

令和 2 年 6 月 1 日現在

機関番号：34315

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K04103

研究課題名(和文) ランダム行列理論モデルの社会調査データへの応用

研究課題名(英文) Application of Random Matrix Theory to social survey data analysis

研究代表者

中井 美樹 (Nakai, Miki)

立命館大学・産業社会学部・教授

研究者番号：00241282

交付決定額(研究期間全体)：(直接経費) 2,500,000円

研究成果の概要(和文)：本研究の成果は、行列理論とそれに基づく計算手法を社会調査データ解析に応用するための新たな手法を提案し、応用を進めたことにある。研究期間には第一に、欠損値を含むデータセットを扱う際のリストワイズ削除法の最適な変数・ケース選択手法を改良した。第二に、変数のウェイト要因を考慮に入れた、分析対象とすべき変数・ケースの選択アルゴリズムを開発し、社会調査データに応用するためコードをR言語により作成した。

研究成果の学術的意義や社会的意義

近年の調査データの統計解析技法の顕著な発展に比して、欠損・欠測を含む社会調査データへの対処手法はあまり注意が払われてこなかったが、その点を考慮しないままだと分析結果は誤ったものになる可能性がある。本研究の成果はそうした起こりうる推定バイアスを回避するための技法上の洗練という点で重要な学術的意義を持つ。また、学際的・国際的な共同研究を進め議論を深めることを通じて、日本のジェンダー不平等や家族にかんするインプリケーションを再考しながら世帯状況を明らかにしたことは社会的意義である。

研究成果の概要(英文)：During the past three years, the project focused on applying matrix theory methods and computational methods to the optimization of the so-called Missing Data problem. The main results are:

- 1) The development of a new algorithm to improve and optimize the listwise deletion method. The work is based on identifying what respondents and variables should be deleted to obtain a fully complete dataset without missing data. The optimization technique is semi-analytical and detects the minimum number of deleted respondents and variables. The algorithm has been implemented on the popular computational platform Matlab/Octave.
- 2) It is important to consider the relative importance of different variables in the analysis before applying complete case analysis. We generalized the algorithm above to include weight factors that quantify what variables cannot be absolutely deleted. The result is a second algorithm that is based on stochastic Monte Carlo optimization. It has been developed in R language.

研究分野：社会学

キーワード：ランダム行列理論 社会調査データ 欠損値を含むデータ カテゴリカル・データ

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

社会学研究において、社会調査データの分析に基づく実証研究は新たな計量モデルの開発や応用とともに進展を見てきた。そこではしばしば、社会学に隣接する社会科学分野だけではなく、医学や工学、生物学といった自然科学の諸領域で有用とされる分析モデルが、主に人を対象とする社会調査データの分析にも様々な利用可能性を持つことが理解され応用されてきた。その一方で、近年、経済学や金融工学などの分野で応用研究が広がっているランダム行列理論は、これまで社会学領域で確率標本抽出にもとづく社会調査データの分析には適用されてこなかった。研究代表者は、科学研究費基盤(C)「社会学的データ分析へのランダム行列理論の応用にかんする研究」(平成26年度～28年度)において社会調査データの相関係数行列の分析へのランダム行列モデルの応用を進めてきた。本課題はその延長線上に位置するものであり、これまで継続してきた研究における残された課題を改善し、技法の公開までをめざす発展的研究である。

ランダム行列理論を応用する新たな手法を応用することにより、社会学における社会調査データ分析に伴う課題を克服し、より精緻な分析モデルに基づく社会調査データ解析からの確で深い社会学的インプリケーションを得ることを目指した。社会調査データ分析に伴う課題とは、具体的には、第一に回収率低下・無回答などといった諸点である。近年の調査無回答の増加傾向を考慮すると欠損値を含むデータ分析手法を慎重に適用しなければモデルの推定値にバイアスを生じうる。第二に、人を対象とした社会調査データは回答に様々な曖昧さや誤差をノイズとして含んでいる。第三に、社会調査データは扱う変数の多くがカテゴリカル変数(離散変数)からなり、データセットは連続変数とカテゴリカル変数の両方が混在したものであるため、異なる種類の変数を適切に分析する必要がある。第四に、社会科学で用いられている概念は直接に観測・測定することが困難なものが多い。こうした諸課題について対処できる分析手法が求められる。こうした多くの課題に応えるため、すでに構築してきた研究組織を基盤として研究を存続させ、社会学的方法論を進化させる必要があると考えるに至った。

2. 研究の目的

近年は社会調査を実施する環境の変化によってサンプルサイズの大きな社会調査を実施する困難さが増し、同時に回答率低下やアイテム非回答などを慎重に考慮して社会調査データを分析する必要性が増している。また、人を対象とした社会調査データは回答に様々な曖昧さや誤差をノイズとして含んでいる。本研究の目的は、確率変数間の相関・共分散のノイズを考慮しランダム行列理論を適用して、確率標本抽出・無作為抽出に基づく社会調査データ分析に伴う諸課題に対処する手法を提案・改善することである。

具体的には以下の3つの課題に取り組む。第一に、ランダム行列理論を応用し、既に考案した共分散行列推定のための手法にともなう課題を改善・改良することである。第二に、新たな社会調査データセットの分析への新たな手法の応用である。改善したモデルを最新の社会調査データセットの分析に適用し、社会学的理論の構築と深化を目指す。第三に、提案する分析を行うためのプログラム・Rパッケージの開発である。本研究で提案するモデルの応用によりデータ分析が可能になるよう手法の公開を行う。

3. 研究の方法

本研究は、統計学や統計物理の分野で発展したランダム行列理論を社会学的方法論に応用するという学際的アプローチをとる。分野横断的な共同研究体制を構築することにより研究を遂行し、社会調査データ解析手法の新たな提案を行う。

過去3年間に開発した手法を新しい社会学的データセットの分析に適用し、2.で述べた研究目的を効率的に遂行するため、主として2つの研究拠点を軸に部門を編成している。具体的には、第一に「A.社会学研究法における幾何学的枠組およびランダム行列理論を応用した手法の改善・改良、および提案モデルの新しい社会調査データへの応用研究」、第二に「B.Rパッケージ開発を通じた手法の公開」からなる。また研究成果を発表し研究を深めるため、研究会を年に数回開催し集中的な議論を行い研究計画の効率的な遂行を行った。全体を通じて研究代表者が統括をつとめ、研究協力者との有益な連携を通じて全体テーマを学際的・総合的に推進した。

本研究ではすでに収集された社会調査データや、二次データに対してランダム行列モデルを応用し、ランダム行列理論の社会学的研究における実証的有效性を提示することとしていた。活用したデータセットは、日本での代表的な社会調査データ「社会階層と社会移動全国調査(SSM調査)」データや「階層と社会意識全国調査(SSP調査)」プロジェクトによる最新のデータSSP-I2015やSSP-W2018などである。

4. 研究成果

(1) 研究会の開催

本研究は、研究代表者が責任を持って研究を遂行することとあわせて、社会学分野における研究での社会調査データ分析にこれまで応用されることがなかったランダム行列理論を応用するため、社会科学領域でのランダム行列モデルの応用研究に精通した海外の研究者の協力のもと、研究成果を蓄積しながら学際的研究を推進してきた。さらに、反復社会調査データを用い

た分析を進める中で、欠損データ由来の推定バイアスを回避するための手法の応用を、統計モデル開発に精通した海外の研究者の協力のもと、共同研究を推進した。研究会を開催し最新の研究情報を得るとともに研究交流と今後の継続的な研究体制の構築を図った。研究会の開催を通じて以下のような課題を整理・検討した。

[2017年度]

- 第1回研究会（2017年9月）社会調査データ分析におけるリストワイズ法の課題の整理
- 第2回研究会（2018年2月）社会調査データ分析におけるリストワイズ法のモデルの改善
- 第3回研究会（2018年3月）反復社会調査データの解析手法における課題の整理と国際学会 ASMOD2018 大会にむけた打ち合わせ

[2018年度]

- 第1回研究会（2018年10月）国際学会 ASMOD2018 大会の発表打合せ
- 第2回研究会（2019年3月）反復社会調査データの解析に関する研究成果のとりまとめと公表に向けた準備

[2019年度]

- 第1回研究会（2019年8月）国際学会 IFCS2019 大会の発表打ち合わせと公表に向けた準備

2017年のCLADAG、2018年のASMOD、2019年のIFCSなどにおける国際学会での研究発表とそれに基づく議論を受け、さらに継続的に本研究を発展・深化させる必要性が研究代表者および研究協力者の間で認識されるに至った。

(2) 国際学会大会への参加と成果発表

本研究期間を通じて達成した研究成果は多く、主要な成果は学会報告の形や論文などにより発表した。以下は、本研究期間を通じて達成した研究成果のうち主なものである。

研究成果の報告として、2017年9月13～15日にミラノビコッカ大学で開催されたThe 11th Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society (CLADAG2017)において研究発表を行った。本研究での提案手法であるランダム行列理論にもとづくデータ分析手法を、幾何学的手法を欠損値処理への対処にも応用・発展させ、「階層と社会意識全国調査 (SSP2015)」の分析に応用した研究成果の発表を行った。社会調査データにはほぼ必然的に発生しうる欠損値を含むデータセットへの対処法の一つとしてよく知られるリストワイズ除去法を適用する場合の最適化手法を開発した。回答者と変数について最適となるデータセットを選択する新規のアルゴリズムを提案した報告に関連して多くの研究者と議論を深め、提案したパッケージへのフィードバックを得る機会を持った。国際学会において様々な専門領域の研究者と慎重に議論を行う中で新たな課題に気づき、さらに研究を深化させることにつながった点もある。欠損値などの不完全データの問題に対処する際には、変数選択を主観的基準や客観的基準に基づいて進めることが一般的だが、こうした判断をいかにモデル化するか、などさらに継続的に本研究を発展・深化させる必要性が研究代表者および研究協力者の間で認識されるに至った。この成果、すなわち拡張的な新たな手法の定式化と社会調査データ分析への応用については、英文論文としてまとめ公表した。

2017年CLADAG2017において、繰り返しのクロスセクション調査を用いた社会学的研究成果を発表した。ジェンダー視点から日本の世帯の特徴とその時代的变化を明らかにし、手法の応用領域および欠損データの対処手法についてなど多様なフィードバックを得た。さらに本研究の意義と課題が浮かび上がり、研究を深化させることに繋がった。具体的には、データの欠測のために生じる可能性のあるバイアスについてサンプリングウェイトを考慮した分析へと解析手法の再検討を促すことに結びついている。

欠測を考慮した潜在構造を分析する計量社会学的モデルに関する成果を、2018年10月24～26日にフェデリコ二世ナポリ大学において開催された国際学会 2nd International Conference on Advances in Statistical Modelling of Ordinal Data (ASMOD2018)において報告を行った。これは、アイテム非回答とユニット非回答の両方の欠測を考慮した手法の提案とその応用となっている。この研究報告についても英語の論文としてまとめ、公表した。

ミラノ・ビコッカ大学のPennoni教授と共同研究により進めてきた研究成果を2019年8月26～29日にギリシャ・テッサロニキにおいて開催されたThe 16th Conference of the International Federation of Classification Societies (IFCS2019)において発表した。この研究は、欠測によって生じるサンプルの偏りの補正手法を用いた推定方法の社会学的データへの応用である。ランダムサンプリングに基づく社会調査において生じるユニット非回答による欠測に起因するバイアスの補正手法を、「社会階層と社会移動全国調査」データおよび「階層と社会意識研究プロジェクト」調査データに応用した。分析結果より、世帯タイプの経年変化について明らかにした。学際的・国際的に研究遂行し議論する中で、互いに文化的背景が異なる研究者とデータ分析に伴う課題や社会における意識構造の解釈・理解や家族・ジェンダーに関連するインプリケーションを深く議論することが可能となり、データ分析結果の持つジェンダー不平等や家族・世帯に関する社会学的インプリケーションを再考する機会を得たことは、今後の研究の進化に結びつくという意味からも大きな意義があった。社会学および統計学からのアプローチにより達成された研究成果はまた、英語による論文として公表した。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 3件/うちオープンアクセス 1件）

1. 著者名 Fulvia Pennoni and Miki Nakai	4. 巻 なし
2. 論文標題 A latent variable model for a derived ordinal response accounting for sampling weights, missing values and covariates	5. 発行年 2018年
3. 雑誌名 Proceedings of the International Conference on Advances in Statistical Modelling of Ordinal Data	6. 最初と最後の頁 155-162
掲載論文のDOI（デジタルオブジェクト識別子） 10.6093/978-88-6887-042-3	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Miki Nakai	4. 巻 なし
2. 論文標題 Changes in couples' bread-winning patterns and wife's economic role in Japan from 1985 to 2015	5. 発行年 2019年
3. 雑誌名 Statistical Learning of Complex Data	6. 最初と最後の頁 133-141
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1007/978-3-030-21140-0_14	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Graziano Vernizzi and Miki Nakai	4. 巻 なし
2. 論文標題 Weighted optimization with thresholding for complete-case analysis	5. 発行年 2019年
3. 雑誌名 Statistical Learning of Complex Data	6. 最初と最後の頁 143-151
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1007/978-3-030-21140-0_15	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Nakai Miki	4. 巻 なし
2. 論文標題 Social Differentiation of Cultural Taste and Practice in Contemporary Japan: Nonhierarchical Asymmetric Cluster Analysis	5. 発行年 2017年
3. 雑誌名 Data Science: Innovative Developments in Data Analysis and Clustering	6. 最初と最後の頁 149-159
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-319-55723-6	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Fulvia Pennoni and Miki Nakai	4. 巻 7(1)
2. 論文標題 A latent class analysis towards stability and changes in breadwinning patterns among coupled households	5. 発行年 2019年
3. 雑誌名 Dependence Modeling	6. 最初と最後の頁 234-246
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1515/demo-2019-0012	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 該当する

[学会発表] 計5件(うち招待講演 0件/うち国際学会 5件)

1. 発表者名 Fulvia Pennoni and Miki Nakai
2. 発表標題 A latent variable model for a derived ordinal response accounting for sampling weights, missing values and covariates
3. 学会等名 International Conference on Advances in Statistical Modelling of Ordinal Data (ASMOD2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Vernizzi Graziano, Nakai Miki
2. 発表標題 Optimization of the Listwise Deletion Method
3. 学会等名 CLADAG 2017 (国際学会)
4. 発表年 2017年

1. 発表者名 Nakai Miki
2. 発表標題 Changes in couples' breadwinning patterns and wife's economic role in Japan
3. 学会等名 CLADAG 2017 (国際学会)
4. 発表年 2017年

1. 発表者名 Nakai Miki
2. 発表標題 Changes in the Gendered Division of Labor and Women's Economic Contributions within Japanese Couples
3. 学会等名 IFCS 2017 (国際学会)
4. 発表年 2017年

1. 発表者名 Fulvia Pennoni and Miki Nakai
2. 発表標題 Assessment of recent social attitudes in Japan: a latent class item response theory model for web survey data
3. 学会等名 16th Conference of the International Federation of Classification Societies (IFCS 2019) (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	Vernizzi Graziano (Vernizzi Graziano)	Siena College・Department of Physics and Astronomy・Professor	
研究協力者	Pennoni Fulvia (Pennoni Fulvia)	University of Milano-Bicocca・Department of Statistics and Quantitative Methods・Associate Professor	