MPhil Health Data Science: Bayesian Statistics. Assessment

Instructions

- All the instructions should be followed and all questions answered.

- The submitted material must include

  - answers to the questions that are asked (in any document format, such as PDF or Word).

  - code, in any plain text format, that runs without error and reproduces all results required to answer the questions. The code may be mixed with the answers, e.g. using a report-generating format such as R Markdown. If using this kind of format, both source and compiled versions should be submitted.

Background to the question

The data in the file `sgaj.csv` records data about 12 patients who have been administered a dose of a drug called SEYAB. The data are measurements of a novel biomarker called SGAJ every 6 hours for the next 24 hours, along with associated demographic data. The first seven rows of the data are shown here:

```
##   patient female age time measured
## 1       1      1  19    0    199.5
## 2       1      1  19    6    239.4
## 3       1      1  19   12    163.5
## 4       1      1  19   18    268.1
## 5       1      1  19   24    228.9
## 6       2      0  19    0    172.4
## 7       2      0  19    6    159.3
```

The variables are

- `patient`: integer code identifying the patient.

- `female`: 1 for female, 0 for male

- `age`: age (years)

- `time`: the number of hours after SEYAB administration

- `measured`: the measured SGAJ (in μmol/L)

The laboratory technician who designed the current SGAJ assay says it is not yet entirely reliable and that the measurement error's standard deviation is 75 μmol/L.

A previous study of a similar cohort of patients measured SGAJ in patients only once, at 12 hours after administration of SEYAB, and estimated that the average SGAJ in 50 year-old females was 225 μmol/L, but this average might have been as high as 265, or as low as 185. At 12 hours, the standard deviation across all individuals in typical SGAJ levels was roughly 90.

The study principal investigator tells you that it is expected that age and sex affect SGAJ, but it is not understood in what way, except that SGAJ in men will be higher.

Questions to answer

1. Illustrate the trajectory of SGAJ results for each patient in the study to explore to what degree age and sex influence SGAJ.

2. We are interested in estimating the rate of change per day of SGAJ for patients over the first 24 hours, and in how this differs across patients.

   Construct a Bayesian regression model to answer these questions, in which between-patient variability is described using a hierarchical model.

   - write down your model in formulae, defining all notation used
   - justify all modelling assumptions including priors, and relate them to the information provided
   - implement and run your model in JAGS and R, summarising the posterior distributions of key quantities of interest
   - state your findings clearly in words, noting any limitations of your model

   Any reasonable model may be used in advance of modelling the data. There is no need to formally search for a model that best fits the data.

3. Use your model to predict how likely it is that SGAJ for patient `id = 12` would be measured above 500 μmol/L at 30 hours after SEYAB administration.

4. Conduct a posterior predictive check for the measurement at 18 hours after SEYAB administration for patient `id = 3`. Visualise and quantify your findings. Clearly explain the meaning and any implications of your findings.