Numerical Assignment

November 2, 2018

1. (a) To estimate π, we usually consider a unit circle and the smallest square containing the circle. Instead of a square, if we consider the smallest equilateral triangle containing the circle and generate observations from a uniform distribution on the triangle, will it lead to a better estimate of π? Justify using Monte Carlo techniques.

   (b) State as many methods as you can to estimate e using both numerical approximations and Monte Carlo methods. Compare them using appropriate simulations. You can also use results from large sample theory like Central Limit Theorem, Extreme Value Theory, etc.

2. Generate observations from a triangle, and a convex polygon using (i) direct geometric approach and alias method, and (ii) accept-reject algorithm.

3. Simulate uniformly from the circle $\{(x_1, \ldots, x_d) : x_1^2 + \cdots + x_d^2 \leq 1\}$ on $\mathbb{R}^d$ for $d = 2, 5, 10, 25$ and 50 using the following methods:
   (i) accept-reject, (ii) MCMC and (iii) spherical symmetry.
   Give relevant plots (on $\mathbb{R}^2$) to summarize your findings for these three methods with varying values of d.

4. Consider minimizing the following (complicated) function in $\mathbb{R}^2$:

   $$(x\sin(20y)+y\sin(20x))^2\cosh(\sin(10x)x)+(x\cos(10y)-y\sin(10x))^2\cosh(\cos(20y)y).$$

   Use deterministic algorithms as well as Monte Carlo techniques. The global minimum for this function is 0, and attained at the point $(0, 0)$. (*Why?*)

5. In a blood donation camp organized at IIT Kanpur, there were N donors. Out of these N donors: n_1 had blood group A, n_2 of them had blood group B, n_3 of them had blood group AB, and the rest had blood group O. Based on this information, estimate the allele frequencies for A, B and O using the following methods: (i) Fisher's scoring, and (ii) EM algorithm. Compare the performance of these two methods theoretically, and validate using an appropriate simulation study.

6. Suppose that $p_n(t) = X_0 + X_1 t + \cdots + X_n t^n$ is a polynomial in t, where each X_i is randomly chosen from the following distributions:

   (i) ± 1 w.p. $1/2$, (ii) $N(0, 1)$, (iii) $C(0, 1)$ and (iv) $Exp(1)$ with n = 5, 10, 25, 50 and 100.

   Use Monte Carlo simulations to answer the following questions:
   (i) How many real roots does the polynomial $p_n(t)$ have, on average?
   Let A_n denote the number of real roots of $p_n(t)$. You need to give idea about how the sequence $E(A_n)$ behaves.
   (ii) More generally, given a subset of the complex plane, how many roots of $p_n(t)$ are in the given subset, on average?

Guidelines and Instructions:

Use of R packages is NOT encouraged. You will get more credits for writing your own codes to generate data and for algorithms.

Each question has 5 points. Total is 30 points.

Please submit the following THREE files ONLY:

(i) writeup (.pdf/word file),
(ii) codes (.R/.C/.m/.py/text file) and
(iii) supplement (scanned .pdf/word file)
to the gmail id: `assignment.stat.iitk@gmail.com`.

DO NOT submit anything to my IITK email/other gmail ids.

In the subject of the email, strictly use the format '`ROLL NUMBER - NAME`**' ONLY (e.g., 12345678 - AAAAAA BBBB).**

Extended deadline : 11.11.2018 (Sunday) strictly by 11:59pm IST.