

論文審査の要旨

博士の専攻分野の名称	博 士 (理 学)	氏名	小田 凌也
学位授与の要件	学位規則第 4 条第 ①・② 項該当		
論文題目	Consistent variable selection criteria in multivariate linear regression even when dimension exceeds sample size (次元数が標本数を超えるときでも多変量線形回帰において一致性をもつ 変数選択規準)		
論文審査担当者	主 査 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		
[論文審査の要旨] 本論文では、複数の予測対象である変数(目的変数)と複数の目的変数に影響を与えると考えられる変数(説明変数)の関係を記述する多変量解析の手法の 1 つである多変量線形回帰において、有効な説明変数を変数選択規準により選択する変数選択問題を取り扱っている。変数選択規準の望ましい性質に、変数選択規準により選ばれた最適な説明変数の組み合わせが真の説明変数となる確率(選択確率)が漸的に 1 となる一致性がある。これまで多くの先行研究により、標本数 n のみを無限大とする大標本漸近理論の下で様々な変数選択規準の一致性が調べられてきた。一方、近年では、標本数 n のみでなく目的変数の個数 p も大きな高次元データを解析する機会が増加している。しかし、このような高次元データでは大標本漸近理論による一致性の評価は妥当でなく、有限標本下においてその選択確率が低くなってしまいう可能性がある。したがって、 n のみでなく p も無限大とする漸近理論、特に p が無限大の場合とそうでない場合の両方を統一的に扱うことが可能な、 $n \rightarrow \infty$, $p/n \rightarrow c \in [0,1)$ となる高次元漸近理論により一致性を評価する必要がある。多変量線形回帰において広く用いられている変数選択規準として、標本共分散行列の逆行列により基準化された多変量残差平方和に任意の正数 α で調整したモデルの複雑さを表す項 “ $\alpha \times$ (平均構造のパラメータ数)” を加えることで定義される一般化 C_p 規準(Generalized C_p ; GC_p)規準 (Atkinson, 1980; Nagai <i>et al.</i> , 2012)がある。Yanagihara (2016)では、真の誤差分布が多変量正規分布であるという仮定の下で、 GC_p 規準が一致性を満たすために必要な α の条件を高次元漸近理論の下で導出し、その条件を用いて高次元下でも一致性が保たれる GC_p 規準を提案している。さらにその提案された GC_p 規準は、真の誤差分布の正規性が崩れた場合でも一致性が保持されることが柳原(2019)により示されている。一方、高次元データの中でも p が n を超えるようなデータも多く見られるようになってきている。しかし、 $p > n$ のときは標本共分散行列が特異となるため、 GC_p 規準が定義できない。そこで、本論文では、 $p > n$ のときでも計算可能な変数選択規準として、特定の重み行列を伴う重み付き多変量残差平方和に GC_p 規準でも用いているモデルの複雑さを表す項を足し合わせた変数選択規準のクラスを考え、重み行列としてスカラー行列を用いて定義されるスカラー型一般化 C_p			

(Scalar-type GC_p ; SGC_p)規準, リッジパラメータ λ によるリッジ型標本共分散行列を用いて定義されるリッジ型一般化 C_p (Ridge-type GC_p ; RGC_p) 規準を提案している. このとき, SGC_p 規準, RGC_p 規準が一致性をもつために必要な α の条件を真の誤差分布が多変量正規分布でないという条件の下で導出している. ここでは, $p > n$ の場合とそうでない場合の両方を統一的に取り扱うことを可能とするために, $n \rightarrow \infty$, $p/n \rightarrow c \in [0, \infty]$ という高次元漸近理論を用いている.

以上のように, 本論文では, 変数選択規準が一致性を持つために必要な α の条件を導出しているが, $p > n$ を許容する高次元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. 特に, リッジ型標本共分散行列を用いた変数選択規準において高次元漸近理論の下で一致性を評価できている研究結果を本論文以外にまだ見たことがない. また, 従来の高次元漸近理論の枠組みでの一致性の研究では, 真の誤差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うことが多い. ところが, 真の誤差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. 真の誤差分布に正規性を仮定しない場合での一致性という挑戦的な研究テーマに取り組み, 解決している点も, 評価に値する点である.

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：

Consistent variable selection criteria in multivariate linear regression even when dimension exceeds sample size

(次元数が標本数を超えるときでも多変量線形回帰において一致性をもつ変数選択規準)

Hiroshima Mathematical Journal, 2020 年, Vol.50 No.2 または Vol.50 No.3 掲載予定

参考論文：

- (1) High-dimensional asymptotic behaviors of the differences between the log-determinants of two Wishart matrices.
Yanagihara, H., Oda, R., Hashiyama, Y. and Fujikoshi, Y.
Journal of Multivariate Analysis, Vol. 157 (2017) 70-86.
- (2) Asymptotic null and non-null distributions of test statistics for redundancy in high-dimensional canonical correlation analysis.
Oda, R., Yanagihara, H. and Fujikoshi, Y.
Random Matrices: Theory and Applications, 1950001 (2019) 26pp.
- (3) Strong consistency of log-likelihood-based information criterion in high-dimensional canonical correlation analysis.
Oda, R., Yanagihara, H. and Fujikoshi, Y.
Sankhya A, (2019) 19pp, <https://doi.org/10.1007/s13171-019-00174-3>.
- (4) A consistent variable selection method in high-dimensional canonical discriminant analysis.
Oda, R., Suzuki, Y., Yanagihara, H. and Fujikoshi, Y.
Journal of Multivariate Analysis, Vol. 175, 104561 (2020) 13pp.
- (5) A high-dimensional bias corrected AIC for selecting response variables in multivariate calibration.
Oda, R., Mima, Y., Yanagihara, H. and Fujikoshi, Y.
Communications in Statistics-Theory and Methods, (2020) 24pp,
<https://doi.org/10.1080/03610926.2019.1705978>.