

科学研究費助成事業 研究成果報告書

令和元年5月29日現在

機関番号：12601

研究種目：研究活動スタート支援

研究期間：2017～2018

課題番号：17H06569

研究課題名（和文）低ランク性を活かした統計手法の開発と基礎理論

研究課題名（英文）Development of statistical methods employing low-rankness and their basic theory

研究代表者

松田 孟留（Matsuda, Takeru）

東京大学・大学院情報理工学系研究科・特任助教

研究者番号：50808475

交付決定額（研究期間全体）：（直接経費） 2,100,000円

研究成果の概要（和文）：近年、さまざまな分野において膨大なデータを解析する必要が生じており、データのもつ特性を活用した統計手法の開発が必須である。本研究では、データのもつ低ランク性を活かした行列補完、ノンパラメトリック回帰、損失推定の手法を開発し、その基礎理論を整備した。現実の多変量データは低ランク性を有することが多いため、これらの手法を用いることでより効果的な統計解析が可能となる。

研究成果の学術的意義や社会的意義

近年、さまざまな分野において膨大なデータを解析する必要が生じており、データのもつ特性を活用した統計手法の開発が必須である。スパース性に着目した統計手法は最近さかんに研究されている一方で、低ランク性を活かした統計手法はまだ十分に研究されているとはいえない。本研究の成果として、低ランク性を活かした行列補完、ノンパラメトリック回帰、損失推定の手法が開発した。現実の多変量データは低ランク性を有することが多いため、これらの手法を用いることで効果的な統計解析が可能となる。

研究成果の概要（英文）：Recently, big data analysis is becoming more and more important in many fields and so we need to develop statistical methods that take advantage of some property of real data. In this study, we developed statistical methods for matrix completion, nonparametric regression, and loss estimation that employ low-rankness in data. Since multivariate data in real world often has low-rankness, the proposed methods lead to more effective data analysis.

研究分野：統計科学

キーワード：低ランク 特異値 行列補完 縮小ランク回帰 ベイズ統計

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

近年、さまざまな分野において膨大なデータを解析する必要が生じている。この課題に対して、データのもつスパース性を活用した統計手法が最近さかんに研究されてきた。一方、現実のデータは低ランク性を有することも多いが、低ランク性を活かした統計手法はまだ十分に研究されているとはいえない。ここで低ランク性とは、行列の特異値がスパースであることを意味する。

2. 研究の目的

本研究の目的は、低ランク性を活かした効果的な多変量解析の手法を開発し、さらに開発した手法の理論的な基礎づけを与えることである。申請者が以前に構成した特異値縮小事前分布に基づいて高次元多変量回帰や行列補完の問題に対する統計手法を開発し、同時に理論保証を与える。また、低ランク性を活かしたモデル選択規準の構成にも取り組む。

3. 研究の方法

高次元多変量回帰や行列補完の問題に対して、申請者が Matsuda and Komaki (2015) で構成した特異値縮小事前分布を土台とした手法を開発する。特異値縮小事前分布は経験ベイズの観点からも自然に解釈できるため、経験ベイズ法の適用にも取り組む。得られた手法に関して、効率的なアルゴリズムや理論保証も与える。数値実験や実データに対する性能評価により提案手法と既存手法を比較し、提案手法の利点を明らかにする。また、低ランク性を活かしたモデル選択規準の構成について、統計的決定理論（特に縮小推定）の専門家である Rutgers 大学の William E. Strawderman 教授と共同で取り組む。

4. 研究成果

低ランク性を活かして行列型データの未観測成分を推定する経験ベイズ行列補完の手法を開発した。提案手法は、Efron and Morris による行列平均に対する特異値縮小ミニマックス推定量の拡張になっている。すなわち、Efron and Morris が用いた階層モデルにおいて一部の成分のみが観測できるという設定を考え、経験ベイズ推定量を導出したのが提案手法である。EM アルゴリズムの各ステップが線形代数の公式を用いて陽に書き下せるため、効率的なアルゴリズムとなっている。また、提案手法は既存手法と比べてパラメータチューニングが不要という利点をもつ。数値実験の結果、提案手法は多くの状況で既存手法と同等以上の補完精度を達成することが確認された（図1）。また、実データに対して提案手法が既存手法を上回る補完精度を達成することも確認された。さらに、観測ノイズの分散が変数によって異なる場合への対処、欠損メカニズムを考慮した拡張、という2つの拡張を行った。

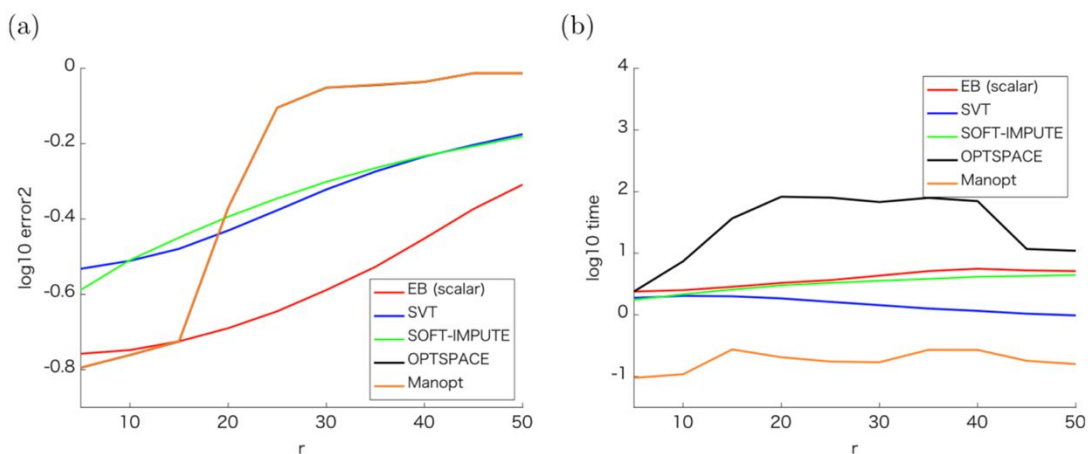


図1：経験ベイズ行列補完アルゴリズムの既存手法との比較。(a)補完精度、(b)計算時間

低ランク性を活かしたノンパラメトリック回帰の手法を開発し理論解析を行った。ノンパラメトリック回帰では関数を基底展開したときの係数列の減衰の速さで滑らかさを表現するのが一般的である。本研究では、複数のノンパラメトリック回帰を同時に行う設定において、各係数列の減衰の速さに加えて係数列の間の相関構造も考慮した関数クラスを定義した（図2）。この関数クラスは、アレイ信号データなど本質的に低次元であることが多いデータを表現するのに有効である。この関数クラスに対して、係数列をブロックに分割して各ブロックに特異値縮小を施す推定量が漸近的ミニマックスリスクを達成することを証明した。

$$\Theta(\beta, Q) = \left\{ \theta \in (l^2(\mathbb{N}))^m \mid \sum_{j=1}^{\infty} a_j^2 \theta_j \theta_j^\top \leq Q \right\}$$

図2：多変量ノンパラメトリック回帰に対するソボレフ楕円の定義

低ランク性を活かした損失推定量の改良に関して、Rutgers大学のWilliam E. Strawderman教授と共同研究を行った。モデル選択や信頼区間の構成においてリスクの不偏推定量は重要な概念である。Jonhstoneは統計的決定理論の枠組みを用いて損失推定の問題を定義し、通常のパラメータ推定問題と同様に、Stein型の縮小によってリスクの不偏推定量を改良できることを示した。我々はこの結果を行列型パラメータが現れる状況に拡張し、特異値縮小によってリスクの不偏推定量を改良できることを証明した。最尤推定量やEfron-Morris推定量に加えて、縮小ランク回帰型の推定量に対しても、リスクの不偏推定量を優越する損失推定量を導出した。数値実験によって、実際に損失推定の精度が改善されることが確認された(図3)。また、本研究において、一般の特異値変換型の推定量に対してSteinが与えたリスクの不偏推定量の公式の厳密な証明を与えた。従来議論では縮小ランク回帰型推定量のような不連続性をもった推定量に対する理論保証があいまいであったが、本研究では必要な正則条件を明確にし、たしかに縮小ランク回帰型推定量に対して従来公式が正しいことを確認した。

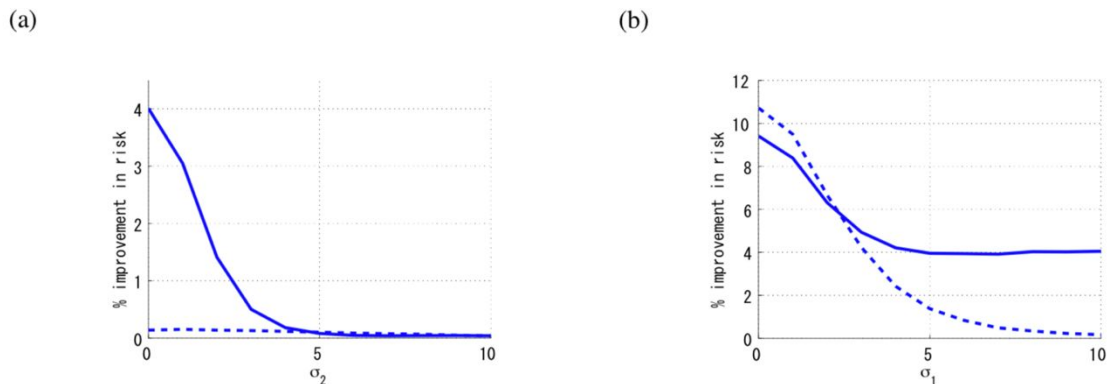


図3：最尤推定量に対する損失推定量のリスクの数値計算結果。
破線が特異値縮小型の損失推定量、実線がJohnstoneによる損失推定量

5. 主な発表論文等

〔雑誌論文〕(計 3 件)

- T. Koyama, T. Matsuda and F. Komaki. Minimax estimation of quantum states based on the latent information priors. *Entropy*, Vol. 19, 618, 2017.
- T. Matsuda and W. E. Strawderman. Improved loss estimation for a normal mean matrix. *Journal of Multivariate Analysis*, Vol. 169, pp. 300--311, 2019.
- T. Matsuda and F. Komaki. Empirical Bayes matrix completion. *Computational Statistics & Data Analysis*, Vol. 137, pp. 195--210, 2019.

〔学会発表〕(計 10 件)

- 2017.9 松田 孟留, 駒木 文保. 経験ベイズ法による行列補完. 2017 年度統計関連学会連合大会.
- 2017.11 松田 孟留, 駒木 文保. 経験ベイズ法による行列・テンソル補完. IBIS2017.
- 2018.6 T. Matsuda. Minimax adaptive reduced-rank regression. 5th IMS Asia Pacific Rim Meeting (ims-APRM 2018), Singapore.
- 2018.7 T. Ohki, T. Matsuda, A. Gunji, Y. Takei, R. Sakuma, H. Takahashi, Y. Kaneko, M. Inagaki, T. Hanakawa and K. Hiraki. Timing of phase amplitude coupling in the temporal pole is essential for neuronal and functional maturation in adolescence. 11th FENS Forum of Neuroscience, Berlin.
- 2018.7 T. Matsuda, F. Homae, H. Watanabe, G. Taga and F. Komaki. Statistical verification of the common oscillatory behaviors in oxy-Hb and deoxy-Hb time series. Toward Understanding "INDIVIDUALITY", Kyoto.

2018.9 松田 孟留, 宮武 勇登. 離散化誤差を考慮した常微分方程式の初期値推定. 日本応用数学会 2018 年度年会.

2018.9 松田 孟留. Information criteria for nonnormalized models. 2018 年度統計関連学会連合大会.

2018.10 T. Matsuda, F. Homae, H. Watanabe, G. Taga and F. Komaki. Statistical verification of the common oscillatory behaviors in oxy-Hb and deoxy-Hb time series. fNIRS 2018, Tokyo.

2019.2 T. Matsuda and Y. Miyatake. Reducing the effect of discretization errors in estimating ODE models by iteratively reweighted least squares. ANZIAM 2019, Nelson.

2019.2 T. Matsuda. Oscillator decomposition of time series data. ICMMA 2018, Tokyo.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称:

発明者:

権利者:

種類:

番号:

出願年:

国内外の別:

取得状況(計 0 件)

名称:

発明者:

権利者:

種類:

番号:

取得年:

国内外の別:

〔その他〕

ホームページ等

6. 研究組織

(1) 研究分担者

研究分担者氏名:

ローマ字氏名:

所属研究機関名:

部局名:

職名:

研究者番号(8桁):

(2) 研究協力者

研究協力者氏名:

ローマ字氏名:

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。