ASSIGNMENT 6 — Advanced Techniques II

For all questions below, provide all programming code and plots in the report. Unless stated otherwise, assume $\alpha = 0.05$

Markov Chain Monte Carlo (MCMC)

1. Use Gibbs sampling to randomly draw numbers from a bivariate normal distribution. The two conditional probabilities are: $p(x_{i+1} \mid y_i) = \mathcal{N}\left(\mu_X + \frac{\sigma_X}{\sigma_Y}\rho(y_i - \mu_Y),\ (1-\rho^2)\sigma_X^2\right)$ and $p(y_{i+1} \mid x_{i+1}) = \mathcal{N}\left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(x_{i+1} - \mu_X),\ (1-\rho^2)\sigma_Y^2\right)$. Pay attention to i versus $i+1$. Use the following constants: $\mu_X = 78.8$, $\sigma_X = 3.668$, $\mu_y = 211$, $\sigma_y = 26.904$, and $\rho = 0.81$ (10 marks).

a. Make an algorithm to perform Gibbs sampling (5 marks).

b. Plot the marginal distributions of X and Y (3 marks).

c. Plot the joint distribution X and Y (2 marks).

Bootstrapping

2. For the following data: [8.453532, 10.025041, 11.495339, 9.367600, 8.333229, 9.788753, 10.883344, 10.543059, 9.869095, 10.799819] (5 marks):

a. Bootstrap the mean and plot the histogram (2 mark).

b. Bootstrap the median and plot the histogram (1 mark).

c. For the mean bootstraps, find its mean and 95% Confidence Intervals (1 mark).

d. For the median bootstraps, find its mean and 95% Confidence Intervals (1 mark).

3. You want to test whether the following *paired* samples are significantly different from one another: pre = [22,25,17,24,16,29,20,23,19,20], post = [18,21,16,22,19,24,17,21,23,18]. Often researchers would run a paired sampled t-test, but you are concerned the data does not follow a normal distribution. The null hypothesis is that there is no difference between the paired samples. (6 marks).

a. Calculate the paired differences, that is post - pre, which will result in a vector of paired differences (pdiff0 = post - pre). (1 mark).

b. Calculate the mean of the paired differences (Xpdiff0). (1 mark).

c. Subtract a) from b) (pdiff1 = pdiff0 - Xpdiff0). (1 mark).

d. Bootstrrap c) *with* replacement (pdiff1) and plot the histogram (should be centered about zero). (2 marks.)

e. Find the two-tailed p-value. (2 marks).

4. Your advisor wants you to find the best fit parameters, and their confidence intervals, for a model. He gives you the following data: x-values = [-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5], y-values = [18.691806, 13.385586, 9.522445, 6.908887, 5.664939, 5.367358, 4.991651, 7.239345, 10.578517, 12.138775, 17.554915]. The theoretical relationship can be described as, $y = a + b \cdot x^2 + \epsilon$ (10 marks).

a. Build a least-squares loss function in R and find a and b using the nlm() function. Initialize your search with $a = 10$ and $b = 10$. (3 mark).

b. Use bootstrapping (10,000 resamples, *with* replacement) to:

   i. Plot the bootstrapped distributions for both a and b. (3 mark).

   ii. Find the mean of the bootstrapped distributions for both a and b. (1 mark).

   iii. Find the confidence intervals for both a and b. (2 mark).

   iv. Plot the original data and best fit curve using the means of the bootstrap distributions. (1 mark).