Project 1

STAT40850 Bayesian Analysis (online)

2024/2025

AUTHOR

Alberto Caimo

Instructions

- Hand-in date: **Wednesday 12th February 2025 at 11:59pm.**
- You should submit it to the Assignment 1 object in Brightspace.
- You should submit a pdf file and corresponding source file (e.g., rmd or qmd file) containing your answers to the questions.
- The maximum length of the PDF document is 4 pages. Any submissions exceeding this limit will incur a penalty.
- You may submit it multiple times before the deadline, but only the last version will be marked.

Dataset

We want to compare the effectiveness of two different teaching methods: the current method (Method A) and a new method (Method B) in helping students pass an exam.

The two main questions of interest concern the estimate of the exam pass rate for Method A and the assessment of whether Method B is more successful than Method A.

To answer this question we have collected a random sample of students who passed and failed using each method and we will model the outcomes for each group using independent binomial distributions.

The dataset is the following:

- Method A: 169 out of 296 students passed the exam;
- Method B: 247 out of 380 students passed the exam;

Questions

Focus on Method A. Use **Base R** to:

1. Specify and plot a Beta prior distribution $p(\theta_A)$ with mean $E(\theta_A) = 0.65$. Calculate the prior probability $\Pr(\theta_A < 0.5)$.

[10]

2. Estimate and plot the posterior distribution $p(\theta_{\mathsf{A}} \mid x_{\mathsf{A}})$. Calculate the posterior probability $\Pr(\theta_{\mathsf{A}} > 0.7)$.

[20]

3. Present a brief commentary on the results obtained focusing on whether there is any evidence against the hypothesis that $\theta_{\mathsf{A}} = 0.60$ a posteriori.

[10]

4. Estimate (via Monte Carlo sampling) and plot the posterior predictive distribution $p(\tilde{x}_{\mathsf{A}} \mid x_{\mathsf{A}})$. Comment on the fit of the model to the observed data. Use the sample simulated from the posterior predictive distribution to estimate the probability $\Pr(\tilde{x}_{\mathsf{A}} \geq 180 \mid x_{\mathsf{A}})$.

[20]

Now consider both Method A and B.

5. Use **Stan** to estimate the posterior distribution for the probability of passing the exam for each group (θ_{A} and θ_{B}) using the same prior adopted above for θ_{A} and a symmetric prior for θ_{B}. Estimate the posterior distribution of the difference $\theta_{\mathsf{diff}} = \theta_{\mathsf{B}} - \theta_{\mathsf{A}}$.

[20]

6. Plot and summarise the posterior distributions estimated in the previous question; estimate of the posterior probability $\Pr(\theta_{\mathsf{diff}} < 0 \mid x_{\mathsf{A}}, x_{\mathsf{B}})$ by using the Stan MCMC output; present a brief commentary on the results obtained.

[20]