Exam of course **Bayesian Data Analysis II**

January 11, 2019

Prof. Emmanuel Lesaffre

L-Biostat, K.U.Leuven, Leuven

Included is the description of your exam project (2 analyses).

Practical arrangements:

1. Establish groups of size **4 to 5**, not more and not less!

2. Deliver your exam project on a **pdf** file. Each document should have a title page with heading: **2019 BDA II UHasselt,** the names of the members of the group and their email addresses. Send the document to me **by email** (email address: **emmanuel.lesaffre@kuleuven.be**) at the latest at **10PM (Belgian time)** on **January 3, 2019.**

   Note that if the document is sent to another email address (do not send it to my UHasselt email address!), there is always the risk I will not see it before the deadline, hence use the KULeuven email address!!!

   Further, note that this is AN ABSOLUTE DEADLINE, submissions after this deadline will NOT be taken into account.

3. Give your pdf file **the name of the first member of the group!**
4. **Send also your programs** to me in the same email, so that I can check that the programs really work!
5. Use WINBUGS/JAGS/R2WinBUGS/R2jags, CODA or BOA in combination with R to solve the questions.
6. For clarity, **first repeat the question** in your document and then give the solution.
7. Describe the flow of your procedures and the reasoning behind. Be clear! **PUT PAGE NUMBERS** on project!
8. Annotate your output, but not everything that you have done needs to be put in the report. **Limit your report to 20 pages (excluding title pages, but including tables and figures)!**
9. Print a version for yourself and bring it with you at the oral defense and **DO NOT annotate your version**. **So DO NOT ANNOTATE YOUR VERSION!!!**

10. **Study also the course material**; you will get general questions on Bayesian statistics at the oral defense.
11. **Good luck!**

Analysis 1: Growth of Nigerian indigenous chickens

The data in the file **NIC_dataset.txt** describes data from experimental birds consisted of 416 progenies produced from mating involving Nigerian indigenous chickens and exotic broiler parent stock. So many genotypes were generated in a straight and reciprocal cross. The progenies produced from the same parent stock (pure breed) consisted of 218 progenies while those from different parents (cross breed) consisted of 198 progenies. The experiment aimed at evaluating the growth of different breeds. The NIC data set is a longitudinal hierarchical data, in which the body weight (BW) of individual chicken clustered within breeds (1= pure breed and 2= cross breed) were measured every week from hatching up to thirteen weeks.

Questions/Tasks

- Fit a linear mixed effects model with a random intercept and slope to the longitudinal weights. Assume normality for the random effects and also for the independent measurement error. Take vague priors for all model parameters.
- Check graphically the assumed normality of the random effect and measurement errors.
- Assess the mixed effects model using the Gelman posterior predictive check (PPC).
- Even when the PPC indicates model appropriateness, it might well be that some part of the model needs to be improved. Check what can be improved in the original mixed effects model. Hint: One can
    - Change the distribution of random effects;
    - Change the distribution of measurement error;
    - Vary the standard deviations of random effects and/or measurement error in different ways;
    - Adapt the fixed effects part of the model.
- Compare the chosen models with DIC and the pseudo-Bayes factor
- Perform a sensitivity analysis on the prior of random effects. Check what happens to the mixing of the chains and the posterior results.

Analysis 2: Evolution of ALS

The dataset is about Amyotrophic lateral sclerosis (ALS), also known as motor neuron disease or Lou Gehrigs disease. It is a progressive, invariably fatal, neurodegenerative disease caused by death of the neurons in the brain and the spinal cord that controls voluntary movement. As these motor neurons degenerate, signaling to voluntary muscles is disrupted, which leads to muscle weakening, atrophy and paralysis. Eventually, all muscles under voluntary control are affected, and patients gradually lose their ability to walk, talk, swallow, and breath. Ten items were used to measure the degeneration over time. Each item ranges from 0 to 4 where a higher score implies a better condition. A total score was calculated by summing up these items, ranging from 0 to 40.

In this project, a binary version of the total score is used. The new variable, called **ALS binary score**, is coded as 1 (normal) if the total score is greater than 30 and 0 (abnormal) otherwise. Other information includes **age** (at baseline), **gender** (female=1), **treatment** (active=1, placebo=0), and **time** (in years). The dataset can be found in **project2.txt.**

The question of interest is how the degeneration evolves over time.

To begin, **create new ID if necessary**.

Questions/tasks:

Part 1:

- Fit a linear mixed effects logistic regression model with random intercept and slope **without any further covariates** to the evolution of the ALS binary score. Assume a bivariate normal distribution N(0, D) for the random effects. Take vague priors for all model parameters.

- Do a sensitivity analysis by considering different prior specifications for the covariance matrix D. **Hint**: Two following specifications for D can be used: (i) IW prior, and (ii) uniform priors for standard deviations and the correlation coefficient.

- Check with DIC whether D can be a diagonal matrix and/or that we can remove the random effects.

Part 2:

- Select the best model from part 1 and include all covariates. Standardize age to improve convergence of the MCMC procedure. Is this model better? Give motivation.

- Do a model selection here and perform posterior predictive checks to evaluate the chosen model.

- Check the logistic link, i.e. verify whether the probit or the complementary log-log link provide a better model.

- Give all necessary posterior summary measures of the relevant parameters.

- Is time associated with a decline of the ALS score? What about the treatment effect?

Hint: To improve convergence of the MCMC procedure, you may give some plausible initial values to the regression coefficients and parameters related to variances, covariance of random effects.

For more information, see **https://nctu.partners.org/ProACT/Document/DisplayLatest/2**.