

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 31 日現在

機関番号：15401

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330038

研究課題名(和文) 多変量モデルの高次元推測と応用に関する研究

研究課題名(英文) High-Dimensional Statistical Inference for Multivariate Models and Its Applications

研究代表者

藤越 康祝 (Fujikoshi, Yasunori)

広島大学・理学(系)研究科(研究院)・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) 3,700,000円

研究成果の概要(和文)：本研究においては、多変量回帰モデルや判別分析モデルを含む多変量線形モデルおよび成長曲線モデルにおいて、モデル選択規準にもとづく変数選択法の高次元での一貫性などの性質を明らかにした。モデル選択規準としては、AIC規準、BIC規準、Cp規準などを取り上げ、標本数の大きさと目的変数の数の比が1より小さい一定値に近づくという高次元漸近的枠組みのもとで、変数選択法が一貫性をもつための十分条件を導出した。また、正準相関分析において、変数の冗長性を検定するための尤度比統計量に対して、高次元での漸近分布とその誤差限界を与えた。

研究成果の概要(英文)：In this project we derived high-dimensional consistency properties for variable selection methods based on model selection criteria in multivariate linear model including multivariate regression model and discriminant analysis and in growth curve model. The model selection criteria treated includes AIC criterion, BIC criterion and Cp criterion. The high-dimensional properties were derived under a high-dimensional asymptotic framework such that the ratio of the number of response variables to the sample size tends to a fixed number less than 1. Further, we derive a high-dimensional asymptotic distribution for the likelihood ratio criterion for testing an additional information hypothesis in canonical correlation analysis and its error bound.

研究分野：統計科学

キーワード：多変量線形モデル 多変量回帰モデル 判別分析モデル 成長曲線モデル 変数選択法 モデル選択規準 高次元一貫性 冗長性検定

1. 研究開始当初の背景

多変量モデルの高次元推測に関して、次の背景が研究目的に繋がっている。

(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、変数をふるいにかけることによって、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、標本共分散行列の最大固有値、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、今後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。また、変数選択法のアプローチとして、冗長性モデルを導入し、これにAIC規準などの情報量規準を適用する方法がある。ここで強調したいのは、AICなどのモデル選択規準は大標本漸近的枠組のもとで提案されたものであり、高次元の場合の性質や、高次元の場合に対しても適当な方法を提案することが重要な課題になっている。

2. 研究の目的

多変量モデルの高次元推測問題として、

- (1) モデル選択規準に基づく変数選択問題、
 - (2) 多変量推測統計量の高次元漸近分布の導出、
 - (3) 得られた結果の数値的検証と応用、
- に取り組む。(1)に関しては、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。(2)に関しては、伝統的な検定統計量や固有値固有ベクトルのみならず、高次元推測に特有な統計量の高次元漸近分布の導出にも取り組む。本研究での高次元漸近的元枠組としては、有用な結果が期待される次元数と標本数の比が一定値に近づく場合を想定している。

3. 研究の方法

(1) モデル選択規準に基づく変数選択問題に関しては、次の問題に対して高次元AIC規準や C_p 規準を提案し、一致性などの性質を明らかにする。

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。

の問題に対して、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。

(2) 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固有ベクトル、高次元規準の高次元漸近分布、行列値データの検定統計量、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。

(3) (1),(2)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。

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も視野においている。研究を進める上での役割分担は、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:柳原)とする。

4. 研究成果

本研究においては、多変量モデルの選択規準に基づく変数選択問題、多変量推測統計量の高次元漸近分布の導出、および、得られた結果の数値的検証と応用に取り組むことを目的にしている。これらの目的に関連した主要結果として、次の成果を得た。

多変回帰モデルにおいて、回帰係数の変数選択問題に対して高次元AIC規準を提案した。通常のAIC規準は、モデルのリスクの大標本漸近的な不偏推定量と提案されるが、本研究では目的変数の次元と標本数が共に大きいという高次元漸近的枠組のもとでの漸近的な不偏推定量として提案している。一般に、モデル選択規準としてのAIC規準や C_p 規準は、大標本の漸近的枠組のもとで一致性をもたないことが知られている。しかし、本研究では、高次元漸近的枠組のもとで一致性があ

ることを示した。これらの成果は、Fujikoshi, Sakurai and Yanagihara (2014, JMA), Yanagihara, Wakaki and Fujikoshi (EJS, 2015)に掲載されている。また、Yanagihara (JJSS, 2015)では、候補のモデルに正規性は仮定するが、真のモデルには正規性を仮定しないという条件の下で、一致性を満たすための罰則項の条件を導出した。目的変数の数が標本数より大きい場合は、今後の課題である。

多変量成長曲線モデルにおいても多項式のモデルの次数の選択、および、個体間説明変数の選択に関して、高次元規準量の提案やAICとCp規準の一致性に関して新たな結果を得た。とくに、前者の問題では、標本数と目的変数の次元を大にしたとき一致性をもたないが、後者の問題では、標本数と個体間説明変数を大にしたとき一致性をもつことを指摘した。これらの成果は、Fujikoshi, Enomoto and Sakurai (JMA, 2013)、Enomoto, Sakurai and Fujikoshi (SUTJM, 2013)に掲載されている。

Fujikoshi and Sakurai (2016, AJMMS)では、主成分分析において、モデル選択規準に基づく次元推定法の大標本漸近的枠組での一致性に関する性質を導出している。また、この推定法の高次元での性質も導出し、投稿中である。AIC規準は大標本漸近的枠組では、一致性をもたないが、高次元漸近的枠組では、一致性をもつ場合があることを指摘している。

Fujikoshi and Sakurai (2016, JMAに掲載決定)では、多変量回帰モデルにおける回帰係数行列のランクの推定に関する情報規準に関して、高次元漸近的枠組のもとで、一致性をもつための条件を与えている。さらに、標本数が目的変数の個数より小さい場合に対してリッジ型の情報規準量を提案し、数値的にその性質を導出している。リッジ型の情報規準量の性質を理論的に明らかにすることは、今後の重要課題である。

Wakaki and Fujikoshi (TPIA, 2016)では、正準相関分析においてある変数の追加情報の有無を調べるための尤度比統計量の高次元漸近分布およびその近似の誤差限界を与えている。Wakaki, Fujikoshi and Ulyanov (HMJ, 2015)では、多変量分散分析モデルにおける尤度比検定統計量に関して、高次元漸近的枠組のもとでの漸近展開を導出している。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計8件)

Yasunori Fujikoshi, Tetsuro Sakurai,

Some properties of estimation criteria for dimensionality in principal component analysis, Amer. J. Math. Manage. Sci., 査読有、35巻、2016、133 ~ 142

Hirokazu Yanagihara, Conditions for consistency of a log-likelihood-based information criterion in normal multivariate linear regression models under the violation of normality assumption, Journal of the Japan Statistical Society, 査読有、45巻、2015、21 ~ 56

Hirokazu Yanagihara, Hirofumi Wakaki, Yasunori Fujikoshi, A consistency property of the AIC for multivariate linear models when the dimension and the sample size are large, Electronic Journal of Statistics, 査読有、9巻、2015、869 ~ 897
DOI: 10.1214/15-EJS1022

Hirofumi Wakaki, Yasunori Fujikoshi, Vladimir V. Ulyanov, Asymptotic expansions of the distributions of MANOVA tests when the dimension is large, Hiroshima Math. J., 査読有、44巻、2014、247 ~ 259

Yusuke Hashiyama, Hirokazu Yanagihara, Yasunori Fujikoshi, Jackkife bias correction of the AIC for selecting variables in canonical correlation analysis under model misspecification, Linear Algebra and Its Application, 査読有、455巻、2014、82-106

Yasunori Fujikoshi, Tetsuji Sakurai Hirokazu Yanagihara, Consistency of high-dimensional AIC-type and Cp-type criteria in multivariate linear regression, Journal of Multivariate Analysis, 査読有、123巻、2014、184 ~ 200
DOI: 10.1016/j.jmva.2013.09.006

Rie Enomoto, Tetsuro Sakurai, Yasunori Fujikoshi, Consistency of AIC and its modifications in the growth curve model when the number of groups and the sample size are large, SUT Journal of Mathematics, 査読有、49巻、2013、93 ~ 107

Yasunori Fujikoshi, Rie Enomoto, Tetsuro Sakurai, High-dimensional AIC in the growth curve model, Journal of Multivariate Analysis, 査読有、122巻、2013、239 ~ 250
DOI: 10.1016/j.jmva.2013.07.006

[学会発表](計8件)

榎本理恵(代表)、櫻井哲朗、藤越康祝、一様および自己回帰共分散構造をもつ成

長曲線モデルに関する高次元推測、統計
関連学会連合大会、2015年9月7日、岡
山大学（岡山県岡山市）

櫻井哲朗（代表）、藤越康祝、主成分分析
における次元の推定、統計関連学会連合
大会、2015年9月7日、岡山大学（岡山
県岡山市）

Yasunori Fujikoshi、Explicit and
computable error bounds for
asymptotic expansions of the
distribution functions of some
multivariate statistics, International
Scientific Conference “Probability
Theory and Its Applications”, 2015年2
月13日、Moscow State University
(Moscow, Russia)

藤越康祝、判別分析における情報量規準
を用いた変数選択法の高次元漸近的性質、
統計関連学会連合大会、2014年9月14
日、東京大学（東京都文京区）

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、2014年9月14
日、東京大学（東京都文京区）

藤越康祝、情報量規準を用いた次元の推
定法の高次元一致性、統計関連学会連合
大会、2013年9月10日、大阪大学（大
阪府・豊中市）

櫻井哲朗（代表）、藤越康祝、多項分布に
おける高次元漸近極限、統計関連学会連
合大会、2013年9月9日、大阪大学（大
阪府・豊中市）

榎本理恵（代表）、櫻井哲朗、藤越康祝、
成長曲線モデルにおける各規準量の高次
元漸近分布、統計関連学会連合大会、
2013年9月9日、大阪大学（大阪府・豊
中市）

〔図書〕（計1件）

杉山高一、藤越康祝、小椋透、朝倉書店、
多変量データ解析、2014、225

6. 研究組織

(1) 研究代表者

藤越 康祝 (FUJIKOSHI YASUNORI)
広島大学・大学院理学研究科・名誉教授
研究者番号：40033849

(2) 研究分担者

柳原 宏和 (YANAGIHARA HIROKAZU)
広島大学・大学院理学研究科・准教授
研究者番号：70342615