Assignments Bayes Statistics 2018

Find below the assignments for the workshop "Bayes Statistics" held in Potsdam 2018.

Please make for each assignment one `R`-script (or alternativly `Rmarkdown`-script) that produces the requested results. You do not need to provide the resulting output in a seperat file. In some cases you have to provide short answers to questions. Please just add in such cases a comment line in your R code for your response. Copy all required `R` and `JAGS` files into one folder and send me a zip-file of this folder via email.

All data sets for the assignments can be found here: `https://github.com/lindemann09/Potsdam-Bayes-2018/tree/master/assignment`

If you have any questions regarding this assignments, please feel free to contact me.

Oliver Lindemann
Email: `lindemann@essb.eur.nl`

Assignment 1: Probabilities

32 of the 120 first year students in psychology were male. Assuming that the gender is equally distributed in the population, what is the probability that psychology is subject with a particular gender bias $\theta = [0, 0.25, 0.5, 0.75, 1]$. Use an uninformative prior.

Assignment 2: Bayesian *t*-test

Data: `exam_anxiety.csv`

> This data set comprises the exam anxiety scores of 103 medicine and sport science students. Higher scores indicate more anxiety.

Examine whether the two groups of students differ in terms of their exam anxiety.

a. Calculate a frequentist and Bayesian t-test using the `BayesFactor`-Package in R.
b. Which hypothesis (null vs. alternative) receives more empirical support? What does the BF tell you, that is, how do you interpret the BF value? (For your short answers just add comment lines in the R code.)

Assignment 3: Bayesian ANOVA

Data: `LooksOrPersonality.Rda`

This data set is taken from the very popular statistics text book of Andy Field (2014). It is descibed in the book as follows:

> Lots of magazines go on and on about how men and women want different things from relationships [...]. The big question seems to be: are looks or personality more important? Imagine you wanted to put this to the test. You devised a cunning plan whereby you'd set up a speed-dating night. [...] Each date varied in their attractiveness (attractive, average or ugly) and their charisma (charismatic, average and dull), and by combining these characteristics you get nine different stooge dates. [...] Obviously you had two sets of stooge dates: one set was male and the other female, so that your participants could match up with dates of their preferred gender.
>
> The participants were not these nine stooges, but 10 men and 10 women who came to the speed-dating event that you had set up. Over the course of the evening they speed-dated all nine stooges of the gender that they'd normally date. After their 3-minute date, they rated how much they'd like to have a proper date with the person as a percentage (100% = '*I'd pay large sums of money for their phone number*', 0% = '*I'd pay a large sum of money for a plane ticket to get me as far away from them as possible*'). As such, each participant rated nine different people who varied in their attractiveness and personality. So, there are two repeated-measures variables: *Looks* (with three levels because the person could be attractive, average or ugly) and *Personality* (again with three levels because the person could have lots of charisma, have some charisma or be a dullard). — Field (2014)

Use the `BayesFator` package to analyse the data with a Bayesian ANOVA with the repeated measures.

a. Perform a Bayesian two-way ANOVA with the repeated measures *Looks* and *Personality* and calculate the bayes factors for all three effects (i.e., *BF*s when omitting from the full model).
b. Make the same analysis with an "ultrawide" Cauchy distribution as prior.
c. Please include also the factor *Gende*r and test if there is evidence for a three-way interaction *Looks* × *Personality* × *Gender*.

Assignment 4: One-sample *t*-test in `JAGS`

Data: `weight.csv`

> A clinic provides a program to help their clients lose weight and asks you to investigate the effectiveness of the program. You have a sample of 15 people, weighing each person before the program begins and 3 months later.

Determine whether the program is effective. To do so, implement the t-test in `JAGS` and use a Cauchy distribution as prior.

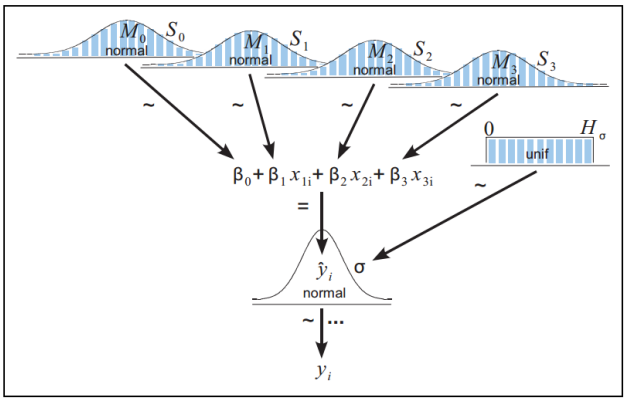
a. Plot the posterior distribution of the δ parameter.
b. Estimate the bayes factor between $\mathcal{H}_0 : \delta = 0$ and $\mathcal{H}_1 : \delta \sim \text{Cauchy}(0, 1)$ by using the Savage-Dickey method (see slides).
c. Use both a normal and an "ultrawide" rscale parameter for the Cauchy prior distribution.

Assignment 5: Modelling Linear Regression

Data: `expenses.csv`

> A researcher is interested in the factors influencing the leisure expenses of 16- and 17-year olds. She collected data from 202 teenagers in order to predict leisure expenses in november (`expenses_nov`) from leisure expenses in July (`expenses_jul`), income from the students' side occupation in november (`salary_nov`), and time spent on studying for high-school (`studytime`).

The following figure depicts a possible regression model with one intercept (β_0 or α) and three beta weights ($\beta_1, \beta_2, \beta_3$) for the three dependent variables:



The priors of these parameters are normally distributed with $M_i = 0$ and large standard deviation (e.g. $\sigma = 100$). The measurements error of the dependent variable is normally distributed with a standard deviation σ (that is, the inverse of the precision parameter λ). The prior for σ is uniform (you might use here `beta(1,1)`).

Task: Please implement this regression model in `JAGS` and apply the data set above. It might be useful to take the `JAGS` model for a simple linear regression with a gamma distributed prior for the precision λ from the workshop (see github) and modify it accordingly. Please standardized the variables (`scale` command in `R`) before entering the data in the regression model.