HOMEWORK 9, STAT 251

Remember to submit your R code along with the completed homework.

(1) A study was done to see if taking a fish oil supplement lowered blood pressure. For the seven subjects the change in blood pressure was: 8, 12, 10, 14, 2, 0, 0. The study organizers felt it was reasonable to assume the data were conditionally *iid* and normally distributed with known mean $\mu = 6.3$ and variance σ^2. For the prior distribution for σ^2, we have chosen an inverse-gamma with shape=2.01 and rate=49.

   (a) Derive the posterior distribution of σ^2. (That is, show that the posterior distribution of σ^2 is an inverse-gamma)

   (b) Refer to part (a). Plot the prior and the posterior distributions for σ^2 on the same graph. Make sure to include a legend to distinguish the prior from the posterior.

(2) Ten policemen working in downtown Cairo had their blood sampled to determine the level of lead concentration. We assume the data are conditionally iid with the $N(\mu = 20, \sigma^2)$ distribution. The ten observed data points were: 17.4, 13.4, 27.3, 25.1, 23.4, 13.6, 38.2, 23.5, 20.7, 28.3. The prior for σ^2 is the inverse gamma distribution with shape=4.5 and rate=240 (i.e., $\sigma^2 \sim IG(4.5, 240)$).

   (a) Plot the prior and posterior distributions for the variance on the same graph. Make sure to include a legend to distinguish the prior from the posterior.

   (b) What is the posterior probability that the variance in blood lead concentration of the population of policemen in downtown Cairo is less than 49?

   (c) Plot the posterior predictive distribution for the *next* downtown Cairo police officer's blood lead concentration.

(3) Here is the total serum cholesterol for 9 urban residents of Guatemala: 197, 199, 214, 217, 222, 223, 227, 228, 234. It is assumed that $Y_{urban,i}|\sigma^2_{urban} \overset{iid}{\sim} N(\mu_{urban} = 220, \sigma^2_{urban})$, where $Y_{urban,i}$ denotes the total serum cholesterol for the i^{th} individual in the sample of Guatemalan urban residents. The prior belief about σ^2_{urban}, the variance of the serum cholesterol of urban residents of Guatemala, is that it has an inverse gamma distribution with shape=2.5 and rate=600.

   (a) What is the posterior distribution of the variance in this case?

   (b) What is the prior expected value of σ^2_{urban}?

   (c) What is the posterior expected value of σ^2_{urban}?

   (d) What is the posterior mode of σ^2_{urban}?

   (e) Plot the posterior and prior distribution of σ^2_{urban} on the same graph. Make sure to include a legend to distinguish the prior from the posterior.

(4) Here is the total serum cholesterol for 10 rural residents of Guatemala: 139, 142, 143, 144, 145, 148, 155, 162, 171, 181. It is assumed that $Y_{rural,i}|\sigma^2_{rural} \overset{iid}{\sim} N(\mu_{rural} = 150, \sigma^2_{rural})$, where $Y_{rural,i}$ denotes the total serum cholesterol for the i^{th} individual in the sample of Guatemalan rural residents. The prior belief about the variance of serum cholesterol in rural Guatemalan residents is an inverse gamma with shape=2.5 and rate=600.

   (a) What is the posterior distribution of the variance in this case?

   (b) Plot the posterior.

(5) Now consider the ratio of variances of the two groups: $\sigma^2_{urban}/\sigma^2_{rural}$.

   (a) Plot the (Monte Carlo-estimated) posterior distribution of

$$\frac{\sigma^2_{urban}}{\sigma^2_{rural}}.$$

   (b) Report and interpret a 95% posterior credible interval for this ratio.

   (c) Explain why it is important to determine whether the CI for the ratio of variances includes 1. (Hint: What does it imply if the ratio equals 1?)

(6) Plot the prior predictive distribution of Y_{urban} and the posterior predictive distribution of Y_{urban} on the same graph. (This of course requires a legend be included in your plot to distinguish one from the other.)