STA 410/2102 — Practice Questions for Third Test

Please note that these questions do **not** cover all the topics that may be on the test.

1. Write an R function that numerically evaluates a one-dimensional integral using the trapezoid rule. The function should take as arguments the function to integrate, the lower bound for the integral, the upper bound for the integral, and the number of trapezoids to use to approximate the integral over this range. You should try to avoid evaluating the function at the same point twice.

2. For legal reasons, it is important to determine what fraction of the air pollution particles at a certain location were emitted by a certain factory. By measuring particles just outside the factory, it has been determined that particles from this factory have diameters that are normally distributed with mean 25 microns and standard deviation 5 microns. There are no other major sources of particulate pollution in the area, but some particles from minor sources are present, whose diameters are known to be uniformly distributed over the range from 10 microns to 100 microns (which is the range of particles that can be measured by the device used).

   Given a vector, x, of independent observations of particle diameters at the location of interest, write an R program to estimate the proportion, p, of the particles at this location that come from the factory. Your program should find the maximum likelihood estimate for p using the EM algorithm, applied to the model in which the data comes from a mixture of the $N(25, 5^2)$ and the Uniform$(10, 100)$ distributions, with the normal distribution having proportion p and the uniform distribution having proportion $1 - p$.

   Your R function should take as arguments the data, x, an initial guess at p, and the number of iterations to do. Show how you derived the appropriate formulas to use in the E and M steps.

3. This question concerns the same problem as Question 2. Suppose that one of the parties to the legal dispute prefers Bayesian methods to maximum likelihood estimation. They claim that based on meteorological and other data, the proportion of particles coming from the factory cannot be less than 0.1 or greater than 0.6. They advocate using a prior distribution for p that is uniform over this range.

   Write an R function to find the posterior mean of p using this prior. You should use the trapezoidal integration function of Question 1 to do the integration (regardless of whether you were able to actually answer that question). Your function should take as arguments the data, x, and the number of trapezoids to use to approximate the integral.

4. Let the state variable, x, for a Markov chain consist of two components, x_1 and x_2. The possible values for x_1 are 0 and 1. The possible values for x_2 are 0, 1, and 2. (There are therefore six possible values for the entire state: 00, 01, 02, 10, 11, and 12.) Define the distribution π by the probabilities

$$\pi(x) \;=\; \begin{cases} 1/4 & \text{if } x_2 = x_1 \\ 1/4 & \text{if } x_2 = x_1 + 1 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Sketch how the Gibbs sampling procedure would work for this distribution, giving in particular the details of what conditional distributions to sample from when, and what these conditional distributions are.

   (b) Write down explicitly the transition probabilities for the Gibbs sampling Markov chain that you described above, in which first x_1 and then x_2 are updated. (Ie, write down the 6 by 6 matrix whose entries are $T(x, x')$ for all x and x'.)

   (c) Show explicitly from the definition of invariance that the Markov chain you described above leaves π invariant.