

統計数理 Vol. 74, No. 1

(通巻 143 号)

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

目次

特集「統計エキスパート人材育成プロジェクト」

「特集 統計エキスパート人材育成プロジェクト」について

中西 寛子・千野 雅人	1
木構造に基づく確率分布推定を用いた DEA 効率性評価 [原著論文]	
趙 宇	5
長崎県の果物生産における多様なデータを基にした価格予測モデルの構築 [原著論文]	
姜 佳明・趙 宇	21
走行環境が自動運転自動車に与えるヒヤリハット要因の分析 [原著論文]	
田島 友祐・笛田 薫・三浦 雄作・三樹 孝博	31
土壌化学組成と母岩・気候条件の関係：主成分分析と線形判別分析による検討 [原著論文]	
張 天逸・中西 寛子	47
脳波位相と正答確率の関数推定に向けたロジット変換を用いたガウス過程回帰 [原著論文]	
小野島 隆之・神保 雅一	61
中高生の「早期 U ターン意向」等に関する分析	
—滋賀県長浜市におけるケーススタディ— [原著論文]	
松原 悠・堀 兼大朗・伊達 平和	81
神経画像解析における統計学的課題とアプローチ：確率場理論 [総合報告]	
内田 航・青木 茂樹・岩崎 学	95
講義形式における統計の入門科目の授業実践とその考察 [研究ノート]	
岡本 雅子・瀧川 一学・中西 寛子	107
学校外教育の累積的效果：利用パターンとタイミングが教育達成に与える影響 [研究ノート]	
真田 英毅	121
共通オッズ比とランク累積分布関数によるランキングデータ解析： 世界大学ランキングを事例に [研究ノート]	
樊 怡舟	135

2026 年 6 月

大学共同利用機関法人 情報・システム研究機構 統計数理研究所

〒 190-8562 東京都立川市緑町 10-3 電話 050-5533-8500(代)

本号の内容はすべて <https://www.ism.ac.jp/editsec/toukei/> からダウンロードできます

ISSN 2760-2125

統
計
数
理

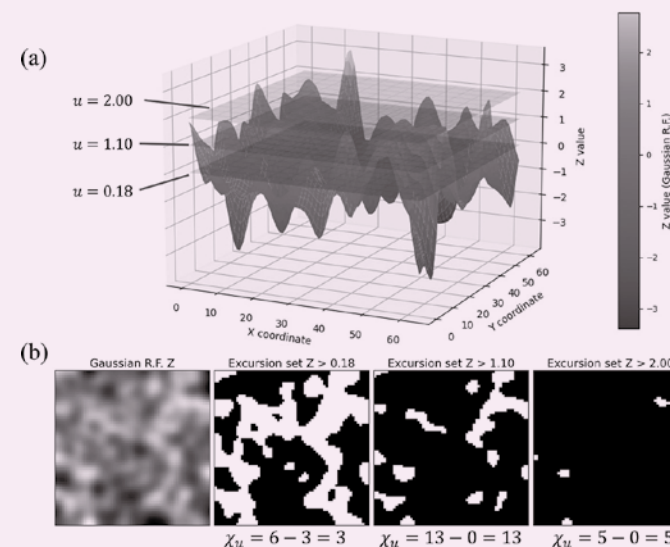
PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

第
74
卷
第
1
号

2026

統計数理 Vol. 74, No. 1

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS



統計数理研究所

統計数理

(年2回発行)

編集委員長 野間 久史

編集委員 庄 建倉

Figueira Lourenço, Bruno

特集担当編集委員 千野 雅人

中西 寛子

編集室

池田 広樹

川合 純華

長嶋 昭子

「統計数理」は、統計数理研究所の研究成果を掲載する機関誌「彙報」として1953年に創刊され、1985年に誌名を現在の「統計数理」に改めました。2025年からはオンラインジャーナルとして新たな形での発行を開始しています。本ジャーナルは、統計科学全般に関する論文を広く受け付け、統計科学の深化と発展、さらに統計科学を通じた社会への貢献を目指しています。

投稿を受け付けるのは、次の6種です。

- a. 原著論文
- b. 総合報告
- c. 研究ノート
- d. 研究詳解
- e. 統計ソフトウェア
- f. 研究資料

投稿された原稿は、編集委員会が選定・依頼した査読者の審査を経て、掲載の可否を決定します。投稿規程、執筆要項は、本誌最終頁をご参照ください。

また、上記以外にも統計科学に関して編集委員会が重要と認める内容について、編集委員会が原稿作成を依頼することがあります。

その他、「統計数理」に関するお問い合わせは、各編集委員にお願いします。

All communications relating to this publication should be addressed to associate editors of the Proceedings.

大学共同利用機関法人 情報・システム研究機構

統計数理研究所

〒190-8562 東京都立川市緑町10-3 電話 050-5533-8500(代)

<https://www.ism.ac.jp/>

© The Institute of Statistical Mathematics 2026

印刷：笹氣出版印刷株式会社

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

Vol. 74, No. 1

Contents

Special Topic : Project for Training Experts in Statistical Sciences

On the Special Topic “Project for Training Experts in Statistical Sciences” Hiroko NAKANISHI and Masato CHINO	1
Tree-based Distribution Estimation for DEA Efficiency Evaluation Yu ZHAO	5
Construction of Price Prediction Models Based on Diverse Data for Fruit Price in Nagasaki Prefecture Jiaming JIANG and Yu ZHAO	21
Analysis of Near-miss Factors Influenced by Driving Environment in Autonomous Vehicles Yusuke TAJIMA, Kaoru FUEDA, Yusaku MIURA and Takahiro MIKI	31
Geochemical Signatures of Soils in Relation to Parent Rock and Climatic Conditions: A Principal Component and Linear Discriminant Approach Tenichi CHO and Hiroko NAKANISHI	47
Gaussian Process Regression with a Logit Link to Estimate the Relationship between Neural Oscillatory Phase and the Probability of a Correct Response Takayuki ONOJIMA and Masakazu JIMBO	61
Analysis of “Early U-turn Intentions” among Junior High and High School Students: A Case Study in Nagahama City, Shiga Prefecture, Japan Yu MATSUBARA, Kentaro HORI and Heiwa DATE	81
Challenges and Advances in Neuroimaging Statistical Analysis : Random Field Theory Wataru UCHIDA, Shigeki AOKI and Manabu IWASAKI	95
A Practice and Consideration of Introductory Statistics in a Lecture Style Masako OKAMOTO, Ichigaku TAKIGAWA and Hiroko NAKANISHI	107
Cumulative Effects of Shadow Education: The Impact of Usage Patterns and Timing on Educational Attainment Teruki SANADA	121
A Method for Analyzing Hierarchical Ranking Data Using Common Odds Ratio and Cumulative Distribution Function of Rank: Application to World Universities Rankings Yizhou FAN	135

June, 2026

Research Organization of Information and Systems

The Institute of Statistical Mathematics

10-3 Midori-cho, Tachikawa, Tokyo 190-8562, JAPAN

表紙の図は本誌 100 ページを参照

「特集 統計エキスパート人材育成プロジェクト」 について

中西 寛子¹・千野 雅人² (オーガナイザー)

最近まで、日本の大学には、統計やデータサイエンスを冠する学部・学科は総合研究大学院大学以外には存在しなかったが、これは世界でも珍しいことであった。欧米はもちろん、アジアや中東など世界の多くの国々の多くの大学には、以前から長きにわたり、統計学部・学科が設置されていたのである。

例えば、米国では、職業としての「統計家」の人気の高いことを反映して、大学での統計学科の人気の高く、その拡大も著しい。米国統計学会の会員誌「AMSTAT NEWS 2022年12月号」の特集は、「力強い成長が続く統計学の学位」であった。これによれば、2021年に米国で、統計学「学士」を授与する大学の数は前年から17校増加して176校、統計学「修士」を授与する大学の数は3校増加して162校となっており、いずれも、その数の多さもさることながら、着実に増加しているのである。

これらの大学の統計学科・統計学研究科の定員の増加は、更に著しい。2021年の1年間に米国で授与された学位の数は、統計学「学士」は前年から7%増加して5,340(うち49は生物統計学)、統計学「修士」は前年から5%増加して5,128(うち917は生物統計学)となっており、特に2010年以降の統計学「学士」の増加は、驚くほどに際立っている。これらの学位取得者の多くが大学院修士課程に進み、その後の修士学位の増加につながっていくものと思われる。

なお、2020年ころから、「データサイエンス」や「データ分析」の学位を授与する大学が現れ始め、その後は、それらの大学の数や学位の数が、劇的に増加している。

ちなみに、米国の政府統計の表章に利用される「米国標準職業分類」には、1980年代以前から、独立した小分類の職業として「15-2040 統計家」(Statisticians)が存在している。また、直近の2018年職業分類改定により、小分類「15-2050 データサイエンティスト」(Data Scientists)が新設された。世界各国の政府統計局が参照する「国際標準職業分類」(ISCO-08)にも、小分類「212 数学家, 保険数理士, 統計家」(Mathematicians, Actuaries and Statisticians)が存在するので、「統計家」という職業の存在は、グローバル・スタンダードと言えるようである。

このような中で、「日本標準職業分類」には、「B051 自然科学系研究者」のような大くくりの小分類しか存在せず、「統計家」のようなプロフェッショナルな職業分類は存在しない。これは、プロフェッショナルな職業を尊重しない印象がある日本の雇用の実情を反映したものであろう。

しかし、日本でも2017年以降は、毎年、データサイエンス系の学部・学科が全国の大学に新設されるようになった。最近では、データサイエンス系学部・学科の新設ラッシュとも言える状況となっており、民間の大学進学情報サイトの一覧情報等を基に「データ」を冠する学科を設置する大学の数を数えると、2025年には40大学に達している。

¹ 統計数理研究所：〒190-8562 東京都立川市緑町10-3; nmf.nakanishi@nifty.com

² 統計数理研究所：〒190-8562 東京都立川市緑町10-3; chino@ism.ac.jp

一方、それまで統計学科のなかった日本の大学では、データサイエンスの基盤となる統計学を体系的に的確に教えることのできる専門教員は極めて少なく、また、急激にそのニーズが高まったからといって、大学教員レベルの人材の育成が一朝一夕に進むわけではない。統計教員の不足は、従来から政府の戦略方針などで指摘されていた問題であったが、このような最近の環境変化により、更に大きな問題として顕在化したと言える。

このような中で、統計数理研究所が中核となり、文部科学省補助事業「統計エキスパート人材育成プロジェクト」が2021年4月に開始された。このプロジェクトは、我が国では統計系の学科を有する大学が米国などの他国に比べて極めて少なく、高度な統計学の専門知識を身に付ける場が非常に少ないという現状認識の下に、高度な統計学のスキルを有する人材の育成と統計人材育成エコシステム(育成された人材が指導者となり持続的に人材育成が進展するシステム)の構築を図ろうとするものである。具体的な目標は、①中核機関のシニア教員の教育指導により、事業期間5年間で少なくとも約30名の大学統計教員を育成すること、②育成された大学統計教員が、所属する大学等において更に人材を育成することにより、10年間で約500名の統計エキスパートを育成することである。

統計数理研究所では、2025年度までに、全国の大学等の各種学術分野の若手研究者を対象に大学統計教員育成研修を3期にわたって実施した。3期にわたる研修の修了生は、目標を大きく超える38名に達した。これらの研修修了生は、大学統計教員としてそれぞれ所属する大学等において大学院生を教育指導し、統計を駆使して産官学で活躍できる統計エキスパートに育成する計画である。

実際の研修を行っている統計数理研究所大学統計教員育成センター研修部には、9名のシニア教員と先に述べた研修生が所属している。研修は2年間(半年を1クールとした4クール)である。具体的な目標は、1)大学院修士相当の学生に対して、「基本的な統計学の講義」、「発展的な統計学の講義」、2種の「専門分野と統計が融合した講義」が各15コマできること、2)専門分野において、統計学に基づく研究指導ができること、3)統計学の知識を十分に活かし自身の研究を発展させることである。

この目標のために、3種類の研修内容(全体研修、グループ研修、シニア教員の講義)が準備されている。「全体研修」では、2年間で8回の模擬講義を行うことになる。他の研修生の模擬講義にも参加することが義務付けられているため、2年間で約100回の模擬講義を聴講することになる。この経験は他にはないもので、本研修で最も大事にしていることである。「グループ研修」では、シニア教員1名に対し研修生2名からなるグループにおいて個人指導を受けることになる。各自の弱点を明らかにし、それを克服できるようシニア教員は指導を行っている。「シニア教員の講義」は2年間で10~12あり、各教員の専門性に準じた統計学の講義が聴講できる。

このバランスの取れた3種類の研修内容を経験するとともに、各自の研究を充実させるため、シニア教員や他の研修生との共同研究が展開される。毎年、約10件の共同研究が、論文、学会の口頭発表、ディスカッションペーパーなどで公開され、そのいくつかは海外において発表されている。他にも教材開発やコンサルテーションの経験もできるようになっており、様々な力をつけて研修を修了する。

今回の特集は、このようにして育成された研修生・修了生による研修成果の報告とも言えるものである。これらの若手研究者の専門分野は多岐にわたるため、投稿の内容も、土壤化学組成、スマート農業、脳波位相、自動運転など、多岐にわたっていて興味深い。

このプロジェクトの推進により、統計教員の不足という重い課題が確実に解消される方向に進むことが期待される。このことは、統計コミュニティにおいても高く評価され、統計数理研究所は、2023年度に「日本統計学会統計活動賞」を受賞した。また、プロジェクトの中間評価で

も、文部科学省から最高位となる総合評価「S」（当初の計画を超える進展があり、事業を継続することでさらに優れた成果が期待される）を頂いた。

しかし、毎年約 5,000 人の「統計学修士」を輩出するなど、多数の高度統計人材を継続的に育成している米国に比べると、まだまだ道半ばである。人材育成の取組は、5 年で完結するような短期的なものではない。最近のデータサイエンス系学部・学科の新設ラッシュや、数理・データサイエンス・AI 教育プログラム認定制度の創設などの現状にかんがみれば、統計学教員の不足はさらに深刻になったとも考えられ、長期にわたる継続的な取組を一層充実させることが必要である。今後も、統計人材育成の歯車が十分に円滑に回るようになるまで、引き続き取り組んでいきたい。

最後になるが、本特集の編集に当たり、寄稿していただいた研修生・修了生、査読者の方々、編集室の皆さま、また、シニア教員として研修生への教育指導のかたわら編集委員を務めていただいた折笠秀樹先生、国友直人先生、神保雅一先生に、お礼を申し上げたい。これを機に、統計科学分野の人材育成に対する理解が進み、その充実に少しでも貢献することができれば、幸いである。

参 考 文 献

- American Statistical Association (2022). AMSTATNEWS Dec. 2022 ISSUE #545, https://magazine.amstat.org/wp-content/uploads/2025/06/12DECEMBER_AMSTAT.pdf (最終アクセス日 2026 年 2 月 2 日).
- International Labour Organization (2008). International Standard Classification of Occupations (ISCO), <https://ilostat.ilo.org/methods/concepts-and-definitions/classification-occupation/> (最終アクセス日 2026 年 2 月 2 日).
- 総務省政策統括官(統計制度担当)(2009). 日本標準職業分類(2009 年 12 月告示), https://www.soumu.go.jp/toukei_toukatsu/index/seido/shokgyou/index.htm (最終アクセス日 2026 年 2 月 2 日).
- U.S. Bureau of Labour Statistics (2018). U.S. Occupational classification, https://www.bls.gov/soc/2018/major_groups.htm (最終アクセス日 2026 年 2 月 2 日).

木構造に基づく確率分布推定を用いた DEA 効率性評価

趙宇†

(受付 2025 年 5 月 29 日; 改訂 8 月 26 日; 採択 9 月 2 日)

要 旨

標準的な Data Envelopment Analysis (DEA) モデルは決定論的であり、観測データに測定誤差や外生的ノイズが存在しないことを前提としている。しかし、実データには確率の変動が不可避であり、これを無視すると効率性推定に深刻な誤差が生じる可能性がある。既存のブートストラップ法や回帰型アプローチは統計的推論を部分的に導入しているものの、データの不確実性を直接的に扱う枠組みは確立されていない。本研究では、DEA における効率性および信頼区間を推定するため、密度木と密度フォレストに基づく新たなサンプリング手法を提案する。提案手法は、観測データをクラスタに自動分類し、各クラスタ内でガウス型エントロピー関数に基づく情報利得を最大化することで局所的な確率構造をモデル化する。さらに、アンサンブル学習により推定の安定性と頑健性を確保し、観測データのばらつきやノイズを反映した擬似データを生成する。これにより、DEA における統計的不確実性を柔軟かつ実証的に評価できる。電力企業データによる実証分析では、観測データに基づく DEA 効率値と、提案手法による推定効率値との間で Kolmogorov-Smirnov 検定を行った結果、統計的に有意な分布差は認められなかった。この結果は、提案手法が観測データの分布的特性を適切に再現し、測定誤差やノイズを考慮した効率性評価を可能にする有効なアプローチであることを示唆している。

キーワード：密度木、密度フォレスト、サンプリング、信頼区間推定、不確実性。

1. はじめに

Data Envelopment Analysis (DEA) は、効率性を評価するためのノンパラメトリック手法であり、経済、医療、教育、エネルギーなど多様な分野でパフォーマンス評価や政策立案に広く用いられてきた。DEA における効率性とは、一般に、多入力・多出力を有する意思決定単位 (Decision Making Unit; DMU) において、最小限の入力で最大限の出力を得るという観点から、入力が出力へと変換される生産過程の効率を指す。しかし、従来の DEA モデルは入出力データに誤差が存在しないことを前提としており、実データに不可避な測定誤差や外生的ノイズなどの確率の変動を考慮していない。これらは効率性推定の精度に大きな影響を及ぼすため、近年では DEA に統計的推論を導入し、不確実性を考慮した効率性評価の研究が進展している。

Simar and Wilson (1998) は、DEA におけるブートストラップ手法を初めて提案し、効率性の信頼区間推定や仮説検定を可能にした。その後、Kneip et al. (2008) が漸近理論を統合して統計的基盤を強化し、Moradi-Motlagh and Emrouznejad (2022) は過去 20 年間のブートストラップ

† 東京理科大学 経営学部; 〒 102-0071 東京都千代田区富士見 1-11-2; yu.zhao@rs.tus.ac.jp

手法の理論的發展を整理するとともに、ロバスト確率的手法 (Aragon et al., 2005; Cazals et al., 2002) や条件付き効率性推定 (Daraio and Simar, 2014; Daraio et al., 2020) の重要性を指摘した。また、Dia et al. (2022), Kerstens et al. (2022), Walheer (2022), Ngo and Tsui (2022), Michali et al. (2023), Lin and Lu (2024), Kang et al. (2024) など近年の主要な發展に位置づけられる。これらの中で、Simar and Wilson (1998) は最も広く用いられているが、本稿の 2.2 節で示すとおり、その手法はデータに内在する誤差構造や異質性を十分に捉えられないという限界がある。このため、観測された入出力ベクトルを確率変数として扱い、その確率構造に基づいた柔軟なサンプリングおよび信頼区間推定手法の開発が求められている。

別の流れとして、Kuosmanen and Kortelainen (2012) による Stochastic Nonparametric Envelopment of Data (StoNED) がある。StoNED は DEA と (Aigner et al., 1977, SFA) を統合し、非効率性とノイズを合成した誤差項をセミノンパラメトリックに推定可能とした。ここでいう非効率性は、入力の上乗れや出力の不足といった改善余地を示す尺度であり、効率性と表裏一体の概念である。これにより確率的変動を考慮した効率性推定が可能となったが、Mergoni et al. (2025) が指摘するように、計算負荷や次元の呪いといった課題が残る。

さらに近年、機械学習の進展も DEA 研究に影響を与えている。Zhu et al. (2021), Aparicio et al. (2023), Boubaker et al. (2025), Esteve et al. (2023), Shi and Zhao (2024) は、DEA と機械学習を統合した効率性測定・予測手法を提案しており、また Valero-Carreras et al. (2021), Aparicio and Esteve (2023), Guillen et al. (2023), Mergoni et al. (2025) はフロンティア推定の精度向上を目指している。しかし、これらの研究はいずれも推定精度やフロンティア構造の改善に主眼を置いており、効率性の信頼区間推定に関する議論はほとんど見られない。

以上を踏まえ、本研究では確率的変動を適切に考慮した DEA 効率性評価手法の開発を目的とし、木構造(密度木および密度フォレスト)に基づく新たなサンプリングアルゴリズムを提案する。本手法は、観測データから高密度領域を自動抽出し、ガウス型エントロピー関数 (Criminisi and Shotton, 2013) による情報利得最大化で局所的確率構造をモデル化する。さらに、アンサンブル学習により推定の安定性と頑健性を高め、入出力データの確率的特性を反映した疑似データ生成を実現する。

本論文の構成は以下のとおりである。第 2 章で DEA の概要と不現実環境下での効率性評価の課題を述べ、第 3 章で密度木に基づくサンプリングアルゴリズムを提案する。第 4 章で効率性の信頼区間構築手法を示し、第 5 章で電力企業データによる実証分析を行い、第 6 章で結論と今後の課題を示す。

2. DEA の概要

2.1 決定論的アプローチとしての DEA

少ない入力から多くの出力を得ようとする多入力・多出力を有する DMU を対象とする。入力および出力のベクトルを $v \in \mathbb{R}_+^d := (x, y) \in \mathbb{R}_+^m \times \mathbb{R}_+^s$ と定義する。ただし、 x は入力、 y は出力を表す。すべての v によって構成される集合 P は、実現可能な入出力の組み合わせを表し、生産可能集合と呼ばれる。この集合 P の境界は生産フロンティアと呼ばれ、その上に位置する DMU は効率的であるとされる。一般に、効率的でない DMU を効率的にするためには、その入出力水準を生産フロンティア上の水準まで改善する必要がある。改善の方向には複数の選択肢が存在するが、本稿では実務で広く用いられている入力指向型モデルに基づく枠組みを採用する¹⁾。以下に、その定式化を示す。

生産可能集合 P に対して、入力指向型の効率性を δ としたとき、効率値 $\hat{\delta}$ は次の線形計画問題を解くことで得られる (Banker et al., 1984) :

$$(2.1) \quad \hat{\delta} := \min\{\delta > 0 \mid (\delta \mathbf{x}, \mathbf{y}) \in P\}.$$

ここで δ は、所与の出力 \mathbf{y} を維持したまま、入力 \mathbf{x} をどれだけ縮小できるかを表す尺度であり、効率的な DMU では $\hat{\delta} = 1$ 、非効率的な DMU では $\hat{\delta} < 1$ となる。

今、 n 個の DMU が観測されたとする。DEA では、複数の観測された DMU の入出力データに基づき、それらを包含する経験的生産可能集合 P^{DEA} を構築する：

$$(2.2) \quad P^{DEA} := \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}_+^{m+s} \mid \sum_{h=1}^n \lambda_h \mathbf{x}_h \leq \mathbf{x}, \sum_{h=1}^n \lambda_h \mathbf{y}_h \geq \mathbf{y}, \sum_{h=1}^n \lambda_h = 1, \lambda \geq \mathbf{0} \right\}.$$

ここで、 $\mathbf{x}_h \in \mathbb{R}_+^m$ および $\mathbf{y}_h \in \mathbb{R}_+^s$ は、第 h 番目の DMU における入力および出力ベクトルを表し、 $\lambda = (\lambda_1, \dots, \lambda_n)^\top$ は n 次元の非負ベクトルである。制約 $\sum_{h=1}^n \lambda_h = 1$ を課すことで、 P^{DEA} は観測された DMU の凸包およびその凸包より大きい入力・小さい出力をもつ点の集合となり、このとき規模に関する収穫は可変 (Variable Returns to Scale, VRS) となる (Banker, 1984)²⁾。

評価対象となる DMU _{o} , $o \in \{1, 2, \dots, n\}$ について、 P^{DEA} に基づき、DEA における入力指向型の効率性は、次の式によって定式化できる：

$$(2.3) \quad \hat{\delta}_o = \min\{\delta > 0 \mid (\delta \mathbf{x}_o, \mathbf{y}_o) \in P^{DEA}\}.$$

この問題は BCC (Banker, Charnes and Cooper) モデル (Banker, 1984) と呼ばれ、各 DMU ごとに個別に解く必要がある。以降、DEA における記法の慣習に従い、混乱の生じるおそれがある場合には、DMU の番号を表す添字を省略して記述する。

2.2 不確実環境下における DEA 効率性評価の課題

従来、DEA に基づく効率性に対して統計的推論を導入する手法として、Simar and Wilson (1998) によるブートストラップ法 (以下、Simar–Wilson 法) が広く採用されてきた。Simar–Wilson 法では、観測された入力および出力に測定誤差や外生的ノイズが含まれる可能性は考慮されておらず、データのばらつきはすべて各 DMU の効率性の違いに起因すると仮定されている。この仮定のもとでは、経験的生産可能集合 P^{DEA} は真の生産可能集合 P の部分集合 ($P^{DEA} \subset P$) となるため、効率値 $\hat{\delta}_o$ は系統的に上方バイアスされることが知られている (Simar and Wilson, 2004)。すなわち、有限サンプルに基づく DEA フロンティアは真のフロンティアより内側に位置し、結果として効率値が過大推定される傾向が生じる。Simar–Wilson 法はブートストラップによるバイアス補正を試みるが、データ生成過程に内在する測定誤差や外生的ノイズを直接モデル化していないため、これらの影響を十分に補正できるわけではない。さらに、Simar–Wilson 法はすべての DMU が同一の効率性分布から独立に抽出されているとする同質性の仮定を前提としている。この仮定のもとでは、効率性のばらつきは主として確率的な変動に起因するものとみなされるが、実際には技術水準や経営環境の異なる集団が混在している場合が多く、一律の効率性分布を仮定することは現実的ではない。Olesen and Petersen (2016) はこの点を批判し、Simar–Wilson 法では効率性のばらつきを全体的な傾向として扱うため、異常値や特異な DMU の構造的特徴が適切に捉えられず、不公平な評価を招く可能性がある指摘している。

Simar–Wilson 法は、DEA に統計的推論を導入する重要な突破口を提供した一方で、観測データに内在する誤差構造や異質性を十分に捉えられないという限界を有する。これらの課題を克服するためには、観測された入出力ベクトルを確率変数として扱い、その確率構造に基づいたより柔軟なサンプリングおよび信頼区間の推定手法の開発が必要である。

3. 木構造に基づくサンプリング

3.1 準備：密度木(Density Tree)

入出力ベクトル $\mathbf{v} \in \mathbb{R}_+^d$ を考える．ここで， $\mathcal{S}_0 := \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ は n 個の DMU(データ点)の集合である．本研究で用いる密度木 (Criminisi and Shotton, 2013) は，頂点(ノード)とそれらを結ぶ辺から構成される階層的な二分木であり，各ノード $j \in \{0, 1, \dots, J\}$ にはデータ集合 $\mathcal{S}_j \subseteq \mathcal{S}_0$ が対応する．

各ノード j では，分割関数 $h_j(\mathbf{v}, \boldsymbol{\theta}) \in \{0, 1\}$ に基づき，データ \mathcal{S}_j を左右に分割する．ここで分割パラメータ $\boldsymbol{\theta} := \{d, t\}$ は， $d \in \{1, \dots, m, m+1, \dots, m+s\}$ (分割方向) および閾値 $t \in \mathbb{R}$ から構成され，定義域は $\mathcal{T} = \{(d, t) \mid d \in \{1, \dots, m, m+1, \dots, m+s\}, t \in \mathbb{R}\}$ である．

ノード j に到達したデータ \mathcal{S}_j は次のように二分分割される：

$$(3.1) \quad \mathcal{S}_j^L = \{\mathbf{v} \in \mathcal{S}_j \mid h_j(\mathbf{v}, \boldsymbol{\theta}) = 0\},$$

$$(3.2) \quad \mathcal{S}_j^R = \{\mathbf{v} \in \mathcal{S}_j \mid h_j(\mathbf{v}, \boldsymbol{\theta}) = 1\}.$$

分割パラメータ $\boldsymbol{\theta}$ は，エントロピー関数に基づく情報利得を最大化するように逐次的に決定される．情報利得 $I(\mathcal{S}_j, \boldsymbol{\theta})$ は以下で定義される：

$$(3.3) \quad I(\mathcal{S}_j, \boldsymbol{\theta}) = H(\mathcal{S}_j) - \sum_{i \in \{L, R\}} \frac{|\mathcal{S}_j^i|}{|\mathcal{S}_j|} H(\mathcal{S}_j^i),$$

ここで， $|\cdot|$ は集合の濃度を表し， $H(\cdot)$ はクラスタリング目的に用いるエントロピー関数である．

最適な分割パラメータ $\boldsymbol{\theta}^*$ は次式で与えられる：

$$(3.4) \quad \boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \mathcal{T}} I(\mathcal{S}_j, \boldsymbol{\theta}).$$

二分木による分割構造を図 1 に示す．情報利得 $I(\mathcal{S}_j, \boldsymbol{\theta})$ が大きいほど，分割により親ノードの不確実性 $H(\mathcal{S}_j)$ が大幅に減少したことを意味し，その分割がデータの内部構造を効果的に捉えていると解釈できる．したがって，密度木は情報利得を最大化する分割を逐次的に適用することで，階層的かつ確率的に意味のあるクラスタ構造を形成できる．

なお，本研究では Criminisi and Shotton (2013) に倣い，各クラスタ(ノード)をガウス分布で近似できると仮定し，エントロピー $H(\cdot)$ にはガウス分布に基づくエントロピーを用いる．すなわち， \mathcal{S}_j に属する各データ \mathbf{v} を多変量正規分布に従う確率変数 \mathbf{V} の実現値とみなす：

$$(3.5) \quad \mathbf{V} \sim \mathcal{N}_d(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j).$$

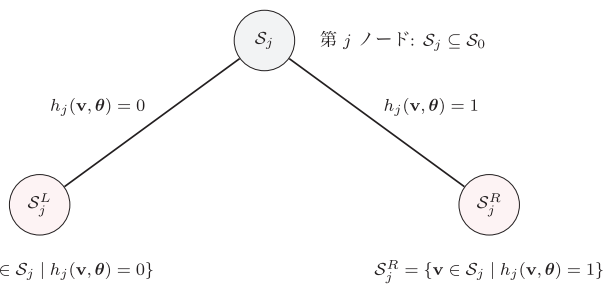


図 1. 二分木による分割構造：ノード \mathcal{S}_j は分割関数 $h_j(\mathbf{v}, \boldsymbol{\theta})$ に基づき左右に分割される．

ここで、 $\boldsymbol{\mu}_j \in \mathbb{R}_+^d$ は平均ベクトル、 $\boldsymbol{\Sigma}_j$ は正定値対称共分散行列である。対応する確率密度関数は以下の通りである：

$$(3.6) \quad p(\boldsymbol{v} | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \frac{1}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma}_j)^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{v} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{v} - \boldsymbol{\mu}_j)\right).$$

このときのエントロピーは次式により与えられる (Cover and Thomas, 2012, p.256)：

$$(3.7) \quad H(\mathcal{S}_j) = \frac{1}{2} \log((2\pi e)^d \det(\boldsymbol{\Sigma}_j)).$$

従って、式(3.3)は次のように書き換えられる：

$$(3.8) \quad I(\mathcal{S}_j, \boldsymbol{\theta}) = \log(\det(\boldsymbol{\Sigma}_j)) - \sum_{i \in \{L, R\}} \frac{|\mathcal{S}_j^i|}{|\mathcal{S}_j|} \log(\det(\boldsymbol{\Sigma}_j^i)).$$

この情報利得を最大化する分割パラメータを逐次的に選定することで、 S_0 を複数のクラスタに分割できる。各クラスタは密度木の葉ノードに対応し、クラスタサイズによる重み付けにより、単点クラスタの生成を抑制する設計となっている。

葉ノードのインデックスを $l(\boldsymbol{v}) : \mathbb{R}_+^d \rightarrow \mathbb{N}$ とすると、密度木全体における混合分布は次式で表される (Criminisi and Shotton, 2013)：

$$(3.9) \quad p(\boldsymbol{v}) = \sum_{l(\boldsymbol{v})} \frac{|\mathcal{S}_{l(\boldsymbol{v})}|}{|\mathcal{S}_0|} p(\boldsymbol{v} | \boldsymbol{\mu}_{l(\boldsymbol{v})}, \boldsymbol{\Sigma}_{l(\boldsymbol{v})}).$$

ここで $\sum_{l(\boldsymbol{v})} \frac{|\mathcal{S}_{l(\boldsymbol{v})}|}{|\mathcal{S}_0|} = 1$ を満たす。これはガウス混合モデルとして解釈できる。

図 2 は、密度木を用いた数値実験の結果を示している。ここでは、 $X \sim \text{Uniform}(0, 1)$ および $U \sim \text{Exp}(1/\mu)$, $\mu = 1/5$ に従う確率変数に基づき、関数 $Y = X^{0.5} \exp(-U)$ により生成された 200 個の二次元データを対象としている。深さ 0 から 5 まで密度木による分割を適用した結果、深さ 0 では全データを 2 つのクラスタに分割し、各クラスタにおける情報利得は 0.78 で

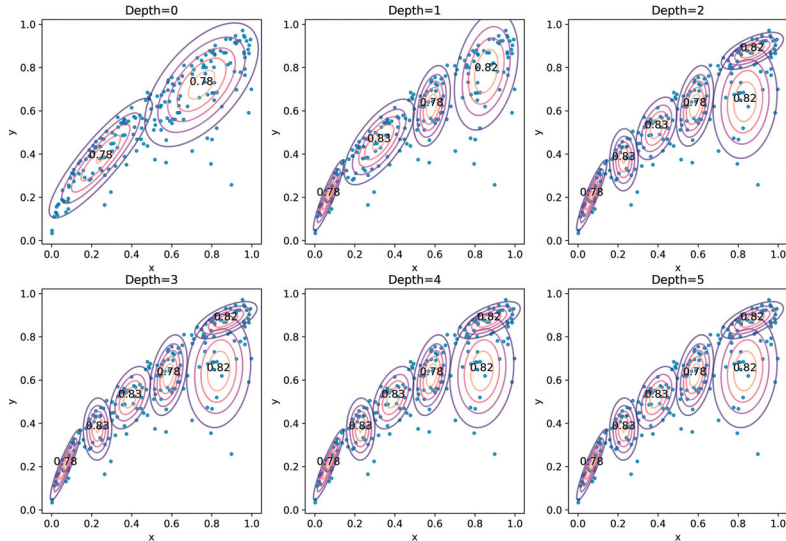


図 2. 密度木によるクラスタリング。

あった。深さが増すごとにクラスタ数は増加し、情報利得の更新が観察されるが、深さ 2 以降は分割構造が安定し、情報利得の増加も停滞する。この結果より、密度木は高密度領域に中心を持つクラスタを形成する傾向があることが確認できる。

なお、初期仮定として $V \sim \mathcal{N}_d(\mu_j, \Sigma_j)$ を置いているが、図 2 の例に見られるように、全体分布がガウス分布から逸脱していても密度木は有効に機能する。これは分割基準の設計がガウス分布を前提としており、各クラスタが局所的にガウス近似されることに起因する。実際、この生成モデルにおいては、 $X = x$ の条件下で Y はスケールされたベータ分布に従い、 $Y | X = x \sim x^{0.5}Z$, $Z \sim \text{Beta}(5, 1)$ が成り立つことは明らかである。

3.2 木構造に基づくサンプリングアルゴリズム

本節では、Simar–Wilson 法のように各 DMU に同一の効率性分布を仮定することなく、観測データに基づき入出力ベクトルの擬似データを生成するための木構造ベースのサンプリング手法を提案する。本手法は、密度木およびそのアンサンプルである密度フォレストに基づき、局所的な確率分布を推定し、これを利用して擬似データを生成するものである。生成されたデータは、後述する効率性の統計的性質(推定値の分布や信頼区間など)の評価に用いられる。

ここで、真の生産フロンティアは一意に存在すると仮定し、入力ベクトル x を確率変数 X の実現値として、また効率値 δ を確率変数 D の実現値として扱う。このとき、以下を仮定する：

$$(3.10) \quad E(D) = \mu > 0, \quad \text{Var}(D) < \infty.$$

この枠組みの下で、所与の出力 y に対する入力 X の不確実性や効率性分布の統計的性質を推定するため、以下のサンプリングアルゴリズムを導入する。

Algorithm 1 密度木に基づくサンプリングアルゴリズム(ガウス混合分布)

Require: 観測された入出力ベクトル集合 $\{(x_h, y_h)\}_{h=1}^n$

Ensure: 生成された擬似入力ベクトル集合 $\{x_h^b\}_{h=1}^n$, $b = 1, \dots, B$

1: 密度木の構築：

2: **for** 各ノード $j = 0, 1, \dots, J$ **do**

3: 情報利得を最大化する分割パラメータを決定：

$$\theta^* = \arg \max_{\theta \in \mathcal{T}} I(\mathcal{S}_j, \theta)$$

4: 分割関数 $h_j(\cdot, \theta^*)$ に基づき、データを左右に分割：

$$\mathcal{S}_j^L = \{v \in \mathcal{S}_j \mid h_j(v, \theta^*) = 0\},$$

$$\mathcal{S}_j^R = \{v \in \mathcal{S}_j \mid h_j(v, \theta^*) = 1\}$$

5: **end for**

6: 葉ノードにおける局所分布推定：

7: **for** 各葉ノード \mathcal{S}_l **do**

8: 平均ベクトル μ_l および共分散行列 Σ_l を推定

9: 葉ノードの事前確率を計算：

$$p(L = l) = \frac{|\mathcal{S}_l|}{|\mathcal{S}_0|}$$

10: **end for**

11: 条件付きサンプリング：

12: **for** 各データ点 (x_h, y_h) **do**

- 13: ベイズの定理により各クラスタに属する事後確率を計算：

$$p(L = l | \mathbf{Y} = \mathbf{y}_h) = \frac{p(\mathbf{y}_h | L = l)p(L = l)}{\sum_{l'} p(\mathbf{y}_h | L = l')p(L = l')}$$

- 14: 事後確率 $p(L = l | \mathbf{Y} = \mathbf{y}_h)$ に基づきクラスタ l^* をサンプリング
 15: クラスタ l^* における条件付き平均および分散を計算：

$$\begin{aligned}\boldsymbol{\mu}_{x|y}^{(l^*)} &= \boldsymbol{\mu}_x^{(l^*)} + \boldsymbol{\Sigma}_{xy}^{(l^*)}(\boldsymbol{\Sigma}_{yy}^{(l^*)})^{-1}(\mathbf{y}_h - \boldsymbol{\mu}_y^{(l^*)}), \\ \boldsymbol{\Sigma}_{x|y}^{(l^*)} &= \boldsymbol{\Sigma}_{xx}^{(l^*)} - \boldsymbol{\Sigma}_{xy}^{(l^*)}(\boldsymbol{\Sigma}_{yy}^{(l^*)})^{-1}\boldsymbol{\Sigma}_{yx}^{(l^*)}\end{aligned}$$

- 16: 条件付き分布 $\mathcal{N}_m(\boldsymbol{\mu}_{x|y}^{(l^*)}, \boldsymbol{\Sigma}_{x|y}^{(l^*)})$ から \mathbf{x}_h^b をサンプリング
 17: end for

アルゴリズム 1 では、情報利得に基づき再帰的に分割を行うことにより、密度木を構築する。その後、各葉ノードに局所的な多変量正規分布を推定し、これらを事前確率に基づいて重ね合わせることで、全体としてガウス混合モデルを構成する。さらに、観測された出力 \mathbf{y}_h に対してベイズの定理を用いてクラスタの事後確率を計算し、擬似入力 \mathbf{x}_h^b をサンプリングする。

このとき、ガウス混合モデルにおける $\mathbf{Y} = \mathbf{y}$ に対する条件付き分布は次式で表される：

$$(3.11) \quad p(\mathbf{X} | \mathbf{Y} = \mathbf{y}) = \sum_l \frac{|S_l|}{|S_0|} p(\mathbf{X} | \mathbf{Y} = \mathbf{y}; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l).$$

しかし、式(3.11)に従って単純にサンプリングを行う場合、クラスタ選択は事前確率 $\frac{|S_l|}{|S_0|}$ のみに依存する。このため、 $\mathbf{Y} = \mathbf{y}$ に対する尤度 $p(\mathbf{y} | L = l)$ が非常に低いクラスタからも \mathbf{X} が生成される恐れがある。特に、データ点がクラスタ境界付近に位置する場合、実際には複数の局所分布の影響を受けるはずであるにもかかわらず、単一クラスタに基づく条件付き分布のみを用いると、隣接クラスタの情報が無視される結果、不自然なサンプルが生成される可能性がある。この問題を回避するため、本研究ではベイズの定理に基づいて事後確率 $p(L = l | \mathbf{Y} = \mathbf{y})$ を計算し、 \mathbf{y} に適合するクラスタを優先的に選択する。これにより、局所的なデータ構造を反映しつつ、多様性と一貫性を兼ね備えた自然なサンプリングが可能となる。

観測データから構築された単一の密度木は、データのばらつきや偶然的な構成に強く依存する。このため、観測されていない領域に対して新規データ点を生成する場合、単一木の分割構造に由来する推定誤差やバイアスが顕著に現れ、推定される分布が不安定となる可能性がある。特に、データ密度が低い領域や、外挿を要する領域では、密度木の局所的な偏りが不自然なサンプル生成を引き起こしやすい。この問題を回避するため、本研究では複数のランダム化された密度木を用いたアンサンブル学習により、分布推定の安定性を向上させる。以下に示すアルゴリズム 2 は、複数の密度木からなるフォレスト(密度フォレスト)に基づくサンプリング手法である。

Algorithm 2 密度フォレストに基づくサンプリングアルゴリズム

Require: 観測された入出力ベクトル集合 $\{(\mathbf{x}_h, \mathbf{y}_h)\}_{h=1}^n$, 木の数 T

Ensure: 生成された擬似入力ベクトル集合 $\{\mathbf{x}_h^b\}_{h=1}^n$, $b = 1, \dots, B$

- 1: 密度フォレストの構築：
- 2: for $t = 1$ to T do
- 3: ブートストラップ標本 $S_0^{(t)}$ を作成
- 4: ランダムな特徴量選択に基づき密度木を構築

- 5: 各葉ノード $S_l^{(t)}$ に対し, 平均 $\mu_l^{(t)}$ および共分散 $\Sigma_l^{(t)}$ を推定
- 6: **end for**
- 7: 各木に対する条件付き分布推定:
- 8: 各木に対して, アルゴリズム 1 を適用
- 9: フォレスト全体の密度推定:
- 10: 各木から得られた条件付き密度を平均化し, フォレスト全体の密度関数を構築:

$$p_{\text{forest}}(\mathbf{X} | \mathbf{Y} = \mathbf{y}) = \frac{1}{T} \sum_{t=1}^T p^{(t)}(\mathbf{X} | \mathbf{Y} = \mathbf{y})$$

- 11: 条件付きサンプリング:
- 12: **for** 各データ点 $(\mathbf{x}_h, \mathbf{y}_h)$ **do**
- 13: 平均化された密度 $p_{\text{forest}}(\mathbf{X} | \mathbf{Y} = \mathbf{y}_h)$ に基づき, \mathbf{x}_h^b をサンプリング
- 14: **end for**

なお, アルゴリズム 2 におけるランダム特徴量選択は, Breiman (2001) において提案された Random Forest の構築原理に基づいている. すなわち, 各ノードにおいて, 全特徴量集合 \mathcal{F} の中からサイズ k の部分集合 $\mathcal{F}_j \subset \mathcal{F}$ を一様ランダムに抽出し, その中で最も情報利得 $I(\mathcal{S}_j, \theta)$ を最大化する分割パラメータ θ^* を選択する:

$$(3.12) \quad \mathcal{F}_j \sim \text{Uniform}(\mathcal{P}_k(\mathcal{F})),$$

$$(3.13) \quad \theta^* = \arg \max_{\theta \in \mathcal{T}_j(\mathcal{F}_j)} I(\mathcal{S}_j, \theta),$$

ここで, $\mathcal{P}_k(\mathcal{F})$ は \mathcal{F} から k 個を選択する部分集合の全体を表し, $\mathcal{T}_j(\mathcal{F}_j)$ は \mathcal{F}_j に対応する分割パラメータ空間である. k の選択については, 例えば $k = \lfloor \sqrt{|\mathcal{F}|} \rfloor$ のように, 特徴量数の平方根に比例させる設定が標準的である (Breiman, 2001). このようなランダム特徴量選択により, 各密度木の分岐構造に多様性を導入でき, 特定の特徴量や局所的なデータ構成に依存するリスクを低減できる. これにより, 単一の密度木に起因する推定誤差やバイアスを緩和し, アンサンブル全体の分布推定の安定性と汎化性能を向上させる効果が期待される.

4. 木構造に基づくサンプリングによる DEA 効率性の信頼区間推定

本節では, 密度木または密度フォレストに基づき生成された擬似入力データを用いて, DEA 効率性の推定および信頼区間を構築する方法について述べる.

まず, 入力指向型 DEA モデル (2.1) は, サンプリングされた入力ベクトル $\{\mathbf{x}_h^b\}_{h=1}^n$ を用いて以下のように表される:

$$(4.1) \quad \delta^b = \min \left\{ \delta^b > 0 \mid \sum_{h=1}^n \lambda_h \mathbf{x}_h^b \leq \delta^b \mathbf{x}^b, \quad \sum_{h=1}^n \lambda_h \mathbf{y}_h \geq \mathbf{y}, \quad \sum_{h=1}^n \lambda_h = 1, \quad \lambda \geq \mathbf{0} \right\}.$$

ここで, $b = 1, \dots, B$ はサンプリングインデックスを表し, $\{\mathbf{x}_h^b\}_{h=1}^n$ はアルゴリズム 1 またはアルゴリズム 2 によって生成された擬似入力データである.

それぞれのサンプルに対して得られる B 個の効率性推定値 δ^b について, その平均を以下により定義する:

$$(4.2) \quad \bar{\delta} = \frac{1}{B} \sum_{b=1}^B \delta^b.$$

前節で述べた通り，入力指向型 DEA モデル(2.1)に関して，入力ベクトル X を確率変数，対応する DEA 効率値を確率変数 D とみなし， $E(D) = \mu > 0$ ， $\text{Var}(D) < \infty$ が成立すると仮定する．この仮定のもとで，大数の法則により，サンプル数 $B \rightarrow \infty$ の極限において， $\bar{\delta}$ は母平均 $E(D)$ に一致し，すなわち真の効率値に収束することが保証される．この性質を利用することで， $\{\delta^b\}_{b=1}^B$ の経験分布に基づき，推定された DEA 効率値に対する信頼区間を構築することが可能となる．

本研究では， $\bar{\delta}$ の信頼区間を構築するために，ノンパラメトリック・ブートストラップ法を適用する．具体的には，効率値の集合 $\{\delta^b\}_{b=1}^B$ から，復元抽出により K 個のブートストラップ標本 $\{\delta^{b,k}\}_{b=1}^B$ ， $k = 1, \dots, K$ を生成する．各標本に対して平均値

$$(4.3) \quad \bar{\delta}^{(k)} = \frac{1}{B} \sum_{b=1}^B \delta^{b,k}$$

を計算し，その経験分布に基づいて信頼区間を構築する．信頼区間の構築には，ブートストラップ平均のパーセンタイル法を用いる．すなわち， $\{\bar{\delta}^{(k)}\}_{k=1}^K$ を昇順に並べ，その 2.5 パーセンタイル値および 97.5 パーセンタイル値をそれぞれ信頼区間の下限および上限として採用する．

これにより，DEA 効率性推定に伴う不確実性をノンパラメトリックな枠組みで定量的に評価することが可能となる．本研究と既存手法である Simar–Wilson 法との本質的な相違は以下の点にある．2.2 節で述べたように，Simar–Wilson 法は観測データに含まれる測定誤差や外生的ノイズを明示的にモデル化せず，データ生成過程に内在する誤差構造を十分に補正するものではない．これに対し，本研究では観測データ x_n に測定誤差や外生的ノイズが含まれる可能性を明示的に考慮する．この場合， P^{DEA} と P の包含関係 ($P^{DEA} \subset P$) は必ずしも成立せず，従来のバイアス補正とは異なるアプローチが必要となる．具体的には，観測データのばらつきやノイズを反映した確率的分布モデルを学習し，このモデルに基づいて多数の擬似データセットを生成する．各擬似データセットに DEA を適用して効率値を再推定することで，効率性推定値に対する経験的な分布を構築し，信頼区間を推定する．このアプローチにより，観測誤差を考慮したうえで，DEA 推定に伴う統計的不確実性を柔軟かつ実証的に評価することが可能となる．

5. シミュレーション実験

本節では，文献 Fare et al. (1989) で用いられた電力会社データ(表 1)を用いて，提案手法の有効性を検証する．このデータセットは，19 の電力会社(DMU)に対し，3 つの入力変数(Labor, Fuel, Capital)と 1 つの出力変数(Output)を含む．

アルゴリズム 1 により構築した密度木と，そこから得られたクラスタごとの DMU の割り当て結果を，それぞれ図 3 および表 2 に示す．表 2 より，クラスタ 1 は入出力変数の水準が低い小規模 DMU，クラスタ 2 は中程度の水準を示す中規模 DMU，クラスタ 3 は Fuel や Capital および Output が顕著に大きい大規模 DMU で構成されていることが分かる．さらに，図 3 に示すように，各クラスタにおける平均ベクトルおよび共分散行列には明確な差異が認められ，密度木による分類がデータの潜在構造を的確に捉えていることが確認できる．

次に，表 3 に，観測データに基づく DEA 効率値($\hat{\delta}$)，アルゴリズム 1 および 2 による推定効率値(それぞれ $\hat{\delta}^1$ および $\hat{\delta}^2$)，ならびにブートストラップ($B = 2000$)による 95% 信頼区間を示す．なお，密度フォレストはブートストラップ標本から 10 本の密度木を生成して構築した．

観測データに基づく DEA 効率値($\hat{\delta}$)と，密度木により生成した擬似データに基づく推定効

表 1. 入出力データ.

DMU の番号	Labor	Fuel	Capital	Output
1	179.0	3969.1	1005.5	3531.4
2	96.0	896.32	194.6	841.3
3	94.0	803.17	200.0	690.5
4	141.0	1629.9	564.1	1498.1
5	147.0	3264.4	617.4	3347.7
6	261.0	1427.5	347.1	1263.2
7	448.0	6365.3	1680.4	5734.0
8	223.0	4911.0	1319.4	4107.7
9	279.0	7779.3	1785.6	6964.2
10	214.0	2307.5	690.0	1961.0
11	367.0	4334.7	1268.9	4207.2
12	97.0	1870.7	416.5	1752.7
13	143.0	2730.6	779.8	2667.6
14	95.0	540.61	305.0	457.2
15	241.0	9509.2	1892.1	9539.9
16	147.0	1561.1	718.5	1360.6
17	306.0	1699.0	306.3	1689.7
18	182.0	881.16	182.3	750.3
19	500.0	2548.9	650.1	2439.5

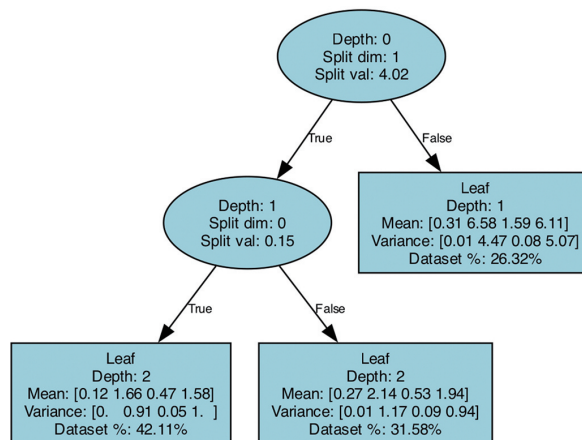


図 3. アルゴリズム 1 で構築した密度木.

表 2. クラスターの割り当て.

クラスター	DMU の番号
1	2, 3, 4, 5, 12, 13, 14, 16
2	1, 6, 10, 17, 18, 19
3	7, 8, 9, 11, 15

表 3. アルゴリズム 1 およびアルゴリズム 2 による効率性推定結果.

DMU 番号	$\hat{\delta}$	アルゴリズム 1(密度木)				アルゴリズム 2(密度フォレスト)			
		$\bar{\delta}^1$	$SD(\bar{\delta}^1)$	2.5%	97.5%	$\bar{\delta}^2$	$SD(\bar{\delta}^2)$	2.5%	97.5%
1	0.8691	0.9151	0.0760	0.9119	0.9183	0.9180	0.0674	0.9149	0.9209
2	1.0000	0.9505	0.0777	0.9470	0.9537	0.9869	0.0339	0.9853	0.9883
3	1.0000	0.9635	0.0715	0.9603	0.9667	0.9914	0.0261	0.9903	0.9925
4	0.9307	0.9425	0.0615	0.9398	0.9453	0.9228	0.0510	0.9206	0.9249
5	1.0000	0.9193	0.0752	0.9160	0.9226	0.9519	0.0606	0.9493	0.9546
6	0.9071	0.9410	0.0663	0.9380	0.9438	0.9206	0.0505	0.9185	0.9228
7	0.8909	0.9321	0.0544	0.9298	0.9345	0.9080	0.0558	0.9056	0.9104
8	0.8208	0.8971	0.0711	0.8940	0.9001	0.9224	0.0654	0.9195	0.9252
9	0.8885	0.9629	0.0421	0.9611	0.9646	0.9185	0.0435	0.9167	0.9205
10	0.8469	0.9377	0.0658	0.9347	0.9405	0.9287	0.0549	0.9263	0.9311
11	0.9531	0.8972	0.0696	0.8942	0.9003	0.9258	0.0643	0.9230	0.9286
12	1.0000	0.9373	0.0636	0.9346	0.9402	0.9424	0.0437	0.9406	0.9442
13	0.9602	0.9346	0.0664	0.9316	0.9374	0.9362	0.0642	0.9334	0.9390
14	1.0000	0.9953	0.0443	0.9931	0.9970	1.0000	0.0006	1.0000	1.0000
15	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000
16	0.8885	0.9416	0.0628	0.9388	0.9444	0.9149	0.0569	0.9124	0.9173
17	1.0000	0.9403	0.0626	0.9375	0.9430	0.9275	0.0508	0.9251	0.9298
18	1.0000	0.9566	0.0762	0.9530	0.9599	0.9859	0.0368	0.9843	0.9874
19	0.9441	0.9346	0.0652	0.9316	0.9375	0.9487	0.0502	0.9466	0.9508
平均	0.9421	0.9420	0.0617	0.9393	0.9447	0.9448	0.0461	0.9428	0.9468

率値($\bar{\delta}^1$)との分布を比較するため、Kolmogorov–Smirnov 二標本検定を適用した。Kolmogorov–Smirnov 検定は、二つの独立標本が同一分布に由来するかを検定するノンパラメトリック手法であり、DEA 効率値の分布形状に仮定を置かない点で本分析に適している。検定の結果、統計量 $D = 0.3684$ 、 $p = 0.1532$ が得られ、有意水準 5% では帰無仮説(両分布が同一である)を棄却できなかった。このことから、密度木による擬似データは観測データに基づく効率性分布を統計的に再現しているといえる。

さらに、密度フォレストについても同様に検定を行った結果、 $D = 0.3684$ 、 $p = 0.1532$ となり、同様に帰無仮説は棄却されなかった。すなわち、密度フォレストによる擬似データも観測データと統計的に同等の効率性分布を保持していることが確認された。特に、密度フォレストは特徴量選択にランダム性を導入するにもかかわらず、効率性分布を保持した上で推定の不確実性を反映した擬似データ生成が可能であることが示された。

最後に、密度木と密度フォレストを比較すると、効率性推定値の標準偏差の平均は密度木で 0.0617、密度フォレストで 0.0461 となり、アンサンブル化による分散低減効果が確認された。また、95% 信頼区間の幅の平均も密度木 ($|0.9447 - 0.9393| = 0.0054$) に比べ密度フォレスト ($|0.9468 - 0.9428| = 0.004$) で狭まり、より安定した推定が得られることが分かった。以上より、密度フォレストを用いることで、単一密度木に依存した場合に生じる推定の不安定性を回避し、より頑健で信頼性の高い DEA 効率性推定が可能であることが実証された。

6. 結論

本研究では、DEA における効率性推定および信頼区間推定に向けて、密度木および密度フォ

レストに基づくサンプリングアルゴリズムを提案した。提案手法は、観測データに内在するばらつきやノイズを明示的に取り込み、局所的な確率構造に基づいてデータ生成を行うことで、DEA 推定に伴う統計的不確実性を柔軟かつ実証的に評価可能とするものである。

本研究の主な貢献は、以下の三点に整理できる。

- (a)従来の Simar–Wilson 法とは異なり、観測された入力ベクトルを確率変数として扱い、測定誤差や外生的ノイズを内在化した擬似データ生成を可能とした点である。これにより、観測データの確率変動を考慮できると同時に、効率性も確率変数として扱うことが可能となり、バイアス補正を要せずブートストラップ法により信頼区間の推定が可能となった。さらに、観測された出力ベクトルを確率変数とすることで、提案手法は出力指向型 DEA モデルにも容易に拡張できる。
- (b)単一の密度木に依存する際に生じる推定の不安定性を回避するため、複数のランダム化された密度木からなる密度フォレストを導入した点である。これにより、標準偏差のばらつきや信頼区間幅のばらつきが抑制され、アンサンブル化による分散低減効果と推定安定性の向上が実証的に確認された。
- (c)実データを用いた検証の結果、Kolmogorov–Smirnov 検定において、観測データに基づく DEA 効率値と提案手法による推定効率値との分布に有意差は確認されなかった。このことは、提案手法が観測データに内在する効率性構造を統計的に保持している可能性を示唆するものである。

さらに、密度木の構築においては、各葉ノード単位で局所的にガウス近似を仮定しているものの、全体分布がガウス分布であることを仮定していない点に注意を要する。実際、図 2 に示したように、全体として非ガウスの分布（たとえばベータ型分布）であっても、局所的なガウス近似に基づいて効果的なモデリングとサンプリングが可能であることが示された。したがって、提案手法は、全体分布がガウス性から逸脱している場合でも有効に機能する柔軟性を有している。

また、本研究で直接比較対象としなかった StoNED や SFA といった回帰型効率性推定手法に対しても、以下の点で提案手法は優位性を有すると考えられる。StoNED や SFA では、非効率性と誤差項を合成した条件付き分布に基づき効率性を点推定するため、追加的なパラメトリック仮定が必要となり、効率性推定値の大きさ自体に実務的意味を持たせることは難しい。一方で、提案手法は効率性そのものの確率分布に着目し、古典的な統計手法を通じた直感的な信頼区間推定を可能としており、実務上の解釈性に優れる。

もっとも、本研究には限界も存在する。提案手法と Simar–Wilson 法に基づく信頼区間推定法との比較は実データに基づいて実施しておらず、その理由は、両手法が前提とするデータ生成過程が本質的に異なること、および実データでは真の投入量・産出量が不明であることに起因する。また、現実の生産可能集合を厳密に定義するための情報が不十分であることも、比較を困難にしている。この点を補完するためには、人工データを用いた多様な実験設計に基づく包括的検証が今後必要である。

今後の研究課題としては、多様な DEA モデルへの拡張、ガウス分布以外の柔軟な局所分布モデルの導入、単一密度木に基づくサンプリングにおける外れ値影響の定量的評価、ならびに密度木構築・サンプリング手順の頑健性と計算効率のさらなる向上が挙げられる。

注.

¹⁾ 指向性(改善の方向)によってさまざまなモデルが拡張可能である(森田浩, 2024)が、本稿

の提案手法は指向性に依存せず適用可能であるため、ここでは簡便のため入力指向を前提に議論を進める。

- ²⁾ 一方、 $\sum_{h=1}^n \lambda_h = 1$ を除去すると、規模に関する収穫一定 (Constant Returns to Scale, CRS) 下の経験的生産可能集合が得られ、この場合は入力を k 倍すると出力も k 倍となる。なお、規模に関する収穫には VRS や CRS のほか、Increase Return to Scale (IRS), Decrease Return to Scale (DRS), および General Return to Scale (GRS) などがある (森田浩, 2024)。本稿の提案手法は、規模に関する収穫の仮定に依存せず適用可能であるため、以下では簡便のため VRS を前提に議論を進める。

参 考 文 献

- Aigner, D., Lovell, C. K. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics*, **6**(1), 21–37.
- Aparicio, J. and Esteve, M. (2023). How to peel a data envelopment analysis frontier: A cross-validation-based approach, *Journal of the Operational Research Society*, **74**(12), 2558–2572.
- Aparicio, J., Kapelko, M. and Ortiz, L. (2023). Enhancing the measurement of firm inefficiency accounting for corporate social responsibility: A dynamic data envelopment analysis fuzzy approach, *European Journal of Operational Research*, **306**(2), 986–997.
- Aragon, Y., Daouia, A. and Thomas-Agnan, C. (2005). Nonparametric frontier estimation: A conditional quantile-based approach, *Econometric Theory*, **21**(2), 358–389.
- Banker, R. D. (1984). Estimating most productive scale size using data envelopment analysis, *European Journal of Operational Research*, **17**(1), 35–44.
- Banker, R. D., Charnes, A. and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, **30**(9), 1078–1092.
- Boubaker, S., Le, T. D., Ngo, T. and Manita, R. (2025). Predicting the performance of MSMEs: A hybrid DEA-machine learning approach, *Annals of Operations Research*, **350**(2), 555–577.
- Breiman, L. (2001). Random forests, *Machine Learning*, **45**, 5–32.
- Cazals, C., Florens, J.-P. and Simar, L. (2002). Nonparametric frontier estimation: A robust approach, *Journal of Econometrics*, **106**(1), 1–25.
- Cover, T. M. and Thomas, J. A. (2012). *Elements of Information Theory*, John Wiley & Sons, New York.
- Criminisi, A. and Shotton, J. (2013). *Decision Forests for Computer Vision and Medical Image Analysis*, Springer Science & Business Media, United Kingdom.
- Daraio, C. and Simar, L. (2014). Directional distances and their robust versions: Computational and testing issues, *European Journal of Operational Research*, **237**(1), 358–369.
- Daraio, C., Simar, L. and Wilson, P. W. (2020). Fast and efficient computation of directional distance estimators, *Annals of Operations Research*, **288**(2), 805–835.
- Dia, M., Takouda, P. M. and Golmohammadi, A. (2022). Assessing the performance of Canadian credit unions using a three-stage network bootstrap DEA, *Annals of Operations Research*, **311**(2), 641–673.
- Esteve, M., Aparicio, J., Rodriguez-Sala, J. J. and Zhu, J. (2023). Random Forests and the measurement of super-efficiency in the context of Free Disposal Hull, *European Journal of Operational Research*, **304**(2), 729–744.
- Fare, R., Grosskopf, S. and Kokkelenberg, E. C. (1989). Measuring plant capacity, utilization and technical change: A nonparametric approach, *International Economic Review*, **30**(3), 655–666.
- Guillen, M. D., Aparicio, J. and Esteve, M. (2023). Gradient tree boosting and the estimation of production frontiers, *Expert Systems with Applications*, **214**, <https://doi.org/10.1016/j.eswa.2022.119134>.

- Kang, H. J., Kim, C. and Choi, K. (2024). Combining bootstrap data envelopment analysis with social networks for rank discrimination and suitable potential benchmarks, *European Journal of Operational Research*, **312**(1), 283–297.
- Kerstens, K., Sadeghi, J., Van de Woestyne, I. and Zhang, L. (2022). Malmquist productivity indices and plant capacity utilisation: New proposals and empirical application, *Annals of Operations Research*, **315**(1), 221–250.
- Kneip, A., Simar, L. and Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models, *Econometric Theory*, **24**(6), 1663–1697.
- Kuosmanen, T. and Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints, *Journal of Productivity Analysis*, **38**, 11–28.
- Lin, S.-W. and Lu, W.-M. (2024). Efficiency assessment of public sector management and culture-led urban regeneration using the enhanced Russell-based directional distance function with stochastic data, *Journal of the Operational Research Society*, **75**(8), 1624–1642.
- Mergoni, A., Emrouznejad, A. and De Witte, K. (2025). Fifty years of data envelopment analysis, *European Journal of Operational Research*, **326**(3), 389–412.
- Michali, M., Emrouznejad, A., Dehnohalaji, A. and Clegg, B. (2023). Subsampling bootstrap in network DEA, *European Journal of Operational Research*, **305**(2), 766–780.
- Moradi-Motlagh, A. and Emrouznejad, A. (2022). The origins and development of statistical approaches in non-parametric frontier models: A survey of the first two decades of scholarly literature (1998–2020), *Annals of Operations Research*, **318**(1), 713–741.
- 森田浩 (2024). DEA の例解, オペレーションズ・リサーチ = Communications of the Operations Research Society of Japan: 経営の科学, **69**(1), 12–19.
- Ngo, T. and Tsui, K. W. H. (2022). Estimating the confidence intervals for DEA efficiency scores of Asia-Pacific airlines, *Operational Research*, **22**(4), 3411–3434.
- Olesen, O. B. and Petersen, N. C. (2016). Stochastic data envelopment analysis—A review, *European Journal of Operational Research*, **251**(1), 2–21.
- Shi, Y. and Zhao, W. (2024). An Integrated machine learning and DEA-predefined performance outcome prediction framework with high-dimensional imbalanced data, *INFOR: Information Systems and Operational Research*, **62**(1), 100–129.
- Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Science*, **44**(1), 49–61.
- Simar, L. and Wilson, P. W. (2004). Performance of the bootstrap for DEA estimators and iterating the principle, *Handbook on Data Envelopment Analysis* (eds. W. W. Cooper, L. M. Seiford and J. Zhu), 265–298, Springer US, Boston, Massachusetts.
- Valero-Carreras, D., Aparicio, J. and Guerrero, N. M. (2021). Support vector frontiers: A new approach for estimating production functions through support vector machines, *Omega*, **104**, <https://doi.org/10.1016/j.omega.2021.102490>.
- Walheer, B. (2022). Global Malmquist and cost Malmquist indexes for group comparison, *Journal of Productivity Analysis*, **58**(1), 75–93.
- Zhu, N., Zhu, C. and Emrouznejad, A. (2021). A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies, *Journal of Management Science and Engineering*, **6**(4), 435–448.

Tree-based Distribution Estimation for DEA Efficiency Evaluation

Yu Zhao

School of Management, Tokyo University of Science

Standard Data Envelopment Analysis (DEA) models are deterministic, assuming that the observed data are free from measurement errors and exogenous noise. However, in real-world data, stochastic variations are inevitable, and ignoring them can lead to substantial errors in efficiency estimation. Although bootstrap-based and regression-based approaches have been proposed to partially introduce statistical inference into DEA, a framework that directly addresses data uncertainty has not yet been established. In this study, we propose a novel sampling method based on density trees and density forests to estimate efficiency and confidence intervals in DEA. The proposed method automatically clusters the observed data and models local probabilistic structures by maximizing information gain based on a Gaussian entropy function within each cluster. Furthermore, by employing an ensemble learning framework, it ensures estimation stability and robustness while generating pseudo-datasets that reflect the variability and noise inherent in the observed data. This enables a flexible and empirical assessment of statistical uncertainty in DEA. An empirical analysis using data from electric utility companies demonstrated the effectiveness of the proposed method. Specifically, a Kolmogorov-Smirnov test comparing the efficiency scores derived from observed data with those estimated by the proposed method revealed no statistically significant differences in their distributions. This result suggests that the proposed approach can appropriately reproduce the distributional characteristics of the observed data and provide an effective means of efficiency evaluation that accounts for measurement errors and noise.

長崎県の果物生産における多様なデータを 基にした価格予測モデルの構築

姜 佳明¹・趙 宇²

(受付 2025 年 6 月 26 日；改訂 9 月 25 日；採択 9 月 26 日)

要 旨

スマート農業とは、ロボット技術、人工知能(AI)、IoT(モノのインターネット)などの先端技術を活用し、農作物の品質向上や農作業の省力化・効率化を目指す新たな農業の形態である。特にデータ活用による農業の効率化が強く求められている。農業の持続的な発展と地域の活性化を目的として、農業従事者の減少や農産物価格の高騰、さらに地方自治体が直面する深刻な労働力不足や高齢化の進行、農業従事者の収入減少といった、日本の農業を取り巻く課題を背景に、本研究では、長崎県における主要果実「びわ」の市場価格を予測するモデルの構築を目指す。具体的に、2011年から2023年までの長崎市中央卸売市場におけるびわの市場販売量、代替品(いちご・メロン)の価格、気象条件を説明変数として、びわの価格を予測するモデルを構築する。これにより、農家による価格リスクの管理を支援し、スマート農業の地域への実装に貢献することを目的とする。

キーワード：実社会問題解決、需要関数、スマート農業、地域経済、データサイエンス。

1. はじめに

2018年、首相官邸に設置された未来投資会議において、2025年までに農業の担い手のほぼすべてがデータを活用した農業を実践するという、スマート農業に関する政府目標が掲げられた。さらに、農林水産省は、スマート農業の普及を促進することを目的に、データ連携基盤の整備や現場への技術実装を支援する取組支援事業を立ち上げている。つまり、この目標を達成するためには、農業生産に関する各種データの利便性や操作性を向上させる工夫が不可欠である(長崎, 2023)。本研究では、先端技術を活用して多様な農業関連データを収集・分析し、生産工程の効率化および経営収益の向上に与える効果を実証的に検討する。

これらの取り組みは、農業分野における後継者や人材確保の効率化を図るとともに、新たなプレイヤーの農業分野への参入を促進し、農業をより魅力あるビジネスへと変革させる契機となっている。また、農業データの利活用を通じて生産性や品質の向上を実現するには、農業分野の特殊性を踏まえた上で、データの円滑な活用促進と営農ノウハウ等の保護を両立させるルール整備が重要である。そのため、政府はAI技術の導入に伴う契約上の課題への対応として、農林水産省は、「農業分野におけるデータ契約ガイドライン」(平成30年12

¹長崎大学 情報データ科学部：〒852-8521 長崎市文教町 1-14; jiaming@nagasaki-u.ac.jp

²東京理科大学 経営学部：〒162-8601 東京都新宿区神楽坂 1-3; yu.zhao@rs.tus.ac.jp

月)を基礎に、「農業分野における AI・データに関する契約ガイドライン」(令和2年3月)(<https://www.maff.go.jp/j/kanbo/tizai/brand/attach/pdf/keiyaku-10.pdf>)を策定し、農業におけるデータ活用の信頼性と安全性を高める環境整備を進めている。

日本の農業は、農業従事者の高齢化が進み、担い手不足が深刻化している。農業就業人口は、1995年の414万人から半減し、2015年には210万人となり、その年齢構成は、65歳以上の割合が全体の6割を超え、50歳未満の割合が1割程度という偏った構造となっている(日本学術会議, 2020)。今後、高年齢層の農業従事者が大量にリタイアすることにより、担い手不足はさらに深刻化することが考えられる(松本, 2020)。この中で、長崎地域は、国内でも人口減少や少子高齢化の課題が顕著な地域のひとつである。長崎大学情報データ科学部が2023年度に実施したアンケート調査(未公開)では、農業従事者の高齢化に加え、農業に対する否定的な意見が半数を超えており、多くの人が農業に対してマイナスのイメージを持っていることが明らかとなった。さらに、多くの農家が人手不足に苦しんでいることも確認された。しかしながら、農業は国民の食料・食料安全保障を支えるだけでなく、経済成長や生物多様性の保全にも寄与する重要な産業であり、その持続的発展のためには社会全体の支援が求められる。近年、気候変動やリソースの制約が厳しい現代において、データに基づく意思決定は農業の効率性と生産性を高める鍵となる。農業従事者の高齢化や担い手不足が深刻化する中で、スマート農業は、作業の自動化やデータ、センシング技術の活用を通じて、生産性の飛躍的な向上や熟練農業従事者の技術継承、精密な栽培管理を実現し、農業に変革をもたらすことが期待されている(松本, 2020)。

特に、データサイエンス技術を活用したスマート農業は、近年急速に進展している。気象・土壌データを用いた精密農業や、AIによる病害虫予測・灌漑最適化といった先端的取り組みが進められている。国外では、スマートセンサーやドローンによるデータ収集、機械学習による作物の成長予測等の研究が進行している(Yang et al., 2021)。国内のデータサイエンス技術を活用したスマート農業に関する研究事例はまだ少ないが、例えば温州みかんの糖度予測においては、前年の糖度および当年の気象データを用いたAIモデルにより高精度な予測が可能となっている(農研機構農業情報センター, 2020)。また、栃木県下野市の「トマトパーク」では、生育・収量予測ツールの活用により10aあたり50t以上の収量に加え、20%以上の増収が実現された。さらに、高知県の次世代型施設園芸プロジェクトでは、AIを用いて高収量農家の気象・出荷データを学習させ、生産者に対する最適な栽培アドバイスを提供することに成功している(長崎, 2023)。また、農業データの連携基盤として機能するWAGRIでは、さまざまな有料API(Application Programming Interface)が提供されて、農業分野におけるデータ利活用が加速している。たとえば、「柿果実収穫適期判定API」では、品種と地点の選択により収穫期の予測が可能である。「NARO 栽培管理支援API」シリーズでは、水稻の発育ステージ(出穂期・成熟期など)を気候平年値も加味して予測する機能があり、生育支援の強力なツールとなっている。

しかし、これらの技術の活用事例や研究結果は、まだ全国的に十分に広まっているとは言えない。特に、長崎地域における具体的な事例報告は、例えばびわシミュレーションシステムの開発による長崎県主産地の動向予測と活性化策(清水, 2012)、ビワ果実の成熟日予測モデルの開発(紺野, 2020)以外は限られており、地域特性に応じた研究やデータ活用のさらなる進展が求められている。このような地域別の研究が進むことにより、日本全体でのスマート農業の発展が期待される。

2. 使用データ

長崎県は、日本の西端に位置し、対馬暖流の影響による温暖多湿な気候(年平均気温 16°C, 年間降水量 2,000mm 前後)を有する(市来・一瀬, 1976)。同県は、びわ(栽培面積全国 1 位)、いちご、ももなど果実の産地として知られている。本研究では、びわ、いちご、メロンといった高付加価値の地域特産果実、特にびわに焦点を当てた価格予測モデルの開発を行う。

長崎県のびわ生産は、栽培面積および生産量が全国の約 3 割を占めて(政府統計の総合窓口 <https://www.e-stat.go.jp/>)、その栽培は特に長崎市茂木地区に集中している。この地域のびわは、長崎県を代表する農産物の一つとして位置づけられており、びわ生産は地域経済にも大きな影響を与えている。しかし、びわの栽培は自然条件の影響を受けやすく、生産量が年によって不安定であることが特徴である(寺床, 2018)。びわは温暖な気候を好み、西日本(特に長崎県)や千葉県房総半島を中心に栽培されている。露地栽培のびわは、特に幼果が低温に弱いため、冬季の寒冷や台風などの自然災害によって、次年度の収量が大きく減少することがある。このため、露地栽培のびわは気象条件に大きく影響され、そのリスクが高いと言える。一方、ハウス栽培のびわは、7月中旬から8月にかけて気温が高くなる時期に、遮光とミスト散水を併用することで、出蕾・開花時期を早め、収穫時期を前倒しすることが可能となり、これにより高単価での販売が実現できるようになった(瀧崎 他, 2008)。本研究では、長崎市中心卸売市場の市場年報を基に、長崎産の露地びわとハウスびわの月別単価(4月から6月)に基づく、価格予測モデルの構築を目指す。市場での取引実績をもとに、価格の変動要因を明らかにし、今後の価格動向を予測することによって、農家や市場関係者がより効率的に価格リスクを管理できるよう支援することが本研究の狙いである。

吉田(2004)は、市場価格を消費支出、市場販売量、代替品の価格、その他の要因に基づいて月別価格の予測を可能にする短期的モデルを構築し、行政部局の日常的な業務において実用可能な成果を示した。このモデルは、果実類の価格変動を的確に捉えるために、消費行動や市場供給状況といった経済的要因に加え、品質や代替性といった市場構造的要因を考慮している。具体的には、通常的需求関数を価格について解いた誘導形を採用し、販売量が価格に対して負の影響を与え、消費支出や代替品価格が正の影響を与えるという前提に基づいてモデルを構築した。とりわけ品質要因は、果実類の価格形成において重要な役割を果たす。みかんでは糖度が品質の指標となり得るが、りんごに関しては、流通段階における腐敗や日持ちの良否といった品質評価指標が確立されておらず、気象条件との関係性が価格形成要因として注目されている。

本研究では、吉田の先行研究の枠組みを踏まえた上で、新たにびわの市場価格予測モデルを構築することを目的とする。びわは生産・出荷の時期が限定される果実であり、その市場価格は供給量や代替品の存在、さらには気象条件によって大きく左右される。本研究では、びわの代替財として、同時期(主に春から初夏)に市場に出回る「いちご」と「キンショウメロン」の2品目を選定し、これらの価格動向がびわの価格に与える影響を分析する。また、果実の決定要因として気象条件の役割に着目し、「気象 → 品質」という関係性を明確に把握するための実証分析を行った。使用した気象データは、2012年から2023年の期間における長崎県内の月別平均気温などの主要気象変数を含んでおり、これらは国土交通省気象庁の公式ウェブサイト(<https://www.data.jma.go.jp/stats/etrn/index.php>)より取得した一次資料に基づいている。分析においては、被説明変数として長崎市中心卸売市場の市場年報に掲載されている露地びわとハウスびわの平均価格を使用した。説明変数としては、びわの市場販売量、代替品であるいちごおよびキンショウメロンの価格、さらに気象変数を取り上げている。これにより、びわの価格形成メカニズムにおける短期的な価格変動を予測可能とするモデルの構築を目指している。

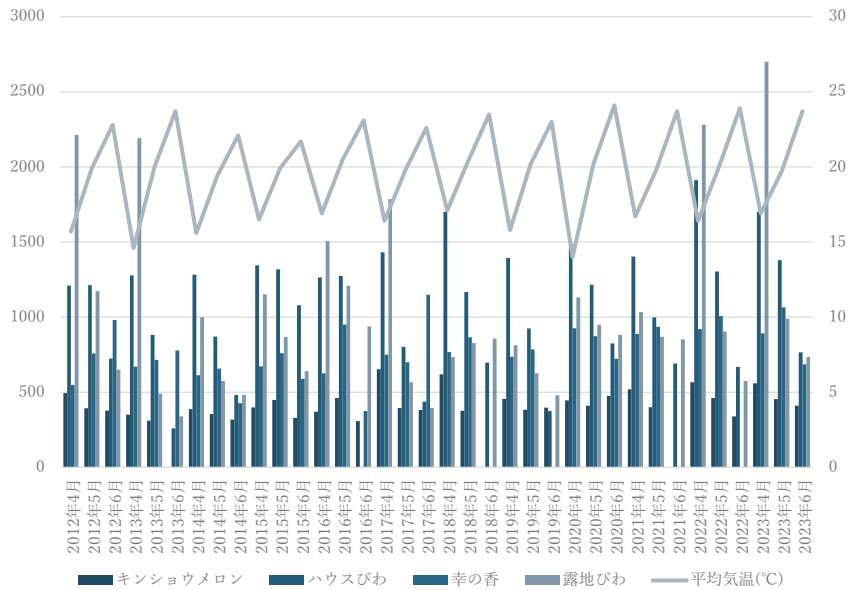


図1. 価格データの時系列推移.

図1は、キンショウメロン、ハウスびわ、「幸の香」いちご、露地びわの4月～6月の価格(左軸：円/kg)と平均気温(右軸：℃)を示している。この図より、基本的に価格が高い順は、露地びわ、ハウスびわ、「幸の香」いちご、キンショウメロンとなっており、全体として年々価格が上昇する傾向も認められる。露地びわの価格は、1kgあたり300円～2,700円の範囲で推移して、ハウスびわについては1kgあたり250円～1,900円の範囲で変動している。いずれも例年4月に価格のピークを迎える傾向があり、その後、5月から6月にかけて供給量の増加等により価格は下落する傾向が見られる。一方、代替品として取り扱われる「幸の香」いちごの価格は、1kgあたり370円～1,148円の範囲で推移しており、主に5月から6月にかけて価格のピークを迎える。また、キンショウメロンは1kgあたり310円～653円と比較的価格が安定して、びわと同様に4月に価格のピークを迎える傾向が確認される。このような中で、露地びわの価格は他品目と比べて特に大きな変動幅を示して、特に2017年から2019年にかけては価格の低迷が顕著であった。対照的に、同期間におけるハウスびわの価格は相対的に安定して、露地びわの価格を上回る傾向も見られた。これらの価格動向は、気温といった気象条件の影響を大きく受けている可能性が考えられる。したがって、次節においては、気象要因と価格変動との関連性について、より詳細な検証を行うこととする。

3. モデルの考え方

需要関数 Q_t を以下のように定式化する。

$$(3.1) \quad Q_t = \gamma_0 + \gamma_1 P_t + \gamma_2 P'_t$$

ここで、 t は時点を表し、 P'_t は代替品の価格(複数)である。この式を価格 P_t について解き、さらに両辺に対数変換を施すことで、価格と他の変数との間に線形な関係をもつ価格と他の変数に関する線形モデルを導出することができる。本研究ではこの枠組みを応用し、果実の品質を

表 1. 説明変数の概要.

記号	定義	経済的・統計的理由
Q_t	月次取扱量	供給量の増加による価格低下という逆需要関係を捉えるため
P_t^s	「幸の香」いちごの単価	近縁果実であるいちごの代替効果を反映するため
$P_{t,sm}^M$	キンショウメロンの単価 × 取扱量を 3 カ月移動平均で平滑化	メロンの代替効果を考慮しつつ、価格の短期的変動を平滑化するため
T_t^{avg2}	平均気温の 2 カ月移動平均	開花から収穫期までの累積気温が品質(糖度)に影響するため
D_t^{season}	4~6 月を 1, それ以外を 0 とするダミー変数	びわの旬(春~初夏)による季節要因を反映するため
$D_t^{pandemic}$	2022~2023 年を 1 とするダミー変数	コロナ禍に伴う物流コストおよび需要変動の影響を補正するため

気温で近似し、さらに季節性や新型コロナウイルス感染症の影響をダミー変数で表現した、以下の価格決定モデル(式(3.2))を構築した。

$$(3.2) \quad \ln P_t = \beta_0 + \beta_1 \ln Q_t + \beta_2 \ln P_t^s + \beta_3 \ln P_{t,sm}^M + \beta_4 T_t^{avg2} + \beta_5 D_t^{season} + \beta_6 D_t^{pandemic} + \varepsilon_t$$

誤差項 ε_t は独立性、等分散性、および正規性を満たすものと仮定する。被説明変数は、露地びわ価格とハウスびわ価格の算術平均値とした。なお、各説明変数の具体的な生成方法およびその経済的・統計的な理由については、表 1 に整理した。

推定に際しては、被説明変数としてのびわの価格、および説明変数である「幸の香」いちご、キンショウメロンの価格および取扱量について、それぞれ自然対数変換を行い、推定係数を価格弾力性として解釈可能な形式とした。また、気温およびキンショウメロンに関しては、それぞれの影響が即時的ではなく時間をかけて現れる点を考慮し、前者には 2 カ月移動平均、後者には単価と取扱量の積を 3 カ月移動平均により平滑化した。さらに、説明変数間の多重共線性の影響を緩和するため、リッジ回帰を推定手法として採用し、正則化項 $\alpha \|\beta\|^2$ を損失関数に付加した。正則化パラメータ α は、時系列データに適した 5 分割交差検証法を用いて、RMSE (Root Mean Squared Error) が最小となる値を探索範囲(0.05, 0.1, 0.2, 0.3, 0.5, 1, 1.2, 1.5, 3, 10, 30, 100)の中から選定した。

4. 推定結果

表 2 に、本モデルの月次予測精度指標を示す。決定係数 $R^2 = 0.671$ 、 $RMSE(\log) = 0.256$ 、および平均絶対パーセント誤差 $MAPE$ (Mean Absolute Percentage Error) = 20.8%、 MAE (Mean Absolute Error) = 419.0 (円/kg) となり、青果物の短期価格予測において実務上許容される精度を有していると評価できる。さらに、年次ベースで予測値を集計したところ、2012~2021 年における実測値との乖離は年間平均で ± 250 円/kg 以内に収まっていた。一方、2022~2023 年において観測された下方乖離は、新型コロナウイルス感染症の影響を捉えるダミー変数の導入により、約 427 から 260 円程度へと大幅に縮小された(表 3 参照)。これらの結果は、モデルが短期および年次の変動を一定程度的確に捉えていることを示している。

表 2. モデル予測精度のサマリ.

指標	値
決定係数 R^2	0.671
RMSE(log)	0.256
MAE (円/kg)	419.0
MAPE (%)	20.8
最適 α (Ridge 回帰)	1.0

表 3. 年度ごとの乖離(円/kg).

年月	乖離(円/kg)
2012-12-31	-95.8
2013-12-31	-59.5
2014-12-31	-46.9
2015-12-31	-327.9
2016-12-31	39.5
2017-12-31	-62.8
2018-12-31	20.9
2019-12-31	146.5
2020-12-31	251.0
2021-12-31	-254.3
2022-12-31	-426.7
2023-12-31	-259.8

表 4. リッジ回帰による推定係数(弾力性).

変数	β
$\ln Q_t$	-0.300
$\ln P_t^s$	+0.269
$\ln P_{t,sm}^M$	+0.189
T_t^{avg2}	-0.191
D_t^{season}	0.000
$D_t^{pandemic}$	+0.211

なお、表 4 に示す通り、本モデルから推定された係数は、複数の重要な経済的関係性を示している。まず、供給量に関する弾力性については、取扱量の 1% 増加に対して価格が約 0.30% 低下するという負の関係が確認された。これは、需給バランスに基づく典型的な逆需要関係を反映している。また、「幸の香」いちごに代表される代替果実の価格が 1% 上昇した場合、びわの価格は 0.27% 上昇しており、消費者の選好における代替効果の存在が示唆された。さらに、キンショウメロンを反映した変数も 0.19 の正の係数を持ち、類似果実との競合関係が確認された。品質要因として導入した平均気温(2カ月移動平均)は、 0.2°C (約 1%)の上昇に対して価格が 0.19% 低下する傾向を示した。これは、気温上昇による収穫集中や品質の均質化が市場価格を押し下げる可能性を示している。加えて、2022~2023 年の新型コロナウイルス感染症の影響

を表すダミー変数は、価格に対して正の効果(+21%)を持ち、この期間において供給制約や物流コスト上昇などの影響が市場価格に反映されたことが確認された。

一方で、4～6月の旬期を表す季節ダミーは推定値が0に非常に近づいた、すなわちリッジ回帰モデルにおいて“重要性がほとんどない”と判断された。このことから、季節性は明示的なダミー変数ではなく、気温や供給量などの実質的な変数によって十分に説明されていると解釈できる。

5. 結びに

農業分野において近年データの重要性が高まる中、“スマート農業”は注目されているトピックの一つである(深津, 2022)。本研究は、農業分野におけるデータ駆動型アプローチの有効性を実証的に示したものである。特に、地方特産品であるびわの価格予測への応用可能性が確認された点は、今後のスマート農業の推進に向けて、農家による価格リスク管理や収益の安定化につながるものである、スマート農業の地域実装に貢献するものであり、実務的な示唆を含むものと位置づけられる。

本研究は、長崎県という少子高齢化や人手不足が深刻な地域において、主要な果実—特にびわ—の価格予測モデルを構築することを目的としている。対数誘導価格モデルとリッジ正規化を組み合わせ、月次のびわ価格変動を解析した結果、決定係数 $R^2 = 0.671$ 、 $RMSE(\log) = 0.256$ 、 $MAPE = 20.8\%$ という実務上許容される精度が得られた。さらに、2012～2021年の実測値と予測値との差が年間平均 ± 250 円/kg 以内に収まり、モデルの整合性が確認された。

また、2022～2023年に観測された価格の下方乖離については、新型コロナウイルス感染症の影響を表すダミー変数の導入により、乖離幅が約427から260程度に縮小され、当該モデルが外的ショックに対して一定の対応力を有していることが示唆された。推定結果からは、供給量、代替果実価格および気温といった変数が、価格決定において統計的な影響を与える主要因子であることが明確となった。一方、4～6月の旬期ダミーの重要性が認められなかった点は、気温や供給量などの実質的な変数に季節性が内包されている可能性を示した。今後のモデル改良の方向性として、品質指標(例：糖度、酸度)や需要面の経済指標(例：家計果実支出)の導入が期待される。さらに、回帰係数の信頼性については、ブートストラップ法(小西 他, 2008)を用いた検証を今後の課題とする。

以上より、本研究は地域農業の持続的な発展に資する価格予測モデルの一例として、農業とデータサイエンスの融合が持つ実践的可能性を実証的に示したものと位置づけられる。将来的には、本研究で構築したモデルを他の農産物や他地域に応用・展開することで、汎用性と実用性をさらに高めていく可能性がある。また、長崎大学情報データ科学部における他分野の研究室との学際的連携、たとえばAIセンサー技術を活用した病害虫の自動検出、土壌環境のリアルタイムモニタリング、あるいは気象データとの統合分析などを通じて、収量および価格予測モデルの精度をより一層高める取り組みも視野に入る。こうした技術的統合と地域連携を通じて、本研究の成果は、地域農業のスマート化・高度化を支える基盤技術として、今後さらなる展開が期待される。

参 考 文 献

- 深津時広(2022). 農業現場におけるデータ利活用の現状と展開, スマート農業に向けたデータの収集と活用の取り組み, (国研)農研機構・農業機械研究部門施設園芸生産システムグループ, https://www.chisou.go.jp/sousei/resas/pdf/24_dataseminar_shiryo.pdf (最終アクセス日 2025

- 年9月30日).
- 市来小太郎, 一瀬 至 (1976). 指定試験の歩みと実績(18): 長崎県果樹試験場におけるカンキツ土壤肥料試験とビワ育種, *農業技術*, **31**(6), 270–271.
- 小西貞則, 越智義道, 大森裕浩 (2008). 『計算統計学の方法—ブートストラップ, EM アルゴリズム, MCMC—』, シリーズ〈予測と発見の科学〉, 5, 朝倉書店, 東京.
- 紺野祥平, 杉浦俊彦, 谷本恵美子, 稗圃直史, 蔦木康德, 山田英尚, 岩田浩二 (2020). ビワ果実の成熟日予測モデルの開発, *生物と気象*, **20**, 41–48.
- 松本賢英 (2020). スマート農業の社会実装に向けた取組み, *日本農薬学会誌*, **45**(2), 141–145.
- 長崎裕司 (2023). 『スマート農業』, 国立研究開発法人 農業・食品産業技術総合研究機構(編著), 成山堂書店, 東京.
- 日本学術会議 (2020). 提言: 人口減少社会に対応した農業情報システム科学の課題と展望, <https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t296-3.pdf> (最終アクセス日 2025年9月30日).
- 農研機構農業情報センター (2020). 「AIによる温州みかん糖度予測手法を開発—適切な栽培管理への活用に期待—」, 2020年農業技術10大ニュース, TOPIC 9, https://www.naro.go.jp/publicity_report/press/laboratory/rcait/134410.html (最終アクセス日 2025年9月30日).
- 清水一也 (2012). びわシミュレーションシステムの開発による長崎県主産地の動向予測と活性化策, *長崎農林技術開発センター研究報告*, **3**, 1–11.
- 瀧崎祥光, 堺一樹, 三原崇史, 奥田良幸 (2008). ハウス栽培ビワ‘長崎早生’の収穫促進に及ぼす遮光およびミスト散水の影響, *熊本県農業研究センター研究報告*, **15**, 149–153.
- 寺床幸雄 (2018). 社会関係資本からみた長崎市におけるビワ栽培の持続性と地域的課題, *経済地理学年報*, **64**(1), 36–54.
- Yang, X., Shu, L., Chen, J., Ferrag, M., Wu, J., Nurellari, E. and Huang, K. (2021). A survey on smart agriculture: Development modes, technologies, and security and privacy challenges, *IEEE/CAA Journal of Automatic Sinica*, **8**(2), 273–302, <https://doi.org/10.1109/JAS.2020.1003536>.
- 吉田泰治 (2004). プロジェクト研究の紹介: 果実価格予測モデルと品質要因, *農林水産政策研究所レビュー (PRIMAFF Review)*, **12**, 46–56.

Construction of Price Prediction Models Based on Diverse Data for Fruit Price in Nagasaki Prefecture

Jiaming Jiang¹ and Yu Zhao²

¹School of Information and Data Sciences, Nagasaki University

²School of Management, Tokyo University of Science

Smart agriculture is a new form of agriculture that utilizes advanced technologies such as robotics, artificial intelligence (AI), and the Internet of Things (IoT) to improve crop quality and streamline farming operations. In particular, the efficient use of data is strongly demanded to enhance agricultural productivity. This study aims to promote the sustainable development of agriculture and revitalize local communities by addressing the challenges facing Japanese agriculture, such as the declining number of agricultural workers, rising agricultural product prices, severe labor shortages and aging populations in local governments, and decreasing farmer incomes. Specifically, the study seeks to develop a price prediction model for “biwa” (loquats), a major fruit produced in Nagasaki Prefecture. Using data from 2011 to 2023 from the Nagasaki City Central Wholesale Market, the model will incorporate explanatory variables such as the market volume of loquats, the prices of substitute products (strawberries and melons), and weather conditions. The goal is to support farmers in managing price risks and contribute to the implementation of smart agriculture in the region.

走行環境が自動運転自動車に与える ヒヤリハット要因の分析

田島 友祐¹・笛田 薫²・三浦 雄作³・三樹 孝博³

(受付 2025 年 7 月 1 日；改訂 9 月 18 日；採択 9 月 25 日)

要 旨

本論文では、自動運転車両の走行実験データを用いて、走行環境がヒヤリハット発生に与える影響を分析した。具体的には、ヒヤリハット発生時の走行環境として、走行モード、道路形状、行動類型、道路情報、周辺状況をカテゴリ変数化し、ゼロ切断ポアソン回帰モデルを構築することで、ヒヤリハット数の予測ならびにその要因を分析した。さらに、それらの2変数間の交互作用項を導入し、モデルの変数選択を行い、解釈性を検討した。その結果、ヒヤリハット発生に有意な影響を与える走行環境の組合せが明らかになった。また、走行モードによらず行動類型と周辺状況の交互作用がヒヤリハットの発生数の増加に影響することが判明した。本研究の成果は、今後、巡回型バスなどの無人自動運転サービスの安全性向上に資する知見を提供するものと思われる。

キーワード：ゼロ切断ポアソン回帰、交互作用、リスク分析。

1. はじめに

近年、自動運転車両(Autonomous Vehicles, AV)の技術は急速に進展しており、レベル4に相当する運転手不要のモビリティサービスの実現に向けた実証実験が国内外で進められている。自動運転の実現に合わせ、法制度の整備も着実に進行している(国土交通省 2023, 2024)。自動運転の実現は、交通事故の削減、巡回バスやタクシーの増加による高齢者や交通弱者への移動支援、さらには、近年需要が高まっている物流や配達などで生じている慢性的なドライバー不足の解消といった、現代社会が抱える多くの課題への解決策として期待されている(Fagnant and Kockelman, 2015; Litman, 2020)。

滋賀大学では、あいおいニッセイ同和損害保険と連携し、群馬大学および同大学発スタートアップである日本モビリティが実施した自動運転車両の走行実験データを用いて、自動運転車両の走行ルートに関するリスク評価モデルの構築に取り組んでいる。自動運転車両においては、人間による運転ミスや判断の遅れといった従来の事故原因は回避可能と考えられる。一方で、自動運転車両に搭載されているセンサーや走行アルゴリズムに基づく判断が環境要因と複

¹ 大妻女子大学 データサイエンス学部：〒102-8357 東京都千代田区三番町 12 番地; yusuke.tajima@otsuma.ac.jp

² 滋賀大学 データサイエンス学部：〒522-8522 滋賀県彦根市馬場 1 丁目 1-1; kfueda@biwako.shiga-u.ac.jp

³ あいおいニッセイ同和損害保険株式会社 商品・CSV × DX 企画部：〒150-8488 東京都渋谷区恵比寿 1-28-1; t-miki@aioinissaydowa.co.jp, yusakumiura1218@aioinissaydowa.co.jp

雑に相互作用することが考えられるが、その相互作用については未だ不明な点が多く、安全に運用するための課題が多く存在するとされている。

実運用に向けて重要なことは、事故には至らなかったものの潜在的な危険が存在する「ヒヤリハット」事象に注目することである。ヒヤリハットは、重大事故に繋がる手前の予兆を示しており、それを分析することによって、重大事故の要因を把握することができ、安全な走行ルートや運用管理が可能になると考えられる。このようなヒヤリハットの要因は、走行環境(道路形状、交通状況、周辺の構造物や人の動きなど)によって大きく影響されるため、データに基づく定量的な分析が求められる(Kim et al., 2017; 伊藤 他, 2020)。

本論文では、今後日本各地で展開される、中型のバスなどの無人自動運転モビリティサービスに向けて、走行環境がヒヤリハット発生に与える影響を明らかにすることを目的とする。具体的には、走行環境に関するカテゴリ変数(走行モード、道路形状、交通情報、周辺状況など)をもとに、ゼロ切断ポアソン回帰モデル(zero-truncated Poisson regression)を用いてヒヤリハットの件数を予測し、各要因およびその交互作用がリスクに与える影響を分析する。なお、モデルの妥当性の確認のためにゼロ切断負の二項回帰モデルによる解析結果も併用する。

2. 自動運転車両におけるヒヤリハット数予測に向けた実験およびデータ概要

2.1 自動運転車両による走行実験の概要

本研究では、群馬大学および日本モビリティ株式会社によって実施された走行実験データを用いる。このデータは、自動運転車両の走行中に発生するヒヤリハット(本論文では、車両に搭載されている加速度センサが予め設定した閾値を超過したイベントと定義する)の予測を目的として、複数地域で実施された走行実験データである。走行実験は都市部や郊外をはじめとする多様な環境で行われている。実験に用いられている自動運転車両には、加速度センサだけでなく、カメラも搭載されており、走行実験中の走行環境についても連続的に記録されている。

2.2 計測データの構成

本研究で用いた走行実験データは、2017年から2021年にかけて実施された、計55箇所での走行実験の記録である。このうち、走行実験の記録の一部に欠損があるものや、記録管理の関係で連携が困難なデータを除外し、最終的に40箇所の走行実験データを対象として分析を行った。

これらの40箇所の走行実験データ全体の総走行距離は約18,600kmであり、そのうち8,828kmが自動運転による走行である。手動運転を行う必要がある場合として、走行実験ルート内での工事区間や、やむを得ない事情によるルート変更など自動運転の実施が困難と判断される場が考えられる。なお、40箇所の走行実験で各ルートの走行距離や自動運転・手動運転の走行距離にばらつきが見られるが、40箇所すべてで見ると自動運転と手動運転がほぼ同程度の割合で実施されている。

走行実験中に用いられる車両には加速度センサおよび周囲の状況を記録するカメラが搭載されている。本研究では、そのうちヒヤリハットが発生した時点の車両情報ならびに走行環境について記録がされているデータを用いる。具体的には、自動運転か手動運転かを制御している走行モード、道路形状、行動類型、道路情報、および周辺状況に着目する。以下に、本研究で着目した各項目について説明する。

- 道路形状：単路、交差点の区別に加え、見通しの良好／不良、走行環境に関する情報を示す。
- 行動類型：車両の挙動として、直進、右折、左折の進行方向を示す。

- 道路情報：歩車区分に関する情報であり、ガードレールや段差の有無などを示す。
- 道路情報2：中央分離帯の有無やその構造に関する情報を含む。歩者区分を表す道路情報と同様にガードレールや段差の有無などを示す。
- 周辺状況：市街地か否かなど、道路周辺の土地利用に関する情報を示す。

ここで、ヒヤリハットとは、車載の加速度センサが事前に設定された閾値を超過したイベントを指し、車両のみだしや歩行者や自転車等の飛び出しに対して、事故を避けるための急ブレーキなどの急減速時に発生する。ハインリッヒの法則(Heinrich's Law) (Heinrich, 1931)やバードの法則(Bird's Accident Ratio) (Bird and Germain, 1985)にあるように1つの大きな事故の背後には軽微な事故があり、その背後にはヒヤリハットがあるということが知られている。本分析で扱う40箇所走行実験データには、合計406件のヒヤリハットが発生しており、このヒヤリハットについて走行環境をもとに発生要因についての分析を実施する。

2.3 データのカテゴリ化

本分析の実施のために、ヒヤリハット発生時の走行モード、道路形状、行動類型、道路情報、道路情報2、および周辺状況についてのカテゴリ定義を表1に示す。

ヒヤリハットとして記録された合計406件の走行環境について表1の数値変換を施し、同一の組合せパターンごとに分類した。表2はその一部であり、ヒヤリハットの数は1以上であり、そのパターン(水準組合せ)は48とおり、ヒヤリハットの数は最小値で1、最大値で74となっている。

表 1. 走行環境に関する変数のカテゴリ化。

項目	カテゴリ	対応する水準
走行モード	ヒヤリハット発生前が手動運転	0
	ヒヤリハット発生前が自動運転	1
道路形状	単路	0
	交差点	1
行動類型	直進	0
	右折	1
	左折	2
道路情報(歩車区分)	何もしないもしくは線のみ	0
	段差もしくはガードレールがある	1
道路情報2(中央分離帯)	何もしないもしくは線のみ	0
	段差もしくはガードレールがある	1
周辺状況	市街地(住宅街)	0
	市街地(住宅街でない)	1
	非市街地	2

表 2. ダミー変数の組合せごとの出現回数。

ID	走行モード	道路形状	行動類型	道路情報	道路情報2	周辺状況	出現回数
1	0	0	2	1	0	1	74
2	1	1	0	1	0	0	41
3	0	0	0	0	0	2	25
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
48	0	1	0	1	1	2	1

3. ゼロ切断ポアソン回帰

自動車の走行運転時のヒヤリハットの予測のような問題を扱う際、予め定めた単位時間あたりにある事象が生じる回数を予測する問題としてポアソン回帰 (Gourieroux et al., 1984; Long, 1997; Long and Freese, 2006; Winkelmann, 2008; Cameron and Trivedi, 2013) が用いられることが多い。ポアソン回帰は、従属変数 Y が非負の整数値をとる場合に適した回帰モデルとして知られている。ポアソン回帰モデルでは、従属変数 Y_i の期待値 $\lambda_i = E[Y_i]$ が説明変数に対して次のように定義される。

$$\lambda_i = \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})$$

これはリンク関数として対数リンク関数(log-link function)を用いていることを意味する。

$$\log \lambda_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

このとき、 Y_i は平均 λ_i のポアソン分布に従うと仮定される。

$$Y_i \sim \text{Poisson}(\lambda_i)$$

ポアソン分布の確率分布は以下の通りである。

$$P(Y_i = y) = \frac{\lambda_i^y e^{-\lambda_i}}{y!}, \quad y = 0, 1, 2, \dots$$

しかし、今回の分析に用いるデータはヒヤリハットが発生したという条件のもとでの走行環境に関するデータであるため、ヒヤリハットが0となるデータは存在しない。そのため、ポアソン回帰を実施することは難しい。そこで、従属変数が常に正の整数をとる場合の回帰モデルであるゼロ切断ポアソン回帰(zero-truncated Poisson regression)を用いる。通常のポアソン回帰では、従属変数 Y_i は0を含む非負の整数であることを仮定しているが、観測対象が「発生した件数」のみに限定されている場合、すなわち $Y_i = 0$ がデータに含まれないとき用いるのがゼロ切断ポアソン回帰 (Gurmu, 1991; Grogger and Carson, 1991) である。

ここでは、 $Y_i > 0$ の条件のもとでの発生件数を X_i とする。 X_i がゼロ切断ポアソン分布に従うとは、次のような Y_i の条件付き分布を意味する。

$$P(X_i = x) = P(Y_i = x | Y_i > 0) = \frac{\lambda_i^x e^{-\lambda_i}}{x!(1 - e^{-\lambda_i})}, \quad x = 1, 2, 3, \dots$$

これは、 $Y_i = 0$ の確率を除いて再正規化したポアソン分布である。このとき、

$$E[X_i] = \frac{\lambda_i}{1 - e^{-\lambda_i}} \quad V[X_i] = \frac{\lambda_i}{1 - e^{-\lambda_i}} \left(1 - \frac{\lambda_i e^{-\lambda_i}}{1 - e^{-\lambda_i}} \right)$$

であるから、

$$\mu_i = \frac{\lambda_i}{1 - e^{-\lambda_i}}$$

とおき、期待値の対数を説明変数で表現する点は通常のポアソン回帰と同様である：

$$\log \mu_i = \beta^T \mathbf{x}_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}.$$

ただし、この場合、 $\mu_i > 1$ であるので、 $\beta^T \mathbf{x}_i > 0$ となるような β を推定する必要がある。ゼロ切断ポアソン回帰における β の推定法については、様々な研究がなされている (Gurmu, 1991; Grogger and Carson, 1991; Long, 1997; Van der Heijden et al., 2003; Simonoff, 2003; Tian et al., 2018; Li et al., 2023; Godwin, 2024)。本論文では、RのパッケージであるVGAM (Yee and Ma, 2024)を用いてゼロ切断ポアソン回帰を実施する。

ヒヤリハット数の検討を進めるにあたり、ゼロ切断ポアソン回帰では過分散の影響などにより十分な当てはまりが得られない可能性がある。そこで、本研究では負の二項回帰も実施し、ゼロ切断ポアソン回帰との比較による検討も進める。負の二項回帰は、ポアソン回帰を一般化したものであり、次の式で示されるものである。ここで、 n はサイズパラメータ、 $\mu_i^{(0)}$ は平均パラメータであり、サイズパラメータによって過分散に対する調整がなされる。

$$\Pr(Y_i = x | \mu, \phi) = \frac{\Gamma(n+x)}{\Gamma(n)x!} \left(\frac{n}{\mu_i^{(0)}+n} \right)^n \left(\frac{\mu_i^{(0)}}{\mu_i^{(0)}+n} \right)^x, \quad x = 0, 1, 2, \dots$$

この場合にも、ゼロ切断ポアソン回帰と同様に観測値に 0 が現れないゼロ切断負の二項回帰モデルを用いる。このとき、

$$\begin{aligned} P(X_i = x) &= P(Y_i = x | Y_i > 0) \\ &= \frac{1}{1 - \left(\frac{n}{\mu_i^{(0)}+n} \right)^n} \frac{\Gamma(n+x)}{\Gamma(n)x!} \left(\frac{n}{\mu_i^{(0)}+n} \right)^n \left(\frac{\mu_i^{(0)}}{\mu_i^{(0)}+n} \right)^x, \quad x = 1, 2, 3, \dots \end{aligned}$$

となる。

$$\mu_i = E[X_i] = \frac{\mu_i^{(0)}}{1 - \left(\frac{n}{\mu_i^{(0)}+n} \right)^n}$$

したがって、リンク関数は

$$\log \mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

である。使用する R のパッケージは VGAM である。

4. ゼロ切断ポアソン回帰を用いたヒヤリハット数の予測

4.1 主効果のみの予測

図 1 は、走行モード、道路形状、行動類型、道路情報、道路情報 2、周辺状況のカテゴリ変数を用いてゼロ切断ポアソン回帰により推定したヒヤリハットの予測値と実測値の散布図であ

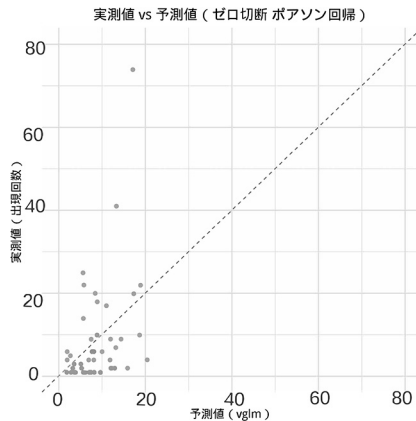


図 1. 主効果のみでのゼロ切断ポアソン回帰。実測値と予測値の比較。

る。横軸は予測値、縦軸は実測値を示しており、斜めの点線は予測値と実測値が一致する線を示している。図より、推定精度は高くなく、すべての実測値に対して予測値が0~20の範囲に集中しており、十分な推定精度をもつ推定ではないことがわかる。

4.2 交互作用項を考慮した予測

よりあてはまりのよい推定とするために、走行モード、道路形状、行動類型、道路情報、道路情報2、周辺状況のカテゴリ変数の2変数間の交互作用項を導入したモデルを用いて、改めてヒヤリハットの推定を実施する。表3は用いた交互作用項をまとめたものである。このうち、ヒヤリハットの数か0となっている交互作用項については考慮から除外している。

本論文において、3変数以上の交互作用については、分析に用いるデータ数が十分でないため、観測値が存在しない水準組合せがあり、推定不能となるため考慮しない。今後、実証実験数の増加や実環境による走行が増えた際には、3因子交互作用が及ぼすヒヤリハットへの影響を考慮することも考えられ、走行環境との影響についてより詳細な分析ができると考えられる。

ポアソン回帰では、過分散問題(over-dispersion)が指摘されることがある。そこで、ゼロ切断負の二項回帰でも推定を行い、比較検討を実施している。図2は、不要な交互作用項を除外

表3. 交互作用項の一覧。

交互作用項
走行モード_ダミー:道路形状_ダミー
走行モード_ダミー:行動類型_ダミー
走行モード_ダミー:道路情報_ダミー
走行モード_ダミー:道路情報2_ダミー
走行モード_ダミー:周辺状況_ダミー
道路形状_ダミー:行動類型_ダミー
道路形状_ダミー:道路情報_ダミー
道路形状_ダミー:道路情報2_ダミー
道路形状_ダミー:周辺状況_ダミー
行動類型_ダミー:道路情報_ダミー
行動類型_ダミー:道路情報2_ダミー
行動類型_ダミー:周辺状況_ダミー
道路情報_ダミー:周辺状況_ダミー
道路情報2_ダミー:周辺状況_ダミー

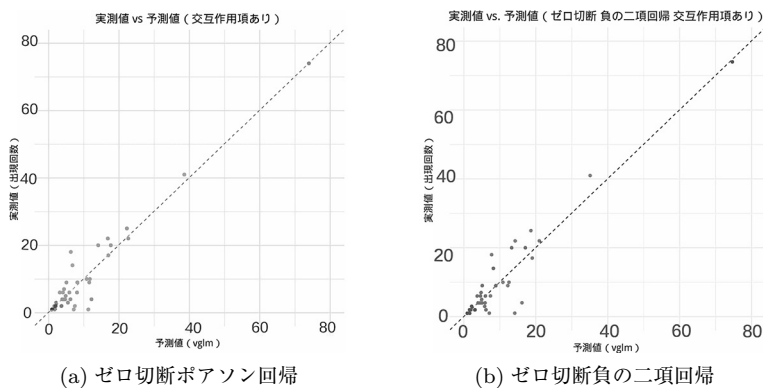


図2. 交互作用項をすべて用いた回帰。実測値と予測値の比較。

表 4. 推定結果と有意性(ゼロ切斷ポアソン回帰).

効果名	推定値	有意性
(Intercept)	2.8374	***
走行モード_ダミー 1	-0.3854	
道路形状_ダミー 1	-0.7338	*
行動類型_ダミー 1	-2.3714	
行動類型_ダミー 2	-1.6812	***
道路情報_ダミー 1	-0.1837	
道路情報 2_ダミー 1	-1.0257	**
周辺状況_ダミー 1	-1.3750	**
周辺状況_ダミー 2	0.2696	
走行モード_ダミー 1:道路形状_ダミー 1	1.1589	***
走行モード_ダミー 1:行動類型_ダミー 1	-1.6288	
走行モード_ダミー 1:行動類型_ダミー 2	-18.1724	
走行モード_ダミー 1:道路情報_ダミー 1	0.8553	*
走行モード_ダミー 1:道路情報 2_ダミー 1	-0.1119	
走行モード_ダミー 1:周辺状況_ダミー 1	-1.0141	**
走行モード_ダミー 1:周辺状況_ダミー 2	-19.1795	
道路形状_ダミー 1:行動類型_ダミー 1	-16.3506	
道路形状_ダミー 1:行動類型_ダミー 2	-2.9793	
道路形状_ダミー 1:道路情報_ダミー 1	0.1066	
道路形状_ダミー 1:道路情報 2_ダミー 1	0.3108	
道路形状_ダミー 1:周辺状況_ダミー 1	-0.2658	
道路形状_ダミー 1:周辺状況_ダミー 2	-15.1947	
行動類型_ダミー 1:道路情報_ダミー 1	0.4531	
行動類型_ダミー 2:道路情報_ダミー 1	0.3345	
行動類型_ダミー 1:道路情報 2_ダミー 1	0.7148	
行動類型_ダミー 2:道路情報 2_ダミー 1	1.2907	.
行動類型_ダミー 1:周辺状況_ダミー 1	16.9911	
行動類型_ダミー 2:周辺状況_ダミー 1	3.2628	***
行動類型_ダミー 1:周辺状況_ダミー 2	1.0536	
行動類型_ダミー 2:周辺状況_ダミー 2	0.3013	
道路情報_ダミー 1:周辺状況_ダミー 1	1.1098	.
道路情報_ダミー 1:周辺状況_ダミー 2	-0.4562	
道路情報 2_ダミー 1:周辺状況_ダミー 1	1.1485	***
道路情報 2_ダミー 1:周辺状況_ダミー 2	-1.1076	

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, .: $p < 0.1$

したモデルでのヒヤリハットの推定結果を示しており、図 1 と同様の形式の散布図となっている。左はゼロ切斷ポアソン回帰、右はゼロ切斷負の二項回帰による推定結果である。

図 2, 表 4 のモデルでは、どちらの回帰においても、交互作用を考慮しないモデル(図 1)と比較して予測値の当てはまりが改善していることが確認できる。また、ゼロ切斷ポアソン回帰ならびにゼロ切斷負の二項回帰の予測値と観測値の関係を比較すると、ほぼ同等の当てはまりであるが、本研究ではヒヤリハットの出現回数が多い観測値の特徴を抽出することが目的であり、その意味ではわずかにゼロ切斷ポアソン回帰の方が当てはまりが良いと判断できる。

4.3 変数選択によるモデルの簡素化

表 4 は、ゼロ切斷ポアソン回帰ですべての 2 因子交互作用までを考慮したモデルの各効果

表5. 交互作用項を含むゼロ切断ポアソン回帰モデルの比較(対数尤度・パラメータ数・AIC).

モデル		対数尤度	パラメータ数	AIC
1	主効果のみ	-286.1136	9	590.2272
2	交互作用項をすべて用いたモデル	-104.8937	34	277.7874
3	モデル2から係数は大きい が有意差がない交互作用項を削除	-135.1193	24	318.2386
4	モデル3から行動類型:道路情報 を削除	-136.3659	22	316.7318
5	モデル4から走行モード:道路情報 を削除	-137.6000	21	317.2000
6	モデル5から走行モード:道路情報2 を削除	-138.2256	20	316.4512
7	モデル6から道路形状:道路情報と 道路形状:道路情報2を削除	-139.3560	18	314.7120

注: モデル3では, 以下の交互作用項を削除した:

走行モード:行動類型, 走行モード:周辺状況, 道路形状:行動類型, 道路形状:周辺状況, 行動類型:道路情報2

(変数)の推定値ならびに対応する p 値をまとめたものである. モデルのあてはまりは大きく改善されたが, ヒヤリハット回数が最も多い(74回)組合せなど, 予測値が実測値と一致しており, 特定のデータに対する過適合(オーバーフィッティング)が生じている可能性が考えられるとともに, 推定モデルの各効果の p 値が大きく, ヒヤリハットの数の予測に寄与しないと考えられる変数も多く含まれていることがわかる. 表4の中で, 推定値は大きい有意でない交互作用項として次のものが挙げられる.

- (i) 走行モード1(自動運転):行動類型2(左折)
- (ii) 走行モード1(自動運転):周辺状況2(非市街地)
- (iii) 道路形状1(交差点):行動類型1(右折)
- (iv) 道路形状1(交差点):周辺状況2(非市街地)
- (v) 行動類型1(右折):周辺状況1(市街地(住宅地でない))

これらの交互作用の水準組合せをもつデータは1個あるいは2, 3個と少ないため, 係数は大きい有意ではないものとなっている. このように有意でない交互作用項を除外し, 推定精度を維持しつつモデルの簡素化と汎化性能の向上を目的として, p 値が小さい交互作用項を残して, AICに基づく変数選択(変数減少法)を実施することとした. すべての交互作用項を考慮したモデルから, AICによるモデルの当てはまりに着目しつつ, 変数の解釈性が高くなるように逐次変数選択を実施する. 表5は, ゼロ切断ポアソン回帰およびゼロ切断負の二項回帰において変数選択の過程をまとめたものである. この結果, 最終的に交互作用項として, 走行モード:道路形状, 行動類型:周辺状況, 道路情報:周辺状況, 道路情報2:周辺状況を含むモデル7を選択した.

図3は, モデル7による推定結果である. 図1ならびに図2と同様の形式の散布図となっている. 左はゼロ切断ポアソン回帰の推定結果, 右はゼロ切断負の二項回帰の推定結果である. ゼロ切断ポアソン回帰での推定のAICは278から315であり, 負の二項回帰での推定のAICは269から263となっており, 過適合はなくなりつつ, 交互作用項を考慮しないモデルよりも予測精度は高くなっている.

表6は, 交互作用項を考慮したゼロ切断ポアソン回帰モデル, 表7はゼロ切断負の二項回帰モデルの各効果の係数と p 値を示しているが, 各効果がヒヤリハットの数の予測に関係しているモデルとなっており, すべての交互作用項を考慮したモデルに対して, 推定精度を維持しつつモデルの簡素化と汎化性能の向上を図ることができた. この場合も, ゼロ切断負の二項回帰

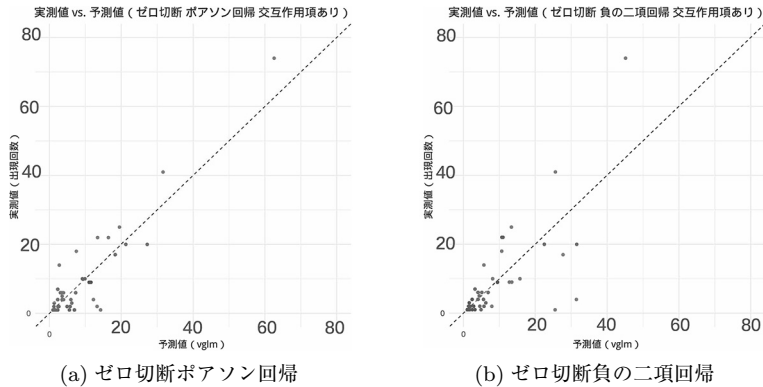


図 3. 逐次変数選択をしたモデルでの予測値と実測値の比較.

表 6. 推定係数と有意性(ポアソン).

変数	推定値	有意性
(Intercept)	2.91142	***
走行モード_ダミー 1	-0.50219	**
道路形状_ダミー 1	-1.54837	***
行動類型_ダミー 1	-3.48629	***
行動類型_ダミー 2	-1.70777	***
道路情報_ダミー 1	0.3957	*
道路情報 2_ダミー 1	-0.85589	***
周辺状況_ダミー 1	-1.66205	***
周辺状況_ダミー 2	0.05855	
走行モード_ダミー 1:道路形状_ダミー 1	2.19994	***
行動類型_ダミー 1:周辺状況_ダミー 1	3.88723	***
行動類型_ダミー 2:周辺状況_ダミー 1	3.5463	***
行動類型_ダミー 1:周辺状況_ダミー 2	2.5152	*
行動類型_ダミー 2:周辺状況_ダミー 2	0.57793	
道路情報_ダミー 1:周辺状況_ダミー 1	0.65437	
道路情報_ダミー 1:周辺状況_ダミー 2	-1.14006	***
道路情報 2_ダミー 1:周辺状況_ダミー 1	1.07014	***
道路情報 2_ダミー 1:周辺状況_ダミー 2	-0.63207	

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, ∴ $p < 0.1$

による予測値と実測値とのプロットを比較するとゼロ切断ポアソンモデルと大差がないことがわかる。そのため、今回はモデルの簡易さならびに解釈のしやすさからゼロ切断ポアソン回帰モデルを採用した。

4.4 主効果と交互作用効果の分析

ヒヤリハットの予測モデル7の各効果についてみる。なお本研究で用いた48パターン(水準組合せ)に対するヒヤリハットの数は1以上であり、ヒヤリハットが起こったもとの解析結果であることに注意されたい。

まず、主効果については、以下のようになっている。ただし、交互作用項についても合わせて考える必要があるため、ここでは主効果の大小関係を述べるに留める。

表7. 推定係数と有意性(負の二項回帰).

変数	推定値	有意性
(Intercept)	3.3191	***
(Intercept2)	0.726	***
走行モード_ダミー1	-1.1006	**
道路形状_ダミー1	-1.7812	***
行動類型_ダミー1	-3.8772	***
行動類型_ダミー2	-2.0443	***
道路情報_ダミー1	0.1293	*
道路情報2_ダミー1	-0.862	***
周辺状況_ダミー1	-2.0463	***
周辺状況_ダミー2	-0.7438	
走行モード_ダミー1:道路形状_ダミー1	2.6716	***
行動類型_ダミー1:周辺状況_ダミー1	3.46	***
行動類型_ダミー2:周辺状況_ダミー1	3.1075	***
行動類型_ダミー1:周辺状況_ダミー2	3.1762	*
行動類型_ダミー2:周辺状況_ダミー2	1.2342	
道路情報_ダミー1:周辺状況_ダミー1	1.3427	
道路情報_ダミー1:周辺状況_ダミー2	-0.6431	***
道路情報2_ダミー1:周辺状況_ダミー1	1.5627	***
道路情報2_ダミー1:周辺状況_ダミー2	-0.6139	

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, .: $p < 0.1$

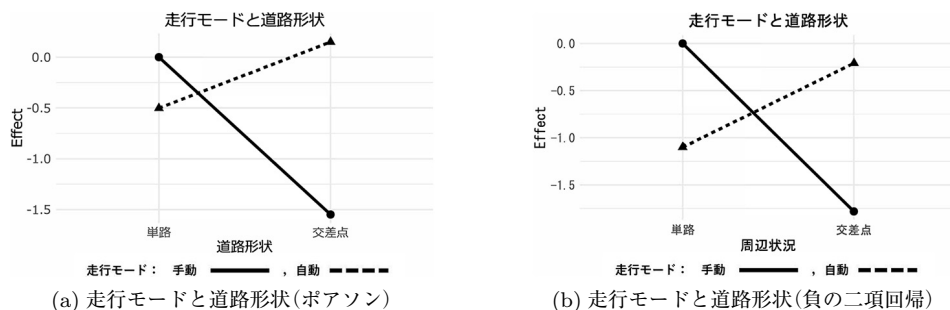


図4. 走行モードと道路形状の比較.

- 走行モード：手動>自動
- 道路形状：単路>交差点
- 行動類型：直進>左折>右折
- 道路情報(歩者区分)：段差などの障害物がある>何も無い
- 道路情報2(中央分離帯)：何も無い>段差などの障害物がある
- 周辺状況：市街地であり住宅街>非市街地>市街地だが住宅街ではない

次に、交互作用項を含めた要因効果についてみていく。図4は走行モードと道路形状、図5は行動類型と周辺状況、図6は道路情報(歩者区分)と周辺状況、図7は道路情報(中央分離帯)と周辺状況について各カテゴリでの要因効果量を示したものである。横軸ならびに各種類の線はそれぞれの説明変数に対応しており、縦軸は要因効果量を表している。それぞれ、左にはゼロ切断ポアソン回帰、右にはゼロ切断負の二項回帰による結果を示している。

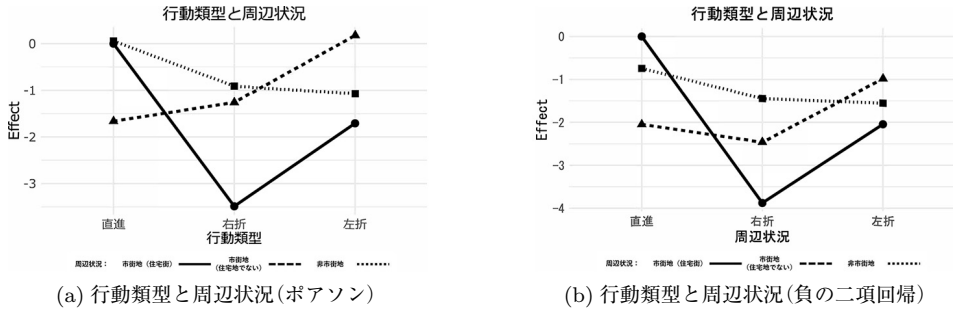


図 5. 走行モードと道路形状の比較.

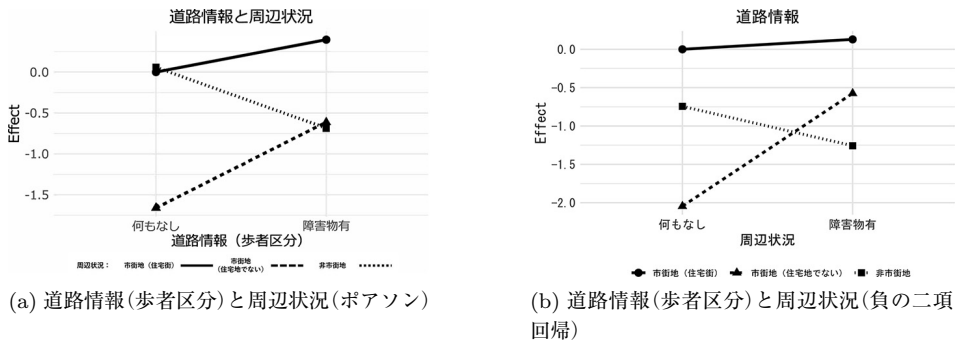


図 6. 道路情報 (歩者区分) と周辺状況.

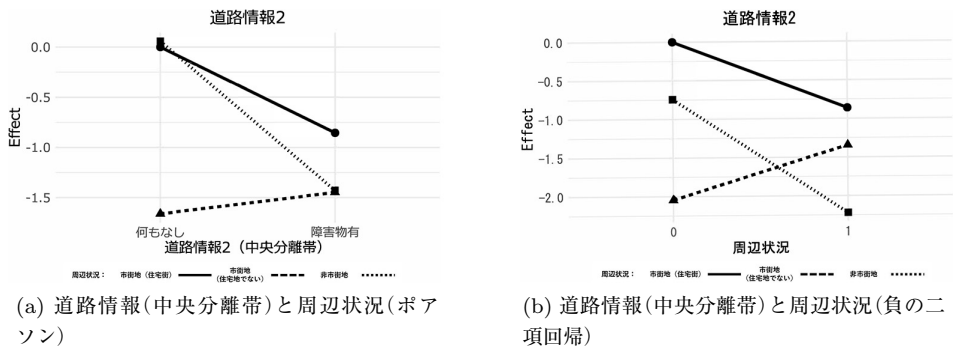


図 7. 道路情報 (中央分離帯) と周辺状況.

(1) 走行モードと道路形状 図 4 より, 走行モード(手動と自動)は道路形状のみと交互作用が有意である. 手動運転と自動運転では走行する道路形状によってヒヤリハットの発生数への影響が異なることがわかる. 交差点の周辺状況の視認に加え, 交差点に入る直前の減速や進行時の徐行などを同時に処理することは, 今回の実証実験中では自動運転よりも手動運転の方がヒヤリハットの数を減少させる傾向にあることがわかる.

(2) 行動類型と周辺状況 図 5 より, 市街地(住宅地でない)における左折については, 他の行動類型と周辺状況の交互作用と異なり, ヒヤリハットを増やす要因になっていることがわかる. ここで市街地(住宅地でない)としては駅近くなどの交通量が多いところが挙げられる. そのような場所では車両だけでなく人通りや自転車も多く, 左折時の周辺状況の確認が困難であるこ

表 8. ダミー変数による走行条件と出現回数.

走行モード	道路形状	行動類型	道路情報	道路情報 2	周辺状況	出現回数
0	0	2	1	0	1	74
1	1	0	1	0	0	41

とが考えられる。また、右左折よりも直進行動の方がスピードが出、飛び出しなど不測の事態が生じることが考えられ、急ブレーキをふむなどヒヤリハットが生じていることが考えられる。

(3)道路情報(歩者区分)と周辺状況 図6より、非市街地ではヒヤリハットを減らす要因になっていることがわかる。一方、市街地では歩者区分のための段差やガードレールなどの障害物はヒヤリハットを増やす要因となっている。

(4)道路情報(中央分離帯)と周辺状況 図7より、市街地(住宅街)と非市街地ではヒヤリハットを減らす要因になっていることがわかる。一方、市街地(住宅地でない)では中央分離帯のための段差やガードレールなどの障害物の有無はヒヤリハットの増減への影響は小さい。

4.5 ヒヤリハットが多いパターンの分析

最後に、ヒヤリハットが多いパターンについてゼロ切断ポアソン回帰モデルを用いて詳細にみていく。

表8はヒヤリハットが多い上位2つのパターンをまとめたものである。まず、ヒヤリハット数が74のパターンの推定値は次のようになる。

$$\log(\mu) = 2.91142 - 1.70777 + 0.3957 - 1.66205 + 3.5463 + 0.65437 = 4.13797$$

この場合、 $\mu = e^{4.14} \approx 62.63$ となる。Interceptである2.91142だけでは平均ヒヤリハット数は $\mu = e^{2.91} \approx 19$ となるが、この場合、主効果と交互作用項を合わせて正となる項目としては

- (1) 行動類型(左折)：周辺状況(市街地(住宅街でない))
- (2) 道路情報(段差もしくはガードレールがある)：周辺状況(市街地(住宅街でない))

の2つの交互作用が影響している。よって、駅前商業地など交通量が多い所であつ、左折時あるいは段差などがある所でヒヤリハットが起りやすいことがわかる。

次に、ヒヤリハット数が41のパターンについて推定値は次のようになる。

$$\log(\mu) = 2.91142 - 0.50219 - 1.54837 + 0.3957 + 2.19994 = 3.4565$$

したがって、 $\mu = e^{3.46} \approx 31.71$ となる。この場合、同じく主効果と交互作用項を合わせ正となる項目としては

- (1) 道路情報(段差もしくはガードレールがある)
- (2) 走行モード(自動運転)：道路形状(交差点)の交互作用

の2つが影響している。すなわち段差やガードレールがあり歩者区分があるところ、あるいは自動運転の際の交差点でヒヤリハットが起りやすいことがわかる。

5. まとめ

本研究では、走行実験においてヒヤリハットが発生した時点での発生状況のデータについて、走行環境がヒヤリハット発生に与える影響を分析した。自動運転車両の走行環境要因(道路形状、行動類型、道路情報、周辺状況)と自動運転モードの交互作用がヒヤリハット発生数に

与える効果を、ゼロ切断ポアソン回帰モデルとゼロ切断負の二項回帰モデルにより検討した。

その結果、道路形状(単路, 交差点)と走行モード(手動, 自動)との相互作用がヒヤリハット発生に有意な影響を与えることが明らかとなった。行動類型や周辺状況との組合せがリスクを増大させる傾向があり、交通量が多い地域とそうでない地域でヒヤリハットとなりうる要因が異なることが見られた。特に、(i) 交通量が多い市街地(住宅街でない)で段差やガードレールによる歩者区分がある所、(ii) 自動運転の際の交差点、(iii) 交通量が多い市街地での左折時にはヒヤリハットが増加することが明らかになった。(iii) で特筆すべき点は交差点での直進や右折よりも左折時のヒヤリハットが多い点である。

今後の課題として、ヒヤリハットが発生した場合の要因としてではなく、走行開始から走行終了までを一定の短時間で区切り、ヒヤリハットが発生するか否かのデータを用いたポアソン回帰などにより、どのような走行環境でヒヤリハットが発生しやすいのかをさらに詳細に分析していく。さらに、交通量などの影響がヒヤリハットにどのように関係するのかも明らかにし、自動運転車両がとっさに判断することが難しい横断についても危険度を推測することで、自動運転車両が直面するヒヤリハット数の予測と削減に繋がると考えられる。

謝 辞

本研究の実施に際し、走行データのご提供および、モデルの評価・解釈に関して実運用の観点から多大なご助言を賜りました国立大学法人群馬大学次世代モビリティ社会実装研究センターおよび当該センターの小木津准教授、ならびに日本モビリティ株式会社に、心より感謝申し上げます。

参 考 文 献

- Bird, F. E. and Germain, G. L. (1985). *Practical Loss Control Leadership*, International Loss Control Institute, Loganville, Georgia.
- Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data*, 2nd ed., Cambridge University Press, Cambridge, UK.
- Fagnant, D. J. and Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations, *Transportation Research Part A: Policy and Practice*, **77**, 167–181.
- Godwin, R. T. (2024). One-inflated zero-truncated count regression models, *arXiv preprint arXiv:2402.02272* (最終アクセス日 2025 年 10 月 3 日).
- Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to Poisson models, *Econometrica*, **52**(3), 701–720.
- Grogger, J. T. and Carson, R. T. (1991). Models for truncated counts, *Journal of Applied Econometrics*, **6**(3), 225–238.
- Gurmu, S. (1991). Tests for detecting over-dispersion in the positive Poisson regression model, *Journal of Business and Economic Statistics*, **9**(2), 215–222.
- Heinrich, H. W. (1931). *Industrial Accident Prevention: A Scientific Approach*, McGraw-Hill.
- 伊藤誠, 堀宗朗, 加藤真平 (2020). 自動運転車両における異常検知と事故予防のためのリスク評価, 人工知能学会論文誌, **35**(4), 562–573.
- Kim, S., Kang, J. Y., Kim, H. M. and Lee, J. (2017). Analysis of near-miss incidents in autonomous vehicles using naturalistic driving data, *IEEE Transactions on Intelligent Transportation Systems*, **18**(4), 807–818.

- 国土交通省 (2023). 道路運送車両法の一部を改正する法律(令和4年法律第36号, 2023年4月施行), <https://laws.e-gov.go.jp/law/326AC0000000185/> (最終アクセス日 2025年10月3日).
- 国土交通省 (2024). 自動運転移動サービス社会実装・事業化の手引き, https://www.road-to-the-l4.go.jp/activity/guideline/pdf/自動運転移動サービス社会実装・事業化の手引き_v2.0_.pdf (最終アクセス日 2025年10月3日).
- Li, Xun-Jian., Sun, Yuan, Tian, Guo-Liang, Liang, Jiajuan and Shi, Jianhua (2023). Mean regression model for the zero-truncated Poisson distribution and its generalization, *Computational Statistics and Data Analysis*, **179**, 107650.
- Litman, T. (2020). Autonomous vehicle implementation predictions, Victoria Transport Policy Institute, Victoria, British Columbia.
- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*, Sage Publications, Thousand Oaks, California.
- Long, J. S. and Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata*, Stata Press, College Station, Texas.
- Simonoff, J. S. (2003). *Analyzing Categorical Data*, Springer-Verlag, New York.
- Tian, G. L., Liu, Y., Tang, M. L. and Jiang, X. (2018). Type I multivariate zero-truncated/adjusted Poisson distributions with applications, *International Journal of Computational and Applied Mathematics*, **344**, 132–153.
- Van der Heijden, P. G., Bustami, R., Cruyff, M. J., Engbersen, G. and Van Houwelingen, H. C. (2003). Point and interval estimation of the population size using the truncated Poisson regression model, *Statistical Modelling*, **3**(4), 305–322.
- Winkelmann, R. (2008). *Econometric Analysis of Count Data*, Springer, Berlin, Heidelberg.
- Yee, T. W. and Ma, C. (2024). Generally altered, inflated, truncated and deflated regression, *Statistical Science*, **39**(4), 568–588.

Analysis of Near-miss Factors Influenced by Driving Environment in Autonomous Vehicles

Yusuke Tajima¹, Kaoru Fueda², Yusaku Miura³ and Takahiro Miki³

¹Faculty of Data Science, Otsuma Women's University

²Faculty of Data Science, Shiga University

³Product & CSV × DX Planning Department, Aioi Nissay Dowa Insurance

This paper analyzes the influence of driving environments on the occurrence of near-miss incidents in autonomous vehicles using data from driving experiments. Specifically, we categorized the driving environment at the time of each incident in terms of driving mode, road geometry, driving behavior, road information, and surrounding conditions. Using these categorical variables, we constructed a zero-truncated Poisson regression model to predict the number of incidents and identify contributing factors. Furthermore, by introducing two-way interaction terms between variables, we conducted variable selection and examined the interpretability of the model. The results revealed specific combinations of driving environments that significantly affect incident occurrence. In particular, the interaction between driving behavior and surrounding conditions was found to increase the number of incidents regardless of the driving mode. The findings of this study are expected to contribute to the enhancement of safety in future autonomous mobility services, such as driverless shuttle buses.

土壌化学組成と母岩・気候条件の関係：主成分 分析と線形判別分析による検討

張 天逸¹・中西 寛子²

(受付 2025 年 6 月 3 日；改訂 9 月 30 日；採択 9 月 30 日)

要 旨

土壌形成には気温や降水量などの気候因子が強く影響するため、土壌の化学組成からその場の気候条件を推定する試みは多くなされてきた。しかし、土壌の化学組成には気候条件のみによって規定されるわけではなく、特に化学風化の程度が弱い環境では母岩の特性が強く反映されると考えられるが、従来モデルではほとんど考慮されてこなかった。そこで本研究では、土壌の化学組成が気候条件および母岩によってどのように特徴づけられるかを明らかにすることを目的に、異なる気候条件下で生成された世界各地の土壌試料の主要元素化学組成データを対象として、多変量解析を実施した。主成分分析(PCA)の結果、土壌化学組成は寒冷・乾燥から温暖・湿潤といった気候条件に加え、化学風化の進行度にかかわらず母岩の影響も強く受けていることが示された。また、母岩タイプの判別を目的とした線形判別分析(LDA)により、土壌の化学組成から母岩タイプが高精度に識別されることが明らかとなった。本研究結果は、温暖・湿潤環境で化学風化が進行した状況下においても、母岩の特性が土壌組成に反映されうること示しており、土壌化学組成を検討する際には気候要因と母岩タイプの影響を明確に区別して評価することで、従来モデルに内在していた潜在的なバイアスを低減できることを示唆する。

キーワード：組成データ、対数比解析、土壌化学組成、化学風化、PCA、LDA。

1. はじめに

土壌は、地球表層において岩石が風化作用を受けて生成される物質であり、その化学組成は母岩を出発物質とし、そこに加わる外的要因によって物理的・化学的な反応が進行した結果として形成される。土壌化作用は、鉱物の溶解・析出、植生・微生物活動、天水・地下水による物質の付加・溶脱などの多数の素過程が相互に影響しながら進行する複雑系であるが、これらは基本的に太陽エネルギーを起点とし、温度や降水量といった気候要因を介して土壌系にエネルギーが供給されることにより駆動される(Rasmussen et al., 2005)。このような背景から、土壌の化学組成から気温や降水量を推定するという試みはこれまでに多く行われており、実際古気候研究にも幅広く活用されてきた(例えば、Sheldon et al., 2002; Rasmussen and Tabor, 2007; Perri, 2018; Mangold et al., 2019; Deng et al., 2022)。しかしながら、先行研究のほとんどは、土壌化作用の出発条件となる母岩の違いをモデル化の際に考慮しておらず、古気候推定の文脈においても、後背地の化学的特性が時間的・空間的に不変であるという仮定のもとで議

¹ 早稲田大学 データ科学センター：〒169-8050 東京都新宿区西早稲田 1-6-1; tianct@fuji.waseda.jp

² 統計数理研究所：〒190-8562 東京都立川市緑町 10-3; nmf.nakanishi@nifty.com

論されてきた。

本研究では、土壌の化学組成が気候条件(温度・降水量)および母岩タイプによってどのように特徴づけられるのかを明らかにすることを目的として、異なる気候帯および母岩条件下で形成された土壌の化学組成データに対し、探索的なデータ解析を実施した。明確な母岩情報を有し、かつ異なる気候帯に位置する405点の土壌試料について、バルク組成でパーセントオーダーで含まれる元素(主要元素)のデータを収集し、これらのデータに対して主成分分析および線形判別分析を適用することで、土壌化学組成と気候・母岩との関係性、ならびに土壌化学組成から母岩タイプ(玄武岩/花崗岩)の識別が可能かどうかを評価した。

2. 土壌のバルク化学組成データ

本研究では、異なる気候帯に存在する母岩が既知の土壌の主要成分に関する全岩化学組成データ($n = 405$)を収集した。対象としたデータは一般的な岩石や土壌にパーセントオーダーで含まれる SiO_2 , TiO_2 , Al_2O_3 , Fe_2O_3 , MgO , CaO , Na_2O , K_2O とした。土壌のサンプルは、アメリカ、プエルトリコ、ブラジル、アルゼンチン、マレーシア、中国、ロシア、スウェーデン、スイス、イタリア、カメルーン、ナイジェリア、エチオピアから得られたものである(USDA, 2009; Land et al., 1999; Norton and von Blanckenburg, 2010; Giovanoli et al., 1988; Dethier and Bove, 2011; Talabi, 2015; White et al., 1998; Yusoff et al., 2013; White et al., 2002; Bao and Zhao, 2008; Liu et al., 2016; Fu et al., 2019; Pokrovsky et al., 2005; Barbera et al., 2008; Manefouet Kentsa et al., 2023; Wu et al., 2021; Li et al., 2018; Liu et al., 2020; Le Blond et al., 2015; Caner et al., 2014; Campodonico et al., 2019; Yuan et al., 2022; Ma et al., 2007; Perri, 2020)。これら8元素(SiO_2 , TiO_2 , Al_2O_3 , Fe_2O_3 , MgO , CaO , Na_2O , K_2O)のうち、検出限界以下となったデータについては0.005 wt.%(質量パーセント)を代入して分析を行った。化学組成データに加えて、各土壌サンプルについて、その場の年平均気温、年間降水量のデータを、観測値に基づく高解像度気候データベース(WorldClim: Fick and Hijmans, 2017)から取得した。

今回用いた土壌の化学組成データの基本統計量を表1に示した。表から、各元素が含まれる割合は様々であるが、変動係数から Al_2O_3 が含まれる割合が最も安定していることがわかる。図1は土壌の化学組成データの対散布図である。花崗岩を母岩とする土壌と玄武岩を母岩とする土壌について、各元素の濃度範囲には重なりが認められ、単一元素の濃度のみから母岩を区別することが困難である一方、元素の組み合わせによっては領域の重なりが少ない組み合わせも存在することがわかる。また、 TiO_2 と Fe_2O_3 などいくつかの成分間に強い正の相関がみられる。一方で、外れ値も多数確認される。

組成データは、各成分(変数)の合計値が100%となる形式のデータである。組成データは自由度が成分の個数よりも少なくなるため、各成分の値は互いに依存しており、1つの成分が増

表1. 土壌の主要元素バルク組成データ(wt.%)の基本統計量。

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	CaO	Na_2O	K_2O
平均値	48.32	2.44	21.05	14.98	1.35	1.36	0.64	1.08
中央値	47.17	2.50	20.19	14.15	0.77	0.37	0.18	1.00
標準偏差	14.67	1.36	5.09	7.88	1.81	2.52	0.88	0.99
変動係数	0.30	0.56	0.24	0.53	1.35	1.85	1.38	0.92

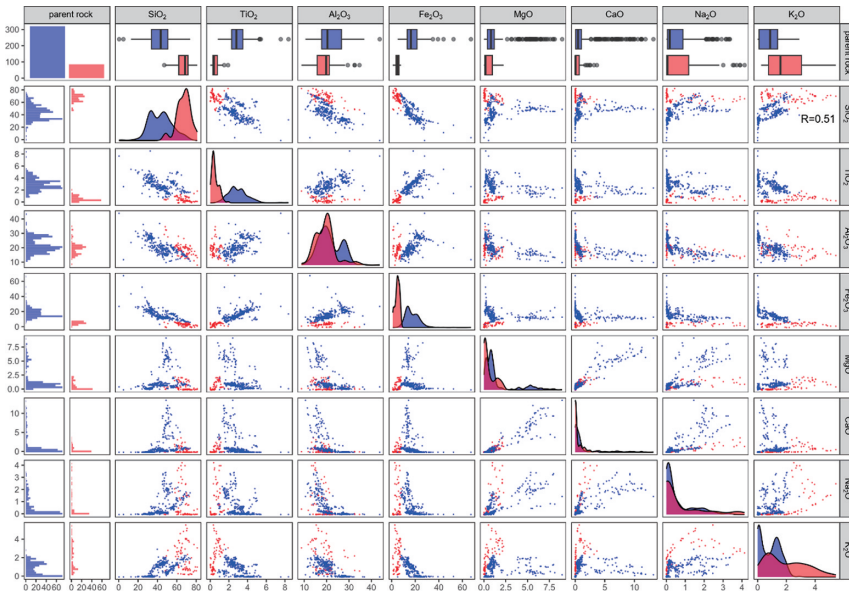


図 1. 土壌の主要元素全岩化学組成データの対散布図. 赤色が花崗岩を母岩とする試料であり、青色が玄武岩を母岩とする試料を示す。

表 2. 化学組成データ (wt.%) を Al_2O_3 で規格化したデータの基本統計量。

	SiO ₂	TiO ₂	Fe ₂ O ₃	MgO	CaO	Na ₂ O	K ₂ O
平均値	2.54	0.12	0.71	0.08	0.08	0.04	0.06
中央値	2.47	0.12	0.73	0.04	0.02	0.01	0.05
標準偏差	1.22	0.06	0.38	0.12	0.17	0.06	0.06
歪度	0.87	-0.07	4.79	2.63	3.07	1.89	1.85
尖度	1.70	0.04	55.85	7.47	10.51	3.14	4.34

加すれば、自動的に他の成分が相対的に減少する。したがって、組成データから 1 成分のみを取り出して変動を議論することは困難である (Aitchison, 1982; 太田, 2023)。また、このような閉じた性質 (成分の値の合計が一定値に制約される性質) は、通常の統計手法や多変量解析技術を直接適用する際にランク落ちの問題を引き起こすことがある。今回扱う土壌の化学組成データも組成データであるため、解析の前処理として、総和が 100% とならないような変数変換を行った。

Al_2O_3 は一般的な岩石や土壌に一定の割合以上に含まれており、多くの分析手法で精度よく定量値が得られる元素である。また、表 1 の考察で述べたように Al_2O_3 は組成データでの変動係数をもっとも小さいため、組成データ化前のデータの絶対量 (母岩に対する元素の移動量) における変動ももっとも小さいと考えられる。そのため、本研究では Al_2O_3 を規格化成分として選定した。 Al_2O_3 を用いて各元素の規格化 (= 各 8 元素の濃度/ Al_2O_3 の濃度) を行ったデータの基本統計量を表 2 に示した。 Fe_2O_3 , MgO , CaO について歪度、尖度ともに大きな値を示し

表 3. 化学組成データ(wt.%)を Al_2O_3 で規格化後対数変換を施したデータの基本統計量.

	SiO_2	TiO_2	Fe_2O_3	MgO	CaO	Na_2O	K_2O
平均値	0.80	-2.42	-0.49	-3.50	-4.50	-4.67	-3.71
中央値	0.90	-2.11	-0.31	-3.32	-3.96	-4.73	-2.99
標準偏差	0.57	0.96	0.62	1.52	2.48	1.92	1.83
歪度	-2.27	-2.23	-1.72	-0.43	-0.38	-0.03	-1.10
尖度	18.13	5.85	4.42	0.71	-0.60	-1.10	0.27

たため、 Al_2O_3 で規格化後の各変数に対数変換(= $\log(\text{各 8 元素の濃度}/\text{Al}_2\text{O}_3 \text{ の濃度})$)を施したデータの基本統計量を表 3 に示した. SiO_2 が尖度 18.13 と大きな値を示した以外は、多くの元素について歪度、尖度がともに正規分布に近づく結果となり、元素間のバランスが改善された. 以上より、本研究の主成分分析では Al_2O_3 で規格化後に対数変換をしたデータを用いて解析を行った.

3. 主成分分析によるデータの特徴の探索

3.1 土壌化学組成データの主成分分析(全データ)

土壌の化学組成データの特徴と土壌と気候条件の関係を検討するために、主要元素を Al_2O_3 で規格化後に対数変換をしたデータについて主成分分析を行った. 範囲が最大の元素は CaO で 10.5, 最小の元素は Fe_2O_3 で 5.18 という様に、元素の濃度の散らばりに偏りがあるため、分析には相関行列を用いた. 第 1 主成分は寄与率 52.2%, 第 2 主成分は寄与率 33.3% となり、第 1 主成分と第 2 主成分での累積寄与率は 85.5% であった. 各主成分の固有ベクトルを表 4 に示した. 第 1 主成分(PC1)は図 2 のように温度(A)と降水量(B)の情報を追加することで、寒冷・乾燥地域と温暖・湿潤地域を示す軸であると考えられる. CaO , Na_2O , MgO , K_2O , SiO_2 は寒冷・乾燥地域での土壌に濃集しているが、これはこれらの元素が温暖・湿潤地域においては化学風化作用のために溶脱され、土壌中で枯渇する現象を反映した結果であると考えられる. 第 2 主成分(PC2)は玄武岩母岩(Δ)と花崗岩母岩(\square)の布置からこれらの母岩の違いを意味していると考えられ、固有ベクトルからは Fe_2O_3 , TiO_2 , MgO は玄武岩を母岩にもつ土壌と関係し、 SiO_2 , K_2O は花崗岩を母岩にもつ土壌と関係していることがみられる. 事実、 Fe_2O_3 , TiO_2 , MgO は苦鉄質鉱物に濃集する元素であり、 SiO_2 , K_2O は珪長質鉱物に濃集する元素である. したがって、今回検討した土壌データにおいて、化学風化作用に加えて母岩の影響も強く残っていることが明らかとなった. また、花崗岩を母岩とする土壌について、第 1 主成分スコアの増加に伴い第 2 主成分スコアが減少する一方で、玄武岩を母岩とする土壌では第 2 主成分スコアが増加するという傾向が見られた. これは化学風化がかなり進行した試料においても、母岩タイプの判別は可能であることを強く示唆している.

3.2 土壌化学組成データの主成分分析(土壌の母岩タイプで層別して分析)

本研究で収集したデータセットは花崗岩を母岩にもつ 78 の土壌と玄武岩を母岩にもつ 327 の土壌から構成されており、中間質の組成の岩石を母岩にもつ土壌は含まれていない. 3.1 節での主成分分析の結果より、母岩の化学組成に由来する差異がデータに大きな影響を与えてい

表 4. 土壌データの相関行列を用いた主成分分析の固有ベクトルと寄与率.

	PC1	PC2
SiO ₂	-0.359	-0.377
TiO ₂	-0.112	0.580
Fe ₂ O ₃	-0.080	0.602
MgO	-0.449	0.234
CaO	-0.492	0.045
Na ₂ O	-0.491	0.049
K ₂ O	-0.409	-0.316
寄与率	52.2%	33.3%

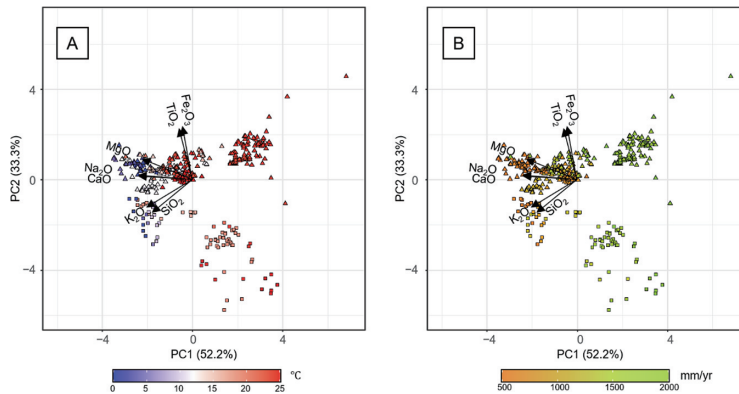


図 2. 土壌化学組成データの相関行列を用いた主成分分析結果のバイプロット (第 1 主成分と第 2 主成分). (A) 年平均気温により色分け. (B) 年間降水量により色分け.

ることが明らかとなったため、母岩の種類(花崗岩, 玄武岩)で層別して、再度、主成分分析を行った(表 5).

MgO, CaO, Na₂O, K₂O は花崗岩を母岩とする土壌, 玄武岩を母岩とする土壌で第 1 主成分に絶対値が大きい固有ベクトルを持つ。これらは風化で溶脱する元素であり、温度と降水量の情報を追加することで、寒冷・乾燥地域と温暖・湿潤地域を示す軸であると考えられる(図 3)。SiO₂ については玄武岩を母岩とする土壌において特に絶対値が大きい第 1 主成分の固有ベクトルを示す一方、花崗岩を母岩とする土壌では比較して絶対値が小さい第 1 主成分の固有ベクトルを示した。これは花崗岩に豊富に含まれる石英の他の鉱物に対する風化抵抗性を反映している可能性がある。Fe₂O₃ と TiO₂ について、玄武岩を母岩とする土壌では第 1 主成分とほとんど関係しない一方、花崗岩を母岩とする土壌では風化で溶脱する元素と同じ方向の寄与がある程度見られた。これは、強風化環境での花崗岩を母岩とする土壌でしばしば観察されるギブサイト Al(OH)₃ の濃集を伴うようなサンプルが影響している可能性が考えられる。

表 5. 母岩で層別したデータについての相関行列を用いた主成分分析の固有ベクトルと寄与率.

	花崗岩母岩土壌		玄武岩母岩土壌	
	PC1	PC2	PC1	PC2
SiO ₂	-0.203	-0.794	-0.435	-0.013
TiO ₂	-0.325	-0.176	0.022	-0.697
Fe ₂ O ₃	-0.308	-0.410	0.083	-0.700
MgO	-0.442	0.233	-0.431	-0.099
CaO	-0.433	0.277	-0.464	0.007
Na ₂ O	-0.447	0.158	-0.461	-0.106
K ₂ O	-0.420	0.122	-0.436	0.048
寄与率	62.6%	14.2%	58.8%	17.7%

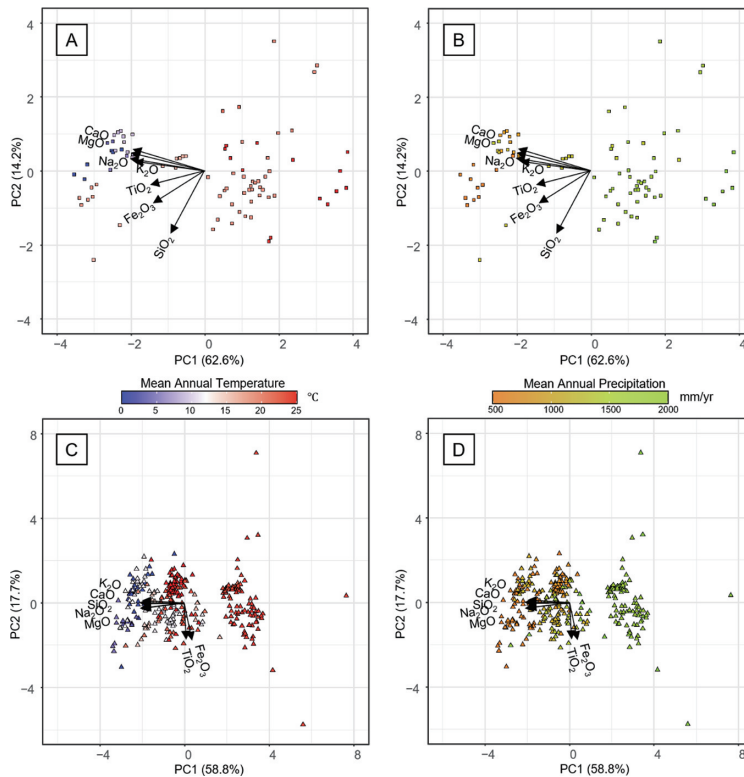


図 3. 母岩組成で層別した相関行列を用いた主成分分析結果のバイプロット(第1主成分と第2主成分). (A) (B) 母岩が花崗岩の試料についての分析結果. (C) (D) 母岩が玄武岩の試料についての分析結果.

以上より、花崗岩を母岩とする土壌、玄武岩を母岩とする土壌に関わらず、第1主成分は温暖・湿潤地域と寒冷・乾燥地域を分離する傾向が強く認められた。したがって、本研究で使用したデータセットについて、気候要因が土壌の化学組成の差異をもたらしていることが改めて示唆された。

4. 線形判別分析による土壌化学組成からの母岩の推定

母岩で層別せずに行った分析について、「 Al_2O_3 で規格化した後に対数変換をしたデータ」による主成分分析における第2主成分の固有ベクトルの結果から、 Fe_2O_3 、 TiO_2 、 MgO は玄武岩を母岩にもつ土壌と関係し、 SiO_2 、 K_2O は花崗岩を母岩にもつ土壌と関係していることがみられた。次に、玄武岩を母岩とする土壌と花崗岩を母岩とする土壌は化学組成を用いて判別可能か、またどの元素が判別に寄与するかを調べることを目的として、線形判別分析を行った。変数変換が判別精度に与える影響を検討するために、分析は「組成データ」、「 Al_2O_3 で規格化したデータ」、「 Al_2O_3 で規格化した後に対数変換をしたデータ」の3通りについて行った。判別精度は、見かけ上の誤判別率(全データに占める誤分類数の割合)によって評価した。

その結果、「組成データ」、「 Al_2O_3 で規格化したデータ」、「 Al_2O_3 で規格化した後に対数変換をしたデータ」の見かけ上の誤判別率は、それぞれ1.5%、1.5%、2.2%となり、いずれも高い精度で判別が可能であった(表6)。各元素の変動が判別関数に与える影響を、p値を用いて評価した結果を順次述べる。

表6. 花崗岩母岩土壌と玄武岩母岩土壌の各変数変換データにおける線形判別分析結果.

	LD1_wt. %		LD1_X/ Al_2O_3		LD1_log(X/ Al_2O_3)	
	判別係数	p 値	判別係数	p 値	判別係数	p 値
SiO_2	0.060	3.9×10^{-9}	0.489	1.5×10^{-9}	0.377	3.4×10^{-2}
TiO_2	-0.872	2.0×10^{-31}	-21.266	1.2×10^{-49}	-1.020	2.2×10^{-22}
Al_2O_3	0.162	2.6×10^{-13}	-	-	-	-
Fe_2O_3	-0.005	7.5×10^{-1}	-0.594	5.5×10^{-3}	-1.589	9.5×10^{-20}
MgO	-0.251	1.2×10^{-2}	-4.747	3.7×10^{-3}	0.148	7.8×10^{-2}
CaO	0.013	8.7×10^{-1}	1.143	3.6×10^{-1}	-0.461	4.8×10^{-15}
Na_2O	0.169	1.8×10^{-1}	-0.175	9.3×10^{-1}	0.244	6.5×10^{-4}
K_2O	0.342	1.0×10^{-3}	4.621	5.2×10^{-3}	0.297	3.9×10^{-6}
判別値 (花崗岩母岩)	> 5.35		> -0.528		> 4.26	
見かけ上の誤判別率	1.5 %		1.5 %		2.2 %	

LD1_wt.%: 組成データでの分析. LD1_X/ Al_2O_3 : Al_2O_3 で規格化したデータでの分析.
LD1_log(X/ Al_2O_3): Al_2O_3 で規格化後に対数変換したデータでの分析.

「組成データ」を用いた場合は、判別関数の値の増加に寄与する有意な元素 (p 値 <0.05)は、 SiO_2 , Al_2O_3 , K_2O という珩長質鉱物に濃集する元素が該当した。判別関数の値の減少に寄与する有意な元素は、 TiO_2 , MgO という苦鉄質鉱物に濃集する元素が該当した。 Fe_2O_3 は SiO_2 との相関係数が -0.84 であり、 CaO は MgO との相関係数が 0.94 であることから、これらは有意な元素から省かれ、 \pm の符号についても不安定であることが見て取れる。

「 Al_2O_3 で規格化したデータ」を用いた場合は、判別関数の値の増加に寄与する有意な元素は、 SiO_2 , K_2O という珩長質鉱物に濃集する元素が該当した。判別関数の値の減少に寄与する有意な元素は、 TiO_2 , Fe_2O_3 , MgO の苦鉄質鉱物に濃集する元素が該当した。 CaO は MgO との相関係数が 0.95 であることから有意な元素から省かれ、 \pm の符号についても不安定であることが見て取れる。

「 Al_2O_3 で規格化した後に対数変換をしたデータ」については、判別関数の値の増加に寄与する有意な元素は、 SiO_2 , Na_2O , K_2O が該当し、判別関数の値の減少に寄与する有意な元素は、 TiO_2 , Fe_2O_3 , CaO が該当した。 MgO は CaO や Na_2O との相関係数が 0.82 , 0.83 であることや CaO と Na_2O の相関係数が 0.86 であることから互いの関係が深く、 MgO が有意な元素から省かれたと考えられる。上記3つの異なるデータ前処理方法での分析について、判別精度に違いは認められず、いずれも見かけ上の誤判別率ではあるが2%程度の高い精度で母岩が判別された。

「組成データ」, 「 Al_2O_3 で規格化したデータ」および「 Al_2O_3 で規格化した後に対数変換をしたデータ」のいずれにおいても MgO , CaO , Na_2O は有意でないことや符号が反転することなどから判別分析に寄与する元素ではない可能性が高い。そこで、これら3元素を除外したデータセットで再度、線形判別分析を行った結果を表7に示した。見かけ上の誤判別率はそれぞれ1.7%, 0.2%, 2.7%となり、8元素すべてを含むデータセットと比べて、「組成データ」と「 Al_2O_3 で規格化した後に対数変換をしたデータ」では見かけ上の誤判別率がわずかに上昇し、「 Al_2O_3 で規格化したデータ」ではわずかに減少したが、見かけ上の誤判別率であるため、意味のある違いとは考えられない。よって、 MgO , CaO , Na_2O は他の主要元素データの存在下では土壤の

表7. MgO , CaO , Na_2O を除外したデータでの花崗岩母岩土壤と玄武岩母岩土壤の各変数変換データにおける線形判別分析結果.

	LD1_wt.%		LD1_X/ Al_2O_3		LD1_log(X/ Al_2O_3)	
	判別係数	p 値	判別係数	p 値	判別係数	p 値
SiO_2	0.077	5.0×10^{-16}	0.315	1.9×10^{-5}	0.522	1.5×10^{-3}
TiO_2	-0.805	1.4×10^{-28}	-21.137	3.7×10^{-49}	-1.015	2.9×10^{-22}
Al_2O_3	0.214	2.1×10^{-31}	-	-	-	-
Fe_2O_3	0.008	6.3×10^{-1}	-0.974	2.3×10^{-6}	-1.661	1.6×10^{-23}
K_2O	0.471	2.8×10^{-8}	4.737	6.1×10^{-4}	0.008	8.7×10^{-1}
判別値 (花崗岩母岩)	> 7.95		> -1.00		> 4.92	
見かけ上の誤判別率	1.7%		0.2%		2.7%	

母岩タイプの判別に寄与しないと結論づけられる。特に CaO と Na₂O は風化の初期段階から容易に溶脱し、化学風化の進行程度を最も鋭敏に反映する元素であるため、これらの元素が母岩判別に寄与しないという結果は、風化プロセスに関する既存の知見とも整合的である (Ohta et al., 2007; Buggle et al., 2011)。

5. 結論

高温・高圧の地球内部で生成された岩石は、低温・低圧の地球表層環境に晒されることで、鉱物の分解・再結晶を経て土壌を形成する。特に、温暖かつ湿潤な環境では化学風化が促進され、土壌の化学的特徴に大きな影響を与えるため、古土壌や陸源性碎屑岩は、その生成環境での気候条件を推定するための記録媒体として活用されてきた。本研究では、異なる気候帯に発達する土壌データについて、主成分分析により化学組成の特徴を探索的に解析し、また花崗岩を母岩とする土壌と玄武岩を母岩とする土壌は化学組成を用いて判別が可能であるか、判別に寄与する元素はなにかを線形判別分析により検討した。これらの解析の結果、気候要因により促進される化学風化とともに母岩も土壌の化学組成を特徴づける主要因であり、風化が進行した試料においても母岩が土壌化学組成に影響を与えていることが示された。

従来、風化度に基づく気候復元の研究においては、上部大陸地殻の平均組成である花崗岩質を後背地と想定して議論することが多かった。しかしながら、本研究の結果は、このような仮定に基づく解析では、母岩の化学組成の影響と化学風化の効果とを区別できず、その結果、気候変動と後背地の地質的变化が同時に生じた場合、それぞれの影響が混在したまま評価され、母岩の差異が気候推定値に系統的なバイアスをもたらしている可能性を示唆している。特に、巨大火成岩岩石区(LIPs)の出現や山脈隆起に伴う環境変動イベントにおいては、後背地が変化し、堆積物の母岩が変化することは十分考えられるため、母岩の変化が堆積物の化学組成に与える影響は適切に評価されるべきである。

また、本研究で用いたデータには定量限界未満の値を 0.005 wt.% として扱ったが、組成データにおけるこのような小さい値の取り扱いについては理論的基盤が十分とは言えない (Greenacre, 2021)。今後は、平方根変換 (中西, 2003) や Box-Cox 変換の利用など、定量限界未満の値や 0 値を含む組成データを適切に処理可能な解析手法の整備が求められる。

謝 辞

統計数理研究所の水田正弘教授には組成データ解析に関して日頃より貴重なご助言を賜り、原稿について有益なアドバイスをいただいた。統計数理研究所の三輪哲久教授には組成データ解析に関して日頃より貴重なご助言を賜った。また、匿名査読者 2 名と編集者のコメントにより本論文は大きく改善された。本研究の一部に JSPS 科研費 25K17503 を使用した。本研究は文部科学省事業「統計エキスパート人材育成プロジェクト」で得られた成果を発展させたものである。

参 考 文 献

- Aitchison, J. (1982). The statistical analysis of compositional data, *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**, 139–160.
- Bao, Z. and Zhao, Z. (2008). Geochemistry of mineralization with exchangeable REY in the weathering crusts of granitic rocks in South China, *Ore Geology Reviews*, **33**, 519–535.

- Barbera, V., Raimondi, S., Egli, M. and Plötze, M. (2008). The influence of weathering processes on labile and stable organic matter in Mediterranean volcanic soils, *Geoderma*, **143**, 191–205.
- Buggle, B., Glaser, B., Hambach, U., Gerasimenko, N. and Marković, S. (2011). An evaluation of geochemical weathering indices in loess–paleosol studies, *Quaternary International*, **240**, 12–21.
- Campodonico, V. A., Pasquini, A. I., Lecomte, K. L., García, M. G. and Depetris, P. J. (2019). Chemical weathering in subtropical basalt-derived laterites: A mass balance interpretation (Misiones, NE Argentina), *Catena*, **173**, 352–366.
- Caner, L., Radtke, L. M., Vignol-Lelarge, M. L., Inda, A. V., Bortoluzzi, E. C. and Mexias, A. S. (2014). Basalt and rhyo-dacite weathering and soil clay formation under subtropical climate in southern Brazil, *Geoderma*, **235–236**, 100–112.
- Deng, K., Yang, S. and Guo, Y. (2022). A global temperature control of silicate weathering intensity, *Nature Communications*, **13**, 1781.
- Dethier, D. P. and Bove, D. J. (2011). Mineralogic and geochemical changes from alteration of granitic rocks, Boulder Creek Catchment, Colorado, *Vadose Zone Journal*, **10**, 858–866.
- Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas, *International Journal of Climatology*, **37**, 4302–4315.
- Fu, W., Li, X., Feng, Y., Feng, M., Peng, Z., Yu, H. and Lin, H. (2019). Chemical weathering of S-type granite and formation of Rare Earth Element (REE)-rich regolith in South China: Critical control of lithology, *Chemical Geology*, **520**, 33–51.
- Giovanoli, R., Schnoor, J. L., Sigg, L., Stumm, W. and Zobrist, J. (1988). Chemical weathering of crystalline rocks in the catchment area of acidic Ticino lakes, Switzerland, *Clays and Clay Minerals*, **36**, 521–529.
- Greenacre, M. (2021). Compositional data analysis, *Annual Review of Statistics and Its Application*, **8**, 271–299.
- Land, M., Ingri, J. and Öhlander, B. (1999). Past and present weathering rates in northern Sweden, *Applied Geochemistry*, **14**, 761–774.
- Le Blond, J. S., Cuadros, J., Molla, Y. B., Berhanu, T., Umer, M., Baxter, P. J. and Davey, G. (2015). Weathering of the Ethiopian volcanic province: A new weathering index to characterize and compare soils, *American Mineralogist*, **100**, 2518–2532.
- Li, J., Gong, Q., Yan, T., Li, R., Liu, N. and Cen, K. (2018). Quantitative description of geochemical backgrounds of gold due to rock weathering in Jiaodong peninsula, China, *Journal of Geochemical Exploration*, **192**, 155–162.
- Liu, J., Zhou, L., Algeo, T. J., Wang, X., Wang, Q., Wang, Y. and Chen, M. (2020). Molybdenum isotopic behavior during intense weathering of basalt on Hainan Island, South China, *Geochimica et Cosmochimica Acta*, **287**, 180–204.
- Liu, W., Liu, C., Brantley, S. L., Xu, Z., Zhao, T., Liu, T., Yu, C., Xue, D., Zhao, Z. and Cui, L. (2016). Deep weathering along a granite ridgeline in a subtropical climate, *Chemical Geology*, **427**, 17–34.
- Ma, J., Wei, G., Xu, Y., Long, W. and Sun, W. (2007). Mobilization and re-distribution of major and trace elements during extreme weathering of basalt in Hainan Island, South China, *Geochimica et Cosmochimica Acta*, **71**, 3223–3237.
- Manefouet Kentsa, B. I., Kamgang Kabeyene, V., Tamen, J., Nzolang, C. and Ndjigui, P. (2023). Geochemistry of major elements in soil-saprolite of different rocks in a microclimate of the Bambouto Mountains (Cameroon), *Applied Geochemistry*, **154**, <https://doi.org/10.1016/j.apgeochem.2023.105701>.
- Mangold, N., Dehouck, E., Fedo, C., Forni, O., Achilles, C., Bristow, T., Downs, R. T., Frydenvang, J., Gasnault, O., L’Haridon, J., Le Deit, L., Maurice, S., McLennan, S. M., Meslin, P.-Y.,

- Morrison, S., Newsom, H. E., Rampe, E., Rapin, W., Rivera-Hernandez, F., Salvatore, M. and Wiens, R. C. (2019). Chemical alteration of fine-grained sedimentary rocks at Gale crater, *Icarus*, **321**, 619–631.
- 中西寛子 (2003). 都道府県別選挙得票率からわかること—統計的データ分析の例として—. *オペレーションズ・リサーチ*, **48**, 17–22.
- Norton, K. P. and Von Blanckenburg, F. (2010). Silicate weathering of soil-mantled slopes in an active Alpine landscape, *Geochimica et Cosmochimica Acta*, **74**, 5243–5258.
- 太田 亨 (2023). 『組成データ解析入門：パーセント・データの問題点と解析方法』, 朝倉書店, 東京.
- Ohta, T. and Arai, H. (2007). Statistical empirical index of chemical weathering in igneous rocks: A new tool for evaluating the degree of weathering, *Chemical Geology*, **240**, 280–297.
- Perri, F. (2018). Reconstructing chemical weathering during the Lower Mesozoic in the Western-Central Mediterranean area: A review of geochemical proxies, *Geological Magazine*, **155**, 944–954.
- Perri, F. (2020). Chemical weathering of crystalline rocks in contrasting climatic conditions using geochemical proxies: An overview, *Palaeogeography, Palaeoclimatology, Palaeoecology*, **556**, <https://doi.org/10.1016/j.palaeo.2020.109873>.
- Pokrovsky, O. S., Schott, J., Kudryavtzev, D. I. and Dupré, B. (2005). Basalt weathering in Central Siberia under permafrost conditions, *Geochimica et Cosmochimica Acta*, **69**, 5659–5680.
- Rasmussen, C. and Tabor, N. J. (2007). Applying a quantitative pedogenic energy model across a range of environmental gradients, *Soil Science Society of America Journal*, **71**, 1719–1729.
- Rasmussen, C., Southard, R. J. and Horwath, W. R. (2005). Modeling energy inputs to predict pedogenic environments using regional environmental databases, *Soil Science Society of America Journal*, **69**, 1266–1274.
- Sheldon, N. D., Retallack, G. J. and Tanaka, S. (2002). Geochemical climofunctions from North American soils and application to Paleosols across the Eocene-Oligocene boundary in Oregon, *The Journal of Geology*, **110**, 687–696.
- Talabi, A. O. (2015). Weathering of meta-igneous rocks in parts of the basement terrain of Southwestern Nigeria: Implications on groundwater occurrence, *International Journal of Scientific and Research Publications*, **5**, 1–17.
- USDA (2009). National soil characterization data. Lincoln, NE, Soil Survey Laboratory, National Soil Survey Center, Soil Conservation Service, <http://soils.usda.gov/survey/nscd/> (最終アクセス日 2025 年 10 月 1 日).
- White, A. F., Blum, A. E., Schulz, M. S., Vivit, D. V., Stonestrom, D. A., Larsen, M., Murphy, S. F. and Eberl, D. (1998). Chemical weathering in a tropical watershed, Luquillo Mountains, Puerto Rico: I. Long-term versus short-term weathering fluxes, *Geochimica et Cosmochimica Acta*, **62**, 209–226.
- White, A. F., Blum, A. E., Schulz, M. S., Huntington, T. G., Peters, N. E. and Stonestrom, D. A. (2002). Chemical weathering of the Panola Granite: Solute and regolith elemental fluxes and the weathering rate of biotite, *Water Rock Interactions, Ore Deposits, and Environmental Geochemistry: A Tribute to David A. Crerar* (eds. R. Hellmann and S. A. Wood), Special Publication 7, 37–59, The Geochemical Society, Alexandria, Virginia.
- Wu, Y., Li, X., Gong, Q., Wu, X., Yao, N., Peng, C., Chao, Y., Wang, X. and Pu, X. (2021). Test and application of the geochemical lithogene on weathering profiles developed over granitic and basaltic rocks in China, *Applied Geochemistry*, **128**, <https://doi.org/10.1016/j.apgeochem.2021.104958>.
- Yuan, W., Gong, Y., Chen, J., Wang, Z., Huang, F., Yang, X., Chetelat, B., Teng, H. and Schott, J. (2022). Gallium isotope constraints on the intense weathering of basalt, *Geochimica et Cosmochimica Acta*, **333**, 22–38.

- Yusoff, Z. M., Ngwenya, B. T. and Parsons, I. (2013). Mobility and fractionation of REEs during deep weathering of geochemically contrasting granites in a tropical setting, Malaysia, *Chemical Geology*, **349**, 71–86.

Geochemical Signatures of Soils in Relation to Parent Rock and Climatic Conditions: A Principal Component and Linear Discriminant Approach

Tenichi Cho¹ and Hiroko Nakanishi²

¹Center for Data Science, Waseda University

²The Institute of Statistical Mathematics

Soil formation is strongly influenced by climatic factors such as temperature and precipitation, suggesting that the chemical composition of soils can serve as a proxy for reconstructing past climatic conditions. However, soil chemistry is not solely governed by climate, and the influence of parent rock properties may become pronounced particularly under conditions of limited chemical weathering. To better understand how soil chemistry is related to climatic conditions and parent rock types, we performed multivariate analyses on chemical composition data from soils formed under diverse climatic regimes. Principal component analysis (PCA) revealed that soil chemical composition is strongly influenced by climatic conditions and, irrespective of the degree of chemical weathering, by parent rock composition. Linear discriminant analysis (LDA) further demonstrated that parent rock type can be accurately identified based on soil chemical composition. Our findings demonstrate that parent rock characteristics are preserved in soil composition even under conditions of intense chemical weathering, and highlight the necessity of incorporating parent rock composition into climate estimation models to reduce biases associated with variations in source rock composition.

脳波位相と正答確率の関数推定に向けたロジット変換を用いたガウス過程回帰

小野島 隆之¹・神保 雅一²

(受付 2025 年 6 月 28 日；改訂 9 月 19 日；採択 9 月 25 日)

要 旨

人間の脳活動は脳波として計測でき、その状態が課題成績に影響を与えることから、特定の周波数帯域の活動の脳波位相が知覚に関係すると考えられている。この関係を推定するには、被験者ごとに対象とする周波数帯域の脳波の位相と課題成績の関数関係を推定する必要がある。近年では、脳波位相依存刺激法などの脳波位相の状態に基づく刺激法も提案されており、脳波位相と脳機能の関係を被験者ごとに柔軟に推定できる手法が求められている。本研究では、刺激提示に対する各試行の二値応答(正答/誤答あるいは検出/未検出)を観測データとし、脳波位相を説明変数とした正答確率を柔軟に推定する枠組みを構築した。具体的には、ロジット変換をリンク関数とし、周期カーネルを用いたガウス過程回帰によって脳波位相に依存して変化する正答確率を推定するベルヌーイ・ロジットモデルを構築した。このモデルにおけるサンプルサイズの増加に伴うグラム行列の退化と計算時間の増加に対処するためにモデルを改良し、新たに二項・ロジットモデルを提案した。本研究は、脳波位相と正答確率の関数関係を推定するための2つの統計モデルを構築し、数値実験により検証したものである。

キーワード：脳波、ガウス過程回帰、ロジット変換、周期カーネル。

1. はじめに

1.1 研究背景

人間の脳活動はしばしば周期的な性質を持ち、頭皮脳波(electroencephalogram: EEG)として非侵襲的に計測することが可能である。この周期的な脳活動は外部刺激や異なる脳領域の活動と同期するとされ、脳機能において重要な神経基盤として議論されている。ここでいう脳波位相とは、特定の周波数帯域の波形に対する瞬間的な位相を指し、刺激提示のタイミングとの関係で知覚や反応に影響を及ぼすと考えられている。

近年では、特定の脳波の位相がある値のときにのみ刺激を与える脳波位相依存刺激法が提案されている(Zrenner et al., 2016)。特に経頭蓋磁気刺激法(Transcranial Magnetic Stimulation: TMS)と組み合わせた場合には、刺激に対する応答が位相依存的に変化することが報告されている(Zrenner et al., 2018)。しかし、脳波位相依存刺激法における応答の位相依存性や最適な位相には個人差があり(Torrecillos et al., 2020; Zrenner et al., 2023; Kirchoff et al., 2024)、対

¹ 滋賀大学 データサイエンス・AI イノベーション研究推進センター：〒522-8522 滋賀県彦根市馬場1丁目1-1; takayuki-onojima@biwako.shiga-u.ac.jp

² 統計数理研究所：〒190-8562 東京都立川市緑町10-3; jimbo@ism.ac.jp

象とする脳波やその位相の状態を事前を選択する方法が実験を遂行する上で課題となっている。このような課題に対し、TMS研究などではベイズ最適化を用いて少数試行から最適な刺激位相を推定する枠組みが提案されている (Tervo et al., 2020, 2022; Kirchoff et al., 2024)。ベイズ最適化は、目的関数の不確実性推定を活用して効率的に次の刺激条件を選択できるため、試行数の限られる行動実験や脳刺激実験で有効である。

一方で、視覚刺激を脳波位相に基づいて提示する実験課題 (Shakeel et al., 2021) や、入眠過程において残存する α 帯域 (8–12Hz) の脳波位相をリアルタイムで検出し、そのタイミングに合わせて音刺激を与える刺激法 (Hebron et al., 2024) も報告されている。これらの視覚・聴覚の感覚モダリティでは、刺激提示直前の脳波位相が課題成績に影響を与えることが報告されており (Mathewson et al., 2009; Busch and VanRullen, 2010; Ng et al., 2012; Zoefel and VanRullen, 2015; Onojima et al., 2017)、脳波と認知機能の関係解明において瞬時位相の状態が重要であることが示唆されている。

このように脳波位相と課題成績の関係や、脳波位相を基にした刺激法などが提案されてきたが、被験者ごとの脳波位相と課題成績の関数関係は不明瞭なままである。特に視覚や聴覚などの感覚刺激を用いた行動実験において、課題成績と脳波位相の関数関係を推定する手法は十分に確立されていない。したがって、この関数を推定するとともに、その信用区間を求めることが重要である。

1.2 本研究の目的

本研究の目的は、脳波位相と課題の応答の関係を柔軟に推定できる統計モデルの構築である。ここで対象とする課題は Yes/No 課題や二者択一課題 (two alternative forced choice task: 2AFC) などの 2 値応答を伴うものであり、各試行において脳波位相 $x \in (-\pi, \pi]$ と、正答 $z(x) = 1$ または誤答 $z(x) = 0$ が観測される。このような課題における正答確率 $p(x)$ の形状は被験者や実験条件によって異なる可能性があるため、特定の関数形を仮定せずに柔軟に表現できるノンパラメトリックな手法が望ましい。

そこで本稿では、ロジット変換をリンク関数とし、周期カーネルを用いたガウス過程回帰 (Logit-Linked Periodic Gaussian Process Regression) によって $p(x)$ を推定する枠組みを構築する (ベルヌーイ・ロジットモデル)。また、この推定手法を用いる際にサンプルサイズの増加に伴って起こるグラム行列の退化と計算時間の増加に対する対策として、新たに二項・ロジットモデルを提案する。本研究は、脳波位相と課題成績の関数関係を推定するための 2 つの統計モデルを構築し、数値実験により比較した。

2. 統計モデルと推定手法

初めに、本研究で推定する関数について説明する。ここでの脳波位相とは、計測された脳波から特定の周波数帯域を抽出し、ヒルベルト変換などを用いて求めた瞬時位相であり、 $(-\pi, \pi]$ の範囲の値を持つ (図 1a 下段)。被験者に課題を提示する場合、刺激直前の脳波位相の状態が課題成績に影響を与えることが報告されている (Mathewson et al., 2009; Busch and VanRullen, 2010; Ng et al., 2012; Zoefel and VanRullen, 2015; Onojima et al., 2017)。これらの研究では脳波計測を行いながら視覚刺激 (ドットや図形) や聴覚刺激 (ビーブ音や音声) を被験者に提示し、その刺激を検出できたか、または、正しく判別できたのかを二値応答として計測している。その上で実験後に刺激提示前の脳波の位相が課題成績に影響を与えるのかを検証している。本研究では、脳波位相と課題の二値応答の正答確率の関係を被験者ごとに関数として推定することを目的とする (図 1b)。

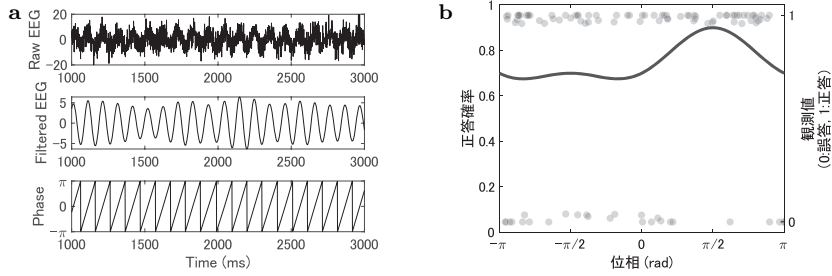


図 1. 瞬時位相と推定する関数の概念図：(a)生の EEG 信号(Raw EEG, 上段), 帯域通過フィルタ後の EEG 信号(Filtered EEG, 中段), およびヒルベルト変換によって推定された瞬時位相(Phase, 下段)の例. (b)脳波位相と正答確率の関数. 位相に依存して変動する正答確率(実線)を, 観測値(灰色の点)から推定することを想定している. この正答確率の関数形は, 3章で説明する数値実験で用いるものと同じであり, シミュレーションにより作成したものである. 観測値は描画時の重なりを避けるため, 縦方向に小さなジッターを加えて表示している. 左軸は正答確率(未知), 右軸は 0/1 の観測値を表す.

本研究が対象とする課題は, 二者択一課題(2AFC)や Yes/No 課題, 検出課題などの二値応答を伴うものであり, 各試行において脳波位相 $x \in (-\pi, \pi]$ と, 正答 $z(x) = 1$ または誤答 $z(x) = 0$ が観測される. Yes/No 課題や検出課題の場合は正答/誤答の代わりに検出/未検出の応答を伴う. この場合は正答確率ではなく検出率(hit rate)や誤報率(false alarm rate)などを推定することになるが, 本研究ではこれらの確率を包括的に正答確率として扱う. 被験者ごとに n 回の刺激を与え刺激に対する応答と刺激提示前の脳波位相を計測することを想定する. 実験試行ごとに観測される $z(x)$ は, 位相依存の正答確率 $p(x)$ に基づくと仮定する.

本稿では, ロジット変換をリンク関数とし, 周期カーネルを用いたガウス過程回帰(Logit-Linked Periodic Gaussian Process Regression)によって $p(x)$ を推定する手法を構築する. 本研究ではベイズ最適化や適応的実験設計などへの応用を目指すものであるが, 推定手法自体の構築と評価に焦点を当て, ベルヌーイ・ロジットモデルと二項・ロジットモデルを構築し, 数値実験により推定結果を比較した.

2.1 ロジット変換によるガウス過程回帰

位相 x は $\Omega = (-\pi, \pi]$ を定義域とし, 観測値 $z(x)$ は $\{0, 1\}$ の 2 値をとるとする. この $z(x)$ は位相 x ごとに $p(x)$ が与えられたもとの, 独立なベルヌーイ分布 $\text{Bin}(1, p(x))$ に従って得られる観測値であるとする. ここで $p(x)$ は位相 x ごとの正答確率であり $[0, 1]$ に値をとる連続的な周期関数とする. $p(x)$ のロジット変換

$$g(x) = \log \frac{p(x)}{1 - p(x)}$$

は Ω 上で $(-\infty, \infty)$ に値をとる定常ガウス過程 $GP(\mu_0, k(\cdot, \cdot))$ に従うとする. ただし, k はカーネル共分散関数であり, $p(x)$ と $g(x)$ は 1 対 1 に対応しており, 逆変換は

$$p(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

と書ける. したがって, $p(x)$ は $[0, 1]$ に値を取る確率過程であり, その見本関数が正答確率関数である. ロジット変換をリンク関数とするガウス過程回帰の統計モデルの推定・予測については Rue et al. (2009); Wang and Shi (2014); Sofro and Oktaviarina (2018) で

議論されており，本稿では $\Omega = (-\pi, \pi]$ 上で周期カーネルを用いた同様のモデルを構築し，このベルヌーイ分布とロジット変換を用いたガウス過程回帰をベルヌーイ・ロジットモデルと呼ぶこととする．位相 x_i ($i = 1, \dots, n$) での観測値 $z_i = z(x_i)$ は正答確率 $p_i = p(x_i)$ のベルヌーイ分布より得られる．また， g_i を p_i のロジット変換とする．ここで $\mathbf{z}_n = (z_1, \dots, z_n)$, $\mathbf{g}_n = (g_1, \dots, g_n)$, $\mathbf{p}_n = (p_1, \dots, p_n)$ とし，標本を $D_n = ((z_1, x_1), \dots, (z_n, x_n))$ とする．我々は任意の x に対して $p(x)$ の事後確率密度関数 $f(p(x) | D_n)$ を得たい．そのためにロジット変換 $g(x)$ の事後確率密度関数 $f(g(x) | D_n) = f(g(x) | \mathbf{z}_n)$ を求める．まず， z_i は p_i が与えられたもとで，独立にベルヌーイ分布 $\text{Bin}(1, p_i)$ に従うため，

$$f(\mathbf{z}_n | \mathbf{g}_n) = f(\mathbf{z}_n | \mathbf{p}_n) = \prod_{i=1}^n p_i^{z_i} (1 - p_i)^{1 - z_i} = \prod_{i=1}^n \frac{e^{g_i z_i}}{1 + e^{g_i}}$$

である．また， \mathbf{g}_n の事前分布は n 変量正規分布に従うことを仮定しているため，

$$f(\mathbf{g}_n) = \frac{1}{(2\pi)^{n/2} |K_n|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{g}_n - \mu_0 \mathbf{1})^\top K_n^{-1} (\mathbf{g}_n - \mu_0 \mathbf{1}) \right]$$

であり， $K_n = (k(x_i, x_j))$ の各要素はカーネル関数である．また， μ_0 は事前分布として考えられる正答確率 p_0 をロジット変換した値 $\mu_0 = \log \frac{p_0}{1 - p_0}$ である．ここでは次の周期カーネルを用いる：

$$(2.1) \quad k(x_i, x_j) = \theta_1 \exp \left[-\frac{(1 - \cos(x_i - x_j))}{\theta_2} \right] = \theta_1 \exp \left[-\frac{2 \sin^2(\frac{x_i - x_j}{2})}{\theta_2} \right]$$

θ_1, θ_2 はハイパーパラメータである．このとき， \mathbf{g}_n の事後分布は

$$f(\mathbf{g}_n | \mathbf{z}_n) = \frac{f(\mathbf{z}_n | \mathbf{g}_n) f(\mathbf{g}_n)}{f(\mathbf{z}_n)} \\ \propto \left(\prod_{i=1}^n \frac{e^{g_i z_i}}{1 + e^{g_i}} \right) \frac{1}{(2\pi)^{n/2} |K_n|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{g}_n - \mu_0 \mathbf{1})^\top K_n^{-1} (\mathbf{g}_n - \mu_0 \mathbf{1}) \right]$$

である．ここで，

$$(2.2) \quad h(\mathbf{g}_n) = \left(\prod_{i=1}^n \frac{e^{g_i z_i}}{1 + e^{g_i}} \right) \exp \left[-\frac{1}{2} (\mathbf{g}_n - \mu_0 \mathbf{1})^\top K_n^{-1} (\mathbf{g}_n - \mu_0 \mathbf{1}) \right]$$

とおくと， $f(\mathbf{g}_n | \mathbf{z}_n) \propto h(\mathbf{g}_n)$ であり， \mathbf{g}_n の事後分布は正規分布とはならない．通常，事後分布を求めるにはマルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo methods: MCMC) などの手法を用いるが⁸，MCMC では状態空間の次元 n はサンプルサイズとともに増加し，計算量もそれともなって増加する (Rue et al., 2009)．また，本研究では将来的にベイズ最適化と組み合わせる際にリアルタイムで高速で関数推定を行うことを想定しているため，ラプラス近似により \mathbf{g}_n の事後分布を正規分布で近似する (ラプラス近似の精度については Rue et al., 2009; Wang and Shi, 2014 及び付録 A を参照)． $h(\mathbf{g}_n)$ の対数をとると，

$$(2.3) \quad \log h(\mathbf{g}_n) = \sum_{i=1}^n g_i z_i - \sum_{i=1}^n \log(1 + e^{g_i}) - \frac{1}{2} (\mathbf{g}_n - \mu_0 \mathbf{1})^\top K_n^{-1} (\mathbf{g}_n - \mu_0 \mathbf{1})$$

であり，(2.3) の極大点，すなわち

$$\nabla \log h(\mathbf{g}_n) = \mathbf{z}_n - \mathbf{p}_n - K_n^{-1} (\mathbf{g}_n - \mu_0 \mathbf{1}) = \mathbf{0}$$

の解を \mathbf{g}_n^* とする． \mathbf{g}_n^* のまわりでラプラス近似を行うと \mathbf{g}_n の条件付き分布は

$$f(\mathbf{g}_n | \mathbf{z}_n) \propto \exp \left[-\frac{1}{2} (\mathbf{g}_n - \mathbf{g}_n^*)^\top A_n^* (\mathbf{g}_n - \mathbf{g}_n^*) \right]$$

となり，多変量正規分布 $\mathcal{N}(\mathbf{g}_n^*, (A_n^*)^{-1})$ で近似できる．ただし， $\mathbf{g}_n^* = (g_1^*, \dots, g_n^*)$ であり，

$$p_i^* = \frac{e^{g_i^*}}{1 + e^{g_i^*}}$$

$$A_n^* = -\text{diag}(p_1^*(1-p_1^*), \dots, p_n^*(1-p_n^*)) + K_n^{-1}$$

である．

2.2 新たな x に対する $p(x)$ の推定

本稿ではベイズ最適化を用いた実験を想定し， \mathbf{z}_n が与えられたもとの，新たな x_{n+1} に対する $p_{n+1} = p(x_{n+1})$ の分布を求めたい．そのために新たな x_{n+1} に対する $g_{n+1} = g(x_{n+1})$ の分布を求める．そこで，Wang and Shi (2014) と同様に g_{n+1} の事後分布 $f(g_{n+1} | \mathbf{z}_n)$ を求め，対応する p_{n+1} の信用区間を求める．まず， $f(g_{n+1} | \mathbf{g}_n)$ は多変量正規分布 $\mathcal{N}(\mu_{n+1} \mathbf{1}, K_{n+1})$ の条件付き分布であり， $\mathcal{N}(\mu_{n+1}(\mathbf{g}_n), \sigma_{n+1}^2)$ となる．ただし，

$$\mu_{n+1}(\mathbf{g}_n) = \mathbb{E}[g_{n+1} | \mathbf{g}_n] = \mu_0 + \mathbf{k}_{n+1}^\top K_n^{-1} (\mathbf{g}_n - \mu_0 \mathbf{1})$$

$$\sigma_{n+1}^2 = \text{Var}[g_{n+1} | \mathbf{g}_n] = k(x_{n+1}, x_{n+1}) - \mathbf{k}_{n+1}^\top K_n^{-1} \mathbf{k}_{n+1}$$

$$\mathbf{k}_{n+1} = (k(x_1, x_{n+1}), \dots, k(x_n, x_{n+1}))^\top$$

である．ここで，

$$\varepsilon_1 = g_{n+1} - \mu_{n+1}(\mathbf{g}_n) = g_{n+1} - \mu_0 - \mathbf{k}_{n+1}^\top K_n^{-1} (\mathbf{g}_n - \mu_0 \mathbf{1})$$

とおくと， $\varepsilon_1 \sim \mathcal{N}(0, \sigma_{n+1}^2)$ に従う．

次に， $f(\mathbf{g}_n | \mathbf{z}_n)$ が $\mathcal{N}(\mathbf{g}_n^*, (A_n^*)^{-1})$ で近似できるので， \mathbf{z}_n が得られたもとので，

$$\varepsilon_2 = \mathbf{g}_n - \mathbf{g}_n^* \sim \mathcal{N}(\mathbf{0}, (A_n^*)^{-1})$$

したがって g_{n+1} は

$$g_{n+1} = \mu_0 + \mathbf{k}_{n+1}^\top K_n^{-1} (\mathbf{g}_n^* - \mu_0 \mathbf{1} + \varepsilon_2) + \varepsilon_1$$

と書ける．また，

$$\begin{aligned} \text{Cov}[\mathbf{g}_n, \varepsilon_1] &= \mathbb{E}[(\mathbf{g}_n - \mu_0 \mathbf{1}) \varepsilon_1] \\ &= \mathbb{E}[(\mathbf{g}_n - \mu_0 \mathbf{1})(g_{n+1} - \mu_0 - \mathbf{k}_{n+1}^\top K_n^{-1} (\mathbf{g}_n - \mu_0 \mathbf{1}))] \\ &= \mathbf{k}_{n+1} - K_n K_n^{-1} \mathbf{k}_{n+1} = \mathbf{0} \end{aligned}$$

であるから， ε_1 と \mathbf{g}_n は無相関(正規性より独立)である．したがって ε_1 と ε_2 も独立であり，また， \mathbf{z}_n と ε_1 も独立であり， $\mathbb{E}[\varepsilon_1 | \mathbf{z}_n] = \mathbb{E}[\varepsilon_1] = 0$ である．ゆえに， $\mathbb{E}[\varepsilon_2 | \mathbf{z}_n] = \mathbf{0}$ に注意すると，

$$\begin{aligned} \mathbb{E}[g_{n+1} | \mathbf{z}_n] &= \mu_0 + \mathbf{k}_{n+1}^\top K_n^{-1} (\mathbf{g}_n^* - \mu_0 \mathbf{1}) = \mu_{n+1}(\mathbf{g}_n^*) \\ \text{Var}[g_{n+1} | \mathbf{z}_n] &= \text{Var}[\mathbf{k}_{n+1}^\top K_n^{-1} \varepsilon_2 + \varepsilon_1] \\ &= \mathbf{k}_{n+1}^\top K_n^{-1} \text{Var}[\varepsilon_2] K_n^{-1} \mathbf{k}_{n+1} + \text{Var}[\varepsilon_1] \\ &= \mathbf{k}_{n+1}^\top K_n^{-1} (A_n^*)^{-1} K_n^{-1} \mathbf{k}_{n+1} + \sigma_{n+1}^2 \end{aligned}$$

となり, $f(g_{n+1} | z_n)$ は正規分布

$$\mathcal{N}(\mu_0 + \mathbf{k}_{n+1}^\top K_n^{-1} (\mathbf{g}_n^* - \mu_0 \mathbf{1}), \mathbf{k}_{n+1}^\top K_n^{-1} (A_n^*)^{-1} K_n^{-1} \mathbf{k}_{n+1} + \sigma_{n+1}^2)$$

で近似できる. これを用いて g_{n+1} の期待値と信用区間を求め, ロジット変換の逆変換によって p_{n+1} の期待値と信用区間を求める.

2.3 グラム行列の退化と正則化

本稿では, 定義域 $\Omega = (-\pi, \pi]$ 上の周期カーネルのグラム行列を考えている. 3章以降で行う数値実験では, サンプルサイズ n の位相 x を一様分布 $U(-\pi, \pi]$ に従って生成し, x_1, \dots, x_n を得ることで, グラム行列 K_n を構成する. しかし, x_i と x_j が近い値を取る場合, グラム行列が退化し, 逆行列が計算できなくなる可能性がある. 特にサンプルサイズが大きくなると, 定義域内に多くの点が集中するため, この問題が顕著になる. また, ベイズ最適化のように関数の最大値などを探索する手法では, 特定の位相 x の周辺にサンプルが集中することがあり, さらに退化のリスクが高まる. このため, 2.1 および 2.2 節で述べた手法を用いる際には, カーネルのグラム行列の退化を想定しておく必要がある.

まず, このような数値的な不安定性を評価するために, 本稿ではカーネルのグラム行列 K の 1-ノルムに基づく条件数の逆数を指標とした(付録B参照). この指標 $\kappa_1^{-1}(K)$ が小さい場合(例えば $\kappa_1^{-1}(K) < 10^{-10}$), 行列 K は数値的に特異または準特異と判断される. 図2に, 周期カーネルについて, 区間 $\Omega = (-\pi, \pi]$ におけるカーネル関数とサンプルサイズの増加に対する条件数の逆数の変化を示す. 周期カーネルの θ_2 が大きくなると x と x' が離れていても相関が大きくなる. さらに, サンプルサイズが数十から百以上になると, 数値的に不安定になる傾向が見られる(図2右参照). カーネル関数のグラム行列 K_n の退化の問題は周期カーネルのように定義域 Ω が有界区間の場合に顕著に現れると思われる. この問題に対して2つの方法でグラム行列の退化の対策を行なった.

- 正則化項を加えることによる退化の防止
- 位相 Ω のビン分けと二項分布を用いた二項・ロジットモデルの構築

まず1つ目の方法では, グラム行列の対角要素に微小な正則化項 εI を加えて退化を防ぐ:

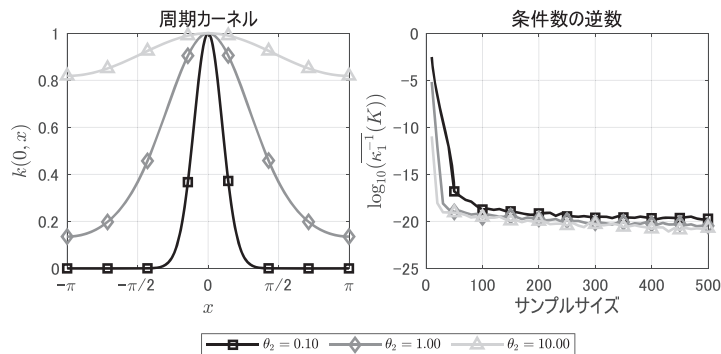


図2. ハイパーパラメータごとの周期カーネル(Circular)の相関関数とグラム行列の条件数の逆数の平均値: 左図は $\theta_1 = 1$ とし, $\theta_2 = 0.1, 1, 10$ のときの周期カーネル関数の形状を示す. 右図は各カーネル関数に対するグラム行列 K における条件数の逆数の10回の試行結果の平均値 $\overline{\kappa_1^{-1}(K)}$ の変化を示す. サンプルサイズが増加するに連れて, 条件数の逆数が小さくなる傾向が見られる.

$$K'_n = K_n + \varepsilon I \quad (\varepsilon = 10^{-6})$$

ここで I は n 次の単位行列である．この正則化により，グラム行列の逆行列が求まり， p_{n+1} の期待値および信用区間の計算が数値的に安定して行えるようになる．ベルヌーイ・ロジットモデルではこの方法でグラム行列の退化を防止している．

2つ目の方法は，次節で詳しく説明するが， Ω をいくつかのビンに分け，ビンごとの正答数が二項分布に従うと仮定することで，サンプルサイズの増加によるグラム行列の退化が発生しないモデルを構築する方法である．

2.4 z_i が二項分布に従う場合

ロジット変換をリンク関数とするガウス過程回帰を用いることで，位相と正答確率の関係を推定できるが，2.3節で述べたようにサンプルサイズが増加すると計算時間の増加とグラム行列の退化の問題がある．そこで， Ω を b 個のビン c_1, \dots, c_b に等間隔に分割する．各ビン $c_i = (c_{iL}, c_{iU}]$ に対して複数回の標本が得られる場合，(i) c_i に属する標本の平均値(あるいは中央値)をビン c_i の位相 x_i とする，あるいは，(ii) 区間中央 $\frac{c_{iL} + c_{iU}}{2}$ を位相 x_i とする．(i)の方が(ii)の代表値の決め方より推定精度が高くなることが期待される．一方，リアルタイムでの高速計算が必要な場合(ii)の方法を用いると事前に逆行列を計算しておく利点があり，本稿では代表値として(ii)の方法を採用する． m_i をその標本のサイズ，その中の正答数を z_i とし，標本を $D'_b = \{(z_1, m_1, x_1), (z_2, m_2, x_2), \dots, (z_b, m_b, x_b)\}$ とする．このとき， z_i は二項分布 $\text{Bin}(m_i, p(x_i))$ に従うため，

$$(2.4) \quad f(\mathbf{z}_b | \mathbf{g}_b) = f(\mathbf{z}_b | \mathbf{p}_b) = \prod_{i=1}^b \binom{m_i}{z_i} p_i^{z_i} (1-p_i)^{m_i-z_i} = \prod_{i=1}^b \binom{m_i}{z_i} \frac{e^{g_i z_i}}{(1+e^{g_i})^{m_i}}$$

である．ここで g_i はベルヌーイ・ロジットモデルのときと同様にガウス過程に従い，観測値 z_b が得られたもとで \mathbf{g}_b の事後分布は

$$\begin{aligned} f(\mathbf{g}_b | \mathbf{z}_b) &= \frac{f(\mathbf{z}_b | \mathbf{g}_b) f(\mathbf{g}_b)}{f(\mathbf{z}_b)} \\ &= \frac{1}{f(\mathbf{z}_b)} \left(\prod_{i=1}^b \binom{m_i}{z_i} \frac{e^{g_i z_i}}{1+e^{g_i}} \right) \frac{1}{(2\pi)^{b/2} |K_b|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{g}_b - \mu_0 \mathbf{1})^\top K_b^{-1} (\mathbf{g}_b - \mu_0 \mathbf{1}) \right] \end{aligned}$$

である． $K_b = (k(x_i, x_j))$ の各要素はカーネル関数であり，カーネル関数は周期カーネルを用いている．ここで，

$$(2.5) \quad h(\mathbf{g}_b) = \left(\prod_{i=1}^b \binom{m_i}{z_i} \frac{e^{g_i z_i}}{1+e^{g_i}} \right) \exp \left[-\frac{1}{2} (\mathbf{g}_b - \mu_0 \mathbf{1})^\top K_b^{-1} (\mathbf{g}_b - \mu_0 \mathbf{1}) \right]$$

とおくと， $f(\mathbf{g}_b | \mathbf{z}_b) \propto h(\mathbf{g}_b)$ となる．この \mathbf{g}_b をベルヌーイ・ロジットモデルの時と同様に，ラプラス近似により正規分布で近似する． $h(\mathbf{g}_b)$ の対数

$$(2.6) \quad \log h(\mathbf{g}_b) = \sum_{i=1}^b g_i z_i - \sum_{i=1}^b m_i \log(1+e^{g_i}) - \frac{1}{2} (\mathbf{g}_b - \mu_0 \mathbf{1})^\top K_b^{-1} (\mathbf{g}_b - \mu_0 \mathbf{1})$$

の極大点，すなわち

$$\nabla \log h(\mathbf{g}_b) = \mathbf{z}_b - M_b \mathbf{p}_b - K_b^{-1} (\mathbf{g}_b - \mu_0 \mathbf{1}) = \mathbf{0}$$

の解を \mathbf{g}_b^{**} とする．なお， $M_b = \text{diag}(m_1, \dots, m_b)$ とする． \mathbf{g}_b^{**} のまわりでラプラス近似を行う

と \mathbf{g}_b の事後分布は

$$f(\mathbf{g}_b | \mathbf{z}_b) \propto \exp \left[-\frac{1}{2} (\mathbf{g}_b - \mathbf{g}_b^{**})^\top A_b^{**} (\mathbf{g}_b - \mathbf{g}_b^{**}) \right]$$

となり、多変量正規分布 $\mathcal{N}(\mathbf{g}_b^{**}, (A_b^{**})^{-1})$ で近似できる。ただし、

$$A_b^{**} = -\text{diag}(m_1 p_1^{**}(1-p_1^{**}), \dots, m_b p_b^{**}(1-p_b^{**})) + K_b^{-1}$$

である。この事後分布を用いることで、新たな x_{n+1} に対する $p_{n+1} = p(x_{n+1})$ を 2.2 節と同様の手順で推定することが可能であり、 $f(g_{n+1} | \mathbf{z}_b)$ は

$$f(g_{n+1} | \mathbf{z}_b) \sim \mathcal{N}(\mu_0 + \mathbf{k}_{n+1}^\top K_b^{-1} (\mathbf{g}_b^{**} - \mu_0 \mathbf{1}), \mathbf{k}_{n+1}^\top K_b^{-1} (A_b^{**})^{-1} K_b^{-1} \mathbf{k}_{n+1} + \sigma_{n+1}^2)$$

である。ただし

$$\begin{aligned} \mathbf{k}_{n+1} &= (k(x_1, x_{n+1}), \dots, k(x_b, x_{n+1}))^\top \\ \sigma_{n+1}^2 &= \text{Var}[g_{n+1} | \mathbf{g}_b] = k(x_{n+1}, x_{n+1}) - \mathbf{k}_{n+1}^\top K_b^{-1} \mathbf{k}_{n+1} \end{aligned}$$

である。この g_{n+1} の事後分布から、 p_{n+1} の期待値と信用区間を推定することが可能である。この手法を二項・ロジットモデルと呼ぶこととする。この場合はサンプルサイズによらずカーネルのグラム行列 K_b は一定であり、計算量の増加や退化の懸念は生じない。ただし、 b を大きく設定すると K_b が退化する可能性がある。これを避けるためには周期カーネルのパラメータ θ_2 を小さく取る必要がある。例えば、各ビンの代表値を区間の中央とする場合、 $b = 12$ のときは $\theta_2 = 1$ でも退化しないが、 $b = 30$ のときは $\theta_2 = 1$ だと退化し、 $\theta_2 = 0.25$ だと退化しない。また、各ビンの代表値を平均値とする場合も同様の結果が得られる。

3. 数値実験

本研究では、ベルヌーイ・ロジットモデルを構築し、2.3 節(図 2)においてグラム行列が退化しやすいことを確認した。その対処として二項・ロジットモデルを新たに構築した。3 章では、これら 2 つのモデルの推定手法を比較・検証するための数値実験について説明する。数値実験の目的は、(i) 提案手法により関数 $p(x)$ を推定することができるかを確認すること、(ii) サンプルサイズの増加に伴う推定結果の変化を明らかにすること、(iii) g_{n+1} の期待値および 95% 信用区間から導かれる $p_{n+1} = p(x_{n+1})$ の信用区間を両モデルで比較することである。検証の流れは以下の通りである。まず、4 章では両モデルにおけるハイパーパラメータの選択方法を説明し、予備実験を想定した数値実験を行い、ハイパーパラメータの選択を行う。次に、5 章で選択したハイパーパラメータを用いてサンプルサイズ $n = 5, 50, 500, 5000$ のデータに対して推定を行い、結果を例示する。ここではサンプルサイズ $n = 500$ が本実験で想定しているデータサイズであり、 $n = 5$ は極端に小さい場合、 $n = 5000$ は極端に大きい場合を想定している。最後に、サンプルサイズの増加に伴う両モデルの推定性能の違いを比較する。

次の条件で数値実験を行い、 g_{n+1} の期待値と 95% 信用区間を推定し対応する $p_{n+1} = p(x_{n+1})$ の信用区間を求めることで推定手法を検証した。まず、サンプルサイズを n とし、一様分布 $U(-\pi, \pi]$ に従って x_1, \dots, x_n を生成して、真の正答確率 $p(x)$ を次のように設定した(図 1b)：

$$p(x) = p_0 + a_1 \sin x + a_2 \cos 2x$$

ここで $a_1 = 0.1$, $a_2 = 0.05$, $p_0 = 0.75$ とした。この p_0 は課題に対する正答確率が 0.75 であることを仮定し、その周りで脳波位相により正答確率が関数として変化していることを想定している。この正答確率は実際の実験課題の難易度に相当しており、予備実験等で事前に正答

確率 p_0 が 0.75 もしくは 0.5 になるように刺激の強度や難易度を調整していることを想定する．本研究では $p_0 = 0.75$ とし事前分布のパラメータ $\mu_0 = \log \frac{p_0}{1-p_0} \simeq 1.099$ としている．この $p(x)$ に基づき，ベルヌーイ分布 $\text{Bin}(1, p(x_i))$ から 0, 1 の値をとる z_i を生成し，データセット $D_n = \{(z_1, x_1), (z_2, x_2), \dots, (z_n, x_n)\}$ を生成した．二項・ロジットモデルの場合はこれらの (z_i, x_i) を $b = 12$ 個のビンに分類して $D'_b = \{(z'_1, m_1, x'_1), (z'_2, m_2, x'_2), \dots, (z'_{12}, m_{12}, x'_{12})\}$ とした．なお各ビンの代表値は区間中央 $x'_i = \frac{c_{iL} + c_{iU}}{2}$ とした．

4. ハイパーパラメータの選択

ベルヌーイ・ロジットモデルと二項・ロジットモデルはハイパーパラメータとして θ_1 と θ_2 を設定する必要がある．なお，付録 A にあるように θ_1 が大きくなるとラプラス近似の精度が悪くなるため，本研究では $\theta_1 \leq 1$ とした．また， θ_2 が大きくなるとグラム行列が退化する確率が大きくなる (図 2)．二項・ロジットモデルでは退化を避けるために $b = 12$ を考慮して $\theta_2 \leq 2$ とした．ベルヌーイ・ロジットモデルも二項ロジットモデルとの比較のために $\theta_2 \leq 2$ とした．本稿では交差検証と経験ベイズ法により，それぞれのモデルのハイパーパラメータを選択した．

4.1 交差検証による選択

d -fold 交差検証によりハイパーパラメータを選択するために誤差関数として次のような平均二乗誤差を用いた．

$$\text{CV-MSE} = \frac{1}{d} \sum_{k=1}^d \left\{ \frac{1}{N_{vld}^{(k)}} \sum_{i=1}^{N_{vld}^{(k)}} (z_i^{(k)} - \hat{p}^{(k)}(x_i^{(k)}))^2 \right\}$$

d はデータを分割するグループ数である． $d = 5$ とし，データを 5 分割したうちの 1 グループを検証データ，残りの 4 グループを訓練データとする． k を検証データのグループ番号とし，そのサンプルサイズを $N_{vld}^{(k)}$ とする． $z_i^{(k)}$ が検証データの観測値， $x_i^{(k)}$ が観測点である．訓練データで求めたモデルから計算された検証データの観測点 $x_i^{(k)}$ の正答確率を $\hat{p}^{(k)}(x_i^{(k)})$ とする．この正答確率と $z_i^{(k)}$ の平均二乗誤差を用いて 5-fold 交差検証を行った．本シミュレーションでは，まず，交差検証に用いるデータとしてサンプルサイズ $n = 200$ のデータを生成し，交差検証によるハイパーパラメータの選択を行った．なお，本研究では本実験で 500 程度のサンプルサイズを得ることを想定し，予備実験として 200 程度のサンプルサイズのデータを用意できることを想定している．

図 3a 左はベルヌーイ・ロジットモデルにおける交差検証によるハイパーパラメータ選択の結果である．ハイパーパラメータは $\theta_1 = [0.05, 1]$ と $\theta_2 = [0.05, 2]$ の範囲で 0.05 刻みで CV-MSE を計算し，値が最小のパラメータ $\theta_1 = 0.3$ ， $\theta_2 = 0.2$ を選択した．次に新たに $n = 200$ のデータを生成し，推定されたハイパーパラメータを用いて正答率の推定を行った (図 3a 右)．ベルヌーイ・ロジットモデルと同じデータセットを用いて二項・ロジットモデルにおける交差検証によるパラメータ選択をした結果が図 3b 左である．二項・ロジットモデルにおいてもほぼ同じようなパラメータ $\theta_1 = 0.3$ ， $\theta_2 = 0.15$ が選択された．このように予備実験等でハイパーパラメータを選択のためのデータセットを用意することで交差検証によりハイパーパラメータを選択することができる．

4.2 経験ベイズ法による選択

d -fold 交差検証を用いればハイパーパラメータの選択は可能であるが， d 回の再学習が必要であり，ある程度の計算時間が必要である．また， d 分割するため，パラメータ選択のために

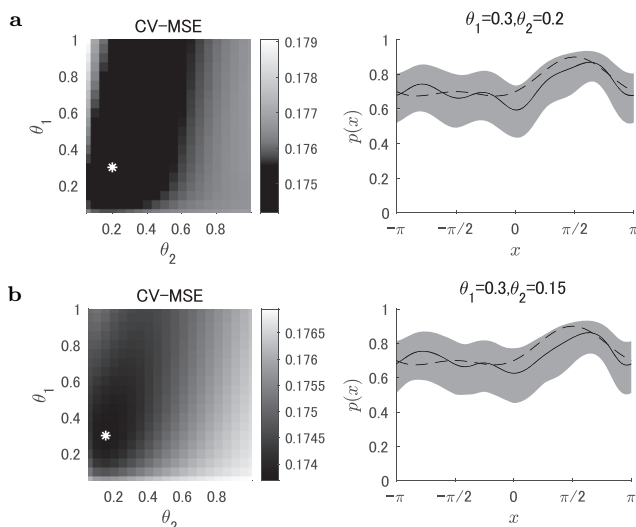


図3. 交差検証によるハイパーパラメータの選択：(a)ベルヌーイ・ロジットモデル．左図は θ_1 および θ_2 (≤ 1) ごとに交差検証で計算した平均二乗誤差を示す．白い星印は、このマップ上での最小値を表す ($\theta_1 = 0.3, \theta_2 = 0.2$)．右図は推定された正答率の平均 (実線) と 95% 信用区間 (灰色帯) を示し、破線は真の正答率を示す．(b)二項ロジットモデル．平均二乗誤差が最小のときのハイパーパラメータは $\theta_1 = 0.3, \theta_2 = 0.15$ ．左図と右図の構成は (a) と同様．

用意したすべてのデータを用いて学習と検証が行えない．本研究では将来的にベイズ最適化により少ないサンプルサイズで関数推定を行うこと目指しており、ハイパーパラメータの選択のために十分なサンプルサイズのデータを用意できないことも想定される．そこで経験ベイズ法を用いてハイパーパラメータ $\theta = (\theta_1, \theta_2)$ の選択を行う．この方法では次の周辺尤度

$$f(z|\theta) = \int f(z_n, g_n|\theta) dg_n = \int f(z_n|g_n, \theta) f(g_n|\theta) dg_n$$

を考え、この関数の値が最大となる θ を選択する．ベルヌーイ・ロジットモデルにおいてラプラス近似を用いると

$$\begin{aligned} f(z_n|\theta) &= \int f(z_n|g_n, \theta) f(g_n|\theta) dg_n \\ &= \int \prod_{i=1}^n \frac{e^{z_i g_i}}{1 + e^{g_i}} \frac{1}{\sqrt{2\pi^n |K_n|^{1/2}}} \exp\left[-\frac{1}{2} (g_n - \mu_0 \mathbf{1})^\top K_n^{-1} (g_n - \mu_0 \mathbf{1})\right] dg_n \\ &= \frac{1}{\sqrt{2\pi^n |K_n|^{1/2}}} \int h(g_n) g_n = \frac{1}{\sqrt{2\pi^n |K_n|^{1/2}}} h(g_n^*) \frac{(\sqrt{2\pi})^n}{|A_n^*|^{1/2}} \\ &= \frac{h(g_n^*)}{|K_n A_n^*|^{1/2}} \end{aligned}$$

となる．なお、二項・ロジットの場合も対応する $h(g_b^{**})$ に置き換わるだけである．ここでベルヌーイ・ロジットモデルの式(2.2)と二項・ロジットモデルの式(2.5)に g_n^* と g_b^{**} を代入した $h(g_n^*)$ と $h(g_b^{**})$ を用いると、ベルヌーイ・ロジットモデルの対数周辺尤度は

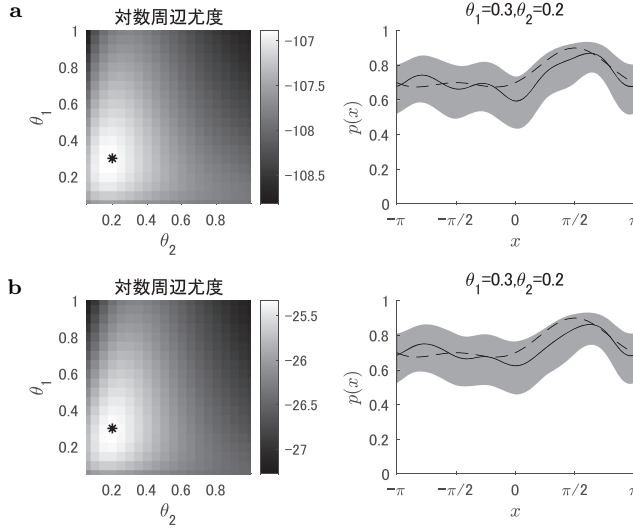


図 4. 経験ベイズ法によるハイパーパラメータの選択：(a)ベルヌーイ・ロジットモデル．左図は θ_1 および θ_2 (≤ 1) ごとの対数周辺尤度を示す．黒の星印はこのマップ上での最大値を表す ($\theta_1 = 0.3$, $\theta_2 = 0.2$)．右図は推定された正答率の平均(実線)と 95% 信用区間(灰色帯)を示し、破線は真の正答率を示す．(b)二項ロジットモデル．対数周辺尤度が最大のときのハイパーパラメータは $\theta_1 = 0.3$, $\theta_2 = 0.2$ ．左図と右図の構成は (a)と同様．

$$\begin{aligned}
 \log f(z_n|\boldsymbol{\theta}) &= \sum_{i=1}^n [g_i^* z_i - \log(1 + e^{g_i^*})] - \frac{1}{2} (\mathbf{g}_n^* - \boldsymbol{\mu}_0)^\top K_n^{-1} (\mathbf{g}_n^* - \boldsymbol{\mu}_0) \\
 &\quad - \frac{1}{2} \log |K_n A_n^*|
 \end{aligned}$$

であり、二項・ロジットモデルの対数周辺尤度は

$$\begin{aligned}
 \log f(z_b|\boldsymbol{\theta}) &= \sum_{i=1}^b \left[\log \binom{m_i}{z_i} + g_i^{**} z_i - m_i \log(1 + e^{g_i^{**}}) \right] \\
 &\quad - \frac{1}{2} (\mathbf{g}_b^{**} - \boldsymbol{\mu}_0)^\top K_b^{-1} (\mathbf{g}_b^{**} - \boldsymbol{\mu}_0) \\
 &\quad - \frac{1}{2} \log |K_b A_b^{**}|
 \end{aligned}$$

である．これらの対数周辺尤度が最大となるハイパーパラメータを選択し同様に推定を行った結果が図 4 である．経験ベイズ法でも交差検証のときに用いた $n = 200$ のデータに対して $\theta_1 = [0.05, 1]$ と $\theta_2 = [0.05, 2]$ の範囲で 0.05 刻みで対数周辺尤度の計算を行い、値が最大となるハイパーパラメータを選択した．ベルヌーイ・ロジットモデルでは、 $\theta_1 = 0.3$, $\theta_2 = 0.2$ が選択された(図 4a)．同様に二項・ロジットモデルでも偶然ではあるが同一の値 $\theta_1 = 0.3$, $\theta_2 = 0.2$ が選択された(図 4b)．5 章では、両モデルにおいて比較のためハイパーパラメータをこれらの値に固定してサンプルサイズの増加に伴う変化の比較を行った．本稿では 1 つのデータセットを用いてハイパーパラメータを選択したが、データセットに依存して選択されるハイパーパラメータは異なる可能性がある．実際に、 $n = 200$ のデータを 100 回生成し、経験ベイズ法によるハイパーパラメータの選択を行った結果、ベルヌーイ・ロジットモデルでは θ_1 の平均 0.31

(標準偏差 0.26), θ_2 の平均 0.55 (標準偏差 0.47), 二項・ロジットモデルでは θ_1 の平均 0.28 (標準偏差 0.23), θ_2 の平均 0.56 (標準偏差 0.48) となり, データごとに一定のばらつきがみられた. 実際の脳波データを用いた推定を行う場合, 検証用データや予備実験から得られたデータに基づき, ハイパーパラメータを事前に選択する必要がある.

5. 推定結果

サンプルサイズ n が, 5, 50, 500, 5000 のデータセットに対してベルヌーイ・ロジットモデル (図 5) と二項・ロジットモデル (図 6) を用いた推定を実施した. 実際の実験ではサンプルサイズ n は数十から数百程度であり $n = 5000$ とすることは困難であると思われるが, 推定精度の確認のために $n = 5000$ も実施した. ベルヌーイ・ロジットモデルの場合は, $\kappa_1^{-1}(K) < 10^{-10}$ を満たすときに特異行列とみなし, サンプルサイズが $n = 5000$ の場合においても行列が退化しないように $\varepsilon = 10^{-6}$ と設定している. $\varepsilon = 0$ の場合のグラム行列の条件数の逆数 $\kappa_1^{-1}(K)$ は $n = 50$ で $\kappa_1^{-1}(K_{50}) = 4.5891 \times 10^{-19} < 10^{-10}$ となり特異行列になる. 一方, $\varepsilon = 10^{-6}$ のグラム行列の条件数の逆数は $n = 5000$ でも $(K'_{5000}) = 1.1827 \times 10^{-9} > 10^{-10}$ となりグラム行列が退化しない. ただし, サンプルサイズが大きくなるにつれて, K_n に εI を加えても κ_1^{-1} は徐々に小さくなる.

どちらの結果もサンプルサイズが $n = 5$ の場合は真の関数の値 (破線) と期待値の値 (実線) のずれが大きいが, 95% 信用区間は広がっており, その中におおむね真の関数を含む結果となった. サンプルサイズが大きくなるにつれて, 期待値は真の値に近づき $n = 5000$ のときにはおおむね一致しておりまた信用区間も狭くなっているのがわかる.

二項・ロジットモデルの場合はピンのサイズを 12 に固定しているため, サンプルサイズが大きくなってもグラム行列のサイズも値も変化しない. そのため, サンプルサイズの増加に伴う行列の退化が起きなかった. 実際, グラム行列の条件数の逆数 $\kappa_1^{-1}(K)$ は $\kappa_1^{-1}(K_{12}) = 0.0582$ である. 図 5 と図 6 は同じデータセットをもとに計算しているため, ベルヌーイ・ロジットモデルと二項・ロジットモデルのいずれの場合も, 期待値と信用区間の推定の結果がおおむね一

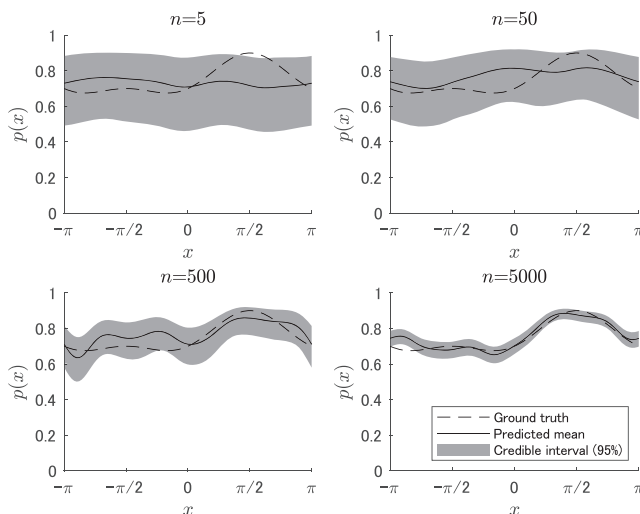


図 5. サンプルサイズごとのベルヌーイ・ロジットモデルによる推定例: 実線は推定された正答率の平均と灰色帯は 95% 信用区間を示し, 破線は真の正答率を示す.

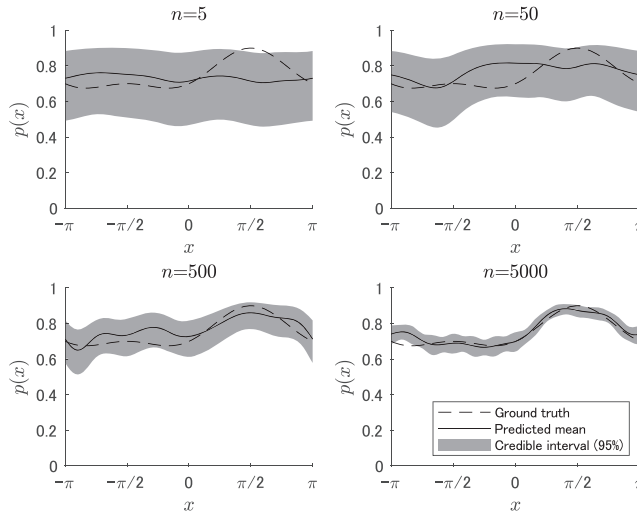


図 6. サンプルサイズごとの二項・ロジットモデルによる推定例：実線は推定された正答率の平均と灰色帯は 95% 信用区間を示し、破線は真の正答率を示す。

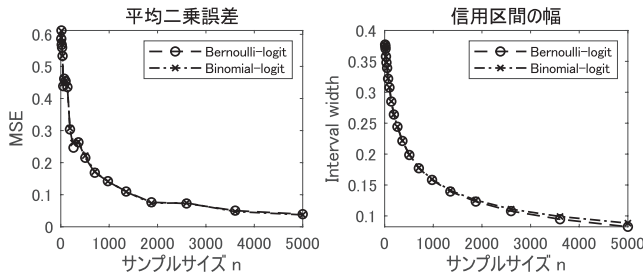


図 7. サンプルサイズの増加に伴う平均二乗誤差と信用区間の変化：左図は推定された正答率の平均と真の正答率との二乗誤差の平均を示し、右図は信用区間の幅の平均を示す (50 回)．サンプルサイズ n は 10 から 5000 の範囲で対数スケールに基づき等間隔に 20 点選び、整数に丸めた値を用いた。

致している。ただし、サンプルサイズ $n = 5000$ の結果を見るとベルヌーイ・ロジットモデルの結果と比較して二項・ロジットモデルの信用区間が揺らいでいるのがわかる。これは 12 個のビンに分けたためにデータが存在する x の値とそうでない x の値で推定値の分散が異なることが原因である。ただし、 $n \leq 500$ ではそのゆらぎが顕著ではない。

最後に、 n を増やしながらか D_n を同様の生成方法で 50 パターンずつ生成し、正答率の真の関数と推定された正答率の平均の二乗誤差の平均と信用区間の大きさの平均の変化を検証した(図 7)．ベルヌーイ・ロジットモデルの場合と二項・ロジットモデルの場合それぞれの結果を比較したがほぼ変わらなかった。ここでベルヌーイ・ロジットモデルでは特異行列になることを避けるために正則化を行っており、これが推定精度に影響を与える可能性がある。また、二項・ロジットモデルでは $b = 12$ としてビン分けを行っており、これが推定精度に影響を与える可能性がある。ビン分けの影響はサンプルサイズが大きいときに信用区間に現れており、ベルヌーイ・ロジットモデルのほうが信用区間の幅が少し小さくなっている。平均二乗誤差に関してはベルヌーイ・ロジットモデルと二項・ロジットモデルはほぼ同じ結果を示した。

ただし、ベルヌーイ・ロジットモデルはサンプルサイズが大きくなるにつれてグラム行列のサイズが大きくなり計算に必要な時間も長くなるのに対して、二項・ロジットモデルはグラム行列のサイズが変化しないので行列計算の演算に必要な時間がほぼ一定である。

6. 考察

6.1 ベルヌーイ・ロジットモデルと二項・ロジットモデルの比較

本稿では、脳波位相と正答確率の関数を推定するためにベルヌーイ・ロジットモデルと二項・ロジットモデルの両者を構築した。ベルヌーイ・ロジットモデルは各試行ごとの 0/1 応答をそのままベルヌーイ分布に従うものとしてモデル化している。このモデルはサンプルサイズの増加に伴い、グラム行列が退化する可能性が高くなり、数値安定性の確保には正則化を必要としている。また、サンプルサイズが増加するにつれてグラム行列のサイズも大きくなるため、計算量も増加する。一方、二項・ロジットモデルでは位相区間を複数のビンに分割し、各ビンで得られた正答数を二項分布に従うと仮定している。これにより、サンプルサイズが増えてもグラム行列のサイズはビン数に依存して一定のままであり、計算量や退化の問題が回避できる。特にリアルタイムでの高速計算に適していると考えている。

また、本研究は将来的に脳波位相依存刺激法におけるベイズ最適化での利用を想定している。ベイズ最適化を用いた先行研究として、脳波位相と TMS による運動誘発電位 (Motor Evoked Potential: MEP) の関数推定に関するベイズ最適化の試みが報告されている (Kirchhoff et al., 2024)。彼らも周期カーネルを用いたガウス過程回帰により脳波位相と MEP の関数を推定しているが、MEP は連続値として扱えるため、通常のガウス過程回帰で実装可能であった。一方、本研究では二者択一課題などの 2 値応答を対象とするため、ロジット変換を導入した統計モデルを構築した。さらに、ベイズ最適化への応用を見据えて、MCMC では計算負荷が大きいという問題に対してラプラス近似を導入し、また、二項・ロジットモデルによるビン分けを用いることでガウス過程回帰を行う際のグラム行列の退化に対する対処と、リアルタイムでの高速計算が行えるように改良した。

6.2 推定方法と数値的安定性の課題

本研究で用いたロジット変換をリンク関数としたガウス過程回帰ではラプラス近似により正規分布として事後分布を求めている。推定したいモデルが解析的に求めることが難しい場合、MCMC などの手法により推定する方法もあるが、ベルヌーイ・ロジットモデルでは状態空間の次元 n はサンプルサイズとともに増加し、計算量もそれに伴って増加する (Rue et al., 2009)。また、二項・ロジットモデルでも将来的にベイズ最適化と組み合わせる際にリアルタイムで高速で関数推定を行うことを想定しており、計算時間が大きくなる MCMC によらない手法として構築した。ただし、ハイパーパラメータである θ_1 が大きいとラプラス近似による正規近似の精度に注意が必要である。本研究の数値実験の設定では、交差検証と経験ベイズ法のどちらの結果も θ_1 が 1 以下の値を選択しており、この範囲においては大きな問題は生じないが、 θ_1 に 1 以上の値を設定したい場合は正規近似の近似精度が低下する可能性がある (付録 A 参照)。

さらに、両モデルは θ_2 が大きいとグラム行列が退化しやすくなる問題がある。このパラメータを小さく設定したとしてもベルヌーイ・ロジットモデルの場合、グラム行列はサンプルサイズの増加に伴い退化する確率が高くなるため ε を設定する必要がある。またベイズ最適化に用いる場合は獲得関数により、意図的に近い観測点をサンプリングする可能性があり、より退化する確率が高くなる。リアルタイムでの高速計算が必要な場合にも、サンプルサイズに依存して行列のサイズが大きくなることは好ましくなく実装上の障害となる。本稿では、この問題に

対してビン分けと二項分布を仮定した二項・ロジットモデルを用いることにより、サンプルサイズが増加しても、グラム行列のサイズが一定であるモデルを構築することで対処している。二項・ロジットモデルはグラム行列のサイズがサンプルサイズに対して一定であるため、計算時間も変化しないことからリアルタイムで実装する適応的な手法に適していると考えている。また、ビン分けを等間隔とし、区間中央を代表値とする場合、グラム行列が対称な循環行列となり、ビンの数が多い場合もその逆行列の計算を高速フーリエ変換(FFT)で計算することが可能となる。

本研究の数値実験においては、位相を12のビンに分割しているにもかかわらず、二項・ロジットモデルはベルヌーイ・ロジットモデルとほぼ同様の性能を示している。平均二乗誤差に関してはサンプルサイズが5000でもほぼ一致しており、信頼区間は僅かにベルヌーイ・ロジットモデルの方が狭いが、二項・ロジットモデルも大きくは異なる(図7)。実データにおいて同様の結果が得られるかは正答率の関数の形にもよるが、実際の被験者ごとの実験において、1つの条件の実験試行が5000を超えることは稀であると考えられる。また、本研究では脳波位相が観測されていることを前提としているが、実際には脳波位相の推定にも不確実性がありビン分けをしていない状態の位相の精度が十分に高いとは限らない。特に脳波位相依存刺激法による実験では、位相が特定の値の近傍で刺激を与える。この場合、ある範囲の位相は同じ状態とみなし刺激を与えているため、位相をビン分けした統計モデルを構築することは実験に用いる上で妥当であると考えられる。

7. おわりに

本稿では、ロジット変換をリンク関数とした周期カーネルを用いたガウス過程回帰により、位相ごとの正答率関数を推定する手法を構築した。

本研究の統計モデルでは脳波の位相を観測点として扱ったが、この位相は必ずしも正確に求められるものではない。特に脳波位相依存刺激法を用いる場合には、リアルタイムで過去の信号から未来を予測し、現在の瞬時位相を決定する必要がある(Chen et al., 2013; Onojima and Kitajo, 2021)。より実践的なモデルを考える場合、脳波位相を確率変数として扱う必要があるが、本研究では観測点として与えられると仮定した。今後は、位相推定の不確実性を含めた統計モデルの構築や実データでの検証を進める必要がある。

さらに、本稿で構築した手法はノンパラメトリックなガウス過程回帰に基づくものであるが、今後は実データにおいてスプライン回帰などの他のノンパラメトリックな手法や、周期関数を用いたパラメトリックな手法との比較を、獲得関数を導入したベイズ最適化の枠組みにおいて検討する必要がある。

謝 辞

本研究は、統計数理研究所が主催する統計エキスパート人材育成コンソーシアムの支援を受けて実施されました。ここに深く感謝申し上げます。また、本研究はJSPS 科研費 JP24K18601, JP22K11943 の助成を受けたものです。有益なご指摘と建設的なコメントをくださった査読者の皆様に感謝申し上げます。特に、ビンの代表値の取り方について、平均値や中央値を用いることにより推定精度が向上する可能性をご指摘くださったコメントに感謝いたします。

付 録

A. ラプラス近似の精度

Rue et al. (2009)は一般の潜在ガウスモデルにおけるラプラス近似の精度の改良について議論している. Wang and Shi (2014)は彼らの結果を用いてロジスティック回帰の場合のラプラス近似に言及しており, 一般には $\log h(\mathbf{g}_n)$ 全体をラプラス近似するより, $\log f(z_i|g_i)$ を1つずつテラー展開を用いて2次近似するほうが近似精度が良くなることを述べている. しかし, 我々のベルヌーイ・ロジットモデルの場合には $\log f(z_i|g_i)$ はすべて同じ形のため全体をラプラス近似することと彼らの逐次近似法は同値である.

なお, $h(\mathbf{g}_n)$ のラプラス近似の精度はハイパーパラメータ θ_1 に依存している. まず $n = 1$ のときにそのことを確認しておこう. 本稿ではラプラス近似を用いて(2.2)式

$$(A.1) \quad h(\mathbf{g}_n) = \prod_{i=1}^n \frac{e^{z_i g_i}}{1 + e^{g_i}} \exp \left[-\frac{1}{2} (\mathbf{g}_n - \mu_0 \mathbf{1})^\top K_n^{-1} (\mathbf{g}_n - \mu_0 \mathbf{1}) \right]$$

の正規近似を考えている. この近似は, もとの非正規分布関数を次のように正規分布で近似している.

$$h(\mathbf{g}_n) \simeq h(\mathbf{g}_n^*) \exp \left[-\frac{1}{2} (\mathbf{g}_n - \mathbf{g}_n^*)^\top A_n^* (\mathbf{g}_n - \mathbf{g}_n^*) \right],$$

$$A_n^* = -H_n^* = \text{diag}(p_1^*(1 - p_1^*), \dots, p_n^*(1 - p_n^*)) + K_n^{-1}$$

\mathbf{g}_n^* は $h(\mathbf{g}_n)$ の極大点である. (A.1)式の正規近似の精度を確認するために $n = 1$ のときの近似を考える. $k(x, x) = \theta_1$ とおくと, (A.1)式は

$$h(g_1) = \frac{e^{g_1 z_1}}{1 + e^{g_1}} \exp \left[-\frac{1}{2} \frac{(g_1 - \mu_0)^2}{\theta_1} \right]$$

と書ける.

$$\nabla \log h(g_1) = z_1 - p_1 - \frac{g_1 - \mu_0}{\theta_1} = z_1 - \frac{e^{g_1}}{1 + e^{g_1}} - \frac{g_1 - \mu_0}{\theta_1} = 0$$

の解を g_1^* とすると,

$$A_1^* = p_1^*(1 - p_1^*) + \frac{1}{\theta_1} = \frac{e^{g_1^*}}{(1 + e^{g_1^*})^2} + \frac{1}{\theta_1}$$

であるから, $h(g_1)$ は $\mathcal{N}(g_1^*, (A_1^*)^{-1})$ の定数倍で近似される. この関数と $h(g_1)$ を重ねて表示すると, $\mu_0 = \log 3 (p_0 = 0.75)$, $\theta_1 = 1$, $z = 0, 1$ のとき, 図 8a, b のように, $z = 0, 1$ のいずれの場合にもほぼ一致しており, よく近似できている. しかし, $\theta_1 = k(x, x) = 4$ の時は, 図 8c, d のように, すこし近似が悪くなる. $\theta_1 = 16$ の時は, 図 8e, f のように $h(g_1)$ の歪が大きくなり, より近似が悪くなる. θ_1 が大きくなるに従い, 近似精度は悪くなる. 次に $n = 20$ とし, 観測点 x_i を $(-\pi, \pi]$ で等間隔にとり, $\mathbf{x} = (x_1, \dots, x_{20})$ での観測値を $\mathbf{z} = (1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1)$ として, 周期カーネルのハイパーパラメータが $\theta_1 = 1$, $\theta_1 = 16$ のそれぞれの場合に MCMC を 100 万回ずつおこない, g_{15} の周辺事後分布を求めた. その結果が図 8g, h である. $\theta_1 = 1$ のときはほぼ正規分布であるが, $\theta_1 = 16$ のときは歪みが生じラプラス近似の精度が悪くなる.

B. 条件数の逆数

特異行列の判定のために MATLAB (The MathWorks Inc., 2022) の組み込み関数 `rcond` を用

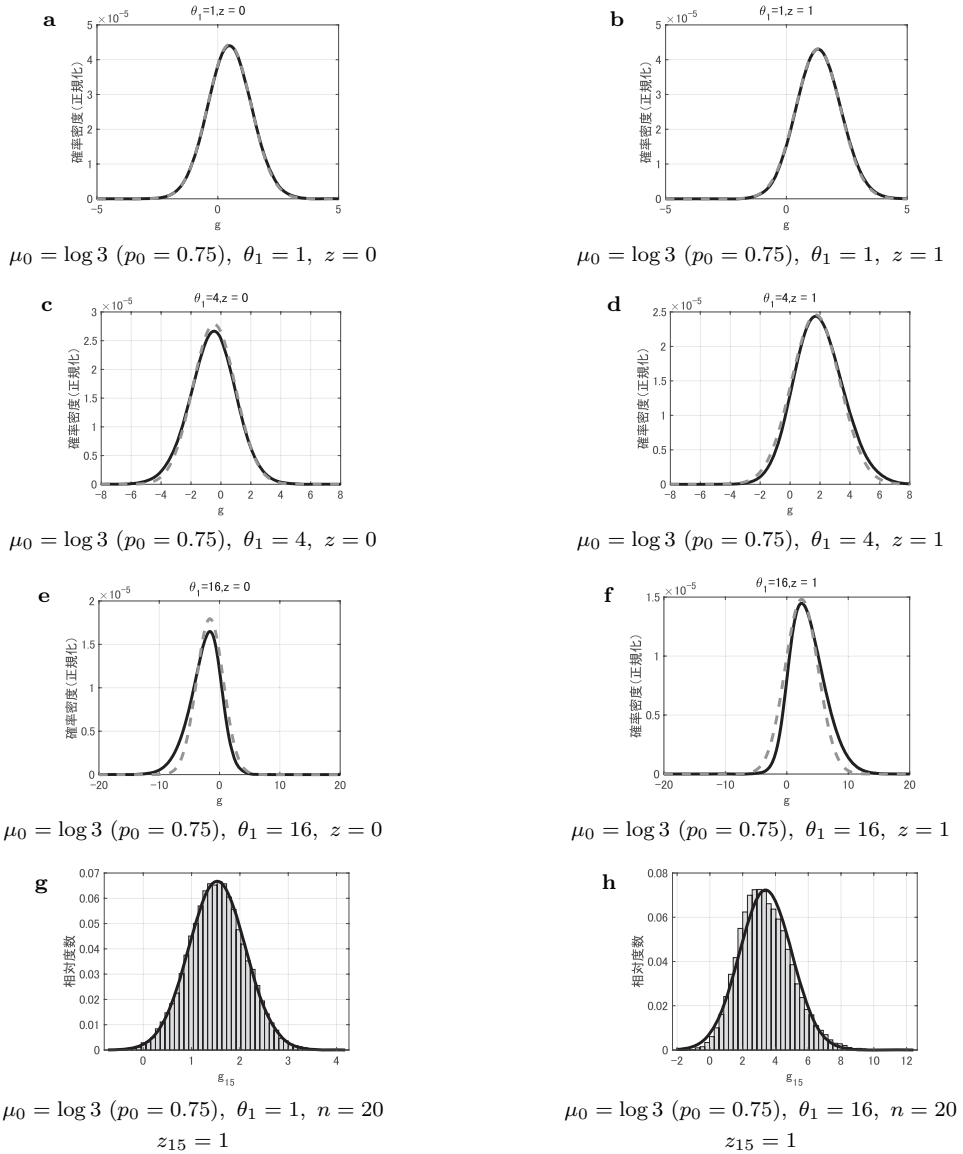


図 8. 異なる θ_1 と z の条件における分布の比較.

いた. この関数は行列 K の 1-ノルムに基づく条件数の逆数を指標とした. 条件数は次式で定義される:

$$\kappa(K) = \|K\| \cdot \|K^{-1}\|$$

ここで $\|\cdot\|$ は行列ノルムであり, 本稿では 1-ノルムを用いた. 1-ノルムは, 行列の各列の絶対値の合計のうち最大のものとして定義される:

$$\|K\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |k_{ij}|$$

この定義に基づき、1-ノルム条件数の逆数は次のように近似される：

$$\kappa_1^{-1}(K) \approx \frac{1}{\|K\|_1 \cdot \|K^{-1}\|_1}$$

`rcond` は、LAPACK (Anderson et al., 1999) の条件数推定アルゴリズムに基づき、LU 分解を用いて $\|K^{-1}\|_1$ を間接的に評価することで、条件数の逆数 $\kappa_1^{-1}(K)$ を高速かつ数値的に安定に算出している。

参 考 文 献

- Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. and Sorensen, D. (1999). *LAPACK Users' Guide*, 3rd ed., Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.
- Busch, N. A. and VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention, *Proceedings of the National Academy of Sciences*, **107**(37), 16048–16053, <https://doi.org/10.1073/pnas.1004801107>.
- Chen, L. L., Madhavan, R., Rapoport, B. I. and Anderson, W. S. (2013). Real-time brain oscillation detection and phase-locked stimulation using autoregressive spectral estimation and time-series forward prediction, *IEEE Transactions on Biomedical Engineering*, **60**(3), 753–762.
- Hebron, H., Lugli, B., Dimitrova, R., Jaramillo, V., Yeh, L. R., Rhodes, E., Grossman, N., Dijk, D.-J. and Violante, I. R. (2024). A closed-loop auditory stimulation approach selectively modulates alpha oscillations and sleep onset dynamics in humans, *PLOS Biology*, **22**(6), <https://doi.org/10.1371/journal.pbio.3002651>.
- Kirchhoff, M., Humaidan, D. and Ziemann, U. (2024). Closed-loop phase selection in EEG-TMS using Bayesian optimization, *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1730–1735, <https://doi.org/10.1109/SMC54092.2024.10831723>.
- Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M. and Ro, T. (2009). To see or not to see: Prestimulus α phase predicts visual awareness, *Journal of Neuroscience*, **29**(9), 2725–2732, <https://doi.org/10.1523/JNEUROSCI.3963-08.2009>.
- Ng, B. S. W., Schroeder, T. and Kayser, C. (2012). A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception, *Journal of Neuroscience*, **32**(35), 12268–12276, <https://doi.org/10.1523/JNEUROSCI.1877-12.2012>.
- Onojima, T. and Kitajo, K. (2021). A state-informed stimulation approach with real-time estimation of the instantaneous phase of neural oscillations by a Kalman filter, *Journal of Neural Engineering*, **18**(6), <https://doi.org/10.1088/1741-2552/ac2f7b>.
- Onojima, T., Kitajo, K. and Mizuhara, H. (2017). Ongoing slow oscillatory phase modulates speech intelligibility in cooperation with motor cortical activity, *PLOS ONE*, **12**(8), <https://doi.org/10.1371/journal.pone.0183146>.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **71**(2), 319–392, <https://doi.org/10.1111/j.1467-9868.2008.00700.x>.
- Shakeel, A., Onojima, T., Tanaka, T. and Kitajo, K. (2021). Real-time implementation of EEG oscillatory phase-informed visual stimulation using a least mean square-based AR model, *Journal of Personalized Medicine*, **11**(1), <https://doi.org/10.3390/jpm11010038>.
- Sofro, A. and Oktaviarina, A. (2018). Gaussian process regression model in spatial logistic regression, *Journal of Physics: Conference Series*, **947**(1), <https://doi.org/10.1088/1742-6596/947/1/012005>.
- Tervo, A. E., Metsomaa, J., Nieminen, J. O., Sarvas, J. and Ilmoniemi, R. J. (2020). Automated

- search of stimulation targets with closed-loop transcranial magnetic stimulation, *Neuroimage*, **220**, <https://doi.org/10.1016/j.neuroimage.2020.117082>.
- Tervo, A. E., Nieminen, J. O., Lioumis, P., Metsomaa, J., Souza, V. H., Sinisalo, H., Stenroos, M., Sarvas, J. and Ilmoniemi, R. J. (2022). Closed-loop optimization of transcranial magnetic stimulation with electroencephalography feedback, *Brain Stimulation*, **15**(2), 523–531, <https://doi.org/10.1016/j.brs.2022.01.016>.
- The MathWorks Inc. (2022). MATLAB version: 9.13.0.2320565 (R2022b), <https://www.mathworks.com/help/releases/R2022b/> (最終アクセス日 2025 年 5 月 1 日).
- Torrecillos, F., Falato, E., Pogosyan, A., West, T., Di Lazzaro, V. and Brown, P. (2020). Motor cortex inputs at the optimum phase of beta cortical oscillations undergo more rapid and less variable corticospinal propagation, *Journal of Neuroscience*, **40**(2), 369–381, <https://doi.org/10.1523/JNEUROSCI.1953-19.2019>.
- Wang, B. and Shi, J. Q. (2014). Generalized Gaussian process regression model for non-Gaussian functional data, *Journal of the American Statistical Association*, **109**(507), 1123–1133, <https://doi.org/10.1080/01621459.2014.889021>.
- Zoefel, B. and VanRullen, R. (2015). The role of high-level processes for oscillatory phase entrainment to speech sound, *Frontiers in Human Neuroscience*, **9**, <https://doi.org/10.3389/fnhum.2015.00651>.
- Zrenner, C., Belardinelli, P., Mueller-Dahlhaus, F. and Ziemann, U. (2016). Closed-loop neuroscience and non-invasive brain stimulation: A tale of two loops, *Frontiers in Cellular Neuroscience*, **10**, <https://doi.org/10.3389/fncel.2016.00092>.
- Zrenner, C., Desideri, D., Belardinelli, P. and Ziemann, U. (2018). Real-time EEG-defined excitability states determine efficacy of TMS-induced plasticity in human motor cortex, *Brain Stimulation*, **11**(2), 374–389, <https://doi.org/10.1016/j.brs.2017.11.016>.
- Zrenner, C., Kozák, G., Schaworonkow, N., Metsomaa, J., Baur, D., Vetter, D., Blumberger, D. M., Ziemann, U. and Belardinelli, P. (2023). Corticospinal excitability is highest at the early rising phase of sensorimotor μ -rhythm, *NeuroImage*, **266**, <https://doi.org/10.1016/j.neuroimage.2022.119805>.

Gaussian Process Regression with a Logit Link to Estimate the Relationship between Neural Oscillatory Phase and the Probability of a Correct Response

Takayuki Onojima¹ and Masakazu Jimbo²

¹Data Science and AI Innovation Research Promotion Center, Shiga University

²The Institute of Statistical Mathematics

Human brain activity can be measured using electroencephalography (EEG), and the ongoing brain state is known to affect task performance. It has been suggested that the phase of EEG oscillations within specific frequency bands is related to perceptual processes. To estimate this relationship, it is necessary to model the functional dependence between EEG phase in a target frequency band and task performance on a subject-by-subject basis. In recent years, stimulation techniques that rely on the ongoing EEG phase, such as phase-dependent stimulation, have been proposed, highlighting the need for methods that can flexibly estimate the relationship between EEG phase and brain function for each individual. In this study, we constructed a statistical framework for flexibly estimating the probability of a correct response, using trial-by-trial binary outcomes (correct/incorrect or detected/undetected) as observed data and EEG phase as an explanatory variable. Specifically, we developed a bernoulli–logit model that employs the logit link function and Gaussian process regression with a periodic kernel to estimate the probability of correct responses as a function of EEG phase. To address matrix degeneration and increased computational cost with larger sample sizes, we further refined the model and proposed a binomial–logit model. This study presents two statistical models for estimating the functional relationship between EEG phase and the probability of a correct response, and verifies their effectiveness through numerical simulations.

中高生の「早期Uターン意向」等に関する分析

—滋賀県長浜市におけるケーススタディ—

松原 悠¹・堀 兼大朗²・伊達 平和²

(受付 2025 年 6 月 27 日；改訂 8 月 21 日；採択 8 月 22 日)

要 旨

本研究では、滋賀県長浜市において中学生と高校生を対象としたUターン意向等に関する調査を実施し、特に、20代や30代といった比較的若い年代でUターンする意向(早期Uターン意向)に注目した分析を行った。長浜市役所と共同して長浜市内の全中学校・高校の生徒に対して回答を依頼し、約3,000名から回答を得た。ロジスティック回帰分析を通じて、市への愛着・地元企業の認知度・まちの満足度・親との関係等が将来居住意向やUターン意向と関連していることがわかった。特に、早期Uターン意向を持つ者は、地元企業の認知度が高い、親と悩み事を気軽に話せる関係にある、といった傾向を有していた。政策面においては、働く場に関する認知度の向上を図る施策や、親子のコミュニケーションを改善するような施策の有用性が示唆された。

キーワード：Uターン、早期Uターン意向、中学生、高校生、ロジスティック回帰分析。

1. はじめに

日本の自治体の多くは、進学や就職に伴う若年層の流出という課題を抱えている。たとえば、2020年の国勢調査のデータ(政府統計の総合窓口(e-Stat), 2021)に基づいて高校卒業前後の17歳人口と19歳人口を比較すると、19歳人口と17歳人口の比率が全国水準を上回るのは1,718市町村中364市町村にとどまり、残りの1,300以上の市町村において若年層の転出超過傾向があることがうかがえる。

自治体にとっては、若年層が一旦転出した後、再び出身自治体に戻ってくる(Uターンする)かどうかが地域の存続にあたっての重要事項の一つとなる。Uターンに関する主要な調査としては、国立社会保障・人口問題研究所が実施している人口移動調査が挙げられる。2016年に実施された第8回人口移動調査の結果(国立社会保障・人口問題研究所, 2018)によると、出生都道府県から県外に移動した後、再び出生都道府県に戻ることをUターンと定義したとき、Uターンを行った者の割合は全国で20.4%となっている。なお、これは全ての年齢層を含んだ値である。たとえば、40～49歳の者に限った場合には同割合は24.9%となり、県内出生者の約4

¹ 滋賀大学 データサイエンス・AIイノベーション研究推進センター：〒522-8522 滋賀県彦根市馬場一丁目1-1; yu-matsubara@biwako.shiga-u.ac.jp

² 滋賀大学 データサイエンス学部：〒522-8522 滋賀県彦根市馬場一丁目1-1; kentaro-hori@biwako.shiga-u.ac.jp, heiwa.date@biwako.shiga-u.ac.jp

分の1がこの年齢層までにUターンを行っていることとなる。また、出生都道府県別に全年齢のUターン率をみると、最小で15.2%(北海道)～最大で30.0%(宮崎県)となっており、県による差がみられることがわかる。

次に、各都道府県の現住者における県内出身者(Uターン者および出生から県内に継続居住している者)という観点から同調査における現住都道府県別の移動歴をみると、県内出身者の割合は全国で59.8%となっている。県別では、最小で44.6%(千葉県)～最大で76.8%(岩手県)と、県によって値が大きく異なり、特に三大都市圏以外の道県で割合が大きいの傾向が存在する。なお、同調査では全体の12.6%が県外移動歴不詳または出生県不明となっているため、実際の県内出身者の割合は上記の数値よりも高い可能性がある。他の類似の調査としては、国土交通省(2015)が存在し、三大都市圏以外の地方では県内出身者の割合が77.9%を占めていると報告されている。割合の値は少し異なるものの、地方では県内出身者が現住者の多くを占めているという傾向は共通している。以上より、地方において県内出身者は地域の将来を担う重要な候補者であり、地域の中高生の将来居住意向やUターン意向は、地域の将来を左右する重要な要素の一つであると考えられる。

以上のような背景から、現在、本研究の研究対象地域である滋賀県長浜市をはじめとする自治体においては、Uターン促進の施策が模索されている。本研究では、滋賀県の北部に位置する長浜市において、Uターン等に関する現状把握の一環として、市内の中高生を対象とした将来居住意向やUターン意向に関する調査を同市役所と共同して実施した。なお、自治体が居住している中高生らの(将来の)Uターンの促進を図ろうとしたとき、中高生らが一度転出した後よりも、その自治体内に居住している期間のほうが日常的な接点を有するため、より多様な施策を講じることができるのではないと思われる。秋山(2017)もUターン促進における長期的な関係構築型のコミュニケーション戦略の必要性を指摘している。Uターン意向は、Uターン行動そのものではなく、あくまで回答時点での意向ではあるものの、その現状や変化を把握することは、今後地域の中高生に対して継続的なはたらきかけを行うにあたり、施策の効果を確認し改善を図っていくうえでも重要であると考えられる。なお、本調査においては、Uターンを、「長浜市内に居住した経験がある者が長浜市以外の地域に住んだ後に長浜市に戻ること」と定義している。上述の人口移動調査においては都道府県単位でUターンが定義されていたことに対し、本調査においては市との共同調査という特性上、市単位でUターンを定義していることに注意されたい。

本研究の目的について述べる。中高生を対象とするUターン意向に関する調査は、次章で述べるように多数の実施事例がある。しかしながら、それらの多くは単純集計結果や少数のクロス集計結果が公表されているのみで、回帰分析等を含めた分析が実施されているものは限られている。また、若年層の流出という課題を考えるにあたっては、Uターン意向の有無そのものだけでなく、比較的若い年齢のうちにUターンする意向があるかどうかを把握することも重要と思われる。本研究ではこれを「早期Uターン意向」と呼ぶこととし、これに注目した回帰分析を行う。そして、分析結果に基づき若年層のUターン促進に資する知見を抽出することを目指す。

以下、2章では先行研究を概観し、本研究の位置付けを明確にする。3章では調査ならびに分析の概要を述べる。4章では分析の結果を示し、5章で考察と全体のまとめを行う。

2. 先行研究の概要と本研究の位置付け

Uターン意向に関する研究について、調査の対象者に着目して分類を行うと、「調査時点で当該自治体に居住している中高生等を対象としたもの」と、「調査時点で既に他の自治体に転出

している当該自治体出身の成人等を対象としたもの」が存在する。本研究は前者に分類されるものの、以下では、U ターン意向との関連が報告されている要素について後者で得られた知見も援用しながら論を進める。

中高生を対象とした U ターン意向に関する調査は、地方自治体によって実施されたものが多い。そのなかで、中学生と高校生の両方を調査対象とした事例としては、伊賀市 (2015)・富士吉田市 (2023)・鶴岡市企画部政策企画課 (2023)がある。これらの事例では、中学生の地元への将来居住意向(継続居住意向+ U ターン意向)が、高校生と比べて大きい傾向がみられている。また、伊賀市 (2015)では生まれたときから伊賀市に住み続けている人のほうが、そうでない人と比べて U ターン意向が大きいことが示されている。以上の事例より、中学生と高校生の別(ないし学年)や居住年数と、U ターン意向との関係を考慮する必要があると考えられる。

高校生のみを対象とした先行事例も多く存在する。そのなかで、本研究の関心である早期 U ターン意向については以下のように様々な形で尋ねられているものの、現時点でわかっていることは限定的である。設問の形式と得られた知見を順に記す。新温泉町 (2015, 2021)は、U ターン意向がある者を対象として希望する U ターン時期を尋ねており、回答の選択肢として、「大学・短大などを卒業したとき」「都会生活に満足したのち(5~10年後)」「結婚や子育てが必要となる時期」「人生の転機(20年後程度)」「定年後」といったものを用意している。ここでは、男性よりも女性のほうが早い時期を選択する傾向がみられている。久慈地域キャリア教育推進研究会・久慈市キャリア教育推進協議会 (2023)は、U ターン意向を尋ねる質問のなかに「進学先卒業後、地元以外の地域で何年か働いてから地元に戻りたい」「定年したら戻りたい」という選択肢を用意している。ここでは、単純集計結果は示されているもののクロス集計結果等は示されていない。新発田市 (2015)は、U ターン意向を持つと回答した者に対して、戻りたい年齢を記入することを求めている。ここでは、回答した高校3年生のうち、20代または30代で U ターンしたいと回答した者が U ターン意向を持つ者全体の約半数だったという結果が示されているものの、クロス集計等の結果は示されていない。樋田 (2018)は、高卒時(18歳時)・25歳時・40歳時・60歳時(老後)の居住希望地域を尋ねており、これらの回答から上記の各年齢時点までに U ターンしたいという意向を読み取ることができる。結果の分析を通じて、学業成績の高い層が将来の比較的早い段階に地元に戻って暮らすことを希望していると考察されている。以上の先行研究より、早期 U ターン意向に関する分析を行うにあたっては、性別や学業成績を考慮する必要があると考えられる。

成人を対象とした U ターン意向に関する代表的な調査事例としては、労働政策研究・研修機構 (2016)がある。三大都市圏以外の地方圏出身で、現在出身県外に居住者している 25~39歳の者 2,027人を対象として実施されたウェブモニター調査の結果のクロス集計から、「地元への愛着」と「高校時代までの地元企業の認知度」という2つの要素が、出身市町村へ戻りたい意向の大きさと正の関連を持っていることが示されている。成人における愛着と U ターン意向の関連については、海野 他 (2022)や松田・埴淵 (2023)も同様の傾向を報告している。中高生を対象とした調査でも同様の傾向がみられており、たとえば、松江市における中学2年生を対象とした調査(松江市, 2024)では、市に愛着を感じている者のほうが、将来居住意向に関する質問で「今の場所に住み続けたい」と回答している傾向があることが示されている。一方、労働政策研究・研修機構 (2016)において愛着と並んで重要な要素とされた地元企業の認知度(調査回答者が出身地を離れる前に地元企業をどの程度知っていたか)に関しては分析事例が少なく、特に、中学生時代における地元企業の認知度と将来居住意向や U ターン意向との関係が取り上げられている先行事例は管見の限り存在しない。

以上に加えて、先述の新発田市 (2015)では現状のまちの満足度が高校生の U ターン意向と関連することが報告されているほか、遠藤・沖 (2017)やマイナビ (2023)では進学や U ターン

就職といった進路選択における両親の影響の存在が示されている。

なお、上述の先行研究は各要素とUターン意向等との関係を個々に分析したものが大半であり、様々な要素がUターン意向等に与える影響を総合的に考慮した分析は希少である。そこで本研究においては、先行研究で指摘されてきた性別・中学生と高校生の別(ないし学年)・居住年数・学業成績といった要素によるUターン意向等の差を総合的に考慮したうえで、政策的な介入が比較的可能と思われる市への愛着・地元企業の認知度・まちの満足度・親との関係といった要素が、将来居住意向やUターン意向、そして早期Uターン意向に与える影響に注目していく。

3. 方法

3.1 調査方法

はじめに、調査対象地域である滋賀県長浜市の概要を述べる。長浜市は、滋賀県の北部に位置する自治体で、福井県と岐阜県に隣接している。現在の長浜市は、2010年1月1日に1市6町が合併して誕生したもので、2020年の国勢調査の人口は113,636人と、一定の人口規模を有している(長浜市, 2022)。市の中心部である長浜地域は相対的に都市的な性格を有しており、黒壁スクエア等の観光スポットも存在する。しかしながら、2008年から人口の自然増減が自然減、2009年からは社会増減が社会減という傾向が続いており、人口は減少基調をたどっている(長浜市, 2020)。また、長浜地域以外の旧町地域で特に人口減少が進んでおり、「過疎地域の持続的発展の支援に関する特別措置法」における一部過疎地域にも指定されている(総務省, 2022)。人口ビジョン(長浜市, 2020)においては、現状における将来推計人口を上回る「2060(R42)年に人口規模88,000人の維持及び人口構造の若返りを目指す」という長期的展望が掲げられており、人口関連施策が現在模索されている。

本調査の調査対象者は、長浜市内の中学校ならびに義務教育学校全12校に在籍している全中学生と、長浜市内に立地する高校ならびに高等養護学校全6校に在籍している全高校生である。以降、前者を対象とした調査を単に「中学生調査」と呼び、後者を対象とした調査を「高校生調査」と呼ぶこととする。

調査方法としては、中学生調査については、各学校(長)宛てに依頼文および回答フォームのQRコードを送り、各生徒のタブレット端末で回答してもらう形をとった。実施期間は令和5年7月14日～9月11日である。高校生調査については、各生徒に回答フォームのQRコード付き案内文を配付し、各自で回答してもらう形をとった。実施期間は令和5年7月20日～9月11日である。調査票の内容に関しては、当初、中学生調査の調査票と高校生調査の調査票をほぼ同一の内容にすることを計画していた。しかしながら、調査実施に際して関係者と内容調整を行う過程で、中学生調査の質問項目を高校生調査よりも大幅に減らすこととなってしまった。このため、次章以降の分析においては、中学生調査と高校生調査の両方で取得できた情報を用いた分析と、高校生調査のみにおいて取得できた情報を用いた分析とを分けて実施している。

回収結果について、中学生調査では2127名から回答が得られた(回収率67.0%)。高校生調査では902名から回答が得られた(回収率34.9%)。

3.2 分析方法の方針

本研究で実施する分析の概要について述べていく。まず、調査の主要な結果として、将来居住意向・Uターン意向・早期Uターン意向に関する設問への回答の分布を次章の冒頭で示す。次に、前章で述べた様々な要素とUターン意向等との関係を総合的に検討するため、ロジスティック回帰分析を行う。将来居住意向の有無については中学生と高校生のそれぞれについて

二項ロジスティック回帰分析を行い、結果を比較する。U ターン意向と早期 U ターン意向の有無については高校生のみにおいてデータが得られており、多項ロジスティック回帰分析を行って「U ターン意向なしの群」「U ターン意向ありかつ早期 U ターン意向なしの群」「早期 U ターン意向ありの群」の 3 群の特徴を比較する。

中学生ならびに高校生の将来居住意向の有無に関する二項ロジスティック回帰分析におけるモデルと、高校生の U ターン意向や早期 U ターン意向の有無に関する多項ロジスティック回帰分析のモデルを数式で表すと次のとおりとなる。式(3.1)が前者に対応しており、式(3.2)が後者に対応している。

$$(3.1) \quad \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_{i=1}^8 \beta_i X_i$$

$$(3.2) \quad \log\left(\frac{\pi_{(2)}}{\pi_{(1)}}\right) = \beta_{02} + \sum_{i=1}^{11} \beta_{i2} X_i, \log\left(\frac{\pi_{(3)}}{\pi_{(1)}}\right) = \beta_{03} + \sum_{i=1}^{11} \beta_{i3} X_i,$$

π : 「将来居住意向あり」に該当する確率

$\pi_{(1)}$: 「U ターン意向なし」に該当する確率

$\pi_{(2)}$: 「U ターン意向ありかつ早期 U ターン意向なし」に該当する確率

$\pi_{(3)}$: 「早期 U ターン意向あり」に該当する確率

$\beta_0, \beta_1, \dots, \beta_8, \beta_{02}, \beta_{12}, \dots, \beta_{112}, \beta_{03}, \beta_{13}, \dots, \beta_{113}$: 切片と回帰係数

X_1, \dots, X_{11} : 次章の表 4 で示す統制変数・説明変数

なお、目的変数が 3 群以上の順序変数の場合には、順序ロジスティック回帰分析が用いられる。本研究においては、「U ターン意向なしの群」「U ターン意向ありかつ早期 U ターン意向なしの群」「早期 U ターン意向ありの群」の 3 群のあいだに一定の順序関係があるとはみなしていない(たとえば、必ずしも早期 U ターン意向ありの者が U ターン意向が最も強いとは限らず、早期 U ターン意向がない者であっても U ターン意向が強い場合もある)ため、順序ロジスティック回帰分析は採用しなかった。

高校生調査においては、長浜市外から長浜市内の高校に通学している者もおり、回答者全員が長浜市に居住した経験を有するわけではない。本研究では U ターン意向に主たる関心があるため、長浜市居住経験を有する者のみを分析対象とすることとする。

3.3 ロジスティック回帰分析で使用する変数

目的変数には、将来居住意向・U ターン意向・早期 U ターン意向に関する設問への回答を使用する。将来居住意向は、「あなたは将来、長浜市に住みたいと思いますか」という質問文を用いて、「住みたい」「どちらかと言えば住みたい」「どちらとも言えない」「どちらかと言えば住みたくない」「住みたくない」という 5 件法で尋ねた。肯定的な回答である「住みたい」と「どちらかと言えば住みたい」を「将来居住意向あり(1)」に、それ以外を「将来居住意向なし(0)」にまとめた 2 値変数を分析で使用する。U ターン意向については、「仮に将来、長浜市以外の地域に住むことになったとして、その後、長浜市に戻りたいと思いますか」という質問文を用いて、「戻りたい」「どちらかと言えば戻りたい」「どちらとも言えない」「どちらかと言えば戻りたくない」「戻りたくない」という 5 件法で尋ねた。「戻りたい」と「どちらかと言えば戻りたい」という回答を「U ターン意向あり(1)」に、それ以外を「U ターン意向なし(0)」にまとめた 2 値変数を分析で使用する。早期 U ターン意向については、「U ターン意向あり」と回答した者のみを対象として、「長浜市に戻るとしたら、いつごろ戻りたいと思いますか」という質問文で尋ねた。回

答の選択肢としては、先行研究を参考としつつより直截的に時期を尋ねる形とし、「20歳代のころ」「30歳代のころ」「40歳代のころ」「50歳代のころ」「60歳代以降」「まだ考えていない」の6つを用意した。「20歳代のころ」と「30歳代のころ」に戻りたいという回答を「早期Uターン意向あり(1)」に、それ以外を「早期Uターン意向なし(0)」にまとめた2値変数を分析で使用する。

説明変数としては、市への愛着(5件法で質問)、地元企業の認知度(4件法で質問)、まちの満足度、親との関係を用いる。愛着の質問方法としては、對間他(2019)が愛着に関する69種類の尺度をレビューしているように、これまで実に多種多様な測定方法が提案されている。一方で、労働政策研究・研修機構(2016)や松田・埴淵(2023)のように、愛着の有無や程度を直接的に尋ねている例も存在しており、本調査でもこれにならって、「長浜市に親しみや好きという気持ちはありますか」という形で質問した。(なお、当初は愛着という言葉を用いて、「愛着がありますか」と質問する予定であったものの、関係者と内容調整を行う過程を通じて、中学生にとってもよりわかりやすい表現になるよう上記の形に変更することとなった。)まちの満足度に関しては、調査票において次の9つの観点、すなわち、若者向けの飲食店やカフェの数・若者向けの服屋の数・若者が集まる遊び場の数・スポーツ施設の数・都会への行きやすさ・勉強ができる場所の数・図書館の使いやすさ・自然の美しさや風景・地元のお祭りやイベント、についてそれぞれ5件法で満足度を尋ねた。そして、これらの9個の満足度を合算し、それを9で割った平均値をまちの満足度に関する説明変数として用いることとした。なお、市への愛着・地元企業の認知度・まちの満足度については、値が大きいほど肯定的な回答になるようコーディングした。そして、これらの変数については量的変数とみなして用いることとした。親との関係としては、将来の長浜市居住を勧められることがあるかを尋ねた。選択肢としては、「よく勧められる」「たまに勧められる」「まったく勧められない」の3つを用意した。これについては、「よく勧められる(2)」「たまに勧められる(1)」「まったく勧められない(0)」とコーディングしたうえで、量的変数とみなして用いることとした。また、高校生対象の調査においては、「悩み事を気軽に話せる人」として親が該当するかどうかを尋ねており、この質問への回答も親との関係に関する質的変数として用いることとした。

統制変数としては、性別・学年・居住年数・居住地域・高校入試の偏差値を用いる。なお、性別に関して「その他・答えたくない」とした回答については、少数であり今回は分析の対象外とした。居住年数については、異常値が入力されていた回答(居住年数が19以上となっていた回答)を分析対象外とした。居住地域に関しては、長浜市は相対的に都市的な性格を有する中心部の長浜地域と、それ以外の旧町地域とで地域特性に違いがあるため、長浜地域に居住しているかどうかのダミー変数を追加した。成績という要素に関連する変数としては、調査票のなかに成績を尋ねる質問を含めることができなかつたため、高校生調査の分析において高校入学時の学力に相当する統制変数として各高校の入試の偏差値を用いることとした。偏差値の値は情報源によって値は若干異なるものの大小の傾向は共通しており、今回は家庭教師のトライ(2025)を参照した。偏差値情報が入手できなかった1校は分析対象外とした。

なお、使用する全ての変数のうちのいずれか一つでも無回答等の欠損値がある回答は分析対象から除外(リストワイズ消去)した。

4. 結果

4.1 分析対象者の属性

本調査の回答者の性別・学年・居住地域の分布を表1に示す。なお、高校生については、3.2節において分析対象外とすることとした長浜市に居住した経験が一度もない者を除いたところ、 $n = 732$ となった。

表 1. 分析対象者の性別・学年・居住地域の分布.

		中学生	高校生
性別	男性	1029	324
	女性	1046	379
	その他・答えたくない	52	29
学年	1年生		381
	2年生	(内訳不明)	245
	3年生		103
居住地域	長浜	1204	382
	びわ	129	44
	浅井	174	92
	虎姫	96	24
	湖北	149	66
	高月	177	54
	木之本	98	31
	余呉	38	7
	西浅井	62	20
	長浜市外(※)	0	11

※過去に長浜市に居住した経験があるものの現在は転出している者
 注) 高校生は、3名が学年未回答、1名が居住地域未回答.

表 2. 将来居留意向の分布.

	住みたい	どちらかと言えば		どちらかと言えば 住みたくない	住みたくない	合計	肯定的な 回答の割合
		住みたい	言えない				
中学生	370	452	949	205	151	2127	38.6%
高校生	99	187	318	83	45	732	39.1%
合計	469	639	1267	288	196	2859	38.8%

表 3. 高校生の U ターン意向と希望する U ターン時期の分布.

Uターン意向	希望するUターン時期						合計	早期Uターン 意向の割合 (※)		
	20歳代 のころ	30歳代 のころ	40歳代 のころ	50歳代 のころ	60歳代 以降	まだ 考えて いない			回答対象外 (Uターン意 向なし)	
戻りたい		41	37	11	5	3	51	-	148	52.7%
どちらかと言えば戻りたい		32	48	17	10	3	81	-	191	41.9%
どちらとも言えない		-	-	-	-	-	-	313	313	-
どちらかと言えば戻りたくない		-	-	-	-	-	-	44	44	-
戻りたくない		-	-	-	-	-	-	36	36	-
合計		73	85	28	15	6	132	393	732	21.6%

※希望するUターン時期が「20歳代のころ」または「30歳代のころ」の者の割合、「Uターン意向」が「戻りたい」または「どちらかと言えば戻りたい」の者の数(339名)を分母とした場合・・・ 46.6%

4.2 将来居留意向・U ターン意向・早期 U ターン意向の回答の分布

本調査の主要な結果の一つとして、将来居留意向に関する設問への回答結果を表 2 に示す。約 4 割の者が、「住みたい」または「どちらかと言えば住みたい」という肯定的な回答を示した。なお、肯定的な回答の割合という観点において、先行研究とは異なり、中学生の将来居留意向が高校生と比べて若干高いという傾向は本調査ではみられなかった。

次に高校生の、U ターン意向と早期 U ターン意向に関する設問への回答結果を表 3 に示す。U ターン意向に関して「戻りたい」または「どちらかと言えば戻りたい」と回答した者の割合は回

表 4. ロジスティック回帰分析に使用する変数の記述統計.

	model 1に 使用するデータ		model 2に 使用するデータ		model 3に 使用するデータ					
	中学生		高校生		Uターン意向 なしの高校生		Uターン意向 ありかつ 早期Uターン意 向なしの高校生		早期Uターン意 向ありの高校生	
	1	0	1	0	1	0	1	0	1	0
【質的変数の度数分布】										
将来居住意向有無 (あり=1)	811	1262	251	399	-	-	-	-	-	-
Uターン意向有無 (あり=1)	-	-	-	-	0	351	157	0	142	0
早期Uターン意向有無 (あり=1)	-	-	-	-	0	351	0	157	142	0
性別 (女性=1)	1045	1028	341	309	190	161	80	77	71	71
学年 (高校 1 年生=1)	-	-	-	-	189	162	81	76	78	64
学年 (高校 2 年生=1)	-	-	-	-	113	238	60	97	47	95
居住地域 (長浜地域=1)	1174	899	341	309	173	178	80	77	88	54
親との関係 悩み事を気軽に話せる (該当=1)	-	-	-	-	181	170	105	52	101	41
【量的変数の平均と標準偏差】	平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
居住年数	12.62	2.32	14.95	3.12	14.58	3.63	15.43	2.23	15.31	2.41
入試の偏差値	-	-	-	-	50.56	8.72	51.47	8.39	50.29	8.64
市への愛着	4.03	1.03	4.00	0.99	3.61	1.05	4.39	0.67	4.54	0.64
地元企業の認知度	2.38	0.75	2.26	0.68	2.10	0.66	2.38	0.60	2.51	0.72
まちの満足度	3.39	0.79	3.22	0.67	3.03	0.60	3.42	0.67	3.48	0.69
親との関係 長浜市居住の勧め	0.43	0.64	0.41	0.63	0.30	0.54	0.50	0.64	0.61	0.75
	n=2073		n=650		n=351		n=157		n=142	

答者の 4 割強 $((148+191)/732 \approx 0.46)$ となった。また、早期 U ターン意向を持つ者の割合は、回答者全体の 2 割程度、U ターン意向を持つ者の 4 割強という結果となった。先行研究と比較すると、U ターン意向を持つ者全体の約半数が早期 U ターン意向を有していた新発田市 (2015) と似た傾向と言える。

4.3 将来居住意向・U ターン意向・早期 U ターン意向に関連する要素の分析

本節ではロジスティック回帰分析の結果を示していく。まず、使用する変数の記述統計量を表 4 に示す。本研究では、3.2 節で述べたように、中学生ならびに高校生の将来居住意向の有無に関する二項ロジスティック回帰分析(model 1・model 2)と、高校生の U ターン意向や早期 U ターン意向の有無に関する多項ロジスティック回帰分析(model 3)を行うため、各分析で使用するデータにおける各変数の記述統計量を分けて記載している。

次に、説明変数間の関係を明らかにするため、各説明変数間の相関係数を表 5、表 6 に示す。これらは表 1 に示した中学生のデータと高校生のデータに関するものである。相関係数の絶対値はいずれも 0.50 を下回っており、強い相関を有する説明変数の組は存在していないことがわかる。

ロジスティック回帰分析の結果を表 7 に示す。計算には IBM SPSS Statistics 29 を用いた。表中の数値は特に注記していない限りオッズ比(式(3.1))と対応させて言えば e^{β_i} の、式(3.2)と

表 5. 中学生のデータにおける説明変数間の相関係数.

	市への愛着	地元企業の 認知度	まちの満足度	長浜市居住の 勤め
市への愛着	-			
地元企業の認知度	0.35	-		
まちの満足度	0.49	0.28	-	
長浜市居住の勤め	0.23	0.20	0.25	-

※カテゴリカルデータについても連続値とみなして計算している

表 6. 高校生のデータにおける説明変数間の相関係数.

	市への愛着	地元企業の 認知度	まちの満足度	長浜市居住の 勤め	悩み事を 気軽に話せる
市への愛着	-				
地元企業の認知度	0.35	-			
まちの満足度	0.47	0.23	-		
長浜市居住の勤め	0.16	0.18	0.14	-	
悩み事を気軽に話せる	0.19	0.12	0.18	0.08	-

※カテゴリカルデータについても連続値とみなして計算している

表 7. ロジスティック回帰分析の結果(オッズ比).

	model 1 中学生の 将来居住意向の 有無	model 2 高校生の 将来居住意向の 有無	model 3 (参照カテゴリ: Uターン意向なし)	
			Uターン意向ありかつ 早期Uターン意向 なしの高校生	早期Uターン意向 ありの高校生
性別 (女性)	1.136 [0.929, 1.389]	1.047 [0.728, 1.507]	0.977 [0.640, 1.490]	1.029 [0.653, 1.622]
学年 (高校 1 年生)	-	-	1.543 [0.763, 3.123]	1.210 [0.588, 2.492]
学年 (高校 2 年生)	-	-	2.401 [1.144, 5.036] *	1.547 [0.715, 3.347]
居住年数	1.044 [0.997, 1.093] †	1.062 [0.994, 1.136] †	1.086 [0.997, 1.182] †	1.053 [0.968, 1.145]
居住地域 (長浜地域)	0.918 [0.750, 1.125]	1.237 [0.865, 1.769]	1.158 [0.765, 1.753]	1.820 [1.160, 2.854] **
入試の偏差値	-	-	1.012 [0.986, 1.038]	0.989 [0.962, 1.016]
市への愛着	2.554 [2.212, 2.949] ***	2.182 [1.687, 2.820] ***	2.359 [1.741, 3.195] ***	3.052 [2.155, 4.322] ***
地元企業の認知度	1.009 [0.871, 1.168]	1.378 [1.031, 1.842] *	1.274 [0.908, 1.787]	1.581 [1.105, 2.262] *
まちの満足度	1.670 [1.430, 1.950] ***	1.957 [1.427, 2.683] ***	1.531 [1.064, 2.202] *	1.490 [1.017, 2.182] *
親との関係				
長浜市居住の勤め	1.466 [1.249, 1.720] ***	1.469 [1.102, 1.958] **	1.484 [1.056, 2.087] *	1.706 [1.197, 2.432] **
悩み事を気軽に話せる	-	-	1.365 [0.882, 2.112]	1.623 [1.004, 2.622] *
定数	0.001 ***	0.000 ***	0.000 ***	0.000 ***
-2 対数尤度	2272.1	712.6		1104.5
Nagelkerke R ²	0.292	0.287		0.310
n	2073	650		650

[凡例] † : p < .10, * : p < .05, ** : p < .01, *** : p < .001, []内の数値は95%信頼区間

対応させて言えば $e^{\beta_{i2}}$ や $e^{\beta_{i3}}$ の推定値と信頼区間)であり, 他の変数を固定したもとで各変数が 1 単位増加した際に(質的変数の場合はそのダミー変数が 0 から 1 になった際に), model 1 と model 2 においては「将来居住意向あり」に該当する確率と「将来居住意向なし」に該当する確率の比, model 3 においては「U ターン意向ありかつ早期 U ターン意向なし」に該当する確率や「早期 U ターン意向あり」に該当する確率と, 参照カテゴリとした「U ターン意向なし」に該当する確率との比が何倍増加するかを示している. すなわち, オッズ比が 1 より大きければ, その

説明変数の増加はUターン意向等が増加する方向に作用し、逆に1未満であればUターン意向等が減少する方向に作用することを意味する。

分析の結果を順に確認していこう。モデルの説明力の程度をあらわすNagelkerke R^2 の値はいずれのモデルも0.3前後となっており、ある程度の説明力を有すると解釈できる。以下、有意水準を5%として有意となった変数について述べる。model 1とmodel 2では、使用する変数を揃えて分析を実施している。model 1では説明変数のうち、市への愛着・まちの満足度・親による長浜市居住の勧めが有意となった。model 2ではそれらの説明変数に加えて、地元企業の認知度も有意となった。

model 3では、「Uターン意向ありかつ早期Uターン意向なし」の高校生において、説明変数のうち、市への愛着・まちの満足度・親による長浜市居住の勧めが有意となった。また、統制変数のなかでは学年(高校2年生)が有意となった。「早期Uターン意向あり」の高校生においては、市への愛着・まちの満足度・親による長浜市居住の勧めが有意となったことに加え、地元企業の認知度・親に悩み事を気軽に話せるが有意となった。また、統制変数のなかでは居住地域(長浜地域に居住)が有意となった。なお、先行研究で報告されていた、性別や成績と早期Uターン意向との関連については、本分析では有意水準に至らなかった。

5. 考察とまとめ

5.1 結果に関する考察

はじめに、前章での分析結果に考察を加える。まず、中学生の将来居住意向の有無に関するmodel 1と高校生の将来居住意向の有無に関するmodel 2を比較すると、model 2のみにおいて地元企業の認知度が有意となった。将来居住意向の規定要因が中学生と高校生とで異なっていることが示唆される結果である。

次に、Uターン意向と早期Uターン意向に関するmodel 3において、「Uターン意向ありかつ早期Uターン意向なし」の高校生に関する結果と「早期Uターン意向あり」の高校生に関する結果を比較すると、地元企業の認知度と親との関係(親に悩み事を気軽に話せること)に関する説明変数が「早期Uターン意向あり」のみにおいて有意になった。これは、Uターン意向の規定要因と早期Uターン意向の規定要因が異なることを示唆する結果である。

早期Uターン意向にのみ関連する要素について、解釈を加えたい。まず、地元企業の認知度に関しては、ここで仮に地元企業の認知度が低い場合を考えたとき、長浜市に愛着があってもつかはUターンしたいという意向があったとしても長浜市周辺で働く具体的なイメージを持つことができず、そう遠くない将来までの具体的なライフプランとして早期Uターンを思い描くことが難しくなることが想定できる。したがって、地元企業の認知度と早期Uターン意向との関連の存在は理解可能なものであると思われる。

次に、親との関係に関しては、早期Uターンを行った場合には親と近い距離感で長く関わり続けるようになることが想定される。したがって、悩み事を気軽に話せるような良好な関係の存在が早期Uターン意向と正の関連を有していることは理解可能であると思われる。

なお、model 3においては、市内中心部の長浜地域に居住している者が早期Uターン意向を有する傾向があることもみられた。この結果の解釈としては、3.1節で述べたように長浜地域は長浜市内において相対的に都市的な性格を有していることから、若いうちは都市的な環境で暮らしたいと考えている者にとっても、長浜地域であれば長浜市が居住の選択肢に比較的に入りやすくなっているのではないかと思われる。

5.2 施策に関する考察

前節での考察を踏まえて、早期Uターン者やUターン者を増やすための施策の方向性につ

いてまとめたい。

本調査を通じて、地元企業の認知度が早期 U ターン意向と関連していることが明らかになった。したがって、これまでに広く実施されてきた中学校における職場体験等（たとえば、文部科学省, 2009）をはじめとする、中学生や高校生に地元における働く場を知ってもらうための取り組みが重要になると考えられる。なお、働く場のイメージを持ってもらうという観点では、企業等のいわゆる事業所内で働く以外に、長浜市内にはワーキングスペース等も存在することから、それらの見学や利用者へのインタビュー等を通じて長浜市内においても多様な働き方が可能であることを知ってもらうような施策も考えられよう。なお、中学生の将来居住意向に関して、地元企業の認知度が今回は有意にならなかったという結果から、中学校時代における職場体験等の取り組みが将来居住の促進において効果が無いと判断するのはやや速断である可能性があると思われる。むしろ、そのような体験でもたらされた地元企業の認知度の向上が、その後高校生になってより具体的に進路等を考えはじめた際に将来居住意向に正の影響を与えてくるという解釈も可能である。今後のさらなる研究が期待される。

本調査では、親との（良好な）関係と早期 U ターン意向や U ターン意向との関連も示唆された。これを踏まえると、たとえば親を対象とした中高生の子どもの接し方に関する講習会の開催等も U ターン施策の一つとして位置付けることができよう。なお、親の勧めの有無や程度は、長浜市が親にとって子に居住を勧めたいような地域であるかどうかにも関連していると思われる。U ターン等に関する親の意向についての先行研究としては山形県の庄内地域において実施された山口 他 (2016) があり、親自身が庄内での生活に魅力を感じているほうが子どもに庄内に残ってほしいと考える傾向があることが示されている。ただし、親の意向と子の意向との関係等は扱われておらず、今後さらなる研究の進展が望まれる。

5.3 研究の限界

最後に、本研究の限界について 2 点整理する。まず 1 点目は、中高生を対象とした先行研究で従来から指摘されているように、本研究が明らかにしたのはあくまで中生時点での U ターン意向であり、実際の U ターン行動との関連は必ずしも明らかではない点である。さらに、U ターン意向という観点だけでみても、松田・埴淵 (2023) が指摘している転出先の居住環境や転出先の地域への愛着度等のように、成人になった後の時点の U ターン意向に関連する要素は転出前の中高生段階では原理的に考慮できないものも多く、中生時点の U ターン意向は成人後に変化する可能性もある。ただし、本稿の冒頭でも述べたように、中高生の U ターン意向の調査や分析は、自治体における中高生を対象とした U ターン関連施策が中高生にどのように受け取られているかという効果を確認するための一つの指標としても有用な可能性があり、今後、何年かおきに実施して経時変化を把握する調査やパネル調査等を実施することでさらなる有用な知見を蓄積できる可能性がある。

2 点目は、本稿では U ターン意向に関連する要素を幅広く検討することを優先したため、愛着をはじめとする個別の要素に関する詳細な分析は実施できていないという点である。たとえば、愛着と関連する要素は先行研究で多数見出されており、「地域風土との接触度（鈴木・藤井, 2008）」「地域の社会的環境や物理的環境への評価（引地 他, 2009）」「居住年数や住居形態（長谷, 2013）」「通学時における人や自然との接触や寄り道場所数（藪谷・阿久井, 2021）」等がある。また、森 他 (2022, 2023) は、富山県高岡市においては愛着が「地域環境」と「文化・歴史」によって醸成されることや、富山県氷見市においては愛着が「魚・海」と「公園・交通」によって醸成されることを論じており、愛着の形成機構に地域性が存在することも示されている。これらの先行研究で得られた知見も活用しつつ、長浜市の地域性に適した愛着の醸成方法を明らかにしていくことも今後の課題である。

謝 辞

調査への協力や回答をいただいた長浜市内の中学校・高校の関係者の皆様に御礼を申し上げます。

参 考 文 献

- 秋山和久 (2017). 移住・定住促進に関連する先行調査のレビュー, 公共コミュニケーション研究, **2**(1), 44-51, https://doi.org/10.34419/pras.2.1_44.
- 遠藤健, 沖清豪 (2017). 地方における高校生の進路選択の特性と要因—「福島県高校生調査」の分析—, 早稲田教育評論, **31**(1), 101-115.
- 富士吉田市 (2023). 第6次富士吉田市総合計画(2022年度改定版), <https://www.city.fujiyoshida.yamanashi.jp/uploaded/attachment/4244.pdf> (最終アクセス日 2025年8月27日).
- 長谷起世子 (2013). 地域への定住と愛着心からみるまちづくりに関する研究—A市C地区における住民の意識分析—, 関西福祉大学社会福祉学部研究紀要, **17**(1), 51-60.
- 樋田有一郎 (2018). 地方郡部(離島・中山間地域)の高校生の地域移動志向に関する研究序説—鳥根県魅力化高校の高校生の都鄙移動とライフコースに注目して—, 早稲田大学大学院教育学研究科紀要別冊, **25**(2), 43-54.
- 引地博之, 青木俊明, 大瀨憲一 (2009). 地域に対する愛着の形成機構—物理的環境と社会的環境の影響—, 土木学会論文集D, **65**(2), 101-110, <https://doi.org/10.2208/jscejd.65.101>.
- 伊賀市 (2015). 伊賀市中学生・高校生世代アンケート調査報告書, <https://www.city.iga.lg.jp/cmsfiles/contents/0000003/3199/2-2-5anke-to.pdf> (最終アクセス日 2025年8月27日).
- 家庭教師のトライ (2025). 滋賀県の高校受験情報, <https://www.trygroup.co.jp/exam/high/shiga/> (最終アクセス日 2025年6月23日).
- 国土交通省 (2015). 国土交通白書 2015(平成26年度年次報告), <https://www.mlit.go.jp/hakusyo/mlit/h26/hakusho/h27/pdf/np102100.pdf> (最終アクセス日 2025年8月27日).
- 国立社会保障・人口問題研究所 (2018). 第8回人口移動調査結果の概要, <https://www.ipss.go.jp/ps-idou/j/migration/m08/ido8gaiyou.pdf> (最終アクセス日 2025年8月27日).
- 久慈地域キャリア教育推進研究会, 久慈市キャリア教育推進協議会 (2023). 令和3年度久慈地域高校生就職・進学意向調査結果報告書, https://www.pref.iwate.jp/_res/projects/default_project/_page_/001/053/433/r5.5_kakutei.pdf (最終アクセス日 2025年8月27日).
- マイナビ (2023). マイナビ 2024年卒大学生Uターン・地元就職に関する調査, <https://career-research.mynavi.jp/wp-content/uploads/2023/05/s-UturnReport-24-001.pdf> (最終アクセス日 2025年6月23日).
- 松田茂樹, 埴淵知哉 (2023). 大都市居住者の出身地へのUターン意向の規定要因, 地域イノベーション, **15**, 1-9, <https://doi.org/10.15002/00026746>.
- 松江市 (2024). 中学生意識調査(まちづくりのための市民アンケート 2019年度 中学生アンケート結果), https://www.city.matsue.lg.jp/material/files/group/2/chuugakusei_anke-to.pdf (最終アクセス日 2025年6月23日).
- 文部科学省 (2009). 中学校職場体験ガイド, https://www.mext.go.jp/a_menu/shotou/career/05010502/026.htm (最終アクセス日 2025年6月23日).
- 森豪大, 籾谷祐介, 宋俊煥 (2022). 高校生のシビックプライドの醸成要因と将来の定住意識に与える影響—富山県高岡市に居住する高校生を対象として—, 都市計画論文集, **57**(3), 933-940, <https://doi.org/10.11361/journalcpj.57.933>.
- 森豪大, 籾谷祐介, 宋俊煥 (2023). 農漁村地域における高校生のシビックプライドの醸成要因と将来の定住意識に与える影響—富山県氷見市に居住する高校生を対象として—, 日本都市計画学会中部

- 支部研究発表会論文集, **34**, 7–12, https://doi.org/10.11361/cpijchubu.34.0_7.
- 長浜市 (2020). 長浜市人口ビジョン, <https://www.city.nagahama.lg.jp/cmsfiles/contents/0000001/1543/jinkouvision.pdf> (最終アクセス日 2024 年 2 月 26 日).
- 長浜市 (2022). 長浜市の概況, <https://www.city.nagahama.lg.jp/0000001835.html> (最終アクセス日 2025 年 6 月 23 日).
- 労働政策研究・研修機構 (2016). UIJ ターンの促進・支援と地方の活性化—若年期の地域移動に関する調査結果, <https://www.jil.go.jp/institute/research/2016/documents/152.pdf> (最終アクセス日 2025 年 8 月 27 日).
- 政府統計の総合窓口(e-Stat) (2021). 令和 2 年国勢調査 人口等基本集計 第 2-5 表 男女, 年齢(各歳), 国籍総数か日本人別人口, 平均年齢及び年齢中位数—全国, 都道府県, 市区町村(2000 年(平成 12 年)市区町村含む), <https://www.e-stat.go.jp/stat-search/file-download?statInfId=000032142408&fileKind=0> (最終アクセス日 2025 年 6 月 23 日).
- 新発田市 (2015). 人口ビジョンおよび総合戦略の策定にかかるアンケート調査報告 若者の仕事創出, 定住に向けたアンケート調査報告書【確報版】, https://www.city.shibata.lg.jp/_res/projects/default_project/_page_/001/001/837/1001837_3_1001837_4.pdf (最終アクセス日 2024 年 2 月 8 日).
- 新温泉町 (2015). 新温泉町まちづくりアンケート《高校生》結果報告書, <https://www.town.shinonsen.hyogo.jp/uppdf/1446801640.pdf> (最終アクセス日 2025 年 8 月 27 日).
- 新温泉町 (2021). 新温泉町のまちづくりに関する高校生アンケート結果報告書, <https://www.town.shinonsen.hyogo.jp/uppdf/1647931026.pdf> (最終アクセス日 2025 年 8 月 27 日).
- 総務省 (2022). 過疎地域市町村等一覧(令和 4 年 4 月 1 日現在), https://www.soumu.go.jp/main_content/000807168.pdf (最終アクセス日 2025 年 6 月 23 日).
- 鈴木春菜, 藤井聡 (2008). 「地域風土」への移動途上接触が「地域愛着」に及ぼす影響に関する研究, 土木学会論文集 D, **64**(2), 179–189, <https://doi.org/10.2208/jscejd.64.179>.
- 對間昌宏, 浅見泰司, 石川徹 (2019). 場所への愛着に関する諸概念の尺度の相互関係, 人間・環境学会誌, **22**(1), 49–58, https://doi.org/10.20786/mera.22.1_49.
- 鶴岡市企画部政策企画課 (2023). つるおか若者意識調査 2023(集計結果), <https://www.city.tsuruoka.lg.jp/kyoiku/school-education/seisaku20231003.files/wakamono20231204.pdf> (最終アクセス日 2025 年 8 月 27 日).
- 海野遥香, 増本太郎, 寺部慎太郎, 柳沼秀樹, 田中皓介 (2022). 若年層に着目した地域愛着・街のシンボルへの意識と UJ ターン行動の関連性, 都市計画論文集, **57**(3), 1180–1185, <https://doi.org/10.11361/journalcpj.57.1180>.
- 藪谷祐介, 阿久井康平 (2021). 高校生の通学時における地域接触が地域愛着形成に与える影響, 都市計画論文集, **56**(3), 772–779, <https://doi.org/10.11361/journalcpj.56.772>.
- 山口泰史, 江崎雄治, 松山薫 (2016). 山形県庄内地域における若年人口の流出と親世代の意識, 地学雑誌, **125**(4), 493–505, <https://doi.org/10.5026/jgeography.125.493>.

Analysis of “Early U-turn Intentions” among Junior High and High School Students: A Case Study in Nagahama City, Shiga Prefecture, Japan

Yu Matsubara¹, Kentaro Hori² and Heiwa Date²

¹Data Science and AI Innovation Research Promotion Center, Shiga University

²Faculty of Data Science, Shiga University

We conducted a survey of junior high school students and high school students in Nagahama City, Shiga Prefecture, Japan, regarding their intentions to make a U-turn, that is, to return to their hometowns after living in other municipalities, focusing especially on the U-turn intention at a relatively young age such as in their 20s and 30s (early U-turn intentions). In collaboration with the Nagahama City Office, questionnaires were distributed to all junior high and high schools in the city, yielding approximately 3,000 responses. Logistic regression analyses revealed that attachment to the city, recognition of local companies, satisfaction with the city, and relationship with their parents were related to future residence intention and U-turn intention. Notably, those who intended to make an early U-turn tended to have a high level of recognition of local companies and have a relationship with their parents in which they can easily talk about their problems. The results suggest the usefulness of measures to increase their awareness of places to work and to foster better relationships between parents and students.

神経画像解析における統計学的課題と アプローチ：確率場理論

内田 航^{1,2}・青木 茂樹^{1,2}・岩崎 学^{1,2}

(受付 2025 年 7 月 1 日；改訂 8 月 23 日；採択 8 月 26 日)

要 旨

磁気共鳴画像法(MRI)に代表される神経画像技術の発展は、脳構造・機能の非侵襲的評価を可能にし、神経科学および臨床医学に飛躍的な進歩をもたらしてきた。特に、全脳探索的に脳の構造・機能的変化を統計学的に評価する手法は解剖学的・病理学的仮説に基づかない疾患評価を可能にし、疾患病態評価法として医学的にインパクトの高い知見が蓄積しつつある。しかし、神経画像データは本質的に高次元であり、全脳探索的解析における統計的推論では深刻な多重比較問題が生じる。また、隣接するボクセルで観測される MRI 特徴量間には複雑な空間的相関構造が存在し、古典的多重比較補正とは異なるアプローチの展開が求められてきた。本総説は、神経画像解析分野に特徴的な統計学的課題と、神経画像分野において長年応用されてきた確率場理論、現在の標準となっている統計学的手法の発展を概観する。

キーワード：神経画像, 多重比較, 確率場理論, permutation test, threshold-free cluster enhancement.

1. 神経画像解析における統計学的課題

磁気共鳴画像法(Magnetic Resonance Imaging; MRI)技術の目覚ましい発展は、ヒトの脳構造および機能を非侵襲的に三次元画像として捉えることを可能にし、神経科学研究や精神・神経疾患の病態解明に革新的な進歩をもたらした。特に、構造的 MRI (sMRI) は脳の形態学的特徴(例: 灰白質容積, 皮質厚などの巨視的な形態変化)の定量化を、拡散 MRI はその信号値の生物物理学的モデルに基づく神経微細構造の定量化, 機能的 MRI (fMRI) は特定の精神活動や刺激呈示に伴う脳血流動態変化の時空間的マッピングを可能にする。これらの情報は、アルツハイマー病, パーキンソン病などの神経変性疾患 (Uchida et al., 2023; Andica et al., 2023), 統合失調症, うつ病, 自閉症などの精神疾患といった (Nagai et al., 2024), 多様な疾患の早期診断マーカー開発や病態基盤の解明に加え、健常脳機能の理解深化に不可欠なものとなっている。例えば、本稿の後半(6.2 節)で紹介するタウオパチー研究のように、拡散 MRI を用いて神経線維の密度や太さといった微細構造を定量化し、臨床症状が類似する疾患群間で脳内のどこに違いが生じるのかを全脳にわたり探索的に比較することが可能となった (Uchida et al., 2023)。こうした探索的アプローチは、疾患の病態解明や客観的な診断マーカー開発への貢献が期待さ

¹ 順天堂大学 健康データサイエンス学部：〒279-0013 千葉県浦安市日の出 6 丁目 8-1; w-uchida@juntendo.ac.jp, saoki@juntendo.ac.jp, m.iwasaki.vs@juntendo.ac.jp

² 順天堂大学 大学院医学研究科：〒113-8421 東京都文京区本郷 2 丁目 1-1

れる。

一方で、これらの神経画像データから科学的に妥当な知見を抽出するためには、統計学的に特有かつ深刻な課題を克服する必要がある。まず、MRI データは本質的に極めて高次元である。例えば、Human connectome project で公開されている拡散 MRI では、1.25mm 立方のボクセル (voxel; 3 次元画像の最小単位) 解像度で脳全体を撮像し、 $145 \times 174 \times 145 = 3,658,350$ 個もの観測点が得られている (Glasser et al., 2013)。fMRI では、これに加えて時間軸方向の観測が数百点加わり、データはさらに高次元化する。神経・精神疾患を対象とした神経画像研究においては、一般に全脳のどこに脳の機能的・構造的変化が存在するのか自明でない場合が多く、探索的な病態評価の側面が強い。この場合、全脳ボクセルに渡る網羅的な統計学的評価が重要になるが、先の特徴から多重比較の問題が常態化していると言えよう。古典的には、関心のある疾患において仮説に基づいて脳の生物学的に意味のある局所的な構造を Region of interest (ROI) として定義し、局所領域における平均的な特徴量を抽出することで、この問題に対処してきた。しかし、このアプローチでは事前に定義した ROI 内の特徴量の平均 (または中央値など) を算出するため、探索的な観点からは不十分であり、ボクセルレベルの探索的アプローチが求められる。以上のように、MRI のような大規模かつ複雑なデータ構造に対して探索的に統計的推論を行う際、各ボクセルで仮説検定を同時に実行することに伴う多重比較問題が深刻な課題となるのである。

MRI 画像に特徴的なもう一つの性質に、各ボクセルにおける信号値間には複雑な空間的相関構造が存在する点が挙げられる。これは、脳組織の解剖学的連続性により隣接する細胞が類似した構造・機能特性を持つこと、撮像時の点拡がり関数による空間的平滑化効果などに起因する。この空間的相関の存在下では、従来の多重比較補正法は適切に機能せず、統計的推論の妥当性が著しく損なわれる可能性がある。また、解剖学的にも、その変化は単一のボクセルではなくある程度の空間的広がりを持つクラスターとして現れることが期待されるため、クラスターレベルでの推論が有効となる。

本稿では、神経画像データにおいて、統計学的に考慮すべき特徴と派生する課題、その課題に対処すべく発展してきた統計学的手法の応用について概説する。

2. 神経画像分野における統計モデリングの基本と前処理

2.1 一般線形モデル (General Linear Model)

神経画像解析における全脳探索的解析においては、各ボクセル k において一般線形モデル (General Linear Model; GLM) が広く用いられる。 n 人の被験者からなるデータセットにおいて、ボクセル k での観測値を要素とするベクトルを \mathbf{Y}_k とすると、GLM は以下のように定式化される：

$$(2.1) \quad \mathbf{Y}_k = \mathbf{X}\boldsymbol{\beta}_k + \boldsymbol{\epsilon}_k, \quad k = 1, 2, \dots, V$$

ここで、

- \mathbf{Y}_k : ボクセル k における $n \times 1$ の観測データベクトル (MRI 特徴量)。
- \mathbf{X} : $n \times q$ の計画行列 (design matrix)。各列が説明変数 (例: 群のダミー変数, 年齢, 性別など) を表し、全てのボクセルで共通に用いられることが多い。群のダミー変数は、例えば、ある疾患群と健常対照群の間で各ボクセルの特徴量に差があるかを検証する場合に設定される。
- $\boldsymbol{\beta}_k$: ボクセル k における $q \times 1$ の未知の回帰係数ベクトル。これが推定対象となる。
- $\boldsymbol{\epsilon}_k$: $n \times 1$ の誤差ベクトル。

誤差ベクトル ϵ_k の各要素は、互いに独立に平均 0, 分散 σ^2 の正規分布に従うと仮定される。パラメータ β_k は通常、最小二乗法(または重み付き最小二乗法)によって推定される： $\hat{\beta}_k = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_k$ 。特定の仮説を検証するためには、コントラストベクトル \mathbf{c} (例えば、2 群間比較の場合、 $[1, -1, 0, \dots, 0]^T$)を用いて、 $\mathbf{c}^T \beta_k$ の値を評価する。帰無仮説 $H_0: \mathbf{c}^T \beta_k = 0$ の下で、検定統計量(例：t 統計量や F 統計量)が各ボクセルで計算される。例えば、t 統計量は、

$$(2.2) \quad T_k = \frac{\mathbf{c}^T \hat{\beta}_k}{\sqrt{\widehat{\text{Var}}(\mathbf{c}^T \hat{\beta}_k)}} = \frac{\mathbf{c}^T \hat{\beta}_k}{\sqrt{\hat{\sigma}_k^2 \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}}$$

として得られ、帰無仮説の下で自由度 $n - q$ の t 分布に近似的に従う。添字 k はボクセルのインデックスを表し、 T_k の値は脳内の全ボクセルについて個別に算出される。各ボクセルで算出された T_k の値を空間座標場に配置した統計マップが、後続の多重比較補正の対象となる。なお、第 4 章で述べる確率場理論では、この統計マップを確率場 $Z(\mathbf{x})$ の実現値として扱う。

2.2 前処理の統計学的意義と影響

生の MRI データは、ノイズやアーチファクトを多く含み、また被験者間で脳の形状やサイズ、位置が異なるため、統計解析を行う前に一連の前処理が必要となる。これらの前処理ステップは、後続の統計解析の妥当性や感度に大きな影響を与える。本稿では、今日まで神経画像解析において広く利用されている Statistical Parametric Mapping (SPM) (Friston, 2003) における一般的な前処理についてまとめる。SPM は神経画像解析に必要な標準的な前処理に加え、統計解析までの一連の解析を実行可能なソフトウェアであり、今日まで広く利用されている。MRI 画像でしばしば見られるアーチファクトは、装置特性や撮像原理に依存して生じる。例えば、磁場不均一に起因する感度ムラや、高周波成分の打ち切りに起因する Gibbs アーチファクトなどがあり、種々補正アプローチが提案されているが (Tustison et al., 2010; Kellner et al., 2016), SPM には実装されていない。実際に解析する MRI 画像に見られるアーチファクトに応じて個別に対処する必要があるが、本稿では詳細な方法論についての説明は割愛する。

2.2.1 空間標準化(Spatial Normalization)

異なる被験者の脳画像を共通の座標系(標準脳テンプレート、例えば MNI 座標系)に位置合わせする処理である。これにより、特定の脳部位に対応するボクセル座標が被験者間で概ね一致し、群レベルでの全脳探索的な統計解析が可能になる。標準化は、通常、線形変換(平行移動、回転、スケールリング、せん断)と非線形変換(局所的な変形)を組み合わせで行われる。標準化の精度は、特に小領域の解析において結果の信頼性に直結し、ボクセル毎の検定を行う場合は 1 ボクセルを構成する数 mm 以内の誤差を担保する必要があるが、技術的に難しく、SPM では次項の空間平滑化で対処する。

2.2.2 空間平滑化(Spatial Smoothing)

画像に対してガウシアンカーネルなどの空間フィルタを適用し、意図的に画像をぼかす処理である。平滑化の主な目的は以下の通りである：

- (1) 信号対雑音比の向上：ランダムノイズを平均化することで、信号の検出感度を上げる。
- (2) データ分布の正規性改善：中心極限定理に似た効果により、データの分布を正規分布に近づける。これは確率場理論(Random Field Theory, 以下 RFT) のガウス性仮定を満たす助けとなる。
- (3) 空間標準化の残存誤差の緩和：標準化後も残る微小な解剖学的位置ずれの影響を軽減し、群内での対応付けを改善する。

(4) **RFT** の仮定充足：RFT が適用可能となるために必要なデータの平滑性を確保する。

しかし、過度な平滑化は空間解像度を低下させ、小さな活性化領域を見逃す原因となる。平滑化カーネルのサイズ選択は、期待される効果の空間的広がりとのトレードオフの関係にあり、結果に大きな影響を与える重要なパラメータである。そのため、目的となる脳構造により最適化が必要であるが、一般に全脳探索的な統計解析を実施する場合には、慣習的に半値全幅(full width at half maximum, 以下 FWHM) = 8mm 程度に設定される場合が多い。

3. 神経画像解析における多重比較問題

神経画像解析における全脳探索的研究では、脳内の各ボクセルにおいて、特定の効果の有無に関する仮説検定を同時に行う。例えば、ある疾患群と健常対照群の間で脳の局所的な灰白質容積に差があるかを検証する場合、数万オーダーのボクセルそれぞれに対して帰無仮説の検定を実施することになる。仮に、各ボクセルでの検定を有意水準 $\alpha = 0.05$ で行うと仮定する。もし全ての帰無仮説が真であるとしても、1回の検定で偽陽性を犯す確率は α であるため、全ボクセル V 個での偽陽性の期待数は、 $E[\text{偽陽性数}] = V \times \alpha$ となる。前述の $V = 145 \times 174 \times 145$ の場合、 $3,679,375 \times 0.05 \approx 183,969$ 個となり、到底容認できない数の偽陽性が期待される。これは、神経画像研究の信頼性を著しく損なう。この問題に対処するため、多重比較補正が実施され、第1種の過誤確率(Family-Wise Error Rate; FWER)の制御が頻用される。

3.1 Bonferroni 補正とその保守性

最も単純な FWER 制御法は Bonferroni 補正である。これは、個々の検定の有意水準を $\alpha_{\text{voxel}} = \alpha_{\text{FWER}}/V$ と設定するものである。これらの手法は、検定統計量がどのような依存構造を持っていても(あるいは独立であっても)FWER を α_{FWER} 以下に制御するが、検定統計量間に正の相関が存在する場合には極めて保守的となる。脳画像データでは、隣接ボクセル間の統計量は生物学的・物理学的要因から強く正に相関している点を考慮すると、実効的に独立な検定の数は実際のボクセル数よりもはるかに少ないはずである。そのため、Bonferroni 補正を適用すると、個々のボクセルに課される有意水準が過度に厳しくなり、真の効果を見逃す第二種の過誤確率が増大し、検出力が著しく低下する。空間相関を考慮する試みの一つとして、実効的な独立検定数を推定し、Bonferroni 補正を行う方法 (Nyholt, 2004)があるが、安定した推定が困難であり、神経画像領域では広く利用されるに至っていない。

3.2 ファミリーワイズエラー率 (Family-Wise Error Rate; FWER)

FWER は、検定のファミリー(ここでは全 V 個の検定)において、少なくとも1つの真の帰無仮説を誤って棄却する確率を指す。ここで、前述の一般線形モデルに基づく検定統計量を T_k とし、ある閾値 u を超える場合は帰無仮説 \mathcal{H}_0 を棄却するとする。この時、FWER は下記で表される。

$$(3.1) \quad \text{FWER} = Pr \left(\bigcup_{k \in \mathcal{H}_0} \{|T_k| \geq u\} \right) = Pr \left(\max_{k \in \mathcal{H}_0} |T_k| \geq u \right)$$

ここで \mathcal{H}_0 は真の帰無仮説の集合(効果が実際に存在しないボクセルのインデックス集合)、 u は棄却のための閾値である。すなわち、FWER は帰無仮説 \mathcal{H}_0 の下で、検定統計量 T_k の最大値分布として表される。これは、閾値 u を超えるクラスターが一つ以上存在することと同値である。次章に示す RFT では、この最大値分布を求める。

4. 確率場理論(Random Field Theory, RFT)

神経画像データの持つ空間的相関構造を考慮し、Bonferroni 補正の保守性を克服する試みとして、1990 年代に RFT が導入された (Worsley et al., 1992; Friston et al., 1994a, 1994b). RFT は、ボクセルごとの統計値 (t 値, F 値など) からなるマップを連続的な確率場 (Random Field) の離散的な実現値とみなし、その確率場のトポロジカルな特性 (特にオイラー標数) を利用して、閾値を超える領域の最大値分布やクラスターサイズの分布を導出することで、空間的相関を考慮した FWER 制御を行う。この手法は SPM などの主要な解析ソフトウェアパッケージに実装され、長らく神経画像統計解析の標準的手法として広く用いられてきた。

4.1 ガウス確率場とオイラー標数

神経画像解析領域においては、ボクセルごとの統計量がガウス確率場 (Gaussian Random Field; GRF) に従うという仮定に依存している。GRF $Z(\mathbf{x}), \mathbf{x} \in S \subset \mathbb{R}^D$ (ここで D は空間次元、神経画像では通常 3 次元) は、探索空間 S (例: 全脳領域) 内の任意の位置 \mathbf{x} (個々のボクセル位置に対応) における統計値 $Z(\mathbf{x})$ が、どの有限個の点の組 $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_m))$ をとっても、それらが多変量正規分布に従うような確率場である。

RFT では、この確率場がある高い閾値 u を超える領域、すなわち excursion 集合 $A_u = \{\mathbf{x} \in S : Z(\mathbf{x}) \geq u\}$ の幾何学的・トポロジカルな特性を利用する。特に重要なのがオイラー標数 (Euler Characteristic; EC) $\chi(A_u)$ である。3 次元の場合、オイラー標数は、

$$(4.1) \quad \chi(A_u) = (\text{連結成分の数}) - (\text{穴の数}) + (\text{空洞の数})$$

と定義される。神経画像解析の文脈では、非常に高い閾値 u に対しては、「穴」や「空洞」は稀であり、 $\chi(A_u)$ は閾値 u を超えるボクセル群が形成する連結した塊、すなわちクラスターの数に近似できる。図 1 では、例としてある 2 次元確率場と 3 つの異なる閾値を適応した場合のクラスターとオイラー標数の変化を示している。図 1(a) において、 x 軸と y 軸に空間的座標が対応し、 z 軸方向に検定統計量が示されている。図 1(b) のように、ある閾値 u におけるオイラー標数 χ_u は閾値 u に依存して変化している様子が見て取れる。2 次元の場合、閾値を超える白い連結成分の数から、その中に存在する黒い領域 (穴) の数を引くことでオイラー標数が計算される。例えば、図 1(b) の $Z > 0.18$ においては、連結した白いクラスタの数は 6 であり、穴の数は 3 であるため、オイラー標数は $6 - 3 = 3$ となる。3 次元の場合も同様に、連結した塊の数から、ドーナツのようなトンネル状の構造 (穴) の数と、球殻のような内部に閉じられた空間 (空洞) の数を加減することで計算される。

Worsley ら (Worsley et al., 1992, 1996) により、GRF のオイラー標数の期待値 $E[\chi(A_u)]$ は、閾値 u 、探索空間 S の体積と形状、そして確率場の平滑性 (smoothness) の関数として解析的に与えられることが示された。実用上、この期待値は、探索空間の体積 $|S|$ と、画像の平滑性を示す RESEL (Resolution Elements) カウント R を用いて表現されることが多い。3 次元の神経画像の場合：

$$(4.2) \quad E\chi(A_u) \approx \frac{R(4 \ln 2)^{3/2}}{(2\pi)^2} (u^2 - 1) \exp\left(-\frac{u^2}{2}\right)$$

RESEL カウントは、 $R = \frac{|S|}{(\text{FWHM}_x \text{FWHM}_y \text{FWHM}_z)}$ で定義され、画像の各方向の半値全幅 (FWHM) を考慮した実効的な空間解像度の単位である。データの空間的平滑性の推定には、一般線形モデル (GLM) の残差の空間勾配から推定されるのが一般的である (Kiebel et al., 1999)。この推定の精度は RFT の妥当性に直接影響するため、適切な平滑化処理と正確な平滑性推定が重要となる。

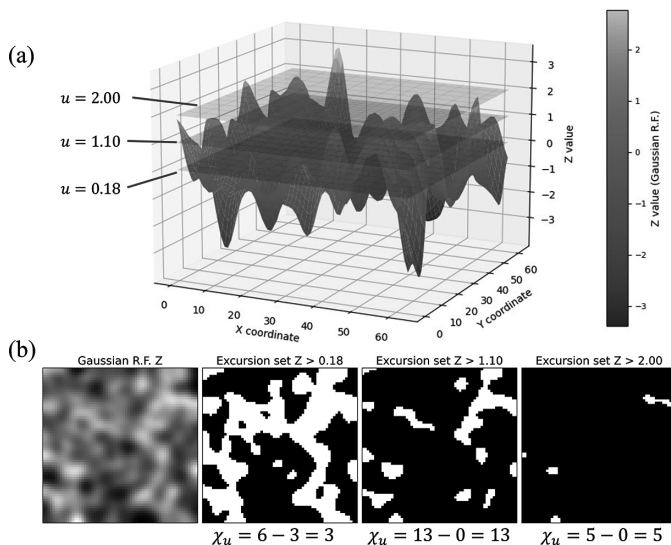


図 1. (a)生成された 2 次元ガウス確率場の 3 次元サーフェスプロット．図中には、3 つの異なる閾値 ($u = 0.18, 1.10, 2.00$) が半透明な水平平面として重ねて表示されている．カラーバーは確率場の Z 値のスケールを示す．(b)左端の図は、(a)で示された元の 2 次元ガウス確率場 (Gaussian R.F. Z) をグレースケールで表示したヒートマップ．続く 3 つの図は、それぞれ閾値 $u = 0.18, 1.10, 2.00$ における excursion 集合 (確率場の値 Z が閾値 u を超える領域、すなわち $Z > u$) を二値画像で示したものである．これらの二値画像では、閾値を超える領域が白で、それ以外の領域が黒で表示されている．各 excursion 集合の下部には、その領域の連結成分数から穴の数を引いて計算されたオイラー標数 (χ_u) が示されている．

4.2 ボクセルレベル FWER 制御 (Peak Height Inference)

RFT によれば、帰無仮説 (すなわち、コントラストベクトル c で定義される効果 β が、全てのボクセルで一様にゼロであるという仮説) の下で、確率場の最大値 $Z_{\max} = \max_{\mathbf{x} \in S} Z(\mathbf{x})$ がある閾値 u を超える確率 $Pr(Z_{\max} \geq u)$ は、高い u に対して $E[\chi(A_u)]$ で近似できる：

$$(4.3) \quad \text{FWER}_{\text{voxel}} = Pr(Z_{\max} \geq u) \approx E[\chi(A_u)]$$

神経画像解析では、 Z_{\max} は全脳で観測された統計値マップ中の最大値 (ピークの高さ) に対応する．この近似を用いて、指定した FWER 水準 α (例：0.05) を与える閾値 u_α を逆算することができる．個々のボクセルの統計値がこの u_α を超えた場合に、そのボクセルは統計的に有意であると判定する．

4.3 クラスターレベル FWER 制御 (Cluster Extent Inference)

ボクセルレベルの推論は、個々のボクセルでの信号強度が非常に高い場合に有効だが、空間的に広がった比較的弱い信号の検出には不向きである．そこで、RFT はクラスターレベルの推論も提供する．これは以下の手順で行われる：

- (1) まず、個々のボクセルで検定統計量 T_k を計算する．
- (2) 次に、任意のクラスター定義閾値 u_{CDT} (CDT, Cluster-Defining Threshold) (例：対応する片側 p 値が 0.001 や 0.01 となるような統計値) を設定する．

- (3) $T_k \geq u_{\text{CDT}}$ を満たすボクセルを特定し、隣接する(例：SPM では 18 連結で定義)有意ボクセルを連結してクラスターを形成する。
- (4) 形成された各クラスターについて、そのサイズ(体積またはボクセル数) S を計算する。
- (5) 帰無仮説の下で、サイズ s 以上のクラスターが偶然生じる確率 $Pr(S_{\text{max}} \geq s)$ を RFT に基づいて評価し、FWER 制御された p 値をクラスターごとに算出する。ここで S_{max} は観測されうる最大のクラスターサイズである。

クラスターサイズ S の分布に関して、Friston et al. (1994a) は、帰無仮説のもとで観測される最大のクラスターサイズが s 以上である確率 $Pr(S_{\text{max}} > s)$ は、

$$(4.4) \quad P(S_{\text{max}} \geq s) = 1 - \exp[-E[m] \cdot e^{-\beta s^{2/D}}]$$

と近似されることを示した。ここで $E[m]$ は CDT の設定による閾値 u_{CDT} を超えるクラスターの期待数(これも $E[\chi(A_{u_{\text{CDT}}})]$ から導かれる)、 $e^{-\beta s^{2/D}}$ はある特定のクラスターのサイズが s を超える条件付き確率であり、近似的に $\Pr(-\beta s^{2/D})$ 、 β は $\beta = [\Gamma(D/2 + 1) \cdot E[m]/E[N]]^{2/D}$ (D は空間次元、 $E[N]$ は閾値を超えるボクセルの期待数)の形で与えられる。この理論に基づき、観測されたクラスターサイズに対する補正済み p 値が計算される。

5. RFT の限界とノンパラメトリック多重比較補正への展開

RFT は神経画像解析における多重比較補正の標準的手法として長年用いられてきたが、その理論的妥当性はいくつかの仮定に依存している。RFT の主な仮定には、(1)統計量が(少なくとも裾の部分で)ガウス確率場(または関連する確率場)として適切に近似できること、(2)確率場が連続かつ十分に滑らかであること(一般に FWHM がボクセルサイズの 2-3 倍以上)、(3)確率場の相関構造が脳全体で均一であること、などがある。これらの仮定、特にガウス性と平滑性の仮定は、実データでは必ずしも満たされない。また、クラスターを定義するための閾値(CDT)設定の恣意性も問題となる。このような背景から、データの分布に関する仮定をほとんど置かないノンパラメトリックなアプローチ、CDT 設定の恣意性を克服するアプローチを組み合わせた手法が注目されるようになった。代表的なものとして、Permutation Test と、それと組み合わせる用いられることが多い閾値フリークラスター強調(Threshold-Free Cluster Enhancement; TFCE)がある。ここでは概要に留めるが、これらの手法は、近年の神経画像コミュニティにおいて標準的な解析手法の一つとなっている。

5.1 Permutation Test

Permutation test は、R. A. Fisher によって提唱されたノンパラメトリックな統計的推論手法であり、検定統計量の帰無分布に関して特定の確率分布を仮定しない(Fisher, 1935; Nichols and Holmes, 2002; Winkler et al., 2014)。具体的には、群ラベルをランダムに並べ替える操作を多数回行い、各データセットに対して検定統計量を計算することで、帰無仮説の下での検定統計量の経験的な帰無分布を構築する。多重比較補正に適用する際は、各データセットにおいて「ファミリー全体で最も極端な統計量」(例えば、全ボクセル中の最大 t 値や、後述する最大 TFCE 値)を求め、その分布を構築する(Westfall and Young, 1993)。観測データから得られた統計値が、この最大統計量の帰無分布の上位 $\alpha \times 100$ パーセンタイル点を超えるかどうかで有意性を判断する。この手順により、データの分布形状や空間相関構造について明示的な仮定を置くことなく、FWER を制御できる。年齢や性別など複数の共変量を含む場合、単純な群ラベルの並べ替えでは説明変数の対応が崩れるため、異なる手法が用いられる(Winkler et al., 2014)。詳細は割愛するが、まず共変量のみで説明される効果を回帰により取り除き、その残

り(残差)を被験者間でランダムに並べ替える。このアプローチにより、共変量の影響を適切に統制しながら、興味のある特定の変数(例えば群のダミー変数)の効果を頑健に検定することが可能となる。

5.2 Threshold-free cluster enhancement (TFCE) の概念

従来法では、クラスターを定義するための閾値(CDT)を研究者が任意に設定する必要があるが、CDT 閾値の選択によって結果が大きく変わりうるという問題があった(Woo et al., 2014)。閾値フリークラスター強調(TFCE)は、Smithら(Smith and Nichols, 2009)によって提案された、CDT の恣意性問題を克服するための手法である。

TFCE の基本的な考え方は、各ボクセルの元の統計値を、そのボクセルが属する全ての可能なクラスターの「高さ(元の統計値)」と「広がり(クラスターサイズ)」の情報を統合的に反映するように変換することである。具体的には、あるボクセルに対して、考える様々な閾値レベルでそのボクセルが含まれるクラスターの大きさとその閾値自体の高さとを掛け合わせ、それらの値を低い閾値からそのボクセルの統計値まで積分する。数式での詳細な定義は原著(Smith and Nichols, 2009)に譲るが、この処理の結果、TFCE マップでは、元の統計マップにおいて局所的に高い信号強度を持つだけでなく、ある程度の空間的連続性・広がりを持って存在する信号領域が特に高い値を示すようになる。すなわち、単に「狭く鋭いピーク」を持つ領域や、「低強度だが広範囲」な領域よりも、信号強度(高さ)と空間的広がり(裾野)の双方をバランス良く兼ね備えた構造が強調される。これにより、特定の CDT を事前に主観的に選択する必要なく、統計的に意味のあるクラスター様の信号を強調し、その後の統計的推論における検出感度を高めることが期待される。

TFCE によって得られた強調統計マップと Permutation test を組み合わせることで、従来法の課題を克服し、近年の神経画像コミュニティにおいて標準的な解析手法の一つとして広く受け入れられつつあり、多くの主要な解析ソフトウェア(例: FMRIB Software Library [FSL] など)に実装されている。ただし、計算コストが高いという課題があり、効率的な実装や計算資源が求められる。

6. 臨床応用研究事例

神経画像統計解析手法は、基礎研究における脳機能・構造の理解深化に貢献するだけでなく、実際の臨床現場における疾患の診断支援や治療効果の客観的評価など、医学的応用も進んでいる。本節では、その代表的な事例を紹介する。

6.1 アルツハイマー病早期診断支援システム VSRAD

アルツハイマー病(AD)は、早期発見と早期介入が進行抑制に重要とされる神経変性疾患である。VSRAD(Voxel-based Specific Regional analysis system for Atrophy Detection)は、確率場理論に基づく多重比較補正は適応されていないが、前述の SPM の技術的基盤に基づいて脳の萎縮の程度を客観的かつ定量的に評価し、AD の早期診断を支援するために日本で開発されたシステムとして臨床的に広く利用されている(Matsuda et al., 2012)。VSRAD は、その客観的な評価指標と比較的簡便な操作性から、日本国内において医薬品医療機器等法(旧薬事法)に基づく医療機器としての承認を受け、2015 年時点で約 3000 の医療機関で AD の診断補助ツールとして臨床応用されている(松田, 2016)。VSRAD は、医師による視覚的読影に加えて定量的な萎縮情報を提供することで、より客観性の高い AD 診断の一助となっている。その有用性は複数の研究で示されており、例えば、Waragai et al. (2014)は健常高齢者と比較した AD の

診断能において感度 84.1%，特異度 84.6% を報告している。

6.2 タウオパチー研究

臨床応用研究の一例として，Uchida et al. (2023) らによる，進行性核上性麻痺 (Progressive Supranuclear Palsy; PSP) および大脳皮質基底核症候群 (Corticobasal Syndrome; CBS) 患者における白質の微細構造変化に関する研究を挙げる。PSP, CBS はいずれもタウオパチー (脳内に異常なタウタンパク質が蓄積し，その神経毒性に関連する神経変性を病態に持つ神経変性疾患群の総称) に分類される神経変性疾患であり，臨床的には病態が重複するため鑑別が困難な場合があるため，両疾患の鑑別に有用な客観的マーカーの開発が試みられている (Ling and Macerollo, 2018)。この研究では，PSP 患者群，CBS 患者群，および健常対照群を対象に，拡散 MRI を用いて計算された神経細胞微細構造を反映する特徴量の全脳探索的解析 (Fixel-based analysis, FBA) を行った。FBA では神経微細構造の特徴量として線維密度 (FD)，線維径 (log-FC)，両者の積 (FDC) が抽出可能であり，従来法との比較として萎縮などの巨視的な変化を捉える Voxel-based morphometry (VBM) と古典的な微細構造指標である Fractional Anisotropy (FA) の探索的な比較も行った。VBM, FA の検定は前述の RFT に基づく。FBA の統計解析においては，FBA で得られた統計値マップに対し，TFCE を拡張した Connectivity-based Fixel Enhancement (CFE) (Raffelt et al., 2015) を適用することでクラスター様の信号を強調し，その後 Permutation Test を実行することで，ファミリーワイズエラー率 (FWER) を厳密に制御した全脳探索的な多重比較補正を行った。その結果，PSP 患者群では健常対照群と比較して，上小脳脚 (SCP) や皮質脊髄路 (CST) を中心に神経細胞構造特徴量の有意な低下が認められた。一方，CBS 患者群では運動野下の白質，CST，上縦束 (SLF)，脳梁などでより広範な特徴量の低下が示された (図 2)。これらの変化は両疾患の病理学的な知見と一致した脳領域に対応しており，非侵襲的 MRI 検査を用いた探索的な統計画像解析により，両疾患に特徴的な神経微細構造変化を捉えられる可能性が示唆された。さらに，SCP における細胞損傷は病理研究で PSP により特徴的であるこ

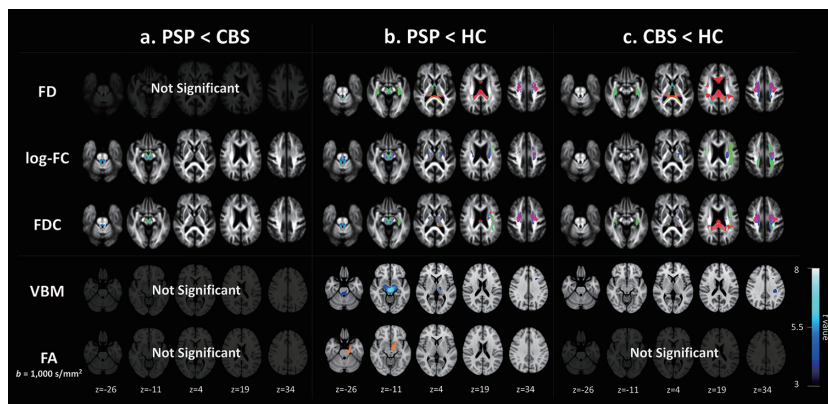


図 2. 進行性核上性麻痺 (PSP) と大脳皮質基底核症候群 (CBS) における白質微細構造変化の比較。(a) CBS と比較して PSP で有意に低値を示した領域，(b) 健常者 (HC) と比較して PSP で有意に低値を示した領域，(c) HC と比較して CBS で有意に低値を示した領域を示している (Uchida et al., 2023 より引用 (CC-BY))。図中上 3 つのパネルの有意領域の色は，神経線維が走行する方向に基づき RGB で示されている (赤：左右方向，青：上下方向，緑：前後方向)。また，図中下 2 つのパネルの有意領域の色は，検定統計量 t に基づいてカラースケール (最大値 8，最小値 3) で示されている。

とが示されており, SCPにおけるMRI指標は臨床的に鑑別が難しいPSPとCBSの鑑別に寄与した。また, その変化は運動機能障害や認知機能障害といった臨床症状の重症度とも関連することが示された。この研究は, FBAとノンパラメトリックな統計手法を組み合わせることで, タウオパチーにおける疾患特異的かつ進行性の白質微細構造変化を感度良く捉え, 疾患の鑑別・予後予測マーカーとしてのその臨床的意義を示した例といえる。

7. まとめと展望

本総説では, 神経画像分野における統計解析手法, 特に多重比較問題への対応に焦点を当て, 古典的な確率場理論(RFT)から, 現代の標準となりつつあるTFCE-Permutation Testに至るまでの発展を概観した。神経画像統計解析の分野における統計学のアプローチは, 今後もさらなる発展が期待される。神経画像統計解析は, データ科学の一分野として, また現代神経科学を支える基盤技術として, 今後ますますその重要性を増していくであろう。本総説が, 当該分野への理解を深め, 統計学と神経科学のさらなる発展を促進するための一助となれば幸いである。

謝 辞

本研究は, 統計エキスパート人材育成プロジェクトの助成を受けた報告である。

参 考 文 献

- Andica, C., Kamagata, K., Uchida, W., Saito, Y., Takabayashi, K., Hagiwara, A., Takeshige-Amano, H., Hatano, T., Hattori, N. and Aoki, S. (2023). Fiber-specific white matter alterations in Parkinson's disease patients with GBA gene mutations, *Movement Disorders*, **38**(11), 2019–2030, <https://doi.org/10.1002/mds.29578>.
- Fisher, R. A. (1935). *The Design of Experiments*, Oliver & Boyd, Edinburgh.
- Friston, K. J. (2003). *Statistical Parametric Mapping*, 237–250, Springer US, Boston, Massachusetts, https://doi.org/10.1007/978-1-4615-1079-6_16.
- Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C. and Evans, A. C. (1994a). Assessing the significance of focal activations using their spatial extent, *Human Brain Mapping*, **1**(3), 210–220, <https://doi.org/10.1002/hbm.460010306>.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D. and Frackowiak, R. S. (1994b). Statistical parametric maps in functional imaging: A general linear approach, *Human Brain Mapping*, **2**(4), 189–210, <https://doi.org/10.1002/hbm.460020402>.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R. et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project, *Neuroimage*, **80**, 105–124, <https://doi.org/10.1016/j.neuroimage.2013.04.127>.
- Kellner, E., Dhital, B., Kiselev, V. G. and Reiser, M. (2016). Gibbs-ringing artifact removal based on local subvoxel-shifts, *Magnetic Resonance in Medicine*, **76**(5), 1574–1581, <https://doi.org/10.1002/mrm.26054>.
- Kiebel, S. J., Poline, J.-B., Friston, K. J., Holmes, A. P. and Worsley, K. J. (1999). Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model, *NeuroImage*, **10**(6), 756–766, <https://doi.org/10.1006/nimg.1999.0508>.
- Ling, H. and Macerollo, A. (2018). Is it useful to classify PSP and CBD as different disorders? Yes, *Movement Disorders Clinical Practice*, **5**(2), 145–148, <https://doi.org/10.1002/mdc3.12581>.
- 松田博史 (2016). VSRAD®: 精神科臨床へのニューロイメージングの最新応用, *総合病院精神医学*, **28**(3), 210–218, <https://doi.org/10.11258/jjghp.28.210>.

- Matsuda, H., Mizumura, S., Nemoto, K., Yamashita, F., Imabayashi, E., Sato, N. and Asada, T. (2012). Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated lie algebra improves the diagnosis of probable Alzheimer disease, *American Journal of Neuroradiology*, **33**(6), 1109–1114, <https://doi.org/10.3174/ajnr.A2935>.
- Nagai, Y., Kirino, E., Tanaka, S., Usui, C., Inami, R., Inoue, R., Hattori, A., Uchida, W., Kamagata, K. and Aoki, S. (2024). Functional connectivity in autism spectrum disorder evaluated using rs-fMRI and DKI, *Cerebral Cortex*, **34**(13), 129–145, <https://doi.org/10.1093/cercor/bhad451>.
- Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples, *Human Brain Mapping*, **15**(1), 1–25, <https://doi.org/10.1002/hbm.1058>.
- Nyholt, D. R. (2004). A simple correction for multiple testing for SNPs in linkage disequilibrium with each other, *The American Journal of Human Genetics*, **74**(4), 765–769, <https://doi.org/10.1086/383251>.
- Raffelt, D. A., Smith, R. E., Ridgway, G. R., Tournier, J.-D., Vaughan, D. N., Rose, S., Henderson, R. and Connelly, A. (2015). Connectivity-based fixel enhancement: Whole-brain statistical analysis of diffusion MRI measures in the presence of crossing fibres, *Neuroimage*, **117**, 40–55, <https://doi.org/10.1016/j.neuroimage.2015.05.039>.
- Smith, S. M. and Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence, and localisation in cluster inference, *NeuroImage*, **44**(1), 83–98, <https://doi.org/10.1016/j.neuroimage.2008.03.061>.
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. and Gee, J. C. (2010). N4ITK: Improved N3 bias correction, *IEEE Transactions on Medical Imaging*, **29**(6), 1310–1320, <https://doi.org/10.1109/TMI.2010.2046908>.
- Uchida, W., Kamagata, K., Andica, C., Takabayashi, K., Saito, Y., Owaki, M., Fujita, S., Hagiwara, A., Wada, A., Akashi, T. et al. (2023). Fiber-specific micro-and macroscopic white matter alterations in progressive supranuclear palsy and corticobasal syndrome, *npj Parkinson's Disease*, **9**(1), <https://doi.org/10.1038/s41531-023-00565-2>.
- Waragai, M., Hata, S., Suzuki, T., Mizukami, K., Asada, T. and Matsuda, H. (2014). Utility of SPM8 plus DARTEL (VSRAD) combined with magnetic resonance spectroscopy as adjunct techniques for screening and predicting dementia due to Alzheimer's disease in clinical practice, *Journal of Alzheimer's Disease*, **41**(4), 1207–1222, <https://doi.org/10.3233/JAD-132786>.
- Westfall, P. H. and Young, S. S. (1993). *Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment*, John Wiley & Sons, Edinburgh.
- Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. and Nichols, T. E. (2014). Permutation inference for the general linear model, *NeuroImage*, **92**, 381–397, <https://doi.org/10.1016/j.neuroimage.2014.01.060>.
- Woo, C.-W., Krishnan, A. and Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations, *NeuroImage*, **91**, 412–419, <https://doi.org/10.1016/j.neuroimage.2013.12.058>.
- Worsley, K. J., Evans, A. C., Marrett, S. and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain, *Journal of Cerebral Blood Flow & Metabolism*, **12**(6), 900–918, <https://doi.org/10.1038/jcbfm.1992.127>.
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J. and Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation, *Human Brain Mapping*, **4**(1), 58–73, [https://doi.org/10.1002/\(SICI\)1097-0193\(1996\)4:1<58::AID-HBM4>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O).

Challenges and Advances in Neuroimaging Statistical Analysis: Random Field Theory

Wataru Uchida^{1,2}, Shigeki Aoki^{1,2} and Manabu Iwasaki^{1,2}

¹Faculty of Health Data Science, Juntendo University

²Graduate School of Medicine, Juntendo University

The advancement of neuroimaging techniques, notably Magnetic Resonance Imaging (MRI), has enabled non-invasive assessment of brain structure and function, bringing remarkable progress to neuroscience and clinical medicine. In particular, statistical methods for whole-brain exploratory analysis of structural and functional brain changes have enabled disease evaluation independent of anatomical or pathological hypotheses, leading to the accumulation of medically impactful findings for disease pathophysiology assessment. However, neuroimaging data are inherently high-dimensional, creating a severe multiple comparisons problem in whole-brain exploratory statistical inference. Furthermore, complex spatial dependencies exist among MRI signals observed at the voxel level, necessitating the development of approaches different from classical multiple comparison corrections. This review provides an overview of the characteristic statistical challenges in neuroimaging analysis, the Random Field Theory that has been applied in the neuroimaging field for many years, and the evolution of current standard statistical methodologies.

講義形式における統計の入門科目の授業実践と その考察

岡本 雅子¹・瀧川 一学²・中西 寛子³

(受付 2025 年 6 月 24 日；改訂 8 月 15 日；採択 8 月 29 日)

要 旨

京都大学は、全学共通科目として統計関係の科目を数多く開講している。その中で、工学部工業化学科(現：理工化学科)2回生向けに「統計入門」が開講されている。本論文では、2023年度の「統計入門」を対象とし、学習者の学習項目ごとの理解度を把握するために、レポート課題および期末テストの結果、理解度アンケートの結果を分析した。その結果、レポート課題では受講者のつまづき箇所が明らかとなった。また、理解度アンケートから、授業前半の学習項目の理解度(学習者の自己評価)は比較的高かったが、後半の理解度は低かったことがわかった。さらに、期末テストでも同様の結果が得られた。加えて、期末テストの信頼性を表す指標として Moodle に実装されているアイテム分析機能を用いて分析したところ、改善が必要な設問があることがわかった。また、受講者の学習動向などから、講義形式で実施した統計入門の授業について課題や問題点などが明らかとなった。

キーワード：統計入門、統計教育、高等教育、初学者、講義形式。

1. はじめに

内閣府の「AI 戦略 2019」(内閣府, 2019)において、デジタル社会の「読み・書き・そろばん」である「数理・データサイエンス・AI」を理解して活躍できる人材を育成する必要があるとし、その人材育成のための目標・仕組みとして、内閣府・文部科学省・経済産業省の3府省が連携し、大学(短期大学含む)・高等専門学校等における数理・データサイエンス・AI教育の取り組みを奨励するための認定制度「数理・データサイエンス・AI教育プログラム認定制度」(文部科学省, 2021)が2021年に創設された。

このような状況の下、京都大学の全学共通科目「統計入門」は、「数理・データサイエンス・AI教育プログラム認定制度」において、リテラシーレベル(MDASH Literacy)の教育プログラムとして認定されている。この「統計入門」は、京都大学における教養・共通教育の自然科学科目群・データ科学群において開講しており、京都大学の教養・共通教育におけるデータサイエンス教育の中心となっている。同科目では、「研究や、生活、社会・経済活動に不可欠な統計を、集計・分析し、理解する力を養うこと」「統計分析手順の機械的な利用や解釈だけを講義す

¹ 京都大学 国際高等教育院附属データ科学イノベーション教育研究センター：〒606-8501 京都府京都市左京区吉田近衛町 69; okamoto.masako.8v@kyoto-u.ac.jp

² 東京大学 大学院新領域創成科学研究科：〒277-8561 千葉県柏市柏の葉 5-1-5; takigawa@k.u-tokyo.ac.jp

³ 統計数理研究所：〒190-8562 東京都立川市緑町 10-3; nmf.nakanishi@nifty.com

るのではなく、その基礎となる考え方を学ぶこと」「統計学的命題について、厳密な数学的証明は避け、あくまで統計・統計学のエンドユーザとして必要とされる直感的な理解を目指すこと」を目的としている。そのため、同科目は、統計の分析手順の機械的な利用や解釈だけを講義するのではなく、統計ならびに統計学に関する基本的な考え方を中心に講義し、その基礎となる考え方を学ぶことに重点を置いている。

本論文では、2023年度に工学部工業化学科(現：理工化学科)2回生向けの「統計入門」を対象とし、受講者の学習項目ごとの理解度を把握することによって、統計入門科目において注意すべきことを考える。そのため、はじめに、本授業の教授全容を説明する(第2章)。次いで、授業で実施されたレポート課題および期末テスト、理解度アンケート調査の分析結果を報告する(第3章)。最後に、受講者の成績や学習動向などから、講義形式で実施した統計入門科目について課題や問題点などについて考察する(第4章)。

2. 授業実践

2.1 対象とした授業

本論文の対象とした授業は、2023年度の前期に京都大学で実施された全学共通科目「統計入門」である。2クラスを対象とし、両クラスとも工学部工業化学科の2回生を対象としたクラス指定の科目となっている。なお、両クラスとも他学部の1回生から4回生および大学院生も受講可能となっている。

2023年度の受講者数はAクラス102名、Bクラス107名の合計209名であった(表1参照)。これらのクラスは、第一著者と第二著者の2名の教員が共同で担当し、受講者がつまづいた場合には、教員やティーチングアシスタント(TA)1名がサポートに入り、直接対応した。

本授業の教授内容である統計や統計学については、膨大な研究の蓄積が有り、その利用はきわめて多分野にわたる。加えて、各分野で独自の発展をとげている部分もあり、本科目のみでそのすべてを扱うことは出来ないため、本科目では、統計ならびに統計学に関する基本的な考え方を中心に講義することで、より発展的な統計・統計学の学習への礎となることを目指している。具体的には、二元分割表(2×2クロス集計表)の独立性の検定と関連性の強さの推定を主な題材として、統計データの収集、チェック、集計、分析、結果の解釈という一連の過程について解説し、統計的仮説検定と推定の考え方に関する内容について理解を深める構成となっている(表2参照)。

本授業では、教科書「講義実録 統計入門」(京都大学データ科学イノベーション教育研究センター, 2023)に概ね準拠したスライドを作成し、講義形式で実施した。授業で使用するスライドについてはPDFに変換し、学内LMS(PandA)経由で受講者に配付した。なお、演習については、授業時間外による個別学習(自習)とした。

また、演習で使用する統計解析ソフトウェアは「JMP(ジャンプ)」を推奨しており、受講者は自身のノートPCにJMPをインストールし、自習課題に取り組んだ。JMPのインストール手順書(Windows版とmacOS版)はPDF版の資料および動画を準備して、受講者に配付した。さらに、JMPを使用した演習用の自習課題についてもPDF版の資料(Windows版とmacOS

表 1. 対象授業の受講者数.

	Aクラス	Bクラス
合格者	87	93
不合格者	15	14
合計	102	107

表 2. 統計入門の授業内容.

	学習内容	レポート課題
第 1 回	概要と導入	1 回目
第 2 回	量的データの確認と要約	
第 3 回	質的データの確認と要約	
第 4 回	確率の基礎・JMP 等によるデータの確認と要約	
第 5 回	カイ二乗検定—適合度と独立性の検定—	
第 6 回	二元分割表とフィッシャーの正確検定	
第 7 回	さまざまな確率分布と統計的検定の考え方	
第 8 回	統計的推定	
第 9 回	二元分割表のリスク比とオッズ比	2 回目
第 10 回	二元分割表における母比率の差の検定・推定	
第 11 回	検定・推定と標本規模, t 検定	
第 12 回	平均の差の検定	
第 13 回	相関と単回帰分析	
第 14 回	重回帰分析・ロジスティック回帰分析	
第 15 回	期末試験	
第 16 回	フィードバック	

版)を作成し、受講者に提供した。

最後に、2023 年度の授業は、原則として対面授業としながら、毎回の授業は、ビデオ会議サービス (Zoom) の録画機能を使って収録し、不要箇所を除くなど編集した上で、動画配信サービス (Microsoft Stream) を用いて復習用教材としてアーカイブ動画を受講者に提供した。

2.2 成績評価

本授業では、受講者の理解度を確認するため、毎回の確認テスト、2 回のレポート課題、最後に期末テストを実施し、総合的に評価した。

なお、今回対象とした 2 クラスは授業内容および成績評価を統一するため、授業スライド、確認テスト、レポート課題、期末テスト等はすべて同じものを使用するなど、2 クラスの授業について同等の質を担保できるように配慮して授業を実施した。

2 クラスの合格者数(表 1)から A クラスと B クラスでほとんど差は見られなかったため、2 クラスをまとめて分析することとする。また、不合格者については、要件(確認テスト、2 回のレポート課題、期末試験)を総合的に評価した結果、合格基準に達していなかったため、不合格となっている。

3. 実践結果

本論文では、受講者の理解度を確認するために実施した 2 回目のレポート課題および期末テスト、第 14 回の授業で実施した理解度アンケート調査を評価の対象とした。なお、理解度アンケート調査については 181 名から回答があった。

理解度アンケート調査は、「統計入門」の学習項目に係る理解度(受講者の自己評価)を受講者に尋ねるものであって、期末テスト前の学習項目の復習を兼ねて、ほぼ理解できているものについてチェックを入れて回答させた。加えて、授業に対する受講者の率直な意見を集めるために、「つまづいた際の対処」に関するアンケートを自由記述形式で尋ねた。また、「統計入門の授業に関する感想」を尋ねたアンケートも行った。これらを理解度アンケートと区別するため、授業アンケートと呼ぶ。具体的な文面は後述する。

3.1 レポート課題

本授業は、前述したように、二元分割表の独立性の検定と関連性の強さの推定を主な題材として、統計的仮説検定と推定の考え方について理解を深める構成となっており、受講者の理解度を確認するため、2回のレポート課題を課している。1回目のレポート課題は、データのチェックや集計、可視化に係る内容で、2回目のレポート課題は、統計的仮説検定と推定に係る内容となっている。本論文では、受講者から提出された2回目のレポート課題について、つまりき箇所とその件数、つまりき内容について調べた。分析対象とした2回目のレポート課題の内容は、授業の初回に実施した「たこ焼き器」に関するアンケートの実際の回答結果(匿名化済み)に対して、以下の課題を実施するものである。

- (1) 住居が大阪府内であればたこ焼き器の所有率が高いことを検定したい。帰無仮説と対立仮説の設定および有意水準 α の設定
- (2) クロス集計表の作成
- (3) 帰無仮説の下で理論度数表の作成
- (4) カイ二乗値の計算
- (5) 標本サイズが十分に大きい場合、(4)で求めたカイ二乗値が従う確率分布
- (6) p 値の計算および有意水準 α との比較
- (7) フィッシャーの正確検定との結果の比較
- (8) オッズ比の区間推定および信頼係数 95% の上限・下限の計算

3.1.1 レポート課題の結果

2回目のレポートは、181名から提出があった。各項目の正誤は表3の通りである。8項目のうち、項目(1), (2), (3), (4), (6), (7)は概ね正しく記述できていた。残りの(5)と(8)はつまりき受講者が多かった。これらについて述べる。

「(5)標本サイズが十分に大きい場合、(4)で求めたカイ二乗値が従う確率分布」の正答は「自由度1のカイ二乗分布」である。計算の必要でない基本的な問題であるが、正答率が64.0%と低かった。誤答を確認したところ、「正規分布」や「標準正規分布」、「自由度3のカイ二乗分布」などが散見された。「正規分布」という解答に関しては、「標本サイズが十分に大きい場合」という文言から「正規分布」を想起してしまったと推察される。授業では、このようなつまりきを考慮してより丁寧な説明が必要であることがわかった。

また、「(8)オッズ比の区間推定および信頼係数95%の上限・下限の計算」では、正答率が56.9%であった。JMP等の統計ソフトを使用すれば容易に計算することができる。しかしながら、レポートの中に計算式を記述している受講者がいたので、それを確認すると、使用する公式が誤っていたり、(代入する数値が異なるなど)計算式が誤っていたりする答案が多くみら

表 3. 2 回目のレポート課題の評価.

	正	誤
(1)	157(86.7%)	21(13.3%)
(2)	175(96.7%)	6(3.3%)
(3)	167(92.3%)	14(7.7%)
(4)	157(86.7%)	24(13.3%)
(5)	116(64.0%)	65(35.9%)
(6)	161(89.0%)	20(11.0%)
(7)	156(86.2%)	25(13.8%)
(8)	103(56.9%)	78(43.1%)

れた。このことから、受講者が区間推定の計算方法を正確に理解していないと思料される。

区間推定については、授業アンケートの自由記述においても、「区間推定の部分でどう役にたつものかあまりイメージがわからず躓いてしまった」といったネガティブな記述がみられ、一部の受講者にとっては推定に関する学習自身に不安を抱えていることがわかった。

レポート課題について授業中あるいは授業外で受講者からの質問が少なかったため、レポート課題でつまずいた際にどのように解決していたのかを知るためにアンケート調査を実施した。具体的には、第14回の授業で行った授業アンケートにおいて、「レポートでつまずいてしまったとき、どのようにされましたか。来年度の参考にさせていただきたいと思いますので、みなさんの経験を教えてください(自由記述)」と問うたところ、81名から回答があった。その内訳は、「つまずかなかった」が1名いたが、「ネットで検索」が32名、「授業資料をみながら一人で解決」が22名おり、合計54名がつまずき箇所を受講者自身で解決していた。他方、「友人に相談」が22名、「教員に相談」が4名おり、合計26名が友人や教員に相談するなどして問題を解決していた。

3.2 期末テスト

2023年度に工学部工業化学科(現：理工化学科)2回生向けに開講された「統計入門」の期末テストは、オンライン化してWeb試験として実施した。期末テストは、問題は全部で43題とし、85点満点のものを100点換算した。受験時間は60分、单元ごとにランダムで表示(出題)されるように設定した。期末テストの内容は、設問1-37は基本的な問題、設問38-43は発展的な問題で構成されている。なお、設問1-10は二択問題、設問11-22は択一式問題、23-30は多肢選択問題、31-37は計算結果や用語等の穴埋め問題、38-43は多肢選択問題となっている。

実際の受験状況は、教員およびTAの監督下の教室で、学生が所持しているノートPCやタブレット等からPandA経由で専用サイト(Moodle)にアクセスし、期末テストを受験している。

本Moodleには、テスト結果を分析する機能があるので、それをを用いて理解度について詳しい分析を次の節で示す。

3.3 Moodleのアイテム分析機能

Moodleにはアイテム分析の機能が備わっており、出題された個々の設問に対してパフォーマンスの分析や判定が可能となっている(姫野 他, 2015)。今回は、(1)ファシリティ指標、(2)識別指数を利用することにする。

3.3.1 ファシリティ指標

「ファシリティ指標」は、設問の最高点(受験者の最高点)に対する平均点が比率で表示され、以下のように解釈される(Moodle, 2022)。このファシリティ指数が35%から65%の範囲のとき、「平均的な学生に対する適切な設問である」と解釈される。

- 5%以下：Extremely difficult or something wrong with the question
- 6%-10%：Very difficult
- 11%-20%：Difficult
- 21%-34%：Moderately difficult
- 35%-65%：About right for the average student
- 66%-80%：Fairly easy
- 81%-89%：Easy
- 90%-94%：Very easy
- 95%-100%：Extremely easy

表 4. 期末テストの結果.

	2023 年度
受験者数	192
平均	54.71
標準偏差	10.44
最小値	16.67
最大値	88.51
中央値	54.60

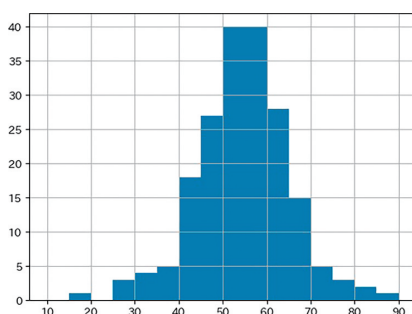


図 1. 期末テストの得点分布.

3.3.2 識別指数

「識別指数」は、設問の得点とテスト全体の得点の間の相関関係である。これは、期末テスト全体において「優秀な受講者」と「優秀でない受講者」の各設問における大まかなパフォーマンスを示す指標である (Johnson, 1951)。同指標は、受講者を上位 1/3 と下位 1/3 に分割し、各設問の上位の受講者の獲得得点の合計 (A) より下位の受講者の獲得得点 (B) の差を全体の数で割る $((A-B)/(A+B))$ ことにより計算され、値は +1 から -1 の値を取る。つまり、優秀な受講者ほど正答しやすく、そうでない受講者ほど誤答しやすい傾向が顕著な場合に、識別指数が高くなる。また、識別指数は、20% 以上であると設問は識別力が高く、40% 以上であると優れた設問であると判断される (Ebel, 1972)。また、マイナスの値は、当該設問では、「優秀な受講者」が「優秀でない受講者」より得点が取れなかったことを示し、良問ではないことになる。

3.3.3 期末テストの結果

期末テストの結果について表 4 および得点の分布を図 1 に示す。なお、ヒストグラムの階級の最小値は 10 点、階級幅は 5 点である。期末テストの受験者数は、A クラスと B クラスを合わせて 192 名であり、平均 54.71、標準偏差 10.44 で、最高値が 88.51、最低値が 16.67、中央値が 54.60 であった。

2023 年度の「統計入門」の期末テストについて、図 1 のヒストグラムの分布の形状を見ると、ほぼ左右対称で単峰性となっており、受講者の成績は平均的な点数を中心に分布している。

3.3.4 Moodle のアイテム分析の結果について

期末テストの設問番号と設問内容について表 5 に示す。

図 2 に設問番号に対するファシリティ指標の数値を示す。ファシリティ指標が 35% を下回る設問、つまり、比較的難しいとされる設問が 8 問あった。それらは、基本的な問題から構成した設問 16, 20, 37 と、発展的な問題から構成した設問 38, 39, 40, 41, 43 である。これら 8 問については本項で細かな考察を行うので、以下に問題内容を示す。

表 5. 期末テストの内容.

設問番号	設問内容	設問番号	設問内容
1	要約統計量	26	分散の計算
2	標本調査	27	正規分布
3	第二種の過誤	28	相関係数
4	区間推定	29	相関係数
5	信頼区間	30	標準化
6	独立性の検定	31	統計学的仮説検定
7	統計的仮説検定	32	片側検定と両側検定
8	統計的仮説検定	33	対数オッズ比
9	t 検定	34	第一種の過誤と第二種の過誤
10	量的変数	35	無相関検定
11	要約統計量	36	統計学的仮説検定
12	要約統計量	37	確率分布
13	要約統計量	38	分散分析
14	確率	39	分散分析
15	確率	40	分散分析
16	確率の計算	41	回帰分析
17	箱ひげ図	42	回帰分析
18	ヒストグラム	43	回帰分析
19	偏差値		
20	データの散らばり具合		
21	標本調査		
22	クロス集計表		
23	記述統計と推測統計		
24	データの種類		
25	量的変数		

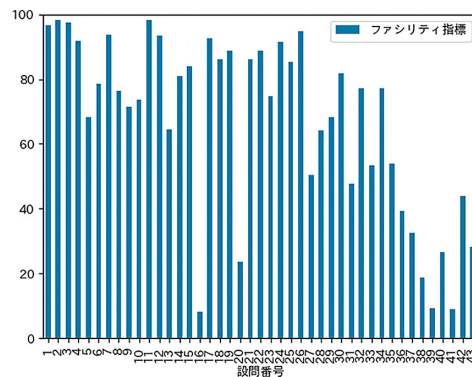


図 2. 設問ごとのファシリティ指標.

- 設問 16(択一式問題) : 「袋の中に赤色のボールが 4 個, 白色のボールが 6 個入っている. 袋からボールを 1 個取り出し, それが赤色のボールならそのボールを袋に戻さず, 白色のボールなら袋にボールを戻す. その後, 再度ボールを 1 個取り出す. 2 回ボールを取り出した結果, 赤色と白色が 1 回ずつとなる確率にもっとも近いのはどれか」という確率の計算問題

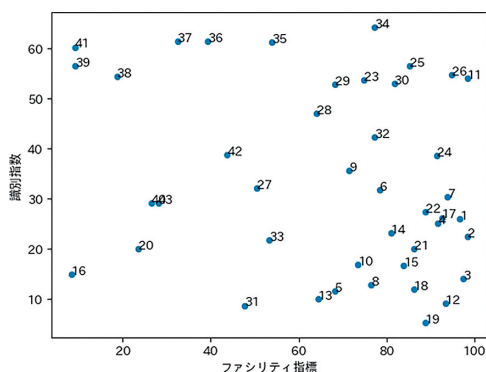


図 3. ファシリティ指標と識別指数の散布図.

- 設問 20(択一式問題)：2 クラスの試験結果について、各クラスの平均値と標準偏差と 2 クラスをまとめたヒストグラムが提示され、それらから情報を読み取る問題
- 設問 37(多肢選択問題)：確率分布の期待値や分散等を計算する問題
- 設問 38(穴埋め問題)：3 群の標本の数値を提示し、「標本平均」、「変動(偏差平方和)」、「不偏分散」をそれぞれ計算する問題
- 設問 39(穴埋め問題)：設問 38 の計算結果を用いて、3 群の一元配置分散分析を行うための分散分析表を完成させる問題
- 設問 40(択一式問題)：設問 39 で計算した分散比を用いて、提示した帰無仮説が棄却できるかどうかを答える問題
- 設問 41(択一式問題)：重回帰分析の出力結果から予測値を計算する問題
- 設問 43(多肢選択問題)：重回帰分析の出力結果から言えることを選択する問題

これらの設問のうち、設問 16, 20, 37 は、基本的ではあるが、授業内容を深く理解していなければ誤答をしやすい傾向のある設問である。このほか、設問 38-40 は一元配置分散分析の計算方法等を、設問 41 および 43 は回帰分析の結果の見方を理解している必要があり、発展的な設問である。

次に、ファシリティ指標と識別指数を散布図として図 3 に示す。ここでは、ファシリティ指標と識別指数からそれぞれの設問の適切さを評価する。2 つの値の相関は -0.22051 となっており、ほぼ無相関であった。散布図の中に点が散在しており、ファシリティ指標と識別指数に直線的な関係はないが、ファシリティ指標が 65% より大きく、識別指数が 30% 以下の範囲で密度が高い。

識別指数は前述したように、20% 以上を識別力の高い設問、40% 以上を優れた設問である基準値とした。この結果、図 3 より識別指数が 40% 以上の優れた設問が 15 問、識別指数が 20% 以上で 40% より小さい値の識別力の高い設問は 16 問、識別指数が 20% 以下を示した設問は 12 問であった。

さらに、ファシリティ指標と識別指数の数値を併用して見ていく。ファシリティ指標が 65% より大きく、識別指数が 40% 以上の設問は 8 問であった。次に、ファシリティ指標が 65% より大きく、識別指数が 20% 以上で 40% より小さい値の識別力の高い設問は 10 問であった。これら 18 問は、比較的易しい問題構成であるにもかかわらず、識別力の高い設問と言える。

ファシリティ指標が 65% より大きい値で、識別指数が 20% より小さい値を示した設問は 9 問であった。これらは、比較的易しい問題構成であるため、識別力は低い設問と言える。これ

らの設問の適正について確認する必要があるかもしれない。

先に示したファシリティ指標が35%を下回る比較的難しい8問について識別指数を含めて場合分けする。ファシリティ指標が35%より小さい値で、識別指数が40%以上の優れた設問は37, 38, 39, 41, また、識別指数が20%以上で40%より小さい値である識別力の高い設問は20, 40, 43であった。先に示したように、設問20は統計指標やヒストグラムから情報を読み解く問題で、それ以外の6問は確率分布、分散分析、回帰分析に関する設問である。これらは、比較的難しい問題構成であり、識別力は高い設問である。これらの設問では授業全体の理解が十分である受講生とそうでない受講生が識別されたと思われる。

ファシリティ指標が35%より小さい値で、識別指数が20%より小さい値を示した設問が1問あり、設問16の確率(復元抽出)であった。この設問は難しく、識別力が低いという結果となった。全体に理解が不十分であるため、今後の授業内容および問題構成について考えていく必要がある。

3.4 学習項目の理解度について

第14回の授業で理解度アンケート調査を実施し、「統計入門」の学習項目に係る理解度(受講者の自己評価)について受講者に尋ねた。理解度については、期末テスト前の学習項目の復習を兼ねて、ほぼ理解できているものについてチェックを入れて回答させた。なお、本項目については、181名から回答があった。

授業前半の学習項目の理解度について、図4に示す。図4の学習項目は上から下に講義した順番となっている。授業の前半では、「代表値」「四分位数」「箱ひげ図」などの理解度が高かった。一方で、「期待値の線形性」「確率分布の和」「ベルヌーイ分布と二項分布」「調査課題の設定、調査対象の設定集計・解析方法」「単純集計」などの理解度が低かった。

次に、授業後半の学習項目の理解度について、図5に示す。図5の学習項目も図4と同様に、「推測統計と検定」から順に講義した順番となっている。授業の後半では、「独立性の検定」「適合度検定」「リスク差、リスク比、オッズ、オッズ比、対数オッズ比」「第一種、第二種過誤と標本サイズ」「前向き研究、後ろ向き研究」などの理解度が高かったが、「ロジスティック回帰分析」「決定係数」「重回帰分析」「分散分析」「多重検定」「中心極限定理」等の理解度が低かった。

3.5 授業アンケートの結果について

授業アンケートにおいて、「統計入門の授業の感想(おもしろいと感じた点や難しいと感じた点など)を教えてください」と受講者に問うたところ、本項目については178名から回答があった。

「JMPによる実習(自習)が面白かった」「ExcelやJMPを用いて様々なことができるようになったことが嬉しいです」など、JMPの演習(自習)に対するポジティブなコメントが7件、「授業で習った内容をレポート課題で実践するのが楽しかったです」「レポート課題で実際に知識を使えるのが楽しかったし理解も進んだので良かったと思う」など、レポート課題に関するポジティブなコメントが10件あった。一方で、「JMPを使ったことがなかったので、JMPによる分析を行うのに苦労しました」「JMPの使い方を理解するのに時間がかかりました」など、JMPの使用に係るネガティブなコメントが11件、「レポートが難しかったです」「レポートは難しいと感じた」「課題をひとりでやるにはすこしむずかしかった」など、レポート課題に係るネガティブなコメントが9件見られた。

加えて、「高校時代にそこまで深く学ばなくて暗記していた部分が統計範囲は多かったが、それが理解に変わるのが楽しかった」「身近なところでもよく統計を使った調査結果を目にするのが統計にも多様な方法や考え方があり面白かった」「日常で何気なくみているグラフや表を

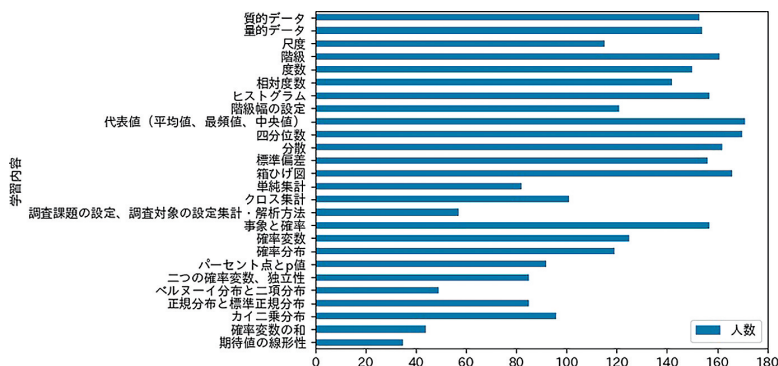


図 4. 授業前半の学習項目の理解度。

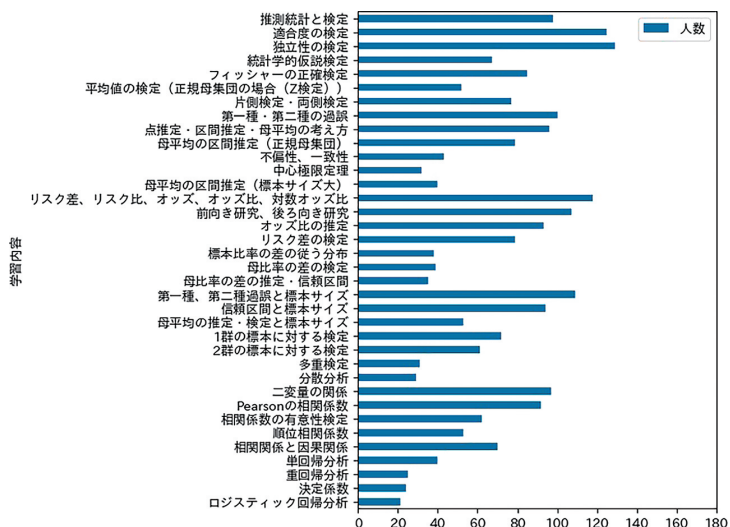


図 5. 授業後半の学習項目の理解度。

批判的に見えるようになるためにはこの統計の視点が必要だと感じた(原文のまま)」など、授業内容に係るポジティブなコメントが 55 件見られた。一方で、「自分がうけた全共(全学共通科目)の中で一番難しかったです」「途中からほとんど理解できなかった」「序盤で簡単だと思っていたら、2 回目のレポートの範囲から急に難しくなった」「授業の後半になるにつれて難しくなった」など、授業内容の難易度に係るネガティブなコメントが 39 件見られた。

4. まとめと考察

ここまで見てきた本授業で課した「レポート課題」と「期末テスト」の傾向をまとめ、そこから考察されることを記す。

4.1 レポート課題

2 回目のレポート課題は、「たこ焼き器」に関するアンケート調査の結果を分析したものであった。全 8 項目中、2 項目の正解率が明らかに低かった。それらは、「(5) 標本サイズが十分

に大きい場合、(4)で求めたカイ二乗値が従う確率分布」と「(8)オッズ比の区間推定および信頼係数95%の上限・下限の計算」であった。3.1.1項で示したように、どちらも正確な理解の欠如と思われ、今後の授業における明確な課題となった。

レポート課題は、受講者が自身で解答してもよいし、他の方法で疑問点を解決してもよい。授業中、授業外の質問件数が少ないので、受講者がレポート課題等でつまづいた際の解決方法をアンケート調査の結果から明らかにしようと試みた。その結果、つまづき箇所を「授業資料やネット検索などを用い受講者自身が解決した層(54名)」と「友人や教員に相談した層(26名)」に類型化されることが明らかとなった。しかしながら、教員に質問したのはわずか4名であったことから、授業中に質問しなかった(できなかった)受講者がいた可能性も同時に浮かび上がった。こうしたことから、講義を中心とした授業の問題点として「質問のしにくさ」をどのように解決していくのかについて、さらなる検討が必要であることが明らかとなった。

4.2 期末テスト

4.2.1 Moodleのアイテム分析結果について

図3に示したように、ファシリティ指標と識別指数の相関はほぼ無相関であった。これは、ファシリティ指標と識別指数は設問に対する評価の視点が異なる独立した評価値が理由であると考えられる。つまり、ファシリティ指標が解答者全員の正誤から評価しているのに対し、識別指数は成績上位者と成績下位者から識別能力を評価しているためである。こうしたことから、設問の難易度の評価については、ファシリティ指標と識別指数を併用して検討した。

期末テストにおいて、設問の難易度を示す指標「ファシリティ指標」が35%を下回る難しいと思われる設問が以下の8問であった(詳しくは3.3.4項)。

- 設問 16：確率(復元抽出)を計算して正しい選択肢を選ぶ問題
- 設問 20：平均値と標準偏差、ヒストグラムの情報から読み取れた内容を選択する問題
- 設問 37：期待値や分散等を選択肢から選ぶ問題
- 設問 38：「標本平均」、「変動(偏差平方和)」、「不偏分散」を計算する問題
- 設問 39：3群の一元配置分散分析を行うための分散分析表を完成させる問題
- 設問 40：計算した分散比を使って提示された帰無仮説が棄却できるかどうかを答える問題
- 設問 41：重回帰分析の出力結果から予測値を計算する問題
- 設問 43：重回帰分析の出力結果を読み解く問題

設問 16, 20, 37 は基本的であるが、授業内容を深く理解していなければならない設問である。設問 38-40 は一元配置分散分析、回帰分析の結果の見方を理解している必要がある発展的な設問である。これらの設問の傾向から見ると、授業後半で学習する問題から構成された設問が多く含まれているので、後半の授業方法の見直しが必要であると思われる。

上記の8問中、設問 16 以外は高い識別率を示しており、受講者の理解度を識別できている良問であると考えられる。しかしながら、設問 16 の識別指数は低く(14.96)、学習内容と期末テストで出題される設問に齟齬がある可能性がある。受講者の正当な評価がなされていないことを示しているため、今後、設問の改善を行う必要があると示唆される。

一方、ファシリティ指標が65%を上回り、識別指数が20%より小さい値を示した設問は9問あった。これらの設問は、学習内容に沿った基本的な事柄を問う設問となっており、ファシリティ指標が高い数値を示している。基本的な内容を問う設問については、受講者が学習内容を理解した上で正答したのか判別が難しいと思われる。

4.2.2 学習項目の理解度について

期末テストの中で発展的な設問であった「分散分析に係る問題」や「回帰分析に係る問題」については、「ほぼ理解できた」と回答した人数が少なく、受講者の自己評価が著しく低くなっていた。一方で、ファシリティ指標と識別指数を考慮した場合、受講者の理解度を識別できている良問であると考えられる。そのため、多くの受講者にとっては、理解できていない難しい学習内容であったことが推察される。

さらに、受講者の反応について見ていくと、自由記述では「高校時代にそこまで深く学ばなくて暗記していた部分が統計範囲は多かったが、それが理解に変わるのが楽しかった」など、授業内容の理解に係るポジティブなコメントが55件あり、多くの受講者から「統計入門」の授業内容についてはポジティブな評価を得られた。一方で、「途中からほとんど理解できなかった」「授業の後半になるにつれて難しくなった」など、授業内容の難易度に係るネガティブなコメントが39件あり、一部の受講者にとっては、授業内容や難易度が適切ではなかったと推察される。こうしたことから、授業内容でつまずいた際に、受講者がより指導を受けやすい環境の整備が必要であるということが明らかとなった。

また、JMPの演習(自習)に対するポジティブなコメントが7件、レポート課題に関するポジティブなコメントも10件あり、自習用の演習やレポート課題についてポジティブな評価を得られた一方で、JMPの使用に係るネガティブなコメントが11件、レポート課題に係るネガティブなコメントが9件あり、一部の受講者にとっては、JMPの使用やレポート課題に不安を抱えていたことがわかった。特に、JMPの演習については授業時間外による個別学習としていたため、JMPの操作方法につまずいた際、個々の疑問点を自力で解消できない場合があったと推察される。

以上の点から、本授業では、授業の後半部分において受講者の理解に対する配慮が十分ではなかったと考えられ、本授業で用いたカリキュラムはこうした点において、改善の余地があると思われる。

とりわけ、授業の前半では、学んだ統計手法を活用してレポート課題に取り組むなど、能動的な学習部分が多かったため、受講者が達成感を感じやすかったのではないかと思料される。一方、授業の後半では、学んだ統計手法を模倣するという比較的、受動的な学習内容であったため、一元配置分散分析の計算方法や回帰分析の結果の見方など、受動的な学習部分において受講者の理解が深まらなかったものと推察される。さらに、授業の後半では、扱った学習内容を活用して、受講者が演習を通して学習内容を理解するまでのプロセスの中で、作業を遂行するために必要となるスキルや知識に対する支援が必要となってくるが、2023年度に用いた講義スライドはこうした側面についての配慮が十分でなかったと考えられる。

5. おわりに

本論文では、2023年度工学部工業化学科(現：理工化学科)2回生向けに開講した「統計入門」について、その実践例を報告するとともに、学習者の学習項目ごとの理解度を把握するために実施したレポート課題、期末テストの結果および理解度アンケートの結果を分析した。その結果、レポート課題では受講者のつまずき箇所が明らかとなった。また、理解度アンケートの結果から、授業前半の学習項目の理解度(学習者の自己評価)は比較的高かったが、後半の理解度は低く、期末テストでも同様の結果が得られた。加えて、期末テストの信頼性を表す指標としてMoodleに実装されているアイテム分析機能を用いて分析したところ、改善が必要な設問があることが明らかとなった。

さらに、授業アンケート結果から、講義を中心とした授業では、受講者がレポート課題等で

つまりいた際の解決方法として、つまりき箇所を「受講者自身で解決した層」と「友人や教員に相談した層」に類型化されることが明らかとなった。

今後は、今回得られた知見をもとに、受講者の理解が深まるように講義スライドを改善するとともに、とりわけ授業の後半で受講者がひとりで演習(自習)できるような教材を開発したい。加えて、講義を中心とした授業の問題点として「質問のしにくさ」があったのではないかと推察されたことから、レポート課題や授業内容に係る質問対応について検討していきたい。また、期末テストで受講者の正当な評価がなされていない設問については、設問の内容や選択肢についての妥当性について検討していきたいと考えている。

謝 辞

本論文の期末テストのための環境をご提供いただいた京都大学国際高等教育院附属データ科学イノベーション教育研究センター FLANAGAN Brendan John 特定准教授に深く感謝申し上げます。査読者の皆様には、貴重なご意見とご指導を賜り、論文をより良いものにすることができました。査読者の皆様に心より御礼申し上げます。

また、本研究は文部科学省事業「統計エキスパート人材育成プロジェクト」の助成を受けたものである。

付 記

本論文は、岡本 他 (2024a, 2024b) で発表した研究を発展させて、その成果をまとめたものである。

参 考 文 献

- Ebel, R. L. (1972). *Essentials of Educational Measurement*, Prentice-Hall, Englewood Cliffs.
- 姫野聡也, 上田浩, 喜多一, 森幹彦 (2015). 学認連携 Moodle における受講者動向分析に向けた小テスト成績と設問に関する一考察, 情報処理学会研究報告, **2015-CE-132**(35), 1-6.
- Johnson, A. P. (1951). Notes on a suggested index of item validity: The U-L index, *Journal of Educational Psychology*, **42**, 499-504.
- 京都大学データ科学イノベーション教育研究センター (2023). 『講義実録 統計入門』, 現代図書, 相模原.
- 文部科学省 (2021). 数理・データサイエンス・AI 教育プログラム認定制度, https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00001.htm (最終アクセス日 2025 年 5 月 23 日).
- Moodle (2022). Quiz Report statistics, https://docs.moodle.org/dev/Quiz_report_statistics (最終アクセス日 2025 年 6 月 19 日).
- 内閣府 (2019). AI 戦略 2019, <https://www8.cao.go.jp/cstp/ai/index.html> (最終アクセス日 2025 年 5 月 23 日).
- 岡本雅子, 瀧川一学, 中西寛子 (2024a). 高等教育における統計初学者のつまりきの分析, 日本教育工学会 2024 秋季全国大会講演論文集, 199-200.
- 岡本雅子, 瀧川一学, 中西寛子 (2024b). 統計の入門科目の授業実践とその考察, 情報処理学会研究報告, **2025-CE-178**(3), 1-6.

A Practice and Consideration of Introductory Statistics in a Lecture Style

Masako Okamoto¹, Ichigaku Takigawa² and Hiroko Nakanishi³

¹Center for Innovative Research and Education in Data Science, Institute for Liberal Arts and Sciences,
Kyoto University

²Graduate School of Frontier Sciences, The University of Tokyo

³The Institute of Statistical Mathematics

Kyoto University offers a number of statistics-related courses as part of the Liberal Arts and Sciences. “Introduction to Statistics” is offered to second-year students in the Faculty of Engineering. This paper analyzes the results of report assignments, the results of final examinations, and the results of comprehension questionnaires for the class of 2023. According to these results, the report assignments clearly revealed items where the student faltered. Furthermore, the results of the comprehension questionnaires showed that the student’s understanding of the learning items in the first half of the class (self-evaluation by the student) was high, but their understanding in the second half of the class was low. Similar results were also found in the final examination. An analysis function implemented in Moodle revealed that some questions on the final examinations could have been improved. Additionally, this paper clarified issues and challenges associated with conducting the introductory statistics course in a lecture style.

学校外教育の累積的効果：利用パターンと タイミングが教育達成に与える影響

眞田 英毅[†]

(受付 2025 年 6 月 27 日；改訂 12 月 9 日；採択 12 月 9 日)

要 旨

本研究は、出身階層が中学・高校段階での学校外教育利用を通じて教育達成に与える効果を、g-computation(反実仮想シミュレーション)の枠組みで検討した。教育達成は、4水準の順序変数として定義し、銘柄大卒の予測確率に着目した。「働き方とライフスタイルの変化に関する全国調査」を用い、出身階層を主成分分析による変数、教育達成を4水準の順序変数として分析した結果、当初の仮説であった「累積的利用」は、必ずしも教育達成を最大化しないことが示された。本研究の主たる発見は第一に、出身階層を問わず、大学受験に直結する「高校段階のみ」の利用と「累積的利用」は同程度の高い効果を示し、進学戦略として高校段階の利用が最も有効であること、第二に、「中学のみ」の利用は進学確率の増加に貢献せず、学力的な困難を抱える生徒が利用を選択している可能性(セレクション効果)が示唆されたことである。そして最も重要な知見として、たとえ同じ学校外教育戦略をとっても、高階層の生徒の進学確率は低階層の約2倍に達し、学校外教育の利用パターンでは説明しきれない階層効果がなお一定程度残存していることが示唆された。

キーワード：教育格差, 主成分分析, 順序ロジットモデル。

1. 研究の意義と背景

近年、学校外教育(塾、家庭教師、通信教育など)¹⁾は、世界各国で広く利用されている(Mori and Baker, 2010; Entrich, 2020)。とりわけ日本では、進学競争が激化する中、家庭が学校外での学習支援に多大な資源を投入する傾向が強まっており、子どもの学力や進学先に与える影響が注目されてきた(Entrich, 2018)。こうした学校外教育は、単なる学力形成の手段にとどまらず、家庭の経済資源や情報資源に基づく教育格差を拡大・再生産する手段として機能している。とくに東アジア諸国では、塾や家庭教師といった受験準備を目的とする補習型学校外教育が、高階層の家庭によって積極的に活用されており、教育達成の格差拡大に寄与していることが指摘されている(Bray and Kwok, 2003; Stevenson and Baker, 1992; Entrich, 2018, 2020; 北村・竹ノ下, 2024)。

これまでの研究は、こうした学校外教育の影響を教育段階ごとに個別に捉える傾向が強く、中学段階での効果(中澤, 2013; 眞田, 2018)や高校段階での効果(眞田, 2022)がそれぞれ論じられてきた。しかし、教育達成の格差形成をより正確に理解するためには、出身階層が学校外教育を通じて教育達成に与えるプロセス全体を、断片的ではなく累積的に捉える視点が不可欠で

[†] 武庫川女子大学 社会情報学部：〒663-8558 兵庫県西宮市池開町 6-46; sanada_teruki_x@mukogawa-u.ac.jp

ある。

なかでも、Fujihara (2024)が示した累積的優位性の議論は、この点に重要な示唆を与える。すなわち、高階層の子どもほど中学・高校を通じて継続的かつ長期的に学校外教育を活用する傾向が強く、その結果、教育達成の格差が教育プロセスの各段階で蓄積されながら拡大していく可能性がある。

とくに日本では、高校段階におけるトラッキング効果(すなわち、高校の種類やレベルによって、その後の進路が制度的に水路づけられる構造)により、教育格差が維持・拡大していることがわかっている(中西, 2018)。そして、それに向けた準備として学校外教育が機能している(眞田, 2018)。さらに高校には大学進学を強く志向する学校とそうでない学校があり、進学率やカリキュラムに大きな違いがある。こうした差異は、高校入学以前の教育投資や学力形成の累積的な影響によって方向づけられている場合が多い(耳塚, 1993; 竹内, 1995)。形式的には早期選抜を行わない日本の制度においても、高校段階での制度的差異には、家庭背景やそれまでの教育投資が色濃く反映されやすい構造があるといえる。

しかし、これまでの学校外教育研究の多くは、中学段階あるいは高校段階といった単発的な効果に注目しており、教育過程を通じて累積的に作用する影響を捉えた分析は十分に行われてこなかった。そこで本研究は、出身階層の違いが中学・高校期における学校外教育の累積的な利用を通じて、最終的な教育達成にいかなる影響を及ぼすのかを、反実仮想的枠組みにより検討することを目的とする。ここでいう累積的效果とは、各教育段階における教育投資がそれぞれ独立に作用するのではなく、先行する段階の投資が後続段階の教育戦略や選抜条件に相互に作用しながら蓄積・連鎖し、最終的に教育達成に結びつく過程をさす(DiPrete and Eirich, 2006)。

本研究における教育達成は、最終学歴に基づいて区分される。とりわけ最上位のカテゴリーとして「銘柄大学」への進学を区別する。銘柄大学とは、荒牧(2000)、Fujihara and Ishida(2016)、中村(2018)に基づき、東京大学や京都大学などの国公立大学に加え、早稲田大学、慶應義塾大学、上智大学、同志社大学など一部の私立大学を含む大学群を指す。こうした銘柄大学への進学は、日本社会における教育格差研究において特に象徴的な意味を持つ到達点である。

このような累積的效果の観点は、社会学で研究が進められてきた、有利・不利の累積(Merton, 1968; DiPrete and Eirich, 2006)とも整合的である。また、高校進学という制度的分岐以前から、家庭の文化資本や情報資源に基づく教育戦略が展開され、非公式な選抜圧力が段階的に積み重ねられていくという点で、社会階層論の観点とも親和性が高い。

本研究の意義は、以下の3点である。第1に、学校外教育の影響を「教育段階での断片的な効果」ではなく、「教育プロセスにおける累積的な効果」としてとらえることにより、断片的だった従来の知見を整理・補完し、教育格差の形成メカニズムをより明確に捉える視点を提供する。第2に、出身階層から教育達成に至る経路における複数の媒介過程を因果的に可視化することで、教育格差の是正に向けた政策的介入の手がかりを提示する。第3に、日本の教育制度において家庭の教育投資と進学機会が累積的に連鎖する構造を実証的に明らかにすることで、教育機会の不平等をめぐる制度的理解に貢献する。

2. 先行研究と仮説

2.1 学校外教育と教育格差に関する実証研究

学校外教育は、世界中に広がっており(Mori and Baker, 2010)、とりわけ東アジアでは塾を中心に早期から普及してきた(Cummings, 1997; Hannum et al., 2019)。近年では英語教育を含む家庭教師の需要など多様化も進んでいる(Yung and Zeng, 2022)。

学校外教育は一般に家庭の経済資源・文化資本と密接に関連しており、高階層出身の子どもがより利用しやすく、教育格差を拡大する手段となっているという指摘がある (Bray and Kwok, 2003; Buchmann et al., 2010; Takashiro, 2021; 北村・竹ノ下, 2024)。Stevenson and Baker (1992)は、日本において模試・塾・家庭教師・通信教育・浪人などが大学進学に与える影響を分析し、学校外教育が制度的な選抜において既に有利な子どもにさらなる優位性を与えていると述べている。また、中澤 (2013)は、塾の進学効果が学力や性別によって異なることを示し、また Kuan (2011)は、台湾における通塾効果が学力や親の教育水準によって異なることを実証している。これらの研究は、学校外教育の効果が一律ではなく、特定の属性を持つ子どもにとってより効果的であることを示唆している。

一方で、学校外教育の機能を必ずしも格差拡大とみなすのではなく、階層的に不利な子どもにとっての挽回手段として捉える研究もある。例えば、眞田 (2018)は、中学段階での家庭教師・通信教育の利用が低階層の子どもの教育達成を促進する可能性があることを指摘している。また、眞田 (2022)は高校段階で不利なトラックの子どもが学校外教育利用で高等教育進学においてトラックの不利を挽回する可能性を指摘している。これは、日本の選抜制度が偏差値による傾斜選抜でありつつも、移行段階での「ご破算性」(竹内, 1991, 1995)を有している点に着目したものであり、学校外教育が進学における「挽回」手段となりうるという視点である。

2.2 先行研究の限界と本研究の位置づけ

先行研究では、学校外教育が教育達成に及ぼす影響について多くの知見が蓄積されてきたが、その多くは中学段階の利用に焦点を当てており、高校段階での学校外教育の役割や効果については、十分に検討されてこなかった。とりわけ、日本の教育制度において高校がトラックとして機能している点を踏まえると、高校段階での学校外教育の利用が、高校トラックの規定力を乗り越え、その後の進路選択にどのように影響するかを検討することは重要である。

また、学校外教育の効果に関する実証研究は、塾や家庭教師、通信教育の利用有無と教育成果との関連を単純に比較するものが多く、それが階層的背景とどのように連動しているかという媒介的メカニズムの検討は限られている。さらに、学校外教育の利用が単発的なものではなく、中学・高校と連続的・累積的である可能性が高いにもかかわらず、その連続性や累積性を考慮した分析も乏しい。

したがって、本研究では、出身階層が中学・高校段階での学校外教育の利用を通じて、どのように教育達成に影響を与えるのかという累積的因果関係を明らかにすることを目的とする。この観点からの分析は、学校外教育が教育機会の不平等を拡大するのか、あるいは縮小する可能性を持つのかという問いに対する新たな貢献となる。

2.3 研究仮説

以上の理論的枠組みおよび実証的知見を踏まえ、本研究では以下の仮説を設定する：

仮説. 出身階層が高い子どもほど、中学および高校段階で学校外教育を累積的に利用し、その結果としてより高い教育達成を遂げる傾向がある²⁾。

本研究では、この仮説を検証するため、出身階層を処置変数とし、中学・高校段階の学校外教育利用を時間的に順序のある媒介変数、最終学歴を基に構成された教育達成水準を順序尺度のアウトカム変数として設定する。

本研究の目的である学校外教育の利用パターンごとの累積的效果を検証するには、従来の媒介分析で用いられる総間接効果の推定だけでは不十分である。

そこで本研究では、VanderWeele (2015)や矢田 他 (2020)の議論に基づき、g-computation の

枠組みを採用する。これにより、学校外教育の利用パターン(「利用なし」「中学のみ」「高校のみ」「両方利用」)ごとに反実仮想的な教育達成の予測確率を算出し、どの利用パターンが最も効果的かを直接比較する。

3. 分析手法とデータ、変数

3.1 分析手法

上述のとおり、本研究では、仮説の検証と学校外教育の累積的効果の検討を行うため、*g-computation* の枠組みを用いる。具体的には、以下の手順で分析を進める。

まず、欠損値を多重代入法により補完したデータセットを用い、出身階層(A)、学校外教育利用パターン(M)、および共変量が教育達成(Y)に及ぼす影響を推定する順序ロジスティック回帰分析を行う。次に、推定されたモデルに基づき、出身階層(A)を高階層と低階層に固定した集団に対し、学校外教育の利用パターン(M)を仮想的に割り当てた場合の銘柄大学進学の予測確率を算出する。最後に、算出した予測確率を比較し、学校外教育の利用パターンごとの累積的効果の有無を直接的に検証する。

この手法により、複数カテゴリーの結果変数と媒介変数を伴う複雑なモデルにおいて、共変量を適切に調整した上での反実仮想的な因果効果を推定する。具体的な推定方法は本節の後半で説明を行う。

3.2 データ

本研究で使用するデータは、東京大学社会科学研究所が実施している4つのパネル調査のうち、若年パネル調査と壮年パネル調査である。若年パネル調査は2007年時点で日本全国に居住する20–34歳の男女を母集団として設定しており、壮年パネル調査は35–40歳の男女を母集団として設定している。これら2つの調査は、対象年齢が異なる以外、質問項目は同じであり、「働き方とライフスタイルの変化に関する全国調査(Japanese Life Course Panel Surveys: JLPS)」として2007年から2025年まで毎年1月から3月にかけて調査が実施された。三輪(2008)によると、標本抽出は層化二段無作為抽出であり、追跡調査であることを事前に対象者に伝えた上で、郵送で調査票を配布し、訪問して回収する方法をとっている。回収率は、若年調査で34.5%、壮年調査で40.4%である。なお、パネル調査データの使用にあたっては東大社研パネル運営委員会の許可を受けた。

JLPSはパネル調査であるので、サンプルの属性や意識、態度の変化を追跡し、その変化の要因を探ることができる。その一方で、JLPSはパネル調査の利点を活かし、子ども期の家庭環境や学校環境、および当時の習慣についての質問が豊富に用意されており、横断調査のように利用することも可能である(例えば、中澤, 2013など)。本研究でも、このようなJLPSの特性を活かし、横断データとして利用することとした。利用したデータは2007年の第1波、2008年の第2波である。

3.3 変数

従属変数である教育達成水準は、本人の最終学歴に基づいて「1 = 高卒以下」, 「2 = 短大・専門・高専卒」, 「3 = (銘柄大以外の)非銘柄大卒」, 「4 = 銘柄大卒」の4カテゴリを作成した。単なる高等教育機関への進学や大学進学ではなく、進学行動の多様性を踏まえた指標にすることで、高等教育における選抜構造をも部分的に反映でき、日本の教育達成における階層的構造を捉えるのに有効である。

処置変数(A)は、中学3年時点の出身階層を表す主成分スコアである。具体的には、眞田・

中西 (2025)を参考に、両親の学歴、父親の職業威信スコア、15歳時の家庭の経済状況(15歳時の暮らし向き・本の冊数・文化資本スコア)の変数から算出する³⁾。この方法は、OECDのPISA調査で長年用いられてきた「社会経済文化的背景(ESCS)」の構成法に準拠している。すなわち、親の教育水準・職業的地位・家庭の文化資源を統合して指標化する点は、教育社会学において国際的に広く認められている標準的な手法であり、本研究の主成分得点もこの枠組みに沿った操作的定義として妥当である(多喜, 2010; Matsuoka, 2014; 鳶島, 2016)。ただし、主成分得点は統計的に構成された連続指標であり、直接的に操作可能な介入変数ではない。この点で、因果推論における「well-defined intervention」とは性質を異にする。本研究では、この指標を「出身階層の代理」として用い、社会経済的背景が学校外教育や教育達成に及ぼす影響を条件付き因果効果として推定するという立場を取る。なお、出身階層スコアの構成要素(親の学歴等)は、処置変数(A)と強く相関するため、多重共線性を避ける観点から、媒介分析モデルにおける共変量としては投入していない。

媒介変数(M)は、学校外教育の利用状況である。中学時点での塾・家庭教師・通信教育のいずれかの利用を示す「中学時学校外教育(M₁)」と、高校時点での同様の利用を示す「高校時学校外教育(M₂)」の2つの二値変数として定義されている。本研究では、これらの利用が教育達成に対して媒介的に機能するかを検討する⁴⁾。

共変量(C)は、年齢、性別、15歳時の成績、大学進学アスピレーション、兄弟姉妹数である。これらは、出身階層によって一意に決定されるものではないが、その後の学校外教育利用と教育達成の双方に影響を与える重要な交絡要因と想定されるため、媒介分析のバイアス除去のために統制した。

本研究の分析対象は、JLPSの継続サンプル(N = 4,800)である。分析に用いた変数には表1に示す通り欠損値が一定程度存在する。これらを含むケースをリストワイズ除去するとサンプルサイズが大幅に減少し、推定にバイアスが生じる懸念があるため、本研究ではRのmiceパッケージを用いたMICE(Multiple Imputation by Chained Equations)による多重代入法により補完を行った。処置変数(A)のスコア算出にあたっては、欠損を含むデータで主成分分析を行うことによるバイアスを避けるため、以下の手順を用いた。まず、分析対象の全サンプルに対し、本研究で使用する全ての変数(処置変数の構成変数を含む)を投入した多重代入モデル(予測平均マッチング法:PMM)を構築し、20個の多重代入データセットを生成した。次に、生成された20個の各データセットにおいて主成分分析を実行し、それぞれの第1主成分得点を「出身階層スコア(A)」として算出し、以降の媒介分析に用いた。

使用した変数の記述統計量(平均値、標準偏差、四分位点等)は、多重代入後のプールした結果として表1に集約して示す。なお、主成分分析の詳細な結果(寄与率、負荷量)は、次節で確認する。

3.4 g-computation による推定手順

本研究では、学校外教育の利用パターンごとの因果効果を推定するため、g-computation (g-formula)を用いたシミュレーションを実施した(Robins, 1986)。具体的な手順は以下の通りである。第一に、欠損値によるバイアスを防ぐため、MICEによる多重代入法を用いてm = 20個の補完済みデータセットを作成した。第二に、各データセットにおいて、出身階層(A)、学校外教育利用パターン(M)および共変量(C)から教育達成(Y)を予測する順序ロジットモデル(式(3.1))を推定した。

$$(3.1) \quad \text{logit}[P(Y \leq j|A, M, C)] = \theta_j - (\delta A + \eta M + \zeta C)$$

ここで、jは教育達成の段階を表す。θ_jは切片のパラメーター、δ、ηはそれぞれ出身階

層と学校外教育の回帰係数, ζ は共変量ベクトル C に対する係数ベクトルである. 学校外教育利用パターン M は, 中学時学校外教育 (M_1), 高校時学校外教育 (M_2) の組み合わせ $(M_1, M_2) \in \{(0, 0), (1, 0), (0, 1), (1, 1)\}$ を表す.

第三に, 推定されたモデル(式(3.1))を用いて, 反実仮想シミュレーションを行った. 具体的には, 全サンプル ($i = 1, 2, \dots, N$) について, 共変量 (C_i) はそのまま, 出身階層を特定の値 (a), 学校外教育を特定のパターン (m^*) に固定した際の, 銘柄大卒の確率 $\hat{P}(Y = 4|A = a, M = m^*, C = C_i)$ を予測した. ここで, 出身階層の値 a は, 高階層のシミュレーションでは標本分布の第3四分位点(75パーセンタイル), 低階層のシミュレーションでは第1四分位点(25パーセンタイル)に設定した. 最後に, 得られた全サンプルの予測確率の標本平均を算出することで, 共変量の分布を周辺化した予測確率(Marginal Predicted Probability)を推定した(式(3.2)).

$$(3.2) \quad \hat{E}[Y(a, m^*)] = \frac{1}{N} \sum_{i=1}^N \hat{P}(Y = 4|A = a, M = m^*, C = C_i)$$

以上の手順を 20 個のデータセットすべてで実行し, 得られた推定値を Rubin のルールに基づき統合した結果を次節で示す.

4. 結果

4.1 分析対象者の記述統計量

本研究の分析に用いた変数の記述統計量を, 表 1(連続変数)および表 2(カテゴリ変数)に示す.

表 1 において, 出身階層 (A) と最も高い相関を示したのは父親学歴 ($r = 0.772$) と母親学歴 ($r = 0.730$) であった. なお, 多重代入前の欠損率は, 父職業威信スコア (20.3%) や 15 歳時の大学進学アスピレーション (18.0%) などで比較的高く, これら全ての変数は多重代入により処理されている. 次に, 表 2 より主要なカテゴリ変数の構成比(プール後の割合)を確認する. 教育達成水準 (Y) については, 高卒が 32.1%, 短大・専門・高専卒が 31.9%, 非銘柄大卒が 27.7% であり, 本研究が注目する銘柄大卒は 8.3% であった. 学校外教育利用パターン (M_1, M_2) については, 中学のみ利用 (1, 0) が 40.1% と最も高く, 次いで両方利用 (1, 1) が 30.8%, 全く利用しない (0, 0) 層は 26.0% であった. 一方で, 高校のみ利用 (0, 1) は 3.1% と極めて少ない分布を示した.

表 1. 使用した連続変数の記述統計と多重代入前の欠損率.

変数名	欠損率(%)	平均	標準偏差	最小	最大	Aとの相関
出身階層 (A)	37.7	0.0	1.6	-4.5	5.6	1.0
年齢 (X)	0.0	32.6	5.9	22.0	42.0	-0.3
15歳時の暮らし向き	1.7	3.1	0.8	1.0	5.0	0.6
15歳時の成績 (X)	3.0	3.2	1.2	1.0	5.0	0.3
15歳時の文化資本スコア	0.0	12.7	3.1	0.0	20.0	0.6
父職業威信スコア	20.3	51.7	8.6	36.7	90.1	0.6
15歳時の本の冊数	14.2	3.2	1.7	0.0	7.0	0.7
兄弟姉妹数 (X)	0.0	1.4	0.9	0.0	9.0	0.0
父親学歴 (年)	14.6	12.6	2.6	9.0	18.0	0.8
母親学歴 (年)	13.8	12.0	1.9	9.0	18.0	0.7

表 2. 使用したカテゴリー変数の記述統計量.

変数名とカテゴリー	構成比 (MICEで プールされた割合)	欠損率(%)	Aとの相関
教育達成水準 (Y) :			
銘柄大卒	8.3%	26.5	0.3
非銘柄大卒	27.7%	26.5	0.3
短大・高専・専門卒	31.9%	26.5	0.0
高卒	32.1%	26.5	-0.4
性別 (X) :			
女性 (0)	50.7%	0.0	0.0
男性 (1)	49.3%	0.0	0.0
学校外教育利用パターン (M1, M2)			
両方利用 (1,1)	30.8%	18.3	0.3
中学のみ利用 (1,0)	40.1%	18.3	-0.1
高校のみ利用 (0,1)	3.1%	18.3	0.1
全く利用していない (0,0)	26.0%	18.3	-0.3
大学進学アスピレーション (X) :			
なし (0)	46.5%	18.0	-0.4
あり (1)	53.5%	18.0	0.4

表 3. 出身階層指標の主成分分析.

変数	第1主成分
父親学歴	0.471
母親学歴	0.445
父職業威信スコア	0.388
15歳時の暮らし向き	0.339
15歳時の本の冊数	0.403
15歳時の文化資本スコア	0.391
分散の寄与率 (%)	44.80%
累積寄与率 (%)	44.80%

4.2 主成分分析による出身階層変数の作成

本分析の処置変数である出身階層(A)は、多重代入で補完されたデータセットごとに主成分分析(PCA)を実施し、第一主成分(PC1)のスコアとして算出された。表3に、PCAの結果として得られた第一主成分の負荷量を示す。表3が示すように、全ての構成変数が正の値の負荷量を持ち、特に父親学歴(0.471)と母親学歴(0.445)が高い負荷量を示した。また、この第一主成分の寄与率は44.8%であり、全変数が持つ分散の約半数がこの単一の次元によって説明されている。

4.3 学校外教育の利用パターンと銘柄大卒の関連

次に、学校外教育の利用パターンが銘柄大卒に及ぼす影響を比較するため、g-computationの枠組みを用いて、出身階層を高階層(第3四分位点)および低階層(第1四分位点)に固定した上で、学校外教育の利用パターンによって銘柄大卒の予測確率がどう変化するかをシミュレーションで検証した(表4)。

結果は以下の通りである。第一に、利用パターン間の比較において、銘柄大卒の確率は「高校のみ利用」と「両方利用」のパターンにおいて最も高い値を示した。「高校のみ利用」の場合、進学

表 4. 出身階層別に見た、学校外教育の利用パターンと銘柄大卒の予測確率。

学校外教育の利用パターン	銘柄大卒の予測確率	効果 (vs 全く利用しない)
高階層 (上位25%)		
全く利用しない (M00)	8.4%	(基準)
中学のみ利用 (M10)	7.6%	-0.8ポイント
高校のみ利用 (M01)	16.2%	7.8ポイント
両方利用 (M11)	14.9%	6.5ポイント
低階層 (下位25%)		
全く利用しない (M00)	4.3%	(基準)
中学のみ利用 (M10)	3.9%	-0.4ポイント
高校のみ利用 (M01)	9.0%	4.7ポイント
両方利用 (M11)	8.2%	3.8ポイント

確率は「全く利用しない」と比較して高階層で +7.8 ポイント (8.4% から 16.2%)、低階層で +4.7 ポイント (4.3% から 9.0%) の増加が見られた。「両方利用」も同様に、高階層で +6.5 ポイント、低階層で +3.8 ポイントの増加を示した。第二に、「中学のみ」の利用における銘柄大卒の確率は、「全く利用しない」と比較して、高階層で -0.8 ポイント、低階層で -0.4 ポイントと、数値上わずかに低い値を示した。第三に、階層間の比較において、全ての利用パターンを通じて、高階層の生徒の銘柄大卒の確率は低階層の生徒の約 2 倍の値を示した。例えば、最も銘柄大卒の確率が高い「高校のみ利用」の場合でも、低階層の確率は 9.0% であり、高階層の「全く利用しない」場合の確率 (8.4%) と近似した値にとどまった。

5. 考察

5.1 結果のまとめと仮説の検証

本研究は、出身階層が中学・高校段階での学校外教育利用を通じて、教育達成 (銘柄大卒) にかかる影響を及ぼすのかを、g-computation (反実仮想シミュレーション) の枠組みで検討した。分析の結果、当初の仮説「出身階層が高い子どもほど、中学および高校段階で学校外教育を累積的に利用し、その結果としてより高い教育達成を遂げる傾向がある」は、完全には支持されず、より複雑な構造を持つことが示された。以上の結果はいずれも、銘柄大卒に限定した分析結果であることに留意しつつ、具体的な内容を検討する。

第一に、銘柄大卒の確率を顕著に高めるのは、出身階層を問わず「高校のみ」で学校外教育を利用するパターンと「両方利用」のパターンであり、その効果はほぼ同程度であった。これまで Fujihara (2024) などが指摘してきた累積的な教育投資の優位性は、少なくとも学校外教育という文脈においては、必ずしも当てはまらない可能性がある。むしろ、大学受験に直結する高校段階での集中的な投資が、進学戦略として効率的であることが示唆された。

第二に、「中学のみ」の利用は銘柄大卒の確率の向上に寄与せず、高階層・低階層ともに 1 ポイント未満のわずかな低下にとどまった。本研究の推定では統計的な有意性や実質的な効果の大きさは限定的であり、この結果のみから中学段階の利用が進学機会を体系的に低下させるとまでは言い切れない。

ただし、中学段階で学校外教育を利用しながら高校段階で利用しない層が存在すること自体は、進学困難などの学力的課題を抱える生徒が早期に支援を求めている可能性 (セレクション効果) の一端を示唆していると考えられる (Entrich, 2020)。

第三に、「両方利用」の効果は「高校のみ」の効果を上回らなかった。したがって、出身階層の高さは学校外教育利用を通じて高い教育達成に結びつくが、そのプロセスは単純な累積の利用によるものではなく、より効果的なタイミング (高校段階) への戦略的投資によって特徴づけら

れる。

5.2 学校外教育の機能の再検討

今回の分析結果は、学校外教育が教育格差の再生産において果たす役割について、重要な示唆を与える。第一に、学校外教育は、大学受験に直結する高校段階での利用が最も戦略的有効性を持つことが示された。Stevenson and Baker (1992)は、日本の学校外教育が進学選抜のための重要な手段であることを指摘したが、本研究の結果は、その機能が特に高校段階において顕著に作動することを示している。第二に、「中学のみ」の利用が正の効果を持たなかった点は、学校外教育の補習的機能を示唆する。Entrich (2018, 2020)は、学校外教育が成績上位層の競争的戦略として機能する側面と、成績下位層の補習的戦略として機能する側面を持つと論じた。実際、Kuan (2011)は台湾のデータを用い、通塾の学力向上効果は、利用確率の低い生徒(成績下位や低階層)においてより大きくなることを示し、学校外教育が潜在的に強力な補習的機能を持つことを実証している。さらにMatsuoka (2018)は、日本の義務教育段階において、親の教育期待や地域の社会経済的背景がホットハウス効果(Hot House Effect)を生み出し、それが学力とは独立に学校外教育への参加を促進することを指摘している。本研究における「中学のみ利用」層もまた、こうした親の期待や不安によって動機づけられたセレクションの結果である可能性が高い。

しかし、本研究が示した最も重要な知見は、たとえ同じ学校外教育戦略をとったとしても、高階層の生徒の銘柄大学進学確率は、低階層の生徒の約2倍に達するという事実である。低階層の生徒が最も効果的な戦略(高校段階での利用)をとった場合の進学確率(約9%)は、高階層の生徒が何も利用しなかった場合の確率(約8%)と大差がなかった。この知見は、学校外教育を利用しても、なお埋められない根源的な階層差の存在を強調するMori and Baker (2010)や北村・竹ノ下 (2024)の議論とも一致する。また、Fujihara (2024)も、高校の学校歴を統制してもなお、出身家庭の経済力が大学進学に及ぼす直接的な影響は極めて大きいことを因果媒介分析により示している。高階層の家庭が持つ豊富な文化資本や情報資源、親の教育期待や家庭内での学習関与(Matsuoka, 2014)といった要素が、学校外教育の効果とは独立に、あるいはその効果を増幅させる形で、子どもの教育達成に影響を与えている可能性がある(DiPrete and Eirich, 2006)。

5.3 本研究の課題と今後の展望

一方で、本研究にはいくつかの限界が存在する。第一に、本研究では従属変数を「銘柄大卒」に限定して分析を行った。これは、Lucas (2001)がEMI仮説として提唱したように、日本社会における教育格差が、量的拡大(大学進学全般)よりも質的差異(難関大学への進学)という競争的な局面において最も顕著に現れると考えたためである(荒牧, 2000; Fujihara and Ishida, 2016; Mori and Baker, 2010)。しかし、学校外教育の効果は、中堅大学卒や、より広範な進路選択においても存在する可能性がある。銘柄大以外の進路に対する累積的效果の検討は、今後の課題である。

第二に、本分析では「高校受験組」と「中高一貫校出身者」を区別していない。Matsuoka (2018)が指摘するように、高校受験が存在しない私立中高一貫校の生徒は、公立中学生と比較して中学段階での学校外教育利用率が著しく低い傾向にある。したがって、中高一貫校出身者がサンプルに含まれている場合、彼らの「中学での非利用」は戦略的な選択や経済的制約ではなく、単に制度的な受験の不在を反映しているに過ぎない可能性がある。本研究で示されたセレクション効果の解釈を精緻化するためにも、学校タイプを考慮した詳細な分析が求められる。Fujihara (2024)が示したように、出身階層による大学進学格差の約25~33%は高校の学校ラン

クによって媒介されており、高校段階の制度的差異を考慮することが今後求められる。

第三に、出身階層指標は観察変数による代理指標であり、本研究で推定された効果は条件付き因果効果として解釈する必要がある。また、結果は g-computation のモデル仮定に依存している。最後に、今後は異なるデータや手法を用いて、本研究で示された知見の一般化可能性を検証する必要がある。特に、利用した学校外教育の種類や質が階層によってどう異なるかを詳細に分析することが、今後の重要な課題となるだろう。

5.4 結論

本研究は、出身階層が中学・高校段階における学校外教育の利用を通じて教育達成に及ぼす影響を、g-computation によって明らかにした。当初の仮説とは異なり、学校外教育の「累積の利用」は必ずしも最大の効果をもたらすわけではなく、「高校のみ」の利用と同程度の効果を持つことが示された。しかし、より重要な知見は、たとえ同じ教育戦略をとったとしても、高階層と低階層の間には学校外教育の利用戦略では埋めがたい進学確率の差が存在する可能性が高い。学校外教育は、銘柄大卒の確率を高める有効な手段の一つではあるが、それだけでは乗り越えられない、より根源的な階層格差の存在が示唆される。この教育格差のダイナミズムを、利用パターン間の効果の比較により実証的に示した点に、本研究の理論的・実践的意義がある。

注.

- 1) 本研究では、学習活動を行う学校外教育に焦点をあてている。
- 2) なお、本研究においては、出身階層を「高い」、「低い」と表現するが、これは社会学において一般的に用いられる表記である（例えば、佐藤, 2009; 平尾・太郎丸, 2011 など）。
- 3) 大学進学アスピレーションは、15歳時にどこまで進学を考えていたかをたずねた質問であり、回答者は高校から大学までの種別を回答している。その質問の回答が大学以上であったか否かでダミー変数を作成した。また、父親の職業威信スコアは、15歳時の父親の職業の回答を用いて、その職業を職業威信スコアに割り当てた。用いた職業威信スコアは2015年のSSM調査で用いられているコード（2015年社会階層と社会移動調査研究会, 2018）である。なお、職業威信スコアとは、各職業への評価の平均によって測定される0点から100点の連続的な値である。詳しくは、脇田（2012）などを参照されたい。15歳時の文化資本スコアは、15歳時に家庭にどのような家具や家財があったかについての20個の質問項目（例えば車やエアコン、ピアノなど）を単純に足し合わせ、数が多いほど文化資本を多く保有しているというように設定した。また、15歳時の成績は、15歳時の成績がどうだったか回顧的に思い出してもらい、その自己評価を「1. 上の方」～「5. 下の方」という5件法でたずねている。本研究ではその値を反転し、値が大きいほど15歳時の成績の自己評価が高くなるようにリコードを行った。15歳時の暮らし向きについても同様に、回顧的に自己評価をたずね、「1. 豊か」～「5. 貧しい」の5件法の回答を反転し用いた。
- 4) 先行研究では学校外教育の詳細（頻度・費用・目的）まで検討したものもあるが（北村・竹ノ下, 2024）、本研究では、媒介効果の累積的構造に焦点をあてており、教育段階ごとの「利用経験の有無」に基づく二値変数を用いている。これは、使用データ（JLPS）において利用頻度や金額などの情報がえられないという制約によるが、教育段階を通じた学校外教育の累積的利用の有無を捉えるという観点からは、代理指標になり得ると考えた。21世紀出生児縦断調査などの別データを用いたより詳細な指標に基づく分析は、今後の課題である。

謝 辞

本研究は、日本学術振興会(JSPS)科学研究費補助金・特別推進研究(25000001, 18H05204), 基盤研究(S)(18103003, 22223005), 研究活動スタート支援研究(22K20190)の助成を受けたものです。東京大学社会科学研究所(東大社研)パネル調査の実施にあたっては、社会科学研究所研究資金、株式会社アウトソーシングからの奨学寄付金を受けました。パネル調査データの使用にあたっては東大社研パネル運営委員会の許可を受けました。執筆にあたり、匿名の2名の査読者および編集委員会の先生方より有益なコメントをいただきました。記して感謝申し上げます。

参 考 文 献

- 2015年社会階層と社会移動調査研究会(2018). 2015年SSM調査コード・ブック, 2015年社会階層と社会移動調査研究会.
- 荒牧草平(2000). 教育機会の格差は縮小したか：教育環境の変化と出身階層間格差, 『日本の階層システム3：戦後日本の教育社会』(近藤博之 編), 15-35, 東京大学出版会, 東京.
- Bray, M. and Kwok, P. (2003). Demand for private supplementary tutoring: Conceptual considerations, and socioeconomic patterns in Hong Kong, *Economics of Education Review*, **22**(6), 611-620.
- Buchmann, C., Condron, C. D. and Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment, *Social Forces*, **89**, 435-461.
- Cummings, W. K. (1997). *Private Education in Eastern Asia, The Challenge of Eastern Asian Education: Implications for America* (eds. W. K. Cummings and P. G. Altbach), 135-152, State University of New York Press, Albany.
- DiPrete, T. A. and Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments, *Annual Review of Sociology*, **32**, 271-297.
- Entrich, S. R. (2018). *Shadow Education and Social Inequalities in Japan*, Springer International Publishing, Cham.
- Entrich, S. R. (2020). Worldwide shadow education and social inequality: Explaining differences in the socioeconomic gap in access to shadow education across 63 societies, *International Journal of Comparative Sociology*, **61**(6), 441-475.
- Fujihara, S. (2024). Identifying the role of high school in educational inequality: A causal mediation approach, *Social Science Research*, **124**, 103077.
- Fujihara, S. and Ishida, H. (2016). The absolute and relative values of education and the inequality of educational opportunity: Trends in access to education in postwar Japan, *Research in Social Stratification and Mobility*, **43**, 25-37.
- Hannum, E., Ishida, H., Park, H. and Tam, T. (2019). Education in East Asian societies: Postwar expansion and the evolution of inequality, *Annual Review of Sociology*, **45**(1), 625-647.
- 平尾一朗, 太郎丸博(2011). 世代間移動レジームにおける非正規雇用の位置, 理論と方法, **26**(2), 355-370.
- 北村友宏, 竹ノ下弘久(2024). なぜ学校外教育に投資するのか：数理モデルによる投資メカニズムの解明, 理論と方法, **39**(1), 79-92.
- Kuan, P. Y. (2011). Effects of cram schooling on mathematics performance: Evidence from junior high students in Taiwan, *Comparative Education Review*, **55**(3), 342-368.
- Lucas, S. R. (2001). Effectively maintained inequality: Education transitions, track mobility, and social background effects, *American Journal of Sociology*, **106**(6), 1642-1690.
- Matsuoka, R. (2014). An empirical investigation of relationships between junior high school students' family socioeconomic status, parental involvement, and academic performance in Japan, *Soci-*

- ological Theory and Methods*, **29**(1), 147–165.
- Matsuoka, R. (2018). Inequality in shadow education participation in an egalitarian compulsory education system, *Comparative Education Review*, **62**(4), 565–586.
- Merton, R. K. (1968). The Matthew effect in science, *Science*, **159**(3810), 56–63.
- 耳塚寛明 (1993). 学校社会学研究の展開, *教育社会学研究*, **52**, 115–136.
- 三輪哲 (2008). 働き方とライフスタイルの変化に関する全国調査 2007 における標本特性と欠票についての基礎分析, 東京大学社会科学研究所パネル調査プロジェクトディスカッションペーパーシリーズ, **10**, https://csrda.iss.u-tokyo.ac.jp/panel/dp/PanelDP_010Miwa.pdf (最終アクセス日 2025 年 8 月 20 日).
- Mori, I. and Baker, D. P. (2010). The origin of universal shadow education: What the supplemental education phenomenon tells us about the postmodern institution of education, *Asia Pacific Education Review*, **11**, 36–48.
- 中村高康 (2018). 相対的学歴指標と教育機会の趨勢分析: 2015 年 SSM 調査データを用いて, 理論と方法, **33**(2), 247–260.
- 中西啓喜 (2018). トラッキングが高校生の教育期待に及ぼす影響—パネルデータを用いた傾向スコア・マッチングによる検証—, *ソシオロジ*, **62**(3), 41–59.
- 中澤涉 (2013). 通塾が進路選択に及ぼす因果効果の異質性: 傾向スコア・マッチングの応用, *教育社会学研究*, **92**, 151–174.
- Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—Application to control of the healthy worker survivor effect, *Mathematical Modelling*, **7**(9-12), 1393–1512.
- 眞田英毅 (2018). 高校進学における学校外教育の効果: 低階層の子どもたちの教育達成, *社会学年報*, **47**, 69–82.
- 眞田英毅 (2022). 高等教育への進学における学校外教育の効果: トラッキング後の挽回, 理論と方法, **37**(2), 184–198.
- 眞田英毅, 中西寛子 (2025). 15 歳時に本を持っていることは何を意味するのか: 社会階層論の観点から, SSE-DP-2025-4, <https://stat-expert.ism.ac.jp/wp/wp-content/uploads/2025/03/SSE-DP-2025-4.pdf> (最終アクセス日 2025 年 6 月 20 日)
- 佐藤嘉倫 (2009). 現代日本の階層構造の流動性と格差, *社会学評論*, **59**(4), 632–647.
- Stevenson, D. L. and Baker, D. P. (1992). Shadow education and allocation in formal schooling: Transition to university in Japan, *American Journal of Sociology*, **97**(6), 1639–1657.
- Takashiro, N. (2021). Determinants of middle school students' participation in shadow education in Japan, *Social Indicators Research*, **155**, 1119–1136.
- 竹内洋 (1991). 日本型選抜の探求: 御破算型選抜規範, *教育社会学研究*, **49**, 34–56.
- 竹内洋 (1995). 『日本のメリトクラシー—構造と心性』, 東京大学出版会, 東京.
- 多喜弘文 (2010). 社会経済的地位と学力の国際比較—後期中等教育段階における教育と不平等の日本の特徴—, 理論と方法, **25**(2), 229–248.
- 鳶島修治 (2016). 読解リテラシーの社会経済的格差—PISA2009 のデータを用いた分析—, *教育社会学研究*, **98**, 219–239.
- VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*, Oxford University Press, New York.
- 脇田彩 (2012). 職業威信スコアのジェンダー中立性—男女別職業評価調査に基づく一考察—, *ソシオロジ*, **57**(2), 3–18.
- 矢田真城, 魚住龍史, 田栗正隆 (2020). 反事実モデルに基づく直接効果と間接効果の推定, *計量生物学*, **40**(2), 81–116.
- Yung, K. W. H. and Zeng, C. (2022). Parentocracy within meritocracy: Parental perspective on lecture-style English private tutoring in Hong Kong, *Language and Education*, **36**(4), 378–394.

Cumulative Effects of Shadow Education: The Impact of Usage Patterns and Timing on Educational Attainment

Teruki Sanada

School of Social Informatics, Mukogawa Women's University

This study examines the effects of shadow education on educational attainment in Japan, focusing on how family socioeconomic background influences attainment through different usage patterns and timings during lower and upper secondary school. Using data from the Japanese Life Course Panel Surveys (JLPS), we apply the g-computation (counterfactual simulation) framework to compare the causal effects of various usage strategies, operationalized as graduation from a “branded” (selective) university. Family background is measured by a composite index derived from principal component analysis, and usage patterns are categorized across two stages: junior high and high school. Contrary to the initial hypothesis, cumulative use (junior high and high school) did not necessarily yield the greatest effect. The primary finding is that, regardless of social origin, using shadow education “only during high school” was as effective as the cumulative pattern for entrance to “branded” universities, suggesting the strategic timing of investment is crucial. Conversely, using shadow education “only during junior high” did not contribute to an increase in attainment probability, which suggests a strong possibility of selection effects, where students facing academic difficulty seek early assistance. Most importantly, even when employing the same shadow education strategy, students from high-SES backgrounds were approximately twice as likely to enter “branded” universities as their low-SES counterparts, highlighting the persistent direct effect of social origin not mediated by shadow education.

共通オッズ比とランク累積分布関数による ランキングデータ解析：世界大学ランキングを事例に

樊 怡舟[†]

(受付 2025 年 6 月 26 日；改訂 12 月 31 日；採択 2026 年 1 月 7 日)

要 旨

高等教育国際化が進んできている中、大学ランキングに大変注目が集まっており、実際の大学運営や政策形成に関わる重要な指標とされてきている。本稿は、こうしたランキングデータの解析方法として、世界大学ランキングにおける各国のランクイン状況を「世界平均水準」と比較可能な形で定量化するためのノンパラメトリックな枠組みを提案する。具体的には、Top100, Top200, ..., Top500 といった複数のベンチマークに対応する 2×2 分割表を構成し、共通オッズ比(Common Odds Ratio, COR)を抽出する手法と、ランク累積分布関数(Cumulative Distribution Function of Rank, CDFR)を導入し、ランキング構造の比較と要約を可能にする。

さらに、CDFR に単峰性仮定を導入することで、各国のランキング分布を「極度先進国型」「中堅先進国型」「発展途上国型」に分類し、国ごとの構造的特徴を明らかにする。分析においては、世界大学学術ランキング(Academic Ranking of World Universities, ARWU)過去 14 年分の Top500 データを用い、中国・アメリカ・イギリスの 3 か国を事例に、ランクイン数・COR・CDFR の年次推移を比較した。

その結果、中国は CDFR が凸型を示し、年を追うごとに COR も上昇する比較的「発展途上国型」、アメリカは明確な凹型と高い COR を持つ「極度先進国型」、イギリスは S 字型の CDFR を示す「中堅先進国型」として特徴づけられることが示された。本研究は、国際ランキングの持つ構造的制約や母集団の仮定による限界も併せて議論しつつ、統計的に理論整合性のある「世界平均水準」の定式化と国際比較の可能性を示唆する。

キーワード：ノンパラメトリック手法、比較的発展途上国型、先進国型、ARWU、共通オッズ比。

1. はじめに

1990 年代後半以降、学術研究と教育活動の国際化が進展する中、各国の高等教育システムのパフォーマンスを国際的に可視化する試みとして、世界大学ランキングは急速に発展・普及してきた (Rauhvargers, 2011, 2013)。現在では、Times Higher Education (THE), QS World University Rankings, ARWU などが広く知られる。これらのランキングは、高等教育機関の国際的な位置づけを示すものとして、大学経営や政策決定に大いに活用される (El Gibari et al., 2018; Williams et al., 2013)。

一方で、ランキングをめぐるのは、「どの国の高等教育システムが相対的に台頭しているの

[†] 広島大学 高等教育研究開発センター：〒739-8512 広島県東広島市鏡山 1-2-2; fanyizhou921106@yahoo.co.jp

か」という、個別の順位づけを超えたマクロな関心が生じる。高等教育システムの成熟段階が異なれば、ランクインの分布や占有の様式が同一であるはずはなく、構造そのものが異なる。政策制定の情景から考えると、個別大学の順位変動よりも、そういったランキングに内在する構造として現れる地域間の差異を把握することが重要である。

世界大学ランキングを用いて国レベルのプレゼンスに着目した分析としては、Top500やTop100にランクインした大学数を国別に集計し、GDPや人口との関連を検討する研究(Günay, 2022; Marginson, 2007)、順位に応じてスコアを割り当てて各国の合計スコアを算出する方法(Pietrucha, 2018)、あるいはランクイン大学数を人口で割ることで効率性を測定する試み(Benito et al., 2020; Lu, 2014)などが挙げられる。

しかし、これらの試みには限界がいくつかある。まず、順位に割り当てるスコアの関数形に理論的根拠が乏しく、分析者の裁量に大きく依存する。また、アメリカなどGDP・ランクイン大学数ともに突出して大きい国が存在し、相関の分析が安定ではない。さらに、スコアの合計や平均といった操作は、ランキングの順序尺度的性質と必ずしも整合しない。

そこで本稿は、個々の大学のランク情報というマイクロデータを出発点とし、国・地域レベルでのランクイン情勢及びその経年変化を記述する枠組みを提案する。具体的には、共通オッズ比(Common Odds Ratio: COR)およびランク累積分布関数(Cumulative Distribution Function of Rank: CDFR)を導入し、先進国型の成熟したシステムと新興国型のシステムとの対比を念頭に、アメリカ、イギリスと中国を代表的事例として分析を行う。

2. 本稿の使用データ

本稿はデータとしてARWUを取り上げる。ARWUは2003年に上海交通大学によって開発された大学ランキングであり、論文被引用数、NatureおよびScience誌への論文掲載数、Highly Cited Researchers (HiCi)の数など、客観指標から算出する点で、QSやTHEといった他の主要なランキングと一線を画している¹⁾。本稿では、2011年から2024年までの各年度のTop500を対象とし、Top100, Top200, Top300, Top400, Top500という5つのベンチマークを設定し、各国のランクイン割合の推移を分析していく。

図1は各ベンチマークにおけるアメリカの大学のランクイン割合の推移を、図2は中国の大学について示したものである。両国のランクイン構造において水準・推移・層別配置の三点で明確な違いが観察される。①どのカテゴリーにおいてもアメリカのシェアが高い。②アメリカ

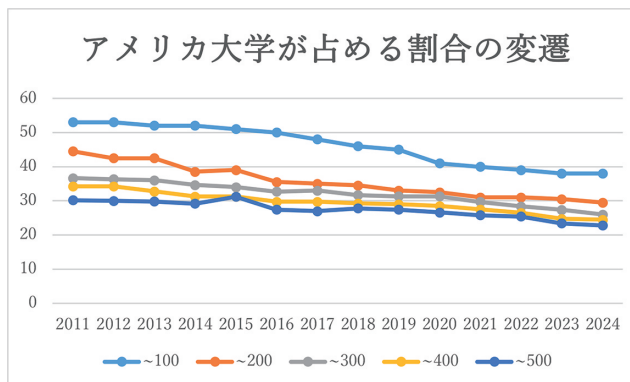


図1. カテゴリー別 ARWU におけるアメリカの大学の割合推移.

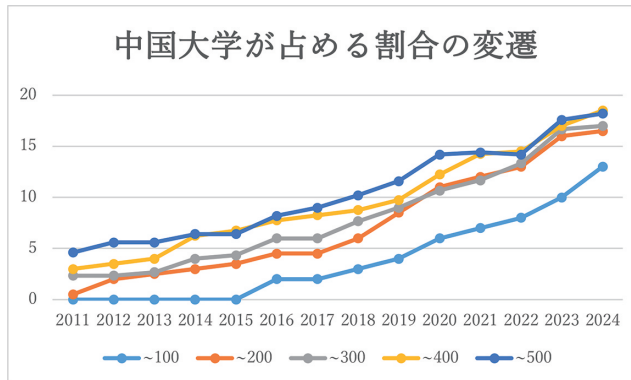


図 2. カテゴリー別 ARWU における中国の大学の割合推移。

は緩やかな減少傾向を示すが、中国は比較的急速な増加を示す。③中国は Top100 から Top500 にかけてベンチマークが下位になるにつれてシェアが拡大する構造に対し、アメリカは逆に上位カテゴリーの方が高いシェアを持つ構造となっている。

次節より、このような構造の違いを読み取るためのツールを導入していく。

3. 共通オッズ比

大学ランキングにおける各国の上位入選状況は分割表として記述することができる。例えば、A 国の Top100 のランクイン事情は以下ようになる。なお、全世界の大学数を N 、A 国の大学数を N_1 とし、Top100 にランクインできた大学のうち t 校が A 国の大学とする。

A 国の入選率を p_A 、そしてその他の国の入選率を p_R とおくと、表 1 の各行はそれぞれ二項分布 $B(N_1, p_A)$ 及び $B(N - N_1, p_R)$ に従う²⁾。A 国とその他の国の間に差が全く存在しないという状態、すなわち $p_A = p_R$ の場合を世界平均水準にあるとする。つまり、入選オッズ比を確認することで、A 国が世界平均水準以上か以下かを記述できる。

一方、本稿のデータは、Top100, Top200, ..., Top500 など複数のカテゴリーから構成されるため、各層の状況から全体のランクイン事情を要約する必要がある。例えば、Top100 に 5 点、Top101-200 に 4 点、以下 3 点、2 点、1 点という順にウェイトづけで各層を評価する。このようなスコア化は計算上の簡便さや直感的理解といった利点を持つ一方、ウェイト設定について

表 1. Top100 ランクイン事情に関する 2×2 の分割表。

	Top100	落選	
A 国	t	$N_1 - t$	N_1
その他	$100 - t$	$N - N_1 - 100 + t$	$N - N_1$
	100	$N - 100$	N

表 2. 第 j カテゴリー (Top K_j) に対応する 2×2 分割表。

	Top K_j	落選	
A 国	t_j	$N_1 - t_j$	N_1
その他	$K_j - t_j$	$N - N_1 - K_j + t_j$	$N - N_1$
	K_j	$N - K_j$	N

は客観的な基準が存在せず、恣意的判断に大きく依存する。

表2は、 j 番目のベンチマークカテゴリー $\text{Top}K_j$ までのランキングにおいて、A国が保有する大学総数 N_1 のうち、 t_j 校がランクインしている状況を表している。ここで、各 $\text{Top}K_j$ を一つの層とみなす。本稿では、ノンパラメトリックな層別解析の枠組みから、次のような COR を定義する。

$$\text{Common Odds Ratio} = \frac{\sum_j \frac{t_j(N - N_1 - K_j + t_j)}{N}}{\sum_j \frac{(K_j - t_j)(N_1 - t_j)}{N}}$$

COR がちょうど1となる際に、次の式が成り立つ。

$$\sum_j \frac{t_j}{N_1} = \sum_j \frac{K_j}{N}$$

つまり、A国のランクイン数の合計が、その大学数の世界シェアと各カテゴリーの総枠数に比例している時に、世界平均と等しい水準と判断されることになる。

ただ、層別解析の場合、層ごとに重複しないサブグループを前提とするが、ランキングデータと構造的に異なる。しかし逆に、この「重複構造」は、上位層をより強く評価するスコア化法と類似した機能を果たしている。さらに、スコア化法に比べて、CORは恣意性を制御しながらもランキングの階層構造をある程度尊重できていると言える。

無論、「Top100」「Top101~200」…で区分することで、入選大学が各層において重複にカウントされない分割表を作成できる。その際の共通オッズは以下のように計算される。

$$\text{Common Odds Ratio} = \frac{\sum_j \frac{t_j(N - N_1 - K_j + K_{j-1} + t_j)}{N}}{\sum_j \frac{(K_j - K_{j-1} - t_j)(N_1 - t_j)}{N}}$$

COR がちょうど1になる際に、 $N \sum_j t_j = N_1 K_J$ となり、 K_J は最終層までのランキングとする。この時、各層のランクイン数 t_j の具体的な構成に依存せず、最後の区切りにおけるランクイン状況しか機能しておらず、有効な統合とは言い難い。

表2に基づいた COR で該当国が世界平均水準以上か以下かについて記述可能である。

4. ランク累積分布関数

本節では、カテゴリーを跨いだ占有率の構造変化を視覚的に記述するランク累積分布関数、すなわちA国の何割の大学がある順位までランクインできたかを示す指標を導入する。

世界ランキングの1位から第 R 位までA国所属大学の累積数を $t(R)$ とする。A国の大学数を N_1 とし、順位 $r \in \{1, 2, 3, \dots, R\}$ において、A国のCDFRは次のように定義される。

$$\text{CDFR}(r) = \frac{t(r)}{N_1}$$

r 位までの占有率 $\frac{t(r)}{N_1}$ は、 $(r, t(r))$ と原点を結ぶ直線の傾きである。世界全体を対象としたCDFRは、 $(0, 0)$ から $(N, 1)$ に至る直線である。これは、どのカテゴリーでも等確率でランクインするという仮想的な世界平均の状況であり、「世界平均水準の基準線」とする。

一方、前節では世界平均水準をCORが1として定義したが、このことはCDFRにおいてどう表現されるだろうか。CORが1の場合、

$$\sum_j \frac{t_j}{N_1} = \sum_j \frac{K_j}{N}$$

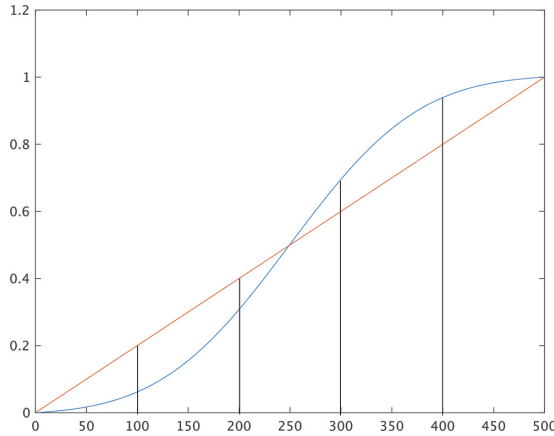


図 3. 共通オッズ比が 1 の集団の CDFR と世界平均水準基準線との関係.

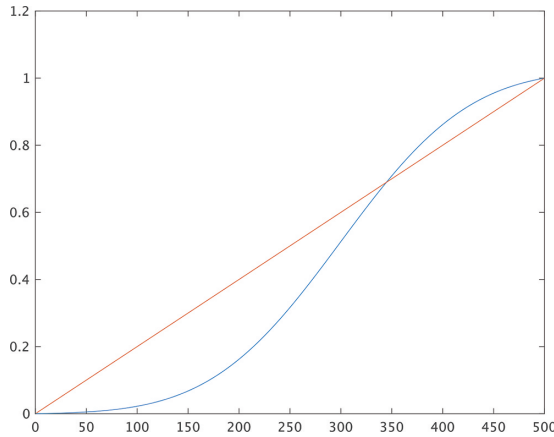


図 4. S 字型 CDFR 曲線.

左辺は各ベンチマークカテゴリー (Top100, Top200, ...) における A 国の CDFR 値の総和, 右辺は対応する世界平均水準基準線上の値の総和である. CDFR 曲線と世界平均基準線との偏差 (上回り・下回り) は全体としてちょうど打ち消し合うことを意味する (図 3).

この関係をさらに細かくみるため, 隣接するベンチマーク K_j と K_{j+1} の間, A 国のランキング占有がどのように変化しているかを考える. その差分を次式で定義する.

$$\Delta_{K_j, K_{j+1}} = \frac{t(K_{j+1})}{K_{j+1}} - \frac{t(K_j)}{K_j}$$

ここで, 単峰性の仮定をおくこととする. すなわち, あるカテゴリー j^* が存在して, 以下の不等式が成り立つ.

$$\Delta_{K_1, K_2}, \dots, \Delta_{K_{j^*-1}, K_{j^*}} \geq 0, \Delta_{K_{j^*}, K_{j^*+1}}, \dots, \Delta_{K_4, K_5} \leq 0$$

K_{j^*} は各国の機関が最も集中する順位帯と解釈でき, その値が大きいくほど, 上位よりも下位に分布しやすいことを意味する. このことに基づいて, CDFR の形を通して該当国が先進国か比較的發展途上かについて記述することが可能である.

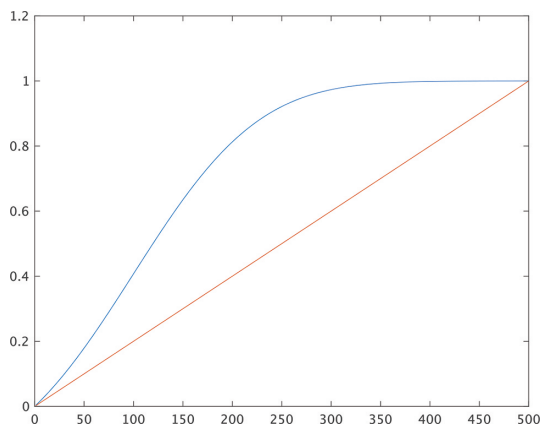


図 5. 凹型 CDFR 曲線.

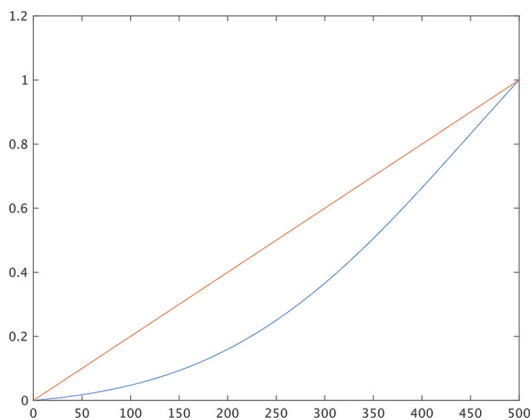


図 6. 凸型 CDFR 曲線.

本稿では 100 位刻みでベンチマーク (Top100, Top200, ...) を設定しているため, 単峰点 $K_{j^*} > 400$ の場合, 観測可能な範囲においてランキング占有率は単調増加関数となる. このときは, A 国の占有割合の単調増加が観測され, CDFR はその区間において凸関数となる (図 4). この凸型 CDFR は「相対的に発展途上」にある国のパターンと解釈できる.

$100 \leq K_{j^*} \leq 400$ であれば, ランキング占有率は単調増加の部分と単調減少の部分とを両方観測できる. A 国の大学のランキング占有率はベンチマークが下位へと進むにつれて一旦上昇し, やがて下降するという推移を示す. CDFR は典型的な S 字型に近い形状を描くと期待される (図 5). そして, 単峰点 K_{j^*} において, CDFR は下に凸から凹に転換する.

一方, 単峰点 $K_{j^*} < 100$ の場合, 観測範囲ではランキング占有率が単調減少関数となる. その結果として, A 国の占有割合が早期から一貫し遞減していくような, 単調減少的な傾向のみが捉えられる可能性がある. このとき, CDFR は全体として凹関数となり (図 6), 初期に急増し, その後漸近的に伸びが緩やかになるような形状を示す.

5. 分析

本節では、中国、アメリカ、イギリスの3か国を代表例として取り上げ、本稿が提案したCORとランク累積分布関数という二つの記述統計量を用い、それぞれの国・地域における世界大学ランキング上のプレゼンスを記述・分析する。

具体的には、以下の三つの作業を行う。

①過去14年間(2011~2024年)におけるARWUのTop500にランクインした各国・地域の大学数を集計し、②その上で、Top500のランクイン情報から各年における該当国のCORを計算し、③併せて、CDFR曲線の推移を可視化する。ただし図の可読性を確保するため、本稿では代表年として2011年(分析の起点)、2017年(中間年)、および2024年(直近年)の3時点についてCDFRを描画し、その形状と傾向を定性的に比較・解釈する。

アメリカの過去14年間のCORはいずれも大幅に1を超える(表3)。2011年時点でCORは2.89と極めて高く、CORは緩やかに低下するが、依然として世界ランキングで圧倒的優位を維持している。CDFRはいずれの年も明確な凹型で、単峰点 K_{j^*} が非常に小さい「極度先進国型」である。下位カテゴリーに向かうほどランクイン占有率が逡減する構造的特徴を有し、図1が示した傾向と整合的である(図7)。つまり、アメリカは極めて高い競争力を保ちつつも、「相対的發展途上国」の台頭に伴い、その相対的優位が徐々に縮小するという「極度先進国型」の動態が捉えられる。

対して、中国の過去14年間のCORは1にまだ未達しているものの年々上昇し、中国が「世界平均水準」に接近しつつあることを示す(表4)。CDFRはいずれも凸型で、下位カテゴリーに向かうほど占有率が増加するという傾向を持っており、図2のシェア推移とも一致する。

表3. 年次別ARWUにランクインできたアメリカの大学数.

アメリカ	~100	~200	~300	~400	~500	COR
2011	53	89	110	137	151	2.8946
2012	53	85	109	137	150	2.8038
2013	52	85	108	131	149	2.5770
2014	52	77	104	125	146	2.2220
2015	51	78	102	125	156	1.6494
2016	50	71	98	119	137	2.2403
2017	48	70	99	119	135	2.3357
2018	46	69	95	117	139	1.8323
2019	45	66	94	116	137	1.7994
2020	41	65	94	114	133	1.8548
2021	40	62	89	110	129	1.7534
2022	39	62	85	106	127	1.6432
2023	38	61	82	99	117	1.8939
2024	38	59	78	98	114	1.8896

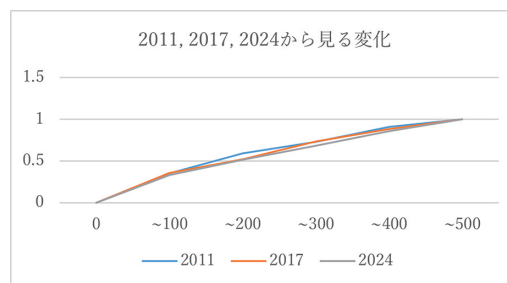


図7. 2011, 2017, 2024におけるアメリカのCDFRの推移.

表 4. 年次別 ARWU にランクインできた中国の大学数.

中国	~100	~200	~300	~400	~500	COR
2011	0	1	7	12	23	0.1685
2012	0	4	7	14	28	0.1868
2013	0	5	8	16	28	0.2387
2014	0	6	12	25	32	0.3686
2015	0	7	13	27	32	0.4460
2016	2	9	18	31	41	0.4530
2017	2	9	18	33	45	0.3941
2018	3	12	23	35	51	0.4332
2019	4	17	27	39	58	0.4823
2020	6	22	32	49	71	0.5001
2021	7	24	35	57	72	0.6434
2022	8	26	40	58	71	0.8087
2023	10	32	50	68	88	0.7544
2024	13	33	51	74	91	0.8280

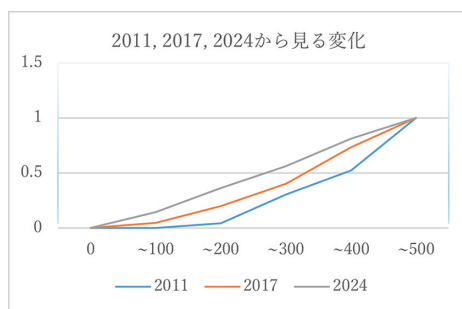


図 8. 2011, 2017, 2024 における中国の CDFR の推移.

表 5. 年次別 ARWU にランクインできたイギリスの大学数.

イギリス	~100	~200	~300	~400	~500	COR
2011	10	19	29	33	37	1.9116
2012	9	19	30	33	38	1.7446
2013	9	19	29	33	37	1.8473
2014	8	20	29	33	38	1.6813
2015	9	21	28	33	37	1.9112
2016	8	21	28	33	37	1.8467
2017	9	20	28	34	38	1.7453
2018	8	21	28	34	39	1.5965
2019	8	21	26	34	36	1.9733
2020	8	20	28	34	36	2.0760
2021	8	20	25	33	38	1.4382
2022	8	21	25	33	38	1.4905
2023	8	20	25	33	38	1.4382
2024	8	20	25	31	35	1.7439

CDFR 曲線の年次推移に注目すれば、2024 年において中国の CDFR 曲線は、世界平均水準の基準線に近接しており(図 8)、つまり、中国は依然平均水準未満ながら、急速に追いつきつつある国として視覚的に確認可能である。

イギリスの過去 14 年間ランクイン数はほぼ横ばいで推移し、COR も 2011 年の 1.91 から 2024 年の 1.74 へと小幅に低下したものの、一貫して世界平均を上回る(表 5)。CDFR はいずれも S 字型を示し、単峰点 K_{j^*} が 100 以上 400 以下、すなわちベンチマークの中間域に位置しており、「中堅先進国型」の分布構造に該当する(図 9)。S 字型の下では占有率が Top200 付近

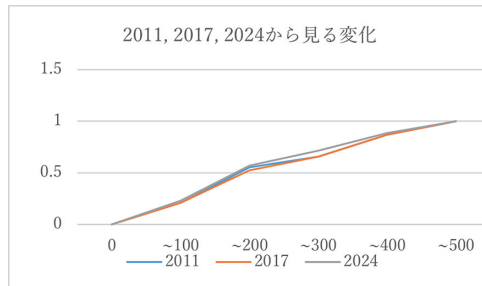


図 9. 2011, 2017, 2024 におけるイギリスの CDFR の推移。

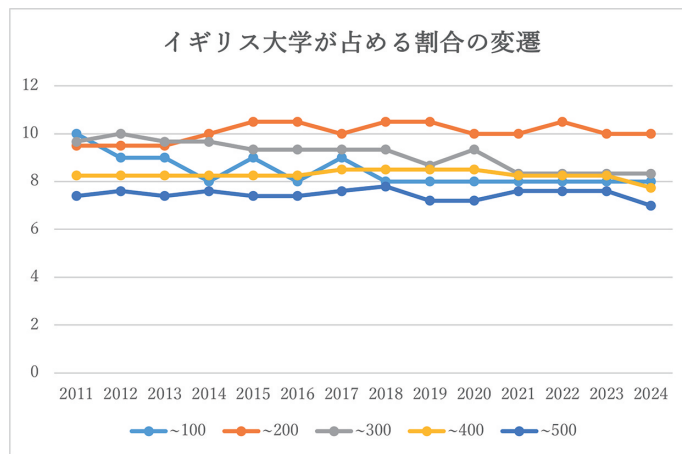


図 10. カテゴリー別 ARWU におけるイギリスの大学の割合推移。

でピークを示し、以降は低下する傾向が観察され、図 10 のシェア推移とも整合的である。以上より、イギリスは長期にわたり世界平均を安定的に上回りつつ、中間順位帯に強みを有する「中堅先進国型」の代表例として位置づけられる。

6. おわりに

本稿は、階層構造をもつランキングデータ全般に適用可能な記述的分析を目指し、複数の 2×2 分割表で表すランキング事情をまとめる COR、そしてベンチマークカテゴリーごとの占有率変化を視覚的記述するランク累積分布関数を提案した。両ツールは整合的であり、ランキングの階層的構造に対し、COR は非対称的なウェイトづけを内在させた統合指標であり、CDFR との関係は図式的に再解釈できる。

応用例として、提案手法を過去 14 年分の世界大学ランキングデータに適用してみた。単峰性仮定を導入した結果、CDFR は極度先進国型(凹型)、中堅先進国型(S 字型)と比較的發展途上型(凸型)の三類型に分類できる。分析結果では、アメリカは高水準の COR を維持しつつも緩やかに低下し、CDFR が凹型を示すことから、極度先進国型の地位を保ちながら相対的優位が縮小している。中国は年々世界平均に追いついてきているが、未だ COR が 1 に達しておらず、凸型 CDFR から読み取れるように下位層中心の比較的發展途上モデルである。イギリスは安定して COR が 1 を上回り、CDFR が S 字型を描くことから、中堅先進国型としての特徴

が明瞭である。

本稿の分析枠組みは、世界大学ランキングに特化したものではなく、階層構造をもつランキングデータ全般に適用可能な汎用的手法である。例えば、映画ランキング、企業の業績ランキングなど、順位やカテゴリーによって区分されたデータに対して、マクロの特徴を記述する手段として有効な視座を提供するものである。

本稿の分析には、いくつかの限界が存在する。

第一に、本研究のデータでは、Top500外の“落選大学”の分布は観測されない。「世界の研究大学群」の規模 N を大きく想定するほど、CDFR(ランク累積分布関数)の傾斜は緩やかになるため、本稿の分析は、事実上「最上位エリート研究大学群」に絞った相対的比較に位置づけられる。ただし、CDFRの形状は母集団サイズ N の設定に依存せず保持され、カテゴリーごとの占有率の変化は N の大小にかかわらず観察可能である。したがって、CDFRを用いた相対的比較は、記述的分析として有効に機能していると言える。

第二に、CORの年次比較には構造的制約が伴う。各年でTop500内の大学構成が変化するため、実質的に母集団が年ごとに異なる。ゆえにCORの上昇や下降が、必ずしも実質的な研究力の向上や衰退を意味するとは限らない。特に、Top401~500のランクイン数が減少すれば、分母構成の変化によってCORが逆に上昇する可能性もある。年次比較を行う場合には、こうした構造的な依存関係を慎重に考慮する必要がある。

第三に、「研究大学」としてカウントされる基準について世界的に統一された定義が存在しない点も留意が必要である。仮に母集団を過度に拡張し、性質の異なる大学群を含めれば、CDFRの単峰性仮定が崩れ、分布構造の比較可能性が損なわれるおそれがある。

最後に、本稿は記述統計の枠組みをとるため、年次変化やカテゴリー間の差異について統計的有意性の検定を行っていない。今後は、Top500ランクイン確率の背後にある母集団分布を仮定できるモデルを構築し、CORやCDFRの変化を推測統計的に検証することが課題である。特に、「ランキングに入る確率分布」を明示的に扱うことで、記述と推測を接続する道が開けるだろう。

注.

- 1) ただし、ARWUが用いる指標の中身は経年的に変化していることに留意しないとイケない。例えば、HiCiやWeb of Scienceの収録範囲は、この20年間で中国発行のジャーナルを多く含むなど、大きく拡大しており、その影響はランキング結果にも反映されている。
- 2) それに加え、表1では、列の周辺度数が制度的に常に100校と固定されており、そして各国の大学数(行の周辺度数)もランキング結果とは無関係に与えられている。A国のランクイン数は超幾何分布 $\Pr(T=t) = \frac{\binom{N_1}{t} \binom{N-N_1}{100-t}}{\binom{N}{100}}$ に従う。

参 考 文 献

- Benito, M., Gil, P. and Romera, R. (2020). Evaluating the influence of country characteristics on the Higher Education System Rankings' progress, *Journal of Informetrics*, **14**(3), 101051, <https://doi.org/10.1016/j.joi.2020.101051>.
- El Gibari, S., Gómez, T. and Ruiz, F. (2018). Evaluating university performance using reference point based composite indicators, *Journal of Informetrics*, **12**(4), 1235–1250, <https://doi.org/10.1016/j.joi.2018.10.003>.

- Günay, A. (2022). Analyzing the correlation between university rankings and the economic performance of countries, *Journal of University Research*, **5**(2), 205–211.
- Lu, C. (2014). University rankings game and its relation to GDP per capita and GDP growth, *International Journal for Innovation Education and Research*, **2**(4), 1–33, <https://doi.org/10.31686/ijer.vol2.iss4.166>.
- Marginson, S. (2007). Global university rankings: Implications in general and for Australia, *Journal of Higher Education Policy and Management*, **29**(2), 131–142.
- Pietrucha, J. (2018). Country-specific determinants of world university rankings, *Scientometrics*, **114**(3), 1129–1139, <https://doi.org/10.1007/s11192-017-2634-1>.
- Rauhvargers, A. (2011). Global university rankings and their impact, European University Association, Brussels.
- Rauhvargers, A. (2013). Global university rankings and their impact: Report II, European University Association, Brussels.
- Williams, R., de Rassenfosse, G., Jensen, P. and Marginson, S. (2013). The determinants of quality national higher education systems, *Journal of Higher Education Policy and Management*, **35**(6), 599–611. <https://doi.org/10.1080/1360080X.2013.854288>.

A Method for Analyzing Hierarchical Ranking Data Using Common Odds Ratio and Cumulative Distribution Function of Rank: Application to World Universities Rankings

Yizhou Fan

Research Institute for Higher Education, Hiroshima University

This paper proposes a statistical framework for analyzing world university rankings by formalizing the notion of the “global average benchmark.” Using the ARWU Top500 data over 14 years, we develop two main tools: (1) the Common Odds Ratio (COR), derived from a series of 2×2 contingency tables; and (2) the Cumulative Distribution Function of Rank (CDFR), a nonparametric measure representing the proportion of ranked universities in each country as a function of rank.

We further introduce a unimodality assumption for the rank probability function to classify countries into structural types—such as “highly advanced,” “mid-range advanced,” or “emerging”—based on the shape of their CDFR curves. Empirical analysis reveals that: China shows a convex CDFR with a rising COR, representing an emerging structure; the U.S. maintains a concave CDFR and a high but slowly declining COR, suggesting a dominant yet gradually waning position; and the U.K. displays an S-shaped CDFR, indicating mid-range strength in global rankings.

While the analysis is limited by the lack of information on unranked institutions and the assumption of a fixed Top500 population, the framework offers a theoretically coherent and visually intuitive method for international comparison of university presence in global rankings.

「統計数理」投稿規程

1. 「統計数理」は、統計科学の深化と発展、そして統計科学を通じた社会への貢献を目指すものである。投稿原稿は、統計科学に関連した内容を持つもので、和文の原稿に限る。
2. 投稿原稿は次の6種とする。
 - a. 原著論文 (Paper)
統計科学の発展に貢献すると考えられる研究結果。
 - b. 総合報告 (Review Article)
特定の主題に関する一連の研究およびその周辺領域の発展を著者の見解に従って総合的、かつ体系的に報告したもの。
 - c. 研究ノート (Letter)
研究速報、新しい発想、提言、問題提起、事例報告など研究上、記録にとどめておく価値があると認められるものや、既発表の論文等に対するコメントで、研究上、記録にとどめておく価値があると認められるもの。
 - d. 研究詳解 (Research Review)
特定の研究領域における理論的あるいは応用的成果を、最近の結果や知見を加えてわかりやすく説明したもの。
 - e. 統計ソフトウェア (Statistical Software)
有用な計算法や解析法に関する短いプログラムおよびサブルーチンのリスト、利用手引き、実行例など。
 - f. 研究資料 (Research Archives)
歴史的なデータ、入手困難なデータや統計的手法の比較検討のために有用なデータ、あるいは、歴史的文献の翻訳や解説など。
いずれも原則として、未発表のものに限る。
3. 投稿された原稿は、編集委員会が選定・依頼した査読者の審査を経て、掲載の可否を決定する。
4. 投稿原稿は電子投稿査読システム <https://www.editorialmanager.com/toukei/> より投稿するものとする。原稿は pdf ファイルとし、必要なフォントはすべて埋め込み、原稿全体を一つのファイルにまとめることとする。論文が採択になった場合、著者は最終稿のソースファイルとハードコピーを提出するものとする。
5. 著作権
 - (1) 掲載される論文等の著作権はその採択をもって統計数理研究所に帰属するものとする。統計数理研究所は、紙媒体の「統計数理」のほか電子媒体などを通じて論文等を公表することができる。特別な事情がある場合は、著者と本編集委員会との間で協議の上措置する。
 - (2) 投稿原稿の中で引用する文章や図表の著作権に関する問題は、著者の責任において処理する。
 - (3) 著者が自分の論文等を複製、転載、翻訳、翻案等の形で利用するのは自由である。この場合、著者は掲載先に出典を明記する。
6. 原稿は次の執筆要項に従って作成する。

「統計数理」執筆要項

1. 原稿は A4 用紙に 1 行 36 字から 40 字で 1 行おき、1 頁あたり 22 行程度とする。原稿の長さは原則として表・図を含めて 30 頁相当以内とし、各ページにページ番号を付す。図表は別紙にまとめ、本文中には挿入箇所のみを指定する。LaTeX で原稿を作成する場合は、「統計数理」スタイルファイルの使用を推奨する。
<https://www.ism.ac.jp/editsec/toukei/>
2. 文体は「である体」とし、句読点は「,」「.」を用いる。
3. 原稿は以下の順に書くものとする。

[第 1 頁] 標題、著者名、所属名および所在地、メールアドレス、和文要旨 (500 字程度、文献の引用および数式は原則として避ける)、和文キーワード (6 語以内)。

[第 2 頁] 英語による標題、著者名、所属名、Abstract (450 ワード程度)、Key words (6 words and phrases 以内)。Abstract は、問題の所在と得られた結果等がそれだけで理解できるようなものとする。

[第3頁以降]

- ① 本文：章、節の番号は、第1章にあたるものは、“1.”、第1章第1節にあたるものは、“1.1”というようにつける。また、式の番号は、章ごとに(2.1), (2.2)のようにし、式の左側に配置する。
 - ② 数式：数式は簡明さを心がけ、添字にさらに添字をつけるのはなるべく避ける。
 - ③ 参考文献：書き方は本要項 第5項を参照。
 - ④ 表：一枚の用紙に一つの表を書く。表の番号は論文中に現れる順に従って、表1, 表2,... または, Table 1, Table 2,... のようにする。
 - ⑤ 図：一枚の用紙に一つの図を描く。図はそのまま写真製版できる鮮明なものを用意する。大きさは印刷出来上りの1~2倍とし、トレースが必要な場合は原則として著者が行うものとする。図の番号は論文中に現れる順に従って、図1, 図2,... または, Fig. 1, Fig. 2,... のようにする。
 - ⑥ 注：本文中の注釈は極力避ける。やむを得ず注釈をつける場合は脚注とせず、論文末尾に後注とする。後注は、順番に“1, 2,...”の番号を付け、本文中では上付きで示す。
4. 本文中での参考文献の引用は、著者名(出版年)とする。たとえば, Efron (1982), 清水・湯浅(1984), Cox and Snell (1981), 坂元 他(2004), Nakano et al. (2000)。
 5. 参考文献の書き方

[雑誌の場合] いずれの場合も雑誌名は省略しないものとする。

① ページ番号がある文献

著者名(出版年). 標題, 雑誌名, 巻, ページ [始-終].

【例】 Chernoff, H. (1973). The use of faces to represent points in k -dimensional space graphically, *Journal of the American Statistical Association*, **68**, 361-368.

【例】 Bligh, E. G. and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification, *Canadian Journal of Biochemistry and Physiology*, **37**, 911-917, <https://doi.org/10.1139/o59-099>.

② ページ番号がない文献

著者名(出版年). 標題, 雑誌名, 巻に続けて, DOIがあるものはDOIを入れ, DOIがない場合はURLと最終アクセス日を記載する。

【例】 Lien, G. Y., Kalnay, E. and Miyoshi, T. (2013). Effective assimilation of global precipitation: Simulation experiments, *Tellus A*, **65**, <https://doi.org/10.1175/WAF-D-13-00032.1>.

【例】 Kiraly, F. J. and Oberhauser, H. (2019). Kernels for sequentially ordered data, *Journal of Machine Learning Research*, **20**(31), <http://jmlr.org/papers/v20/16-314.html> (最終アクセス日 2023年10月1日)。

[叢書の中の一巻の場合]

著者名(出版年). 書名 (編集者名), 叢書名, 発行所名, 発行地名。

【例】 Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1983). *Akaike Information Criterion Statistics, Mathematics and Its Applications*, Reidel, Dordrecht.

[単行本等の場合]

著者名(出版年). 書名, 発行所名, 発行地名。

【例】 Cressie, Noel (1993). *Statistics for Spatial Data*, Wiley, New York.

[編集書の中の一部の場合]

著者名(出版年). 標題, 編集書名 (編集者名), 巻, ページ, 発行所名, 発行地名。

【例】 Akaike, H. (1980). Likelihood and the Bayes procedure, *Bayesian Statistics* (eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), 143-166, University Press, Valencia, Spain.

なお、同じ著者によるものが同一年に複数個現れる場合には、(1980a), (1980b) などとして区別する。文献は、日本人も含め、著者名のアルファベット順に並べる。

6. 著者校正は原則として一回とする。その際、印刷上の誤り以外の字句や図版の訂正、挿入、削除等は原則として認めない。