# 大標本・高次元漸近理論による情報量規準の一致性の評価について

広島大学　大学院理学研究科　柳原 宏和

正規性を仮定した多変量線形回帰モデルにおいて, Nishii [2] により提案された $-2\times$ (最大対数尤度) にモデルのパラメータ数の $\alpha$ 倍 ($\alpha > 0$) を加えることで定義される一般化情報量規準 (Generalized Information Criterion; GIC) や, Atkinson [1] により提案されたすべての説明変数を用いたフルモデルで推定した分散共分散行列により基準化された最小残差平方和に平均に関するパラメータ数の $\alpha$ 倍 ($\alpha > 0$) を加えることで定義される一般化 $C_p$ (Generalized $C_p$; $GC_p$) 規準の最小化により最適な変数を選ぶ変数選択法を取り扱う. 変数選択法の望ましい性質の一つとして, 真の変数の組み合わせが最適な変数として選ばれる確率が漸近的に 1 となる性質, 即ち, 一致性がある. 一致性は, 多くの場合, 標本数 $n$ のみを無限大とする漸近理論である, 大標本漸近理論により評価されている. 大標本漸近理論により評価すれば, 以下の条件を満たせば GIC, $GC_p$ ともに一致性を持つことになる.

$$\lim_{n\to\infty} \alpha = \infty, \quad \lim_{n\to\infty} \frac{\alpha}{n} = 0.$$

一方で, 近年, ハードウェアの発展により, 蓄積・解析できるデータの数が爆発的に増大し, 目的変数の次元数 $p$ が大きいデータである, 高次元データの解析の需要が高まっている. 本論文で取り扱う高次元データとは, 次元数 $p$ は大きいが標本数 $n$ よりも小さいとする適度な高次元 (moderately high-dimensional) データ (Yao *et al.* [3] 参照) である. このような高次元データでは, 大標本漸近理論ではなく, 次元数 $p$ も $p/n$ が 1 未満の定数に収束するという条件の下で $n$ と共に無限大となる漸近理論により一致性を評価した方が妥当である. しかしながら, $p$ はいつでも大きいわけではない. そこで, $p$ が小さいときでも対応できるように, 以下のような漸近枠組みによる大標本・高次元漸近理論により一致性を評価することを考える.

$$n \to \infty, \quad p/n \to c_0 \in [0, 1). \tag{1}$$

上記の漸近枠組みにおいて, 次元数 $p$ は有限でも無限でもどちらでもよいため, 大標本漸近理論を特別な形として含むものになっている. ここで, 以下のような規準により異なる定数 $m$ を定義する.

$$m = \begin{cases} -\dfrac{n}{p}\log\left(1-\dfrac{p}{n}\right) & (\text{GIC の場合}) \\ \dfrac{n}{n-p} & (GC_p \text{ の場合}) \end{cases}$$

このとき, (1) 式の大標本・高次元漸近理論により評価すれば, 説明変数行列と真の分布に関する適当な仮定の下, 以下の条件を満たせば GIC, $GC_p$ が一致性を持つことになる.

$$\alpha = m + \beta, \quad \beta > 0 \quad s.t. \lim_{n\to\infty, p/n\to c_0} \sqrt{p}\beta = \infty, \lim_{n\to\infty, p/n\to c_0} \frac{p}{n}\beta = \infty. \tag{2}$$

(2) 式は, 真のモデルの分布に正規性を仮定することなく導出することができる. 本発表では, (1) 式に基づく大標本・高次元漸近理論による GIC と $GC_p$ の一致性の評価方法と実際に (2) 式がどのように導出されるかを紹介する.

## 引用文献：


[1] Atkinson, A. C. (1980). A note on the generalized information criterion for choice of a model. *Biometrika*, **67**, 413–18.

[2] Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.*, **12**, 758–765.

[3] Yao, J., Zheng, S. & Bai, Z. (2015). *Large Sample Covariance Matrices and High-dimensional Data Analysis.* Cambridge University Press, New York.