EE 511 Simulation Methods for Stochastic Systems
Project #3: Working with Mixtures

Problem 1 [EM]

- Write a 2-dimensional RNG for a Gaussian mixture model (GMM) pdf with 2 sub-populations. Use any function/sub-routine available in your language of choice.
- Implement the expectation maximization (EM) algorithm for estimating the pdf parameters of 2-D GMMs from samples.
- Compare the quality and speed your GMM-EM estimation on 300 samples of different GMM distributions (e.g. spherical vs ellipsoidal covariance, close vs well-separated subpopulations).

Problem 2 [Testing Faith]

Download the "old faithful" data set from blackboard. This has samples of a 2-D random variable: the first dimension is duration of the geyser eruption, the second is waiting time for the next eruption. Apply your GMM-EM algorithm to fit the data to a GMM pdf.

How many EM iterations are needed for convergence? Plot a contour plot of your final GMM pdf. Overlay the contour plot with a scatterplot of the data set. How would you use the GMM pdf estimates to cluster the data?

Problem 3 [Noise in GMM-EM] (Extra Credit)

Modify your GMM-EM routine by sampling and injecting Gaussian noise into the old faithful data at each iteration. Scale the noise to a fraction of the standard deviation in each dimension. And let the noise standard deviation decay at each iteration (e.g. inversely proportional to the square of the iteration counter).

Compare the average convergence time of the GMM-EM with and without noise. Plot the average convergence time for different initial noise standard deviations.

Turn in:

- A summary of your experiments including any relevant plots
- brief discussions of the results
- a print out of your code.