Name: ______________________________

Bayesian Statistics, 22S:138
Midterm 1, Fall 2010

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

1. The following statements are based on the American Pet Products Manufacturers Association 2009-2010 National Pet Owners Survey.

   - The probability that a randomly-selected U.S. household owns at least one dog is 0.39.
   - Given that a household owns at least one dog,
     - the probability that it owns only 1 dog is 0.67
     - the probability that it owns exactly 2 dogs is 0.24

   (a) Given that a household owns at least one dog, what is the probability that it owns 3 or more dogs? Numeric answer; show your work.

   (b) What is the probability that a household randomly selected from among all U.S. household owns 3 or more dogs? (Numeric answer; show your work.)

2. You wish to infer about the proportion p of Iowa high school football coaches who practice yoga. You decide to carry out your study in the following way. You obtain a computerized list of all Iowa high school football coaches. You use a computer to draw one name at random. You interview the first coach and ask whether he practices yoga. If that coach does NOT practice yoga, you draw another coach at random to interview. You keep on drawing names and interviewing coaches until you get the first who says he does practice yoga. You stop sampling after the first "yes."

   (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 coaches and had found 2 of them to practice yoga . 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 failures occur before the first success. The geometric pmf is

$$f(x) = \begin{cases} p(1-p)^x, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

   Suppose you interview 6 coaches who say they do not practice yoga, and the 7th coach says he does, so $x = 6$. 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 credible sets constructed using this method will contain the true population parameter p.
iii. 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.
iv. none of the above