Stat 474/574 Homework 1 - Due Wed., February 8

Unless instructed otherwise, please return a printed version or email a pdf version of your assignment during the class period.

Problem 1: In the General Social Survey, $n = 129$ females aged 65 and older were asked whether they were generally happy. Let $Y_i = 1$ if the ith woman reported being generally happy and let $Y_i = 0$ otherwise. One hundred and eighteen women (91%) reported being generally happy and 11 women (9%) said that they were not generally happy.

1.1 Consider first a uniform prior distribution for θ, the probability that a woman will report being generally happy. Simulate 1000 values from the posterior distribution of θ and summarize results.

1.2 Consider now a conjugate beta prior, such that $E(\theta) \approx 0.6$. You are not very confident that 0.6 is a good guess for θ. Again simulate 1000 values from the posterior distribution of θ. Summarize your results.

1.3 Suppose that some additional information has become available and you are now quite comfortable about 0.6 as a guess for θ. Modify your prior accordingly and repeat the analysis. Comment on what you observe.

1.4 Using the posterior distribution in 1.3, compute the posterior predictive distribution of observing between 80 and 90 women who report being happy if an additional 100 women were to be interviewed.

Problem 2: (To do this problem, you can use the function `BernBeta` in the `LearnBayes` package). An election is approaching and the latest poll of 100 people shows that 58 prefer candidate A and 42 prefer candidate B. You are curious about the chances that candidate A will prevail.

2.1 Before you read the results of the poll, you had not thought much about the chances that candidate A would win (or a priori, you had no information about the probability of success for candidate A). What is the 95% Highest Posterior Density set (HPD set) on the probability that A will win after you have read the results of the poll?

2.2 Using simulation, compute a 95% credible set for the probability that A will win. How do the two intervals compare?

2.3 Just to confirm, you poll a second set of 100 randomly chosen persons from the same population. You find that in this second set, 57 prefer candidate A and the rest prefer B. How does the 95% HPD set you computed in part a change?

2.4 Based on the two surveys, is it reasonable to conclude that the population is evenly divided in terms of preferences for candidates A and B?

2.5 Based on the two surveys, what is the probability that candidate A will win by at least a 10% margin?

Problem 3: Suppose that two different persons (Joe and Sam) are interested in estimating the proportion θ of students at a college who commute to school. Joe uses the discrete prior shown in Table 1, while Sam decides instead to use a Beta$(3, 2)$ prior for the proportion of commuters θ.

θ	0.1	0.2	0.3	0.4	0.5
$p(\theta)$	0.5	0.2	0.2	0.05	0.05

Table 1: *Joe's discrete prior for* θ

3.a Use R to compute the mean and standard deviation of θ for Joe's prior and for Sam's prior. Do Joe and Sam have similar prior beliefs about the proportion of commuters?

3.b Suppose that you carry out a survey of the students in the college and find that out of 100 student surveyed, 30 of them commute. Use the function `pdisc` in `LearnBayes` to find the posterior probability under **each of the priors** that no more than 20% of the students in the college are commuters.

3.c Find a 90% credible set for θ under each of the two models.