Stat 406: Homework 7

issued Novemver 3, 2014

due November 10, in class/6PM via Ctools

1. The LAI data set in Ctools contains leaf area index values $L(i, j, k)$ for locations (i, j) and month k. Derive a Monte Carlo method for calculating the accumulated LAI for the North American continent in a year, say year 2005. Your MC algorithm should involve drawing n locations uniformly from within the North American continent, where $n \ll 60 \times 120 = 720$, the total number of locations (image pixels) where we have collected the LAI values. Describe your method, and produce a plot demonstrating the behavior of the MC estimate and MC error as n increases: $n = 10, 20, 30, \ldots, 200$. (The "truth" can be obtained by taking sum over all locations).

2. Consider the following integral:

$$C = \int_0^\infty (\sin x)^2 x^2 e^{-x^2/2} dx.$$

   (a) To estimate C using MC integration, there may be multiple choices of distributions according to which you may draw samples from for a MC method. Two such choices are Gaussian and gamma distributions. Derive your algorithms accordingly.

   (b) Implement the two algorithms by R programs, and calculate the Monte Carlo errors for both methods. The Monte Carlo errors should be plotted as in page 17 in the lecture notes (e.g., for sample sizes $n = 1, 100, 200, \ldots, 5000$). Of the two which one is better? Please explain.

   (c) Given the following function defined on $[0, \infty)$:

$$f(x) = \frac{1}{C}(\sin x)^2 x^2 e^{-x^2/2} dx,$$

   Note that C is also called the normalizing constant so as to make f a valid density on the positive reals. Derive a Monte Carlo method based on important sampling for approximating the following quantity:

$$I = \int_1^{10} (x-1)^2 f(x) dx.$$

   Be sure to provide the sample coefficient of variation and the MC error for different choices of sample size.