STAT 440/840 - CM 761 - Assignment 2 - Spring 2019

Name

Due: Tuesday, June 25 at 9:00am on Crowdmark

- **Assignment format policy**: Using RMarkdown or LaTeX is required and no hand written and/or imported screenshots will be accepted in the assignments. A mark of 0% will be assigned to the questions which were not complied in RMarkdown or LaTeX, and/or those which include hand-written solutions and/or screenshots.

---

1. The Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\sigma > 0$ is given by

$$f(x\,|\sigma, \alpha) = \begin{cases} \frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha-1} e^{-(x/\sigma)^{\alpha}} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

The MLE for the σ given α is

$$\widehat{\sigma} = \left[\frac{\sum_{i=1}^{n} x_i^{\alpha}}{n}\right]^{1/\alpha}$$

- Assuming $\alpha = 5.2$ use the following methods to construct confidence interval for σ
  i) standard bootstrap confidence interval (assuming normality),
  ii) bootstrap percentile interval,

  iii) bootstrap-t interval using the observed information,
  iv) bootstrap-t interval using the double bootstrap,
  v) the asymptotic normality of the MLE and
  vi) the log-likelihood ratio statistic.
- **Note:**
  - For each method that requires bootstrapping, sample from the data with replacement.

  - A clear answer with appropriate conclusions will obtain full marks.

a) [**8 Marks**] Construct confidence interval for σ using the following data and each of the method given above.

```
temp = read.csv( "eng-monthly-011942-112007.csv" )
speed = temp$Spd.of.Max.Gust..km.h.
```

b) [**12 Marks**] Using a simulation study with $(\sigma, \alpha) = (64, 5.2)$ and $m = 1000$ datasets, compare the coverage probabilities of the above confidence intervals.

2. In this question you will derive and implement an EM algorithm to fit a multivariate-normal distribution to Ozone (z) and Wind (x) from the air quality dataset. Make sure you define any notation that you introduce.

```
data(airquality)
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

a) [**1 Mark**] What is the joint distribution of the missing data and the observed data?

b) [**2 Marks**] State the conditional distribution of the missing data given the observed data.

c) [**2 Marks**] State the complete data likelihood

d) [**4 Marks**] E-step: Derive the expected complete data log-likelihood.

e) [**4 Marks**] M-step: Derive the updates for the parameters.

f) [**6 Marks**] Implement the above EM in `R` for the air quality dataset. Use starting values based on the complete cases. Give the MLE and plot the observed log-likelihood evaluated at each iteration.

g) [**1 Mark**] Plot the imputed dataset.