

Name: _____

Bayesian Statistics, 22S:138

Midterm 1, Fall 2006

PRACTICE PROBLEMS for Midterm 1, Fall 2007

Show any computations that you carry out. Use the back of your exam paper if you run out of space. Point values for each question are shown in parentheses.

1. The following facts described the students who took 22S:138 in a recent year:

- 35% of the students were statistics grad students
- 25% of the students were biostatistics grad students
- 40% of the students were in neither of the above categories
- 60% of the statistics students were women
- 75% of the biostatistics students were women
- 40% of the students in other categories were women

(a) You drew a student at random from the class list for 22S:138 in that year. What is the probability that the student you drew is a woman?

(b) Suppose that the student you drew was a woman. What is the probability that she was **not** a statistics grad student and **not** a biostatistics grad student, given that she was a woman?

2. A social worker did a Bayesian analysis to estimate the proportion p of patients discharged from a community hospital who were satisfied with the care they received while in the hospital.

(a) Before collecting any new data on the subject, she studied information on complaints filed with the hospital and talked with nurses, chaplains, and others who interacted with patients in order to get a rough idea about p . Based on the preliminary assessment, she expressed her knowledge about p in the form of a Beta prior:

$$p \sim \text{Beta}(4,6)$$

i. What is the mean of the prior distribution? (numeric answer)

(b) The social worker then chose a simple random sample of 30 patients who had been discharged from the hospital during the last 12 months. She interviewed each one, and asked whether they were satisfied with the care they received. 18 patients said “yes” and 12 said “no.”

i. What parametric family is appropriate for the likelihood for these data? (just the name)

(c) The social worker combined her prior with the likelihood and obtained the following posterior distribution:

$$p(p|y) = \text{Beta}(22,18)$$

She used R to calculate a 95% equal-tail credible set for p , with the following results:

```
> qbeta(c(0.025, 0.975), 22,18)
[1] 0.3962080 0.6990522
```

i. Circle **all** the statements below that are true regarding this Bayesian interval.

- A. 95% of hospital patients have a satisfaction score between 0.396 and 0.699
- B. For a person who agreed with the social worker’s prior, the probability is .95 that the population proportion p lies in the interval (0.396,0.699).
- C. 95% of simple random samples will produce a sample proportion in the interval (0.,396,0.699)
- D. If we took many, many simple random samples of size 30 from the population of patients discharged from this hospital, 95% of the samples would produce a Bayesian interval that contained the true p

- E. none of the above
- ii. The social worker decides that she would like to estimate p more precisely — that is, she wants a narrower credible set. So she plans to draw a new simple random sample of 10 patients. She wants to calculate an 80% prediction interval for the number of “yeses” that she will get in the new sample. She uses the `pbetap` function, with the following output:

```
> pbetap( 22,18,10,0:10)
      s      pred
[1,]  0 0.001026584
[2,]  1 0.008364758
[3,]  2 0.033298173
[4,]  3 0.085243322
[5,]  4 0.155391473
[6,]  5 0.210791911
[7,]  6 0.215582637
[8,]  7 0.164253437
[9,]  8 0.089312807
[10,] 9 0.031337827
[11,] 10 0.005397070
```

- A. What is the probability that she will get 3 “yeses” in her new sample of 10 people? (numeric answer)
- B. What are the endpoints of the 80% prediction interval? (numeric answer; show your work)

3. The data value y is a random draw from a probability distribution depending on a parameter R :

$$p(y|R) \propto R^{1/2} \exp\left(-\frac{R(y-6)^2}{2}\right) \quad -\infty < y < \infty \\ R > 0$$

- (a) (3) What parametric family is this? (Name and parameter(s)).
- (b) (3) To what parametric family of distributions would the conjugate prior for R belong? (Just name the family.)
4. Researchers choose a simple random sample of counties from amongst all the counties in the U.S. The variable reported for each of the selected counties is the area of the county in square miles.
- (a) Which of the following parametric families would be most likely to be appropriate for the likelihood in their analysis?
- beta
 - binomial
 - gamma
 - normal
 - poisson
- (b) Justify your answer in a sentence or two.

5. Two fair dice, one red and one green, are rolled. Let the event A be “the sum of the faces showing is an even number.” Let event B be “the sum of the faces showing is divisible by 3.”

(a) List the items in A , and find $P(A)$.

(b) List the outcomes in B , and find $P(B)$.

(c) List the outcomes in $A \cap B$ and find $P(A \cap B)$.

(d) Are the events A and B independent? Briefly explain why or why not.