GMANOVA モデルでの仮定緩和のさらなる可能性

中京大学 教養教育研究院 永井 勇

n 個の各個体に対して, 全ての個体で測定時点を揃った状態で p 回測定して得られる経時測定データの分析を本講演では考える．このようなデータの分析の主な目的は, データに潜む経時変動と呼ばれる時間的な変動を上手く捉えることである．そのために, Pothoff and Roy (1964) で提案された次の一般化多変量分散分析 (GMANOVA) モデルがよく用いられる;

$$\boldsymbol{Y} = \boldsymbol{1}_n \boldsymbol{\mu}' \boldsymbol{X}' + \boldsymbol{A} \Xi \boldsymbol{X}' + \boldsymbol{\mathcal{E}}, \tag{1}$$

ここで $\boldsymbol{1}_n$ は全ての成分が 1 からなる n 次元ベクトル, $\boldsymbol{0}_r$ は全ての成分が 0 からなる r 次元ベクトル, $\boldsymbol{Y}$ は各行が各個体で測定して得られた経時測定データからなる $n \times p$ 行列, $\boldsymbol{A}$ は各行が各個体の特徴を表す測定時点に無関係な k 個の変数からなる $n \times k$ 行列とし, $\boldsymbol{A}'\boldsymbol{1}_n = \boldsymbol{0}_k$ (各変数で中心化されている) を満たしているとし, $\boldsymbol{X}$ は後述のように各行が測定時点の関数からなる $p \times q$ 行列であり, これらは既知である．また, $\boldsymbol{\mu}$ は q 次元未知ベクトル, $\boldsymbol{\Xi}$ は $k \times q$ 未知行列であり, $\boldsymbol{\mathcal{E}}$ は $E[\boldsymbol{\mathcal{E}}] = \boldsymbol{0}_n \boldsymbol{0}_p'$, $\mathrm{Cov}[\mathrm{vec}(\boldsymbol{\mathcal{E}})] = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$ の $n \times p$ 誤差行列とし, $\boldsymbol{\Sigma}$ は未知の $p \times p$ 正定値行列とする．ここで, p 回の測定時点を $t_1, \ldots, t_p$ とし, 例えば $\boldsymbol{X}$ の i 行目を $(t_i^0, t_i^1, \ldots, t_i^{q-1})$ とすると, $\boldsymbol{\mu}$ や $\boldsymbol{\Xi}$ は測定時点の $(q-1)$ 次多項式の係数を表しており, $t_1, \ldots, t_p$ の多項式を用いて経時変動の推定をしていることに対応している．また, 過剰適合の問題もあるが, より柔軟な関数を用いて経時変動を推定することも可能である．その際は, 用いる関数の重み付き和で経時変動を推定することとなり, $\boldsymbol{\mu}$ や $\boldsymbol{\Xi}$ は重みに対応している．そのため, GMANOVA モデル (1) において, 未知ベクトル $\boldsymbol{\mu}$ や未知行列 $\boldsymbol{\Xi}$ を推定することで経時変動が推定できる．したがって, この GAMOVA モデルにおいて, これらを上手く推定することが重要である．これらの推定は, 次のリスクを最小にすることでよく行われている;

$$R(\boldsymbol{\mu}, \Xi) = \mathrm{tr}[\{\boldsymbol{Y} - (\boldsymbol{1}_n \boldsymbol{\mu}' \boldsymbol{X}' + \boldsymbol{A} \Xi \boldsymbol{X}')\} \boldsymbol{\Sigma}^{-1} \{\boldsymbol{Y} - (\boldsymbol{1}_n \boldsymbol{\mu}' \boldsymbol{X}' + \boldsymbol{A} \Xi \boldsymbol{X}')\}'].$$

このリスクを最小にする $\boldsymbol{\mu}$, $\boldsymbol{\Xi}$ を $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Xi}}$ とし, $\boldsymbol{A}'\boldsymbol{1}_n = \boldsymbol{0}_k$ に注意すると, $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\hat{\boldsymbol{\mu}} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'\boldsymbol{1}_n/n$, $\boldsymbol{A}'\boldsymbol{A}\hat{\boldsymbol{\Xi}}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} = \boldsymbol{A}'\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$ の解として, それぞれ得られる．これらの解は, $\mathrm{rank}(\boldsymbol{X}) = q$ を仮定すれば $\hat{\boldsymbol{\mu}} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'\boldsymbol{1}_n/n$ となり, さらに $\mathrm{rank}(\boldsymbol{X}) = q$ を追加で仮定すれば $\hat{\boldsymbol{\Xi}} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$ となる．そのため, 従来の分析においては, モデル (1) において $\mathrm{rank}(\boldsymbol{A}) = k$ かつ $\mathrm{rank}(\boldsymbol{X}) = q$ が仮定されている．

一方, モデル (1) で $\boldsymbol{X} = \boldsymbol{I}_p$ としたモデルで, $\mathrm{rank}(\boldsymbol{A}) < k$ で無罰則・反復不要な推定量を永井 (2021) で, 仮定を追加した下での不偏推定量を永井 (2022a) で提案した．これらを拡張し, 本講演ではモデル (1) で $\mathrm{rank}(\boldsymbol{A}) < k$ かつ $\mathrm{rank}(\boldsymbol{X}) = q$ などの下での推定量を提案する．

一方で, $\mathrm{rank}(\boldsymbol{A}) = k$ で $\mathrm{rank}(\boldsymbol{X}) < q$ での推定量を永井 (2022b) で提案し, 同じ状況で仮定を追加した不偏推定量を永井 (2022c) で提案した．本講演では GMANOVA モデル (1) でのこれらの推定量や本講演で提案する推定量などについてまとめ, このモデルにおけるさらなる仮定の緩和の可能性などについても触れる予定である．

引用文献:

[1] Pothoff, R. F. & Roy, S. N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.

[2] 永井 勇 (2021) 高次元小標本における多変量線形回帰モデルでの推定法, 2021 年度統計関連学会連合大会.

[3] 永井 勇 (2022a) 説明変数がランク落ちしている状況での多変量線形回帰における不偏推定量, 2022 年度統計関連学会連合大会.

[4] 永井 勇 (2022b) GMANOVA モデルにおける新たな推定方法とその解釈, . 多様な分野における統計科学の理論とその応用.

[5] 永井 勇 (2022c) GMANOVA モデルにおける新たな推定方法と解釈, 大規模複雑データの理論と方法論～新たな発展と関連分野への応用～.