Exercise J-4.1: A classic example of maximum likelihood estimation is due to Fisher (1925, *Statistical Methods for Research Workers*. Oliver and Boyd: Edinburgh.) and arises in a genetic problem. Consider a multinomial observation $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ with class probabilities given by

$$
\begin{aligned}
p_1 &= \frac{2+\theta}{4} \\
p_2 &= \frac{1-\theta}{4} \\
p_3 &= \frac{1-\theta}{4} \\
p_4 &= \theta/4,
\end{aligned}
$$

where $0 < \theta < 1$. The parameter θ is to be estimated using maximum likelihood estimation based on the observed frequencies $x_1 = 1997,\ x_2 = 907,\ x_3 = 904, x_4 = 32$.

a) [10 points] The EM algorithm for this problem was derived in class. Write an R function that implements the EM algorithm obtained, implementing the following specific instructions: Stop your algorithm when either of the following criteria is satisfied:
   i. The number of iterations reaches 200
   ii. The modified relative error (as defined below) is less than *tollerr.* The tolerance values should be an input to your program.

   Run your program using the starting value $\theta^{(0)} = 0.02$ and *tolerr = 1e-6*. Your printed output should nicely include the following quantities at iteration $n = 1, 2, \ldots$:
   i. Iteration number n (2 digits, no decimals)
   ii. Value of $\theta^{(n)}$ (12 decimal places)
   iii. Value of the modified relative error (use 1 decimal with exponent notation, e.g. 2.0e-07)

$$\text{Relative error} \approx \frac{|\theta^{(n+1)} - \theta^{(n)}|}{\max(1, |\theta^{(n+1)}|)}$$

b) [3 points] The exact solution to this problem is $\theta^* = \left(-1657 + \sqrt{3728689}\right)/7680$. Numerically determine whether the EM algorithm is linearly, super-linearly, or quadratically convergent for this problem.

Exercise J-4.2: Suppose that $y_1, y_2, \cdots, y_n$ is a set of data from a three component mixture density

$$\alpha f_1(y; \boldsymbol{\theta_1}) + \beta f_2(y; \boldsymbol{\theta_2}) + (1 - \alpha - \beta) f_3(y; \boldsymbol{\theta_3}),$$

where $f_i(y; \boldsymbol{\theta_i}) = (2\pi\sigma^2)^{-1/2} e^{-(y-\mu_i)^2/2\sigma^2}$ is the $N(\mu_i, \sigma^2)$ density. Here $\boldsymbol{\theta_i} = (\mu_i, \sigma^2)$ for $i = 1,2,3$.

a) [3 points] Plot a density histogram of the variable y in the dataset ExJ42, and superimpose it by a Kernel smoother.

b) [5 points] Derive the EM algorithm for estimating the mixture parameters α and β and the parameters μ_1, μ_2, μ_3, and σ^2. Write your pseudo algorithm.

c) [10 points] Write an R function to code the EM algorithm in part (b) with the following input variables:

i. Data: $\boldsymbol{y} = (y_1, \cdots, y_n)$, an $n \times 1$ vector of data
ii. Parameter initial values: $\boldsymbol{\theta} = (\alpha, \beta, \mu_1, \mu_2, \mu_3, \sigma^2)$, a 6 by 1 vector
iii. Maximum number of iterations: An upper-bound for the maximum number of iterations
iv. Tolerance: a for the maximum relative error (MRE) stopping criterion as a stopping rule.

Your function should output the iteration process including iteration number, value of the lolog-likelihood at each iteration, and the MRE. At the end the MLE estimates for α and $\beta, \mu_1, \mu_2, \mu_3$, and σ^2should be printed. Then, use your function to fit the three component mixture model to the data given in the dataset ExJ42.

d) [4 points] Again graph a density histogram of the data, and superimpose the histogram by the three component normal mixture density with parameters obtained as in part (c).

e) [5 points] Compute the posterior probability of belonging to group 1, 2, and 3 for each case, and classify each case to one of the three groups based on the highest posterior probability. Then show your result graphically as follows: Draw a graph, with the x-axis indicating case number, and the y-axis including the group numbers 1, 2, 3. Then use red, green, and blue dots to indicate cases that are assigned to Groups 1, 2, and 3 respectively on the graph. Do you see a pattern, and if so, what does the pattern indicate about the classification of the cases?