Computational Inference
Resit Project

Hand-in deadline: 5 p.m. Sunday 18th May
This assignment is worth **15%** of your total mark on the module.

Presentation
For each question, you should provide a clear explanation of the method you have used (including any working) together with your results. The discussion of your results should be brief (no more than a paragraph is needed). Annotated R listings must be included in an appendix. You are not required to explain any methodology described in the lecture notes.

1. Download the file `mixturenormal.txt` from MOLE. Read this file into R with the command
   ```
   x<-dget("mixturenormal.txt")
   ```
   (modifying the path as appropriate). The vector `x` contains data sampled from a mixture of k normal distributions, each with variance 1, but unknown mean. Plot a histogram estimate based on this data to estimate (visually) the value of k. Then use the E-M algorithm to estimate the mean and weight of each normal distribution. Plot the estimated density function using the maximum likelihood estimates of the unknown parameters, together with your histogram.

   You are expected to write your own EM code and not simply use a pre-written R library e.g. `mixtools`

2. Variance Test
   The MOLE file `ceramic.csv` contains information gathered by Said Jahanmir of the NIST Ceramics Division on the strength of 480 ceramic materials that were created in two different batches. It is of interest to discover whether the variance in ceramic strength depends upon the batch.

   - Write down a test statistic that can be used to assess if the variances of the two batches are the same. What is the observed value based upon the data?

   We will perform bootstrapping to test whether the variances of the two batches are the same. We do not wish to make an assumption about the distribution of the ceramic strength of either batch (e.g. that they have the same mean) and so we do not pool the batches when doing the bootstrapping.

- Perform non-parametric bootstrapping creating each new bootstrap dataset by resampling from batch 1 and batch 2 separately i.e. form the empirical cdf $\hat{F}_i$ for batch 1 using only the data from batch 1, similarly for batch 2.
- Plot a histogram of the resampled values of your test statistic and find a 95% percentile confidence interval. Hence test whether there is evidence that the variances in ceramic strength are dependent upon the batch number.