非正規多変量回帰モデルにおける検定統計量および固有値の高次元かつ大標本のもとでの漸近分布

広島大・理・名誉教授　藤越　康祝

本研究では, 高次元多変量回帰モデルを考えるが, 目的変数の数 p は大きいが標本サイズ n より小さい場合を扱う．このとき, 説明変数のいくつかが有意であるかどうかを調べる検定統計量, および, 関連する固有値の漸近分布の導出に関心がある．ここでは, 高次元かつ大標本漸近的枠組, すなわち, p/n がある一定値 $c \in (0,1)$ に近づくもとでの漸近分布を考える．基礎分布が正規分布の場合は Wakaki, Fujikoshi and Ulyanov (2014) により導出されている．また, MANOVA モデルの場合には, Bai, Choi and Fujikoshi (2017) によって, ランダム行列論を利用して, 基礎分布が非正規の場合にも導出されている．本報告では, このような非正規のもとでの結果が多変量回帰モデルの場合に拡張できることを示す．このときの漸近分布は, 正規分布の場合と同じになり, 高次元かつ大標本においてロバスト性をもつことが指摘される．

p 個の目的変数をもつ大きさ n の観測行列を Y とし, 説明変数が $(1+q_1+q_2)$ 個の多変量回帰モデルは

$$\mathrm{E}(\mathsf{Y}) = \mathbf{1}_n\boldsymbol{\theta}' + \mathbf{A}_1\boldsymbol{\Theta} + \mathbf{A}\boldsymbol{\Theta}_2$$

と表せる．ここに, $\mathbf{1}_n$ は成分がすべて 1 の n 次元列ベクトル, $\mathbf{A}_i, i = 1,2$ は $n \times q_i$ の既知行列, $\mathrm{rank}(\mathbf{1}_n, \mathbf{A}_1, \mathbf{A}_2) = 1 + q_1 + q_2$, $\boldsymbol{\theta}$ は $p \times 1$ のパラメータベクトル, $\Theta_i, i = 1,2$ は $q_i \times p$ の未知パラメータ行列である．仮説 $\boldsymbol{\Theta}_2 = \mathbf{O}$ の検定に対する尤度比統計量の分布を扱う場合, 一般性を失うことなく, $\mathbf{A}_i'\mathbf{A}_i = \mathbf{I}_{q_i}$, $\mathbf{A}_i'\mathbf{1}_n = \mathbf{0}$, $\mathbf{A}_1'\mathbf{A}_2 = \mathbf{O}$, $i = 1,2$ としてよい．このとき, 仮説による平方和積和行列 S_h, 誤差平方和積和行列 S_e, 全平方和積和行列 S_t は次のように表せる．

$$\mathsf{S}_h = \mathsf{Y}'\mathbf{A}_2\mathbf{A}_2'\mathsf{Y}, \quad \mathsf{S}_e = \mathsf{Y}(\mathbf{I}_n - \mathbf{P}_0 - \mathbf{A}_1\mathbf{A}_1' - \mathbf{A}_2\mathbf{A}_2')\mathsf{Y}, \quad \mathsf{S}_t = \mathsf{S}_h + \mathsf{S}_e.$$

ここに, $\mathbf{P}_0 = (1/n)\mathbf{1}_n\mathbf{1}_n'$. 統計量 $|\mathsf{S}_e|/|\mathsf{S}_t|$ などの仮説のもとでの高次元漸近分布を考えるとき, $\mathrm{E}(\mathsf{Y}) = \mathbf{O}$, Y の各行は独立同一分布としてよい．Y の各行の共分散行列は $\mathbf{I}_p$ としてよいが, 各行の成分は, 独立で一様な 4 次のモーメントをもつと仮定する．このとき, 高次元漸近的枠組のもとで次が成り立つ．

$$(\mathsf{Y}'\mathbf{A}_2)'\mathsf{S}_t^{-1}\mathsf{Y}'\mathbf{A}_2 = \frac{p}{n}\mathbf{I}_{q_2} + \frac{1}{\sqrt{n}}\mathsf{W}_{22} + \mathrm{o}(n^{-1/2}).$$

ここに, W_{22} は $q_2 \times q_2$ の対称な確率行列で, 異なる要素は互いに独立で, 対角要素は $\mathrm{N}(0,2)$ に従い, 非対角要素は $\mathrm{N}(0,1)$ に従う確率変数である．

上記の摂動展開を利用して, $|\mathsf{S}_e|/|\mathsf{S}_t|$, $\mathrm{tr}\mathsf{S}_h\mathsf{S}_e^{-1}$, $\mathrm{tr}\mathsf{S}_h\mathsf{S}_t^{-1}$ の仮説のもとでの高次元漸近分布が非正規のもとで導出される．また, $\mathsf{S}_h\mathsf{S}_e^{-1}$ の固有値の高次元漸近分布を求めることができる．さらに, これらの統計量は, 高次元かつ大標本において, ロバスト性をもつことを指摘する．

参考文献

1. Bai, Z. D., Choi, C. P. and Fujikoshi, Y. (2017). Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT. Submitted.

2. Wakaki, H., Fujikoshi, Y. and Ulyanov, V. V. (2014). Asymptotic expansions of the distributions of MANOVA tetst statistics when the dimension is large. *Hiroshima Math. J.*, **44**, 247–259.