Instructions

(1) Homework must be typed and answered in the order given (problem 1(a)(b)(c)(d) first, problem 2(a)(b)... second, etc...)

(2) Undergrads and grads will answer all questions.

(3) Include in each part of the homework only the answer. R code and R output (without mistakes), must be included in the appendix to the question. For example, for question 1.a, write only the answer and your comments. The code and output for that part of the question will be in the appendix (the last part of question 1).

(4) No late homework under any circumstances.

(5) Write your name and ID this way: Last name, first name, UCLA ID, date, Homework number.

(6) Do not just give a number as an answer. For example, if asked for probability that posterior proportion is larger than 0.7, write $Prob(p > 0.7) = 0.3$, say and write comments or explanations if needed.

(7) The homework must be turned in in lecture (no mail box, no e-mail).

Problem 1. Suppose we are interested in the estimation of the prevalence of a disease, for example, coronary heart disease, for men aged 30-40. Data from a prospective study of sample size equal to 1250 men have indicated 5 men experienced at least one incident. We are interested in estimating the prevalence rate of the disease for this specific age group. Do the following:

(a) Specify a non-informative Beta prior for the prevalence rate (U(0,1) written as a Beta -see lesson 3). Use R to find the prior mean, standard deviation, and probability that the prevalence rate is less than 2 incidents per 1000 males (hint: lessons 3 and 4).

(b) Specify a Binomial likelihood function that reflects what you observed in the data (hint: lesson 4).

(c) Specify the type of distribution that is your posterior distribution.

(d) Use R to find the posterior mean, the posterior standard deviation, the probability that the prevalence rate is less than 2 incidents per 1000 males and a 95% posterior interval.

(e) Make a plot that shows the prior and posterior distribution together in the same frame. Put a legend and distinguish the two curves. Discuss your results.

(f) Appendix: include documented R code (see web site for course, computing section files) and output here at the end of the problem. Use courier font for code and output. It must be clear what code is for which part of the problem. Write code in the same order that you answer the questions.

Problem 2. Borrowing owls sometimes build their nests in holes that were dug by prairie dogs, coyotes, or badgers and have been since abandoned. They sometimes line the nests with cattle or horse dung. Why? Perhaps it insulates the nest from temperature extremes. A possibility suggested by biologist Dennis Martin is that the owls use the dung to keep predators away. To test this theory, biologist Gregory Green observed lined and unlined owl nests in the Columbia River basin. He recorded whether the nest was raided or not in both samples. We will deal with both samples later in the quarter, but for now consider only the lined nests. Green identified $n = 25$ nests lined with dung and observed that only $s = 2$ were raided. Assume exchangeability of the 25 nests.

Green's prior probabilities of the proportion of nest lined with dung being raided are given below

Proportion of Raided nests (p)	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Prior probabilities	0.08	0.12	0.08	0.04	0.12	0.28	0.12	0.04	0.04	0.04	0.04

(a) Summarize in one sentence Green's prior probability distribution and enter it into R.

(b) Write down the likelihood function and enter it into R.

(c) Find the posterior distribution with R. Write it in a table that also contains the prior probabilities. Has Green changed his views very much after observing the data? Explain your answer.

(d) Plot the posterior and prior distributions.

(e) Appendix: include documented R code (see web site for course, computing section files) and output here at the end of the problem. Use courier font for code and output. It must be clear what code is for which part of the problem. Write code in the same order that you answer the questions.

Problem 3. In the matter of why burrowing owls line their nests with dung, suppose biologist Green's prior opinion concerning the proportion of dung-lined nests that are raided by badgers is given by the Beta(2,2) density.

(a) Specify the posterior density.

(b) Find his predictive probability that the next lined owl nest he observes will be raided. Show work.

Problem 4. Suppose that if $\theta = 1$, then y has a normal distribution with mean 1 and standard deviation σ, and if $\theta = 2$, then y has a normal distribution with mean 2 and standard deviation σ. Also, suppose $Pr(\theta = 1) = 0.5$ and $Pr(\theta = 2) = 0.5$

(a) For $\sigma = 2$, write the formula for the marginal probability density for y and plot it using R.

(b) What is $Pr(\theta = 1 \mid y = 1)$, again supposing $\sigma = 2$? Show work.

(c) Describe how the posterior density of θ changes in shape as σ is increased and as it is decreased.

Problem 5. Since the posterior distribution is a compromise between prior information and the information provided by the new data, then it is interesting to compare the relative strengths. Perform an experiment where you flip a coin 10 times, recoding the data as zeros and ones. Produce the posterior expected value (mean) for two priors on p (the probability of a head): a uniform distribution between zero and one, and a beta distribution, with parameters [10,1]. Which prior appears to influence the posterior mean more than the other? Explain and show your work.

Problem 6. Assume that 2% of the population of the United States are members of some extremist Militia group (Prob(M)=0.02), a fact that some members might attempt to hide and therefore not readily admit to an interviewer. A survey is 95% accurate on positive classification, $P(C \mid M) = 0.95$, and the unconditional probability of classification (i.e, regardless of actual militia status) is given by $P(C) = 0.05$. The survey is 97% accurate on negative classification ($P(C^c \mid M^c) = 0.97$).

(a) Use the law of total probability (lesson 2), and information given in the problem, to show that P(C) is the normalizing constant obtained by accumulating over all possible events. (Hint: $P(C \mid M^c) = 1 - P(C^c \mid M^c)$.

(b) Find the probability that someone positively classified by the survey as being a militia member really is a militia member