Bayesian Statistics Homework

Due: March 28, 2024

Copy of the homework of others is extremely prohibited!

1. Let $y_1, ..., y_n \sim Gamma(a, a\theta)$ where a is known and $\theta > 0$.
   (a) Compute the Jeffreys prior for θ.
   (b) Identify a conjugate prior and resulting posterior for θ.
   (c) Now assume that a is not known, but it is to be fixed using an empirical Bayes approach. Define and justify a value for a in terms of the sample mean and variance of $y_1, ..., y_n$, denote $\bar{y}$ and s^2, respectively.

2. Suppose to toss a coin 12 times with the result of 10 tails and 2 heads, in the uniform prior case, predict the number of tails if we toss the coin 5 times more.

3. Suppose the random variables

$$X_1, \ldots, X_n|\mu \overset{iid}{\sim} N(\mu, \sigma_1^2) \quad \sigma_1^2 \text{ known}$$
$$\mu \sim N(\mu_0, \sigma_0^2)$$

   What is the posterior predictive distribution $p(X_{n+1|X_1 \ldots X_n})$? Show the calculation.

4. I recorded the attendance of students at tutorials for a module. Suppose that we can, in some sense, regard the students as a sample from some population of students so that, for example, we can learn about the likely behaviour of next year's students by observing this year's. At the time I recorded the data we had had tutorials in Week 2 and Week 4. Let the probability that a student attends in both weeks be θ_{11}, the probability that a student attends in week 2 but not Week 4 be θ_{10} and so on. The data are as follows.

Attendance	Probability	Observed frequency
Week 2 and Week 4	θ_{11}	$n_{11} = 25$
Week 2 but not Week 4	θ_{10}	$n_{10} = 7$
Week 4 but not Week 2	θ_{01}	$n_{01} = 6$
Neither week	θ_{00}	$n_{00} = 13$

   Suppose that the prior distribution for $(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$ is a Dirichlet distribution with density proportional to

$$\theta_{11}^3 \theta_{10} \theta_{01} \theta_{00}^2$$

- (a) Find the prior means and prior variances of $\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}$.
- (b) Find the posterior distribution.
- (c) Find the posterior means and posterior variances of $\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00}$.
- (d) Using the relevant R function (either from a package or otherwise), find a 95% posterior interval, based on the exact posterior distribution, for θ_{00}

5. Ten measurements are made using a scientific instrument. Given the unknown value of a quantity θ, the natural logarithms of the measurements are independent and normally distributed with mean $log\theta$ and known standard deviation 0.05. Our prior distribution is such that $log\theta$ has a normal distribution with mean 2.5 and standard deviation 0.5. The logarithms of the measurements are as follows.
   2.99 3.03 3.04 3.01 3.12 2.98 3.03 2.98 3.07 3.10

   - (a) Find the posterior distribution of $\log\theta$.
   - (b) Find a symmetric 95% posterior interval for $\log\theta$
   - (c) Find a symmetric 95% posterior interval for θ.
   - (d) Find the posterior probability that $\theta < 20.0$.