For this assignment you can do it in a group of maximum two people. Objectives of this assignment:

- To gain experience in using the EM algorithm.
- To learn about the inverse transform method and the rejection sampling method.

Altogether there are five questions. You will need to submit your homework via SmartSite.

1. Consider the multinomial distribution with four outcomes, that is, the multinomial with pdf

$$p(x_1, x_2, x_3, x_4) = \frac{n!}{x_1!x_2!x_3!x_4!} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}, \quad \sum_{i=1}^{4} x_i = n, \quad \sum_{i=1}^{4} p_i = 1.$$

Suppose the probabilities are related by a single parameter $0 \le \theta \le 1$:

$$\begin{aligned} p_1 &= \frac{1}{2} + \frac{1}{4}\theta \\ p_2 &= \frac{1}{4} - \frac{1}{4}\theta \\ p_3 &= \frac{1}{4} - \frac{1}{4}\theta \\ p_4 &= \frac{1}{4}\theta. \end{aligned}$$

Given an observation $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$, the log-likelihood is

$$l(\theta) = x_1 \log(2 + \theta) + (x_2 + x_3) \log(1 - \theta) + x_4 \log \theta + c. \tag{1}$$

To use the EM algorithm on this problem, consider a multinomial with five classes formed from the original multinomial by splitting the first class into two with probabilities $1/2$ and $\theta/4$. The original variable x_1 is now split into $x_1 = x_{11} + x_{12}$. Under this reformulation, we now have a MLE of θ by considering $x_{12} + x_4$ to be a realization of a binomial with $n = x_{12} + x_4 + x_2 + x_3$ and $p = \theta$. However, we do not know x_{12}, and the complete data log-likelihood is

$$l_c(\theta) = (x_{12} + x_4) \log \theta + (x_2 + x_3) \log(1 - \theta). \tag{2}$$

   (a) Suppose $\boldsymbol{x} = (125, 18, 19, 35)$. Find the MLE of θ by maximizing (1).

   (b) Using (2), develop an EM algorithm for estimating θ. Note: you should be able to combine the E-Step and the M-Step together; i.e., $\hat{\theta}^{(t+1)}$ can be expressed in terms of $\hat{\theta}^{(t)}$.

   (c) Compare your answers obtained in (a) and (b).

2. Consider a sample of failure time measures $x_i = (\min\{y_i, c_i\}, \delta_i)$, where $\delta_i = 1$ if we have observed the actual failure time y_i and $\delta_i = 0$ if we observe the censoring time c_i. The distribution of failure times is modelled using a Weibull distribution with parameter $\alpha > 0$ whose density is

$$W(y|\alpha) = \frac{2y}{\alpha^2} \exp\left(-\frac{y^2}{\alpha^2}\right), \quad y > 0.$$

Suppose the sample observed is $\{x_1 = (2.3, 1), x_2 = (2.5, 0), x_3 = (1.1, 0), x_4 = (3.1, 1)\}$.

   (a) Develop an EM algorithm for estimating α.

(b) Starting with the initial estimate $\alpha^{(0)} = 1$, obtain the MLE for α with the above sample.

3. Use the inverse transform method to sample from the density

$$f(x) \propto e^{-x}, \quad 0 < x < 2.$$

Note that x is less than 2. Detail your algorithm. Draw a sample of 5000 observations and plot the estimated density of the sample.

4. Consider the following probability density function:

$$f(x) \propto q(x) = \frac{e^{-x}}{1 + x^2}, \quad x > 0.$$

Use rejection sampling to sample from $f(x)$ with the following envelope density functions:

$$g_1(x) = e^{-x}, \quad g_2(x) = \frac{2}{\pi(1 + x^2)}, \quad x > 0.$$

(a) For each density function ($f(x)$, $g_1(x)$ and $g_2(x)$), draw a sample of 5000 random observations and plot the estimated density function for $0 < x < 5$.

(b) Comments on the speeds of sampling and the results using $g_1(x)$ and $g_2(x)$.

5. Design a rejection algorithm to sample from the following density on the upper right quarter of the unit disc:

$$f(x, y) \propto x^\alpha y, \quad x > 0, \quad y > 0, \quad x^2 + y^2 \leq 1.$$

You don't need to run simulations for this problem. Just describe your algorithm in detail.

—— End of Assignment 3 ——