BIOS 560R Advanced Statistical Computing
Fall 2012
Homework 1
Due 9/25/2012 at 4pm before the class

Answer all question. For programming questions, provide a short description of the method/algorithm. There is no requirement for programming language, but an interpreted language (R, Matlab, SAS) is highly recommended. The codes for all problems need to be saved in a **single** file named NAME_hw1.EXT. Replace NAME by your name, and EXT by proper extension name, e.g., R, sas, m, etc. Provide adequate comments in the codes to clearly mark the section for different questions. The codes should generate all results and figures in the homework. Please make sure the codes are self-contained, e.g., does not depend on platform, can be run at any other machine in any subdirectory, and does not require user input.

To get full credit for your solutions of the homework problems, please:
* Hand in a hard copy of your answers including code, results, and plots.
* Submit an electronic version of the codes online at blackboard.emory.edu.

Problem 1. Implement Buffon's needle idea computationally and present the estimate of π (mean and standard deviation) using sample sizes 1,000, 10,000 and 100,000.

Problem 2. Prove that the Gibbs sampler is a special case of Metropolis-Hastings Algorithm. (Hint: show that the acceptance ratio is 1.)

Problem 3. Evaluate integral $\int_0^1 sin^2(1/x)dx$ using grid point evaluation, vanilla Monte Carlo and importance sampling approaches. With the same number of samples, compare their estimation accuracy.

Problem 4. Design a slice sampler scheme to sample random variates from Inverse Gaussian distributions with density $f(x) \propto \sqrt{\frac{1}{x^3}} e^{-x-\frac{1}{x}} I(x > 0)$. check the convergence of the Markov chain using trace plots, autocorrelation plot and Gelman-Rubin statistic.

Problem 5. Implement a Gibbs sampler haplotype phasing program. Write down the conditional distributions. Perform s simulation study with 5 SNPs to evaluate the performance of your program.
(Hint: Step 1: specify "true" haplotype frequency θ, assume sparsity and non-uniformity. Step 2: simulate 50 individuals by randomly select 2 haplotypes from the pool of the haplotypes according to their frequencies. Step 3: form genotype. Step 4: run your MCMC program to get haplotype phase and estimated θ. Step 5: report the accuracy of both.