Extended GMANOVA モデルにおける 罰則付推定量の提案と罰則パラメータの最適化

中京大学 国際教養学部 国際教養学科 永井 勇

各個体に対して経時的に測定して得られるデータ (経時測定データ) は, 様々な分野で収集され分析されている. このようなデータに対する分析の目的の一つは, 平均的な経時変動や共変量ごとの経時変動などを上手く捉えることである. 本講演では, 全ての個体において測定時点が揃っている経時測定データに着目する.

このような経時測定データの分析においては, 一般化多変量分散分析モデル (GMANOVA モデル; Potthoff & Roy, 1964) を用いることで目的である経時変動を推定できるため, このモデルが広く用いられている. このモデルは, 経時変動をどのように捉えるかを一つの行列を用いて表すことで, 経時変動を推定するモデルである (例えば, Nagai (2011) 参照). つまり, このモデルを用いて経時変動を推定することは, 平均的な経時変動および全ての共変量に関する経時変動に対して, 全て共通の次数の多項式などを用いて推定していることを表している. そのため, この GMANOVA モデルを用いた経時変動の推定においては, 例えば, 平均的な経時変動を (q_0-1) 次の多項式を, いくつかの共変量 $\boldsymbol{A}_1$ に対して (q_1-1) 次の多項式を, 別の共変量 $\boldsymbol{A}_2$ に対して (q_2-1) 次の多項式などをそれぞれ用いて全体の経時変動を推定することができない.

そこで本講演では, Kollo and von Rosen (2005) の Definition 4.1.3 に基づいた次の Extended GMANOVA モデルを用いることを考える;

$$\boldsymbol{Y} = \boldsymbol{1}_n\boldsymbol{\mu}'\boldsymbol{X}_0' + \boldsymbol{A}_1\Xi_1\boldsymbol{X}_1' + \boldsymbol{A}_2\Xi_2\boldsymbol{X}_2' + \boldsymbol{\mathcal{E}},$$

ここで, $\boldsymbol{Y}$ は各行が各個体の経時測定データからなる $n\times p$ の既知行列, $\boldsymbol{1}_n$ は全ての成分が 1 の n 次元ベクトル, $\boldsymbol{\mu}$ は q_0 次元未知ベクトル, $\boldsymbol{X}_i$ は $p\times q_i$ 既知行列 $(i=0,1,2;\ q_i\le p)$, $\boldsymbol{A}_i$ は中心化された共変量からなる $n\times k_i$ 既知行列 $(i=1,2;\ k_i\le n)$, Ξ_i は $k_i\times q_i$ 未知行列 $(i=1,2)$, $\boldsymbol{\mathcal{E}}$ は $E[\boldsymbol{\mathcal{E}}]=\boldsymbol{O}_{n,p}$, $\mathrm{Cov}(\mathrm{vec}(\boldsymbol{\mathcal{E}}))=\boldsymbol{\Sigma}\otimes\boldsymbol{I}_n$ の誤差行列, $\boldsymbol{\Sigma}$ は未知の $p\times p$ 正定値行列である. このモデルにおいて, $\boldsymbol{1}_n\boldsymbol{\mu}'\boldsymbol{X}_0'$ が平均的な経時変動を, $\boldsymbol{A}_i\Xi_i\boldsymbol{X}_i'$ が共変量 $\boldsymbol{A}_i$ ごとの経時変動をそれぞれ表している. よって, このモデルを用いて, j 番目の測定時点を t_j $(j=1,\ldots,p)$, $\boldsymbol{X}_i$ の j 行目を $(1,t_j,\ldots,t_j^{q_i-1})$ とし, この $\boldsymbol{X}_i$ と既知の $\boldsymbol{Y}$ と $\boldsymbol{A}_i$ を用いて未知の $\boldsymbol{\mu}$ と Ξ_i を推定することで, 平均的な経時変動やそれぞれの共変量 $\boldsymbol{A}_i$ ごとの経時変動に対して, t_j の (q_i-1) 次多項式を用いて推定できる.

しかし, 測定時点の多項式をいくつか用いるだけでは経時変動を捉えきれない場合がある. そこで, 多項式だけを用いるのではなく, 基底関数などの柔軟な関数を用いて経時変動を推定することが考えられる. そこで本講演では, 上記のモデルにおいて, $\boldsymbol{X}_0$ や $\boldsymbol{X}_1$ を測定時点の (q_0-1) 次および (q_1-1) 次の多項式を用いる項, $\boldsymbol{X}_2$ を柔軟な関数を用いる項として考える. これにより, 平均的な経時変動やいくつかの共変量 $\boldsymbol{A}_1$ に対しては測定時点の多項式を用い, 別の共変量 $\boldsymbol{A}_2$ に対しては柔軟な関数を用いて経時変動が推定できる.

しかしながら, 多項式の部分を単に柔軟な関数に置き換えて従来の手法で未知の $\boldsymbol{\mu}$ や Ξ_i を推定すると, 目的である経時変動ではなく経時測定データへ過剰に適合してしまうという問題が起きる. この問題を回避するために本講演では, 多変量一般化リッジ回帰を用いた推定法 (Yanagihara, Nagai & Satoh, 2009) を適用した罰則付推定量をまず提案する. その後, 提案した罰則付推定量において導入した罰則パラメータを最適化するための C_p 型の情報量規準を構築する. 数値実験による比較などについては当日報告する.

引用文献:

[1] Kollo, T. & von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*, Springer.

[2] Nagai, I. (2011). Modified C_p criterion for optimizing ridge and smooth parameters in the MGR estimator for the nonparametric GMANOVA model. *Open J. Stat.*, **1**, 1–14.

[3] Yanagihara, H., Nagai, I. & Satoh, K. (2009). A bias-corrected C_p criterion for optimizing ridge parameters in multivariate generalized ridge regression. *Jpn. J. Appl. Stat.*, **38**, 151–172 (in Japanese).

[4] Potthoff, R. F. & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 2 –326.