CS 760 Homework 1 by Cheng-Wei Lu

Problem 1.1

Suppose there are two vectors $u, v \in R^D$, and $a, b \in R$.

$$u = \begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_D \end{pmatrix}, v = \begin{pmatrix} y_1 \\ . \\ . \\ . \\ y_D \end{pmatrix} .x_n, y_n \in R, \forall n = 1...D.$$

$$au + bv = \begin{pmatrix} ax_1 + by_1 \\ . \\ . \\ . \\ ax_D + by_D \end{pmatrix} \in R^D \subseteq R^D$$

Q.E.D

Problem 1.2

(a)

Suppose $u \in R^D$,

$$u = \begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_D \end{pmatrix}, x_n \in R, \forall n = 1...D.$$

Suppose $x_1 < 0$, element-wise square root of u

$$= \begin{pmatrix} \sqrt{x_1} \\ . \\ . \\ . \\ \sqrt{x_D} \end{pmatrix} \notin R^D, \text{since } \sqrt{x_1} \text{ is not a real number}$$

Q.E.D

(b)

0^D is closed under element-wise square root, and is a subspace for R^D. Suppose $u \in 0^D$,

$$u = \begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_D \end{pmatrix} = \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \end{pmatrix},$$

element-wise square root of u

$$= \begin{pmatrix} \sqrt{0} \\ . \\ . \\ . \\ \sqrt{0} \end{pmatrix} = \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} \in 0^D,$$

and if there is another $v \in 0^D$,

$$au + bv = a \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} \in 0^D.$$

Q.E.D

Problem 1.3

Suppose there are two vectors $w, v \in U$,

$$w = \sum_{i=1}^{R} a_i u_i, \text{and } v = \sum_{i=1}^{R} b_i u_i, \text{where } a_i, b_i \in R, \forall i, i = 1...R.$$

$$xw + yv = \sum_{i=1}^{R} (xa_i + yb_i) u_i \in U, \text{where } x, y_i \in R.$$

Q.E.D

Problem 1.4

(a)

$$P(diabetes|inactive) = \frac{P(inactive|diabetes) \times P(diabetes)}{P(inactive)} = \frac{9.3\% \times 95\%}{P(inactive)}$$

(b)

I still need to know the portion of the population who has the 3 genes inactive.

(c)

If the portion of the population who has the 3 genes inactive is low, then I should be concerned.

Problem 1.5

I think the distribution can be the right half of normal distribution times 2, where $\mu = t_0$ and $\sigma = \theta$.

$$X \sim N(\mu = t_0, \sigma = \theta)$$

To be more specific,

$$P(x|\theta) = \begin{cases} \frac{2}{\theta\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-t_0}{\theta}\right)} & , x \geq t_0, \\ 0 & , x < t_0. \end{cases}$$

Note that $\int_{t_0}^{\infty} \frac{2}{\theta\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-t_0}{\theta}\right)} \, dx = 1$. I choose normal distribution because, as described in the problem, the probability of delay time is strictly decreasing as delay time grows, which is also the case of the right half of a normal distribution.

hw1

September 24, 2020

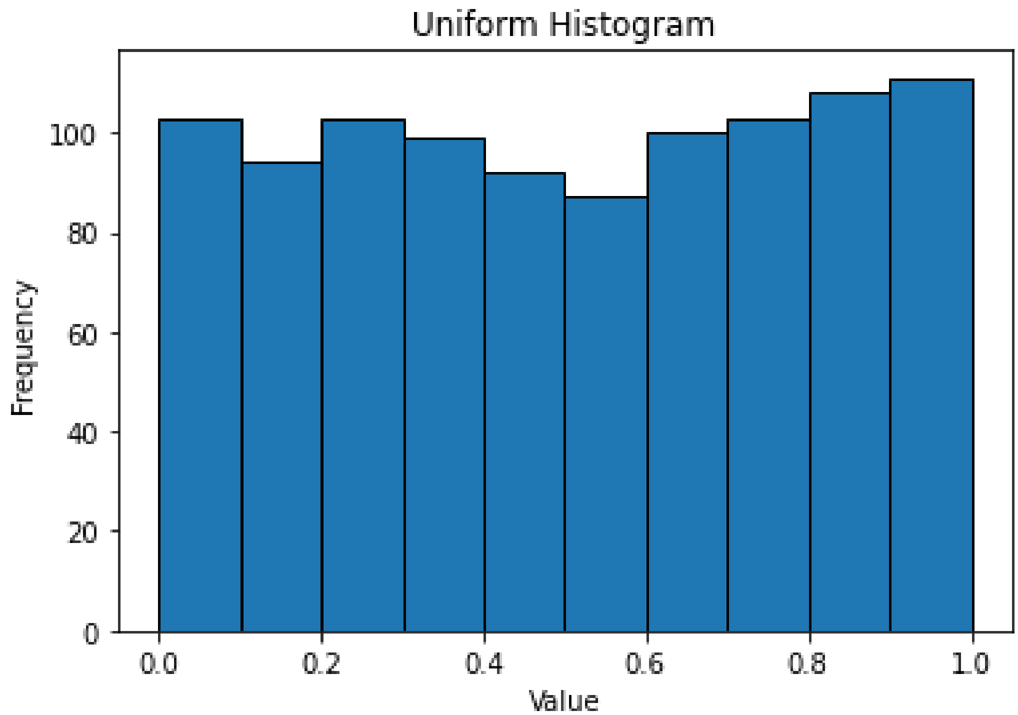
[3]:
```
import math
import matplotlib.pyplot as plt
import random
```

Problem 1.6

(a) Yes, it looks quit uniform to me

[12]:
```
#(a)

N = 1000
histogram_interval = 0.1

sample = []
interval_num = math.floor(1/histogram_interval)
interval_count = [0] * interval_num

for i in range(N):
    temp = random.random()
    sample.append(temp)
    interval_count[math.floor(temp/histogram_interval)] += 1
```

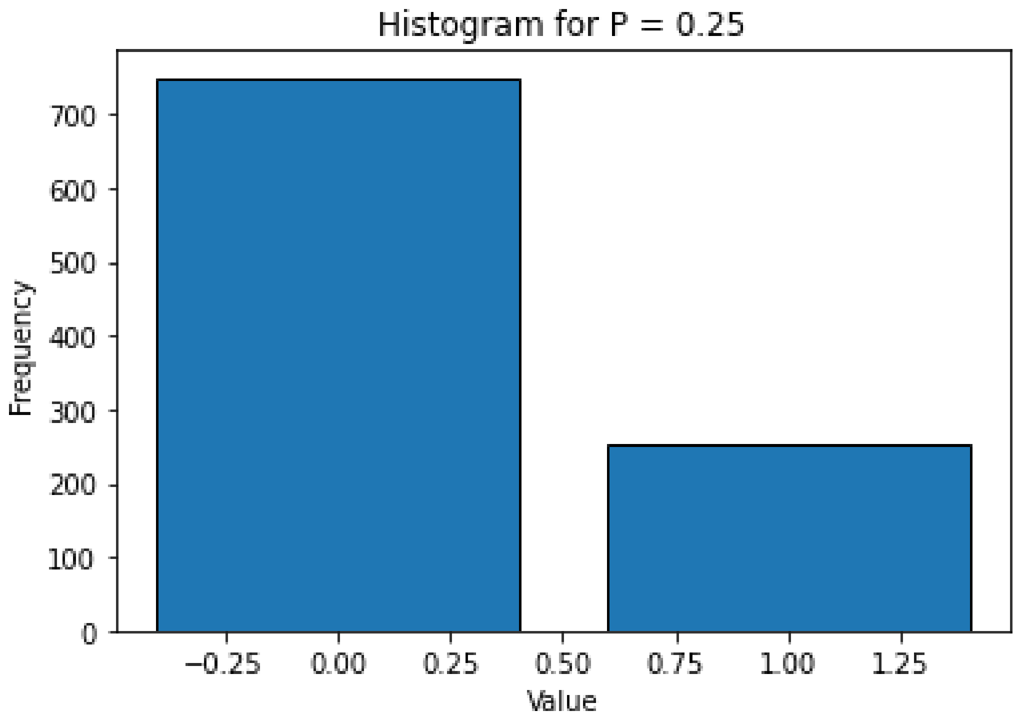
[13]:
```
plt.hist(sample,edgecolor = 'black')
plt.title("Uniform Histogram")
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
```

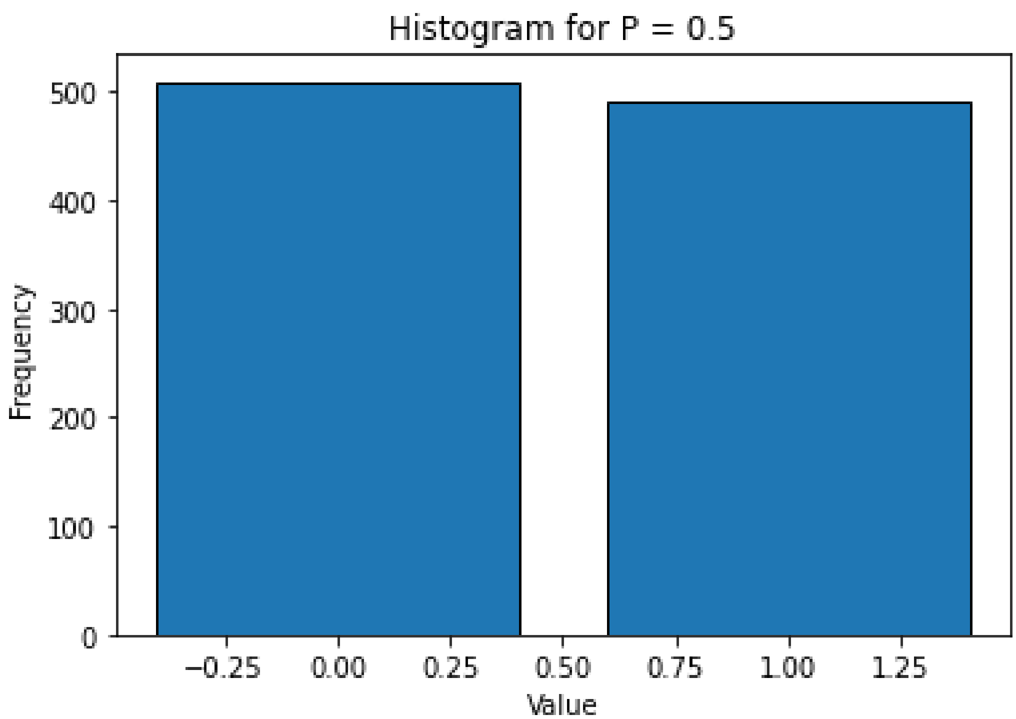


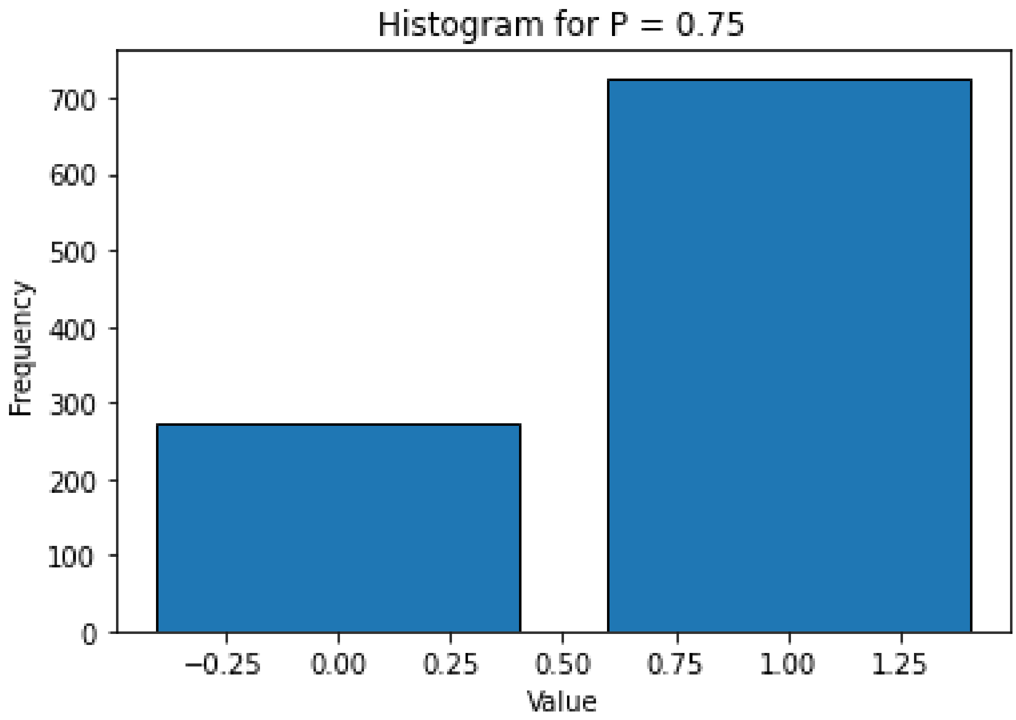
(b) I think it would be a bernoulli distribution where p of the samples would be 1, and (1-p) of them will be 0.

(c) Yes, they match the distribution of my answer.

[15]:
```
#(c)

P = [1/4,1/2,3/4]
for p in P:
    temp = []
    for i in range(N):
        if sample[i] <= p:
            temp.append(1)
        else:
            temp.append(0)
    plt.bar([0,1],[N-sum(temp),sum(temp)],edgecolor = 'black')
    plt.title("Histogram for P = "+str(p))
    plt.xlabel("Value")
    plt.ylabel("Frequency")
    plt.show()
```


Histogram for P = 0.25
Frequency
Value
0
100
200
300
400
500
600
700
−0.25
0.00
0.25
0.50
0.75
1.00
1.25


Histogram for P = 0.5
Frequency
Value
0
100
200
300
400
500
−0.25
0.00
0.25
0.50
0.75
1.00
1.25

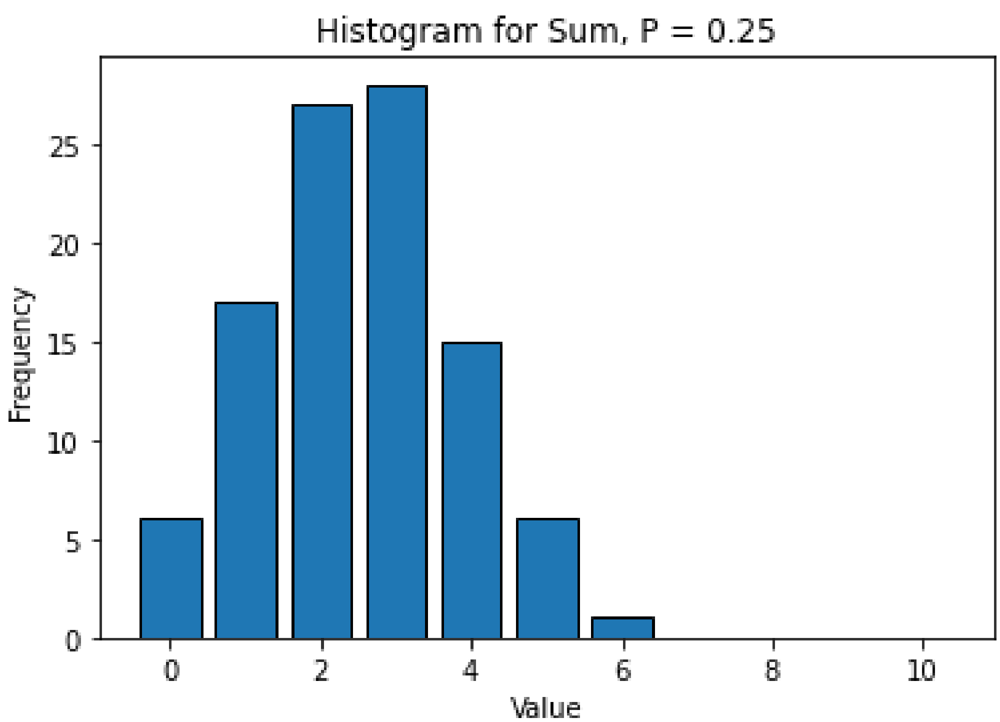


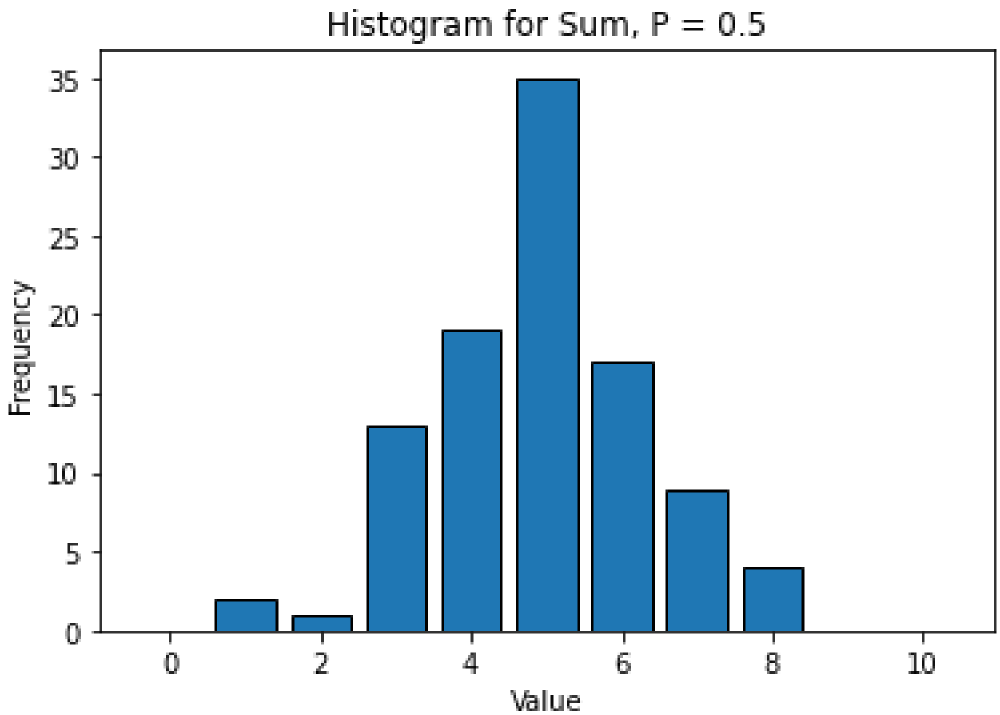
(d) It would be a Binomial distribution, where the sum will be centered at lower number if p is smaller, and centered at higher number if p is higher.

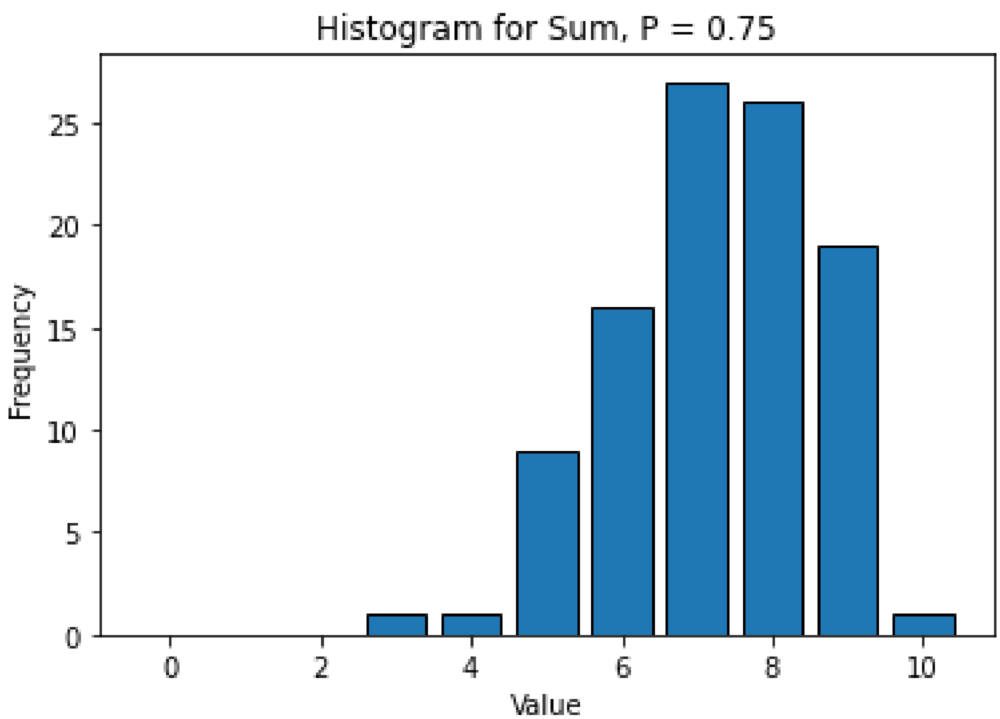
(e) Yes, they match the distribution of my answer.

[16]:
```
#(e)

batch_size = 10

batch_num = math.ceil(N/batch_size)

sum_distribution = []


P = [1/4,1/2,3/4]
for p in P:
    sum_distribution = []
    temp = []
    for i in range(N):
        if sample[i] <= p:
            temp.append(1)
        else:
            temp.append(0)
```

```
for k in range(batch_num):
    batch = temp[batch_size*k:batch_size*(k+1)]
    sum_distribution.append(sum(batch))

#x_axis= [0,1,2,3,4,5,6,7,8,9,10]
x_axis= list(range(0,batch_size+1))
bar_dist = [0]*(batch_size+1)
#bar_dist = [0]*11
for i in sum_distribution:
    bar_dist[i] += 1

plt.bar(x_axis,bar_dist,edgecolor = 'black')
plt.title("Histogram for Sum, P = "+str(p))
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
plt.clf()
```




Histogram for Sum, P = 0.5
Frequency
Value
0
5
10
15
20
25
30
35
0
2
4
6
8
10


Histogram for Sum, P = 0.75
Frequency
Value
0
5
10
15
20
25
0
2
4
6
8
10

<Figure size 432x288 with 0 Axes>

Problem 1.7

(a)

$$l(\theta) = \sum_{i=1}^{N} y_i log((\frac{1}{1+e^{-\theta^T x_i}})) + (1-y_i)log(1-(\frac{1}{1+e^{-\theta^T x_i}}))$$

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial \theta^T} = \sum_{i=1}^{N} y_i(1+e^{-\theta^T x_i}) \times \frac{e^{-\theta^T x_i} \times x_i}{(1+e^{-\theta^T x_i})^2} + (1-y_i) \times \frac{1+e^{-\theta^T x_i}}{e^{-\theta^T x_i}} \\
\times \frac{e^{-\theta^T x_i} \times (-x_i)(1+e^{-\theta^T x_i}) - e^{-\theta^T x_i} \times (-x_i) \times e^{-\theta^T x_i}}{(1+e^{-\theta^T x_i})^2} \\
= \sum_{i=1}^{N} y_i \times \frac{e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}} x_i + (y_i - 1)\frac{1}{1+e^{-\theta^T x_i}} x_i \\
= \sum_{i=1}^{N} y_i \times x_i - \frac{1}{e^{-\theta^T x_i}} x_i \\
= \underline{\underline{\sum_{i=1}^{N} x_i(y_i - \frac{1}{e^{-\theta^T x_i}})}}
\end{aligned}$$

(b)

$$\frac{\partial l(\theta)}{\partial \theta \partial \theta^T} = \sum_{i=1}^{N} -x_i(\frac{-e^{-\theta^T x_i} \times (-x_i^T) \times 1}{(1+e^{-\theta^T x_i})^2}) = \underline{\underline{\sum_{i=1}^{N} \frac{-e^{-\theta^T x_i}}{(1+e^{-\theta^T x_i})^2} x_i x_i^T}}$$

(c)

$l(\theta)$ is a scalar, gradient is a vector, and Hessian is a matrix.