Bayesian Analysis: Practical Sessions

Session 2

Goals:

- Implement a model with JAGS

Exercise 2.1 Systolic blood pressure: Estimating the mean of a normal location model. It has been established that the standard deviation of the systolic blood pressure everywhere in the world is 13, but its mean varies slightly from country to country around an overall worldwide mean equal to 125. The department of Health of Andorra wants to:

a) compute a point estimate for the mean of the systolic blood pressure of the inhabitants of Andorra, μ,
b) predict the systolic blood pressure of one resident of Andorra,
c) determine whether the mean of the systolic blood pressure of Andorrans is higher than the overall world mean of the systolic blood pressure.

For this purpose, they obtain a sample of the systolic blood pressure of twenty Andorrans, which is: 98, 160 , 136 , 128 , 130 , 114 , 123 , 134 , 128 , 107, 123 , 125 , 129 , 132 , 154 , 115 , 126 , 132 , 136 and 130 .

Exercise 2.2 Linear Regression. Brain weight. Build a linear model to try to explain the brain weight of a mamal as a function of its body weight through the data for 62 mammals and give a credible interval for the brain weight of an animal whose body weight is 100 Kg.

The data is in *Brain.txt* (the body weight is measured in kilograms and the brain weight in grams).

Exercise 2.3 Weight and height. In the file WeightHeight.txt there are data about the weight, height and sex of several students. Implement the next models:

a) A model where the weight is explained as a function of the height.
b) A model where the weight is explained as a function of the height and the sex.
c) A model where the weight is explained as a function of the height, the sex and their interaction.

Exercise 2.4 Yield of Potatoes. A researcher is investigating the relationship between yield of potatoes (y) and level of fertilizer (x). She divides a field into eight plots of equal size and applied fertilizer at a different level to each plot. The level of fertilizer and yield for each plot is recorded below:

Fertilizer level x	Yield y
1	25
1.5	31
2	27
2.5	28
3	36
3.5	35
4	Not available
4.5	34

Suppose that we know that yield given the fertilizer level is $Normal(\beta_0 + \beta_1 x, \sigma)$.

a) Using non-informative priors for the parameters find the posterior distribution of β_1.
b) Find a 95% credible interval for y given x=4.

Exercise 2.5 Final grades from Bayesian analysis course. The lecturers of the Bayesian analysis course want to know if the grades get by the boys and the girls are equal. The final grades are in the next table:

boys	girls
9.6	6.1
7.0	9.1
5.0	8.8
8.0	5.7
8.4	8.9
6.4	6.1
	6.5

The lectures have chosen this statistical model:

$$y \mid \beta_0, \beta_1, \sigma \sim Normal(\beta_0 + \beta_1 sex, \sigma),$$

where *sex* is a dichotomy variable which equals 1 for the boys and 0 for the girls.

Answer the following questions:

a) Write the bayesian model and justify the chosen prior distribution.
b) Update the model and draw the posterior distribution for every parameter.
c) Do you think there are differences between the grades get by the boys and the grades get by the girls?
d) A girl could not take the exam because she stayed trapped in an elevator. Calculate a 90% credibility interval for this student's final grade?.

Exercise 2.6 Leukemia: Time to Event Data. Feigl and Zelen (1965) present data on the survival times in weeks of patients who were diagnosed with leukemia. The patients were classified according to one characteristic of white cells referred to as AG+ and AG-. The n_1=17 times from diagnosis to death for the AG+ group are: 65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65, and the n_2=16 observations for the AG- group are: 56, 65, 17, 7, 16, 22, 3, 4, 2, 3, 8, 4, 3, 30, 4, 43. No prior information is available. Suppose a two-sample exponential model for handling the leukemia data.

a) Draw the survival function for every treatment jointly with their 95% credible interval.
b) Calculate a 95% credible interval for the difference of the 24-week (approximately 6-month) probabilities of survival for the two groups.

Exercise 2.7 Dugong: Non-linear model. To model the length (y) of an individual of the species *Dugong dugong* as a function of its age (x) one can use the data in *Dugong.dat*, obtained by Ratkowsky (1983):

Dugong	1	2	3	4	5	....	26	27
Age (X)	1.0	1.5	1.5	1.5	2.5	....	29.0	31.5
Length (Y)	1.80	1.85	1.87	1.77	2.02	....	2.27	2.57

One possible model is:

$$y_i = \beta_0 - \beta_1 e^{-\beta_2 x_i} + \varepsilon_i,$$

where ε_i follows a *Normal* $(0,\sigma)$. This model is a growth model without inflexion point that is recognized as the Mitscherlich law, the Von Bertalanffy curve or the asymptotic regression.

Information from Wikipedia (https://en.wikipedia.org/wiki/Dugong, visited in april 2017): *The **dugong** (/ˈduːɡɒŋ/, /ˈdjuːɡɒŋ/; Dugong dugon) is a medium-sized marine mammal .... An adult's length rarely exceeds 3 metres (9.8 ft) .... Dugongs are long lived, and the oldest recorded specimen reached age 73.*

Exercise 2.8 Challenger: logistic model. Two fuel tanks of the Challenger space shuttle had three rings each. We want to find out whether there is a relation between the existence of damaged rings in the 23 launchings that preceded the one that exploded, *incident*, and the temperature on the day of the launch, *T*. Build a model for *incident* as a function of the temperature. This set of data is analyzed in Dalal, Fowlkes and Hoadley (1989) and in Lavine (1991), and you can find it in *challenger.dat.*

Exercise 2.9 Nicotine: logistic model. In *Nicotine.dat* we get the results of an experiment on the effect of nicotine on the mortality of aphids. The response is the proportion of dead aphids and the explanatory variable is the concentration of nicotine (*Dose*). Build a model and look for the dose under which the probability of death becomes higher than 0.5.

Exercise 2.10 Surgical: Institutional ranking. This exercise considers mortality rates in 12 hospitals performing cardiac surgery in babies. The data are shown below:

Hospital	No of ops	No of deaths
A	47	0
B	148	18
C	119	8
D	810	46
E	211	8
F	196	13
G	148	9
H	215	31
I	207	14
J	97	8
K	256	29
L	360	24

The objective of this study is to know the probability of death around all the hospitals in the country, not only in the hospitals that are in the sample.