Introduction to Bayesian Data Analysis

Tutorial 5

(1) Consider the 1998 General Social Survey, which recorded the number of children for 921 women over the age of 40. Let's take these N=921 women as our population and consider estimating the mean number of children for this population from a sample size of n=45. The number of children per women is assumed to be normally distributed. Evaluate the approriateness of the normal distribution assumption. Assume the following values for parameters of the conjugate prior distribution: $\kappa_0 = 1$, $\mu_0 = 2$; $\nu_0 = 1$, $\sigma_0^2 = 1$

Read in the data with the following commands. (The file `alldata.R` must be stored in your working directory)

```
load("alldata")
CHILDS<-Y$CHILDS[Y$FEMALE==1&Y$YEAR==1998 & Y$AGE>=40  ]
CHILDS<-CHILDS[!is.na(CHILDS) ]
```

(2) Problem 5.1 (Hoff) The files `school.dat` , `school2.dat` and `school3.dat` contain data on the amount of time students from three high schools spent on studying homework during an exam period. Analyse the data from each of these schools separately, using the normal model with a conjugate prior distribution, in which $\{\mu_0 = 5, \sigma_0^2 = 4, \kappa_0 = 1, \nu_0 = 2\}$ and compute or approximate the following:

(a) posterior means and 95% confidence intervals for the mean θ and standard deviation σ from each school

(b) the posterior probability that $\theta_i < \theta_j < \theta_k$ for all six permutations $\{i, j, k\}$ of $\{1, 2, 3\}$

(c) the posterior probability that $\tilde{Y}_i < \tilde{Y}_j < \tilde{Y}_k$ for all six permutations $\{i, j, k\}$ of $\{1, 2, 3\}$, where $\tilde{Y}_i$ is a sample from the posterior predictive distribution of school i.

(d) Compute the posterior probability that θ_1 is bigger than both θ_2 and θ_3, and the posterior probability that $\tilde{Y}_1$ is bigger than both $\tilde{Y}_2$ and $\tilde{Y}_3$

(3) Problem 5.2 (Hoff) Sensitivity analysis: Thirty-two students in a science classroom were randomly assigned to one of two study methods, A and B, so that $n_A = n_B = 16$ students were assigned to each method. After several weeks of study, students were examined on the course material with an exam designed to give an average score of 75 with a standard deviation of 10. The scores for the two groups are summarized by the statistics $\{\bar{y}_A = 75.2, s_A = 7.3\}$ and $\{\bar{y}_B = 77.5, s_B = 8.1\}$. Consider independent, conjugate normal prior distributions for each of θ_A and θ_B, with $\mu_0 = 75$ and $\sigma_0^2 = 100$ for both groups. For each $(\kappa_0, \nu_0) \in \{(1,1), (2,2), (4,4), (8,8), (16,16), (32,32)\}$, obtain $Pr(\theta_A < \theta_B | \mathbf{y}_A, \mathbf{y}_B)$ via Monte Carlo sampling. Plots these probabilities as a function of $\kappa_0 = \nu_0$. Describe how you might use this plot to convey the evidence that $\theta_A < \theta_B$ to people of a variety of prior opinions.

(4) Comparison of two multinomial observations: On the eve of a presidential campaign debate, ABC news conducted a survey of registered voters; 639 persons were polled before the debate and 639 different persons were polled after. The results were:

Survey	Candidate X	Candidate Y	No opinion/other	Total
pre-debate	294	307	38	639
post-debate	288	332	19	639

Assume the surveys are independent simple random samples from the population of registered voters. Model the data with two different multinomial distributions. For $j = 1, 2$, let α_j be the proportion of voters who preferred Candidate X, out of those who had a preference for either Candidate X or Candidate Y at the time of survey j. Plot a histogram of the posterior density $\alpha_2 - \alpha_1$. What is the posterior probability that there was a shift toward Candidate X?