Problem 1. Consider a locus on the X chromosome with two alleles, A and B. Recall that in class, we derived the approach of \( f_n \) and \( m_n \) to equilibrium in an idealized population, where \( f_n \) and \( m_n \) are the female and male frequencies of the A allele at generation \( n \), respectively. (See p. 5 of the lecture notes for more details.)

(a) In class, someone asked a question about the frequencies of the female genotypes A/A, A/B, and B/B at generation \( n \). I answered incorrectly that the frequency of A/A is \( f_n \); this is only true after equilibrium has been reached. Derive the correct female genotype frequencies at generation \( n \) for the idealized population.

(b) Using the relations

\[
\begin{align*}
m_n &= f_{n-1} \\
f_n &= \frac{1}{2}f_{n-1} + \frac{1}{2}m_{n-1}
\end{align*}
\]

and starting with the (rather extreme) values \( f_0 = .8 \) and \( m_0 = .2 \), plot the values of \( f_n \) and \( m_n \) against \( n \) for \( 0 \leq n \leq 10 \). Plot the values on the same graph, using some means to distinguish the male from the female. Please use a computer to produce the plot; the point of this problem is to familiarize yourself with some simple graphical commands in the language of your choosing.

Problem 2. The Rhesus blood group of an individual (usually written as either Rh+ or Rh−) is determined by an autosomal locus that we may assume to be diallelic. Denote the alleles as R and r. The R allele is dominant and gives rise to the Rh+ phenotype whenever it is present. Let \( n_R \) and \( n_r \) denote the number of Rh+ and Rh− individuals in a random sample of size \( n = n_R + n_r \) from a population assumed to be in Hardy-Weinberg equilibrium. Let \( p_R \) denote the frequency of the R allele.

(a) Give the loglikelihood function \( l(p_R) \) for the observed data.

(b) Maximize the answer to part (a) to derive an expression for the MLE \( \hat{p}_R \).

(c) Let \( (n_{R/R}, n_{R/r}, n_{r/r}) \) be the complete data. Give the complete data loglikelihood \( h(p_R) \).

(d) Suppose we observe \( n_R = 132 \) and \( n_r = 23 \). Using part (c), construct an EM algorithm to estimate \( p_R \). Start with \( p_R^{(0)} = .5 \) and record the values of \( p_R^{(k)} \) and \( l(p_R^{(k)}) \) in a table for \( 0 \leq k \leq 10 \). Verify that the eventual solution to which the EM converges agrees with the answer to part (b).

The assignment continues on the next page.
Problem 3. In a genetic linkage experiment, AB/ab animals are crossed to measure the recombination fraction $\theta$ between two loci with alleles A and a at the first locus and alleles B and b at the second locus. A and B are dominant.

(a) Verify that the offspring of an AB/ab × AB/ab mating have phenotypes AB, Ab, aB, and ab with probabilities

\[
\frac{1}{2} + \frac{(1 - \theta)^2}{4}, \quad \frac{1 - (1 - \theta)^2}{4}, \quad \frac{1 - (1 - \theta)^2}{4}, \quad \text{and} \quad \frac{(1 - \theta)^2}{4},
\]

respectively.

(b) Devise an EM algorithm to estimate $\theta$ and apply it to the counts

\[
(y_1, y_2, y_3, y_4) = (125, 18, 20, 34).
\]

Record the iterates $\theta^{(k)}$ until convergence occurs.

Tips: Split the first category into two so that there are five categories for the complete data. The calculations are simpler if you work with $\phi = (1 - \theta)^2$ as the parameter. You should obtain $\theta = .2082$. 