ST4234: Bayesian Statistics

Tutorial 5, AY 19/20

Instructions: You do not need to submit your work. I will discuss the solution on **6pm Tuesday, 10 March**.

1. The following table gives the records of accidents in 1998 compiled by the Department of Highway Safety and Motor Vehicles in Florida. Denote the number of accidents and fatalities

Safety Equipment in Use	Injury	
	Fatal	Nonfatal
None	1601	162527
Seat belt	510	412368

when no safety equipment was in use by n_N and y_N, respectively. Similarly, let n_S and y_S denote the number of accidents and fatalities when a seat belt was in use. Assume that y_N and y_S are independent with $y_N \sim \text{Binomial}(n_N, p_N)$ and $y_S \sim \text{Binomial}(n_S, p_S)$. Assume $p(p_N, p_S) \propto 1$.

(a) Show that p_N and p_S have independent beta posterior distributions.

(b) Use the function `rbeta` to simulate 10000 values from the joint posterior distribution of (p_N, p_S). Using your sample, construct a histogram of

(i) the relative risk p_N/p_S. Find a 95% quantile-based credible interval of this relative risk.

(ii) the difference in risks $p_N - p_S$. Compute the posterior probability that the difference in risks exceeds 0.

2. The files `school1.txt`, `school2.txt`, `school3.txt` contain data on the amount of time students from three high schools spent on homework during an exam period. Analyze data

from each school separately, using the normal model $\mathrm{N}(\theta, \sigma^2)$ with a conjugate prior distribution $\theta \sim \mathrm{N}(\mu_0, \sigma^2/n_0)$ and $\sigma^2 \sim \text{Inv-Gamma}(\nu_0/2, S_0/2)$, in which $\{\mu_0 = 5, n_0 = 1, \nu_0 = 2, S_0 = 8\}$. Compute or approximate the following:

(a) posterior means for the mean θ and standard deviation σ from each school.

(b) the posterior probability that $\theta_3 < \theta_2 < \theta_1$.

(c) the posterior probability that $\tilde{Y}_3 < \tilde{Y}_2 < \tilde{Y}_1$, where $\tilde{Y}_i$ is a sample from the posterior predictive distribution of school i.

(d) 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 $\tilde{Y}_3$.

Hint: To read the `.txt` files in R, you can first change the R directory to the folder where the datasets are located, and then use the following command

```
school1 <- read.table("school1.txt",header=FALSE)
```