MAT5182 Assignment 2 - Optimization and Simulating random variables
Due Date: February 15, 2018

NOTE:

1. Please see the guidelines on the class website for formatting your assignment. If the assignment is not formatted adequately, you may lose marks. In particular, the marker won't be searching your assignment for the answer/figure/code. You need to clearly indicate each of these.

2. Feel free to work in pairs on any of the programming portions of the question. Just be sure to write up your solution separately.

Questions:

1. Recall the following distributional relationships:

   i. For $U \sim U(0,1)$, $X = -\frac{1}{\lambda} \ln U \sim Exp(\lambda)$.

   ii. For $X_1, X_2, \ldots, X_n$ independent $Exp(\lambda)$ random variables, $\sum_{i=1}^{n} X_i \sim Gamma(n, \lambda)$ (note n is integer).

   iii. For $X \sim Gamma(a, 1)$ and $Y \sim Gamma(b, 1)$ (a and b both integer; X and Y are independent), $\frac{X}{X+Y} \sim Beta(a, b)$.

   (a) Propose a method and write a corresponding function to sample from the beta distribution with integer-valued parameters for a and b. NOTE: Your function should depend only on a $U(0,1)$ random number generator (i.e., don't use `rgamma` or `rexp`)

   (b) Generate 1000 random variables using your function and compare the distribution of your generated values to the true distribution for 4 sets of parameter values (generate 4 plots and put them all on the same page).

   (c) Propose an algorithm for sampling from the $F_{m,n}$ distribution. In what cases does your proposed algorithm apply?

2. Generating Normal random variables

   (a) Let $X_1 = \sqrt{-2logU_1}\cos(2\pi U_2)$ and $X_2 = \sqrt{-2logU_1}\sin(2\pi U_2)$, where U_1 and U_2 are iid uniform(0,1). Show that X_1 and X_2 are iid N(0,1).

   HINT: Find two functions of the form $h_1(X_1, X_2) = U_1$ and $h_2(X_1, X_2) = U_2$ using trigonometric identities.

   (b) Use (a) to implement an algorithm to sample $X \sim N(\mu, \sigma^2)$

   (c) Implement an accept-reject method to sample $X \sim N(\mu, \sigma^2)$ using a double exponential distribution envelope. NOTE: You will probably have to write a function to sample from a double exponential.

   HINT: Both the double exponential and normal are symmetric about 0.

(d) Compare the run time of the two approaches. Use plots to compare your samples to their true distributions.

3. The file bivariatenormal.txt contains 50 samples of 2 variables with some missing data (coded as NA). Assume that each data point (row) is biariate normal ($N_2(\boldsymbol{\mu}, \Sigma)$),

   (a) Derive the EM algorithm for joint maximum likelihood estimation of the parameters. Hint: conditional and marginal distributions of the bivariate normal distribution will be normal.

   (b) Determine the MLEs from a suitable starting point.

4. The file oilspills.txt contains data on oil spills in US waters during the 1974-1999 period. Each row of the file gives:

   (1) The year; (2) N: the number of spills in that year; (3) X_1: the amount of oil shipped from international sources in a given year; (4) X_2: the amount of oil shipped from US sources in a given year.

   We expect that the amount of oil shipped in a year is related to the number of oil spills. We therefore fit the following model:

$$N_i \sim Poisson(\lambda_i)$$

   where

$$\lambda_i = \alpha_1 X_{i,1} + \alpha_2 X_{i,2}$$

   (a) Derive and implement a NR scheme for finding the MLEs of α_1 and α_2

   (b) Derive a Fisher scoring scheme for finding the MLEs of α_1 and α_2

   (c) Compare the ease of implementation and performance of the two methods

   (d) Estimate the standard errors of the MLEs of α_1 and α_2