Name: Solutions

Bayesian Statistics, 22S:138
Midterm 1, Fall 2009

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

1. You are in a chess tournament and will play your next game against either Joe or Mary, depending on the results of other games. Here are your subjective probabilities of getting to play each of the players, and of your chances of winning against each one:

   Your probability of playing Joe is 0.25, and your probability of playing Mary is $1 - 0.25 = 0.75$. $P(J) = 0.25$

   Your probability of winning against Joe is 0.7. $P(W|J) = 0.7$

   Your probability of winning against Mary is 0.2. $P(W|M)$

   (a) What is the probability that you will win your next game?

   $P(W) = P(W \cap J) + P(W \cap M)$
   $= P(J)P(W|J) + P(M)P(W|M)$
   $0.25\ (0.7) + 0.75\ (0.2) = 0.325$

   (b) Now suppose you tell your friend that you did win your next game, and your friend wishes to calculate the probability that Joe was your opponent. Do the calculation.

   $P(J|W) = \frac{P(W \cap J)}{P(W)} =$

   $= \frac{0.25(0.7)}{0.325} = \frac{0.175}{0.321} = .538$

2. The job of a court reporter is to transcribe the words spoken during court proceedings so as to produce the official written records of the proceedings.

   A court reporter in Iowa City wishes to use Bayesian methods to estimate the average number of errors she makes per hour of transcribing work. She will call this unknown parameter θ. As data for her self-study, she will randomly select three different days on which she has worked, and will randomly select one hour on each of those days. For each of these three hours, she will compare the tape recording of the hour with her transcript, and will count the number of errors she made. Thus her data will consist of three counts, which she will call y_1, y_2, and y_3.

   (a) Which of the following distributions is most appropriate for the the distribution of the data?
   i. beta
   ii. binomial
   iii. gamma
   iv. inverse gamma
   v. normal
   vi. poisson

   (b) Write the mathematical form of the joint distribution of the three data values.

   $p(y|\theta) = \frac{e^{-3\theta}\theta^{\sum y_i}}{\prod y_i!}$

   (c) The court reporter doesn't want to let her subjective opinions about her own abilities affect her analysis. Therefore, she uses the following noninformative prior:

   $$p(\theta) \propto \frac{1}{\theta^{1/2}},\ 0 < \theta < \infty$$

   Is this prior proper or improper? Briefly explain.

   Improper. It is like the limit of a Gamma density as the 2nd parameter goes to 0. Proper gamma density has both parameters strictly positive.

   (d) Now the court reporter gathers her data. The results are $y_1 = 2$, $y_2 = 7$, $y_3 = 4$.
   Write an expression that is proportional to the posterior density of the unknown parameter, $p(\theta|y)$.

   $p(\theta|y) \propto \frac{1}{\theta^{1/2}}\theta^{13}e^{-3\theta}$

   $= \theta^{12\frac{1}{2}}e^{-3\theta}$

(e) Do you recognize your answer to the previous question as the kernel of a particular parametric density? (yes/no) If so, name the density and its parameters.

$N(13\frac{1}{2}, 3)$

(f) What are the posterior mean and variance of θ? (numeric answers)

$E(\theta|y) = \frac{13\frac{1}{2}}{3} \cdot 4^{\frac{1}{2}}$ $Var(\theta|y) = \frac{13\frac{1}{2}}{3^2} = 1\frac{1}{2}$

(g) Write the line of R code that the court reporter should use to compute the endpoints of a 90% equal-tail credible set for θ.

qgamma(c(0.05, 0.95), 13.5, 3)

(h) Suppose that the credible set produced in the previous question is (2.69, 6.69). The Bayesian interpretation is (circle one):

i. There is 90% probability that the number of errors that the court reporter makes in a randomly sampled hour of work is between 2.69 and 6.69.

ii. There is 90% probability that the court reporter's true average number of errors per hour is between 2.69 and 6.69.

iii. We have used a procedure that, for 90% of samples, will produce an interval that contains the true parameter value.

3. Y is a random variable with density proportional to the following expression:

$$p(y) \propto \frac{1}{y^7} exp\left(-\frac{2}{y}\right), \quad 0 < y < \infty$$

What density is this? (Name the parametric family, and give the values of the parameter or parameters.)

$IG(6, 2)$ 3

4. You are helping your sister assess her subjective probability that she will get an A in the sociology class she is taking. Let's call this probability $P_S(A)$. You offer her a hypothetical choice of two ways of winning \$100: (1) You will flip a fair coin and she gets the \$100 if it comes up heads, or (2) She gets the \$100 if she gets an A in the course.

(a) Your friend chooses the coin flip. This tells you that $P_S(A)$ is in what interval? (Numerica answer)

$[0, .5)$?

(b) Describe how you will set up the next step in assessing your sister's subjective probability, to make the process as efficient as possible.

(1) I will flip a fair coin twice; she gets \$100 if both results are heads
(2) She gets \$100 if she gets an A in the course. 2

(c) A procedure such as the one being used here to assess someone's subjective probability of an event is called (circle one):

i. prior probability

ii. exchangeability

iii. likelihood

iv. calibration experiment

v. none of the above 1