Written Examination for Linear Mixed Models (MSA650 and MVE210)

Teacher: Ziad Taib, +46 708 46 73 56
Jour: Juan Inda ankn. 5325
Date and place: 2019-06-10
Rules: This is a closed book exam. Only simple pocket calculators are allowed.
Grades: The written exam is worth 24 scores (80%) while the computer assignments are worth 6 scores (20%). The two add up to 30 scores (100%). Excellent (≥ 26 scores), pass (≥ 16 scores) and do not pass (< 16 scores).

---

1. In the setting of mixed models, we can work with different types of residuals e.g. marginal residuals and conditional residuals. Define these in a detailed manner and explain what they can be used for. (3p)

2. Patients with chronic headache are randomized into two groups whereas both groups receive an active drug (LNMMA) and placebo, on two different days, with a suitable wash-out period in-between. Group G1 was treated first with placebo (period 1), and then with LNMMA (period 2) and Group G2 was treated first with LNMMA (period 1), and then with placebo (period 2). Pain was measured subjectively on a VAS-scale (small is good), at baseline and at 30, 60, 90 and 120 minutes after treatment. Specify a model with the following fixed effects: treatment, period, treat*period. Also include a random subject effect. Such a design is called a crossover design. What assumptions would make on the model. What are the advantages of such a model compared to a parallel design? What hypotheses would you, ideally, want to test? (3p)

3. To compare the efficacy and safety of two alternative antiretroviral drugs, didanosine (ddI) and zalcitabine (ddC) based on data from 467 HIV infected patients, randomized to the two treatments, the outcomes of interest were: (i) CD4 cell count measurements at baseline, 2, 6, 12 and 18 months and (ii) previous opportunistic infections (prevOI:). The most important research questions were (i) How CD4 cell count evolves over time for this cohort of patients? (ii) Does treatment improve average longitudinal evolutions?
    a. Specify a marginal model to address these questions which includes, among other things, different average longitudinal evolutions per treatment and group.
    b. Specify a linear mixed model assuming different average longitudinal evolutions per treatment group (fixed part), random intercepts and random slopes (random part) i.e. without a main effect for treatment.
    c. What assumptions are implied by the model you specified and what additional assumptions need to be made to be able to analyse data based on the model.

    (4p)

4. Consider the hypothesis test between the random intercepts and the random intercepts & random slopes models random intercepts model

$$y_{ij} = X\beta + b_{i0} + \varepsilon_{ij}, \qquad b_{i0} \sim \mathcal{N}(0, \sigma^2_{b_1})$$

And the random intercepts & random slopes model

$$y_{ij} = X\beta + b_{i0} + b_{i1}t + \varepsilon_{ij}, \qquad b_{i0} \sim \mathcal{N}(0, D) \text{ where } D = \begin{bmatrix} \sigma^2_{b_1} & \sigma_{b_{12}} \\ \sigma_{b_{12}} & \sigma^2_{b_2} \end{bmatrix}$$

Hence, the hypotheses to be tested are

$$H_0: \quad \sigma^2_{b_2} = \sigma_{b_{12}} = 0$$

$$H_a: \quad \sigma^2_{b_2} \neq 0 \text{ or } \sigma_{b_{12}} \neq 0$$

If you were told that he null hypothesis for σ^2_{b2} is on the boundary of its corresponding parameter space, would that be a major problem? (3p)

5. We want to compare two mixed models:
   - M1: random intercepts & linear random slopes, with an unstructured matrix for these random effects
   - M2: random intercepts, & nonlinear random slopes with splines, with a diagonal matrix for these random effects

   a) Can a likelihood ratio test be used for such a comparison? Motivate your answer. Alternatively, is it possible to compare the two models using the AIC and BIC Values.
   b) If a) is (rightly or wrongly) used. Which model fits the data better?

	df	logLik	AIC	BIC
M_1	10	−1522.38	3064.75	3120.45
M_2	10	−1438.53	2897.06	2952.76

(3p)

6. To estimate Estimation of random effects based on a fitted mixed model, estimates for the random effects are based on the posterior distribution:

$$p(b_i \mid y_i; \theta) = \frac{p(y_i \mid b_i; \theta)\, p(b_i; \theta)}{p(y_i; \theta)} \propto p(y_i \mid b_i; \theta)\, p(b_i; \theta),$$

In which the parameter q is replaced by its ML estimate. This distribution can be written in a closed form as

$$[b_i \mid y_i; \theta] \sim \mathcal{N}\Big\{ DZ_i^\top V_i^{-1}(y_i - X_i\beta),\ DZ_i^\top K Z_i D \Big\},$$

   a) Discuss how this result can be used to "estimate" the random effects
   b) Discuss how this result can be used to "predict" the response values (y_i) of individual observations.
   c) Compare the predictor from b) with the corresponding one obtained from the marginal model.
   d) Th prediction in b) is called the Best Linear Unbiased Predictor. Explain why it is called in that way.
   e) As an example of the difference between the marginal and subject-specific predictions in the previous question, we compare the two sets of predictions for 16 randomly selected patients with responses that follow a linear mixed model such that
      - red lines denote the marginal predictions,
      - blue lines denote the subject-specific predictions
      - black circles the observed data

Discuss what you see in the following graph?

(8p)