Homework 5 – Part I: 13 points

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

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

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 using 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)

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

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. Justify your answer using your pogram.