Estimation of varying coefficients for a growth curve model

佐藤健一（広島大学・原爆放射線医科学研究所）, 柳原宏和（広島大学・理学研究科）

経時測定データの解析に有用な成長曲線モデルの紹介をし, このモデルにおいて変化係数を導入した. その推定, 信頼区間の構成および検定について議論した.

1. 成長曲線モデル

個体$i = 1, \cdots, n$の時点$t = t_{i1}, \cdots, t_{ip_i}$での観測値, $y_i(t)$に対して, 時間に関する説明変数$\boldsymbol{x}(t)$, 個体ごとに異なる回帰係数$\boldsymbol{\alpha}_i$を用いて, 次の成長曲線を考える.

$$y_i(t) = \boldsymbol{\alpha}_i'\boldsymbol{x}(t) + \varepsilon_i(t), \tag{1}$$

ここで, $\varepsilon_i(t)$は観測誤差をあらわし, 個体間で独立とする. 時間に関する説明変数は, 例えば, 経時トレンドを多項式曲線で記述するならば, $\boldsymbol{x}(t) = (1, t, \cdots, t^{q-1})'$, と書ける. 個体間で測定時点がすべて等しい場合, すなわち, $p_1 = \cdots = p_n = p$, $t_{1j} = \cdots = t_{nj} = t_j$, $j = 1, \ldots, p$, ならば, 測定時点のデザインはバランス型と呼ばれ, 成り立たないときは, アンバランス型と呼ばれる. ほとんどの個体の測定時点が等しく, 欠測値がある場合もアンバランス型と考えることができる.

次に, 個体iの回帰係数$\boldsymbol{\alpha}_i$が, k個の説明変数$\boldsymbol{a}_i = (a_{i1}, \cdots, a_{ik})'$, によって回帰されるとする. すなわち,

$$\boldsymbol{\alpha}_i = \boldsymbol{\Theta}'\boldsymbol{a}_i, \tag{2}$$

ただし, $\boldsymbol{\Theta}$は未知パラメータ行列. まとめると, 成長曲線モデルは以下のようにかける.

$$y_i(t) = \boldsymbol{a}_i'\boldsymbol{\Theta}\boldsymbol{x}(t) + \varepsilon_i(t). \tag{3}$$

2. 未知パラメータの推定と解釈

未知パラメータ行列$\boldsymbol{\Theta}$の推定量は, バランス型の場合, Potthoff and Roy (1964)によって最尤推定量が求められている. また, 一般化線形モデルを経時測定データに拡張するためにLiang and Zeger (1986)によって提案された一般化推定方程式を用いても推定可能である. アンバランス型の場合には, Laird and Ware (1982)によって変量効果モデルのもとで提案された制限付き最尤推定量として得られ, 特に, 経時データに特化する形で成長曲線モデルに対して, Vonesh and Carter (1987)が推定量を提案している. このうち, Liang and Zeger (1986)およびLaird and Ware (1982)による推定量は繰り返し計算を必要とする. アンバランス型の経時データで提案された推定量は

バランス型のデータにも適用可能であり, 実用的にはVonesh and Carter (1987) の推定量を使えばよい. それぞれのモデルごとに観測誤差に仮定された分散共分散構造は異なるが, 推定量$\hat{\boldsymbol{\Theta}}$と漸近分布$\hat{\boldsymbol{\Theta}} \xrightarrow{d} N_{k\times q}(\boldsymbol{\Theta}, \boldsymbol{\Omega})$ が得られる.

したがって, 未知パラメータ行列の成分ごとに, 推定値, 標準誤差および有意性などが検討可能である. ところが, 解析者にとって, この推定値の解釈は以下の理由により大きな負担となっており, 成長曲線モデルを利用する上で障害になっていると思われる. 1) 各成分が説明変数だけでなく, 時間共変量の交互作用に対応しているため直感的に理解しにくい, 2) パラメータ数は経時トレンドに関するデザインの次数qと説明変数の個数kの積, kqとなり, それぞれの標準誤差, 有意性などを検討するには, 一般的にパラメータが多すぎる, 3) 経時トレンドとしてBスプライン, ガウス基底などの非線形デザインを仮定した場合に, 各成分の解釈が意味をなさない. それゆえ, 未知パラメータを適当に要約し, 解釈が容易な形にする必要がある.

3. 変化係数の導入

成長曲線モデルの平均構造を, Hastie and Tibshirani (1993) らが提案した時間と共に変化する回帰係数, 変化係数を用いて表現することを考える. 個体ごとの回帰係数$\boldsymbol{\alpha}_i$ではなく, 次のように, $\boldsymbol{\beta}(t) = \boldsymbol{\Theta}\boldsymbol{x}(t)$ に焦点を当てる.

$$\begin{aligned} E[y_i(t)] &= \boldsymbol{\alpha}_i'\boldsymbol{x}(t) \\ &= \boldsymbol{a}_i'\boldsymbol{\Theta}\boldsymbol{x}(t) \\ &= \boldsymbol{a}_i'\boldsymbol{\beta}(t) \\ &= \sum_{j=1}^{k} a_{ij}\beta_j(t). \end{aligned} \tag{4}$$

こうして, $\hat{\boldsymbol{\Theta}}$の漸近分布から, 変化係数の推定量, $\hat{\boldsymbol{\beta}}(t)$ の議論ができる. 変化係数は, 各説明変量に対応する形で与えられ, 説明変数の効果を容易に視覚化でき, 統計的な有意性を検討するうえでも有用である.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser.* **B**, **55**, 757–796.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrika*, **38**, 963–974.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.

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

Vonesh, E. F. and Carter, R. L. (1987). Efficients inference for random-coefficient growth curve models with unbalanced data. *Biometrics*, **43**, 617–628.