Probability and Statistics (ECE 3710)

Project 1

Instructor: Dr. Mohammad Shekaramiz

In this project, you are asked to perform some simple tasks using MATLAB/PYTHON.

Submission type: Online, Canvas

Reading: **Chapters 1 and 2 of our textbook, presentation slides and lecture notes on Canvas.**

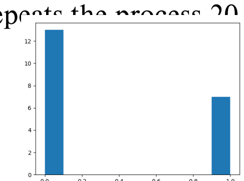
Q1. **(50 points)** An engineer or scientist is measuring the output of a pin of an IC. It turns out that the outputs are either 0 or 1 (one can think of 0 as 0 volts and 1 as the 5 volts). The engineer /scientist is informed that the output is governed by a probability mass function. Therefore, the output can be treated as a random variable and since the outcomes belong to the set {0,1}, it is a discrete random variable (we denote it by X). It turns out that the probability of getting X=1 is 0.3, i.e., p(X=1)=P, where P=0.3. The probability mass function (pmf) is given as follows

$$f(x) = \begin{cases} 1-P, & x = 0 \\ P, & x = 1, \end{cases}$$

where P = 0.3. This pmf can also be represented as

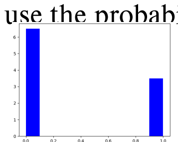
$f(x)=P^{x}(1-P)^{1-x}$.

a) Find the mean (expected value) of X using the given pmf. This means that what do we get in the output, on the average.
b) Find the variance of the random variable X using the given pmf.
c) Now the engineer/scientist, starts measuring the output. He/she repeats the process 20 times. The outcomes are as follows.
   X = [1 0 0 0 0 0 0 0 1 0 1 1 1 1 0 0 0 0 1 0]



c) Plot the histogram of data
c) Normalize the histogram and plot it.
c) Find the mean and variance of the collected samples. Find the probability of p(X=1) from the collected measurements (is this the same as the actual probability P?). Compare the results with the actual mean and variances. Are they different? Explain what you learned from this problem.
c) Plot the data with the approximated pmf (look at the actual pmf and instead of P use the probability of p(X=1) you found from the data). Plot the approximated and the actual pmf.

Data Mean = 0.35
Data Variance = 0.2395
The two sets of figures are very close, but not identical. This shows that the formula and figures are both good and realistic.



Note: The random variable in this problem is called Bernoulli. Meaning that X is governed by a Bernoulli distribution. In most real world problems, we do not now the actual distribution of our data. However, we can model our data via a famous distribution (such as Bernoulli) based on the samples we get (look at what you did in part c).

Q2. **(50 points)** An engineer/scientist is working on a set of samples from a blood test. It turns out that the samples can be treated as a continuous random variable and governed by the following probability density function (pdf)

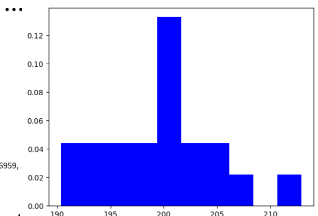
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_x)^2},$$

where σ^2 and μ are the variance and the mean of the distribution (This is referred to as Gaussian distribution). However, the engineer/scientist does not know the mean and variance of this distribution, but rather is given the data (blood glucose mg/100ml) below.

X = [201.4579, 200.9891, 207.9385, 195.9777, 203.4831, 204.1754, 198.7814, 201.0784, 194.1708, ...

194.2602, 200.5244, 203.6113, 212.9275, 196.6655, 200.9367, 199.5875, 190.3349, 197.8052, ...

191.0266, 204.2019].



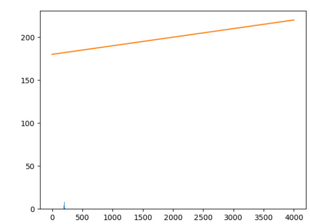
Mean = 199.9967

Variance = 30.0239

a) find the mean and variance of this data.
b) write down the pdf of data based on the mean, variance, and the given pdf.

PDF = [-9625142452.326687, -9580398446.997234, -10254355611.236843, -9108605096.325226, -9819632879.606, -9886564200.273756, -9371088743.876358, -9588913532.0279, -8941418051.959976, -8949653551.786959, -9536148639.54779, -9832010059.81279, -10752317272.32228, -9172652099.386997, -9575403674.26011, -9447246284.05604, -8591627061.034912, -9279273371.473755, -8654186548.7173, -9889130728.40058]

c) plot the histogram of data with the bin size of 5.
d) plot the pdf based on the **equation given for f(x) and the obtained mean and variance from part (a). For the plotting purposes, define x=180:0.01:220 (in MATLAB) or x=np.arrange(180,220.01,0.01) when using PYTHON, and plot f(x) vs. x.**



e) Explain what you learned from this problem.

I hope this data comes from subjects who are already diabetic because every part of this problem suggests that it is normal to have uncared for diabetes (A1C = ~8.6; this is not good).

Aside from that, this problem shows the relation between histogram and pdf of a given data set.

Good luck