STAT743 FOUNDATIONS OF STATISTICS (Part II)

Assignment 2 Due in class at 12:30pm on March 1, 2018

Instructions:

1. Ensure that your full name and student number are clearly marked on your solutions.

2. Solutions need not be typed (although they can be) but must be readable.

3. Start each question on a new page and clearly indicate where each part of the question begins.

4. Submit questions in the same order as given below.

5. 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.

6. You must include a commented listing of your R code for any question which requires the use of computing.

7. No extensions to the due date and time will be given except in extreme circumstances and late assignments will not be accepted. Assignments are due at the **start** of class.

8. Students are reminded that submitted assignments must be entirely their own work. Submission of someone else's solution 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 We often observe data coming from more than one population but we do not know which population an observation comes from. This gives rise to *mixture distributions*. For two populations this has a density of the form

$$f_X(x) = \pi f_1(x) + (1-\pi) f_2(x)$$

where f_1 and f_2 are known densities such that $f_1(x) \neq f_2(x)$ for all x. In this question we shall look at estimation of π based on a sample $X_1, \ldots, X_n$ from this distribution.

a) Give the update equation of a Newton–Raphson algorithm to find the maximum likelihood estimator of π.

b) Now consider completing the data by adding $Z_1, \dots, Z_n$ such that $Z_i = 1$ implies that X_i comes from population 1 and $Z_i = 0$ implies that X_i comes from population 2. Z is often called a *latent variable*. Find the log-likelihood for π based on the complete data $(x_1, z_1), \dots, (x_n, z_n)$.

c) Using the complete data log-likelihood found in (b), calculate the EM quantity $Q(\pi, \hat{\pi}^{(r)})$ and find the EM update equation giving $\hat{\pi}^{(r+1)}$ in terms of the observed data and $\hat{\pi}^{(r)}$.

d) Observations in the following dataset are either from a normal($\mu = 10, \sigma = 1$) or a normal($\mu = 13, \sigma = 1$) distribution. Use both the EM algorithm and the Newton–Raphson method to estimate the probability that a random individual in the population comes from the distribution with mean 10.

9.29	9.73	10.13	9.00	12.74	9.70	9.08	9.33	10.10	10.86
12.86	11.45	9.55	9.78	9.49	13.38	13.35	14.31	10.03	9.76

Q. 2 Casella and Berger 8.17

Q. 3 Casella and Berger 8.33

Q. 4 **a)** Suppose that $X_1, \dots, X_n$ are a random sample from the an exponential family distribution with a single parameter θ and natural parameter $\eta = w(\theta)$ where w is a monotone function. Prove that this family has monotone likelihood ratio and hence find the uniformly most powerful test of

$$H_0 : \theta \geqslant \theta_0 \quad \text{V} \quad H_1 : \theta < \theta_0.$$

b) Suppose $X_1, \dots, X_n$ are a random sample from a population whose density function is

$$f_X(x) = \frac{\beta}{x^{\beta+1}} \qquad 1 < x < \infty, \ \beta > 0.$$

Use the result of part (a) to find the uniformly most powerful test of

$$H_0 : \beta \geqslant 1 \quad \text{V} \quad H_1 : \beta < 1$$

Conduct the test, giving a p-value and stating your conclusions clearly, based on the following observed data.

1.23	2.74	4.99	5.11	5.55	6.55	138.80	1.90	4.74	2.53
2.41	1.21	26.55	6.81	1.17	1.39	1.08	4.87	2.91	2.57