Bayesian Statistics, Assignment 1
Collegio Carlo Alberto
Due date: November 2^{nd}, 2022

Question 1 (15 points): The Galenshore distribution (Hoff 3.9).

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,

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

a) Identify a class of conjugate prior densities for θ. Plot a few (e.g. 4 or 6) members of this class of densities.

b) Let $Y_1, \dots, Y_n \mid \theta \overset{iid}{\sim}$ Galenshore(a, θ). Find the posterior distribution of θ given $Y_{1:n} = y_{1:n}$, using a prior from your conjugate class.

c) Write down $\frac{p(\theta_a|y_{1:n})}{p(\theta_b|y_{1:n})}$ and simplify. Identify a sufficient statistic.

d) Determine $\mathbb{E}[\theta \mid y_{1:n}]$.

e) Determine the form of the posterior predictive density $p(y_{n+1} \mid y_{1:n})$.

Question 2 (15 points): Tumor counts (Hoff 3.3, 4.2, 4.3).

Part 1 (5 points): Tumor counts (Hoff 3.3).

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 (you can find them in the file *dataAssignment1.RData*)

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

a) Find the posterior distributions, means, variances and 95% quantile-based credible intervals for θ_A and θ_B, assuming a Poisson sampling distribution for each group and the following prior distribution:

$$\theta_A \sim \text{Gamma}(120, 10), \quad \theta_B \sim \text{Gamma}(12, 1), \quad p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$$

b) Compute and plot the posterior expectation of θ_B under the prior distribution $\theta_B \sim \text{Gamma}(12 \times n_0, n_0)$ for each value of $n_0 \in \{1, 2, \dots, 50\}$. Describe what sort of prior beliefs about θ_B would be necessary in order for the posterior expectation of θ_B to be close to that of θ_A.

c) Should knowledge about population A tell us anything about population B? Discuss whether or not it makes sense to have $p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$.

Part 2 (5 points): Tumor count comparison (Hoff 4.2).

d) For the prior distribution given in point a), obtain $\mathbb{P}(\theta_B < \theta_A \mid \boldsymbol{y}_A, \boldsymbol{y}_B)$ via Monte Carlo sampling.

e) For a range of values of n_0, obtain $\mathbb{P}(\theta_B < \theta_A \mid \boldsymbol{y}_A, \boldsymbol{y}_B)$ for $\theta_A \sim \text{Gamma}(120, 10)$ and $\theta_B \sim \text{Gamma}(12 \times n_0, n_0)$. Describe how sensitive the conclusions about the event $\{\theta_B < \theta_A\}$ are to the prior distribution on θ_B.

f) Repeat points d) and e), replacing the event $\{\theta_B < \theta_A\}$ with the event $\{\tilde{Y}_B < \tilde{Y}_A\}$, where $\tilde{Y}_A$ and $\tilde{Y}_B$ are samples from the posterior predictive distribution.

Part 3 (5 points): Posterior predictive checks (Hoff 4.3).

Let's investigate the adequacy of the Poisson model for the tumor count data. Generate posterior predictive datasets $\boldsymbol{y}_A^{(1)}, \ldots, \boldsymbol{y}_A^{(1000)}$. Each $\boldsymbol{y}_A^{(s)}$ is a sample of size $n_A = 10$ from the Poisson distribution with parameter $\theta_A^{(s)}$, $\theta_A^{(s)}$ is itself a sample from the posterior distribution $p(\theta_A \mid \boldsymbol{y}_A)$, and $\boldsymbol{y}_A$ is the observed data.

g) For each s, let $t^{(s)}$ be the sample average of the 10 values of $\boldsymbol{y}_A^{(s)}$, divided by the sample standard deviation of $\boldsymbol{y}_A^{(s)}$. Make a histogram of $t^{(s)}$ and compare to the observed value of this statistic. Based on this statistic, assess the fit of the Poisson model for these data.

h) Repeat the above goodness of fit evaluation for the data in population B.