Analysis of Longitudinal Data SoSe 2016 **Sheet 6**

Prof. Dr. Sonja Greven and Dipl. Stat. Jona Cederbaum 05.07.2016

This exercise sheet will familiarize you with generalized linear mixed models (GLMMs). The exercises refer to the content of the eighth and ninth lecture slides.

Exercise 1: GLMM for binary data

The data set `madras` included in the package `drm` comprises monthly records (presence/absence) of the psychiatric symptom "thought disorder" of 86 patients over the first year after initial hospitalization for disease.

(1) Load the data set and take a first look at the data.

(2) How many observations are available for each patient? Is the data set balanced?

(3) Convert the variable `id` into a factor variable with levels 1 to 86.

(4) Now calculate for men and women separately the relative frequencies of occurrence of "thought disorder" over time and plot the curves. What can be seen from the plot?

(5) Since we are interested in individual-specific effects for this data, a GLMM including fixed effects for `month` and `gender` as well as a random intercept per patient shall be estimated. Formulate the model and specify the distributional assumptions, the systematic component, as well as a suitable link function.

(6) What differentiates the estimation of a GLMM from that of an LMM? What makes the estimation so complex?

(7) Which three options for estimating GLMMs did you get to know? What is approximated in each case?

(8) Estimate the model with each of these three options. Use the functions of the packages `MASS` and `lme4` which were presented in the lecture. Use 11 quadrature points for the adaptive Gaussian quadrature (AGQ). Compare the estimates of the fixed effects.

(9) Why does the estimation with AGQ take longer than the Laplace estimation?

(10) What should you consider when interpreting the results of PQL estimation for this data set?

(11) Interpret the AGQ estimates of the fixed effects.

- **(a)** Why can the coefficient estimates not be interpreted marginally (in general)?
- **(b)** Plot the individual probabilities $P(Y_{ij} = 1|b_i)$ over time (`month`) by gender.
- **(c)** Add the probability curve over time for $b_i = 0$ to both plots.

(d) Now calculate the average probabilities $P(Y_{ij} = 1)$ separately for men and women and add these to the plots as well. How can the resulting plots be interpreted?

(12) You now want to perform an LR test to check whether there is an interaction between the sex of the patient and time.

(a) Why is an estimate based on PQL not suitable for this?

(b) Why does the R-function `summary` give no value (`NA`) for the AIC for the estimation via PQL?

(13) Use the AGQ estimation instead and perform the test proposed in (12). What is your test decision?

(14) An alternative approach to analyze longitudinal non-normal data is the use of marginal models (next sheet). Which are the main differences between these two approaches?