AUSTRALIAN NATIONAL UNIVERSITY

RESEARCH SCHOOL OF FINANCE ACTUARIAL STUDIES, AND APPLIED STATISTICS

INTRODUCTION TO BAYESIAN DATA ANALYSIS (STAT3016/4116/7016)

SEMESTER 2 2017

ASSIGNMENT 1

DUE DATE: Thursday 17 August 2017, by 3pm

(12.5% of total course grade)

INSTRUCTIONS:

1. All students must hand in an assignment of their own writing.
2. The assignment should be handed in to the assignment box for STAT3016/4116/7016 available on level 4 of the ANUCBE Building 26C. There will be no online submission facility.
3. Ensure you also complete and attach a cover sheet to your assignment (available on the course website)
4. Begin each question on a new page.
5. Where required, provide sufficient computer output to support your answers. Provide enough intermediate numerical calculations to justify working for your final answer.
6. Computer output must be interpreted in written format. A solution solely highlighting the computer output is not acceptable.
7. No late assignments will be accepted.

COLLABORATION POLICY (as stated in the course outline)

University policies on plagiarism will be **strictly** enforced. You are encouraged to (orally) discuss your assignments with your classmates, but each student must write up solutions separately. Be sure that you have worked through each problem yourself and that all answers you submit are the results of your own efforts. This includes all computer code and output.

Problem 1

Construct a Monte Carlo study (that is, use computer simulation) that investigates how the probability of coverage depends on the sample size and true proportion value. In the study, let n be 10, 25, and 100 and let p be 0.05, 0.25, and 0.50. Write an R function that has three inputs, n, p, and the number of Monte Carlo simulations m, and will output the estimate of the exact coverage probability. Implement your function using each combination of n and p and m = 1000 simulations. Describe how the actual probability of coverage of the traditional interval depends on the sample size and true proportion value.

Problem 2

A hypothetical study is performed to estimate the effect of a simple training program on basketball free-throw shooting. A random sample of 100 college students is recruited into the study. Each student first shoots 100 free-throws to establish some baseline success probability. Each student then takes 50 practice shots each day for a month. At the end of that time, he or she takes 100 shots for a final measurement. Let θ be the average improvement in success probability. Give three prior distributions for θ, explaining each in a sentence:

(a) A noninformative prior,

(b) A subjective prior based on your best knowledge, and

(c) A weakly informative prior.

Problem 3

Suppose that there is a Beta(4,4) prior distribution on the probability θ that a coin will yield a "head" when spun in a specified manner. The coin is independently spun 10 times, and "heads" appears fewer than 3 times. You are not told how many heads were seen, only that the number is less than 3 times. Calculate your exact posterior density, posterior mean and posterior variance of θ. Provide a sketch of your posterior density and obtain a 95% posterior interval for θ. Discuss how your results vary if a Beta (20,20) prior is assumed.

Problem 4

Suppose there are n taxis in Canberra numbered sequentially from 1 to N. You see a taxi at random; it is numbered 159. You wish to estimate N.

(a) Assume your prior distribution on N is geometric with mean 200; that is,

$$p(N) = (1/200)(199/200)^{N-1} \text{ for } N = 1, 2, ...$$

What is your posterior distribution for N?

(b) What are the posterior mean and standard deviation of N? (Sum the infinite series analytically or find an approximation using a computer).

(c) What would be a reasonable 'noninformative' prior distribution for N?

Problem 5

Suppose you own a trucking company with a large fleet of trucks. Breakdowns occur randomly in time and the number of breakdowns during an interval of t days is assumed to be Poisson distributed with mean $t\lambda$. The parameter λ is the daily breakdown rate. The possible values for λ are 0.5, 1, 1.5, 2, 2.5, and 3 with respective probabilities 0.1, 0.2, 0.3, 0.2, 0.15, and 0.05. If one observes y breakdowns, then the posterior probability of λ is proportional to

$$g(\lambda) \exp(-t\lambda)(t\lambda)^y$$

where $g(\lambda)$ is the prior probability.

(a) If 12 trucks break down in a six-day period, find the posterior probabilities for the different values of λ.

(b) Find the probability that there are no breakdowns during the next week.

Problem 6 [STAT4116/STAT7016 only]

Consider two coins C_1 and C_2, with the following characteristics: Pr(heads $|C_1$)=0.25 and Pr(heads $|C_2$)=0.75. Choose one of the coins at random and imagine spinning it repeatedly. Given that the first two spins from the chosen coin are tails, what is the expectation of the number of additional spins until a head shows up?