線形混合モデルのモデル選択規準について
(On a model selection criterion for a linear mixed model)

広島大　理　　若木　宏文

Let Y_{ij} $(i = 1, \dots, N;\ j = 1, \dots, p)$ be an observation about the ith subject at time t_j. Consider a polynomial regression model of which the intercept and the first regression coefficient are random:

$$Y_{ij} = (1\ t_j)\boldsymbol{b}_i + \sum_{k=2}^{q} \beta_k t_j^k + \varepsilon_{ij},$$

where, $\boldsymbol{b}_i$ $(i = 1, \dots, N) \overset{i.i.d.}{\sim} N_2(\boldsymbol{\mu}, \boldsymbol{\Xi}_r)$, ε_{ij} $(i = 1, \dots, N;\ j = 1, \dots, p) \overset{i.i.d.}{\sim} N(0, \sigma^2)$ with unknown parameters $\boldsymbol{\mu}, \sigma^2 > 0$ and $\boldsymbol{\Xi}_r$. We note that $\boldsymbol{\Xi}_r$ is positive semidefinite.

In order to obtain a canonical model, we use the QR decomposition:

$$(\boldsymbol{t}_0, \dots, \boldsymbol{t}_q) = \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_3 \\ \mathbf{O} & \mathbf{C}_2 \end{pmatrix} [\underset{p\times 2}{\mathbf{H}_1} \quad \underset{p\times(q-1)}{\mathbf{H}_2}],$$

where $\boldsymbol{t}_k = (t_1^k, \dots, t_p^k)^{\mathrm{T}}$, $(k = 0, 1, \dots, q)$, and $\mathbf{H}_1, \mathbf{H}_2$ are taken such that $[\mathbf{H}_1\ \mathbf{H}_2\ \mathbf{H}_3]$ is an orthogonal matrix of size p. Then

$$\boldsymbol{Z}_i := \mathbf{H}\mathbf{Y}_i = \begin{pmatrix} \boldsymbol{Z}_{i1} \\ \boldsymbol{Z}_{i2} \\ \boldsymbol{Z}_{i3} \end{pmatrix} \sim N_p \left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Xi} + \sigma^2 \mathbf{I}_2 & & \mathbf{O} \\ & \sigma^2 \mathbf{I}_{q-1} & \\ \mathbf{O} & & \sigma^2 \mathbf{I}_{p-q-1} \end{pmatrix} \right],$$

where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iN})^{\mathrm{T}}$, $\boldsymbol{\mu}_1 = \mathbf{C}_1\boldsymbol{\mu} + \mathbf{C}_3\boldsymbol{\beta}$, $\boldsymbol{\mu}_2 = \mathbf{C}_2\boldsymbol{\beta}$ and $\boldsymbol{\Xi} = \mathbf{C}_1\boldsymbol{\Xi}_r\mathbf{C}_1^{\mathrm{T}}$ with $\boldsymbol{\beta} = (\beta_2, \dots, \beta_1)^{\mathrm{T}}$.

The sufficient statistics are given by

$$\bar{\boldsymbol{Z}}_1 = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}_{i1}, \quad \bar{\boldsymbol{Z}}_2 = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}_{i2}.\ s_2 = \sum_{i=1}^{N} \left\{ (\boldsymbol{Z}_{i2} - \bar{\boldsymbol{Z}}_2)^{\mathrm{T}} (\boldsymbol{Z}_{i2} - \bar{\boldsymbol{Z}}_2) + \boldsymbol{Z}_{i3}^{\mathrm{T}} \boldsymbol{Z}_{i3} \right\}$$

$$\mathbf{S}_1 = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} = \sum_{i=1}^{N} (\boldsymbol{Z}_{i1} - \bar{\boldsymbol{Z}}_1)(\boldsymbol{Z}_{i1} - \bar{\boldsymbol{Z}}_1)^{\mathrm{T}}.$$

The MLE (maximum likelihood estimators) of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are $\hat{\boldsymbol{\mu}}_1 = \bar{\boldsymbol{Z}}_1$ and $\hat{\boldsymbol{\mu}}_2 = \bar{\boldsymbol{Z}}_2$, respectively. While the MLE of $(\sigma^2, \boldsymbol{\Xi})$ is given as follows.

(i) $s_2 \le (p-2) l_2 \quad \Rightarrow \quad \hat{\sigma}^2 = \dfrac{s_2}{N(p-2)},\ \hat{\boldsymbol{\Xi}} = \dfrac{1}{N}\mathbf{S}_1 - \hat{\sigma}^2 \mathbf{I}_2$

(ii) $(p-2) l_2 < s_2 < (p-1) l_1 - l_2 \quad \Rightarrow \quad \hat{\sigma}^2 = \dfrac{s_2 + l_2}{N(p-1)},\ \hat{\boldsymbol{\Xi}} = \left(\dfrac{l_1}{N} - \hat{\sigma}^2 \right) \boldsymbol{h}_1 \boldsymbol{h}_1^{\mathrm{T}}$

(iii) $(p-1) l_1 - l_2 \le s_2 \quad \Rightarrow \quad \hat{\sigma}^2 = \dfrac{s_2 + l_1 + l_2}{Np},\ \hat{\boldsymbol{\Xi}} = \mathbf{O}$

where $l_1 > l_2$ are the eigen values of $\mathbf{S}_1$, and $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are the corresponding eigen vectors.

Denote $\mathbf{Z} = (\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_N)$ and let $\hat{\boldsymbol{\theta}}(\mathbf{Z})$ be the MLE based on $\mathbf{Z}$. We consider a risk of the model given by $\mathrm{Risk}(f) = -2\mathrm{E}_{\mathbf{Z}}\mathrm{E}_{\tilde{\mathbf{Z}}}[\log f(\tilde{\mathbf{Z}}; \hat{\boldsymbol{\theta}}(\mathbf{Z}))]$ based on the Kullback–Leibler divergence, where $\tilde{\mathbf{Z}}$ is a copy of $\mathbf{Z}$ and $f(\mathbf{Z}, \boldsymbol{\theta})$ is the joint probability density function of $\mathbf{Z}$. We found that when $\boldsymbol{\Xi}$ is near $\mathbf{O}$, the usual AIC criterion is not asymptotically unbiased estimator of $\mathrm{Risk}(f)$. We gave an approximation formula of the bias of AIC and discussed how to modify the bias.