Statistics 5444: Homework 1

For each homework assignment, turn in at the beginning of class on the indicated due date. Late assignments will only be accepted with special permission. Write each problem up *very* neatly (LaTeX is preferred). Show all of your work.

Problem 1

Let $x_i \sim N(\mu, \sigma^2)$, for $i = 1, \ldots N$. Assume that σ^2 is a known parameter.

In this problem, we will infer from the observed $x_i's$ the posterior sampling distribution $p(\mu|\boldsymbol{X})$, where $\boldsymbol{X} = \{x_1, \ldots, x_N\}$.

Part 1

What it is the likelihood function $L(\mu|\boldsymbol{X})$?

Part 2

Under the *reference* prior $p(\mu) \propto 1$, find the posterior distribution for μ.

Problem 2

This problem is very similar in spirit to Problem 1. Let $X \sim \text{Bin}(N, p)$, so that $p(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$.
Consider the *reference* prior $p \sim \text{Beta}(\frac{1}{2}, \frac{1}{2})$ (We will motivate this prior distribution later). Recall that the pdf for $z \sim \text{Beta}(\alpha, \beta)$ follows as

$$p(z) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where $B(\alpha, \beta) = \int_0^1 z^{\alpha-1}(1-z)^{\beta-1} dz$.

Specify the posterior distribution $p(p|x)$. That is, explicitly name the type of distribution that the posterior follows and find the parameters of the distribution.

Problem 3

This problem is again very similar to problem 1. This time we will be doing posterior inference on the ordinary least squares (simple regression) problem. Recall the basic set up of the regression problem:

$$y_i = \sum_{j=1}^{p} x_{i,j} b_j + \epsilon \qquad \text{for } i = 1, \ldots, N,$$

where $\epsilon \sim N(0, \sigma^2)$. Generally $(x_1 = 1)$, so that b_1 is interpreted as an intercept term. In matrix notation, we can write down the system of equations as

$$Y = \boldsymbol{X}\beta + \boldsymbol{\epsilon},$$

where $Y = (y_1, \ldots, y_n)^T$ is a vector of $y's$, $\beta = (b_1, \ldots, b_p)^T$ is a vector of (unknown) coefficients, and $\boldsymbol{X}$ is the corresponding $(N \times p)$ matrix of regressors. The error term follows the multivariate normal distribution:

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \Sigma),$$

where $\mathbf{0}$ is the $(N \times 1)$ zero vector, and $\Sigma = \sigma^2 I_{N\times N}$.

Recall that the density for the multivariate normal distribution (with mean vector μ and covariance matrix Σ) is written

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)},$$

where $|\cdot| = Det(\cdot)$.

Part 1

Under the reference prior $p(\beta) \propto 1$, find the posterior distribution for β.

Part 2

Recall that the MLE estimate is $\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Y$. Find $\mathrm{Var}(\hat{\beta})$ and compare the posterior estimate to MLE estimate.

Problem 4

See P.M. Lee, page 10. The top paragraph illustrates the *prosecutor's fallacy*. After reading, the Author states that the two (under the scenario illustrated in the paragraph) conditional distributions $p(E|I)$ and $P(I|E)$ are "equal if and on if the prior probability $P(I)$ of innocence is taken to be $\frac{1}{2}$". Justify his claim (i.e. show that the prior must equal 1/2 (approximately??)).

Problem 5

Consider $X \sim$ Bin(N,p), So that $p(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$. Consider the uniform prior $p(p) = 1$, where $0 \leq p \leq 1$. The posterior sampling distribution for this should be clear from Problem 2 (note that this prior is a Beta$(1, 1)$).
The prior that you have specified places an equal amount of weight on every possible value of p.

However, some people like to work with the log-odds, which we write as

$$\Lambda = \log\left(\frac{p}{1-p}\right).$$

Part 1

Find $p(\Lambda)$. That is, find the pdf for Λ (this is just a simple transformation problem).

Part 2

Now, do the problem the other way around. Consider placing a uniform prior on (Λ). That is, let $p(\Lambda) \propto 1$. What is the implied prior distribution on p?

Problem 6

Consider the binomial sampling distribution $p(x) = \binom{N}{x}p^x(1-p)^{N-x}$. In this problem, we'll compare Frequentist *Asymptotic* Intervals (coverage intervals) to Bayesian Intervals (credible intervals), and verify that a Bayesian can be a rather good "frequentist".

Consider the two different types of intervals:

1. Asymptotic Confidence Interval (Wald Type):
$$\hat{p} \pm 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}, \tag{1}$$
where $\hat{p} = \frac{x}{N}$. This of course follows from the asymptotic normality assumption of the estimator $\hat{p}$, and holds for sample sizes of about 30 or larger (t-intervals are used for smaller sample sizes).

2. Credible Interval (equi-tail):
Construct the posterior distribution ($\pi(p|x)$), and form a 95% credible interval. Any interval containing 95% of the area under the posterior density is a 95% credible interval. I suggest you use the equal tail interval since it is the easiest one to construct.

   Simulating a 95% credible equal-tail interval: sample $p_k \sim \pi(p|x)$, $k \in \{1, \ldots, 10,000\}$. That is, sample 10,000 times from your posterior. Now, order them and use the 250th biggest sample as the lower bound and the 9750th biggest sample as the upper bound.

Part a

Generate a uniform spacing of p's from .01 to .99, with a step size of .01. Under each value of p, generate a Binomial random number with $N = 30$. Repeat this 1,000 times. You'll have, for each value of p, 1,000 replicates from a Binomial distribution, with $N = 30$.

For each value of p, compute the relative coverage frequency using both credible intervals and confidence intervals and plot your results (on a single graph). The y-axis will have coverage frequency (based on your 1,000 Binomial replicates) and the x-axis will be over p. Be sure to label your graph and distinguish results from the credible and confidence intervals.

Part b

Repeat the exercise with $N = \{50, 100, 1000\}$.

Part c

If true $p = .001$, how large would N need to be so that the 95% confidence interval covered within 1% of the 'nominal rate'?

Part d

If true $p = .001$, how large would N need to be so that the 95% credible interval covered within 1% of the 'nominal rate'?

Conclude with your thoughts on the experiment. Are you surprised?