1. Recall the genetic linkage model. For the data \( Y = (125, 18, 20, 34) \) implement the EM algorithm to find the MLE of \( \theta \). Try starting your algorithm at \( \theta = 0.1, 0.2, 0.3, 0.4, 0.6, \) and 0.8. Did the algorithm convergence for all of these starting values? How many iterations were required for convergence?

2. Consider the \( k \)-component mixture model for observations \( y_1, \ldots, y_n \):

\[
f(y_i) = \sum_{j=1}^{k} p_j \phi(y_i; \mu_j, \sigma),
\]

for \( i = 1, \ldots, n \), where

\[
\phi(y_i; \mu_j, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{(y_i - \mu_j)^2}{2\sigma^2} \right),
\]

the \( \mu_j, p_j \) and \( \sigma \) are unknown, \( \sum_{j=1}^{k} p_j = 1 \) and \( 0 < p_j < 1 \). For \( 1 \leq i \leq n \), let

\[
z_{ij} = \begin{cases} 
1 & \text{if } y_i \text{ belongs to component } j, \\
0 & \text{otherwise.}
\end{cases}
\]

(Note that exactly one of \( z_{ij} \) is 1, \( 1 \leq j \leq k \), for each \( i \).) The augmented likelihood is then

\[
\prod_{i=1}^{n} \prod_{j=1}^{k} p_j^{z_{ij}} \phi(y_i; \mu_j, \sigma)^{z_{ij}}.
\]

(a) Derive the \( Q \) function for the problem. What is the E-step? What is the M-step?

(b) The data set \textit{galaxies.dat} represents the velocities at which 82 galaxies in the Corona Borealis region are moving away from our galaxy. If the galaxies are clustered, then the density of the velocities will be multimodal. Fit the above model to these data using the EM algorithm. Take \( k = 3 \). How did you assess convergence? How many iterations were required for convergence? Try several randomly selected starting points for the algorithm. Is there evidence that the observed likelihood is multimodal?