

## 論文審査の要旨

博士の専攻分野の名称	博 士 ( 理 学 )	氏名	福井 敬祐
学位授与の要件	学位規則第4条第①・2項該当		
<p>論文題目 Consistency of log-likelihood-based information criteria for selecting variables in high-dimensional canonical correlation analysis under nonnormality (非正規性の下での高次元正準相関分析における変数選択のための対数尤度関数に基づく情報量規準の一致性)</p>			
論文審査担当者			
主 査	准教授	柳原 宏和	
審査委員	教 授	若木 宏文	
審査委員	教 授	井上 昭彦	
審査委員	教 授	田丸 博士	
<p>〔論文審査の要旨〕</p> <p>本論文では、<math>p</math>次元ベクトル <math>\mathbf{y} = (y_1, \dots, y_p)'</math> と <math>q</math>次元ベクトル <math>\mathbf{x} = (x_1, \dots, x_q)'</math> の間の相関を調べる手法である正準相関分析 (Canonical Correlation Analysis: CCA) において、<math>x_1, \dots, x_q</math> の内の有効でない変数である冗長な変数を取り除く、つまり有効な変数を選ぶ変数選択問題を取り扱っている。そのような変数選択問題において最も広くもちいられている手法として、Akaike (1973) により提案された赤池情報量規準 (Akaike's Information Criterion: AIC) に代表される情報量規準の最小化により最適な変数を選ぶ、情報量規準最小化法がある。Fujikoshi (1985) により、CCA における変数選択問題は、<math>\mathbf{x}</math> と <math>\mathbf{y}</math> を並べた <math>(q+p)</math>次元ベクトル <math>\mathbf{z} = (\mathbf{x}', \mathbf{y}')'</math> の分散共分散行列の構造選択問題に帰着することが報告された。特に、<math>\mathbf{z}</math> の確率分布として正規分布を仮定すれば、<math>\mathbf{z}</math> から <math>n</math> 個の独立標本ベクトル <math>\mathbf{z}_1, \dots, \mathbf{z}_n</math> の標本共分散行列 <math>\mathbf{S}</math> に <math>n-1</math> をかけた統計量がウィッシュャート分布に従うことから、CCA における情報量規準として、ウィッシュャート分布の確率密度関数に基づく <math>\mathbf{S}</math> のカルバックライブラー情報量にモデルの複雑さに対する罰則項を加えた統計量がもちいられる。そのような規準量を Log-Likelihood-Based Information Criterion (LLBIC) と呼ぶ。LLBIC の族は、AIC、バイアス補正 AIC (AIC<sub>c</sub>; Fujikoshi, 1985), Bayesian Information Criterion (BIC; Schwarz, 1978), Consistent AIC (CAIC; Bozdogan, 1987), Hannan-Quinn Information Criterion (HQC; Hanna &amp; Quinn, 1979) などを含む広いものである。情報量規準最小化法による変数選択において重要な特性の一つに、一致性、つまり真の変数に変数選択法により最適な変数として選ばれる確率が漸近的に 1 になるという特性が挙げられる。Nishii <i>et al.</i> (1988) により、標本数 <math>n</math> を <math>\infty</math> とする大標本漸近理論において、AIC, AIC<sub>c</sub> を用いた変数選択法は一致性を持たず、BIC, CAIC, HQC を用いた変数選択法は一致性を持つことが報告された。しかしながら、近年のコンピュータの発達により、蓄積・解析できるデータ数は増え、<math>p</math> が大きい、いわゆる高次元データの解析が必要となってきた。高次元データに対して大標本漸近理論を適用する</p>			

と漸近近似が悪くなるため、高次元データには標本数だけでなく次元数  $p$  も  $\infty$  とする高次元大標本漸近理論により統計量の分布を評価した方がよいことが知られている。 $p/n \rightarrow c_0 \in [0,1)$  となる条件の下での高次元大標本漸近理論の枠組みでは、Yanagihara *et al.* (2014) により AIC, AIC<sub>c</sub> をもちいた変数選択法が一致性を持つことがあり、逆に BIC, CAIC, HQC をもちいた変数選択法は一致性を持たないことがあることが示された。以上の結果は  $z$  の真の分布に正規分布を仮定して得られたものであるが、真の分布が正規分布であるかどうかは誰にもわからない。そこで真の分布に対する正規性の仮定を外した下での一致性の議論が重要となる。多変量線形回帰 (Multivariate Linear Regression: MLR) モデルにおける変数選択において、Yanagihara (2013) で高次元大標本漸近理論の枠組みで一致性の評価を行っている。本論文では Yanagihara (2013) で行っている証明のアイデアを CCA における変数選択問題に適用し、CCA での変数選択において、真の分布に対する正規性の仮定を外した下で LLBIC をもちいた変数選択法が一致性を持つための罰則項の十分条件を得ている。

以上のように、本論文では、情報量規準が一致性を持つための LLBIC の罰則項に関する十分条件を導出しているが、高次元大標本漸近理論の枠組みでの一致性に関する研究は最先端の研究であり、参考となる先行研究もあまりない状況で結果を得ることができたことは、評価に値する点である。また、従来の高次元大標本漸近理論の枠組みでの一致性の研究では、観測値ベクトルが従う真の分布に正規性を仮定することで、ウィッシュャート分布に従う確率変数行列の行列式の収束問題に帰着させ、その収束問題をウィッシュャート分布の特性により解決することにより一致性の評価を行っている。ところが、観測値ベクトルの真の分布に正規性を仮定しないときには、先のウィッシュャート分布に従う確率変数行列の特性を使用することができず、問題は一気に難しくなる。本論文では、Yanagihara (2013) の手法を応用しているが、MLR モデルでは  $x$  を非確率変数とし取り扱っていることに対し、CCA では確率変数として取り扱われるため、Yanagihara (2013) の手法を拡張する必要があった。その拡張に本論文の独創性がある。

以上のことから、審査の結果、本論文の著者は博士(理学)の学位を授与される十分な資格があるものと認める。

主論文:

Consistency of log-likelihood-based information criteria for selecting variables  
in high-dimensional canonical correlation analysis under nonnormality

(非正規性の下での高次元正準相関分析における変数選択のための対数尤度関数に基づく  
情報量規準の一致性)

*Hiroshima Mathematical Journal* に掲載予定.

参考論文:

- (1) Choosing the number of repetitions in the multiple plug-in optimization method for the ridge parameters in multivariate generalized ridge regression.  
Isamu Nagai, Keisuke Fukui and Hirokazu Yanagihara,  
*Bulletin of Informatics and Cybernetics*, Vol.45, 2013, 25-35.
- (2) Selecting a shrinkage parameter in structural equation modeling with a near singular covariance matrix by the GIC minimization method.  
Ami Kamada, Hirokazu Yanagihara, Hirofumi Wakaki and Keisuke Fukui,  
*Hiroshima Mathematical Journal*, Vol.44 (3), 2014, 315-326.
- (3) Comparison with RSS-based model selection criteria for selecting growth functions.  
Keisuke Fukui, Mariko Yamamura and Hirokazu Yanagihara,  
*FORMATH*, Vol.14 に掲載予定.