Biostat 682 Homework 2

Due: Thursday, October 2, 2025 (23:59 pm)

Note: you may use computing program (R or Python) to answer questions related to numerical calculations and visualization. Please include code in your submission.

1. Recall the Cancer at Slater School example in Lecture 2 note. A total of 145 employees were at risk and eight cases of invasive cancer were found. We assume that the number of cancers, y, among the 145 employees, has a binomial distribution with parameter θ representing the chance of cancer. Suppose we want to perform Bayesian analysis for this problem.

   (a) What is the natural conjugate prior family for θ? Suppose we believe that the prior mean of θ is 0.03 and the prior sample size is 100. Determine the prior distribution for θ and find the posterior mean.

   (b) Suppose we would like assign a flat prior to the log odds of cancers, $\text{logit}(\theta)$, i.e. $\pi\{\text{logit}(\theta)\} \propto 1$. Find the prior distribution for θ and find the posterior mean.

   (c) Suppose another school, Lawton, has 100 employees with the same chance of cancers as the Slater school. Find the predictive distribution of the number of cancers for Lawton given the data from Slater.

2. In the Birth Rate example of the Lecture 3 note, consider an alternative prior specification: θ_1 and θ_2 are independent with $\theta_1 \sim \text{G}(2, 1)$ and $\theta_2 \sim \text{G}(200, 100)$ and answer the three questions on Page 11.

   (a) How likely the average number of children for women without college degree is large than the average number of children for women with college degree given the data?

   (b) How likely the number of children of a woman with college degree is strictly larger than that of a woman without college degree given the data?

   (c) How likely the number of children of a woman with college degree is exactly the same than that of a woman without college degree given the data?

3. Suppose $y_1, \ldots, y_m \overset{\text{i.i.d.}}{\sim} \text{Binomial}(\theta, n)$, where $\theta \in (0, 1)$ and m and n are positive integers.

   (a) Find the Jeffrey's prior for θ and the posterior distribution;

   (b) Find the Jeffrey's prior for $\text{logit}(\theta)$ and the posterior density function;

   (c) If we assign a uniform prior on θ, i.e. $\theta \sim \text{Uniform}(0, 1)$, find the prior and posteiror distribution for $\text{logit}(\theta)$.

   (d) When $m = 4$, $n = 100$, $y = (2, 0, 2, 4)$, compare the posterior density plot for $\text{logit}(\theta)$ between part (b) and part (c).

4. We study whether the dosage x_i of a new treatment affects the clinical response y_i among n subjects ($i = 1, \ldots, n$). The model is

$$y_i = \beta\delta x_i + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \tau^{-2}),$$

   where $\beta \in \mathbb{R}$, $\delta \in \{0, 1\}$ and $\tau^2 > 0$ are unknown. The latent indicator δ indicates whether dosage has an effect, while β represents the effect size if $\delta = 1$. Independent priors are assumed:

$$\pi(\beta) \propto 1, \qquad \pi(\tau^2) \propto 1/\tau^2, \qquad \delta \sim \text{Bernoulli}(0.5),$$

   A dataset (`treatment_data.csv`) with (x_i, y_i) is provided on Canvas Files (`data` folder)

   (a) Is dosage clinically effective?

i. Derive the posterior distribution of δ given the data, expressing $\Pr(\delta = 1 \mid \text{data})$ in terms of marginal likelihoods $m_0(y)$ and $m_1(y)$, where $m_k(y) = \mathbb{E}\{\prod_{i=1}^{n} \pi(y_i \mid \beta, \tau^2, \delta = k)\}$ for $k = 0, 1$ and $\mathbb{E}(\cdot)$ is taken with respect to prior distribution of β and τ^2.

ii. Using the provided dataset, compute $\Pr(\delta = 1 \mid \text{data})$.

iii. Interpret your results: what is the posterior evidence that dosage affects patient response?

(b) Predictive comparison for two new patients. Consider two new patients: Patient A with low dosage: -2 and Patient B with higher dosage: 2.

i. Derive the posterior predictive distribution of y^* for a new patient with dosage x^*, showing it is a mixture of t distributions under $\delta = 0$ and $\delta = 1$.

ii. Compute posterior predictive means and 95% predictive intervals for Patients A and B.

iii. Estimate the posterior probability $\Pr(y_B^* > y_A^* \mid \text{data})$.

iv. Discuss: how do these predictive results inform clinical decision-making about whether dosage should be adjusted in practice?