Homework 4

Shao-Ting Chiu (UIN:433002162)

9/23/22

Description

- Course: STAT638, 2022 Fall

  Read Chapter 4 in Hoff.

  Then, do the following exercises in Hoff: 4.1, 4.2, 4.6, 4.8

  All datasets in the Hoff book can be downloaded from https://pdhoff.github.io/book/ (Links to an external site.).

- Deadline: `Oct 4 by 12:01pm`

Problem 4.1

Posterior comparisons: Reconsider the sample survey in Exercise 3.1. Suppose you are interested in comparing the rate of support in that county to the rate in another county. Suppose that a survey of sample size 50 was done in the second county, and the total number of people in the sample who supported the policy was 30. Identify the posterior distribution of θ_2 assuming a uniform prior. Sample 5000 values of each of θ_1 and θ_2 from their posterior distributions and estimate $Pr(\theta_1 < \theta_2 |$ the data and prior $)$.

Problem 4.2

Tumor count comparisons: Reconsider the tumor count data in Exercise 3.3:

(a)

For the prior distribution given in part (a) of that exercise, obtain $Pr(\theta_B < \theta_A | y_A, y_B)$ via Monte Carlo sampling.

(b)

For a range of values of n_0, obtain $Pr(\theta_B < \theta_A | y_A, y_B)$ 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.

(c)

Repeat part (a) and (b), 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.

Problem 4.6

Non-informative prior distributions: Suppose for a binary sampling problem we plan on using a uniform, or $beta(1, 1)$, prior for the population proportion θ. Perhaps our reasoning is that this represents "no prior information about θ." However, some people like to look at proportions on the log-odds scale, that is, they are interested in $\gamma = \log \frac{\theta}{1-\theta}$. Via Monte Carlo sampling or otherwise, find the prior distribution for γ that is induced by the uniform prior for θ. Is the prior informative about γ?

Problem 4.8

More posterior predictive checks: Let θ_A and θ_B be the average number of children of men in their 30s with and without bachelor's degrees, respectively.

(a)

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 5000 samples of $\bar{Y}_A$ and $\bar{Y}_B$ from the posterior predictive distribution of the two samples. Plot the Monte Carlo approximations to these two posterior predictive distributions.

(b)

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 of interest to you.

(c)

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?

(d)

For each of the 5000 θ_B-valeus you sampled, sample $n_B = 218$ Poisson random variables and count the number of 0s and the number of 1s in each of the 5000 simulated datasets. You should now have tow sequences of length 5000 each, one sequence counting the number of people having zero children for each of the 5000 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 adequency of the Poisson model.

References