

令和元年6月13日現在

機関番号：15401

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00047

研究課題名(和文)高次元多変量モデルに関する変数選択法と応用

研究課題名(英文) Variable selection methods in high-dimensional multivariate models and their applications

研究代表者

藤越 康祝 (Yasunori, FUJIKOSHI)

広島大学・理学研究科・名誉教授

研究者番号：40033849

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：多変量解析、とくに、判別分析、多変量回帰分析、主成分分析、正準相関分析、などにおいて、変数選択や次元推定のための情報量規準に関して、標本数や変数の数が大きい状況で一致性をもつための条件を与えた。また、これらの規準は、変数の数が増えると、計算上の問題点が生じるが、多変量回帰モデルや判別分析においては、この問題点を解消した「一般化一つ取って置き法」を提案し、その高次元一致性を持つための条件を与えた。さらに、多変量線形モデルにおける次元の推定に関しては、標本数が変数の数より少ない場合にも利用できる正則化情報量規準を提案し、その高次元性質を調べた。

研究成果の学術的意義や社会的意義

多変量解析においては情報が入手しやすくなったこともあって、変数の次元が大きい場合の統計的方法や、膨大な変数の中から有用な変数を抽出する変数選択法に高い関心がよせられている。本研究は、このような高次元データ分析における重要な課題における基礎的研究のみならず応用に焦点を当てており、統計科学分野における理論や応用に関して高い貢献が期待される。

研究成果の概要(英文)：In multivariate analysis, especially, discriminant analysis, multivariate regression models, principal component analysis, canonical correlation analysis, etc., we derived sufficient conditions for consistency of variable selection methods based on information criteria when the sample size and the dimension are large. The methods have a computational problem when the number of variables are large. In order to avoid its computational problem, we proposed a generalized kick-one-out (KOO) method which shared the high-consistency property in discriminant analysis and multivariate regression model. Further, in the problem of rank estimation in multivariate linear model, we proposed a regularized information criterion, which can be used for the case that the sample size is smaller than the dimension of variables, and studied its high-dimensional properties.

研究分野：統計科学

キーワード：多変量回帰モデル 変数選択法 主成分分析 判別分析 情報量規準 一つ取って置き法 高次元漸近枠組 多変量モデル

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

1. 研究開始当初の背景

(1) 多変量解析においては情報が入手しやすくなったこともあって、変数の次元が大きい場合の分析法に関心がよせられている。例えば、DNN マイクロデータ、ファイナンスデータや画像データのように、次元が標本数より遥かに大きい超高次元データの分析にも関心がある。一方、超高次元とはいかなくても、変数の次元が標本数以下である場合や、変数のスクリーニング法によって、次元を標本数以下にすることができる場合がある。

(2) このような状況において、次元数と標本数の比が一定値に近づく高次元漸近的枠組のもとでの漸近分布が求められるようになった。例えば、多変量正規性のもとで、共分散行列に関するいくつかの検定統計量、線形判別関数、高次元特有の統計量、などに関する結果がある。しかし、現状ではまだ発展段階であって、とくに、非正規での進展が期待されている。

(3) 多変量モデルにおいて変数が増えると、冗長な変数も増え、そのような冗長な変数を除き本質的に必要な変数を抽出するための変数選択問題は極めて重要である。そのアプローチとして、高次元漸近的枠組において一致性をもつように、情報量規準を一般化する研究が一部のモデルで展開されていた。

2. 研究の目的

(1) 本研究では、多変量解析とくに、判別分析、多変量回帰分析、主成分分析、正準相関分析、多変量逆回帰分析、などにおいて、目的変数と説明変数の数が大きい場合における変数選択法と次元選択法に関して研究することを目的としている。具体的には、次の(2)～(4)に関して研究する。

(2) AIC や C_p などの情報量規準を高次元の場合へ拡張し、それらの性質を解明する。また、関連する推測問題にも取り組む。とくに、種々のモデルにおいてリスクの高次元漸近不偏推定量としてのモデル選択規準の構成と、それらの性質を解明する。また、関連した推測統計量の高次元推測分布を扱う。

(3) 罰則付最適化法に基づく変数選択法を開発し、修正モデル選択法との関係を解明する。とくに、標本数が変数の次元より少ない場合にも利用可能となる正則化情報量規準について研究する。

(4) 得られた結果を数値実験により検証し、さらに実データへの応用に取り組む。

3. 研究の方法

(1) 情報量規準に基づく変数選択問題に関しては、次の問題に対して一致性などの性質を持つ高次元情報量規準を提案する。

変数の次元は標本数より小さいとし、多変量回帰モデルにおける回帰係数の選択、多変量回帰モデルにおける目的変数の選択、多変量多項式成長曲線モデルにおける次数の選択、判別分析における変数の選択、判別分析における次元の推定、正準相関分析における変数選択、正準相関分析における次元の推定、などを扱う。

(2) 標本数が次元より小さい場合を扱う。この場合には通常の共分散行列の推定量が特異となるため、例えばリッジ型推定量に置き換えた方法などについてモデル選択規準量の高次元問題を扱う。

(3) 多変量推測統計量の高次元漸近分布については、次の問題に取り組む。

基礎分布が多変量正規分布である場合の多変量推測統計量の高次元分布を導出する。例えば、判別分析および正準相関分析における冗長性検定統計量、MANOVA 問題における固有値、などに対して取り組む。

高次元分布に関する研究成果の多くは、モデルの基礎(誤差)分布として多変量正規性を想定しての展開であるが、基礎分布が非正規の場合への拡張を試みる。

(4) (1)、(2)、(3)で得られた理論的結果に関して、シミュレーションによる数値的検討や実データへの適用を考える。

(5) 本研究では理論的研究が中心になるが、シミュレーションによる数値的検討や実データへの適用に関しても研究する。研究を進める上での役割分担は、高次元変数選択法・高次元推測法の漸近理論及び総括(研究代表者: 藤越) 数値的検証と応用(研究分担者: 櫻井) とする。

4 . 研究成果

Fujikoshi and Sakurai (2019, JJSD)では、2群の判別分析における変数選択問題において、一般化一つ取って置き法(検定型法)による変数選択法を提案した。また、高次元一致性を持つための条件を与えた。この変数選択法は、変数が多くても実行可能であるという大きな利点を持っている。

Bai, Choi and Fujikoshi(2018, Ann. Statist.)では、MANOVAにおける固有値の高次元漸近分布を導出した。その結果は、ランダム行列論を発展させることによって導出された。

主成分分析における次元推定のためのAICとBICについて、高次元漸近的枠組の基で一致性をもつための十分条件を与えた。この結果はBai, Choi and Fujikoshi(2018, Ann. Statist.)によって導出され、非正規で、標本数が変数の数より小さい場合にも与えられている。

正準相関分析において、一方の変数の組の一部が冗長であるかどうかの検定糖計量について、高次元漸近的枠組のもとで仮説および対立仮説の基で漸均分布を導出した。この結果は、Oda, Yanagihara and Fujikoshi(2019, Random Matrices: Theory and Applications)に記載されている。

Fujikoshi(2017, SUT Journal of Mathematics)では、正準相関分析における次元推定のためのAIC、BIC、 C_p 規準の高次元一致性のための十分条件を与えている。

Fujikoshi and Sakurai(2016, JMA)では、多変量線形モデルにおける次元推定のための情報量規準の高次元一致性のための十分条件を与えている。さらに、標本数が変数の数より小さい場合に、正則化法を利用した規準量についてもその高次元一致性に関する結果を与えている。

Fujikoshi and Sakurai(2016, Amer. J. Math. Manage. Sci.)では、主成分分析の次元推定のための情報量規準について、大標本の枠組のもとでの性質を研究している。

5 . 主な発表論文等

[雑誌論文](計7件)

Yasunori Fujikoshi, Tetsuro Sakurai,

Consistency of test-based method for selection of variables in high-dimensional two-group discriminant analysis, Japanese Journal of Statistics and Data Science, 査読有、Vol.2、2019、155-171

DOI: 10.1007/s42081-019-00032-4

Zhidong Bai, Kwork P. Choi, Yasunori Fujikoshi, Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT, Ann. Statist., 査読有、Vol.46、2018、2985-3013
DOI: org/10.1214/17-AOS1646

Zhidong Bai, Kwork P. Choi, Yasunori Fujikoshi, Consistency of AIC and BIC in estimating the number of significant components in high-dimensional principal component analysis, Ann. Statist., 査読有、Vol.46、2018、1050-1076
DOI: org/10.1214/17-AOS1646

Ryoya Oda, Hirokazu Yanagihara, Yasunori Fujikoshi, Asymptotic null and non-null distributions of test statistics for redundancy in high-dimensional canonical correlation analysis, Random Matrices: Theory and Applications, 査読有、Vol.8、2019、1950001-26
DOI: org/10.1142/S2010326319500011

Yasunori Fujikoshi, High-dimensional properties of AIC, BIC and C_p for estimation of dimensionality in canonical correlation analysis, SUT Journal of Mathematics, 査読有、Vol.53、2017、59-72

<http://www.rs.tus.ac.jp/sutjmath/>

Yasunori Fujikoshi, Tetsuro Sakurai, High-dimensional consistency of rank estimation criteria in multivariate linear Model, Journal of Multivariate Analysis, 査読有、Vol.149、2016、199-212

DOI: org/10.1016/j.jmva.2016.04.005

Yasunori Fujikoshi, Tetsuro Sakurai, Some properties of estimation criteria for dimensionality in principal component analysis, Amer. J. Math. Manage. Sci., 査読有、Vol.35、2016、133-142

DOI: org/10.1080/01966324.2015.1121848

[学会発表](計11件)

櫻井哲朗(代表)、藤越康祝、多変量回帰分析や判別分析などにおける新たな変数選択法の提案、統計関連学会連合大会、2018

鈴木裕也(代表)、小田凌也、柳原宏和、藤越康祝、正準判別における一致性を持つ高次元

変数の選択法、統計関連学会連合大会、2018

八木文香（代表）、瀬尾隆、藤越康祝、単調欠測データをもつ成長曲線モデルに関する AIC 型選択規準、統計関連学会連合大会、2018

藤越康祝（代表）、2 次判別関数に関する高次元漸近近似の誤差限界、2018 年度統計関連学会連合大会、2018

山田隆行（代表）、櫻井哲朗、藤越康祝、2 群の線形判別法に関する誤判別確率の高次元漸近ロバストネスについて、統計関連学会連合大会、2018

山田隆行（代表）、櫻井哲朗、藤越康祝、W および Z 判別法について～大標本かつ高次元の下で考察、統計関連学会連合大会、2017

藤越康祝（代表）、非正規多変量回帰モデルにおける検定統計量および固有値の高次元かつ大標本のもとでの漸近分布、統計関連学会連合大会、2017

櫻井哲朗（代表）、藤越康祝、共分散構造をもつ多変量回帰モデルにおける Cp 型の変数選択規準の高次元一致性、統計関連学会連合大会、2017

八木文香（代表）、瀬尾隆、藤越康祝、単調型欠測値をもつ成長曲線モデルにおける推定、統計関連学会連合大会、2017

藤越康祝（代表）、多変量線形モデルにおける次元および変数選択のための情報量規準の高次元性質、統計関連学会連合大会、2016

櫻井哲朗（代表）、藤越康祝、高次元の場合における次元の推定法と漸近的性質、統計関連学会連合大会、2016

〔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：櫻井 哲朗

ローマ字氏名：(SAKURAI, tetsuro)

所属研究機関名：公立諏訪東京理科大学

部局名：共通・マネジメント教育センター

職名：講師

研究者番号 (8 桁)：60609741

(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。