

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 22 日現在

機関番号：15401

研究種目：基盤研究（C）

研究期間：2010 ～ 2012

課題番号：22500259

研究課題名（和文） 多変量データ推測法の新展開と応用

研究課題名（英文） New Development of Statistical Theory in Multivariate Data Analysis and Its Applications

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

研究成果の概要（和文）：本研究では、多変量統計解析における次の重要課題に取り組んだ。

(1) 数量化法における推測理論の展開、(2) 多変量モデルの高次元選択規準の提案と性質、(3) 経時データ分析における推測法の展開、(4) 多変量非線形に関する推測法、(5) 高次元推測法と高次元漸近理論の展開。各課題について、研究成果を得て、それらを国際的雑誌に発表している。とくに、高次元 AIC 規準などを提案し、それらが、一致性をもつことを示した。

研究成果の概要（英文）：In this project we tackled the following important problems in multivariate statistical analysis: (1) Development of statistical inference in quantification methods, (2) Proposals of high-dimensional model selection criteria and study on their properties, (3) Development of statistical inference in analysis of repeated measurement data, (4) Development of statistical inference for multivariate non-linear models, (5) Development of high-dimensional statistical methods and asymptotic theory. Especially, we proposed high-dimensional AIC, etc. Further, we showed that they have consistency property.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010 年度	1,000,000	300,000	1,300,000
2011 年度	900,000	270,000	1,170,000
2012 年度	700,000	210,000	910,000
年度			
年度			
総計	2,600,000	780,000	3,380,000

研究分野：統計科学

科研費の分科・細目：情報学・統計科学

キーワード：モデル選択、多変量回帰モデル、成長曲線モデル、正準相関分析、判別分析、高次元漸近理論、多項ロジット回帰、環境データ分析

1. 研究開始当初の背景

多変量データの推測に関して、次の背景が研究目的に繋がっている。

(1) 多変量データの推測理論は多くの場合、多変量正規性を想定して発展している。一方、実際に応用されるとき、そのような仮定を無視して用いられる場合も多い。このような非正規データに対しても、正規性のもとでの推測法、あるいは、それを適当に修正した推測法が近似的にも適用できるかどうかに関心がある。

(2) 多変量データ分析法には、解析の目的に対してもとの変数の1次結合である正準変数を考えて、少数個の正準変数で分析する方法が提案されている。選ばれた少数個の正準変数に関して、もとの変数の選択問題の推測理論を発展させることが求められている。

(3) 推測法の多くは、集団の特性を明らかにするための方法として発展している。しかし、実際には個人、あるいは、それらの集まりであるクラスターに対して関心がある場合もある。このよう問題を発展させ、適用することも重要である。

(4) 多変量の非線形構造探る方法として、層別逆回帰法、部分線形最小二乗法が提案されている。前者は複数個の線形関数を用いた非線形モデルの探索であり、その線形関数の推測法が研究されている。後者は複数個の潜在変数を想定した非線形モデルである。この問題において、線形関数の個数、潜在変数の数を決める問題が重要であり、モデル選択の観点からのアプローチを発展させることが望まれている。

(5) 近年、電子化された測定技術の進歩により科学の諸分野で、大量なデータの蓄積が可能になっている。この結果、大規模なデータの統計解析や、標本数より変数の数の方が大きい高次元小標本データの解析などが重要になっている。このため、高次元推測法、および、高次元漸近理論の発展が望まれている。

2. 研究の目的

多変量データ分析の新展開として、次の5つの問題を取り上げて研究する。

(1) 数量化法は記述的な統計的方法と位置付けられているが、数量化法において変数の冗長性や選択のための推測法を発展させる。

(2) 多変量正準分析法において、選ばれた正準変数に対して、変数の冗長性や選択のため

の推測法を発展させる。

(3) 個体変動を取り入れた経時データモデルなどにおいて、各個体の分析に焦点を当てた分析法を発展させる。

(4) 多変量の非線形構造を抽出するための推測理論を深化させるとともに、新たな非線形構造を開発する。

(5) 高次元推測法および高次元漸近理論を発展させる。

3. 研究の方法

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、多変量データ推測法の新展開及び総括（研究代表者：藤越）、モデル選択基準の構成及び数値的検証（研究分担者：柳原）とする。次の5つの研究課題を取り上げ、研究を進める。

(1) 数量化2類において、追加情報仮説の検定統計量を、回帰的アプローチおよび判別得点法から導入し、その近似分布を提案する。この有効性を数値的かつ理論的に検討し、具体的な分析への適用を試みる。この検定をもとにした変数選択問題に取り組み、実データへの応用も試みる。さらに、数量化2類での成果を数量化3類などに発展させる。

(2) 主成分分析、正準判別分析、正準相関分析において、選択された少数個の正準変数に関して、もとの変数の選択問題を研究する。逐次法を発展させるための追加情報の検定、モデル選択基準の開発に焦点を当てて研究する。さらに、新たなモデル選択基準を開発し、逐次法との比較を試みる。

(3) 線形混合モデルにおいて考察されている条件付AICを取り上げる。個体内共分散行列と誤差分散が未知の場合に、分散パラメータの簡便的推定法を提案して問題の解決を試みる。また、周辺尤度に基づくAIC基準との比較について検討する。さらに、個体効果共分散構造モデルのもとで、条件付尤度に基づく推測法を発展させる。

(4) 層別逆回帰法と部分線形最小二乗法における次元の推測問題について、モデル選択の立場からの推定法の導入を検討する。計算機手法に基づく推定法と比較する。さらに、平均構造と共分散構造の両方を取り入れた

多変量構造を導入し、推測法とその応用を発展させる。

(5)多変量解析の伝統的手法である、主成分分析、判別分析、多変量検定、正準相関分析、経時データ解析、潜在変数モデルなどにおいて、高次元多変量推測法を発展させる。さらに、非正規の場合の高次元推測法を発展させる。

3. 研究成果

(1)数量化2類において、多変量正規モデルのもとで発展している推測法を適用することの妥当性について、理論的並びに実験的に検証した。これらの結果の1部は、出版した図書にも紹介している。数量化法の利用が高まることが期待される。

(2)モデル選択問題に関して、 C_p 規準は標準化した予測2乗誤差の推定量として構成されたものであるが、本研究では、基準化しない予測2乗誤差の推定量としての基準を提案し、その特徴を明らかにした。また、多変量回帰モデルへの拡張も与えている。これらの方法と従来の方法との使い分けに対して、新たな問題を投げかけており、さらなる研究が望まれる。

(3)高次元の誤差限界問題に関して、一様構造検定統計量の分布に関して、高次元近似を導出し、その近似の誤差限界を与えている。誤差限界の導出を高次元漸近的枠組みで与えた論文は少ない。

(4)経時データ分析問題に関連して、ランダム係数モデルにおける線形仮説の検定統計量の高次元漸近分布を導出した。この結果より、観測時点が大きい場合にも適用できる検定法が可能になった。プロファイル分析法は通常の MANOVA モデルで発展しているが、これらの手法を成長曲線モデルに拡張した。

(5)モデル選択規準は、リスクの推定量として構成されるが、その際バイアス項を減らすことが重要になる。多群の判別問題において、AIC 規準の高次元の枠組みでの修正基準を導出し、その基準のよさを数値実験により確認した。このような研究が他の多変量モデルでも展開されることが期待される。また、cross-validation 規準の高次バイアス補正を、バイアスの漸近展開式を使用することなく、cross-validation 法により構成した。さらに、多項ロジット回帰モデルにおける変数選択のための AIC のバイアスを補正した Bias-corrected AIC (CAIC)を提案した。

(6)正準相関分析のアプローチとして、各組

の変数を主成分に置き換えて、正準相関分析を行う主成分正準相関分析法が提案されている。この方法における主成分の選択法を提案した。また、主成分分析において、何個の主成分を用いればよいかに関係して、寄与率の分布が重要となるが、スパイクモデルのもとで、寄与率の高次元漸近分布を導出した。これらは、高次元の場合に有用な結果である。

(7)この他、学会や国際研究集会において発表した結果、準備中の論文、投稿中の論文が多くある。これらの中には、AIC などのモデル選択規準が高次元の枠組みのもとで一致性をもつという、斬新な結果もある。このような結果は、多変量回帰モデル、成長曲線モデル、判別分析、などにおいても展開し、大きなインパクトを与えている。

4. 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)

- ① Sakurai Tetsuro、Nakata Takashi、Fujikoshi Yasunori、High-dimensional AICs for selection of variables in discriminant analysis、Sankhya、査読有、75 巻、pp.1-25、2013
DOI: 10.1007/s13171-013-0025-0
- ② Ogura Toru、Fujikoshi Yasunori、Sugiyama Takakazu、A variable selection criteria for two sets of principal component scores in principal canonical correlation analysis、Communications in Statistics=Theory and Methods、査読有、42 巻、pp.1-18、2013
DOI: 10.1080/0360926.2011.605235
- ③ Yanagihara Hirokazu、他 3 名、Bias-corrected AIC for selecting variables in multinomial logistic regression models、Linear Algebra and Its Applications、査読有、436 巻 11 号、pp. 4329-4341、2012
- ④ Yanagihara Hirokazu、Fujisawa Hironori、Iterative bias correction of the cross-validation criterion、Scandinavian Journal of Statistics、査読有、39 巻、pp.116-130、2012
DOI: 10.1111/j.1467-9469.2011.00754
- ⑤ Seo Takashi、Sakurai Tomoko、Fujikoshi Yasunori、LR tests for two hypotheses in profile analysis of growth curve data、SUT Journal of Mathematics、査読有、47 巻、pp.105-118、2011
- ⑥ Sakurai Tetsuro、Igeta Masataka、

Fujikoshi Yasunori, High-dimensional asymptotic distribution of test statistic for multivariate linear hypothesis in random-coefficient model, Journal of Combinatorics, Information and System Sciences、査読有、36 巻、pp.49-62、2011

- ⑦ Fujikoshi Yasunori、他 3 名、Prediction error criterion for selecting variables in a linear regression model、Ann. Inst. Statist. Math.、査読有、63 巻、pp.387 - 403、2011
- ⑧ Kato Nohiro, Yamada Tomoyuki, Fujikoshi Yasunori、High-dimensional asymptotic expansion of LR statistic for testing the intraclass correlation structure and its error bound、Journal of Multivariate Analysis、査読有、101 巻、pp.101-112、2010

[学会発表] (計 6 件)

- ① 藤越康祝、高次元モデル選択規準、第 7 回統計学会春季集会、2013 年 3 月 3 日、学習院大学
- ② Fujikoshi Yasunori、On the consistency of model selection criteria under a high dimensional Framework、International Conference on Advances in Interdisciplinary Statistics and Combinatorics、2012 年 10 月 5-7 日、ノースカロライナ大学グリーンズボロ校、米国
- ③ Fujikoshi Yasunori、High-dimensional approximations of the characteristic roots and vectors in discriminant analysis and canonical correlation analysis、第 2 回環太平洋地域数理統計会議、2012 年 7 月 1 日~4 日、筑波国際会議場
- ④ 藤越康祝、判別分析および正準相関分析における固有値と固有ベクトルの高次元漸近分布、2011 年度統計関連学会連合大会、2011 年 9 月 5 日、九州大学
- ⑤ Yanagihara Hirokazu, A comparison of three estimators in GMANOVA model when the sample size is fewer than the dimension、The 9th Tartu Conference on Multivariate Statistics & The 20th International Workshop on Matrices and Statistics、2011 年 6 月 29 日、Tartu University, Estonia
- ⑥ 藤越康祝、正規対称行列変数に関する推測統計量の漸近的精密化、2010 年度統計関連学会連合大会、2010 年 9 月 7 日、早稲田大学

[図書] (計 3 件)

- ① 藤越康祝、杉山高一、多変量モデルの選択、朝倉書店、2012、208
- ② 吉本敦、加茂憲一、柳原宏和、R により環境データの統計分析、朝倉書店、2012、202
- ③ 菅民郎、藤越康祝、質的データの判別分析：数量化 2 類、現代数学社、2011、264

[その他]

ホームページ等

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

(3) 連携研究者

()
研究者番号：