Homework 4

Instructions: This homework is due in class on Friday May 10.

Please read the following guidelines for presenting your work and follow them diligently.

- Write your full name clearly on the top right of the first page. **Staple** pages on the left hand corner. Write neatly in complete sentences.
- You are required to work all the problems, however, only 5 will be graded. The page numbers below refer to the fifth edition of the text.
- Answer the questions in the order in which they are posed. Clearly number the questions as I have.
- You must first work independently on the homework. Please post questions on the discussion board or come to office hours once you have tried the problems.
- Be sure to show/explain your work thoughtfully. How you write your answers is important.
- If you use R to make plots or as a calculator, it is enough to simply include the output (e.g., appropriately labeled plot) in the main part of your homework without the R code.

---

1. Let $X_1, X_2, \ldots, X_n \sim$ *i.i.d* $N(\theta, 1)$.

   (a) Simulate (generate) a data set using $n = 100$ and $\mu = 5$. Document your work on this part by showing your lines of code and also by producing a summary of your dataset using "summary" in R.

   (b) Consider the *improper* non-informative prior for θ:

   $$f(\theta) \;=\; 1, \quad -\infty < \theta < \infty.$$

   Find the posterior density $\pi(\theta|x_1, x_2, \ldots, x_n)$ and plot it. (The flat prior on θ is conceptually a normal prior with variance $\tau_0^2 = \infty$.)

   (c) Calculate a 95% credible interval for θ. Show your calculations, a numerical interval along with a sentence interpreting the interval.

   (d) Simulate 1,000 draws from the posterior density in part (b). Plot a histogram of the simulated values and compare the histogram to the posterior density in (b).

   (e) Find a 95% credible interval for θ by sorting your 1,000 simulated values of θ from part (d) in order from smallest to largest and using the 2.5th and 97.5th percentiles as the lower and uppper end points. Report this interval. How does it compare with the *actual* credible interval in part (c)?

2. Suppose that X is a discrete random variable which takes values 1 and 2. Let $p_\theta(x)$ denote its PMF indexed by parameter θ which can take values 1, 2 or 3:

Table 1: PMF of X

	$x=1$	$x=2$
$p_1(x)$	$\frac{1}{2}$	$\frac{1}{2}$
$p_2(x)$	$\frac{1}{3}$	$\frac{2}{3}$
$p_3(x)$	$\frac{3}{4}$	$\frac{1}{4}$

We use the prior on θ as specified below:

	$\theta=1$	$\theta=2$	$\theta=3$
$\pi(\theta)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$

(a) Create a table showing the likelihood function $L(\theta|x_1, x_2)$ for every possible sample of size 2. (You may assume x_1 and x_2 are two independent realizations of X.)

(b) Create a table showing the posterior distribution $\pi(\theta|x_1, x_2)$ for every possible sample of size 2.

(c) Calculate the posterior mean $E(\theta|x_1 = 1, x_2 = 2)$.

3. A random variable U follows an inverse gamma(a, b) (IG(a,b)) distribution if $V = \frac{1}{U}$ follows a gamma(a, b) distribution:

$$f(v) \;=\; \frac{b^a}{\Gamma(a)} v^{a-1} exp(-b\,x), \qquad a > 0, b > 0, v > 0.$$

Find the density of U by using the CDF method. Please follow these steps.

- Use the relationship between U and V to show the CDF of U is given by

$$F_U(u) \;=\; \begin{cases} 0 & if\ u \le 0, \\ 1 - \int_0^{1/u} \frac{b^a}{\Gamma(a)} x^{a-1} e^{-b\,x}\, dx & u > 0 \end{cases}$$

- Then find the PDF of U by $\frac{\partial}{\partial u} F_U(u)$ keeping in mind that

$$\frac{\partial}{\partial u} \int_0^{1/u} g(x)dx \;=\; g(1/u)\, \frac{\partial}{\partial u} \left(\frac{1}{u}\right).$$

4. A manufacturer believes that a machine produces rods with lengths X in centimeters distributed $N(\theta_0, \sigma^2)$, where θ_0 is known and $\sigma^2(> 0)$ is unknown and that the prior distribution $\sigma^2 \sim IG(a, b)$ is appropriate.

(a) Show that the posterior density of σ^2 based on a sample $x_1, x_2, \ldots, x_n$ is

$$IG\left(\frac{n}{2} + a,\ \frac{n}{2}s^2 + b\right)$$

where

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \theta_0)^2.$$

(*Hint:* just focus on the terms that involve σ^2 in the likelihood function and the prior.)

(b) Determine the mode of the posterior distribution of σ^2.

5. The dataset https://www.stat.berkeley.edu/~statlabs/labs.html shows birth weights (in ounces) for babies born to mothers who smoke during pregnancy (smoke=1) and non-smokers.

(a) Extract the birthweights of babies born to smokers. Document your work by showing the code and also making a histogram of these birthweights. Mark 88 ounces – the weight below which a baby is considered a low birthweight baby – on this graph.

(b) The % of low birthweight babies in the group born to non-smoking mothers is about 3%. What fraction of babies born to smokers, p have low birthweight? State the fraction and include the lines of code.

(c) Let X denote the number of low birthweight babies among smokers. Assuming $X \sim \text{Binomial}(n, p)$, calculate the exact Binomial p-value for testing whether p is significantly higher than 3%. Write your conclusion in context. (You may use binom.test in R but be sure to explain the calculation of the p-value).

(d) Use a uniform prior on p to conduct a Bayesian test of $H_0 : p \leq 0.03$ versus $H_1 : p > 0.03$. Write your conclusion in context. (Be sure to explain your calculations.)