Exam project:

Bayesian statistics using JAGS and R

Teacher: Prof. dr. Dries F. Benoit
Assistant: Wai Kit Tsang
Course: Bayesian Statistics

Practical arrangements:

- This is an individual assignment, not a group work.
- Deadline (strict): Saturday June 4 at 24:00 (upload via Dropbox).
- Deliver your project as a pdf file (use your family name as name for the document).
- Make sure the final document clearly explains how you solved the questions (Sweave, RMarkdown, Knitr, . . . is ideal for this).
- All your R/JAGS code has to be in the final report and should be well explained and documented.
- Also provide a separate R-file (use your family name as name for the document) containing the entire code you wrote for this assignment, so that your code can easily checked.

Exercise 1:

Create a slice sampler to sample from mixture of normals distribution given here:

```
dbinorm <- function(x, p1=.5, p2=1-p1, m1=-1, m2=2){
  p1 * dnorm(x, m1) + p2 * dnorm(x, m2)
}
```

- Report the R code with detailed comments.
- Simulate many draws from your sampler and plot the histogram.
- Overlay the histogram with the true posterior distribution.

Exercise 2:

Load the **prostate cancer dataset** from the **bayesQR** R-package. These data come from a study that examined the correlation between the level of prostate specific antigen and a number of clinical measures in men who were about to receive a radical prostatectomy. It is a data frame with 97 rows and 9 columns. Use JAGS to solve this exercise.

- estimate a regression with *lcavol* as dependent variable
- all other variables are predictors (also add an intercept)
- Check convergence and interpret the results
- Make a prediction for the average observation
- Investigate and summarize the predictive distribution

Exercise 3:

Load the **Churn dataset** from the **bayesQR** R-package and read the documentation (e.g. `?Churn`). Use JAGS to solve this exercise.

- estimate a logistic regression with *churn* as dependent variable
- all other variables are predictors (also add an intercept)
- Check convergence and interpret the results
- redo the analysis with a probit model
- what is the probability that $\beta_{\text{recency}} \leq 0$?
- what is the probability that $\beta_{\text{gender}} > \beta_{\text{lor}}$?
- Make a prediction for the average observation

Exercise 4:

For this exercise you have to read the paper (you can find it on Minerva):

Yu, K and Moyeed, R (2001). Bayesian quantile regression, *Statistics & Probability Letters*, 54(4), 437–447.

Answer the following questions:

- Write an R function that estimates the parameters of the model proposed in the paper.

- Tip: the Metropolis algorithm with normal proposal density should work fine (as well as the slice sampler).
- Do the simulation exercise in *5.1 Simulated data* and compare your results with the results in the paper.
- Simulate 200 observations from the following model:

$$y = x\beta + \epsilon$$
$$\text{with} \qquad \beta = 2, \qquad x \sim U(0, 10) \qquad \text{and} \qquad \epsilon \sim N(0, 0.6x)$$

- Plot the data together with the OLS regression line.
- Estimate the following quantiles $q = \{.05, .25, .5, .75, .95\}$ and plot the regression lines based on the Bayes estimates on the same plot.

Exercise 5:

The dataset 'falcon_data.RData' was collected about Belgian falcons. The dataset contains information about the mass (in grams), the region (two levels) and the population (three levels) of the falcons.

- Calculate a frequentist t-test and ANOVA and interpret the results
- Calculate a Bayesian t-test and ANOVA, using:
  - parameter estimation
    - with single point as null hypothesis
    - with rope (discuss prior and posterior probability of the null hypothesis)
    - investigate prior sensitivity
  - Bayes factors (see document on Minerva for more information on how to calculate these)
    - investigate prior sensitivity
    - interpret your results
- Compare the findings of the frequentist and the two Bayesian approaches