

令和 5 年 5 月 8 日現在

機関番号：12601

研究種目：若手研究

研究期間：2020～2022

課題番号：20K19752

研究課題名（和文）代数構造と幾何構造に基づく統計的推測と生存時間解析の研究

研究課題名（英文）A study on statistical inference and survival analysis based on algebraic and geometric structures

研究代表者

小川 光紀 (OGAWA, Mitsunori)

東京大学・大学院情報学環・学際情報学府・特任講師

研究者番号：50758290

交付決定額（研究期間全体）：（直接経費） 2,800,000円

研究成果の概要（和文）：局外パラメータを含む離散指数型分布族において、局外パラメータの十分統計量を条件付けた条件付き分布の族における複合局所Bregmanダイバージェンスに基づく構造パラメータの推定手法の開発に取り組んだ。特に、ある種のロバスト性が期待される推定方法を開発した。また、Firthのバイアス補正推定量の推定値の存在性について、ロジスティック回帰モデルの場合に理論的結果を得た。二項ロジスティック回帰の場合には推定値の存在が保証されることを数学的に示し、多項ロジスティックの場合には先行研究の議論が不成立である反例を見つけた。

研究成果の学術的意義や社会的意義

局外パラメータを含む統計モデルの推定手法の開発は統計学における古典的問題であり、特に計算負荷の軽減と統計的性質の両方をあわせもつ手法の開発には意義がある。本研究で用いた枠組みは近年の代数統計とダイバージェンスに基づく方法の両者の利点を活用したものであり、学術的にも意義あるものと考えられる。Firthのバイアス補正推定値の存在性に関する研究は、複数の応用領域において理論保証が不十分なまま普及している利用方法を検証するものであり、理論と応用の両面から意義あるものである。

研究成果の概要（英文）：We studied the problem of estimating the structural parameter of discrete exponential family models under the presence of nuisance parameters based on the composite local Bregman divergences on the conditional distributions obtained by conditioning sufficient statistics for the nuisance parameters. In particular, we developed estimation methods with a certain robustness property. We also studied the existence of estimates for Firth's bias-correction method in logistic regression models. We proved the existence of the estimates in the binomial logistic regression case, while we found a counterexample for the discussion of previous research in the multinomial logistic regression case.

研究分野：統計学

キーワード：統計的推測 ダイバージェンス バイアス補正 推定値の存在性

## 1. 研究開始当初の背景

本研究の背景の一つは、局外パラメータを含む統計モデルの推測問題である。統計モデルが局外パラメータを含み、観測ごとに局外パラメータの値が異なる場合の統計的推測は、Neyman-Scott 問題とよばれる古典的な問題である。1980年代から1990年代にかけて、Amari とその共著者らは情報幾何学の観点からこの問題に取り組み、推定方程式により適切な推定量が得られるための条件や、適切な推定方程式が存在する場合の最適な推定方程式の特徴付けを行った。これらの一連の研究により、局外パラメータに対する適切な十分統計量が存在するときには、その十分統計量を所与とした条件付き分布に基づく推定量が漸近的に最適であることが明らかになった。しかしながら、現実の問題設定では条件付き分布に基づく推測は計算コストの観点から実用的でない場合も多い。その大きな原因の一つは、条件付き分布の規格化定数の計算が困難なことにある。一方、1990年代から計算代数統計学とよばれる研究領域が大きく発展し、フィッシャーの正確検定のように、条件付き分布に基づく仮説検定の文脈において、マルコフ連鎖モンテカルロ法に基づいて規格化定数計算を回避する手法が研究されてきた。

次に、医療統計・疫学領域に関わる研究に関する本研究の背景について述べる。これらの領域では、打ち切りを伴うデータを扱う生存時間解析やデータ数の少ない小標本など、複雑な状況や有限標本特有の困難を伴う場合も多い。たとえば、生存時間をアウトカムとする代表的な回帰モデルとしてコックス回帰モデルがあるが、これは比例ハザード性とよばれる強い仮定を課したセミパラメトリックモデルであり、標準的な部分尤度の最大化に基づく推定量の漸近的性質はよく知られているものの、有限標本のもとでは推定値の存在性に問題を生じる状況がある。また、欠測を伴うデータを用いた解析など、適切な対処が求められる複雑な状況が様々存在する。

## 2. 研究の目的

本研究における当初の目的は、局外パラメータを含む離散指数型分布族における構造パラメータの推定問題に対して、統計的性質と計算の実施可能性の両面から有用な方法を発展させること、および、生存時間解析を含む医学・疫学領域に現れる難しい設定を克服する方法論の開発の二つである。前者については、研究代表者が本課題の当初から取り組んでいた、代数統計的方法とダイバージェンスに基づく推定方法の組合せで得られる推定手法を発展させることを目指した。後者については、特に小標本のもとで実用上の障害になりやすい推定値の存在性に関する問題を研究対象とした。当初は生存時間解析および欠測データ解析に関わる研究を中心に考えていたが、より基本的な設定においても先行研究における課題が残っていることが研究期間中に明らかになったことから、より基本的設定での理論的考察を主眼に据えるという研究目標の調整を行った。生存時間解析に関しては、近年その利用がひろがっている制限付き平均生存時間に焦点を当て、特に小標本の場合に起こりうる一般的な推測方法の不具合への対処について検討した。

## 3. 研究の方法

前半の研究課題では、本研究課題の前身となる研究代表者の研究で提案した枠組みを引き続き利用した。その枠組みは、Kanamori & Takenouchi (2017) によって整備された複合局所 Bregman ダイバージェンスの枠組みを、代数統計学において発展してきたマルコフ基底の理論と併用することで実現されるものである。この枠組みで得られる推定量は、条件付き分布の規格化定数を回避できるという特徴をもつ。また、既存の代数統計学の文脈では主に条件付き分布に基づく正確検定を実施する際の道具として用いられていたマルコフ基底に対し、推定に関する応用を与えている点でも特徴的である。この枠組みで得られる推定量を具体的に定めるためには、ダイバージェンスを構成するもとになるスコアを指定する必要がある。その選択によって得られる推定量に付加的な統計的性質を期待できることから、この方針を引き続き採用した。適用対象としては、マルコフ基底の構造について蓄積の多い分割表の統計モデルを中心に検討した。

後者の課題に関しては、Firth のバイアス補正法について理論的検討を主に行った。特に、ロジスティック回帰モデルの場合の推定値の存在性について、一連の先行研究における議論の直観的方針を引き継ぎつつ、論理的ギャップや証明の不備を補うという方針で、推定値の存在性に関する理論保証を試みた。また、大学院生の協力を得て、制限付き平均生存時間に基づく推測で最も基本的なカプラン・マイヤー曲線の積分に基づく方法について、特に小標本で生じうる推定値の決定に事前規定が必要となる事象の確率を考察した。既存研究で検討されているものを含め、いくつかの自然な対処法の統計的性能を数値的に調べた。

#### 4. 研究成果

局外パラメータを含む統計モデルの推定問題については、有限標本空間上の離散指数型分布族において、関心のある一部の構造パラメータの推定方法の構築に関する研究に取り組んだ。特に、モデルに付随するトーリックイデアルという代数構造の生成系であるマルコフ基底について結果の蓄積の多い、分割表の対数線形モデルを中心に検討を行った。複合局所 Bregman ダイバージェンスの枠組みで妥当な推定量を得るためには、局外パラメータの十分統計量を所与とした条件付き分布における標本空間に適切な連結性をもつグラフ構造を導入することが重要である。どのような連結性を要するかは、Bregman ダイバージェンスを定めるスコアの選択に応じて変わり得る。本研究では、特に外れ値に対するロバスト統計で用いられるべき密度スコアや擬球スコアを元に複合局所 Bregman ダイバージェンスを構成することで、局外パラメータを含む設定でもある種のロバスト性を維持できることを、数値実験によって経験的に確認した。

Firth のバイアス補正法を適用した際の推定値の存在性については、ロジスティック回帰モデルの場合に理論的結果が得られた。まず、本研究期間中に、先行研究における存在性に関する議論を精査し、把握できる限りの先行研究において存在性の理論保証としては不完全な議論にとどまっていることを確認した。そのうえで、先行研究の技術的不備を補う形で、二項ロジスティック回帰モデルにおける Firth の補正推定値がデザイン行列のランクに関する仮定のみを課したもつで必ず存在することを証明した。証明における議論によって、推定値の存在性が保証される基本原理が、罰則付き最尤法として定式化した際の罰則項が、パラメータが発散する際にある意味で一様に負の無限大に発散する点にあることがわかった。この議論は、先行研究において直観的になされた予想が二項ロジスティック回帰の場合に正しかったことを意味している。一方で、多項のロジスティック回帰モデルにおいても類似の直観的議論が先行研究によってなされていたが、罰則項が負の無限大に発散しない例が存在することを発見した。以上の結果は、応用分野で既にひろがっている Firth のバイアス補正法の利用目的を二項ロジスティック回帰の場合に支持するとともに、あくまでその性質はバイアス補正法の副産物として得られるものであり、存在性の理論保証についてはモデルの設定ごとに精査すべきであることを示唆している。医療統計や疫学の応用の観点からは、生存時間解析で用いられる統計モデルの場合など、さらなる理論研究が必要である。

最後に、制限付き平均生存時間については、通常のカプラン・マイヤー曲線の積分に基づく推測方法を採用した際、計算結果の一意性を維持するために恣意的な指定を要する事象の起こる確率を表した式を導出し、現実的な設定の下でその大きさを評価した。その結果、特に小標本の状況では、推測に必要な計算に事前規定すべき点を要する確率が無視できないほど大きいことを確認した。また、対処法の事前規定として、先行研究から自然と思われる複数の方法についてその統計的性能を評価し、並び替え検定を用いた方法で安定的な結果が得られることを確認した。ただし、本研究で数値的に確認した設定はやや限定的である点には留意する必要がある。

5. 主な発表論文等

〔雑誌論文〕 計2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 1件）

1. 著者名 Mitsunori Ogawa and Yui Tomo	4. 巻 arXiv:2304.07484
2. 論文標題 On the existence of Firth's modified estimates in logistic regression models	5. 発行年 2023年
3. 雑誌名 arXiv	6. 最初と最後の頁 1-7
掲載論文のDOI（デジタルオブジェクト識別子） 10.48550/arXiv.2304.07484	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 Takeuchi Yoshinori、Ogawa Mitsunori、Hagiwara Yasuhiro、Matsuyama Yutaka	4. 巻 30
2. 論文標題 Non-parametric approach for frequentist multiple imputation in survival analysis with missing covariates	5. 発行年 2021年
3. 雑誌名 Statistical Methods in Medical Research	6. 最初と最後の頁 1691 ~ 1707
掲載論文のDOI（デジタルオブジェクト識別子） 10.1177/09622802211011197	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 -

〔学会発表〕 計4件（うち招待講演 0件/うち国際学会 1件）

1. 発表者名 Mitsunori Ogawa
2. 発表標題 Estimating parameter of discrete exponential families using Markov bases under the presence of nuisance parameters
3. 学会等名 Algebraic Statistics 2022（国際学会）
4. 発表年 2022年

1. 発表者名 榎本駿平，野村尚吾，小川光紀
2. 発表標題 制限付き平均生存時間が推定不可能となる確率の定式化と対処策の検討
3. 学会等名 計量生物学会年会
4. 発表年 2022年

1. 発表者名 塘由惟, 小川光紀, 大庭幸治, 松山裕, 片井みゆき
2. 発表標題 ロジスティック回帰モデルにおける安定な非凸スパース正則化法
3. 学会等名 2022年度統計関連学会連合大会
4. 発表年 2022年

1. 発表者名 塘由惟, 小川光紀
2. 発表標題 ロジスティック回帰におけるFirthの罰則付き最尤法に基づく信頼区間構成法の改良
3. 学会等名 日本計算機統計学会第35回シンポジウム
4. 発表年 2021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