DDA 4010 − Bayesian Statistics

Exercise Sheet 2

This exercise is due on **Oct. 21th, 5:00 pm.**

Assignment A2.1 (3.9 **in Textbook):**
Galenshore distribution: An unknown quantity Y has a Galenshore (a, θ) distribution if its density is given by

$$p(y) = \frac{2}{\Gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2}$$

for $y > 0, \theta > 0$ and $a > 0$. Assume for now that a is known. For this density,

$$\mathrm{E}[Y] = \frac{\Gamma(a + 1/2)}{\theta \Gamma(a)}, \quad \mathrm{E}\left[Y^2\right] = \frac{a}{\theta^2}$$

- Identify a class of conjugate prior densities for θ. Plot a few members of this class of densities.
- Let $Y_1, \ldots, Y_n \sim$ i.i.d. Galenshore (a, θ). Find the posterior distribution of θ given $Y_1, \ldots, Y_n$, using a prior from your conjugate class.
- Write down $p\left(\theta_a \mid Y_1, \ldots, Y_n\right) / p\left(\theta_b \mid Y_1, \ldots, Y_n\right)$ and simplify. Identify a sufficient statistic.
- Determine $\mathrm{E}\left[\theta \mid Y_1, \ldots, Y_n\right]$.
- Determine the form of the posterior predictive density $p\left(\tilde{Y} \mid Y_1, \ldots, Y_n\right)$.

Assignment A2.2 (3.13 **in Textbook):**
3.13 Improper Jeffreys' prior: Let $Y \sim \text{Poisson}(\theta)$.

- Apply Jeffreys' procedure to this model, and compare the result to the family of gamma densities. Does Jeffreys' procedure produce an actual probability density for θ ? In other words, can $\sqrt{I(\theta)}$ be proportional to an actual probability density for $\theta \in (0, \infty)$?
- Obtain the form of the function $f(\theta, y) = \sqrt{I(\theta)} \times p(y \mid \theta)$. What probability density for θ is $f(\theta, y)$ proportional to? Can we think of $f(\theta, y) / \int f(\theta, y) d\theta$ as a posterior density of θ given $Y = y$?

Assignment A2.3 (4.2 **in Textbook):**
Tumor counts: A cancer laboratory is estimating the rate of tumorigenesis in two strains of mice, A and B. They have tumor count data for 10 mice in strain A and 13 mice in strain B. Type A mice have been well studied, and information from other laboratories suggests that type A mice have tumor counts that are approximately Poisson-distributed with a mean of 12. Tumor count rates for type B mice are unknown, but type B mice are related to type A mice. The observed tumor counts for the two populations are

$$\boldsymbol{y}_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6)$$
$$\boldsymbol{y}_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)$$

- For the prior distribution given in part a) of that exercise, obtain $\Pr(\theta_B < \theta_A \mid \boldsymbol{y}_A, \boldsymbol{y}_B)$ via Monte Carlo sampling.
- For a range of values of n_0, obtain $\Pr(\theta_B < \theta_A \mid \boldsymbol{y}_A, \boldsymbol{y}_B)$ for $\theta_A \sim$ gamma $(120, 10)$ and $\theta_B \sim$ gamma $(12 \times n_0, n_0)$. Describe how sensitive the conclusions about the event $\{\theta_B < \theta_A\}$ are to the prior distribution on θ_B.
- Repeat parts a) and b), replacing the event $\{\theta_B < \theta_A\}$ with the event $\left\{\tilde{Y}_B < \tilde{Y}_A\right\}$, where $\tilde{Y}_A$ and $\tilde{Y}_B$ are samples from the posterior predictive distribution.

Assignment A2.4 (4.8 **in Textbook):**
More posterior predictive checks: Let θ_A and θ_B be the average number of children of men in their 30 s with and without bachelor's degrees, respectively.

- Using a Poisson sampling model, a gamma $(2, 1)$ prior for each θ and the data in the files menchild30bach. dat and menchild30nobach. dat, obtain 5,000 samples of $\tilde{Y}_A$ and $\tilde{Y}_B$ from the posterior predictive distribution of the two samples. Plot the Monte Carlo approximations to these two posterior predictive distributions.
- For the moment, suppose you believed that $\theta \in \{0.0, 0.1, \ldots, 0.9, 1.0\}$. Given that the results of the survey were $\sum_{i=1}^{100} Y_i = 57$, compute $\Pr\left(\sum Y_i = 57 \mid \theta\right)$ for each of these 11 values of θ and plot these probabilities as a function of θ.
- Find 95% quantile-based posterior confidence intervals for $\theta_B - \theta_A$ and $\tilde{Y}_B - \tilde{Y}_A$. Describe in words the differences between the two populations using these quantities and the plots in a), along with any other results that may be of interest to you.
- Obtain the empirical distribution of the data in group B. Compare this to the Poisson distribution with mean $\hat{\theta} = 1.4$. Do you think the Poisson model is a good fit? Why or why not?
- For each of the $5,000 \theta_B$-values you sampled, sample $n_B = 218$ Poisson random variables and count the number of 0 s and the number of 1 s in each of the 5,000 simulated datasets. You should now have two sequences of length 5,000 each, one sequence counting the number of people having zero children for each of the 5,000 posterior predictive datasets, the other counting the number of people with one child. Plot the two sequences against one another (one on the x-axis, one on the y-axis). Add to the plot a point marking how many people in the observed dataset had zero children and one child. Using this plot, describe the adequacy of the Poisson model.

Please also submit the R code for Question 4.8. The data can be found in the attachment.

Sheet 2 is due on **Oct. 21th**. Submit your solutions before **Oct. 21th**, **5:00 pm**.