EE 511 -

PROJECT 1 - Probability & Random Nos. Generator

Name - Tamoghna Chattopadhyay
Student ID - 8541324935

1. ## PROBLEM STATEMENT

(1) Let $X \sim U(0,1)$. Evaluate the mean μ and variance σ_X^2.

(2) Generate a sequence of $N = 100$ random numbers between $[0,1]$ and compute the sample mean,

$$m = \sum_{i=1}^{100} \frac{X_i}{N}$$ and sample variance,

$$S^2 = \frac{\sum_{i=1}^{N} (X_i - m)^2}{N-1}$$ and compare to μ and σ^2.

Also, estimate the sample variance of the sample mean. Repeat for $N = 10{,}000$.

(3) The Central limit Theorem says that $m = \frac{\sum_{i=1}^{n} X_i}{n}$

$\longrightarrow N(\mu, \sigma^2/n)$. Repeat the experiment in (2) (for $N = 100$) 50 times to generate a sample of means $\{m_j, j = 1, \cdots 50\}$. Do they appear to be approximately normally distributed values with mean μ and variance σ^2/n ?

(4) We want to check if there are any dependency between X_i and X_{i+1}. Generate a sequence of $N+1$ random numbers that are $\sim U[0,1]$ for $N = 1000$

Compute

$$Z = \left[\frac{\sum_{i=1}^{N} X_i X_{i+1}}{N}\right] - \left[\frac{\sum_{i=1}^{N} X_i}{N}\right]\left[\frac{\sum_{j=2}^{N+1} X_j}{N}\right]$$

Comment on what you expect and what you find.

2. THEORETICAL ANALYSIS

- Sample Mean could be defined as the mean of a random sample of any size (say N). It is given by,

$$\bar{X} = \sum_{i=1}^{N} \frac{X_i}{N}$$

- Variance can be defined as average of the squared differences of the variables from the means. It is given by

$$s^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N-1}$$

- We can say that the sample mean $\bar{X}$ and sample variances s^2 can be considered as the 'best estimates' of the mean μ and variance σ^2.

- Consider a discrete random variable X taking one of the values $x_1, x_2, \ldots$, then the expected value of X or the mean of X is denoted as $E[X]$ and is given by the equation

$$E[X] = \sum_{i} x_i * P[X = x_i]$$

It is the weighted average that all the possible values of X being weighted by the probability mass function of x.

- For a random variable X, the variance is given by,

$$Var(X) = E[(X-\mu)^2] = E[X^2] - E[X]^2$$

- Covariance is the measure of how much two random variwe see,

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

Here $X \sim [0, 1]$

So,

$$f_X(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{otherwise.} \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$

$$= \int_a^b x \cdot \frac{1}{b-a} \, dx$$

$$= \int_0^1 x \cdot 1 \cdot dx$$

$$= \frac{x^2}{2} \Big|_0^1 = \frac{1}{2}$$

(n)

and

$$V(X) = E[(X - E(X))^2]$$

$$= \int_a^b \left(x - \left(\frac{a+b}{2}\right)\right)^2 \cdot \frac{1}{b-a} \, dx$$

let $z = \dfrac{x - (a+b)/2}{b-a}$

$$dz = \frac{dx}{b-a}$$

So, $$V(X) = (b-a)^2 \int_{-1/2}^{1/2} z^2 \, dz$$

$$= (b-a)^2 \left[\frac{z^3}{3}\right]_{-1/2}^{1/2}$$

$$= \frac{(b-a)^2}{12}$$

-ables vary together. It is simply similar to the variance and is given by,

$$\text{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= E[XY] - E[X]E[Y]$$

An expression for variance can also be given by,

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 * \text{Cov}(X,Y)$$

- The Central limit Theorem states that the sum or average of an infinite sequence of independent and identically distributed random variables, when suitably rescaled, tends to a normal distribution.

Thus, $m = \sum_{i=1}^{n} \frac{X_i}{n} \longrightarrow N(\mu, \sigma^2/n)$.

3. SIMULATION METHODOLOGY

The 'rand' function in MATLAB helps in generating a uniformly distributed sequence of random numbers between 0 and 1.

Using the random numbers generated, the sample mean and sample variance were determined using the 'mean' and 'var' commands in MATLAB.

If we look theoretically, then, for

$$X \sim U[0,1]$$

So, here,

$$V(X) = \frac{(1-0)^2}{12} = \frac{1}{12}.$$

So,

$$\boxed{\begin{aligned} E[X] &= \frac{1}{2} = 0.5 \\ V(X) &= \frac{1}{12} = 0.0833 \end{aligned}} \longrightarrow$$ Answer to Part (1) of the project.

Also, for calculating the variance of sample mean of a sequence of random numbers, we can say,

$$Var(\bar{x}) = Var\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)$$

$$= Var\left[\frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n\right]$$

$$= \frac{1}{n^2}Var(x_1) + \frac{1}{n^2}Var(x_2) + \cdots + \frac{1}{n^2}Var(x_n)$$

When x_i are identically distributed, and have same variances σ^2, then,

$$Var(\bar{x}) = \frac{1}{n^2}[\sigma^2 + \sigma^2 + \cdots + \sigma^2]$$

$$= \frac{1}{n^2}[n\sigma^2]$$

$$= \frac{\sigma^2}{n}$$

$$= \frac{Var(X)}{n}.$$

4. EXPERIMENTS AND RESULTS

(1) The mean, μ and variance, σ_X^2 for $X \sim U[0,1]$ are given by,

$$\mu = \frac{1}{2}$$

$$\sigma_X^2 = \frac{1}{12}$$

(2) When we generated $N = 100$ random numbers between $[0,1]$, we get,

mean 1 = 0.5280
var 1 = 0.0882

When we generated $N = 10000$ random numbers between $[0,1]$, we get,

mean 2 = 0.4991
var 2 = 0.0829

In case of $N = 100$, variance of sample mean,

$$\text{Var\_est} = 8.82 \times 10^{-4}$$

In case of $N = 10000$, variance of sample mean,

$$\text{Var\_est} = 8.29 \times 10^{-6}$$

(4) When $N = 1000$ and generating $N+1$ random numbers with $\sim U(0,1)$, we get value of,

$$Z = \left[\frac{\sum_{i=1}^{N} x_i x_{i+1}}{N}\right] - \left[\frac{\sum_{i=1}^{N} x_i}{N}\right]\left[\frac{\sum_{j=2}^{N+1} x_j}{N}\right]$$

$$\Rightarrow \quad Z = 0.0010$$

(5) For the χ^2 Goodness of Fit test

	1	2	3	4	5	6	7	8	9	10
EXPECTED	100	100	100	100	100	100	100	100	100	100
OBSERVED	101	89	112	107	96	86	117	98	115	79

$$\chi^2 = \sum_{i=1}^{10} \frac{(o_i - e_i)^2}{e_i}$$

$$= \frac{(1)^2}{100} + \frac{(11)^2}{100} + \frac{(12)^2}{100} + \frac{(7)^2}{100} + \frac{(4)^2}{100} + \frac{(14)^2}{100}$$

$$+ \frac{(17)^2}{100} + \frac{(2)^2}{100} + \frac{(15)^2}{100} + \frac{(21)^2}{100}$$

$$= \frac{1 + 121 + 144 + 49 + 16 + 196 + 289 + 4 + 225 + 441}{100}$$

$$= 14.86$$

And, degrees of freedom $= 10 - 1 = 9$

From the table of χ^2 values vs p-value, if we compare the values, we can see that, the closest value for χ^2 is 14.68. The corresponding value of Probability,

$$\boxed{\hat{p} \text{ value} = 0.10}$$

When we perform a hypothesis test in statistics, a p-value helps determine the significance of the results. The claim that on's trial, is called the null hypothesis.

In case, p-value > 0.05, indicates weak evidence against the null hypothesis, so we cannot reject the null hypothesis. This is the case here.

- The graphs for part (3) and part (5) are as follows: (after conclusions section).

In case of (5), the different values of random numbers in each bin, we get as:

N ⇒

0 - 0.1	101
0.1 - 0.2	89
0.2 - 0.3	112
0.3 - 0.4	107
0.4 - 0.5	96
0.5 - 0.6	86
0.6 - 0.7	117
0.7 - 0.8	98
0.8 - 0.9	115
0.9 - 0.10	79

5. CONCLUSIONS DRAWN

(2) The values of the means and variance after simulation turned out to be close to the values calculated in (1). As the length of the random sequence increased, the values of mean and variance were found to be closer to the inputed mean value and expected value of variance.

The value of the variance of sample mean was close to σ^2/n.

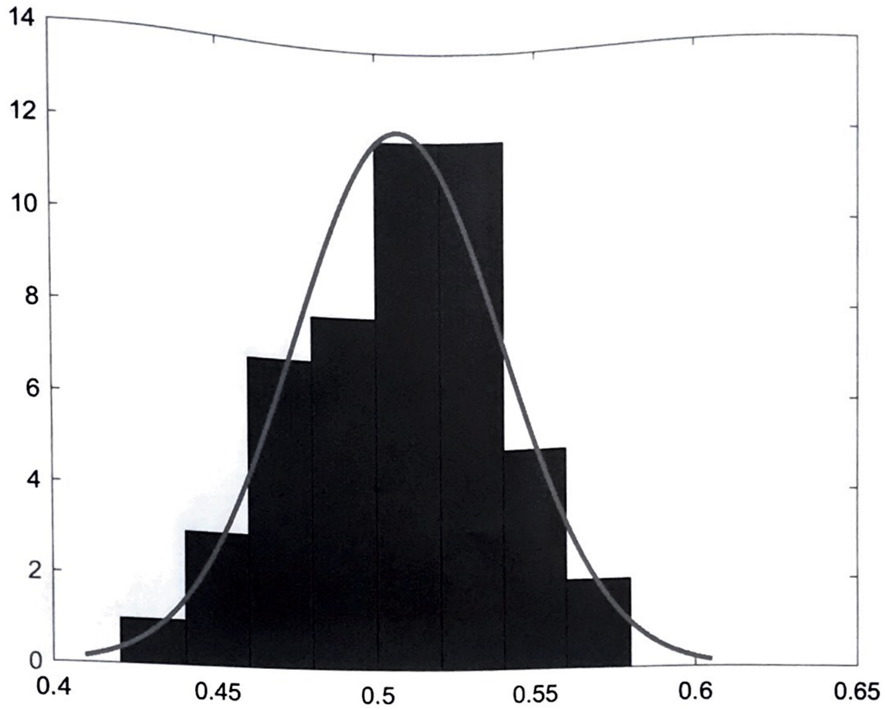
(3) The graph fits a normal distribution $N(\mu, \sigma^2/n)$ when we take 50 samples.

(4) The formula
$$z = \left[\frac{\sum_{i=1}^{N} x_i x_{i+1}}{N}\right] - \left[\frac{\sum_{i=1}^{N} x_i}{N}\right]\left[\frac{\sum_{j=2}^{N+1} x_j}{N}\right]$$
$$= \mathrm{cov}(x_i, x_{i+1})$$

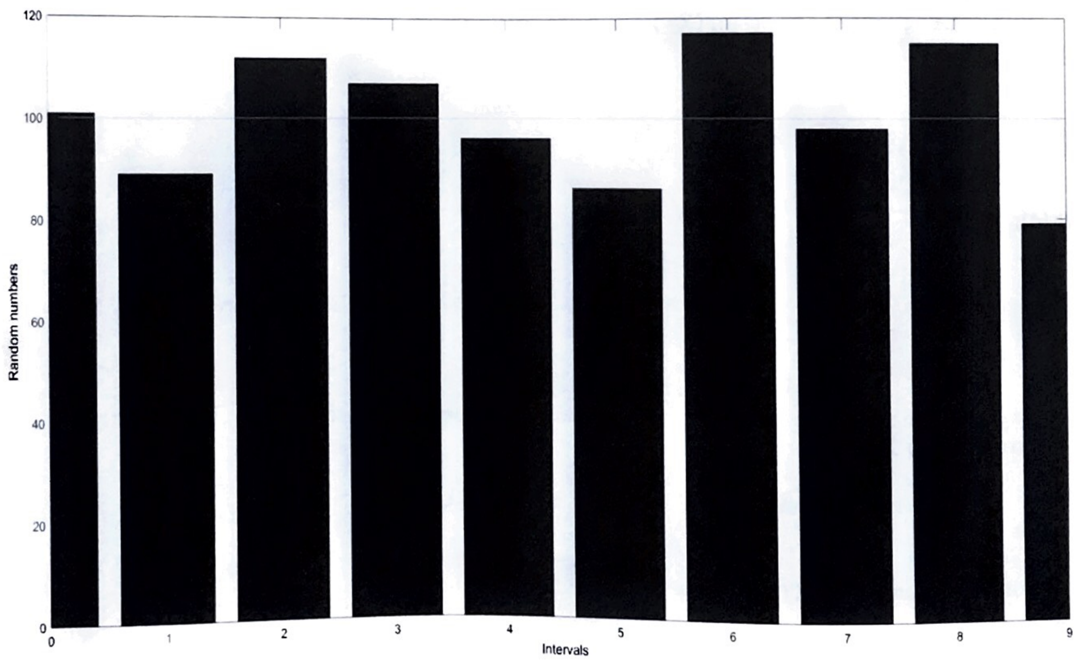
For independent random variables, covariance $= 0$. In our result too, we find value of $z \approx 0$.
So, covariance $= 0$.

(5) For $N = 1000$, the chi-squared values estimates to be around 14.68 for above shown graph. For this, there are 9 degrees of freedom and hence from the lookup table, corresponding P value $= 0.10$. So there is weak evidence against the null hypothesis.

Graph for part (3) : Central Limit Theorem



Graph for part (5) : Histogram of Distribution of 1000 Random Variables



6. CODES :

Part (2):

```
% Part (2) : Genrate uniform random variables in [0,1] and find it's mean
%            and variance.

i = input('Enter the number of random numbers between 0 and 1: ')
X = rand(i,1);

% Find the mean and variance of the generated numbers
stats = [mean(X) std(X) var(X)]

% Expected value of mean and variance
exp_mean = (0+1)/2;
exp_var = (1-0)^2/12;

% Estimate of variance of sample mean
var_est = var(X)/i;
```

Part (3):

```
% Part (3) : Genrate N=100 uniform random variables in [0,1] 50 times to
%            get a set of sample means. Compare it to a normal curve.

N = 100;
no_trials = 50;

% Generate the 50 means

for m = 1:no_trials
    X = rand(N,1);
    Mean_trials(m) = mean(X);
end

% Fit Normal Distribution
histfit(Mean_trials)
```

Part (4):

```
% Part (4) : Genrate uniform random variables in [0,1] and find it's
%            co-variance relation.


i = input('Enter the number of random numbers between 0 and 1: ')
X = rand(i+1,1);

% The different summations used in the value
sum1 = 0;
for m = 1:i
    sum1 = sum1 +  X(m)*X(m+1);
end

sum2 = 0;
```

```
for m = 1:i
    sum2 = sum2 + X(m);
end

sum3 = 0;
for m = 2:i+1
    sum3 = sum3 + X(m);
end

% The value of z

z = ((sum1)/i) - ((sum2*sum3)/(i^2))
```

Part (5):

```
% Part (5) : Bonus Points

i = input('Enter the number of random numbers between 0 and 1: ')
X = rand(i,1);

% Generation of histogram
[N, edges]= histcounts(X,10);
bar(0:1:9,N)
xlim([0 9])
xlabel('Intervals')
ylabel('Random numbers')
hold on
A(1:1, 1:10)=i/10;
plot(0:1:9,A,'green')
xlim([0 9])
```