Name: Solutions

Bayesian Statistics, 22S:138
Midterm 1, Fall 2007

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

1. Your friend is planning to drive from Iowa City to San Francisco. He is trying to assess $P_s(A)$ – his subjective probably of event A, where A is the event that his 1973 VW Beetle (a kind of car) will make it to San Francisco without breaking down.

   You are using a calibration experiment to help him assess $P_s(A)$. For the first step, you offer him a choice two ways of winning a free tank of gas: (a) he gets the free tank of gas if event A occurs; (b) you will flip a fair coin and he gets the free tank of gas if the coin comes up heads.

   (a) Your friend chooses method (b). You now know that $P_s(A)$ lies in what interval? (Give numeric endpoints.)

   1 $(0, 0.5)$

   (b) Briefly describe the two choices you would give your friend for the second step of the calibration experiment. Design the choices to make the calibration experiment efficient.

   (a) win tank of gas if event A occurs

   2 (b) win tank of gas if 2 fair coins are flipped and both come up heads

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

$$p(y|R) \propto Rexp(-R\,y) \quad 0 < y < \infty \qquad (1)$$
$$R > 0$$

   (a) What parametric family is this? (Name and parameter(s)).

   1 exponential (R)

   (b) If we consider (1) as the density for random variable Y, which term(s) in (1) are

      i. the normalizing constant

      1 R

      ii. the kernel

      1 e^{-Ry}

6

(c) If we consider (1) as the likelihood for R, with y known, which term or terms are:

i. a constant

none

1

ii. the kernel

1

Re^{-Ry}

(d) To what parametric family of distributions would the conjugate prior for R belong? (Just name the family.)

1

gamma

3. 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 number of new cases of retinoblastoma (a form of cancer) reported in 2006.

(a) Which of the following parametric families would be most likely to be appropriate for the likelihood in their analysis?

i. beta
ii. binomial
iii. gamma
iv. inverse gamma
v. normal
vi. poisson
vii. t

1

(b) Justify your answer in a sentence or two.

Poisson is appropriate for counts of rare events occurring per unit time, space, etc. This is counts of a rare disease per county per year.

2

4. An instructor is giving a multiple choice test to a student.

- Before seeing the student's answer to the first question, the instructor believes that the probability that the student knows the right answer is 1/3.
- If the student knows the right answer, the probability that she will answer the question correctly is 1.
- If the student doesn't know the right answer and just guesses, the probability that she will answer the question correctly is 1/4.

(a) Before the instructor sees the student's answer, what is the instructor's probability that the student will answer the first question correctly? (Numeric answer; show your work.)

3

$$P(C) = P(C \cap K) + P(C \cap \bar{K})$$
$$= P(C|K)P(K) + P(C \cap \bar{K})\, P(\bar{K})$$
$$= 1 \cdot \tfrac{1}{3} + \tfrac{1}{4} \cdot \tfrac{2}{3}$$
$$= \tfrac{1}{2}$$

(b) If the student does answer the first question correctly, what is the instructor's probability that she just guessed (that is, that she did *not* know the right answer? (Numeric answer; show your work.)

$$P(\bar{K}|C) = \frac{P(C \cap \bar{K})}{P(C)} = \frac{\tfrac{1}{4} \cdot \tfrac{2}{3}}{\tfrac{1}{2}} = \frac{1}{3}$$

C event student answers correctly

K event student knows the answer

$\tfrac{1}{3} + \tfrac{2}{12} = \tfrac{2}{6} + \tfrac{1}{6} = \tfrac{3}{6} = \tfrac{1}{2}$

6

5. A statistician believes that the distribution of heights of undergraduate men at UI is normal with known population mean $\mu = 69$ inches. He wishes to estimate the population *precision* (inverse of the variance) of this distribution. He will use the symbol τ^2 for the precision.

The sampling density for a single observation y drawn from this normal distribution is

$$p(y) = \frac{\tau}{\sqrt{2\pi}} exp\left(-\frac{\tau^2 \, (y-69)^2}{2}\right)$$

For data, the statistician will measure two randomly selected UI undergrad men.

(a) Write the likelihood of τ^2 if there are two data values, represented symbolically as y_1 and y_2.

2

$\prod_{i=1}^{2} \frac{\tau}{\sqrt{2\pi}} \exp\left[-\frac{\tau^2 (y_i - 69)^2}{2}\right] \propto \tau^2 \exp\left[-\frac{\tau^2 \sum_{i=1}^{2} (y_i - 69)^2}{2}\right]$

(b) The statistician puts the following prior on τ^2:

$$p(\tau^2) = Gamma\left(\frac{1}{2}, \frac{9}{2}\right)$$

This prior has as much information as how many prior observations? (In other words, what is the equivalent prior sample size?)

2

1 (First argument is $\frac{n_0}{2}$)

prior $\propto (\tau^2)^{\frac{1}{2}-1} e^{-\frac{9}{2}\tau^2}$

(c) When the statistician measures his two subjects, their heights are:

$$y_1 = 68 \; in$$
$$y_2 = 72 \; in$$

Write a mathematical expression to which the posterior distribution, $p(\tau^2|y_1, y_2)$, is proportional. First write it with the data values represented symbolically as y_1 and y_2. Then plug in the numeric values.

$p(\tau^2 | y_1, y_2) \propto \underbrace{(\tau^2)^{-\frac{1}{2}} e^{-\frac{9}{2}\tau^2}}_{prior} \; \tau^2 e^{-\tau^2 \frac{\sum_{i=1}^{2} (y_i - 69)^2}{2}}$ ~~likelihood~~

$= (\tau^2)^{\frac{1}{2}} e^{-\tau^2\left(\frac{9}{2} + \frac{\sum (y_i - 69)^2}{2}\right)}$

$= (\tau^2)^{\frac{1}{2}} e^{-\tau^2\left(\frac{19}{2}\right)}$

(d) To which parametric family does this posterior distribution belong?

\

gamma

(e) What are the parameters of the posterior distribution (numeric values)?

2

$\mathcal{IG}\left(\frac{3}{2}, \frac{19}{2}\right)$

(f) What are the mean and variance of this posterior distribution?

1

$E(\tau^2 | y) = \dfrac{\frac{3}{2}}{\frac{19}{2}} = \dfrac{3}{19}$

$Var(\tau^2 | y) = \dfrac{\frac{3}{2}}{\left(\frac{19}{2}\right)^2} = \dfrac{6}{(19)^2}$

(g) Write the R (or Splus) function that you would use to compute a 95% posterior credible set for τ^2. Include the appropriate arguments.

1

qgamma(c(0.025, 0.975), 1.5, 9.5)

(h) Suppose that the 95% posterior credible set for τ^2 turned out to be (0.1, 0.5). What is the correct Bayesian interpretation of this interval?

2

For someone who agreed with the prior distribution that was used, there is 95% probability that τ^2 is in (0.1, 0.5).

6