Bayesian Analysis I, FALL 2024

Lab Exercise 5: on **Wednesday Nov. 27, 14-16**

Genera Instruction : Give your BUGS (or JAGS or STAN) codes, output and brief comments for each problem and keep all in a word document, or any other text editors, or latex, etc. For submission, the converted pdf file (from a word document) is preferred in general.

1. Data from a 1991 Gallup opinion poll about the morality of President Bush's not helping Iraqi rebel groups after the formal end of the Gulf War.

Of the 751 adults responding, 150 thought the President's actions were not moral. Suppose we assume a noninformative Beta(0.01, 0.01) prior on the probability π that a randomly sampled adults would respond 'immoral'.

a) Run the samples more than 5000 iterations. Obtain the posterior mean and 95% posterior credible interval on π. Examine the posterior density estimate for π. Repeat with the uniform (0,1) prior. Set the initial values: $\pi = 0.5$.

b) Suppose now our sample size was only 8 with 0 adults considering the Presidential action immoral. Using 3 different priors on π: Beta (0.01,0.01), uniform(0,1)=Beta(1,1), and Beta (0.5,0.5), compare posterior means and 95% posterior credible intervals on π. What do you find when the sample size is small?

c) Use the model to predict the number of replying immoral in new survey. Then, compute the new probability for group responding 'immoral'.

2. The time required for light to travel 7442m were measured. The data indicate deviations from 24,800 nanoseconds: n=66

28,26,33,24,34,-44,27,16,40,-2, 29,22,24,21,25,30,23,29,31,19, 24,20,36,32,36,28,25,21,

28,29, 37,25, 28,26,30,32,36,26,30,22,36,23,27,27,28,27,31,27,26,33, 26,32,32,24,39,28,

24,25,32,25,29,27,28,29,16,23

The usual model that can be used for this data is the normal (μ, τ). However, the data include two large outliers. Thus, we could use robust t distribution with 4 degrees of freedom for the likelihood not excluding those outliers. ('dt(center, scale, df)' in BUGS)

a) Using the noninformative normal(0, 1000000) prior for μ, gamma(0.0001, rate= 0.0001) for $1/\tau$ (precision parameter), give the posterior inference summary for the parameters. Set the initial values as $\mu = 10$ and $1/\tau = 1$(precision) for one chain out of 3 chains. (The initial values for other chains can be selected on your own. Note that gamma(0.0001, scale=1000) is the parameterization we use in our course.)

b) Compare your results in a) with those using the usual normal likelihood.

c) Obtain the predictive distribution of a future observation y_{new} (i.e. 67th observation)

The posterior predictive distributions can be easily obtained using BUGS/JAGS. You can simply add a line indicating the likelihood for y_{new} at the end. (Thus, the generated/updated posterior parameters will be naturally put into (mixed with) this likelihood, resulting in the posterior predictive distribution for y_{new}.)

3. The annual number of airline fatalities can be modelled with a Poisson(λ) distribution : airline data (given in the moodle). Take a Gamma (0.1, rate= 0.1) for the prior distribution for λ.

```
 > getwd()    # To check your current working directory
# To read the dataset in R if it is in the directory:  C:\\Users\\documents\\
 > airline.dat=read.table(file="C:\\Users\\documents\\airline.txt",header=T)
 > head(airline.dat)      # check the data set
```

a) With McMC methods, obtain the posterior inference summary and (approximate) density plot for the mean number of airline fatalities. Give brief comments.

b) We can use this model to predict the future number of fatalities as well. One way (trick) is to use the data statement as follows:

Change the input data - vector of fatal airline accidents by expanding it with a NA for prediction of the number in 2002. Run JAGS via rjags for this modified model and set up a new parameter as below.

```
> a1.par <- c("lambda","fatal[27]")    # For prediction
```

Obtain the posterior summary information/plots as before.

4. A previous study shows that there seems to be a linear relationship between the body length and the body mass of a snake species. Here, we also want to study if the relationship between the body length and the body mass differs between males and females of the snake species. In this data set, the gender of a snake is expressed as 0 (female) or 1 (male).

We assume that each value of body mass follows from a Normal distribution whose mean linearly depends on the body length and the gender of the snake.

We choose uninformative priors for all parameters, using flat distributions which make all plausible values approximately equally likely. Since the standard deviation of a normal distribution can only be positive, we use a uniform(0,100) distribution for the variance parameter.

a) Write the codes for McMC analysis and run 3 chains with different sets of initial values. Obtain the posterior inference summary for the parameters (also with the posterior density estimates).

b) You can check the autocorrelation of samples drawn with McMC using

```
> autocorr.diag(res)
> autocorr.plot(res)
```

Do the diagnostics for convergence of the chains using the Gelman's statistics

```
> gelman.diag(res)
```

Give your output and brief comments.