様式 C-19

科学研究費補助金研究成果報告書

平成 22 年 3 月 31 日現在

研究種目：基盤研究(C)
研究期間：2007～2009
課題番号：19500243
研究課題名（和文）　高次元情報量基準の構成と応用に関する研究

研究課題名（英文）　Study on construction and applications of high-dimensional information criteria

研究代表者
藤越　康祝（FUJIKOSHI YASUNORI）
中央大学・理工学部・客員教授
研究者番号：40033849

研究成果の概要（和文）：本研究では，種々の多変量モデルにおいて，次元は標本数より小さいが共に大きくなるという高次元漸近的枠組で，モデル選択基準の導出や，関連する基本統計量の漸近分布を導出した．具体的には，判別分析における追加情報モデルに対して高次元 AIC 基準，MANOVA モデルにおける平行性モデルに対する CAIC 基準，多変量回帰モデルにおけるリッジパラメータ選択に対する修正 Cp 基準を導出し，高次元のみならず大標本の状況においてもよい基準であることを数値的に確認した．また，MANOVA の固有値，正準相関係数などの高次元漸近分布を導出した．

研究成果の概要（英文）：In this study, we considered to derive model selection criteria for various multivariate models under high-dimensional situations where the dimension is smaller than the sample size, but both of them tend to infinity. Concretely we derived high-dimensional AIC for additional information models in discriminant analysis, CAIC criterion for parallelism models in MANOVA and modified Cp criterion for selection of ridge parameters in multivariate regression model. Further, we derived high-dimensional asymptotic distributions of the roots in MANOVA, canonical correlations, etc.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2007 年度	1,100,000	330,000	1,430,000
2008 年度	900,000	270,000	1,170,000
2009 年度	700,000	210,000	910,000
年度			
年度			
総　計	2,700,000	810,000	3,510,000

研究分野：総合領域
科研費の分科・細目：情報学
キーワード： 高次元情報量基準，多変量モデル，判別分析，変数選択， 高次元小標本，正準相関，経時データモデル，共分散構造

１．研究開始当初の背景

統計的モデルに対する情報量基準の代表的なものとし，赤池氏によって提案された AIC 基準がある．これはモデルのよさ，実際にはモデルの悪さ(リスク)を"-2 log（予測尤度)"で測り，その漸近

的不偏推定量として提案されたものであって,
"AIC=-2 log (最大尤度) + 2 (独立パラメータ数)"
として定義される．第１項はリスクのナイーヴな推定量で，第２項はその推定量のバイアス(補正)項とよばれる．この基準あるいは適当に修正された基準は，種々の統計的モデルの選択に応用され，有用な成果をもたらしている．モデルのリスクを測る情報量基準としてはこの他に，回帰型のモデルに対して予測誤差平方和に基づく Cp 基準が用いられ，また，判別モデルでは誤判別率確率が用いられる．さらに，漸近的に真のモデルを選ぶ基準として，AIC におけるバイアス項の"2"を"log n"に代えたものである BIC 基準がある．

統計的モデルの選択に用いられる AIC, Cp, BIC 基準などは，提案されて以来，国内外の研究者によって理論および応用の両面において研究されている．多変量の場合にも AIC 基準などが適用されているが，この場合には新たな問題が提起されることを注意したい．今，p 個の変数をもつ多変量確率変数が n 個の個体に対して観測されることによって，n x p のデータ行列 X が与えられているとする．X の統計モデルに関して AIC 基準が形式的に求められるが，これは大標本漸近的枠組，すなわち p を固定して n を大としたもとでのリスクの漸近的不偏推定量になるように工夫されたものであることを注意したい．従って，変数の次元 p が n に比べてかなり小さければ，通常の AIC 基準で問題は生じないであろう．しかし，たとえば n=100 で p=50 の場合，大標本のもとで構成された AIC 基準が適切であるかについて，大きな疑問が生じる．

ところで，多変量解析においては情報が手に入りやすくなったこともあって，最近 DNA マイクロデータのように，n << p のような超高次元データの分析にも関心がある．また，超高次元とはいかなくても，p < n ではあるが，p～n/2 のような場合も多い．また，n < p のようなデータに対しても、変数をふるいをかけることによって，p～n/2 の状況になる場合もある．いずれにせよ，p < n で p が比較的大きい状況は，多変量解析においては非常に重要な状況である考えられる．このようなとき，大標本近似はまったく利用しがたいものになっていることも見てきている．

従って，次元が標本数に比べて大きいような多変量データに対して AIC 基準の構成を見直すことは重要である．その際，最近，p/n → c とした高次元漸近的枠組での漸近理論が有効であることが指摘されている．

２．研究の目的

本研究では，多変量モデルの選択に対して変数の次元 p が大きい場合の情報量基準の構成に焦点を当てている．ここでは，まず次元 p と標本数 n について，p < n で両者は同程度に大きくなると言う高次元漸近的枠組のもとで検討する．具体的に取り上げる多変量モデルとしては，共分散構造にも関連している次のモデルを取り上げる．

（１）多変量回帰モデル，判別モデル，正準相関モデルにおける追加情報および次元に関するモデル
（２）主成分の次元に関する共分散構造モデル
（３）経時データに関する平均・共分散構造モデル．

（１）に関するモデルは変数選択問題と密接に関連している．（２）における経時データとは経時的に繰り返し測定されることによって得られたデータであって医学・薬学・生物分野を始めとして多くの分野で重要になってきている．

次に p < n の場合の成果をもとに，n < p の場合に対しても適用可能な高次元情報基準量の構成に取り組む．

高次元情報量基準を構成するために，大標本漸近的アプローチを高次元漸近的アプローチ，すなわち p/n → c で展開することを試みる．

なお，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．

３．研究の方法

統計的モデル選択問題に関する高次元情報量基準の構成に向けて，具体的な多変量モデルでの研究を進める．取り上げるモデルは，研究目的で述べているモデルであって，（１）多変量モデル；多変量回帰モデル，判別モデル，正準相関モデルにおける追加情報および次元に関するモデル，（２）主成分の次元に関する共分散構造，（３）経時データに関する平均・共分散構造モデル，などである．

研究を進めるにあたって，これらのモデル選択問題のうちのいくつかについては，大標本漸近的枠組のもとでの

情報量基準の構成が研究されていているので，そこでの考え方を生かすことを考える．まず，これらの大標本情報量基準の構成において，大標本漸近理論を高次元漸近理論で展開することを試みる．

（2）に関しては，大きい方のいくつかが意味のある主成分で，残りの主成分の分散は同一であるというモデルに対する情報量基準を構成することを考えている．

初年度においては，各モデルには多変量正規性を想定し，真のモデルは候補のモデルに含まれる場合を扱う．研究は，理論面のみならず，シミュレーション及び実データへの応用に関しても研究する．

研究分担者や連携研究者と協力して，高次元クロスバリデーションの構成，リッジ法に基づくモデル選択基準の構成，パーミュテイション近似の開発，数値的検証，などについても取り組む．

研究を進める上での役割分担は，高次元情報量基準の構成及び総括(研究代表者：藤越)，高次元情報量基準に対するパーミュテーション近似(研究分担者：杉山)，リッジ法に基づく構成及び数値的検証(連携研究者：柳原)とする．上記の役割分担に沿って研究を進めるが，その際，研究代表者，研究分担者，連携研究者が必要に応じて適宜集まり，それぞれの研究課題における最新の問題を確認し，お互いに協力しながら，当該研究目的を遂行していく．また，国内の専門家を訪問し，最新の情報やコメントを得ながら研究を行う．さらに，国内での学会や研究集会において、研究成果の発表を行う．研究代表者は国際会議に出席し，研究成果の発表や，当該の研究に関して最新の情報を得ることに努める．平成 20 年度以降も平成 19 年度に沿って進める．

４．研究成果

本研究では，多変量モデルの選択基準問題において，変数の次元が標本数と同程度に大きいとする漸近的枠組のもとで高次元情報量基準の構成に焦点を当てている． 以下では，本研究目的に直結した成果と，関連した成果について述べる．

（1）判別分析における変数選択問へのアプローチとして，追加情報モデルの選択による方法がある．AIC 型リスクに対して大標本枠組での AIC 基準が提案されているが，本研究においては，高次元枠組での漸近的不偏推定量の構成に成功した．この結果に基づいて，高次元 AIC 基準を提案した．また，この基準が高次元の場合に有効であることを数値的に確認した．本結果は，学会等で発表している．

次元が大きいが標本数より小さい場合の高次元 AIC 基準の導出の先駆的成果である．次元が標本数より大きい場合への拡張が期待される．

（2） 説明変数の次元が大きい場合のアプローチの１つとしてリッジ回帰法を適用する方法がある．この場合，リッジパラメータの選択問題が生じるが，これをモデル選択の立場から考察した．リッジパラメータ選択のための Cp 規準に対して，バイアスを完全に除去した新しい修正 Cp 規準を提案した．さらに，これらの結果を多変量回帰モデルへ拡張した．

多変量回帰モデルの場合には，高次元を考慮した CAIC 基準が知られており，本研究で提案した方法との理論的比較は今後の課題である．

（3）主成分における小さい固有値の同等性に関する次元，判別分析における次元，および，正準相関分析における次元の推定問題を大標本漸近的枠組のもとで考察した．これまで，データ行列の尤度に基づくモデル選択基準が提案されていたが，本研究では，次元が固有値のみに依存することから，対応する標本固有値の尤度に基づく AIC 基準を導出した．本結果は学会等で発表した．

本結果は大標本漸近的枠組での結果であって，高次元枠組での展開が期待される．また，新たな基準が，従来の基準より，より有効な次元の推定法であると期待されるが，その検証も今後の課題である．

（4）多変量モデル選択基準の導出と関連して，プロファイル分析における平行性モデルの選択問題に取り組み，AIC 基準，並びに，次元の影響を考慮した CAIC 基準を導出した．CAIC 基準は変数の次元が大であることを考慮した基準になっている． 関連して平行性水準の信頼区間について，新たな構成法を提案した．また，経時データに対するモデルである成長曲線モデルにおいて，平行性問題を考察した．平行性仮説に対する尤度比検定統計量とその帰無分布を導出した．これらの結果の一部は投稿中である．

成長曲線モデルにおける平行性モデルに対する CAIC 基準の導出は研究中である．

（５）高次元情報量基準の導出と関連して，いくつかの基本的統計量の高次元漸近分布の導出に成功した．これらの中には，MANOVA モデルにおける固有値，正準相関係数，一様共分散構造の検定統計量が含まれている．なお，正準相関係数の変換に関して， フイシャーの z 変換を高次元の場合へ拡張した．

高次元漸近近似と大標本漸近近似の間に興味ある関係があることを個別の問題において明らかにしている． これらの結果から一般的な関係を見いだすことが期待される．

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４件）

① 藤越康祝， Statistical inference for parallelism hypothesis in growth curve model，SUT Journal of Mathematics，45 巻，137-148，2009，査読有
② 藤越康祝，櫻井哲朗，High-dimensional asymptotic expansions for the distributions of canonical correlations，J. Multivariate Anal.，100 巻，101-112,2009，査読有
③ 柳原宏和，佐藤健一， An unbiased Cp criterion for multivariate ridge regression，Journal of Multivariate Analysis，101 巻，2010，1226-1238，査読有
④ 藤越康祝，姫野哲人，若木宏文，Asymptotic results in canonical discriminant analysis when the dimension is large compared to the sample，J. Statist. Plann. Inf.，138 巻，2008，3457-3466，査読有

〔学会発表〕（計２件）

① 藤越康祝，同時方程式モデルにおける高次元漸近理論，統計関連学会，2009 年 9 月 8 日，同志社大学
② 藤越康祝，高次元での判別分析における冗長性モデルの選択，統計関連学会，2008 年 9 月 8 日，慶応義塾大学

〔図書〕（計２件）

① 藤越康祝，Vladimir V. Ulyanov，清水良一 ， Wiley， INC， Publication，Multivariate Analysis: High-dimensional and Large-Sample Approximations，2010，1-533.
② 藤越康祝，朝倉書店，経時データ解析の数理，2009，207.

６．研究組織
(1)研究代表者
藤越　康祝（FUJIKOSHI YASUNORI）
中央大学・理工学部・客員教授
研究者番号：40033849

(2)研究分担者
杉山　髙一（SUGIYAMA TAKAKAZU）
中央大学・理工学部・教授
研究者番号：70090371

(3)連携研究者
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