Homework 5

Shao-Ting Chiu (UIN:433002162)

10/2/22

Table of contents

Description

- Course: STAT638, 2022 Fall

  Read Chapter 5 in the Hoff book.

  Then, do the following exercises in Hoff: 5.1, 5.2, 5.5.

  For 5.2, _A denotes the mean exam score for students assigned to method A, and _B denotes the mean exam score for students assigned to method B.

  Be careful: Some of the notation differs between the lecture slides and the Hoff book.

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

Computational Envirommnent Setup

Third-party libraries

```
%matplotlib inline
import sys # system information
import matplotlib # plotting
import scipy # scientific computing
import random
import pandas as pd # data managing
from scipy.special import comb
from scipy import stats as st
from scipy.special import gamma
import numpy as np
import matplotlib.pyplot as plt
# Matplotlib setting
plt.rcParams['text.usetex'] = True
matplotlib.rcParams['figure.dpi']= 300
np.random.seed(20220928) # Consistent random effect
```

Version

```
print(sys.version)
print(matplotlib.__version__)
print(scipy.__version__)
print(np.__version__)
print(pd.__version__)
```

```
3.8.12 (default, Oct 22 2021, 18:39:35)
[Clang 13.0.0 (clang-1300.0.29.3)]
3.3.1
1.5.2
1.19.1
1.1.1
```

Problem 5.1

Studying: The files `school1.dat`, `school2.dat` and `school3.dat` contina data on the amount of time students from 3 high schools spent on studying or homework during an exam period. Analyze data from each of these schools separately, using the normal model with a conjugate prior distribution, in which $\{\mu_0 = 5, \sigma_0^2 = 4, \kappa_0 = 1, \nu = 2\}$ and compute or approximate following:

(a)

Posterior means and 95% confidence intervals for the mean θ and standard deviation σ fomr each school;

(b)

The posterior probability that $\theta_i < \theta_j < \theta_k$ for all siz permutataions $\{i, j, k\}$ of $\{1, 2, 3\}$;

(c)

The posterior probability that $\tilde{Y}_i < \tilde{Y}_j < \tilde{Y}_k$ for all siz permuatations $\{i, j, k\}$ of $\{1, 2, 3\}$, where $\tilde{Y}_i$ is a sample from the posterior predictive distribution of school i.

(d)

Compute the posterior probability that θ_1 is bigger than both θ_2 and θ_3, and the posterior probability that $\tilde{Y}_1$ is bigger than both $\tilde{Y}_2$ and $\tilde{Y}_3$.

Problem 5.2

Sensitivity analysis: Thirty-two students in a science classroom were randomly assigned to one of two study methods, A and B, so that $n_A = n_B = 16$ students were assigned to each method. After several weeks of study, students were examined on the course material with an exam designed to give an average score of 75 with a standard deviation of 10. The scores for the two groups are summarized by $\{\bar{y}_A = 75.2, s_A = 7.3\}$ and $\{\bar{y}_B = 77.5, s_b = 8.1\}$. Consider independent, conjugate normal prior distributions for each of θ_A and θ_B, with $\mu_0 = 75$ and $\sigma_0^2 = 100$ for both groups. For each $(\kappa_0, \nu_0) \in \{(1,1),(2,2),(4,4),(8,8),(16,16),(32,32)\}$ (or more values), obtain $Pr(\theta_A < \theta_B | y_A, y_B)$ via Monte Carlo Sampling. Plot this probability as a function of $\kappa_0 = \nu_0$. Describe how you might use this plot to convey the evidence that $\theta_A < \theta_B$ to people of a variety of prior opinions.

Problem 5.5

Unit information prior: Obtain a unit information prior for the normal model as follows:

(a)

Reparameterize the normal model as $p(y|\theta,\psi)$, where $\psi = \frac{1}{\sigma^2}$. Write out the log likelihood $l(\theta,\psi|y) = \sum \log p(y_i|\theta,\psi)$ in terms of θ and ψ.

(b)

Find a probability density $P_U(\theta,\psi)$ so that $\log p_U(\theta,\psi) = \frac{l(\theta,\psi|y)}{n} + c$, where c is a constant that does not depend on θ or ψ.

Hint: Write $\sum(y_i - \theta)^2$ as $\sum(y_i - \bar{y} + \bar{y} - \theta)^2 = \sum(y_i - \bar{y})^2 + n(\theta - \bar{y})^2$, and recall that $\log p_U(\theta,\psi) = \log p_U(\theta|\psi) + \log p_U(\psi)$.

(c)

Find a probability density $p_U(\theta,\psi|y)$ that is proportional to $p_U(\theta,\psi) \times p(y_1,\ldots,y_n|\theta,\psi)$. It may be convenient to write this joint density as $p_U(\theta|\psi,y) \times p_U(\psi|y)$. Can this joint density be considered a posterior density?

References