

Name: _____

Bayesian Statistics, 22S:138
PRACTICE PROBLEMS for MIDTERM 1, 2005
Midterm 1, Fall 2004

Show any computations that you carry out. Use the back of your exam paper if you run out of space.

1. Your friend asks you to help her assess her own subjective probability that a woman will be elected president of the United States in the next 20 years. Let A denote the event that a woman will be elected president in the next 20 years, and let $P_S(A)$ denote your friend's subjective probability of this event.

To devise a calibration experiment, you ask your friend to imagine that she has given you a coin that she believes to be fair (that is, she believes that the probability of obtaining a head on a flip is 0.5 for the coin).

For the first step of the calibration experiment, you offer your friend two hypothetical ways of trying to win a 2-week vacation in Hawaii:

Game 1: Win if a woman is elected president in the next 20 years.

Game 2: Win if you flip the coin and it comes up a head.

Your friend chooses Game 2.

- (a) From this, you conclude that $P_S(A)$ lies in what interval?
- (b) For the next step of the calibration experiment you will again offer your friend two choices of ways to try to win the vacation. Game 1 will remain as above. Which of the following is the best way to define Game 2? (Circle one.)
- i. Win if you flip the coin two times and get one head and one tail in either order.
 - ii. Win if you flip the coin twice and get heads both times.
 - iii. Win if you flip the coin three times and get at least two heads.
 - iv. Win if you flip the coin four times and get two heads and two tails in any order.
- (c) If your friend chooses Game 2 (as you defined it in the previous question) at the second step, you conclude that $P_S(A)$ lies in what interval?

2. You wish to infer about the proportion p of doctors affiliated with the University of Iowa Hospitals and Clinics who are vegetarians. You decide to carry out your study in the following way. You obtain a computerized list of all the doctors

affiliated with UIHC. You use a computer to draw one name at random. You interview the first doctor and ask whether he/she is a vegetarian. If that doctor is NOT a vegetarian, you draw another doctor to interview. You keep on drawing names and interviewing doctors until you get the first doctor who says he/she IS a vegetarian.

- (a) You choose to use a Beta distribution as a prior for the population proportion p . Suppose you believe that p is probably around .1 and your belief is as strong as if you previously had done a study of 20 doctors and had found 2 of them to be vegetarians. Give one appropriate choice of parameters for the Beta prior.
- (b) The geometric probability mass function can be used when the experiment is a sequence of independent Bernoulli trials, all with the same success probability, and the random variable of interest X is how many trials it takes to get the first success. It is a special case of the Negative Binomial distribution and has the following PMF:

$$\theta \sim \text{Geom}(p)$$
$$p(\theta|p) = p(1-p)^{\theta-1}, \quad \theta = 1, 2, \dots, \quad 0 \leq p \leq 1$$

Suppose you interview 6 doctors who say they are not vegetarians, and the 7th doctor says he is a vegetarian, so $x = 7$. Write the likelihood for p based on your data.

- (c) Write an expression to which the posterior distribution $p(p|x)$ is proportional.
- (d) Can you recognize your result in the previous problem as the kernel of a parametric density? If so, identify the parametric family and give the numeric values of its parameters.

NOTE: If you could not get an answer to the previous question, then answer

the next three questions supposing that the posterior distribution for p is Beta(4, 5).

(e) What are the posterior mean and variance?

(f) Is the Beta the conjugate prior for the geometric likelihood? Briefly justify your answer.

(g) Write the R function or functions, including arguments, that you would use to compute a 90% posterior credible set for p .

(h) Suppose that the 90% posterior credible set for p turned out to be (0.05, 0.5). What is the correct Bayesian interpretation of this interval? (Circle one.)

- i. p has different values at different times. 90% of the time, p is between 0.05 and 0.5, and 10% of the time it is not.
- ii. 90% of doctors have a vegetarianism score in this interval.
- iii. 90% of credible sets constructed using this method will contain the true population parameter p .
- iv. For a person who agreed with the prior distribution on p , given this data the probability is .9 that the true population proportion p is between 0.05 and 0.5.
- v. none of the above

3. The following paragraphs are taken from the webpage of Zogby International, a polling organization. The URL for this article is

<http://www.zogby.com/Soundbites/ReadClips.dbm?ID=9701>

Women also are more security-conscious in the post-9/11 world. 54% of women say people need to be willing to give up some rights to increase their overall security. But as with the presidential race, the marriage gap is fully on display here. 57% of married women say people need to be willing to give up their rights, while 53% of single women are opposed to this premise.

Zogby International conducted interviews of 1001 women likely voters chosen at random nationwide. All calls were made from Zogby International headquarters in Utica, N.Y., from 9/20/04 thru 9/23/04. The margin of error is +/- 3.1 percentage points.

(a) What proportion of the women in the survey were married? (Assume that there were only two possible answers to the security question.)

(b) Suppose you draw a woman at random from the Zogby poll respondents. If she said that people need to be willing to give up rights to increase security, what is the probability that she is single?