

高次元多変量モデルにおける非正規下での変数選択法の一致性

広島大・理 小田 凌也 柳原 宏和

本報告では、目的変数が複数個ある場合の多変量線形回帰モデルにおいて、有効な説明変数を選ぶ変数選択問題を扱う。そのような変数選択問題における変数選択法の代表例としては、AIC や C_p 規準などの変数選択規準を用いた変数選択法が挙げられる。真の説明変数の組み合わせが最適な変数の組み合わせとして選ばれる確率（選択確率）が漸近的に 1 となる性質、即ち、一致性は変数選択法に望まれる性質の 1 つとして挙げられ、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枠組みの下で様々な変数選択規準を用いた変数選択法の一致性が調べられてきた。さらに近年では、標本数 n のみでなく目的変数ベクトルの次元数 p もしくは説明変数ベクトルの次元数 k も大きな高次元データを扱う重要性が高まっており、そのような高次元データに対して大標本漸近枠組みを用いた一致性の評価は妥当でない。そのため、近年の先行研究（例えば、[1], [2] 参照）では、 n は必ず無限大であり $p+k$ は n より小さいが p, k は無限大でもそうでなくてもよいという以下のような漸近枠組みの下で一致性を持つ変数選択規準を用いた変数選択法が提案されている。

$$n \rightarrow \infty, (p+k)/n \rightarrow c_0 \in [0, 1).$$

しかしながら、これまでに提案されている変数選択規準を用いた変数選択法の多くは、真のモデルにおける誤差ベクトルに正規性を仮定しており、さらに標本共分散行列の逆行列を用いているため目的変数ベクトルの次元数 p は標本数 n よりも小さい必要がある。そこで本報告では、そのどちらの条件も緩めることを考え、非正規かつ $p > n$ のときでも一致性を持つ変数選択規準を用いた変数選択法を提案する。そのためにまず、 C_p 規準の罰則項の大きさを正数 α により調整した一般化 C_p (Generalized C_p ; GC_p) 規準において、基準化に用いた標本共分散行列をリッジパラメータ λ を用いたスカラー行列を足し合わせた行列に置き換えたリッジ型標本共分散行列により基準化した GC_p (Ridge-type GC_p ; RGC_p) 規準を考える。さらに、 RGC_p 規準を用い、[3] で提案された以下のような簡便な変数選択法を適用する。

$$\hat{j} = \{l \in \omega \mid RGC_p(\omega \setminus \{l\} | \alpha, \lambda) > RGC_p(\omega | \alpha, \lambda)\}. \quad (1)$$

ただし、 $\omega = \{1, \dots, k\}$ は全ての説明変数の添え字を要素に持つ集合、 $RGC_p(j | \alpha, \lambda)$ は ω の部分集合 j に対するパラメータ α, λ を持つ RGC_p 規準を表す。このとき、真のパラメータに関する適当な仮定の下、以下のような大標本・高次元漸近枠組みを用いて、(1) が一致性を持つための α, λ に関する十分条件を導出する。

$$n \rightarrow \infty, p/n \rightarrow c_1 \in [0, \infty], k/n \rightarrow c_2 \in [0, 1).$$

ここで、 $c_1 = \infty$ は $p/n \rightarrow \infty$ を意味する。報告当日は、十分条件を満たす α, λ を用いた提案法が非正規の下でも高い選択確率を持つことを数値的に示す。

引用文献：

- [1] Bai, Z. D., Fujikoshi, Y. & Hu, J. (2018). Strong consistency of the AIC, BIC, C_p and KOO methods in high-dimensional multivariate linear regression. TR No. 18–9, *Statistical Research Group*, Hiroshima University.
- [2] Oda, R. & Yanagihara, H. (2019). A fast and consistent variable selection method for high-dimensional multivariate linear regression with a large number of explanatory variables. TR No. 19–1, *Statistical Research Group*, Hiroshima University.
- [3] Zhao, L. C., Krishnaiah, P. R. & Bai, Z. D. (1986). On detection of the number of signals in presence of white noise. *J. Multivariate Anal.*, **20**, 1–25.