

令和 5 年 6 月 9 日現在

機関番号：32682

研究種目：若手研究

研究期間：2018～2022

課題番号：18K18008

研究課題名（和文）オッズ比に基づく統計的手続きの拡張と一般化

研究課題名（英文）Statistical procedures based on odds ratios

研究代表者

廣瀬 善大（Hirose, Yoshihiro）

明治大学・総合数理学部・専任准教授

研究者番号：10637391

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：オッズ比と呼ばれる量の統計学的な側面について調べた。オッズ比は、イベントの起こりやすさの指標であるオッズを考え、2つのオッズの比をとった量である。本課題ではたとえば、オッズ比についてベイズ統計におけるマッチング事前分布と呼ばれる確率分布を考察した。また、一般的なパラメータの扱いとして、オッズ比が自然パラメータと呼ばれるパラメータになることを利用して、統計学的なスパース推定の方法を提案し、数値実験によりその性能を評価した。

研究成果の学術的意義や社会的意義

オッズ比は統計学の理論においても実用においても重要な量である。また、ある問題設定でオッズ比を含む自然パラメータと呼ばれるパラメータも、理論と応用の両面で頻りに利用される。本課題では、統計学の問題として分割表や回帰、時系列などを想定し、これらのパラメータの推定を考察することにより、推定問題の性質の調査や、それを活かした推定方法を提案したという点で学術的意義がある。また、応用分野（物理化学）への推定の適用にも取り組んだ点にも学術的意義がある。

研究成果の概要（英文）：A statistical aspect of a quantity called the odds ratio was investigated. The odds ratio is a quantity obtained by taking the ratio of two odds. The odds is an index of the likelihood (probability) of an event occurring. In this study, for example, we considered probability distributions called matching priors in Bayesian statistics for odds ratios. We also proposed a statistical sparse estimation method using the fact that the odds ratio is a parameter called a natural parameter. The method's performance was evaluated through numerical experiments.

研究分野：統計科学

キーワード：数理統計学 スパース推定 情報幾何学 ベイズ統計

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

(1) 研究代表者は数理統計学の訓練を受けており、特に統計的パラメータ推定と情報幾何学に関して研究成果があった。また、分割表における推定において、パラメータや統計量のとり方に恣意性があることに疑問を抱いていた。さらに、ベイズ統計学に関して研究成果があり、特に事前分布に興味があった。これらのことが、本研究課題を着想し、重要性を見出す上で大きく貢献した。

(2) 学術的な背景としては、オッズ比は統計学の理論においても応用においても非常に重要な量であり、特に医学統計において頻繁に用いられている。また、医学統計などの場面においては、データを収集するための様々な制約があり、統計学的な推定に都合のよい方法が利用できるとは限らない。そのため、制約のあるデータ収集方法で得られたデータから、推定しやすい方法で収集されたデータのように推定を行うことができるとしたら、推定の性能をほとんど犠牲にせずにデータ収集の手間を減らすことができる。さらに、オッズ比は尤度の比として扱うこともでき、理論的な関心が大きい。これらの点に関連する比較的新しい成果としてベイズ統計の枠組みで得られた結果が知られていた。

(3) 上記2点が組み合わさることにより本研究課題の核心となる以下の問いに結びついた。

- オッズ比の一致と統計モデルの関係は？
- オッズ比の一致と推定法の関係は？
- オッズ比の一致のベイズ統計的な側面は？

2. 研究の目的

(1) 研究の目的は、主に上記の問いに答えることであった。特に問いをより具体化し、統計学で扱うような問題において、統計的パラメータ推定やベイズ統計、情報幾何学の観点から見た問題の解決を目指した。

(2) さらに、可能な範囲で、応用分野での利用に資するような方法論の提供、あるいはそれにつながる成果を得ることを目指した。

3. 研究の方法

(1) ひとつには理論的・数理的な考察を行った。統計的推定・情報幾何学・ベイズ統計などの観点から、フィッシャー情報量、事後分布、曲率のような重要な量を理論的に評価することで、問いに関係する点を調べた。

(2) また、理論的な考察を進めるまえに計算機による実験を行った。さらに、提案手法の性能評価においては計算機によるシミュレーション・数値実験を行った。

4. 研究成果

(1) オッズ比を考察する上で比較的取り組みやすい設定において、オッズ比に対してベイズ統計学におけるマッチング事前分布を導出した。分割表を想定することで、データ収集方法とマッチング事前分布との関連を調べた。具体的には、前向きにデータを収集する場合のオッズ比と、後向きにデータを収集する場合のオッズ比、さらに同時収集とでも呼ぶべき理想的なデータ収集の場合とを対象とした。

(2) 統計的パラメータ推定、特にスパース推定の観点から、特殊な正規分布を含むような設定において自然パラメータの推定手法を提案した (Härkönen, Sei & Hirose, 2020)。提案手法は、対象とする統計モデルのもつある量が特殊な微分方程式を満たすことを利用しており、さらに情報幾何学的な考察から得られたものである。提案手法は、一般の統計モデルでは計算が困難になるところを、微分方程式を利用した計算法で効率的な計算を実施し、統計モデルの幾何学的な性質を反映した自然パラメータのスパース推定を可能とした。

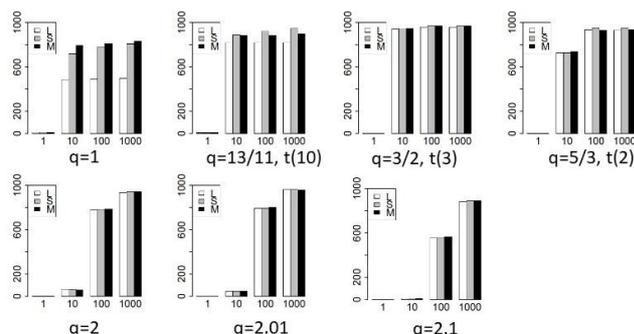
(3) 時系列モデルに対するパラメータ推定法を提案した。基本的なアイデアを(2)の推定法と共有しており、時系列モデルに対する情報幾何学的な考察から得られている。代表的な時系列モデルである自己回帰モデルと移動平均モデルは、どちらも情報幾何の観点からよい性質をもっていることが知られていた。そこで、その幾何学的な性質を反映したパラメータ推定手法を提案した。このモデルと推定法の組み合わせは、時系列におけるスパース構造について、よく扱われているものとは異なるモデリングを提供している。通常スパース構造では時系列モデルを回帰の形式で表現した際の回帰係数に相当するパラメータがスパースであると仮定する。しかし、本

提案では、自己共分散のスパース構造を利用しており、従来のスパース構造よりも細やかなスパース性を推定することができるようになった。

(4) L_q -尤度と呼ばれる量にもとづくパラメータの推定方法を提案した (Hirose, 2020)。 L_q -尤度は、統計学分野で重要な量である対数尤度を、対数関数の代わりにべき関数の族を用いることで定義されている。一般には L_q -尤度は必ずしも扱いやすい式にはならないが、 q -正規分布と呼ばれる確率分布族を仮定することで、 L_q -尤度が計算しやすい形となるような問題設定を導入した。ここで、 q は実数値をとる統計モデル・尤度のパラメータであるため、推定の文脈では問題の難しさを反映する量とみることができる。確率分布としては、裾の重さを調整している量である。提案手法の評価のため

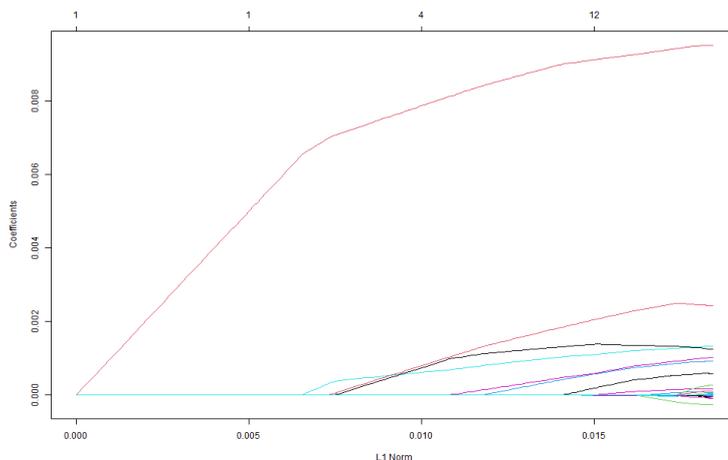
数値実験：結果 (変数選択)

• $d=10, n=100, r=0.4$



にシミュレーションにもとづく数値実験を行った。確率分布のパラメータ推定には通常の線形回帰におけるスパース推定法を形式的にそのまま利用することができ、提案手法の新たな実装は不要であった。推定性能は q の値に依存し、大きな値であるほど推定結果が悪くなる傾向があった (右図参照)。裾の重い確率分布に関する推定が難しく、ロバスト推定の観点からも予想される傾向となった。一方で、提案手法は、データを生成していると想定される統計モデルについて q の真の値を知る必要がないため、適用しやすいという利点がある。

(5) 物理化学における問題に対して、統計的推定の方法を適用した (Toda et al., 2022)。金属におけるある種の光の放出・吸収の現象について、分子において関連する軌道ペアを抽出する問題を考えた。分子全体でのエネルギーの出入りと、軌道ペアごとのエネルギーの出入りに関するデータを扱った。この問題は統計学における線形モデルとしてモデリングすることができることが分かったため、回帰係数のスパース推定が軌道ペアの選択に相当するような定式化を行った。スパース推定法として正則化を検討し、実際に発光・吸光に関するシミュレーションデータに対してスパース推定を行うことにより軌道ペアの選択を行った (右図参照)。



(6) なお、本研究課題の成果はパラメータ推定に関するものが多くなった。本研究課題により応募時やそれ以降に関心をもった疑問のすべてに答えを与えることができたわけではない。特に、ベイズ統計の観点からの研究については今後取り組むべき余地が大きく、本研究課題の後継の位置づけにある科研費課題において、引き続き検討を続けている。また、オッズ比や関連する量を推定するための統計モデル・推定法を提案・整備することで、応用分野におけるモデル・推定方法の選択肢を増やせるような成果につなげることを、今後の目標として目指している。

< 引用文献 >

Härkönen, Sei and Hirose (2020). Holonomic Extended Least Angle Regression. *Information Geometry*, Volume 3, 149-181.

Hirose (2020). Regularization Methods Based on the L_q -Likelihood for Linear Models with Heavy-Tailed Errors. *Entropy*, Volume 22, Article 1036.

Toda, Hirose, Kazuma, Kim, Taketsugu and Iwasa (2022). Excited States of Metal-Adsorbed Dimethyl Disulfide: A TDDFT Study with Cluster Model. *The Journal of Physical Chemistry*, Volume 126, 4191-4198.

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 1件/うちオープンアクセス 3件）

1. 著者名 Toda Keijiro, Hirose Yoshihiro, Kazuma Emiko, Kim Yousoo, Taketsugu Tetsuya, Iwasa Takeshi	4. 巻 126
2. 論文標題 Excited States of Metal-Adsorbed Dimethyl Disulfide: A TDDFT Study with Cluster Model	5. 発行年 2022年
3. 雑誌名 The Journal of Physical Chemistry A	6. 最初と最後の頁 4191-4198
掲載論文のDOI (デジタルオブジェクト識別子) 10.1021/acs.jpca.2c02354	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 -
1. 著者名 新沼 広大, 廣瀬 善大, 今井 英幸	4. 巻 63
2. 論文標題 ハンディキャップのある対戦に対する Bradley-Terry モデルの適用	5. 発行年 2022年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 704-712
掲載論文のDOI (デジタルオブジェクト識別子) 10.20729/00216283	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 Yoshihiro Hirose	4. 巻 22
2. 論文標題 Regularization Methods Based on the Lq-Likelihood for Linear Models with Heavy-Tailed Errors	5. 発行年 2020年
3. 雑誌名 Entropy	6. 最初と最後の頁 1-16
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/e22091036	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 -
1. 著者名 Harkonen Marc, Sei Tomonari, Hirose Yoshihiro	4. 巻 3
2. 論文標題 Holonomic extended least angle regression	5. 発行年 2020年
3. 雑誌名 Information Geometry	6. 最初と最後の頁 149-181
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41884-020-00035-1	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13件（うち招待講演 1件 / うち国際学会 4件）

1. 発表者名 廣瀬 善大
2. 発表標題 統計的推定の情報幾何
3. 学会等名 明治大学MIMS共同研究集会「幾何学・連続体力学・情報科学の交差領域の探索（ ）」
4. 発表年 2022年

1. 発表者名 廣瀬 善大
2. 発表標題 q-指数型分布族をもつ線形モデルに対するLq-尤度を利用したスパース推定
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 廣瀬 善大, 長谷川 凌大, 今井 英幸
2. 発表標題 移動平均モデルの情報幾何を利用したパラメータ推定
3. 学会等名 2021年度統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 Yoshihiro Hirose, Ryota Hasegawa, Hideyuki Imai
2. 発表標題 An Information-Geometric Method for Parameter Estimation on Moving-Average Models
3. 学会等名 EcoSta2021 (国際学会)
4. 発表年 2021年

1. 発表者名 廣瀬 善大
2. 発表標題 Lq-尤度を利用した正則化法
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 長谷川凌大, 廣瀬善大, 今井英幸
2. 発表標題 移動平均モデルに対する情報幾何学的パラメータ推定
3. 学会等名 情報処理北海道シンポジウム2020
4. 発表年 2020年

1. 発表者名 廣瀬 善大
2. 発表標題 統計モデルの接空間を使ったLARS法
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 Yoshihiro Hirose
2. 発表標題 Geometrical Approach to Data Science
3. 学会等名 2019 Winter International Symposium on Big-Data, Cybersecurity and IoT (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Yoshihiro Hirose
2. 発表標題 LARS-type Estimation Procedure for Generalized Linear Models
3. 学会等名 Data Science Statistics & Visualisation (DSSV2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Danilo Guimaraes Goncalves and Yoshihiro Hirose
2. 発表標題 A Geometric Estimation Method for Autoregressive Models
3. 学会等名 Data Science Statistics & Visualisation (DSSV2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Danilo Guimaraes Goncalves and Yoshihiro Hirose
2. 発表標題 An Information-Geometric Method for Parameter Estimation on Autoregressive Models
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 Danilo Guimaraes Goncalves, Yoshihiro Hirose, Hideyuki Imai
2. 発表標題 A Method for Information-Geometric Estimation on Autoregressive Models
3. 学会等名 情報処理北海道シンポジウム2019
4. 発表年 2019年

1. 発表者名 廣瀬 善大
2. 発表標題 オッズ比のマッチング事前分布の比較
3. 学会等名 2018年度統計関連学会連合大会
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

廣瀬善大のホームページ
<https://sites.google.com/view/yoshihirohirose/japanese>

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
---------	---------