DUE DATE: PARTS I & II FRIDAY, SEPTEMBER 18 2020.

TYPE YOUR SOLUTION USING WORD OR LaTeX,
SUBMIT A PDF DOCUMENT ON CANVAS
HANDWRITTEN SOLUTIONS WILL NOT BE GRADED.

The purpose of this homework is to analyze theoretically and observe experimentally the concentration of distances and angles in the unit hypercube and unit sphere centered at the origin.

PART I: CONCENTRATION IN THE UNIT HYPER-CUBE

We recall the following facts about the uniform probability distribution.

Theorem 1 *Let* x *be a random variable uniformly distributed on the interval* $[a, b]$, *with distribution function*

$$f(x) = \begin{cases} \frac{x-a}{b-a} & \textit{if } x \in [a, b], \\ 0 & \textit{otherwise.} \end{cases} \tag{1}$$

The mean (expected value) of x *is* $\mathbb{E}[x] = (a + b)/2$, *and the variance is* $(b - a)^2/12$.

We begin with simple calculations on one-dimensional random variables.

Assignment [20 = 10 + 10]

1. Let x and y be independent random variables with uniform distribution on $[-1, 1]$. What are the expected values
   - $\mathbb{E}\left[x^2\right]$,
   - $\mathbb{E}[x - y]$,
   - $\mathbb{E}[xy]$,
   - $\mathbb{E}\left[(x - y)^2\right]$.
2. Let $X = (x_1, \ldots, x_n)^T$ and $Y = (y_1, \ldots, y_n)^T$ be two independent random vectors with uniform distribution in the unit cube centered at the origin in $\mathbb{R}^n$,

$$K_n = \left[-\frac{1}{2}, \frac{1}{2}\right] \times \cdots \times \left[-\frac{1}{2}, \frac{1}{2}\right]. \tag{2}$$

   (a) What is the expected squared distance, $\mathbb{E}\left[\|X - Y\|^2\right]$?

   (b) What is the expected angle between X and Y, $\mathbb{E}[\angle X, Y]$? you can compute $\mathbb{E}[\cos \angle X, Y]$.

Assignment [20=10+10] We consider the main diagonal of the cube defined by the unit vector

$$\vec{u} = \begin{bmatrix} 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{bmatrix}. \tag{3}$$

3. Use Hoeffding inequality to show that

$$\mathbb{P}\left(|\langle x, \vec{u}\rangle| > \varepsilon\right) \leq 2e^{-2\varepsilon^2} \tag{4}$$

4. Take $\varepsilon = \sqrt{\log 20/2} \approx 1.2$ to conclude that the points in K_n concentrate in the thin slab, of thickness ε, centered around the hyperplane $\vec{u}^\perp$ (that is the plane going through the origin and orthogonal to $\vec{u}$.)

Assignment [30]

5. Randomly generate 400 points inside K_{100} and plot the histogram of the distances between points and the angle between the vectors from the origin to the points, for all pairs of points. Explain your findings in terms of the theoretical analysis that you performed in question 2.

Part II: Sampling the unit ball: a naive approach

We consider the following problem: how can we generate $N = 10,000$ samples of points drawn uniformly inside the unit ball in $\mathbb{R}^{400}$. This problem is equivalent to constructing 10,000 vectors with $n = 400$ co-ordinates and a norm less than 1.

Let $B^n(1)$ be the ball of unit radius centered at the origin in $\mathbb{R}^n$. We consider a subset $A \subset B^n(1)$, and we measure its size using the uniform measure inside the unit ball $B^n(1)$ in $\mathbb{R}^n$,

$$\mu_n(A) = \frac{\text{vol}(A)}{\text{vol}(B^n(1))} \tag{5}$$

In order to generate samples according to the measure μ_n, we can use a standard rejection method. The principle of the rejection method is to sample from a uniform distribution in $Q = [-1, 1]^n$, and keep only the points that fall inside the ball. The algorithm is described below.

Algorithm 1 Generate N samples in the ball $B^n(r)$ with the rejection method

Require: N: number of samples
Require: n: dimension of the space

for $i = 1$ to N **do** // For each point in the cube Q, generate
  for $k = 1$ to n **do** // the coordinates, uniformly one at a time
    $u_i[k] \leftarrow U[0,1]$ // $u_i[k]$ is uniformly distributed in $[0,1]$
    $x_i[k] \leftarrow r(1 - 2 * u_i[k])$ // $x_i[k]$ is uniformly distributed in $[-r, r]$
  end for
  $\|x_i\| \leftarrow \sqrt{x_i[1]^2 + \cdots x_i[n]^2}$ // Euclidean norm of x_i
  if $\|x_i\| \leq r$ **then return** x_i // if x_i is inside $B^n(r)$, we keep it
  end if
end for

Facts: the volume of a ball

We recall that the volume of a ball of radius r for the l^2 norm in $\mathbb{R}^n$ is given by,

$$\operatorname{vol} B^n(r) = \frac{(\sqrt{\pi}\, r)^n}{\Gamma(n/2+1)}, \tag{6}$$

Assignment [50=5+25+20]

6. Derive the expression for the volume $V(Q)$ of

$$Q = [-1, 1] \times \cdots \times [-1, 1]\,. \tag{7}$$

7. Prove that

$$\frac{V\left(B^n(1)\right)}{V(Q)} \longrightarrow \frac{1}{n^n} \quad \text{as} \quad n \to \infty. \tag{8}$$

HINT: use Stirling formula.

8. Generate 10,000 sample uniformly in $[-1, 1]^n$ with algorithm 1, for $n = 1, \cdots, 400$. Plot the number of points rejected in algorithm 1 as a function of n. These points are inside the cube $[-1, 1]^n$, but not inside the ball $B^n(1)$. Comment.