EE 114 - Coding Assignment II

The goal of this assignment is developing the ability to numerically calculate probabilities and probability density functions. We will consider simple events and random variables that we can numerically simulate to estimate their properties.

Deliverable: You are expected to return

- a report containing your findings and all necessary figures
- all your code as a separate attachment

You should upload your report/code in iLearn. Graded out of 10 points.

1 Random points over a square

We will study the Euclidian distance between two random points generated over the unit square.

• Generate a random point P_1 over the unit square $[0,1] \times [0,1]$ uniformly at random. In other words, x and y coordinates of P_1 are uniform random variables over the interval $[0,1]$.

• Generate another random point P_2 with same distribution independent of P_1.

• Let X be the random variable which is the Euclidian distance between P_2 and P_1. In other words

$$X = \text{distance}(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

where (x_i, y_i) are the (x, y) coordinates of P_i.

• (2 pts) What is the distribution of X? This is really though to calculate by hand. However, we can run a lot of independent experiments by generating P_1, P_2 and calculate resulting X. This way, we can numerically estimate the probability distribution.
Your task: Plot the distribution of X by running 1,000,000 independent experiments.

Note that the range of X is $[0, \sqrt{2}]$ since two points over the square are at most $\sqrt{2}$ apart.
Hint: Split this interval into small chunks (such as $[0, 0.01]$, $[0.01, 0.02]$, $\dots$) and calculate the probability of X landing on each of this pieces.

• (2 pts) Let A be the event that the two points are at least 0.5 apart (i.e. $X \geq 0.5$). Let us calculate the probability of A. Again, tun 1,000,000 trials to numerically calculate $P(A)$. Let p_N denote the numerical probability which is the # of times A happened in the first N experiments. Plot p_N as a function of N for $1 \leq N \leq 1,000,000$. Here, p_N should be y-axis and N should be x-axis. What is $P(A)$?

2 Three random points over a cube

This part is similar to above, but we work with 3 points over a 3-D space.
• Generate three independent random points P_1, P_2, P_3 over the unit cube. In other words, each have coordinates randomly generated between [0,1].
• Let X be the random variable that is the minimum Euclidian distance between these points given by

$$X = \min(\text{distance}(P_1, P_2), \text{distance}(P_1, P_3), \text{distance}(P_2, P_3)).$$

• **Your tasks** are same as above.

1. (2 pts) Plot the distribution of X by running 1,000,000 independent experiments.
2. (1 pts) Let A be the event that distance between all points are at least 0.5 i.e. $X \geq 0.5$. Let p_N be the numerical estimate of $P(A)$ after N trials. Plot p_N as a function of N. What is $P(A)$?

3 Volume of a simplex

Consider a d-dimensional space where each point has d coordinates $(x_1, \ldots, x_n)$. A simplex is the region of the space defined as the following set of points

$$S = \{(x_1, \ldots, x_n) \mid \sum_{i=1}^{d} x_i \leq 1 \quad \text{and} \quad x_i \geq 0\}$$

In words, this is the region where coordinates are positive and they add up to at most 1. In case of a 2D plane, S is the triangle induced by points $(1,0)$, $(0,1)$, $(0,0)$.
• Using the ideas above, we can find the volume of a simplex as follows. Generate d random variables $(X_i)_{i=1}^{d}$ over $[0,1]$. Note that $(X_1, \ldots, X_d)$ is the coordinates of a random point over the d dimensional unit cube. Define the sum

$$Y = \sum_{i=1}^{d} X_i.$$

If $Y > 1$, it means point is outside of simplex, otherwise it is inside. With this observation, the probability of $Y \leq 1$ is actually equal to the volume of the simplex divided by volume of the unit cube. Luckily, volume of the unit cube is simply equal to 1! Hence

$$P(Y \leq 1) = \frac{\text{Vol}(S)}{\text{Vol}([0,1]^d)} = \text{Vol}(S).$$

• **Your task:** (2 pts) Find the volume of a simplex by doing 1,000,000 experiments for $d = 1, 2, 3, 4, 5$.

• (1 pts) Do you think there is an analytic formula for the Vol(S)? Write down your best guess.

Remark: You can chat with your friends on basic Matlab/Python functionalities such as plotting tools. However, you are expected to do all probability related questions/calculations yourself.
Cheating will result in disciplinary action which may result in failing the class.