STAT4CI3/6CI3 Computational Methods for Inference

Assignment 4 — Due at 1:30am on Monday, April 8, 2019

Instructions:

1. Ensure that all R code is properly commented and attach a print out with your written solution. Also mail your R code as a **single plain text file** to `cantya@mcmaster.ca` using the subject
   `S4CI3 Assignment 3: <Name> <Student ID>`
2. Start each question on a new page and submit questions in the same order as given below.
3. You are expected to show all details of your solution and any results taken from my notes or the textbook must be clearly and properly referenced.
4. No extensions to the due date and time will be given except in extreme circumstances and late assignments will not be accepted.
5. Students are reminded that submitted assignments must be their own work. Submission of someone else's solution (including solutions from the internet or other sources) under your name is academic misconduct and will be dealt with as such. Penalties for academic misconduct can include a 0 for the assignment, an F for the course with an annotation on your transcript and/or dismissal from your program of study.

Q. 1 A classic example of a hierarchical structure was used to model the number of failures of pumps in a nuclear reactor. Data on 10 such pumps from a paper by Gaver and O'Muirchartaigh in 1987 was

Pump i	1	2	3	4	5	6	7	8	9	10
Time t_i	94.32	15.72	62.88	125.76	5.24	31.44	1.05	1.05	2.10	10.48
# Failures X_i	5	1	5	14	3	19	1	1	4	22

where t_i is the observation time (in thousands of hours) for each pump.

If we let λ_i be the failure rate for the i^{th} pump then we can use the model

$$X_i \sim \text{Poisson}(\lambda_i t_i)$$

A reasonable prior for λ_i is a gamma(α, β) and we may assume that the $\lambda_1, \ldots, \lambda_n$ are independent. In this analysis we will take α fixed but will put an improper prior on β

$$\pi(\beta) \propto \frac{1}{\beta} \qquad 0 < \beta < \infty$$

a) Show that the joint posterior distribution of the eleven parameters is given by

$$\pi(\lambda_1, \ldots, \lambda_{10}, \beta) \propto \prod_{i=1}^{n} \left\{ \lambda_i^{x_i+\alpha-1} \mathrm{e}^{-(t_i+1/\beta)\lambda_i} \right\} \beta^{-(n\alpha+1)}$$

b) Show that conditional on β, the $\lambda_1, \ldots, \lambda_{10}$ are independent and that the conditional posterior distribution of λ_i given β is a gamma distribution giving the parameters of the distribution.

c) Show that the conditional posterior of β given $\lambda_1, \ldots, \lambda_{10}$ is an inverse gamma distribution and give the parameters of the distribution.

d) Implement a Gibbs Sampler to simulate from the joint posterior distribution with $\alpha = 1.8$. Run your Gibbs Sampler for $15,000$ iterations.

e) Based on the final $10,000$ iterations of your sampler give 95% equitailed Bayesian Credible intervals for the failure rates and for β.

Q. 2 Suppose that the we have the following data which comes from a Beta$(\theta, 1)$ distribution.

0.580	0.730	0.986	0.691	0.532	0.819	0.382	0.706	0.310	0.791
0.462	0.653	0.589	0.664	0.750	0.786	0.590	0.892	0.712	0.713

a) Estimate the bias and standard error of the methods of moment estimator

$$\hat{\theta} = \frac{\overline{X}}{1 - \overline{X}}$$

in the following ways

(i) The parametric bootstrap assuming the data come from the above distribution

(ii) The nonparametric bootstrap,

b) Give confidence intervals based on a normal approximation to the distribution of $\hat{\theta} - \theta$ for each of the methods in part (b).

c) For the parametric and nonparametric bootstrap also calculate the basic bootstrap and the bootstrap percentile confidence intervals for θ.

Q. 3 **a)** Show that the number of distinct bootstrap samples that can be drawn using the nonparametric bootstrap on a random sample of size n is given by

$$m_n = \binom{2n-1}{n-1}$$

Hint: Relate the problem to that of arranging n balls and $n-1$ sticks in a sequence.

b) Suppose that $n = 3$ and the observed data points are $(x_1 = 2, x_2 = 3, x_3 = 7)$. Give the m_3 possible nonparametric bootstrap samples and the probability associated with each. Hence give the distribution of the bootstrap mean $\overline{X}^*$ and use it to calculate the bootstrap bias and variance estimates for this example.

c) Show that, for an *iid* sample of size n, the bootstrap bias and variance from a nonparametric bootstrap are

$$\begin{aligned} b_{boot}(\overline{X}) &= \mathrm{E}^*[\overline{X}^* - \overline{x} \mid X_1^*, \ldots, X_n^* \overset{iid}{\sim} \hat{F}] &= 0 \\ v_{boot}(\overline{X}) &= \mathrm{Var}^*[\overline{X}^* \mid X_1^*, \ldots, X_n^* \overset{iid}{\sim} \hat{F}] &= \frac{\sum (x_i - \overline{x})^2}{n^2} \end{aligned}$$

Q. 4 Consider the `cd4` data in library `boot` which gives counts of CD4 for a number of HIV+ patients taking an experimental drug. Suppose that we are interested in the correlation coefficient, ρ, between the baseline counts and the counts after one year on the drug.

a) Assume that the two measures for each individual come from a bivariate normal distribution with unknown means, variances and correlation. Implement a parametric bootstrap to estimate the bias and standard error of the sample correlation coefficient, r and construct a bootstrap normal confidence interval for ρ.
The function `mvrnorm` *in the* `MASS` *library may be used to generate bivariate normal observations.*

b) The Fisher's transformation

$$\psi = \frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right)$$

is often used for inference about the correlation coefficient. Construct a bootstrap normal confidence interval for ψ and transform the interval to one for ρ. Compare the interval with the one you found in part (a).

c) Repeat parts (a) and (b) but now use a nonparametric bootstrap. How do the results from the nonparametric bootstrap compare to those from the parametric bootstrap?

Q. 5 STATS 6CI3 ONLY

Suppose that $X_1, \ldots, X_n$ is a random sample from a Normal(μ, σ^2) distribution where both parameters are considered unknown. Suppose we take an improper prior distribution for μ and σ^2 with joint density

$$\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \qquad -\infty < \mu < \infty, \; 0 < \sigma^2 < \infty$$

a) Show that the conditional posterior distribution of μ given the data and σ^2 is normal and find the mean and variance of this conditional posterior distribution.

b) Show that the marginal posterior distribution of μ is such that the posterior distribution of the random variable

$$\frac{\sqrt{n}(\mu - \overline{x})}{s}$$

(where s is the observed sample standard deviation) is a Student's t distribution with $n-1$ degrees of freedom. The density of a Student's t distribution with ν degrees of freedom is

$$f(x; \nu) \;=\; \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \, \frac{1}{(1 + x^2/\nu)^{(\nu+1)/2}} \qquad x \in \mathbb{R}$$

c) Show that the posterior distribution of $(n-1)s^2/\sigma^2$ is chi-squared with $(n-1)$ degrees of freedom (that is a gamma distribution with shape $\alpha = (n-1)/2$ and scale $\beta = 2$).

d) Carefully explain the distinction between these results and the usual frequentist results that

$$\frac{\sqrt{n}(\overline{X} - \mu)}{S} \sim t_{n-1} \quad \text{and} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$