Computational Statistics
Assignment 3

by James D. Wilson (University of San Francisco)

Directions: For all questions in this assignment, fully answer any question that is asked. Late assignments will automatically have 10 points deducted for each day that they are late.

Quantitative Questions

1. A random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\overline{y} = 150$ pounds. Assume the weights in the population are normally distributed with unknown mean θ and known standard deviation 20 pounds. Suppose that the prior distribution for θ is normal with mean 180 and standard deviation 40.

   (a) Verify that weight distribution ($N(\theta, 20)$) is an exponential family.

   (b) What is a sufficient statistic for θ?

   (c) Calculate the posterior distribution for θ

   (d) A new student is sampled at random from the same population and has a weight of $\tilde{y}$ pounds. Give a posterior predictive distribution for $\tilde{y}$

   (e) For n = 10, give a 95% posterior interval for θ and a 95% posterior predictive interval for $\tilde{y}$

2. Suppose that there are N cables cars in San Francisco, numbered sequentially from 1 to N. You see a cable car at random and it is numbered 203. You wish to estimate N.

   (a) Assume your prior distribution on N is geometric with mean 100; that is,

   $$\pi(N) = (1/100)(99/100)^{N-1},\ N = 1, 2, \ldots$$

   What is your posterior distribution for N?

   (b) What are the posterior mean and standard deviation of N? (sum the infinite series analytically or approximate them using appropriate simulations)

   (c) Choose a reasonable "noninformative" prior distribution for N and give the resulting posterior distribution, mean, and standard deviation for N. Check that the posterior distribution is proper!

3. Show that if $y \mid \theta \sim Exp(1/\theta)$ (i.e., the rate is θ), then the gamma prior distribution is conjugate for inference about θ given an independent and identically distributed sample of y values.

Computational Questions

1. Table 1 below provides the number of fatal accidents and deaths on scheduled airline flights per year over a ten-year period.

   (a) Assume that the number of fatal accidents in each year are independent with a Poisson(θ) distribution. Construct a prior distribution for θ and determine a posterior distribution based on the data from 1976 through 1985. Under this model, use simulations to give a 95% predictive interval for the number of fatal accidents in 1986.

Year	Fatal accidents	Passenger deaths	Death rate
1976	24	734	0.19
1977	25	516	0.12
1978	31	754	0.15
1979	31	877	0.16
1980	22	814	0.14
1981	21	362	0.06
1982	26	764	0.13
1983	20	809	0.13
1984	16	223	0.03
1985	22	1066	0.15

(b) Assume that the number of fatal accidents in each year follow independent Poisson distributions with a constant rate and an exposure in each year proportional to the number of passenger miles flown. That is, for each year let x_i = number of passenger miles flown in year i and θ = expected accident rate per passenger mile. Then, assume that

$$y_i \mid (x_i, \theta) \sim Poisson(x_i \theta)$$

Set a prior distribution for θ and determine the posterior distribution based on the data from 1976 - 1985. Using simulation, provide a 95% predictive interval for the number of fatal accidents in 1986 given that 8 $\times 10^{11}$ passenger miles are flown that year.

(c) Repeat (a) above, replacing 'fatal accidents' with 'passenger deaths.'

(d) Repeat (b) above replacing 'fatal accidents' with 'passenger deaths.'

(e) In which of the cases (a) - (d) above does the Poisson model seem the most reasonable? Discuss why.