

令和 3 年 6 月 18 日現在

機関番号：32629

研究種目：基盤研究(C)（一般）

研究期間：2016～2020

課題番号：16K00056

研究課題名（和文）統計分析手法の自動導出に関する理論構築と技術開発

研究課題名（英文）Theoretical and applied research on automatic derivation of statistical analysis methods

研究代表者

田中 研太郎（TANAKA, Kentaro）

成蹊大学・経営学部・教授

研究者番号：00376948

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究では統計分析手法の自動導出技術の開発を行いました。その大きな成果の一つとして、実験計画法における計画行列と呼ばれるものを自動的に作成する方法の構築とその動作に対する理論的な補助を行ったことが挙げられます。また、これに派生する研究成果として、3水準の場合のロバストなパラメータ設計において、カンファレンス行列というものを利用したときの最適な推定量の導出をおこないました。さらに、その推定量の精度の保証などについて、理論的な結果を与えることに成功しました。

研究成果の学術的意義や社会的意義

情報技術の発達とともに、今後、多種多様な大量のデータが独自の困難性を伴って得られるようになっていくと思われれます。つまり、データ量の増加だけではなく、その形式や種類の細分化が指数的速度に進むことが予測されます。そういった状況において、それぞれの分野でそれぞれのデータ形式に特化した分析のエキスパートを育成していくことは、限られた人的資源の中では不可能に近いと考えられます。本研究は統計分析手法の自動導出を目的としており、その成果を応用することで、エキスパートでなければ扱えなかったようなデータに対しても、誰でも手軽に分析ができるようになり、非常に大きな意義があると考えられます。

研究成果の概要（英文）：In this research, we developed a technique to derive statistical analysis methods automatically. As one of the major results, we obtained a method to automatically create a design matrix in the field of design of experiments. We also clarified the theoretical properties of the method. Furthermore, as an application of this research, we derived the optimal estimator for robust parameter design in the three-level case using a conference matrix.

研究分野：統計学

キーワード：実験計画法 計算代数統計

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

様々な分野において自動化の波が押し寄せています。統計学の分野においても、情報技術の発達とともに、ブートストラップや MCMC などのコンピューターの利用を前提とした分析方法が開発されてきましたが、近年ではさらに、統計分析の方法論自体を構築するためにコンピューターが使用されるようになってきています。情報技術の発達はさらに加速度を増している状況で、今後も多種多様な大量のデータが独自の困難性を伴って得られるようになっていくと思われれます。この先も、データの量が著しく増加していくことが想定されますが、それだけではなく、データの形式や種類についても指数的な速度で細分化されていくといったことが予測されます。そういった状況において、それぞれの分野でそれぞれのデータ形式に特化した分析のエキスパートを育成していくことは、限られた人的資源の中では不可能に近いと考えられます。統計分析の手法自体を自動的に導出できるようになれば、そのようなエキスパートでなければ扱えなかったようなデータに対しても、誰でも手軽に分析ができるようになります。本研究は、統計分析の手法の自動導出の理論と応用に貢献するもので、データ分析技術の発展において非常に大きな意義があると考えられます。

2. 研究の目的

本研究の目的は、広範囲に適用可能な統計分析手法の自動導出技術の開発です。本研究では推定量の「最小単位」を用いることで統計分析の手法自体を自動的に導出する手法を構築することを考えました。推定量が「最小単位」の組み合わせで得られる場合には、その導出の自動化が容易になり、さらに、自動定理証明などによる推定量の理論的な性質の保証も容易になると考えられます。これらの自動導出技術の開発により、統計学の理論と応用の両方に貢献することを目標としました。本研究においては、「最小単位」を用いた方法論が、統計分析手法の導出を自動化するうえで、どこまで有効に作用するのかを明らかにすることを考えました。さらに、本研究では、統計学において比較的新しい手法である計算代数統計の手法を使用しました。統計学の基本的な部分を、代数学の言葉で翻訳することにより、新たな発見を得ることも目的としました。この研究は、統計学と代数学のさらなる融合を促し、両分野の理論と応用の発展に大きく寄与するものだと考えられます。

3. 研究の方法

本研究においては、広範囲に適用可能な統計分析手法の自動導出技術の開発を行いました。従来、推定量の構成方法は状況に応じてそれぞれに構築されることが多かったのですが、本研究では推定量の「最小単位」を用いることで統一的な手法を構築することを目指しました。ここで、「最小単位」とは、例えば、単回帰分析の場合であれば「2つのデータ点より得られる切片と傾き」に相当します。より一般の場合には、計算代数におけるグレブナー基底の理論を用いて「最小単位」を構成できることができたりします。まず、一番簡単な、回帰分析における回帰係数の線形不偏推定量の「最小単位」の構築からスタートし、その適用範囲を広げつつ、さらに、理論的性質の自動導出へと繋げる形で、理論と応用のバランスをとりながら有効性を検証していきました。そして、統計的品質管理のいくつかの手法の中でも極めて重要なツールである実験計画法において、計画行列(実験計画)と呼ばれるものを自動的に作成する方法へとその理論を応用していきました。

4. 研究成果

本研究では統計分析手法の自動導出技術の開発を行いました。その大きな成果の1つとして、実験計画法における計画行列と呼ばれるものを自動的に作成する方法の構築とその動作に対する理論的な補助を行ったことが挙げられます。製造業やサービス産業などの分野においては、非常に多くの場面で統計的品質管理の手法が利用されています。それらの統計的品質管理の手法の中でも特に重要なツールである実験計画法において、計画行列(実験計画)と呼ばれるものを自動的に作成する方法を本研究では新たに考案し、数値実験な

どを通してその有用性も明らかにしています。

実験計画法の目的の一つとして、少ない実験回数でなるべく多くの因子の効果の大きさを精度よく推定したいという目的が挙げられますが、そのような実験計画的に良い実験方法を求める問題が、機械学習の分野で精力的に研究されている、スパースモデリングの枠組みとして捉えられることを示しています。

さらには、2次錘計画問題として定式化できることなどを示しています。

これらの定式化が行われた問題は、数理計画問題ではよく扱われるもので、それらを効率よく解くための方法やソフトウェアなどが多数存在しています。

そのようなソフトウェアを活用して最適化問題を解くことで、自動的に良い実験計画を生成することが可能であることを示しています。

実際に、ある実験条件を設定して、良い実験計画を自動生成させると、直交表と呼ばれる非常に良い性質を持つ実験計画が得られたりするなど、既存の理論的な結果とも合致する結果が得られるなど、有用性が高いことが分かりました。

また、既存の理論では実験計画を構成しづらい実験条件の設定、つまり、人力では良い実験計画を作成することが困難な設定においても、最適な実験計画を自動生成できることも明らかにできました。

一方で、提案している実験計画を自動生成する手法においては、事前いくつかのパラメーター(実験条件)を設定しておく必要があるのですが、実は、定式化後の問題が対称性を持つような形でこれらのパラメーターを設定してしまうと、少ない実験回数を実現してくれるような実験計画を自動生成できないことが明らかになりました。

対称性と統計的推測の間には、昔から様々な厄介な問題があることが知られていまして、今回の新手法においてもその問題の一端が表れていることは、数理統計学の理論において非常に興味深いことだと考えられます。

そのような問題点を回避するために、今回の研究で注目している手法の1つである推定量の「最小単位」が役立つこと、そして、その役立つ方の指針を示すことができました。

さらに、この研究を他の先行研究と対比させることで興味深い関係性を得ることもできました。これにより、実験計画(計画行列)の見方として、ある種の双対的な視点の違いによって、実験計画法に対する計算代数の応用方法にも2つの違った手法が見いだせるということも明らかにすることができました。

また、これらの研究成果の応用として、3水準の場合のロバストなパラメーター設計において、カンファレンス行列というものをういたときの最適な推定量の導出を行っています。

さらに、その推定量の精度の保証などについて、理論的な結果を与えることにも成功しました。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 森 輝雄、貞松 伊鶴、松浦 峻、田中 研太郎	4. 巻 49
2. 論文標題 カンファレンス行列と2水準ノイズを用いた直交計画によるパラメータ設計	5. 発行年 2019年
3. 雑誌名 品質	6. 最初と最後の頁 266 ~ 278
掲載論文のDOI (デジタルオブジェクト識別子) 10.20684/quality.49.3_266	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 Kentaro Tanaka, Masami Miyakawa	4. 巻 50
2. 論文標題 An Application of the Lasso in Experimental Design	5. 発行年 2019年
3. 雑誌名 成蹊大学経済学部論集	6. 最初と最後の頁 13-28
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
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 -

1. 著者名 田中 研太郎	4. 巻 2124
2. 論文標題 実験計画法における計画行列の生成について	5. 発行年 2019年
3. 雑誌名 数理解析研究所講究録	6. 最初と最後の頁 1-5
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
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 -

1. 著者名 田中 研太郎	4. 巻 46
2. 論文標題 カンファレンス行列と実験計画法	5. 発行年 2016年
3. 雑誌名 品質	6. 最初と最後の頁 51--54
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
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 -

〔学会発表〕 計6件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 田中研太郎
2. 発表標題 実験計画法における計画行列の生成について
3. 学会等名 最尤法とベイズ法(RIMS共同研究)
4. 発表年 2019年

1. 発表者名 田中 研太郎
2. 発表標題 実験計画法における最適化と数理
3. 学会等名 日本統計学会春季集会2017
4. 発表年 2017年

1. 発表者名 田中 研太郎
2. 発表標題 Data Visualization in Virtual Reality
3. 学会等名 統計関連学会連合大会2017
4. 発表年 2017年

1. 発表者名 Kentaro TANAKA
2. 発表標題 Symmetries in Experimental Design and Linear AEstimators
3. 学会等名 The 4th Insitute of Mathematical Statistics Asian Pacific Rim Meeting (国際学会)
4. 発表年 2016年

1. 発表者名 田中 研太郎
2. 発表標題 仮想空間における データとモデルの表現
3. 学会等名 統計関連学会連合大会2016
4. 発表年 2016年

1. 発表者名 田中 研太郎
2. 発表標題 実験計画法における最適化と数理
3. 学会等名 第11回日本統計学会春季集会
4. 発表年 2017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

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関