THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT6011/7611/8305 COMPUTATIONAL STATISTICS (2021 Fall)

Assignment 1, due on October 5

All numerical computation MUST be conducted in Python, and attach the Python code.

1. **Question 2 (Metropolis-Hasting algorithm)**

   (a) Assume $\mathbf{X} = (X_1, \ldots, X_{10})$ follows a multivariate standard normal distribution $\pi(\mathbf{x})$. Build a Random Walk algorithm with the Gaussian kernel to estimate $E[D]$, where $D = f(\mathbf{X}) = \sqrt{\sum_{j=1}^{10} X_j^2}$. Firstly, tune the step size such that you have roughly 23.4% acceptance rate. What step size do you choose? Then, draw 10000 samples to do the estimation, set the initial point $\mathbf{x}_0 = \mathbf{0}$, set the burn-in parameter to be 1000 and set the thinning parameter to be 5.

   (b) A simple techique called batching can help us to build a confidence interval with MCMC samples. We divide the above obtained 10000 samples $\{d_1, \ldots, d_{10000}\}$ into 20 batches, where $d_i = f(\mathbf{x}_i)$, and $\mathbf{x}_i = (x_{i1}, \ldots, x_{i10})$ is a sample in the MCMC sequence. Do the estimation in each batch:

   $$\overline{d}_b = \frac{1}{500} \sum_{i=500(b-1)+1}^{500b} d_i,$$

   for $b = 1, \ldots, 20$. Estimate that

   $$s^2 = \frac{1}{20(20-1)} \sum_{b=1}^{20} (\overline{d}_b - \overline{d})^2.$$

   So, the 95% confidence interval would be

   $$\overline{d} \pm t_{(19)}^{0.975} s.$$

   Please give this interval.

   (c) In the self-normalized IS, the optimal proposal is propotional to $\pi(\mathbf{x})|f(\mathbf{x}) - E[D]|$. Replace $E[D]$ by the estimate $\hat{E}[D]$ obtained in (a), draw samples from the optimal proposal based on the Random Walk, and weight each obtained sample by the weighting function

   $$w(\mathbf{x}) = \frac{1}{|f(\mathbf{x}) - \hat{E}[D]|}.$$

Follow a similar procedure of (a) and (b) to calculate the estimate of $E[D]$ and the 95% confidence interval by the self-normalized IS. Does your confidence interval become wider or narrower? (hints: (1) Remember to ensure the 23.4% acceptance rate; (2) You don't need to consider the weights of the averages for each batch in this question.)