繰り返し測定データに基づく非線形混合効果モデリング

野中 美佑 (久留米大学バイオ統計センター)
松井 秀俊 (九州大学大学院数理学府)
小西 貞則 (九州大学大学院数理学研究院)

1　はじめに

現在, 医学の分野において経時的に観測されたデータを解析するための統計的手法が注目されている. 特に, モデルの中に固定効果と変量効果を取り入れた混合効果モデルを用いて統計解析を行う方法は実際に様々な臨床試験データに応用されている. 本研究では, 基底関数展開に基づく非線形混合効果モデル (Rice and Wu (2001)) において, モデルに含まれる未知パラメータの推定法と基底関数の個数を選択するためのモデル評価基準について報告する.

2　基底関数展開に基づく非線形混合効果モデル

本節では, 反復測定データの枠組みにおける混合効果モデルについて説明する. 説明変数 X と目的変数 Y に関して, n 組のデータ $\{(x_{ij}, y_{ij}); i = 1, 2, \cdots, n; j = 1, 2, \cdots, t_i\}$ が観測されたとする. ただし, x_{ij}, y_{ij} はそれぞれ i 番目の個体の第 j 回目の測定における X と Y の実現値とする. いま, データは次のモデルに従って観測されたと仮定する.

$$\begin{aligned} y_{ij} &= m(x_{ij}) + r_i(x_{ij}) + \varepsilon_{ij} \qquad (1) \\ &= \sum_{k=1}^{m_f} \beta_k b_k^{(F)}(x_{ij}) + \sum_{l=1}^{m_r} u_{il} b_l^{(R)}(x_{ij}) + \varepsilon_{ij} = \boldsymbol{\beta}' \boldsymbol{b}^{(F)}(x_{ij}) + \boldsymbol{u}_i' \boldsymbol{b}^{(R)}(x_{ij}) + \varepsilon_{ij}. \end{aligned}$$

ただし, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_{m_f})'$, $\boldsymbol{u}_i = (u_{i1}, \cdots, u_{im_r})'$, $\boldsymbol{b}^{(F)}(x) = (b_1^{(F)}(x), \cdots, b_{m_f}^{(F)}(x))'$, $\boldsymbol{b}^{(R)}(x) = (b_1^{(R)}(x), \cdots, b_{m_r}^{(R)}(x))'$ とする. ここで, $m(x)$ は固定効果, $r_i(x)$ は変量効果を表し, それぞれ m_f, m_r 個の基底関数 $b_k^{(F)}(x)$, $b_l^{(R)}(x)$ の線形和で表されるとする. また, $\boldsymbol{u}_i \sim N_{m_r}(\boldsymbol{0}_{m_r}, \Gamma)$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ を仮定し, $\boldsymbol{u}_i$ と ε_{ij} はそれぞれ互いに独立であるとする. このとき, i 番目の個体に対する観測値 $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{it_i})'$ に対してモデル (1) 式は次のようにまとめられる.

$$\boldsymbol{y}_i = \boldsymbol{B}_i^{(F)} \boldsymbol{\beta} + \boldsymbol{B}_i^{(R)} \boldsymbol{u}_i + \boldsymbol{\varepsilon}_i. \qquad (2)$$

ただし, $\boldsymbol{B}_i^{(F)} = (\boldsymbol{b}^{(F)}(x_{i1}), \cdots, \boldsymbol{b}^{(F)}(x_{it_i}))'$, $\boldsymbol{B}_i^{(R)} = (\boldsymbol{b}^{(R)}(x_{i1}), \cdots, \boldsymbol{b}^{(R)}(x_{it_i}))'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{it_i})'$ とする.

3　未知パラメータの推定

モデル (2) 式に含まれる未知パラメータ $\sigma_\varepsilon^2, \Gamma, \boldsymbol{\beta}$ を最尤法に基づいて推定する. ここで, $\boldsymbol{u}_i$ と ε_{ij} の分布の仮定から $\boldsymbol{y}_i \sim N_{t_i}(\boldsymbol{B}_i^{(F)} \boldsymbol{\beta}, \sigma_\varepsilon^2 I_{t_i} + \boldsymbol{B}_i^{(R)} \Gamma \boldsymbol{B}_i^{(R)'})$ が成り立つので, $\boldsymbol{y}_i$ の密度関数 $f(\boldsymbol{y}_i | \sigma_\varepsilon^2, \Gamma, \boldsymbol{\beta})$ は次のように表わされる.

$$f(\boldsymbol{y}_i|\sigma_\varepsilon^2,\Gamma,\boldsymbol{\beta}) = \frac{1}{(2\pi)^{\frac{t_i}{2}}|\sigma_\varepsilon^2 I_{t_i} + \boldsymbol{B}_i^{(R)}\Gamma\boldsymbol{B}_i^{(R)'}|^{\frac{1}{2}}}$$
$$\times \exp\left\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{B}_i^{(F)}\boldsymbol{\beta})'(\sigma_\varepsilon^2 I_{t_i} + \boldsymbol{B}_i^{(R)}\Gamma\boldsymbol{B}_i^{(R)'})^{-1}(\boldsymbol{y}_i - \boldsymbol{B}_i^{(F)}\boldsymbol{\beta})\right\}.$$

よって, $f(\boldsymbol{y}_i|\sigma_\varepsilon^2,\Gamma,\boldsymbol{\beta})$ を対数変換することによって次の対数尤度関数が得られる.

$$l(\sigma_\varepsilon^2,\Gamma,\boldsymbol{\beta}) = -\frac{1}{2}\sum_{i=1}^{n}\left\{t_i\log(2\pi) + \log|\sigma_\varepsilon^2 I_{t_i} + \boldsymbol{B}_i^{(R)}\Gamma\boldsymbol{B}_i^{(R)'}|\right\}$$
$$-\frac{1}{2}\sum_{i=1}^{n}\{(\boldsymbol{y}_i - \boldsymbol{B}_i^{(F)}\boldsymbol{\beta})'(\sigma_\varepsilon^2 I_{t_i} + \boldsymbol{B}_i^{(R)}\Gamma\boldsymbol{B}_i^{(R)'})^{-1}(\boldsymbol{y}_i - \boldsymbol{B}_i^{(F)}\boldsymbol{\beta})\}.$$

対数尤度関数 $l(\sigma_\varepsilon^2,\Gamma,\boldsymbol{\beta})$ を最大にするパラメータ $\sigma_\varepsilon^2,\Gamma,\boldsymbol{\beta}$ を EM-アルゴリズムを用いて計算し, 推定量を $\hat{\sigma}_\varepsilon^2$, $\hat{\Gamma}$, $\hat{\boldsymbol{\beta}}$ と置く. このとき, 変量効果の係数 $\boldsymbol{u}_i$ と $\boldsymbol{y}_i$ の予測値はそれぞれ次のように与えられる.

$$\hat{\boldsymbol{u}}_i = (\hat{\sigma}_\varepsilon^2\hat{\Gamma}^{-1} + \boldsymbol{B}_i^{(R)'}\boldsymbol{B}_i^{(R)})^{-1}\boldsymbol{B}_i^{(R)'}(\boldsymbol{y}_i - \boldsymbol{B}_i^{(F)}\hat{\boldsymbol{\beta}}),$$
$$\hat{\boldsymbol{y}}_i = \boldsymbol{B}_i^{(F)}\hat{\boldsymbol{\beta}} + \boldsymbol{B}_i^{(R)}\hat{\boldsymbol{u}}_i.$$

4 モデル選択

前節で推定された非線形混合効果モデル $f(\boldsymbol{y}_i|\hat{\sigma}_\varepsilon^2,\hat{\Gamma},\hat{\boldsymbol{\beta}})$ は, 固定効果の基底関数の個数 m_f と変量効果の基底関数の個数 m_r に依存する. m_f, m_r を選択するための方法として, ベイズアプローチに基づくモデル評価基準 MBIC(Konishi *et al.*(2004)) を以下のように導出する.

$$MBIC = -2\sum_{i=1}^{n}\log f(\boldsymbol{y}_i|\hat{\boldsymbol{\theta}}) + p\{\log n - \log(2\pi)\} + \log|R(\hat{\boldsymbol{\theta}})|.$$
$$\boldsymbol{\theta} = (\sigma_\varepsilon^2, vech(\Gamma)', \boldsymbol{\beta}')' : p\times 1, \qquad p = m_f + m_r(m_r+1)/2 + 1,$$
$$vech(\Gamma) = (\Gamma_{11},\cdots,\Gamma_{m_r1},\Gamma_{22},\cdots,\Gamma_{m_rm_r})' : m_r(m_r+1)/2\times 1,$$
$$\Gamma = \{\Gamma_{jk}\}_{1\le j,k\le m_r} : m_r\times m_r \qquad (\Gamma_{jk}=\Gamma_{kj}),$$
$$f(\boldsymbol{y}_i|\boldsymbol{\theta}) = f(\boldsymbol{y}_i|\sigma_\varepsilon^2,\Gamma,\boldsymbol{\beta}), \qquad R(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log f(\boldsymbol{y}_i|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}.$$

MBIC の値を最小にする λ_f, λ_r をそれぞれ選択する.

参考文献

Konishi, S., Ando, T. and Imoto, S. (2004). Baysian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika*, **91**, 27-43.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled nois curves. *Biometrics*, **57**, 253-259.