MATHEMATICS FOR DATA SCIENCE

UMA028

LABORATORY FILE

B.E. Electrical and Computer Engineering

(3rd Year)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Submitted To: Mrs. Sumitra

Submitted By: Aditi Goyal

(102169013)

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

PATIALA, PUNJAB

(July 2024 - December 2024)

INDEX

Sr. No.	Experiment	Date	Remarks	Signature
1.	Lab Experiment – 1	31.07.2024		
2.	Lab Experiment – 2	07.08.2024		
3.	Lab Experiment – 3	21.08.2024		
4.	Lab Experiment – 4	28.08.2024		
5.	Lab Experiment – 5	11.09.2024		
6.	Lab Experiment – 6	09.10.2024		
7.	Lab Experiment – 7	16.10.2024		
8.	Lab Experiment – 8	16.10.2024		
9.	Lab Experiment – 9	06.11.2024		
10.	Lab Experiment – 10	13.11.2024		

EXPERIMENT – 1

1. Birthday Paradox: Suppose there are n people in a room. Assume that $n < 365$ and that the people are unrelated in any way. Find the probability of the event A that at least 2 people have the same birthday. (Assume a non-leap year). Write a MATLAB code to find n such that the probability of A is 0.99.

Code:

```
clc;

clear all;

prod=1;

k=0;

prob=0;

while(prob<0.97)

   prod=prod*((365-k)/365);

   k=k+1;

   prob=1-prod;

end

fprintf("no. of persons required for 0.99 chances are %d",k)
```

Output:

```
Command Window
fx no. of persons required for 0.99 chances are 50>>
```

2. Suppose there are n people in a room. Assume that $n < 365$ and that the people are unrelated in any way. Find the probability of the event A that atleast 1 person shares your birthday. (Assume a non-leap year). Write a MATLAB code to find n such that the probability of A is atleast 0.5 and 0.99.

Code:

```
clc;

clear all;

sum=0;

mybirth=4;
```

```
for val=1:3000
    a=randsample(365,253,true);
    A=find(a==4);
    if numel(A)>=1
        c=1;
    else
        c=0;
    end
sum=sum+c;
end
probability=sum/3000;
fprintf("result is:")
display(probability);
```

Output:

```
Command Window

probability =

    0.4933
```

EXPERIMENT – 2

1. Coupon's Collector Problem: Each packet of an injurious product is equally likely to contain any one of n(= 365) different types of coupons. Write a MATLAB code to find the minimum expected number of packets to buy in order to collect all the coupons.

Code:

```
n = 365;

expected_packets = 0;

for i = 1:n

    expected_packets = expected_packets + (1 / i);

end

expected_packets = expected_packets * n;

fprintf('Expected number of packets to buy to collect all %d coupons is: %.2f\n', n, expected_packets);
```

Output:

```
Command Window
>> Ass2Ques1
Expected number of packets to buy to collect all 365 coupons is: 2364.65
fx >>
```

2. The following game is played. The player randomly draws from the set integers {1,2,3,...20}. Let x denote the number drawn. Next the player draws at random from the set {x,.....25}. If on the second draw, he draws a number greater than 21 he wins; otherwise, he loses. Find The probability that the player wins.

Code:

```
clc;

clear all;

sum=0;

for simul=1:1500

    a=randsample(20,1,true);

    b=randsample(a:25,1,true);

    if(b>21)

        sum=sum+1;
```

```
    end
end
prob=sum/15000;
fprintf("Probability of A wins %f\n",prob)
```

Output:

```
Command Window
  Probability of A wins 0.030400
fx >>
```

3. Person Atosses a coin and then person B rolls a die. This is repeated independently UNTIL a head or one of the numbers 1,2,3,4 appears, at which time the game is stopped. Person A wins with the head and B wins with one of the numbers 1,2,3,4. Simulate the game by a MATLAB code and compute the probability that

a. A wins the game if it starts.

Code:

```
clc;
simul=10000;
winA=0;
winB=0;
for i= 1:10000
    Game=true;
    while(Game)
        a=randi([0,1]);
        if(a==1)
            winA=winA+1;
            Game=false;
        else
          b=randi([1,6]);
          if(b<=4)
               Game=false;
            end
```

```
        end
    end
end

probA=winA/simul;
disp(probA);
```

Output:

```
Command Window
  Probability =     0.5919
```

b. B wins the game, if it starts.

Code:

```
clc;
simul=10000;
winA=0;
winB=0;
for i= 1:10000
    Game=true;
    while(Game)
       b=randi([1,6]);
           if(b<=4)
           winB=winB+1;
           Game=false;
        else
            a=randi([0,1]);
            if(a==1)
               Game=false;
           end
        end
```

```
    end
end

probB=winB/simul;
fprintf('Probability =');
disp(probB);
```

Output:

```
Command Window
Probability =    0.8018
```

EXPERIMENT – 3

1. Gambler's Ruin: You enter a casino with Rs k, and on each spin of a roulette wheel you bet Re 1 at evens on the event R that the result in RED. The wheel is FAIR i.e. P(RED)=1/2. If you lose all k, you leave, and if you ever possess N ≥ k, you leave immediately. What is the probability that you leave with nothing? (Assume spins of the wheel are independent). Verify the theoretical probability with the help of a MATLAB simulation of the problem.

Code:

```
clc;
simul= 10000;
fav=0;
for i= 1:simul
   money= 30;
   N=50;
   while(money>0 && money<N)
      a=randi([0:1],1);
      if a==1
         money= money+1;
      else
         money=money-1;
      end
   end
   if money==0
      fav=fav+1;
   end
end
prob=fav/simul;
fprintf('Probability =');
disp(prob);
```

Output:

```
Command Window
Probability =    0.3971
```

2. Monty Hall Problem: Suppose yourself to be participating in the following con test. You have a choice of three doors. Behind one door is a costly gift, behind the other two doors are worthless items (goats). You choose the FIRST door, where upon the host MONTY HALL who knows where the prize is, opens the THIRD door to reveal a goat. He then offers you the opportunity to change your choice of door. Would you do this change? i.e., Given the sequence of events, simulate the contest on MATLAB to find the Probability that the gift lies behind the SECOND door?

Code:

```
clc;
simul=10000;
fav=0;
for i=1:simul
  a=zeros(1,3);
  a(randsample(3,1))=1;
  door=randsample(3,1);
  prize=find(a==1);
  v=setdiff([1 2 3],[door,prize]);
  if numel(v)==1
    revealed=v;
  else
    revealed=randsample(v,1);
  end
  decision=1;
  if(decision==1)
    latestdoor=setdiff([1 2 3],[door,revealed]);
    if a(latestdoor)
      fav=fav+1;
  end
```

```
    end
end
fprintf('Probability =');
prob=fav/simul
```

Output:

```
Command Window
Probability =
prob =

    0.6671
```

3. Polya's Urn: An urn contains b blue balls and c cyan balls. A ball is drawn at random, its colour is noted and it is returned to the urn with d further balls of the same colour. This process is repeated twice. Simulate this experiment on MATLAB to find

a. P(Second Ball is Cyan)

Code:

```
clc;
simul=10000;
h=0;
j=0;
for i=1:simul
    b=5;
    c=5;
    d=20;
    a=randi([0,1]);
    if(a==0)
        b=b+d;
    else
        c=c+d;
    end
    g=randi([0,1]);
    if(g==0)
```

```
        h=h+1;
    else
        j=j+1;
    end
end
prob=j/simul;
fprintf('Probability =');
display(prob);
```

Output:

```
Command Window
Probability =
prob =

    0.4962
```

b. P(First is cyan |Second is Cyan)

Code:

```
clc;
simul=10000;
b=5;
c=5;
d=20;
fav=0;
fav1=0;
for i=1:simul
    urn=[repmat('B',1,b) repmat('C',1,c)];
    firstballindex=randi(length(urn));
    firstballcolor=urn(firstballindex);
    urn=[urn repmat(firstballcolor,1,d)];
    secondballindex=randi(length(urn));
```

```
    secondballcolor=urn(secondballindex);

    if secondballcolor=='C'
        fav=fav+1;
        if firstballcolor=='C'
            fav1=fav1+1;
        end
    end
end

probsecondcyan=fav/simul;
prob=fav1/fav;

fprintf('Probability of Second Cyan Ball =');
disp(probsecondcyan);
fprintf('Probability of First Cyan Ball =');
disp(prob);
```

Output:

```
Command Window
Probability of Second Cyan Ball =    0.4985

Probability of First Cyan Ball =    0.8325
```

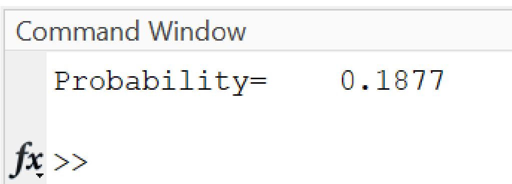
EXPERIMENT – 4

1. Chips: Five cards are dealt at random without replacement from a deck of 52 cards. A be the event that A is the event of atleast 4 spades in hand, B is the event of all spades in hand. Find P(A|B) using conditional probabaility.

Code:

```
clc;
sim=85000;
fav=0;
sample=[1:52];
sample=sample(randperm(52));
spades=1:13;
chosenspades=randsample(sapdes,4,false);
leftfav=setdiff(spades,chosenspades);
leftcards=setdiff(sample,chosenspades);
for i=1:sim
   pick=randsample(leftcards,1);
   if numel(intersect(pick, leftfav))==1
      fav=fav+1;
   end
end
prob=fav/sim;
disp(prob)
prob=prob/4;
```

Output:

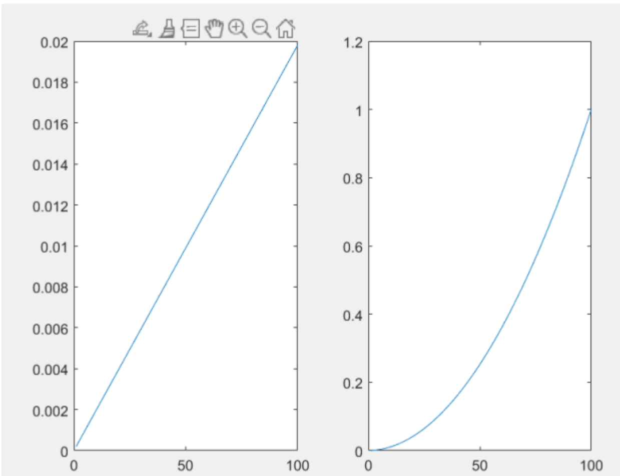


2. Slips: Consider an urn that contains slips of paper each with one of the numbers 1, 2,..., 100 on it. Suppose there are i slips with the number i on it for i = 1,2,...,100. For example, there are 25 slips of paper with the number 25. Assume that the slips are identical except for the numbers. Suppose one slip is drawn at random. Let X be the number on the slip. Find PMF and CDF of X.

Code:

```
clc;
x_vec = [1:100];
pmf= x_vec/5050;
cdf= cumsum(pmf);
subplot(1,2,1);
plot(pmf);
subplot(1,2,2);
plot(cdf);
```

Output:



3. Let X be a random variable with pdf

$$f(x) = 2x, 0 < x<1$$

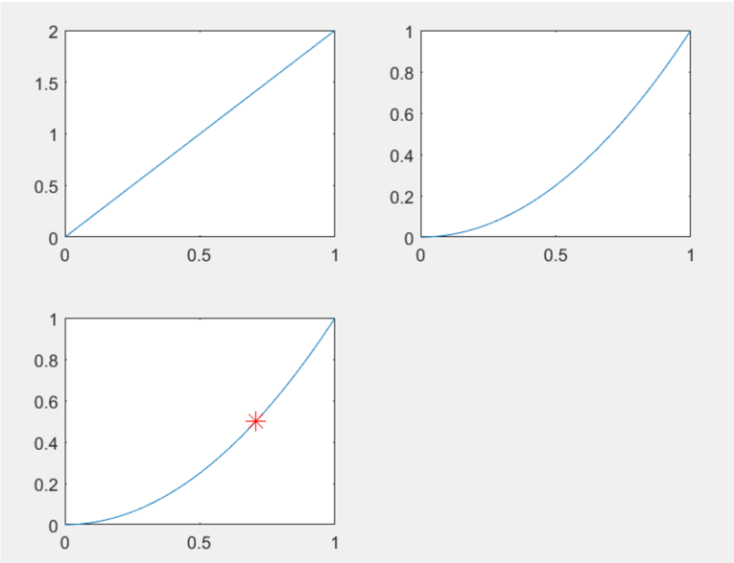
Plot pdf and cdf. Find 0.5 quantiles of the cdf of X.

Code:

```
clc;
n = 100;
xvec = linspace(0, 1, n);
```

```
pdf = 2 * xvec;
cdf = cumtrapz(xvec, pdf);
cdf = cdf / max(cdf);

subplot(2, 2, 1);
plot(xvec, pdf);

subplot(2, 2, 2);
plot(xvec, cdf);

[u, ind] = ismembertol(0.5, cdf, 0.01);
med = xvec(ind);

subplot(2, 2, 3);
plot(xvec, cdf);
hold on;
plot(med, cdf(ind), 'r*', 'MarkerSize', 10);
```

Output:



EXPERIMENT – 5

1. Flight Passengers: A flight company knows that 5% of the people buying tickets do not end up boarding the flight, hence it sells 52 tickets for a 50-seater flight. What is the probability that every person who turns up will get a seat. Simulate this situation on Matlab.

Code:

```
clc;
clear all;


s=0;
for i=1:5000
    k=0;
    for j=1:52
        r=randsample([1:100],1);
        if(r<=95)
            k=k+1;
        end
    end
    if(k<=50)
        s=s+1;
    end
end
prob=s/5000
```

Output:

```
Command Window
Probability:
prob =

    0.7338
```

2. Suppose we roll a fair six-sided die 3 times.
   (a) What is the probability of getting exactly 2 sixes?
   (b) X be the random variable that counts the number of sixes. Calculate μX.
   (c) Simulate this game and verify (a) and (b)

Code:

```
clc;
s=0;
n=3;
sum=0;
p=1/6;
for i=1:5000
    r=binornd(n,p,1);
    sum=sum+r;
    if(r==2)
        s=s+1;
    end
end
prob=s/5000;
mean=n*p;
aver=sum/5000;
disp(aver)
disp(mean)
disp(prob)
```

Output:

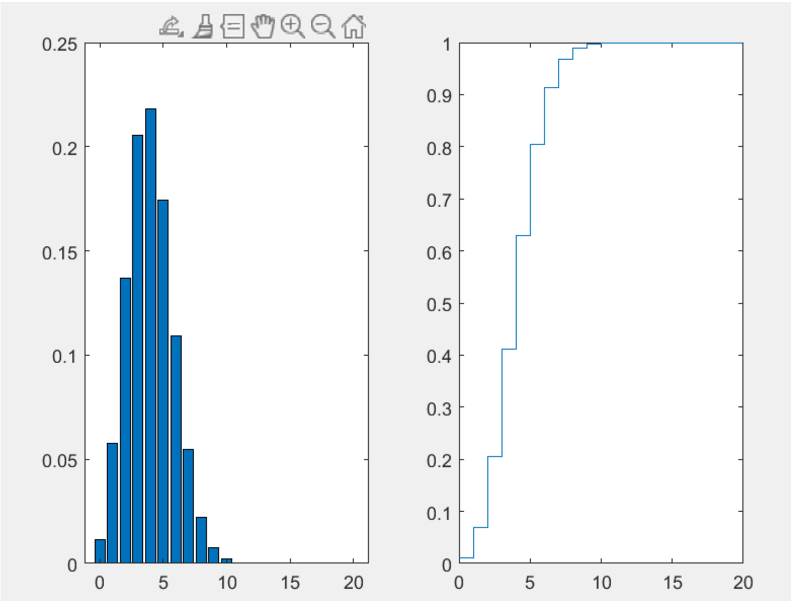
```
Command Window
Average:     0.4994

Mean:     0.5000

Probability:     0.0772
```

3. Plot the c.d.f and p.m.f (bar plot) of Binomial distribution with n = 20,p = 0.2,0.5,0.7. What is the difference in the p.m.f. plots indicating?

Code:

```
clc;
n=20;
p=0.2;
x=0:20;
subplot(1,2,1);
pmf=binopdf(x,n,p);
bar(x,pmf)
subplot(1,2,2);
cdf=binocdf(x,n,p);
stairs(x,cdf);
```

Output:



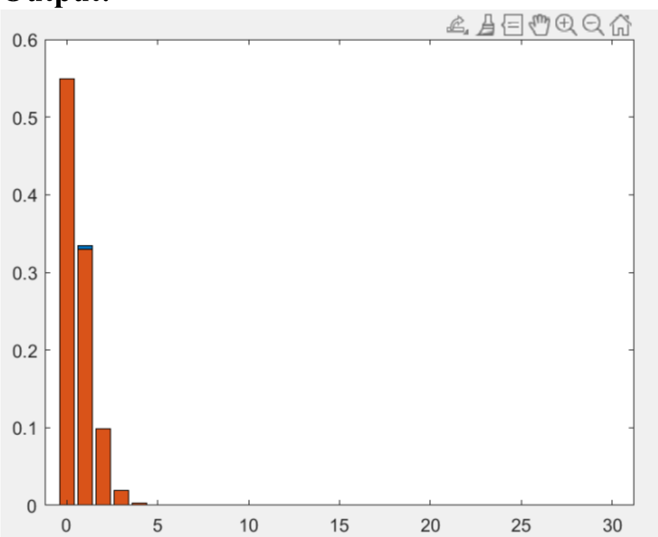
EXPERIMENT – 6

1. Poisson Approximation to Binomial:
   (a) Plot the Binomial p.m.f for n = 30,p = 0.02.
   (b) Plot the Poisson p.m.f for λ = 0.6. Analysing the pm.f, What is the Mode of the distribution.
   (c) What similarity do you see in their plot? Verify with the theoretical result. What deductions can you make on values of n.

Code:

```
clc;
n=30;
l=0.6;
p=0.02;
X= 0:n
pmf=binopdf(X,n,p);


bar(X,pmf);
hold on


pmf1=poisspdf(X,l);
bar(X,pmf1);
```

Output:



2. Generate 10,30,50,60 random numbers for the Poi(λ) distribution with λ = 2. What is the mean of these numbers in each case. What deduction can you make from this sample about λ.

Code:

```
clc;
l=2;
sample=[10 30 50 60];
means= zeros(1, length(sample));
for i= 1: length(sample)
    sample_size= sample(i);
    random_num=poissrnd(l, 1, sample_size);
    mean_val= mean(random_num);
    means(i)= mean_val;
end
means
```

Output:

Command Window

```
means =

    2.5000    2.1000    1.8200    1.8833
```

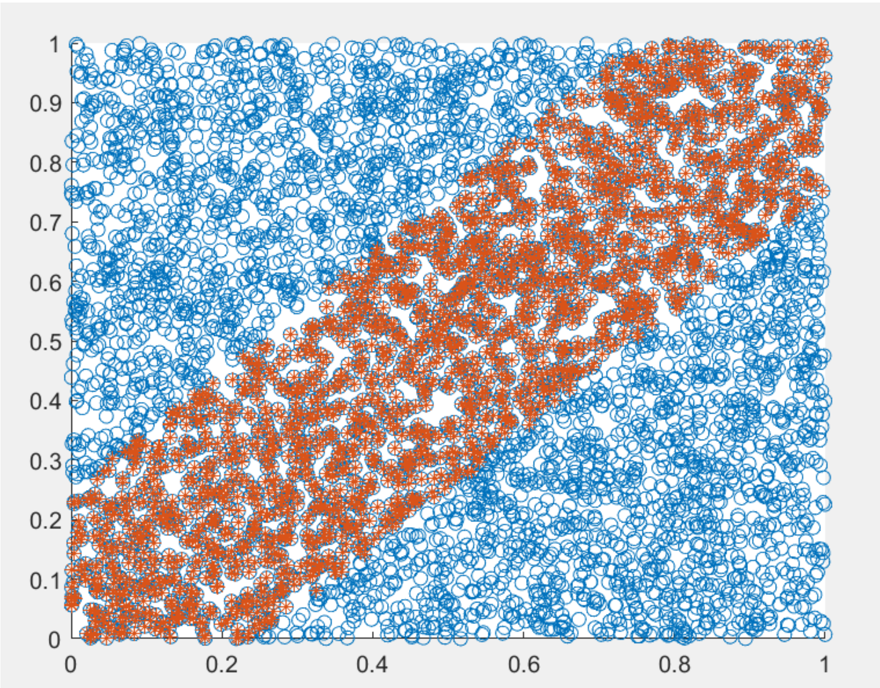
EXPERIMENT – 7

1. A and B are friends. They decide to meet between 1 PM and 2 PM on a given day. There is a condition that whoever arrives first will not wait for the other for more than 15 minutes. Find the probability that they will meet on that day. Simulate this situation on MATLAB.

Code:

```
clc;
a= rand(1,5000);
b= rand(1,5000);
fav= find(abs(a-b)<0.25);
prob= fav/5000;
figure
hold on;
scatter(a,b);
scatter(a(fav),b(fav), '*');
```

Output:



2. Find the Spearman rank correlation for the data set

Person	A	B	C	D	E	F
Math Marks	110	100	140	120	80	90
Hindi Marks	70	20	10	65	60	80

Code:

```
clc;

a=[110 100 140 120 80 90];

b=[70 20 10 65 60 80];

[RHO, PVAL] = corr(a', b', 'type','Spearman');

disp(PVAL);

disp(RHO);
```

Output:

```
Command Window

RHO =

   -0.3714


PVAL =

    0.4972
```

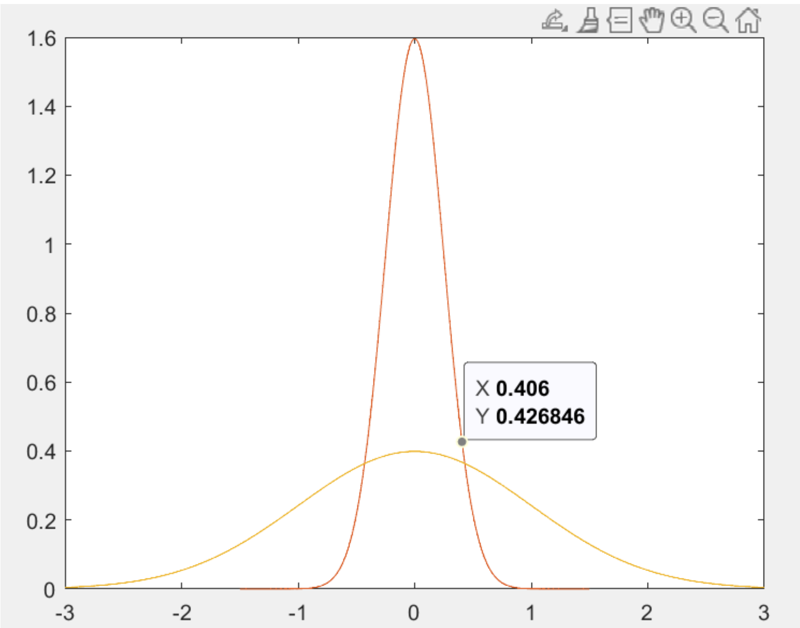
EXPERIMENT – 8

1. Plot the p.d.f. of N(0,0.25),N(0,1) in the range ($\mu-3\sigma,\mu-3\sigma$). Analyse the pdf's with respect to σ.

Code:

```
clc;

sd=0.5;

mean=0;

x=[mean-3*sd: 0.001 :mean+3*sd];

y=normpdf(x, mean,sd^2);

plot(x,y);


hold on

sd1=1;

x=[mean-3*sd1: 0.001 :mean+3*sd1];

y=normpdf(x, mean,sd1^2);

plot(x,y);
```

Output:

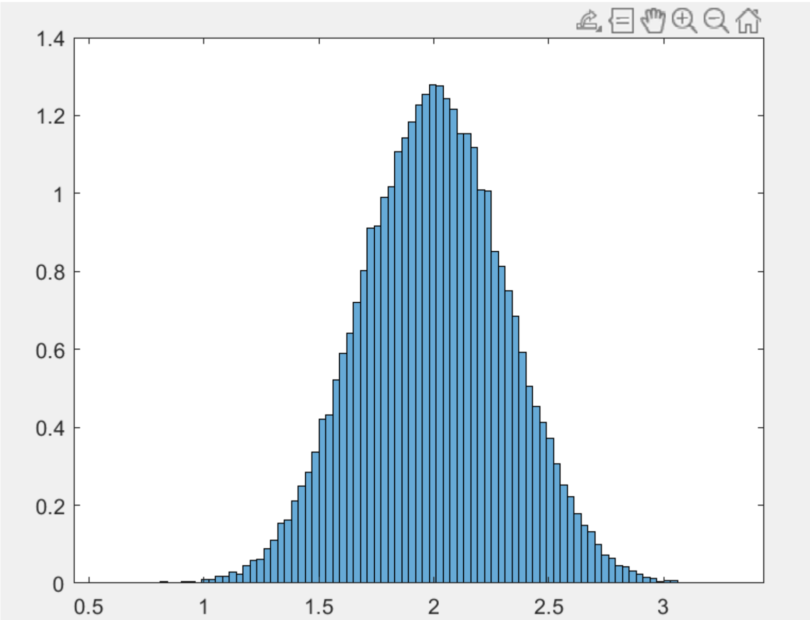


2. Let Xi ~ N(μ,σ2) for i = 1,2,...n. Using Simulation Verify Central Limit Theorem.

Code:

```
clc;
n=10;
for i=1:n
   v(:,i)=normrnd(2,1,[1,50000]);
end
samplesum=sum(v,2);
samplebar=samplesum/n;
var(samplebar);
histogram(samplebar, 'Normalization','pdf');
```

Output:



3. Repeat Question 2 For Xi ~ U(0,1), Xi ~ exp(2), Xi ~ Poiss(2). Are there any changes with respect to n.

Code:

For Uniform Distribution

```
clc;
n=10;
```

```
for i=1:n
    v(:,i)=rand(1,50000);
end
samplesum=sum(v,2);
samplebar=samplesum/n;
var(samplebar);
histogram(samplebar, 'Normalization','pdf');
```

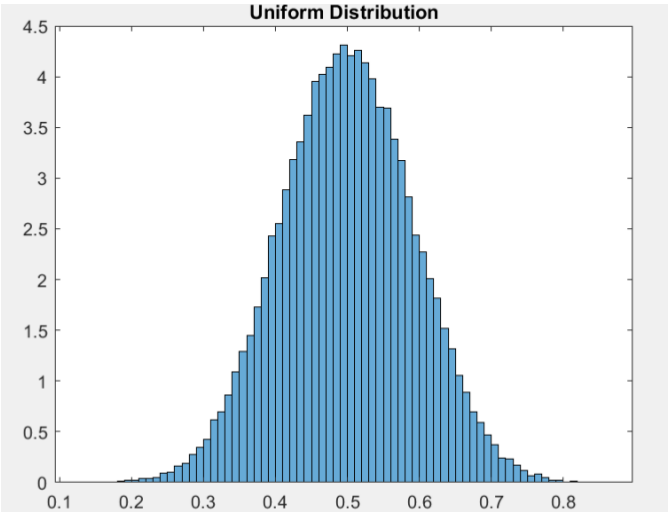
For Exponential Distribution

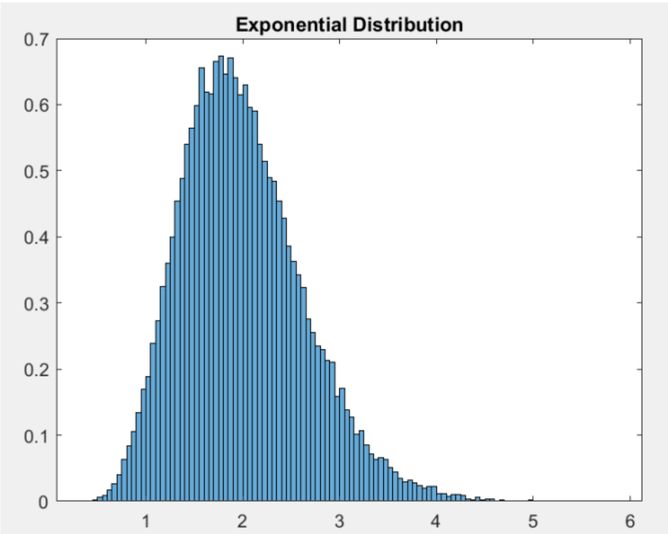
```
clc;
n=10;
for i=1:n
    v(:,i)=exprnd(2,[1,50000]);
end
samplesum=sum(v,2);
samplebar=samplesum/n;
var(samplebar);
histogram(samplebar, 'Normalization','pdf');
```

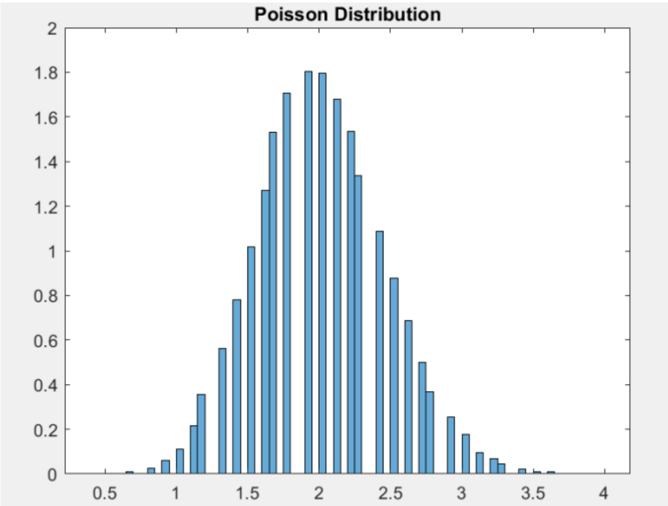
For Poisson Distribution

```
clc;
n=10;
for i=1:n
    v(:,i)=poissrnd(2,[1,50000]);
end
samplesum=sum(v,2);
samplebar=samplesum/n;
var(samplebar);
histogram(samplebar, 'Normalization','pdf');
```

Output:


Uniform Distribution
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8


Exponential Distribution
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6


Poisson Distribution
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.5
1
1.5
2
2.5
3
3.5
4

EXPERIMENT – 9

1. Gambler's Ruin: FortheQuestionin Experiment-3, Compare the theoretical probability obtained with the simulated. Based on 2500,2500 simulations, construct a 95% confidence interval interval for this proportion.

Code:

```
clc;
simul= 10000;
fav=0;
for i= 1:simul
   money= 30;
   N=50;
   while(money>0 && money<N)
      a=randi([0:1],1);
      if a==1
         money= money+1;
      else
         money=money-1;
      end
   end
   if money==0
      fav=fav+1;
   end
end
prob=fav/simul;
fprintf('Probability= ')
disp(prob);


alpha=0.05;
zalphaby2=norminv(0.025);
L=prob + zalphaby2*sqrt(prob*(1-prob)/simul);
```

```
U=prob - zalphaby2*sqrt(prob*(1-prob)/simul);


display(L);

display(U);
```

Output:

```
Command Window
Probability=     0.3965


L =

    0.3869


U =

    0.4061
```

2. For Question 1 in Experiment 7, Compare the theoretical probability obtained with the simulated. Based on 5000,1000 simulations, construct a 95% confidence interval interval for this proportion.

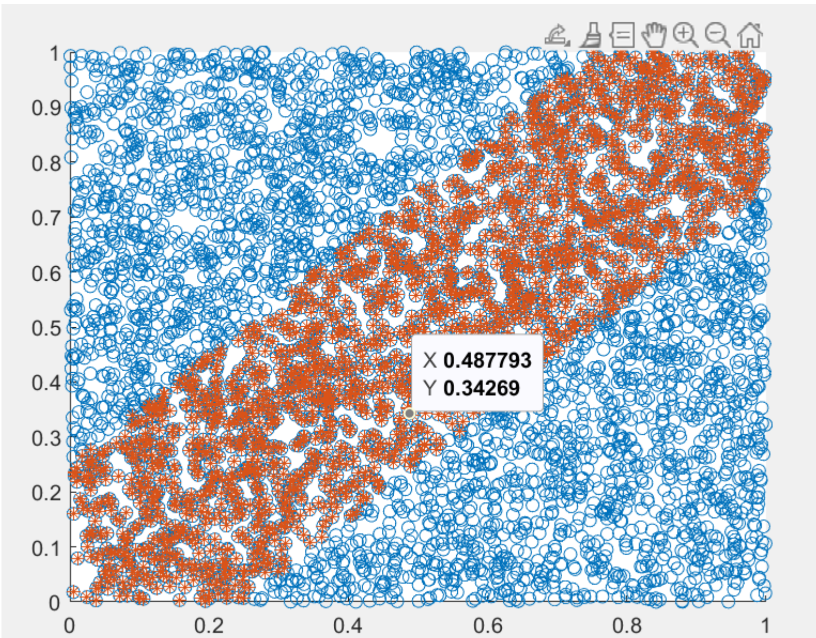
Code:

```
clc;

a= rand(1,5000);

b= rand(1,5000);

fav= find(abs(a-b)<0.25);


figure

hold on;

scatter(a,b);

scatter(a(fav),b(fav), '*');


prob1=numel(fav)/5000;


alpha=0.05;
```

```
zalphaby2=norminv(0.025);
L=prob1 + zalphaby2*sqrt(prob1.*(1-prob1)/5000);
U=prob1 - zalphaby2*sqrt(prob1.*(1-prob1)/5000);

display(L);
display(U);

display(prob1);
```

Output:

```
Command Window

L =

    0.4298

U =

    0.4574

prob1 =

    0.4436
```



3. Calculate the Maximum Likelihood Estimator of θ in U(0,θ) based on a random sample from this distribution.

Code:

```
clc;
x=rand(1,5)
MLE=mle(x,'distribution','unid');
display(MLE);
```

Output:

```
Command Window

x =

    0.4590    0.2199    0.4431    0.2562    0.2355


MLE =

    0.4590
```

4. Let X1,X2,X3,..,Xn be i.i.d from N(μ,σ2). Obtain MLE of μ,σ2, (1−α)100% confidence interval for μ,σ^2.

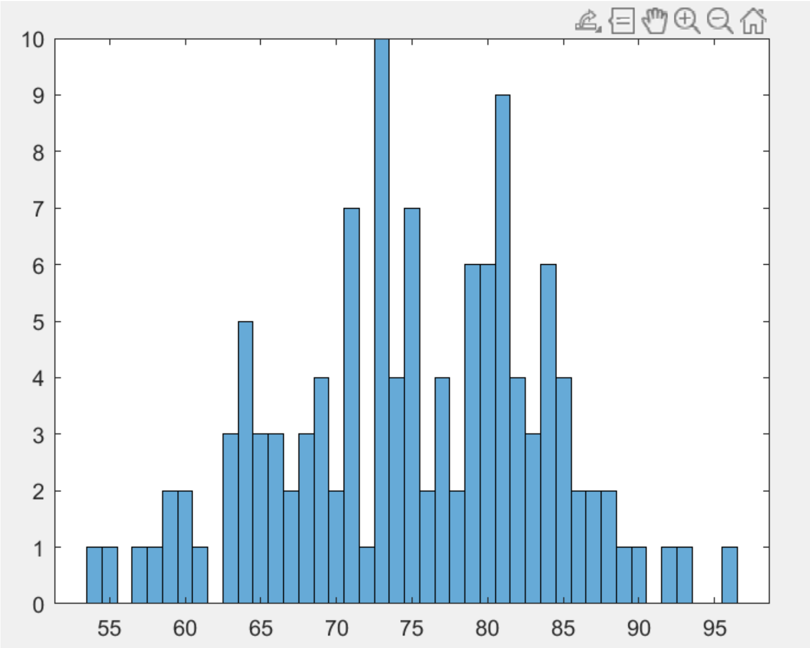
Code:

```
clc;
load examgrades
x=grades(:,1)
histogram(x);
MLE= mle(x,'distribution','normal')
[meanest stdest mci sci] = normfit(x, 0.05)
```

Output:

```
MLE =

   75.0083    8.6838


meanest =

   75.0083
```

```
stdest =

    8.7202


mci =

   73.4321
   76.5846


sci =

    7.7391
    9.9884
```



EXPERIMENT – 10

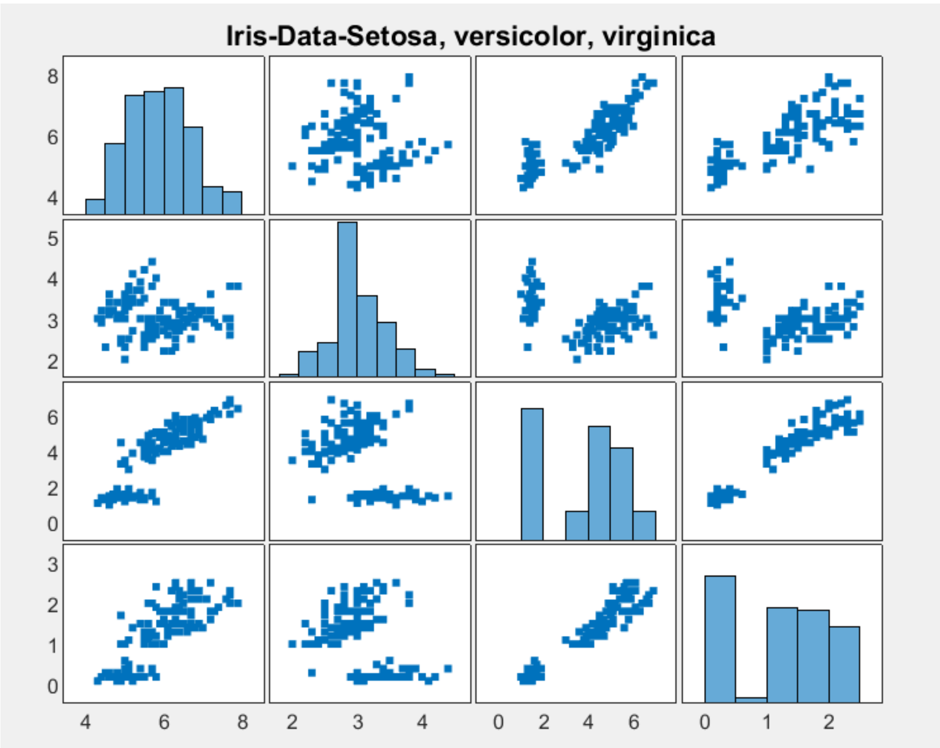
1. Analyse the 'Iris' data set in Matlab with respect to scatterplot, Normality, Correlation, Hypothesis Testing about the mean. What inferences do you draw from the results and plots.

Code:

For Iris Data Plot

```
clc;

load iris_dataset.mat

irisInputs=irisInputs';

irisInputs;


numel(irisInputs)


plotmatrix(irisInputs);

title('Iris-Data-Setosa, versicolor, virginica')
```

Output:



Code:

To find Correlation coefficient matrix

corrcoef(irisInputs)

Output:

```
Command Window
ans =

    1.0000   -0.1094    0.8718    0.8180
   -0.1094    1.0000   -0.4205   -0.3565
    0.8718   -0.4205    1.0000    0.9628
    0.8180   -0.3565    0.9628    1.0000
```

Code:

To extract individual plant matrix and their correlation coefficient

setosa=irisInputs(1:50, 1:4)

versicolor=irisInputs(51:100, 1:4)

verginica=irisInputs(101:150, 1:4)

corrcoef(verginica)

corrcoef(setosa)

corrcoef(versicolor)

Output:

```
setosa
ans =

    1.0000    0.4572    0.8642    0.2811
    0.4572    1.0000    0.4010    0.5377
    0.8642    0.4010    1.0000    0.3221
    0.2811    0.5377    0.3221    1.0000

versicolor
ans =

    1.0000    0.7468    0.2639    0.2791
    0.7468    1.0000    0.1767    0.2800
    0.2639    0.1767    1.0000    0.3063
    0.2791    0.2800    0.3063    1.0000
```

```
virginica
ans =

    1.0000    0.5259    0.7540    0.5465
    0.5259    1.0000    0.5605    0.6640
    0.7540    0.5605    1.0000    0.7867
    0.5465    0.6640    0.7867    1.0000
```

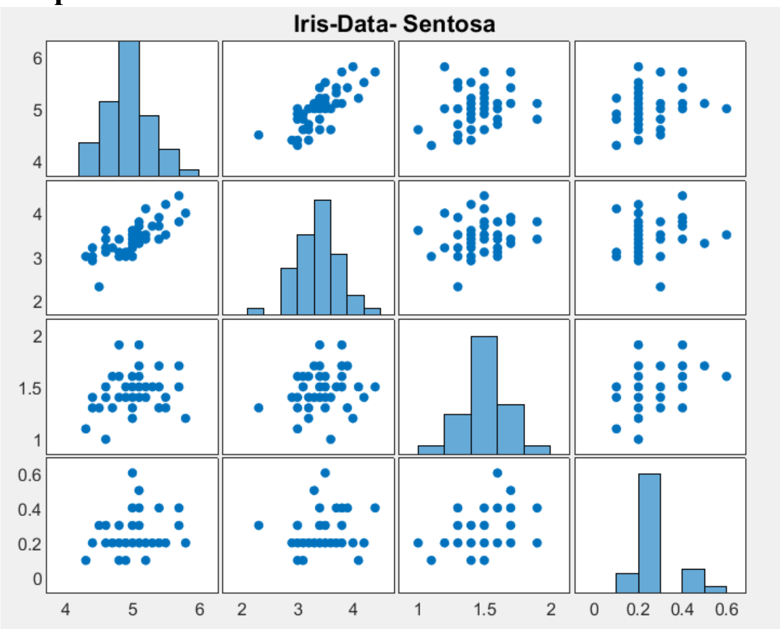
Code:

To plot individual plant matrix plot

plotmatrix(setosa)

title('Iris-Data- Sentosa');

Output:



Code:

To extract 'sepal length' of first feature

setosaSL=setosa(:, 1);

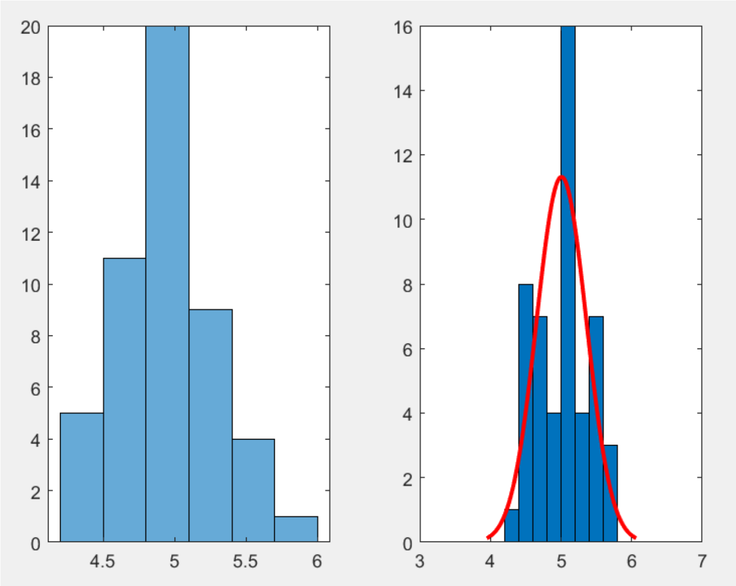
subplot(1,2,1);

histogram(setosaSL);

subplot(1,2,2);

histfit(setosaSL)

Output:



Code:

Hypothesis Test

```
% when H1=H0
[hyp, pv, ci]=ztest(setosaSL, 5, 0.3)

%when h1>h0
[hyp, pv, ci]=ztest(setosaSL, 5, 0.3, 'Tail','right')

%when h1<h0
[hyp, pv, ci]=ztest(setosaSL, 5, 0.3, 'Tail','left')
```

Output:

When H1=H0	When H1>H0	When H1<H0
hyp = 0 pv = 0.8875 ci = 4.9228 5.0892	hyp = 0 pv = 0.4438 ci = 4.9362 Inf	hyp = 0 pv = 0.5562 ci = -Inf 5.0758

Code:

For Normality

```
qqplot(setosa(:,1))
```

Output: