2}$
 但しxは共変量、Sはその標本共分散行列

22

マッチングの欠点

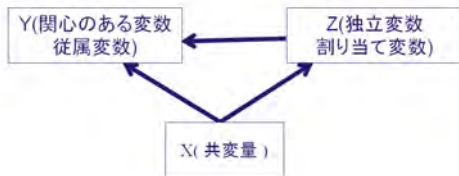
マッチングの欠点

- 共変量に連続変数が存在するとき(血圧など) → 完全マッチングはできない(最近傍マッチング)
- 共変量の数が多いと実行が難しい(次元問題) マハラノビスマッチングでは離散変数を含むと問題
- 距離の定義の方法によって結果が大きく変わる
- 共変量の分布が処置群と対照群で重ならない部分があるとき 例) 血圧が一定以上ならば手術できない = サポート問題(Support Problem)
- 健常群と疾患群の比較 疾患群の数が少ない → 健常群のデータが無駄になる * 多対多マッチングによる対処は可能

23

回帰分析モデル

応用研究で最もよく利用される手法は回帰分析的な手法 共分散分析、ロジスティック回帰、Cox回帰、パス解析、グラフィカルモデリング、構造方程式モデリング、...



24

共分散分析モデルとは?

「潜在的な結果変数」について対象者iのデータは
 $y_{i1} = \tau_1 + \beta^t x_i + \epsilon_{i1}$ $z = 1$ 処置群に所属するとき
 $y_{i2} = \tau_2 + \beta^t x_i + \epsilon_{i2}$ $z = 0$ 対照群に所属するとき

観測される従属変数は $y_i = z_i y_{i1} + (1 - z_i) y_{i2}$

$y_i = \tau_2 + (\tau_1 - \tau_2) z_i + \beta^t x_i + \epsilon_{i2} + (\epsilon_{i1} - \epsilon_{i2}) z_i$
 誤差が2群で同じ分布に従えば、結果として

$y_i = \tau_2 + \tau_z z_i + \beta^t x_i + e_i$ (共分散分析モデル)

因果効果は $E(y_1 | x) - E(y_2 | x) = \tau_1 - \tau_2 = \tau_z$

5. セミパラメトリックベイズの枠組みによる因果効果の推定

25 「ランダムな欠測」が成立するならば

特にこのモデル設定では共変量xによってどちらの群に所属するかが決まっている

$$p(z | y_1, y_2, x) = p(z | x, \phi)$$

とする「強く無視できる割り当て(Strong Ignorability)」条件
 * 但しφは群所属に関するパラメータ
 が成立すれば、関心のあるパラメータθを求めるため尤度は

$$p(y_{obs}, z | \theta, \phi) = \prod_{\text{処置群}} \int p(z_i = 1 | y_{i1}, y_{i2}, x_i) p(y_{i1}, y_{i2} | x_i, \theta) dy_{i2}$$

$$\times \prod_{\text{対照群}} \int p(z_i = 0 | y_{i1}, y_{i2}, x_i) p(y_{i1}, y_{i2} | x_i, \theta) dy_{i1}$$

$$= \prod_{\text{処置群}} p(z_i = 1 | x_i, \phi) p(y_{i1} | x_i, \theta) \times \prod_{\text{対照群}} p(z_i = 0 | x_i, \phi) p(y_{i2} | x_i, \theta)$$

26 共分散分析モデルと「ランダムな欠測」

従って関心のあるパラメータθだけについて考えればよい
 具体的にはこのモデルでの尤度(観測データの尤度)は

$$p(y_{obs} | \theta) = \prod_{\text{処置群}} p(y_{i1} | x_i, \theta) \times \prod_{\text{対照群}} p(y_{i2} | x_i, \theta)$$

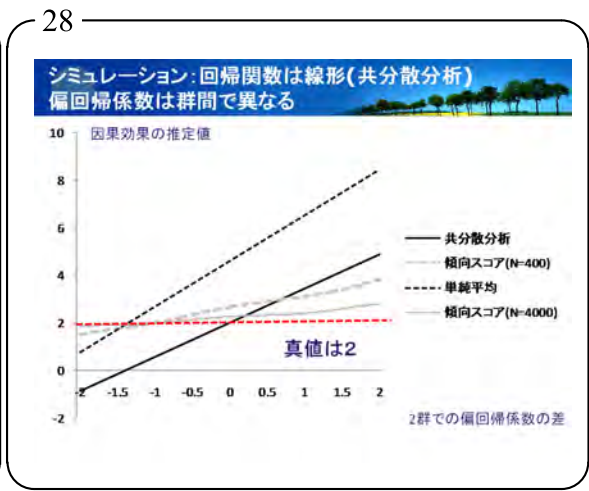
を最大化すれば(正しい)最尤推定値が得られる
 ところでこれは通常の「共分散分析モデルでの尤度」
 ⇒普通の共分散分析で良い!
 (注意)この場合最尤法も最小二乗法も同じ

27 回帰分析的モデルの前提条件

回帰分析的モデルを用いてRubinの因果効果を推定する場合には

- [1] 共変量への潜在的な結果変数の回帰関数が線形(または既知の関数形)である
- [2] 上記の線形関数が潜在的な結果変数(y1,y2)間で共通である
- [3] 誤差変数の分布が潜在的な結果変数(y1,y2)間で共通である

という強い仮定を暗黙の内に仮定している!



29 シミュレーション: 回帰関数が2次項を含む

因果効果の推定値

7 回帰分析的なモデル(事前に共変量と従属変数の回帰関係を仮定する必要があるすべてのモデル)は**モデル仮定に誤りがあると因果効果の推定に大きなバイアス**

6 ⇒回帰関数を仮定しない方法

5 =セミパラメトリックな解析手法が近年利用

4 その一つが「傾向スコアを用いた解析」

3

2

1

0

30 ではノンパラメトリック回帰は?

Cheng(1995,JASA)ではカーネル回帰による因果効果推定法が提案

$$\hat{E}(y_1) = \frac{1}{N} \sum_{i=1}^N (z_i y_{i1} + (1 - z_i) \hat{E}[y_{i1} | x_i])$$

ここで $\hat{E}[y_{i1} | x_i]$ は条件付き平均のNadaraya-Watson推定量

Xの次元が大きいと「次元の呪い」=サンプルサイズ大でも推定値にバイアス
 * 正則化? ⇒実用的?

31



32

傾向スコア

Rosenbaum & Rubin (1983)
無作為割り当てされていない2群の様々な共変量についての群間差を単一の指標にまとめた値

第*i*対象者の共変量の値を X_i 、割り当て変数の値を Z_i とするとき、その対象者が群1へ割り当てられる確率 e_i :

$$e_i = \Pr(Z_i = 1 | X_i)$$

を第*i*対象者の傾向スコアという。
(具体的には)傾向スコアは割り当てと共変量とのロジスティック回帰モデルによって推定されることが多い。

33

傾向スコア解析のモデル設定

Y(関心のある変数 従属変数)
Y:多次元
X:多次元

Z(独立変数 割り当て変数)
Z:1次元
X:多次元

X(共変量)

(1) $p(y, z, x)$
の同時分布の形はわからないし、関心がない

(2) $p(y|x)$ の条件付き分布もわからない
(仮定が難しい)

34

傾向スコア解析のモデル設定

(3) 代わりに以下の分布・モーメント構造は分かっている又は関心がある

$$p(Y) \text{ and } p(Z|X)$$

分布をすべて仮定する「パラメトリック」な手法から、なるべく仮定を少なくする「セミパラメトリック」な手法へのシフトが、様々な分野に広がっている

⇒傾向スコア法などもその有力な手法として、医学や経済学・経営学・社会学・教育学で利用されている

35

マッチング・層化・共分散分析 (Rosenbaum & Rubin, 1983,)

Rubinの因果効果の推定法として3つ紹介

1: マッチング
傾向スコアが等しい処置群と対照群の被験者をペアにして、その差の平均を用いて推定値とする。

2: 層別解析(Subclassification)
傾向スコアの大小によってサブクラスにわけ(通常5クラス)、そのサブクラスごとに処置群と対照群の平均の比較、その結果の併合を行う
⇒複数の共変量の一次元化(傾向スコアとして)
⇒既存のマッチングと層別の問題点の解決

3: 傾向スコアを共変量とした共分散分析

36

IPW推定量: Inverse Probability Weighting

(Rubin, 1985, Robinsら1994)「傾向スコアの逆数による重み付け平均」による推定法

$$\hat{E}(y_1) = \sum_{i=1}^N \frac{z_i y_{1i}}{e_i} / \sum_{i=1}^N \frac{z_i}{e_i}$$

傾向スコア

$$\hat{E}(y_2) = \sum_{i=1}^N \frac{(1-z_i) y_{2i}}{1-e_i} / \sum_{i=1}^N \frac{1-z_i}{1-e_i}$$

なぜなら $E[\frac{z_i}{e_i}] = 1$ より

5. セミパラメトリックベイズの枠組みによる因果効果の推定

37

傾向スコアによるM推定量の重み付け

傾向スコアの推定値を重みとした
M推定量(最尤法や最小二乗法含む)の評価関数

$$Q^W(\theta, \hat{\alpha}) = \sum_{i=1}^N \sum_{j=1}^J \frac{z_{ij}}{e_j(x_i, \hat{\alpha})} m(y_{ij} | \theta_j(\theta))$$

を最大化する θ は一致性と漸近正規性がある(Hoshino, Kurata & Shigemasa, 2006)
 ⇒「因果効果(=平均についての推論)」だけでなく、
 等質でない2群の相関や回帰、SEMなど様々なモデルに利用可能

例) 非実験データから「無作為割り当てされた時の相関係数の差」の推定、など

38

カーネルマッチング

Heckmanら(1998)
 (局所線形)カーネルマッチング

$$K_{ij} = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{1}{2h^2}(e_i - e_j)^2\right)$$

$$K_{ij} = K\left(\frac{e_i - e_j}{h}\right)$$

対照群での傾向スコア
 変数 y_2 がある

処置群での傾向スコア
 変数 y_2 がない

39

カーネルマッチング

カーネルマッチング
 $z=1$ なら y_2 欠測
 $z=0$ なら y_2 観測

$$\hat{y}_{2i} = \frac{\sum_{j=1}^N (1-z_j)w_{ij}y_j}{\sum_{j=1}^N (1-z_j)w_{ij}}$$

重み $w_{ij} = K_{ij}$

第 i 対象者が $z_i=1$ の時
 「もしこの人が対照群に割り付けられていたら」
 を重み付け推定

$$\hat{E}(y_2) = \frac{1}{N} \sum_{i=1}^N ((1-z_i)y_{1i} + z_i\hat{y}_{2i})$$

$$K_{ij} = K(e_i, e_j) = (2\pi)^{-1/2} \exp\left(-\frac{1}{2}\left(\frac{e_i - e_j}{h}\right)^2\right)$$

* 共変量が多い場合は
 一度「傾向スコアに次元を落として」カーネル回帰
 ⇒一般化して局所線形回帰(Fan, 1993, Heckman, 1998)



41

冠動脈疾患患者に対するアスピリンの有用性
 (Gumら, 2001; アメリカ医学会誌)

アスピリンは血小板凝集抑制効果のある鎮痛剤であるが、近年抗動脈硬化作用なども期待
 (しかし)冠動脈疾患患者へのコホート研究で単純な比較ではアスピリン服用群と非服用群で死亡率にはなかった。

服用群と非服用群への割り当ては無作為ではない
 (服用群の方が病状が重い人が多い?)

⇒年齢や他の病気・治療薬の服用など34変数を共変量として傾向スコアを算出して調整
 ⇒服用群と非服用群では差が2倍と大きく開く

42

「銃犯罪」という環境の影響
 (Bingenheimerら2005, Science)

縦断データを用いた研究
 銃犯罪が発生した地域の青少年がその後重大犯罪を犯す比率は、そうでない地域より3倍以上
 (しかし銃犯罪⇔悪い環境というだけでは?)

⇒傾向スコアを用いて調整(属性・薬物使用・家族環境・仲間集団での犯罪)
 ⇒共変量を調整しても、重大犯罪を犯す比率は2倍以上
 ⇒「もともと悪い環境で育ったから」という可能性を排除

43

定期借家権制度と賃貸住宅市場
(大竹・山鹿, 2003)

賃貸住宅供給の促進と不動産証券化の促進を目的として2000年に定期借家権制度が整備

定期借家権制度の方が新しい物件が多くなり(⇒家賃高い)、単純な比較ができない

⇒物件の交通の便や築年数など8変数を共変量として傾向スコアを算出して調整

⇒定期借地権制度の方が家賃が安くなる法制度が有効に機能しているようだ

44

学力別授業の効果
(Congdonら2008, Sociological Quarterly)

Early Childhood Longitudinal Study

全米からの21260人のデータ 小学校1・3年生

【従属変数】読解力得点

【独立変数】学力別授業をしている群vsしていない群 (低・中・高能力群) 2群に分けると差は少ないが...

【共変量】SES、人種、年齢、養子か、家族構成、親の子供への学習期待、学校についての変数など29変数(ダミー変数あり)

【結果】低読解力群になると読解力は低下

高読解力群になると読解力は向上

⇒学力別授業は「差をますますつける授業だった」

45



46

Hecmanのプロビット選択モデル
(Tobit Type-V)

2つの条件に対応する潜在的な結果変数がある

$$y_{i1} = x_i' \beta_1 + u_{i1} \quad y_{i2} = x_i' \beta_2 + u_{i2}$$

$$y_{ie} = x_i' \beta_e + u_{ie}$$

$y_{ie} > 0$ なら y_{i1} が観測 $y_{ie} \leq 0$ なら y_{i2} が観測

このとき因果効果は

$$E(y_1) - E(y_2) = E(x)'(\beta_1 - \beta_2)$$

* 誤差 u_1, u_2, u_e に通常は3変量正規分布を仮定する (この制約は外すことができるが特定の分布仮定は必要)

⇒最尤推定かHeckmanの2段階推定が利用

47

プロビット選択モデルと無視できない欠測

このモデルで y_2 が観測されない確率は?

$$P(y_2 > 0 | y_1, y_2, x) = \Phi\left(\frac{1}{\sqrt{1 - \sigma_{e1}\sigma_{e2}}} \{x\beta_e + \sigma_{e1}\sigma_{e2}^{-1}(y_1 - x\beta_1)\}\right)$$

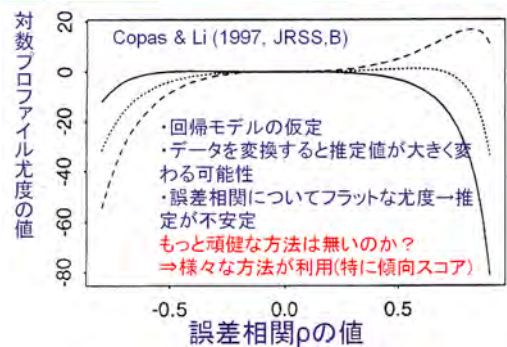
$$\text{但し } \sigma_{ab} = \text{cov}(u_a, u_b)$$

$y_{ie} > 0$ では観測されない y_2 に依存する

⇒「欠測値に依存する」ランダムでない欠測

48

プロビット選択モデルの問題点




5. セミパラメトリックベイズの枠組みによる因果効果の推定

49



ノンパラメトリックベイズ推定の
近年の動向

50




ベイズ推測の利点と欠点

この研究会では私があえて述べるほどのことでは
ありませんが...

【利点】
局外要因の周辺化
局外パラメータの積分消去
事後予測分布⇔plug-in method
主観確率の導入
データの継時的利用とupdate
モデル比較についての整合性

【欠点】
分布仮定を明確に行う必要がある

51



混合分布による密度関数の近似


あらゆる分布は「ある要素数を持つ基底分布の混合分布で表現可能」(Sethuraman, 1994)

例: 正規混合分布を用いて、任意の分布 $p(y)$ は

$$p(y) \approx \sum_{l=1}^{\infty} p_l N(y | \mu_l, \sigma_l^2)$$

従って「基底分布」として特定の分布(例えば正規分布)の混合分布を利用したノンパラメトリック密度推定が可能
但し 要素数は未知
通常の「要素数既知」のモデル+モデル選択で良いか?
⇒要素数が数十になりえる...モデル選択は難しい

52




既存の混合分布モデリングの問題点

(1) 混合要素数(潜在クラス数)の推測が難しい
境界問題を含むために尤度比検定は問題があると以前から指摘されている
⇒モデル比較(周辺尤度・AIC・BIC・DIC...)

(2) 混合要素数の候補が複数ある場合、その要素数ごとにモデリング→推定
→モデルの適合度の計算
⇒計算量が非常に多くなる

53




問題点に対するこれまでの解決法

(1) に対してはベイズ的手法
⇒モデルの事後確率(又は周辺尤度)の計算
* 情報量基準の多くは要素数を過大に推定することが知られている

特にReversible Jump MCMC法 (Richardson & Green, 1997, JRSS, B) が有力な手法
* 複数のモデル(要素数)を移りながら、各モデルでの母数推定を実行
* 各モデルを訪れた割合=モデルの事後確率

54



Reversible Jump MCMC法

Green(1995, Biometrika)
複数のモデルを移動しながらMarkov Chain Monte Carlo法を実行する方法

(1) 各モデル内でMCMCを行い、条件付き事後分布から母数を発生
(2) そのモデルから別のモデルへ移る確率を計算し、その確率で移動
(3) 移動しなければ、そのまま元のモデルに留まる

55

Reversible Jump MCMC法

具体的に混合分布の例なら



56

Reversible Jump MCMC法の問題点

- これにも様々な問題点が指摘
- ❖ 各要素数(モデル)へのジャンプの提案確率の設定が恣意的になる
⇒適切なジャンプ確率の設定の問題
- ❖ 候補となるモデル数(要素数)はせいぜい数個程度まで
- ❖ 別々にモデルを構成する必要がある
⇒プログラミングの問題
- ❖ 計算量が膨大になる((2)は解決されず)

57

ディリクレ過程とは

Stick-breaking表現(Sethuraman,1994)

ある確率変数 θ が分布 G_0 から可算無限個発生されると考え、
そのうち θ_i が実現する確率は p_i であるとする。

$$G(\theta) = \sum_{i=1}^{\infty} p_i \delta_{\theta_i}, \quad \theta_i \sim G_0$$

但し δ_{θ_i} : デルタ関数

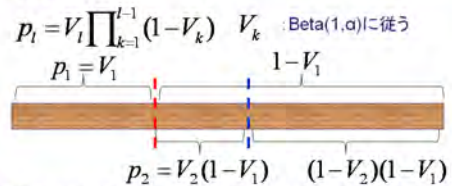
$$p_i = V_i \prod_{k=1}^{i-1} (1 - V_k) \quad V_k \sim \text{Beta}(1, \alpha)$$

他にもChinese restaurant表現などがある

58

Stick-breakingとは

文字通り有限長(=確率1)の棒を段々折っていくということ...



無限の分割が可能・pの総和も1になり矛盾が生じず

p_i は1が小さくなるほど小さくなる傾向

⇒無限次元ディリクレ分布 $E(p_i) = (1/(1+\alpha))(\alpha/(1+\alpha))^{i-1}$

59

ディリクレ過程混合モデル

セミパラメトリックなベイズモデルの構成法
(Ferguson,1973; Walkerら1999)

Dirichlet Process Mixture (DPM) Model

⇒パラメータの事前分布にディリクレ過程事前分布利用

$$G(\theta) = \sum_{i=1}^{\infty} p_i \delta_{\theta_i}, \quad \theta_i \sim G_0$$

Yの確率密度関数は $y \sim f(y|\theta) \quad \theta \sim G(\theta)$

ここで $f(y|\theta)$ を基底分布とする(例:正規分布)

60

その意味

すべての分布は「ある要素数の混合分布」で表現可能

⇒DPMは要素数を事前に仮定しないでよい

(例) 偏回帰係数と切片が異なる複数の

ロジスティック回帰分析モデルの混合モデル

=任意の2値変数の回帰モデルになる

$$p(M=1|X, Z=l) = \frac{1}{1 + \exp(-(a_l X + b_l))}$$

$$p(M=1|X) = \sum_{l=1}^K \pi_l \times \frac{1}{1 + \exp(-(a_l X + b_l))}$$

但しKは未知であり、仮定しない(Ishwaran&James,2001)

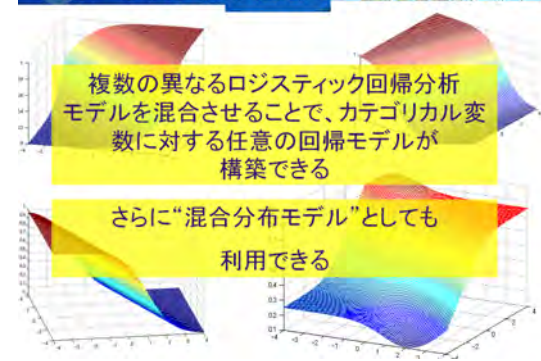
カーネル回帰⇒「サンプルサイズと同じ数の混合モデル」

ディリクレ過程混合⇒サンプルサイズ以下の(未知な数の)混合

5. セミパラメトリックベイズの枠組みによる因果効果の推定

61

例: ロジスティック混合モデル(2次元)



複数の異なるロジスティック回帰分析モデルを混合させることで、カテゴリカル変数に対する任意の回帰モデルが構築できる

さらに“混合分布モデル”としても利用できる

62

DPMモデルに対する推定法

これまで提案されてきた主なものは

- ❖ Escobar & West (1995)
- ❖ MacEachern & Müller(1998)
- ❖ Neal(2000)

⇒MCMCで推定が可能に。

しかしステップ内で積分計算が必要(共役事前分布をうまく選ぶ必要)

- ❖ Ishwaran & James (2001)

Blocked Gibbs Samplerの提案

「有限ディリクレ過程事前分布」を用いることと要素への所属のインディケータ変数を利用

63

有限ディリクレ混合事前分布の利用

Finite-dimensional DPM (Ishwaran&Zarepour,2000)

最大の要素数 $L(\leq N)$ を上限とする

Yの分布関数を $G(\theta) = \sum_{i=1}^L p_i \delta_{\theta_i}, \theta_i \sim G_0$

Yの確率密度関数は $y \sim f(y|\theta), \theta \sim G(\theta)$

但し G_0 :事前分布 $\delta_{\theta_i}(\cdot)$:デルタ関数

$p_i = V_i \prod_{k=1}^{i-1} (1-V_k)$ V_k :Beta分布Beta(1, α)に従う

*Lが大なら通常のディリクレ過程事前分布に等しくなる

64

有限ディリクレ仮定混合モデルの表現

ここで $K = (K_1, \dots, K_N)$ を群別インディケータ θ を要素間で共通の母数

Z を要素によって値が変わる母数とする

(1)インディケータを導入するとYの分布は

$$Y_i | Z, K, \theta \sim p(Y_i | Z_{K_i}, \theta)$$

(2)インディケータは $K | \mathbf{p} \sim \sum_{i=1}^L p_i \delta_i(\cdot)$ 多項分布

(3)比率と母数の事前分布は

$$\mathbf{p}, Z \sim p(\mathbf{p})p(Z|\gamma)$$

$$\theta \sim p(\theta) \quad \tau \sim p(\tau)$$

65

Blocked Gibbs Sampler (Ishwaran & James, 2001)

基本は混合分布、潜在クラスの存在するモデルでの Gibbs samplerと同じ

* Label switchingも発生しない(pの単調性より)

違う点は

(a)各被験者ごとに $\frac{p_i p(Y_i | Z_i, \theta)}{\sum_{k=1}^L p_k p(Y_i | Z_k, \theta)}$ の確率で $K_i = l$ とする

⇒ステップごとに要素数が異なる。後で要素数ごとに分けて事後分布を計算

66

Blocked Gibbs Sampler (Ishwaran & James, 2001)

(b)上記からL個中m要素が存在 $K_j^* (j=1 \dots m)$ 誰も所属しない要素lに対して Z_l は $p(Z|\gamma)$ から発生

存在する要素 K_j^* に対しては

$$p(Z_{K_j^*} | K, \theta, Y) \propto p(Z_{K_j^*} | \gamma) \prod_{i: K_i = K_j^*} p(Y_i | Z_{K_j^*}, \theta)$$

(c)pの完全条件付き事後分布は $p_l = V_l \prod_{k=1}^{l-1} (1-V_k)$ よりVを発生させることで代用

$$V_k \sim \text{Beta}(a_k + M_k, b_k + \sum_{l=k+1}^L M_l)$$

但し M_k は要素kに所属する人数

67

まとめ

ディリクレ過程混合モデルは

- ❖ 複数の要素数を持つモデルの母数を一度に推定する
- ❖ モデル比較の指標の計算が不要
直接「要素数の事後確率」を計算
- ❖ 要素数の候補を事前に特定しなくてよい
- ❖ 結果としてノンパラメトリックなモデリングを可能にしている

Blocked Gibbs samplerは非常に簡単なアルゴリズム(潜在クラスでのGibbs samplerに毛が生えた程度)

68

シミュレーションの例

69

シミュレーションの設定: 混合正規分布

- ❖ 被験者数をN=500, N=1000の2タイプ
- ❖ 要素数は3、混合比率は2種類
 $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}), (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$

❖ 同じデータセットを400個発生としてシミュレーションデータを発生し、それぞれに対して

- 1: 提案手法⇒直接要素数事後分布推定
但しL(=要素数の上限)を10
- 2: 要素数ごとに(上限5)通常のGibbs Samplerで推定 ⇒BICでモデル比較し要素数を決定

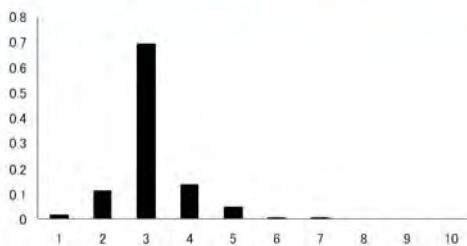
70

結果2(1/3, 1/3, 1/3)

	α	DPM			BIC	
		3	8	20		
N=500	1	1	1	0	20	
	2	21	20	18	95	N=500では 提案法81.25% BIC59.75% 要素数正判定
	3	325	323	320	239	
	4	38	39	44	45	
	5	14	15	15	1	
	6	1	2	3	0	
	7	0	0	0	0	
	8	0	0	0	0	
	9	0	0	0	0	
	10	0	0	0	0	
N=1000	1	0	0	0	4	
	2	13	12	11	22	N=1000では 提案法94.25% BIC84.75% 要素数正判定
	3	377	376	375	339	
	4	10	11	13	34	
	5	0	0	1	1	
	6	0	0	0	0	
	7	0	0	0	0	
	8	0	0	0	0	
	9	0	0	0	0	
	10	0	0	0	0	

71

要素数の事後分布の平均



直接「要素数の事後分布」が導出できる

72

「ランダムでない欠測」に対する
ディリクレ過程混合モデルを
利用した調整と共変量選択

5. セミパラメトリックベイズの枠組みによる因果効果の推定

73

まずはHeckmanのプロビット選択モデルから

3つの方程式を考える

$$y_{i1} = x_i' \beta_1 + u_{i1} \quad y_{i0} = x_i' \beta_0 + u_{i0}$$

$$y_{iz} > 0 \text{ なら } y_{i1} \text{ が観測} \quad y_{iz} \leq 0 \text{ なら } y_{i0} \text{ が観測}$$

* 誤差 u_{i1}, u_{i0} に通常は3変量正規分布を仮定
(この制約は外すことができるが特定の分布仮定は必要)

尤度

$$= \prod_{i: y_{iz} > 0} p(y_{i1}, y_{i0}, y_{iz} | x_i) \prod_{i: y_{iz} \leq 0} p(y_{i1}, y_{i0}, y_{iz} | x_i)$$

$$= \prod_{i: y_{iz} > 0} p(y_{i1}, y_{iz} | x_i) \times \prod_{i: y_{iz} \leq 0} p(y_{i0}, y_{iz} | x_i)$$

y_{i1} と y_{i0} の共分散は消える⇒推定できない

74

なぜ推定できないのか？

モデルの識別性がない
⇒同じデータを様々なパラメータで表現できる
⇒可能な制約条件は？

$$p(y_1, y_0, y_z | x) = p(y_1, y_0 | x) p(y_z | y_1, y_0, x)$$

具体的には

(1) $p(y_z | y_1, y_0, x) = p(y_z | y_0, x)$
(2) 条件付き独立 $p(y_1, y_0, x) = p(y_1 | x) p(y_0 | x)$

のどちらかで良い⇒大幅に仮定緩和
⇒前者のモデルを考える

75

一番簡単な場合: 正規分布モデル

正規分布の場合

$$y_{i1} = \mu_1(x) + \beta_1 y_{i0} + u_{i1}$$

$$y_{iz} = \mu_z(x) + \beta_z y_{i0} + u_{iz}$$

関心があるのは $Cov(y_1, y_0) = \beta_1 V(y_0)$

ここで β_1 は $Cov(y_0, y_z)$ から推定できる
⇒前者のモデルを利用=「条件付き独立」は不要

76

セミパラメトリックモデルの利点

フルモデルは

$$p(y_1, y_0, z, x) = p(y_1, y_0) p(x | y_1, y_0) p(z | y_0, x)$$

関心がある、モデルがある⇒パラメトリック
関心がない⇒ノンパラ⇒ディリクレ過程混合モデルの利用

セミパラメトリックモデル=パラメトリック+ノンパラ
の場合、「パラメトリック」部分のパラメータの収束は
いくつかの条件の下で通常のオーダー($1/\sqrt{N}$)になる
(部分線形モデルではRobinson, 1986より一般的な議論は
Lee, 2009; Otsu, 2010など)
⇒まだセミパラメトリックベイズについての包括的な議論はない
が、上記の因果モデルでも同様の結果が期待できる
* 実際の「ディリクレ過程混合モデルは？」xの一部は連続？

77

従属変数が連続変数の場合

混合回帰分析モデル 但し要素数は未定
連続変数を従属変数とする場合には、通常のディリクレ過程
混合は役に立たない(Dunson, 2007)

【理由】線形回帰分析モデル

$$E(x | y, r = l) = \beta_l' y$$

の混合はただの線形回帰モデル

$$E(x | y) = \left[\sum_{l=1}^{\infty} p_l \beta_l \right]' y$$

* 但しYとXの同時分布の混合によって回帰モデルは
非線形になる(Müllerら1996)
⇒因果モデルではyの周辺分布はパラメトリックのままに
したいので利用できない...

78

Probit Stick-breaking process mixture (PSBPM)モデル

Chung and Dunson(2009, JASA)

重みが説明変数に依存するディリクレ過程混合モデル

$$X \sim \sum_{l=1}^{\infty} \tilde{p}_l(y) f(X | y, \theta_l)$$

$$p_l(y) = \Phi(g_l(y)) \prod_{k=1}^{l-1} (1 - \Phi(g_k(y)))$$

但し Φ は正規累積関数

$$g_l(y) = \alpha_l + f_l(y | \phi), \alpha_l \sim N(\mu, 1)$$

ϕ, μ は超パラメータ, f_l はリンク関数

$$E(x | y) = \left[\sum_{l=1}^{\infty} p_l(y) \beta_l' \right] y \quad \text{はyについて非線形}$$

79

Probit Stick-breaking process を用いたモデリング

フルモデル表現

$$p(y_1, y_0, z, x) = p(y_1, y_0) p(x | y_1, y_0) p(z | y_0, x)$$

関心がある、モデルがある⇒パラメトリック

関心がない⇒ノンパラ⇒Probit Stick-breaking process

$p(y_1, y_0)$: 多変量正規分布や多変量プロビットモデル
* 効用のモデル表現を行う

$$p(x | y_1, y_0) p(z | y_0, x)$$

: 混合要素数が未知の混合多変量正規分布
とプロビットモデル(zの背後に潜在変数仮定)

80

PSBPMでのMCMC

ほとんどIshwaran&JamesのBlocked Gibbs samplerと同じ

変更点

(a) 所属確率がデータに依存する
 $p_i \rightarrow p_i(y)$

(b) リンク関数の母数である ϕ, μ についても
サンプリングが必要

⇒各要素への所属確率が順序プロビットモデルになっ
ている、というような理解をすればよい

81

因果推論でのモデル設定



「ランダムでない欠測」の場合
割り当てモデル $p(z | y_1, y_0, x)$ モデリングが必要
フルモデルは

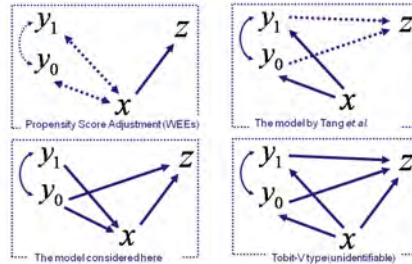
$$p(y_1, y_0, z, x) = p(y_1, y_0) p(x | y_1, y_0) p(z | y_1, y_0, x)$$

関心がある、モデルがある⇒パラメトリック

関心がない⇒Probit stick-breaking過程混合モデルの利用

82

いくつかのモデルの比較



- 1) " $a \rightarrow b$ " represents the regression of b on a, and " $a \leftrightarrow b$ " represents the correlation between a and b
- 2) The solid arrow represents that there exists dependence of variables, and if the model specification is made
- 3) The dashed arrow represents that there exists dependence of variables, but the model specification is not made

83

シミュレーション

潜在的結果変数はそれぞれ1次元で2変量正規分布

$$\begin{pmatrix} y_1 \\ y_0 \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_0 \\ \rho\sigma_1\sigma_0 & \sigma_0^2 \end{pmatrix} \right)$$

割り当てはyAと共変量x(10次元)に依存する

* 識別性のためにyBには依存しないとする

$p(z | y_0, x)$ はyAとxの2次項、交互作用項を含む
ロジスティック回帰モデル

$p(x | y_1, y_0)$ はyAとyBの2次項、交互作用項を含む
回帰分析モデル(誤差は多変量正規分布)

* 但し共変量候補は10個、そのうち6個はXと相関(0.4)
を持つ変数(zとの関連は高いがyとはあまり相関が
ない)。

84

シミュレーション結果

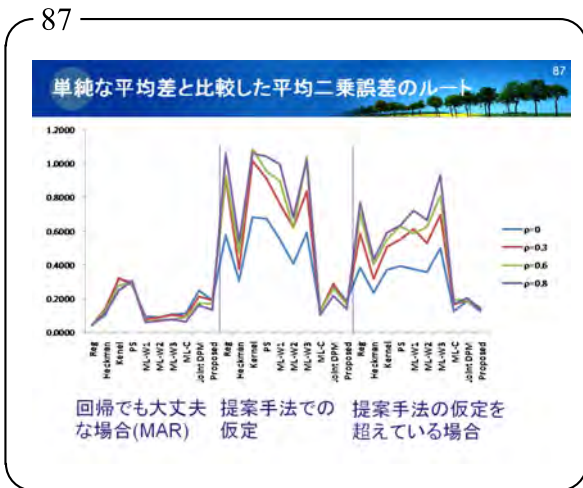
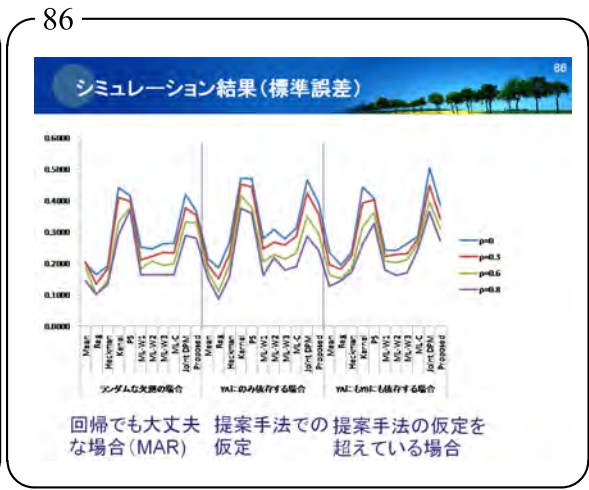
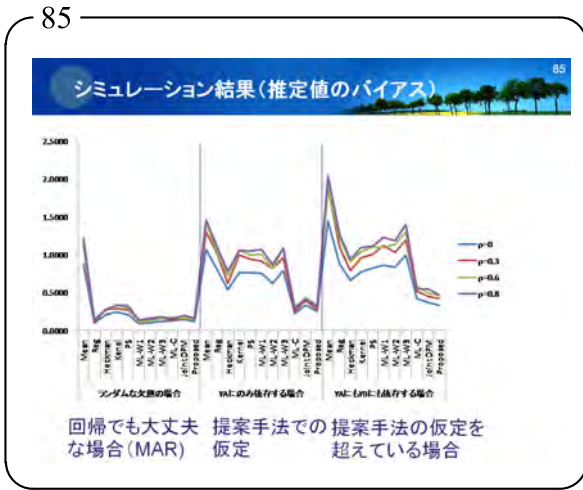
因果効果

$$\mu_1 - \mu_0 \text{ の推定}$$

N=400のサンプルを1000発生し、

- (1) Mean: 観測値の単純な平均値差
- (2) Reg: 単純な回帰モデルによる推定
- (3) Heckman: セレクションモデルによる推定
- (4) Kernel: カーネル回帰による推定
- (5) PS: 傾向スコアマッチング
- (6) ~ (9) ML: 様々な仮定での最尤法
- (10) Joint DPM: 同時分布をディリクレ過程混合モデル
にしたノンパラメトリックベイズ
- (11) 提案手法(PSBPM)

5. セミパラメトリックベイズの枠組みによる因果効果の推定



89

NSW職業訓練プログラム

National Supported Work Demonstration
70年代の米での職業訓練プログラム

- ❖ 職業訓練を受ける群(処置群)と受けない群(対照群)は無作為割り当てされている実験研究
- ❖ 関心は職業訓練の再就職後の賃金への効果
- ❖ 実施前の1974年、1975年と実施後の78年の3時点のデータがある

Lalonde(1986,American Economic Review)ではあえて対照群を"Panel Study of Income Dynamics(PSID)"データと置き換え、無作為で無い形にして共変量調整(Heckmanモデルを用いた推定)を実施

(結果は?)調整しても実験研究での結果と乖離している
⇒共変量調整の問題点(実験の必要性)を主張

90

傾向スコアを用いた調整では?

Dehejia & Wahba(1999,JASA)はLalondeの解析の問題点を指摘

- ❖ Heckmanモデル(またはTobit TypeV)を用いた解析は「あまりに仮定が強い」
- ❖ 利用した共変量が少ない、データの問題など

研究	Lalonde(1986)	Lalonde(1999)
方法	実験研究からの78年での単純平均の差*	78年と75年での処置群と対照群の平均推定量*
推定値(%)	1794	1750
研究	Lalonde(1986)	Dehejia-Wahba(1999)
方法	PSIDを対照群としたときの78年での単純平均の差*	PSIDを対照群としたときのDID推定量*
推定値(%)	-15208	-582
研究	Lalonde(1986)	Dehejia-Wahba(1999)
方法	Heckmanモデルを用いた78年での賃金への因果効果	傾向スコアマッチングを用いた78年での因果効果の推定値
推定値(%)	-667	1691
研究	Abadie	Guzman(2007)
方法	Abadieによる傾向スコアを用いたDID推定量	DIDによる単純平均を用いたDID推定量
推定値(%)	1767902	1748345

その後のいくつかの推定法の比較
* *は共変量を利用していない

91

提案手法との比較

Table 4: Various estimates of causal effect for NSW data

Method	Benchmark	Unadjusted	Reg	PS	Proposed
Estimate	1794	-15205	731	1608	1760
s.e.	833	1154	886	1371	1004
Z value	2.834	-13.176	0.825	1.024	1.753

かなり実験で得られた結果を復元している

Table 5: Various estimates when some of covariates are dropped

Dropped	RE74	RE75	RE74/75	Black-Hispanic
Estimate	1597	1563	1486	1684
s.e.	1007	1009	1013	1005
Z value	1.586	1.549	1.467	1.676

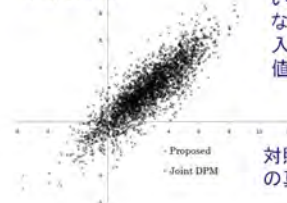
Dropped	Schooling, College	Marital	Marital RE74/75	other than RE74/75
Estimate	1637	1649	1451	1530
s.e.	1006	1007	1023	1032
Z value	1.627	1.638	1.418	1.483

いろいろな共変量をわざと削除しても頑健(結局事前の収入RE74,75があればかなりきれいに推定できるようだ)

92

ベイズ推定の利点

ここで取り上げるまでもないが、予測値がPlug-in estimatorではない(=他の要素の変動を考慮)

 y_1 の事後予測分布の平均

いわゆるfull-nonparametricな方法よりは一部モデルが入っている分、よりよい予測値が得られる

対照群での真の y_1

93

潜在的結果変数のパラメトリックモデルによって

例えばこんなモデルを推定可能
 Y を再就職後の収入、 w を事前の収入
 $Z=1$ (職業訓練群)、 $Z=0$ (給付群)

$$\begin{pmatrix} y_1 \\ y_0 \end{pmatrix} \sim N \left(\begin{pmatrix} \alpha_1 + \beta_1 w \\ \alpha_0 + \beta_0 w \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{10} \\ \sigma_{10} & \sigma_0^2 \end{pmatrix} \right)$$

 β の群間の差

=事前の収入と失業給付の形態の交互作用なども推定できる。

⇒これについてはもう少し解析例を蓄積してから今後紹介したい

94

まとめと議論

社会科学領域(経済学、社会学、マーケティング、心理学など)での因果効果推定のためには

- ❖「反実仮想」に基づく欠測データとしての理解が有用
- ❖共変量がきわめて多い(通常20変数ほどが想定)
⇒モデル誤設定の危険性があるパラメトリック、次元の呪いのあるノンパラでなく、セミパラメトリック解析が有用
- ❖「ランダムな欠測」の場合の共変量調整には共変量選択が非常に重要
- ❖「ランダムでない欠測」状況に対してはベイズセミパラメトリック解析が有用

引用文献

- [1] P. W. Airasian, G. F. Madaus, and T. Kellaghan. *Consequences of Introducing Educational Testing in Northern Ireland, 1973–1977*. Inter-university Consortium for Political and Social Research, Ann Arbor, MI, 1981.
- [2] J. H. Albert. Bayesian estimation of normal ogive response curves using Gibbs sampling. *Journal of Educational Statistics*, Vol. 17, pp. 251–269, 1992.
- [3] T. W. Anderson. Maximum likelihood estimates for the multivariate normal distribution when some of the observations are missing. *Journal of the American Statistical Association*, Vol. 52, pp. 200–203, 1957.
- [4] A. A. Beguin and A. A. W. Glas. MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, Vol. 66, pp. 541–562, 2001.
- [5] A. Birnbaum. Some latent trait models and their use in inferring an examinee’s ability. In F. M. Lord and M. R. Novick, editors, *Statistical Theory of Mental Test Scores*. Addison-Wesley, Reading, MA, 1968.
- [6] E. Bladlow, H. Wainer, and X. Wang. A Bayesian random effects model for testlets. *Psychometrika*, Vol. 64, pp. 153–168, 1999.
- [7] G. E. P. Box and G. C. Tiao. *Bayesian inference in Statistical Analysis*. Addison-Wesley, Reading, MA, 1973.

- [8] D. K. Dey and C. R. Rao, editors. *Bayesian Thinking: Modeling and Computation*, Vol. 25 of *Handbook of Statistics*. North Holland, Amsterdam, 2005.
- [9] G. T. Duncan and M. W. J. Layard. A Monte-Carlo study of asymptotically robust tests for correlation coefficients. *Biometrika*, Vol. 60, pp. 551–558, 1973.
- [10] Y. Fujikoshi, T. Kanda, and N. Tanimura. The growth curve model with an autoregressive covariance structure. *Annals of the Institute of Statistical Mathematics*, Vol. 42, pp. 533–542, 1990.
- [11] A. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, pp. 721–741, 1984.
- [12] A. L. Gross and R. Torres-Quevedo. Estimating correlations with missing data, a Bayesian approach. *Psychometrika*, Vol. 60, pp. 341–354, 1995.
- [13] R. H. Hoyle. *Statistical Strategies for Small Sample Research*. Sage Publications, Thousand Oaks, CA, 1999.
- [14] J. C. Lee. Prediction and estimation of growth curves with special covariance structures. *Journal of the American Statistical Association*, Vol. 83, pp. 432–440, 1988.
- [15] J. C. Lee and C. H. Chang. Bayesian analysis of a growth curve model with a general autoregressive covariance structure. *Scandinavian Journal of Statistics*, Vol. 27, pp. 703–713, 2000.
- [16] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data, 2nd Edition*. Wiley-Interscience, Hoboken, NJ, 2002.
- [17] F. M. Lord and M. R. Novick. *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA, 1968.

-
- [18] K. Pearson. Mathematical contributions to the theory of evolution XI. On the influence of natural selection on the variability and correlation of organ. *Philosophical Transactions of the Royal Society, London, Series A*, Vol. 200, pp. 1–66, 1903.
- [19] D. Rizopoulos. ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, Vol. 17, No. 5, pp. 1–25, 2006.
- [20] P. R. Rosenbaum. Item bundles. *Psychometrika*, Vol. 53, pp. 349–359, 1988.
- [21] H. Swaminathan, R. K. Hambleton, S. G. Sireci, D. Xing, and S. M. Rizavi. Small sample estimation in dichotomous item response models: Effect of priors based on judgmental information on the accuracy of item parameter estimates. *Applied Psychological Measurement*, Vol. 27, pp. 27–51, 2003.
- [22] D. Thissen, L. Steinberg, and J. A. Mooney. Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement*, Vol. 26, No. 3, pp. 247–260, 1989.
- [23] H. Wainer, E. Bladlow, and X. Wang. *Testlet Response Theory and Its Applications*. Cambridge University Press, Cambridge, 2007.
- [24] H. Wainer and G. L. Kiely. Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, Vol. 24, pp. 185–201, 1987.
- [25] W. C. Wang and M. Wilson. The Rasch testlet model. *Applied Psychological Measurement*, Vol. 29, pp. 126–149, 2005.
- [26] B. Zheng. Bayesian estimation of multidimensional bibitem response theory modeling using Gibbs sampling. *Communications in Statistics-Theory and Methods*, Vol. 29, pp. 1405–1417, 2000.
- [27] 天野郁夫. 大学入学者選抜論. IDE 現代の高等教育, Vol. 338, pp. 5–12, 1992.

5. セミパラメトリックベイズの枠組みによる因果効果の推定

- [28] 伊藤圭. 大学入試における総合試験および適性試験の動向. 大学入試研究ジャーナル, Vol. 16, pp. 149–155, 2006.
- [29] 伊庭幸人, 種村正美, 大森裕浩, 和合肇, 佐藤整尚, 高橋明彦. 計算統計 II –マルコフ連鎖モンテカルロ法とその周辺–. 岩波書店, 東京, 2005.
- [30] 岩崎学. 不完全データの統計解析. エコノミスト社, 東京, 2002.
- [31] 岡田謙介, 繁榊算男. 小標本における選抜効果を補正する相関係数の推定について–最尤推定法とベイズ推定法のシミュレーションによる比較. 日本テスト学会誌, Vol. 6, No. 1, pp. 63–74, 2010.
- [32] 近藤治. 多様化する大学入試とその課題. 工学教育, Vol. 57, pp. 10–14, 2009.
- [33] 先崎卓歩. 高大接続の現状課題–「大学全入」時代の学力把握と選抜. 大学と学生, Vol. 61, pp. 48–57, 2008.
- [34] 繁榊算男. ベイズ統計入門. 東京大学出版会, 東京, 1985.
- [35] 晶文社学校案内編集部. 全国大学受験案内 2011 年度用. 晶文社, 東京, 2010.
- [36] 晶文社出版編集部. 全国大学受験案内 '91 年度用. 晶文社, 東京, 1990.
- [37] 晶文社出版編集部. 全国大学受験案内 '96 年度用. 晶文社, 東京, 1995.
- [38] 晶文社出版編集部. 全国大学受験案内 2001 年度用. 晶文社, 東京, 2000.
- [39] 晶文社出版編集部. 全国大学受験案内 2006 年度用. 晶文社, 東京, 2005.
- [40] 滝紀子. 入学者選抜多様化の現状と問題点. IDE 現代の高等教育, Vol. 416, pp. 30–35, 2000.
- [41] 豊田秀樹. 項目反応理論 [事例編] –新しい心理テストの構成法–. 朝倉書店, 東京, 2002.

-
- [42] 豊田秀樹. 項目反応理論 [入門編] –テストと測定の科学–. 朝倉書店, 東京, 2002.
- [43] 中村高康. 大学大衆化時代における入学者選抜に関する実証的研究–選抜方法多様化の社会学的分析. 東京大学大学院教育学研究科紀要, Vol. 37, pp. 77–89, 1997.
- [44] 西井泰彦. 全入時代と私学経営. IDE 現代の高等教育, Vol. 491, pp. 27–35, 2007.
- [45] 野家彰. 大学入学者選抜の現状 (資料). IDE 現代の高等教育, Vol. 416, pp. 63–71, 2000.
- [46] 星野崇宏, 岡田謙介, 前田忠彦. 構造方程式モデリングにおける適合度指標とモデル改善について: 展望とシミュレーション研究による新たな知見. 行動計量学, Vol. 32, pp. 209–235, 2005.
- [47] 星野崇宏. 調査観察データの統計科学. 岩波書店, 東京, 2009.
- [48] 文部省高等教育局大学課大学入試室. 大学入試の多様化と新教育課程入試. 大学と学生, Vol. 368, pp. 28–36, 1996.
- [49] 山村滋. 多様な大学入学者選抜方法の実態分析. 山村滋, 鈴木規夫, 濱中淳子 (編), 学生の学習状況からみる高大接続問題, pp. 81–109. 大学入試センター研究開発部, 2009.
- [50] 大学評価・学位授与機構. 国立大学法人等の中期目標期間に係る教育研究の状況 (平成 16 年度～平成 19 年度), 2008.

独立行政法人大学入試センター 入学者選抜研究機構入試評価部門報告書
「大学入試の標準化、多様化、および精密化」

発行 平成 23 年 8 月 8 日

編集・発行 独立行政法人大学入試センター入学者選抜研究機構
〒153-8501 東京都目黒区駒場 2-19-23
電話：03-3468-3311（代）

印刷 株式会社 コームラ
