Math 459 Midterm 1

Due Thursday, Feb. 25

Guidelines:

- You must show your work to get credit.
- Include your **R** code and the output (just copy+paste into a text file).
- No questions are allowed (no office hours on Mon. Feb. 22 or Tues. Feb. 23, and no lecture on Tues. Feb. 23).
- You are not allowed to work with anyone else.

1. Install the `HistData` package. For this question you will use the `Michelson` dataset, which consists of 100 measurements of the velocity of light in air (as opposed to in a vacuum). The data frame is just a single numeric vector `Michelson$velocity`. These measurements were made by Albert Michelson from June 5th to July 2nd, 1879. Call the values in the dataset x_i, $i = 1, \ldots, 100$. These values can be interpreted in the following way. For each i, $x_i + 299,000$ gives the velocity in km/sec.

Suppose you model the velocity of light in air, in these units, as a random variable from a hypothetical normal population distribution with variance 50 and unknown mean θ. The random sample $\{X_1, \ldots, X_n\}$ is assumed to be i.i.d. from $f(x|\theta)$.

In these units, the 'true' speed of light in a vacuum is 734.5 (i.e. 299,734.5 km/sec). After adjusting for the effects of air, the 'true' value for the velocity of light in air would be 299, 792.5 km/sec. Thus we could think of these 100 observations as being generated from a population distribution with true mean $\theta = 792.5$.

In practice, we usually do not know the true value and wish to make inferences about the unknown true value of the velocity of light in air.

(a) Find the MLE. Use `R` to construct a 95% frequentist confidence interval for θ from the observed sample based on the large sample approximation to the sampling distribution of the sample mean $\bar{X}$.

(b) Assume a normal prior for θ with hyperparameters the mean and variance, $\mu = 800$ and $\sigma^2 = 50$. Find the posterior mean. Use `R` to construct a 95% Bayesian equal-tailed credible interval for θ given the observed sample. Do not use a package which does MCMC sampling.

(c) Plot the posterior. For the same data and parameter values, suggest a non-equal-tailed 95% credible interval which is narrower than the equal-tailed 95% credible interval.

2. Consider an exponential likelihood; inference is to be made about the unknown rate parameter λ. A random sample of size n is observed.

(a) Find a conjugate prior.

(b) Find a non-informative prior which differs from the prior in part (a).

(c) Using `R`, generate a random sample of size $n = 50$ from an $\text{Exp}(\lambda = 3)$ distribution. Compute the MLE. Also compute the posterior mean for both parts (a) and (b).

(d) Construct 95% equal-tailed credible intervals using the generated sample; do this for both parts (a) and (b). Which interval is narrower? Why?

3. Consider Bayesian point estimation of the parameter θ corresponding to the proportion in a binomial likelihood with fixed n. Using a squared-error loss, and a prior that is conjugate to the binomial likelihood, find the **Bayes risk** of the posterior *median* (not the mean) as an estimator for θ.

4. Compare and contrast the principle of falsifiability with the principle of induction *as they are used in statistical inference*. Give an example for each, and state whether the examples follow a frequentist or Bayesian approach. Type your response (it should be no more than a half-page).