Exam of course **Bayesian Data Analysis II – part 2**

January 2022

Prof. em. Emmanuel Lesaffre

L-Biostat, K.U.Leuven, Leuven

Included is the description of the 2nd part of your exam project.

Practical arrangements:

1. **Take the same groups as for part 1!**
2. Deliver your exam project on a **pdf** file. Each document should have a title page heading: **2022 BDA II part 2 KU Leuven,** the group members' names, and their email addresses. Send the document to me **by email** (email address: **emmanuel.lesaffre@kuleuven.be**) at the latest on **Thursday 27 January 2022 at 11.59 PM (Belgian time).**
3. Give your pdf file **the name of the first member of the group!**
4. **Send also your program(s)** to me in the same email!
5. This exam project is of a different nature than part 1. Now we are asking you to be creative and show the capabilities of the package **`NIMBLE`** for the Bayesian analyses.
6. Annotate your output, but not everything you have done needs to be put in the report. **Limit your report to 20 pages (excluding title pages and including tables and figures)!**
7. **You do not need to bring your solution to the oral exam!**
8. **See you in January.**
9. **Good luck!**

A Bayesian longitudinal & survival analysis

- **Description of the data**

Mechanical ventilation (MV) is one of the most common life support procedures in the intensive care unit (ICU). The goals of MV are to improve gas exchange, reduce the work of breathing, and improve patient comfort. Appropriate interaction with the mechanical ventilator is very important throughout ventilatory support in critically ill patients. Patient-ventilator asynchronies (PVAs) appear when the timing of the ventilator cycle is not simultaneous with the timing of the patient's respiratory cycle. PVAs can occur during all phases of the respiratory cycle, including breath initiation, flow and pressure delivery, the transition from inspiration to expiration, and throughout expiration. PVAs are associated with a longer duration of MV, longer stay in the ICU, higher incidence of respiratory muscle injury and tracheostomy, and lower probability of successfully weaning from MV. In the ICU setting, it is common to measure the severity of illness using severity scores. An AI measures PVAs and assesses overall severity together with the SOFA score. SOFA measures the degree of organ dysfunction and assesses the evolution of the patient's severity over the ICU stay. The longitudinal nature of the SOFA index and its strong association with illness severity and the two possible outcomes of interest at the ICU, death or alive discharge, is a suitable framework for quantitative studies that can assess these relationships to characterize patients' evolution and prognosis.

The data was collected from 139 mechanically ventilated patients admitted to four Spanish ICUs from July 2009 to May 2016. All patients were followed from the first day in MV until ICU discharge or day 30 after MV initiation, whichever occurred first. Two main outcomes were of interest in the study: death in or discharge alive from the ICU. Of the 139 studied patients, 28 (20.1%) died, 97 (69.8%) were discharged alive, and 14 (10.1%) were administratively censored. The AI was defined as the proportion of asynchronous events among the total number of ventilator cycles. Patients' overall severity was measured using the SOFA, which assigns from 0 (normal) to 4 (most abnormal) points to the function of each of the following six organ systems: respiratory, circulatory, renal, hematology, hepatic, and central nervous system, giving a possible score of 0 to 24. Both the AI and the SOFA scores were measured daily.

The following data were recorded:

Variable name	Description
`id`	Patient id
`day`	Time (in days) in the Intensive Care Unit since the study entry
`icustaymv`	Number of days in mechanical ventilation
`fail`	Type of failure (1: dead in the ICU; 2: alive discharge; 0: censored at day 30)
`sofa`	Sequential Organ Failure Assessment (score with range 0-24)
`di`	Asynchrony index is defined as the proportion of asynchronous events among the total number of ventilator cycles (range 0-1)
`age`	Age (yrs)

The data set can be found in **AIdataset.Rds** and imported with `readRDS`).

Aim of the research:

The ultimate aim of the statistical analysis is to assess the impact of AI and SOFA on the patient's survival: this involves joint modelling of the two longitudinal processes (i.e., AI and SOFA) and the survival process with outcome death or alive at discharge or censored at day 30.

To do:

Three analyses are asked to perform:

use splines for fixed effects...

- **Analysis 1**: Fit a linear mixed model for SOFA with covariates *age* and *day*. Check the functional form of the covariates in the model. Hint: also, look at a smoothing technique.
- **Analysis 2**: Since AI is measured as a proportion (variable *di*), a Beta(α(*day*), β(*day*)) seems reasonable. Here α(*day*) and β(*day*) are functions of *day* and *age* (possibly depending on random effects). You need to regress the mean bounded outcome (i.e., AI) as a function of the covariates of interest (similar to SOFA). Hint: Have a look at the beta regression model by Ferrari and Cribari-Neto (2004); citation: Ferrari, S. and Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. Journal of Applied Statistics, 31(7), 799-815.

mean bounded outcome over time to model am besten nochmal anschauen

- **Analysis 3**: Simplify the response *fail* by collapsing outcomes 0 and 2 into 0. Perform a proportional hazards model with a baseline Weibull hazard function and covariates the predicted values for SOFA and AI at the global grid of observed times. Check the importance of these covariates.
- **Possible analysis**: For those of you brave enough can try to perform a joint bivariate longitudinal & survival model jointly combining the two longitudinal markers with the survival outcome.

Remember

- To provide illustrative graphics to show the observed observations
- To compute the predicted outcomes and show them graphically
- To motivate the choice of your priors
- To select the most appropriate model with a Bayesian criterion
- To check the fit of our model using a Bayesian criterion
- To include the `NIMBLE` program and make sure I can run it swiftly only needing to adapt the working directory

Note

- Add a small positive value to *`di`*
- Some of the models could generate highly correlated Markov chains. Then apply appropriate thinning to access the chains later on.