On the Behavior of Marginal and Conditional Akaike Information Criteria in Linear Mixed Models

Sonja Greven
Department of Biostatistics, Johns Hopkins University

joint work with
Thomas Kneib
Department of Mathematics, Carl-von-Ossietzky-Universität Oldenburg

IMS New Researchers Conference, Baltimore 2009

Overview

- Additive mixed models offer growing flexibility in modeling, making model choice increasingly important.
- Two Akaike Information Criteria (AICs) have been used in mixed models: the marginal and the conditional AIC.
- We shed light on their theoretical and practical properties when selecting random effects.

1. Background: Mixed Models and the AIC
2. Marginal and Conditional AIC
3. Application: Childhood Malnutrition in Zambia

Linear and Additive Mixed Models

Linear Mixed Models (LMMs): widely used class of flexible regression models

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{b} + \varepsilon, \qquad \begin{pmatrix} \boldsymbol{b} \\ \varepsilon \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2 \boldsymbol{I_n} \end{pmatrix} \right).$$

Special case I: Random intercept / random slope models for longitudinal or grouped data → model intra-subject correlations.

Special case II: Penalized spline smoothing using a particular LMM:

- polynomial part (degree d) of smooth function corresponds to fixed effects
- deviations modeled using random effects.

Similar approaches for varying coefficients, surfaces, spatial effects, ...
→ Additive mixed models can combine all of these modeling components.

Application: Childhood Malnutrition in Zambia

- 1,600 observations from the 1992 Zambia Demographic and Health Survey
- Additive mixed model for stunting, measuring insufficient height for age:

$$zscore_i = \mathbf{x}_i'\beta + m_1(cage_i) + m_2(cfeed_i) + m_3(mage_i) + m_4(mbmi_i) + m_5(mheight_i) + b_{s_i} + \varepsilon_i,$$

  - Categorical covariates: gender of the child, education and employment status of the mother
  - Continuous covariates: age of the child, duration of breastfeeding, age, height and body mass index of the mother - smooth or linear effects?
  - Random intercept: b_{s_i} captures spatial variability between residential districts.

- Model selection involves random effects.
  - Random intercept
  - Random effects modeling smooth deviations from linearity of the $m_k(\cdot)$.

The Akaike Information Criterion (AIC)

- Data $\boldsymbol{y}$ generated from a true underlying model with density $g(\cdot)$.
  Approximate by a parametric class of models $f_\psi(\cdot) = f(\cdot;\psi)$, $\psi \in \boldsymbol{\Psi}$.
- Kullback-Leibler (KL) distance: $K(f_\psi, g) = \mathrm{E}_{\boldsymbol{z}}\left[\log(g(\boldsymbol{z})) - \log(f_\psi(\boldsymbol{z}))\right]$.
- Want to minimize (relative) expected KL distance using estimator $\widehat{\psi}(\boldsymbol{y})$

$$\mathrm{E}_{\boldsymbol{y}}[K(f_{\widehat{\psi}(\boldsymbol{y})}, g)] \to \min \quad \Leftrightarrow \quad -2\,\mathrm{E}_{\boldsymbol{y}}[\mathrm{E}_{\boldsymbol{z}}\left[\log(f_{\widehat{\psi}(\boldsymbol{y})}(\boldsymbol{z}))\right]] \to \min \qquad (1)$$

  Predictive quantity, with $\boldsymbol{y}$ and $\boldsymbol{z}$ independent replications.
- $AIC = -2\log(f_{\widehat{\psi}(\boldsymbol{y})}(\boldsymbol{y})) + 2k$ is an estimator of (1), $k = \dim(\psi)$.
  - $2k$ is correction term: $-2\log(f_{\widehat{\psi}(\boldsymbol{y})}(\boldsymbol{y}))$ not predictive quantity.
  - unbiased for (1) under regularity conditions: $\boldsymbol{\Psi} = \mathbb{R}^k$ and $y_1, \ldots, y_n$ are i.i.d.

...and for the linear mixed model?

Marginal and Conditional Perspective

- Marginal perspective: Random effects induce a correlation structure.

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\beta, \boldsymbol{V}), \qquad \boldsymbol{V} \equiv \boldsymbol{V}(\theta) = \sigma^2 \boldsymbol{I_n} + \boldsymbol{ZDZ}'.$$
$$mAIC = -2l(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) + 2(p+q), \qquad \text{where}$$

  $l(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$ is the maximized marginal log-likelihood, $p = \dim(\boldsymbol{\beta})$, $q = \dim(\boldsymbol{\theta})$.

  Typically returned by statistical software (R, SAS) and widely used for selection of random and fixed effects.

- Conditional perspective: Random effects are regression coefficients estimated subject to regularization.

$$\boldsymbol{y}|\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{X}\beta + \boldsymbol{Zb}, \sigma^2 \boldsymbol{I_n}).$$

The Conditional AIC

- Vaida & Blanchard (Biometrika, 2005): Conditional AIC

$$cAIC = -2l(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\theta}}) + 2(\rho + 1), \quad \text{where}$$

  $l(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\theta})$ is the conditional log-likelihood and ρ is the trace of the hat matrix, the effective degrees of freedom, when $\boldsymbol{D}$ known.

  Authors recommend the cAIC with estimated $\widehat{\boldsymbol{D}}$ when $\boldsymbol{D}$ is unknown: $\widehat{\rho}$.

- Liang, Hu & Zou (Biometrika, 2008): Corrected cAIC replaces ρ by $\Phi_0 = \text{trace}\left(\frac{\partial \widehat{\mathbf{y}}}{\partial \mathbf{y}}\right)$ for known σ^2. (Otherwise, Φ_1 involves second derivatives.)

  No closed form, numerical approximation of Φ_0 requires n extra model fits. In our example: ca. 110 days (!) to compute corrected cAIC for all models.

Questions

- Theoretical and practical properties of the marginal and conditional AIC when focus is on the random effects?
- Can we use the computationally less demanding cAIC of Vaida & Blanchard?

1 Background: Mixed Models and the AIC

2 Marginal and Conditional AIC

3 Application: Childhood Malnutrition in Zambia

Theoretical Results

Marginal AIC

- The marginal AIC is positively biased for (1).
- The bias is dependent on the true unknown $\boldsymbol{\theta}$, does not vanish asymptotically.
- The marginal AIC favors smaller models excluding random effects.

Reasons: $\boldsymbol{\Psi} \neq \mathbb{R}^{p+q}$ and $y_1, \ldots, y_n$ correlated $\Rightarrow$ bias correction not $2\,\mathrm{E}[\chi^2_{p+q}]$.

Compare testing for zero variances (Crainiceanu & Ruppert, 2004; Greven et al, 2008).

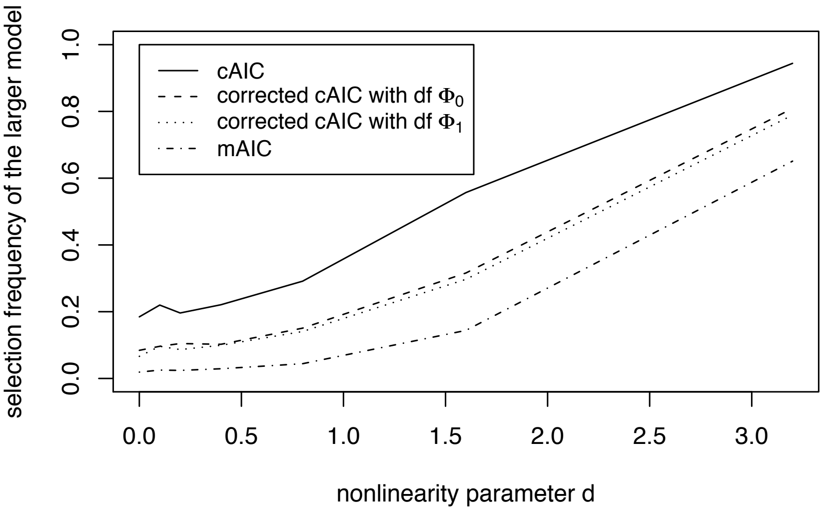
Conditional AIC of Vaida & Blanchard with estimated ρ

- The larger model including the random effect is chosen whenever $\hat{\boldsymbol{b}} \neq 0$.
- If $\hat{\boldsymbol{b}} = 0$, the cAICs of the two models coincide.

Reasons: Bias correction ρ again estimated from the same data $\boldsymbol{y}$.

Simulations - Example

$y_i = m(x_i) + \varepsilon_i, i = 1, \ldots, n = 30$, where $m(x) = 1 + x + 2d(0.3 - x)^2$.



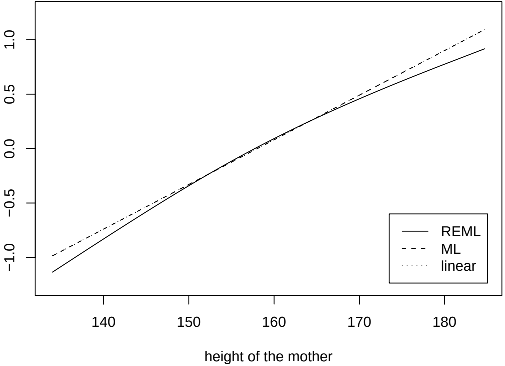
1 Background: Mixed Models and the AIC

2 Marginal and Conditional AIC

3 Application: Childhood Malnutrition in Zambia

Application: Childhood Malnutrition in Zambia

Simple example: $zscore_i = m(mheight_i) + \varepsilon_i$.



ML: mAIC prefers smaller, linear model; tie for cAIC.
REML: mAIC prefers smaller, linear model; cAIC chooses larger, non-linear model.

In the full model the functions are either linear or clearly non-linear . . . .

Summary

- The marginal AIC is affected by the boundary of the parameter space, similarly to likelihood ratio tests.
  It is biased towards simpler models excluding random effects.
- The conventional conditional AIC tends to select models that are too large.
  It includes any random effect not estimated to be exactly zero.
- The corrected conditional AIC rectifies this difficulty but comes at a high computational price.
  $\longrightarrow$ Work in progress: Representation that is computationally feasible?

References

- Greven, S. & Kneib, T. (2009): On the Behavior of Marginal and Conditional Akaike Information Criteria in Linear Mixed Models. *Johns Hopkins University, Department of Biostatistics Working Papers, Paper 179.* http://www.bepress.com/jhubiostat/paper179/

- Akaike, H. (1973): Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *2nd International Symposium on Information Theory, 267-281.* Akademiai Kiado.
- Crainiceanu, C. & Ruppert, D. (2004): Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B 66(1), 165-185.*
- Greven, S., Crainiceanu, C., Küchenhoff, H. & Peters, A. (2008): Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics 17 (4): 870-891.*
- Liang, H., Wu, H. & Zou, G. (2008): A note on conditional AIC for linear mixed-effects models. *Biometrika 95, 773-778.*
- Vaida, F. & Blanchard, S. (2005): Conditional Akaike information for mixed-effects models. *Biometrika 92, 351-370.*