# **SDA Assignment 1**

Lushaank K - 20170010082

Vineesh S - 20170010132

Hemanth P G - 20170010048

Sumanth J - 20170010135

# **Question 1A (Discrete Distributions)**

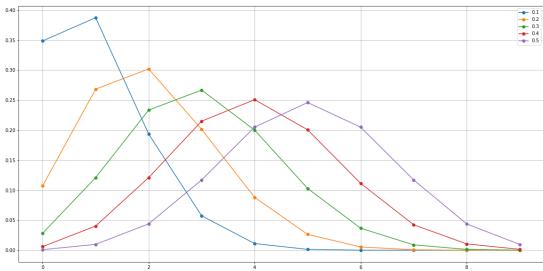
### **Binomial Distribution:**

The binomial distribution is a probability distribution that summarizes the likelihood that a value will take one of two independent values under a given set of parameters or assumptions.

PDF function:

$$\binom{n}{x} p^x q^{n-x}$$

Parameters of the binomial distribution are n and p.



This the graph of binomial distribution for p=[0.1,0.2,0.3,0.4,0.5] And n=10.

As we can see from the graph that as p value increases the local maximum shifts to the right side.

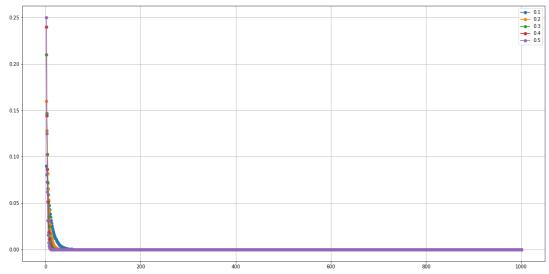
	p	Mean	Varirance
0	0.1	1.000000	0.900000
1	0.2	1.999999	1.599994
2	0.3	2.999941	2.099764
3	0.4	3.998951	2.397902
4	0.5	4.990234	2.499905

These are the practical values of mean and variance of binomial distribution, which are approximately equal to theoretical values as the mean is "np" and variance is "p(1-p)".

### **Geometric Distribution:**

The geometric distribution represents the number of failures before you get a success in a series of Bernoulli trials.

The pdf function is:  $q^{(x-1)}p$  where p is the probability of success on each trail.and q=1-p Here the parameter is "p"



This the graph of geometric distribution for p=[0.1,0.2,0.3,0.4,0.5]

	Lambda	Mean	Varirance
0	0.1	9.900000	91.890000
1	0.2	4.800000	21.760000
2	0.3	3.033333	9.387778
3	0.4	2.100000	5.190000
4	0.5	1.500000	3.250000

These are the practical values of mean and variance of geometric distribution, which are approximately equal to theoretical values as the mean is "1/p" and variance is " $q/(p^{**}2)$ ".Here q=1-p

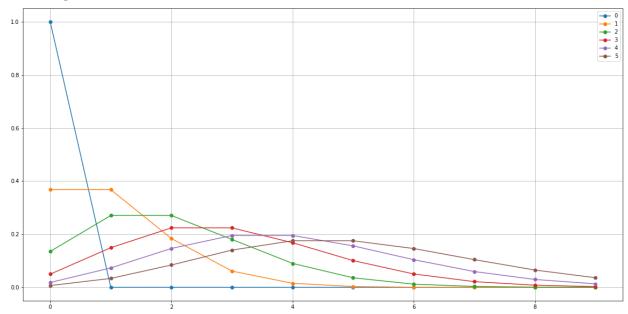
### **Poisson Distribution:**

Poisson distribution is a statistical distribution that shows how many times an event is likely to occur within a specified period of time.

It is used for independent events which occur at a constant rate within a given interval of time.

The pdf function is: $^{(e^{-\mu})}(\mu^X)/x!$  .here "x" is the actual number of successes that result from the experiment.

The parameter is : " $\lambda$ " or it can also be written as " $\mu$ "



This the graph of geometric distribution for  $\lambda$ =[0,1,2,3,4,5]

As we can see from the above graph that as the value of  $\lambda$  increases the local maximum shifts to the right side.

And here as the value of  $\lambda$  tends to infinity the distribution always tends to "0".

	Lambda	Mean	Varirance
0	0	0.000000	0.000000
1	1	0.999999	0.999991
2	2	1.999525	1.997038
3	3	2.988591	2.949774
4	4	3.914546	3.772736
5	5	4.659532	4.614003

These are the practical values of mean and variance of poisson distribution, which are approximately equal to theoretical values as the mean is " $\lambda$ " and variance is " $\lambda$ ".

## **Negative Binomial:**

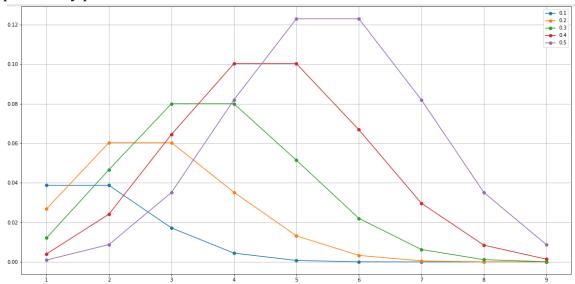
The negative binomial distribution is a discrete probability distribution that models the number of failures in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of successes (denoted r) occurs.

$$\binom{x-1}{r-1}p^r(1-p)^{x-r}$$

The pdf function is:

It has two parameters , they are "stopping parameter r" and probability p".

"success



This the graph of negative binomial distribution for p=[0.1,0.2,0.3,0.4,0.5] And r=[1,2,3,4,5,6,7,8,9]

And n=10 As we can see from the above graph that as the value of  $\lambda$  increases the local maximum shifts to the right side.

And here as the value of  $\lambda$  tends to infinity the distribution always tends to "0". **p** r Mean Variance

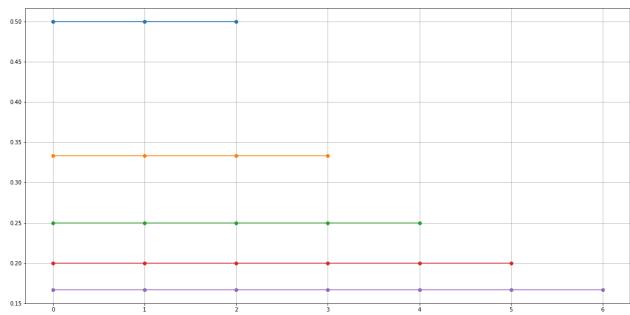
	р	r	Mean	Varirance
0	0.1	1	0.111111	0.123457
1	0.1	2	0.222222	0.246914
2	0.1	3	0.333333	0.370370
3	0.1	4	0.44444	0.493827
4	0.1	5	0.555556	0.617284
5	0.1	6	0.666667	0.740741
6	0.1	7	0.777778	0.864198
7	0.1	8	0.888889	0.987654
8	0.1	9	1.000000	1.111111
9	0.2	1	0.250000	0.312500
10	0.2	2	0.500000	0.625000
11	0.2	3	0.750000	0.937500
12	0.2	4	1.000000	1.250000
13	0.2	5	1.250000	1.562500
14	0.2	6	1.500000	1.875000
15	0.2	7	1.750000	2.187500

Here mean is:rp/q;q=1-p. Here variance is  $rp/(q^{**2})$ .

### **Discrete Uniform:**

The discrete uniform distribution is a symmetric probability distribution wherein a finite number of values are equally likely to be observed; every one of n values has equal probability 1/n.

The pdf function is:1/N where N is maximum in the uniform distribution  $\{0,1,2,3,....N\}$ 



This is the graph of Discrete Uniform Distribution with N=[3,4,5,6,7] Here as the value of N continues to increase the value of the distribution function continues to decrease.

And here as the value of N tends to infinity the distribution always tends to "0".

	N	Mean	Varirance
0	3	1.5	0.250000
1	4	2.0	0.666667
2	5	2.5	1.250000
3	6	3.0	2.000000
4	7	3.5	2.916667

These are the practical values of mean and variance of Discrete Uniform distribution, which are approximately equal to theoretical values as the mean is "(N+1)/2" and variance is "((N\*\*2)-1)/12".

# **Question 1B (Continuous Distributions):**

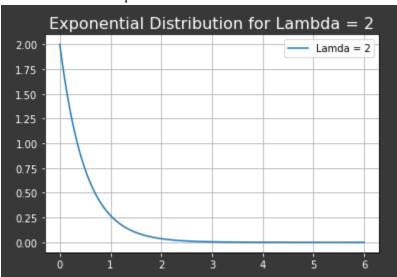
### **Exponential Distribution:**

The exponential distribution is the probability distribution of the time between events in a Poisson point process

#### PDF function:

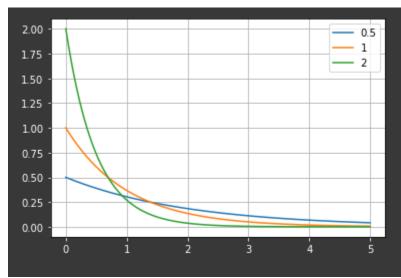
$$f(x;\lambda) = \left\{ egin{array}{ll} \lambda e^{-\lambda x} & x \geq 0, \ 0 & x < 0. \end{array} 
ight.$$

Parameters of the exponential distribution is  $\boldsymbol{\lambda}$ 



Exponential distribution graph is convex in nature

It has only one parameter Lambda ( $\lambda$ )



As we can observe from the graph

As  $\lambda$  decreases, the graph is stretched towards right.

As  $\lambda$  increases, the graph is constricts towards left side

The value of the pdf function is always equal to the value of  $\lambda$  at t=0

As  $\lambda$  tends to infinity, f(x) tends to zero

Mean and variance:

$$\mathrm{E}[X] = rac{1}{\lambda}. \quad \mathrm{Var}[X] = rac{1}{\lambda^2}$$

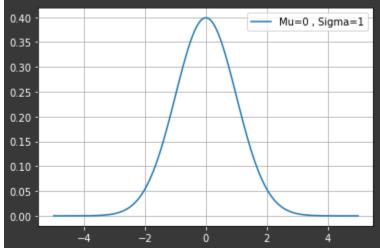
	Lambda	Mean	Varirance
0	0.5	1.918970	3.319808
1	1.0	0.999401	0.995462
2	2.0	0.499950	0.250000
3	5.0	0.199980	0.040000
4	10.0	0.099989	0.010000
5	15.0	0.066659	0.004445

These are the practical values of mean and variance of exponential distribution, which are approximately equal to theoretical values as the mean in " $1/\lambda$ " and variance is " $1/\lambda$ ".

### **Normal Distribution:**

PDF of normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



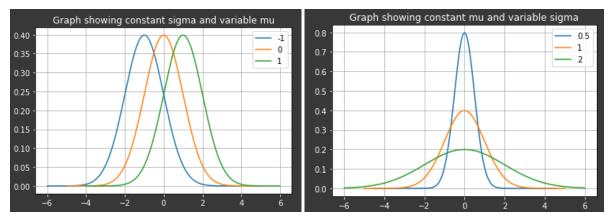
The above graph is plot of normal distribution for mu=0 and sigma = 1

The normal distribution graph is a bell shaped curve having peak at mean ( mu ) mean and median are equal; both located at the center of the distribution  $\approx$ 68% approximately equals, 68% of the data falls within 111 standard deviation of the mean

 $\approx\!\!95\%$  approximately equals, 95% of the data falls within 222 standard deviations of the mean

 $\approx$ 99.7% approximately equals, 99.7, percent of the data falls within 333 standard deviations of the mean

#### Effect of parameters:



As from above graphs we can see that:

As mu increases the graph shifts towards left

As sigma increases the graph height increases

#### Mean and variance:

In normal distribution mean is equal to  $\mu$  and variance is equal to  $\sigma^2$ 

	Parameters	Mean	Varirance
0	mu = -1 , sigma = 0.5	-1.11111	0.15432
1	mu = -1 , sigma = 1	-1.00000	0.99999
2	mu = -1 , sigma = 2	-0.83108	3.37608
3	mu = 0 , sigma = 0.5	-0.00000	0.27778
4	mu = 0 , sigma = 1	-0.00000	0.99998
5	mu = 0 , sigma = 2	0.00000	3.23569
6	mu = 1 , sigma = 0.5	1.11111	0.15432
7	mu = 1 , sigma = 1	1.00000	0.99999
8	mu = 1 , sigma = 2	0.83108	3.37608

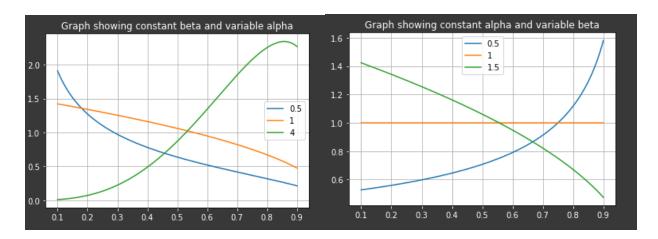
These are the practical values of mean and variance of exponential distribution, which are approximately equal to theoretical values as the mean in " $\mu$ " and variance is " $\sigma^2$ ".

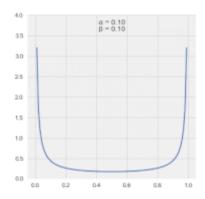
### **Beta Distribution:**

The PDF of beta distribution:

$$f(x) = rac{x^{p-1}(1-x)^{q-1}}{B(p,q)} \qquad 0 \leq x \leq 1; p,q > 0$$

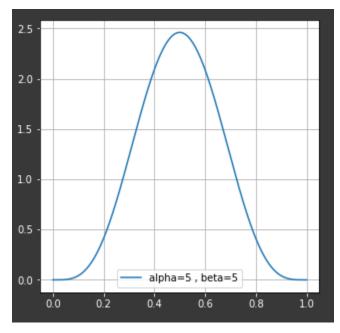
There are two parameters  $\alpha$  and  $\beta$ , where  $\alpha, \beta > 0$ 



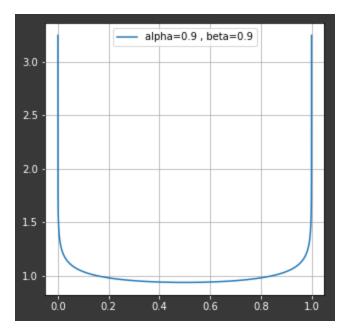


Source: wikipedia

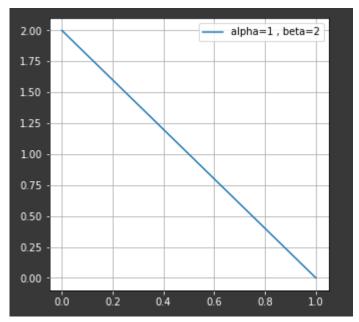
The PDF of Beta distribution can be U-shaped with asymptotic ends, bell-shaped, strictly increasing/decreasing or even straight lines. As you change  $\alpha$  or  $\beta$ , the shape of the distribution changes.



• The PDF of a beta distribution is approximately normal if  $\alpha + \beta$  is large enough and  $\alpha \& \beta$  are approximately equal.



When  $\alpha$  <1,  $\beta$ <1, the PDF of the Beta is U-shaped.



If either of parameters are 1 then beta distribution is a straight line

# Mean and Variance:

$$E[X] = \frac{\alpha}{\alpha + \beta} \qquad Var[X] = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}$$

	Parameters	Mean	Varirance	
0	alpha = 0.5 , beta = 0.5	0.36896	0.09284	
1	alpha = 0.5 , beta = 1	0.34258	0.07396	
2	alpha = 0.5 , beta = 1.5	0.27997	0.06160	
3	alpha = 1 , beta = 0.5	0.44799	0.09603	
4	alpha = 1 , beta = 1	0.50000	0.05333	
5	alpha = 1 , beta = 1.5	0.45379	0.04424	
6	alpha = 4 , beta = 0.5	0.35183	0.14965	
7	alpha = 4 , beta = 1	0.59049	0.09420	
8	alpha = 4, beta = 1.5	0.70559	0.00957	

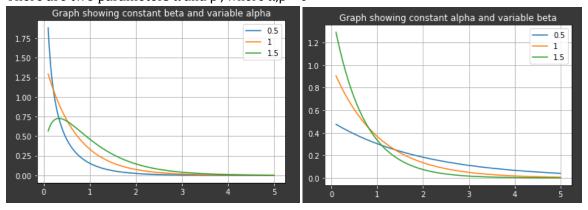
These are the practical values of mean and variance of exponential distribution, which are approximately equal to theoretical values of mean and variance

### **Gamma Distribution:**

The pdf of gamma distribution is:

$$f(x;lpha,eta)=rac{eta^{lpha}x^{lpha-1}e^{-eta x}}{\Gamma(lpha)}\quad ext{ for }x>0\quadlpha,eta>0$$

There are two parameters  $\alpha$  and  $\beta$ , where  $\alpha, \beta > 0$ 



When  $\alpha$ < 1, the Gamma distribution is exponentially shaped and asymptotic to both the vertical and horizontal axes

When  $\alpha$ =1 then graph is same as exponential distribution

When  $\alpha>1$  the Gamma distribution assumes a mounded (unimodal), but skewed shape.

The skewness reduces as the value of a increases.

 $\beta$  has the effect of stretching or compressing the range of the Gamma distribution

### Mean and Variance:

$$E[X] = \alpha/\beta$$
  $Var[X] = \alpha/\beta^2$ 

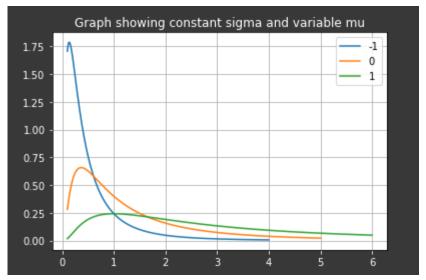
	Parameters	Mean	Varirance
0	alpha = 0.5 , beta = 0.5	1.00000	2.00000
1	alpha = 0.5 , beta = 1	0.50000	0.50000
2	alpha = 0.5 , beta = 1.5	0.33333	0.22222
3	alpha = 1, beta = 0.5	2.00000	4.00000
4	alpha = 1 , beta = 1	1.00000	1.00000
5	alpha = 1, beta = 1.5	0.66667	0.44444
6	alpha = 1.5 , beta = 0.5	3.00000	6.00000
7	alpha = 1.5 , beta = 1	1.50000	1.50000
8	alpha = 1.5 , beta = 1.5	1.00000	0.66667

### **Lognormal Distribution:**

PDF of log normal distribution is:

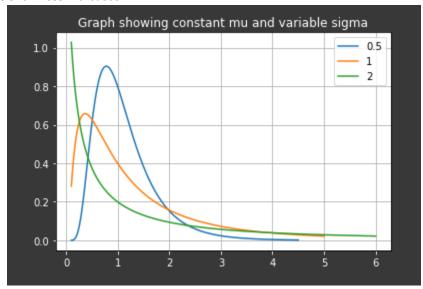
$$\frac{1}{x} \cdot \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln x - \mu)^2}{2\sigma^2} \right)$$

Lognormal Distribution has two parameters mu and sigma



As mu increases the peak of the graph decreases

As mu increases skewness increases



The lognormal distribution is a distribution skewed to the right.

As sigma increases the graph shifts towards left

The degree of skewness increases as sigma increases.

Mean and Variance:

Mean = 
$$\exp\left(\mu + \frac{\sigma^2}{2}\right)$$
 variance =  $\left[\exp(\sigma^2) - 1\right] \exp(2\mu + \sigma^2)$ 

	Parameters	Mean	Varirance
0	mu = -1 , sigma = 0.5	0.41686	0.04936
1	mu = -1 , sigma = 1	0.60653	0.63212
2	mu = -1 , sigma = 2	2.71828	396.03974
3	mu = 0 , sigma = 0.5	1.13315	0.36470
4	mu = 0 , sigma = 1	1.64872	4.67077
5	mu = 0 , sigma = 2	7.38906	2926.35984
6	mu = 1 , sigma = 0.5	3.08022	2.69476
7	mu = 1 , sigma = 1	4.48169	34.51261
8	mu = 1 , sigma = 2	20.08554	21623.03700

# **Question 2**

#### **Exponential Distribution:**

This is the part where we check the application of Central Limit Theorem with an exponential function of lambda being 0.2. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.1. That is what Central Limit Theorem states.

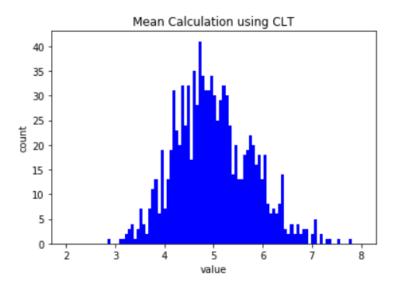


Fig 2.1: Plotting the graph of values from Exponential Distribution

From the given data, we could find the theoretical data mean and variances. Then, we could also compute mean and variance from the normal data generated.

```
Exponential Data Mean is: 5.0
Exponential Data Variance is: 0.625
Experimental Normal Data Mean is: 5.0
```

Experimental Normal DataVariance is: 0.5851

Fig 2.2 and 2.3: Shows results between exponential and experimental normal values

We would get an exponential graph from the data we generated, if we did average it out as follows in Fig 2.4

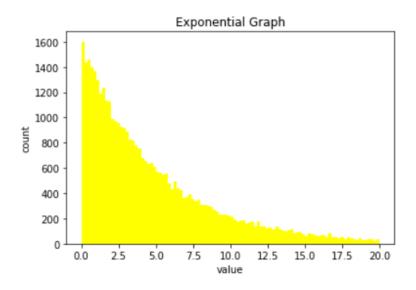


Fig 2.4: Exponential Graph for 40,000 values

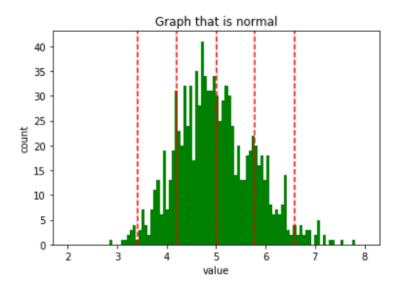


Fig 2.5: Exponential graph uses CLT and becomes a normal graph

#### **Normal Distribution:**

This is the part where we check the application of Central Limit Theorem with a normal function of mu being at 10 and sigma at 0.3. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.6. That is what Central Limit Theorem states.

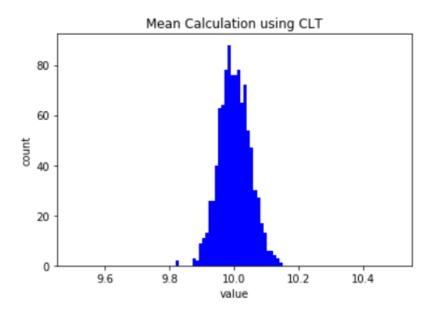


Fig 2.6: Plotting the graph of values from Normal Distribution

```
Normal Data Mean is: 10
Normal Data Variance is: 0.09

Experimental Normal Data Mean is: 10.0
Experimental Normal Data Variance is: 0.0023
```

Fig 2.7 and 2.8: Shows results between normal and experimental normal values

We would get an normal graph from the data we generated, if we did average it out as follows in Fig  $2.9\,$ 

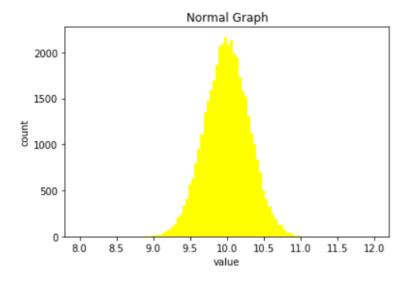


Fig 2.9: Normal Graph for 40,000 values

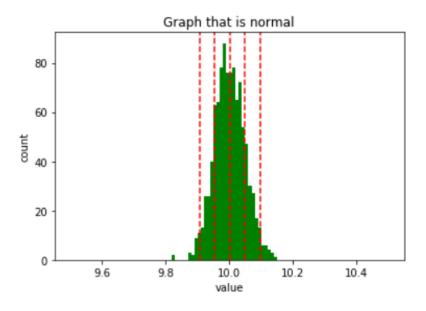


Fig 2.10: Normal graph uses CLT and becomes a normal graph

#### **Beta Distribution:**

This is the part where we check the application of Central Limit Theorem with a beta function of alpha being at 10 and beta at 7. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.11. That is what Central Limit Theorem states.

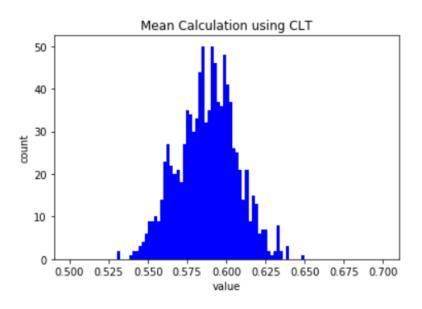


Fig 2.11: Plotting the graph of values from Beta Distribution

From the given data, we could find the theoretical data mean and variances. Then, we could also compute mean and variance from the normal data generated.

```
Beta Data Mean is : 0.59
Beta Data Variance is : 0.0135
```

Experimental Normal Data Mean is: 0.59
Experimental Normal Data Variance is: 0.0004

Fig 2.12 and 2.13: Shows results between beta and experimental normal values

We would get an beta graph from the data we generated, if we did average it out as follows in Fig 2.14

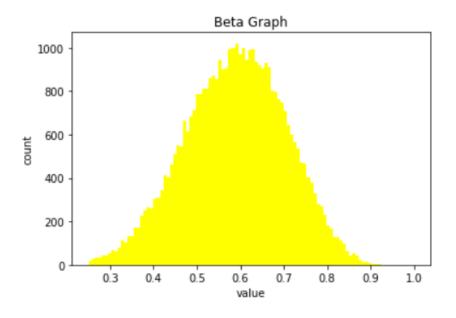


Fig 2.14: Beta Graph for 40,000 values

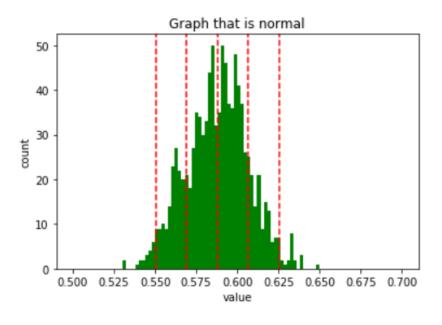


Fig 2.15: Beta graph uses CLT and becomes a normal graph

#### **Gamma Distribution:**

This is the part where we check the application of Central Limit Theorem with a gamma function of a being at 10 and b at 2. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.16 . That is what Central Limit Theorem states.

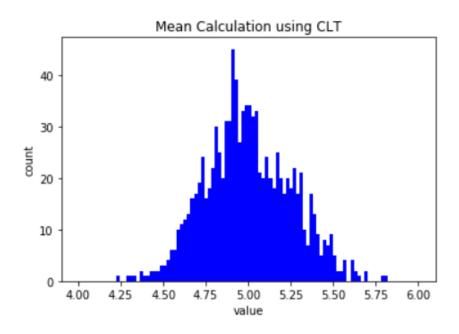


Fig 2.16: Plotting the graph of values from Gamma Distribution

```
Gamma Data Mean is: 5.0
Gamma Data Variance is: 2.5

Experimental Normal Data Mean is: 5.0

Experimental Normal Data Variance is: 0.0636
```

Fig 2.17 and 2.18: Shows results between gamma and experimental normal values

We would get an gamma graph from the data we generated, if we did average it out as follows in Fig 2.19

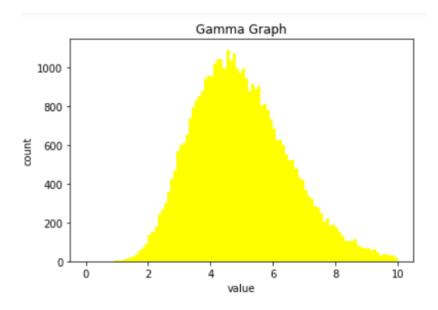


Fig 2.19: Gamma Graph for 40,000 values

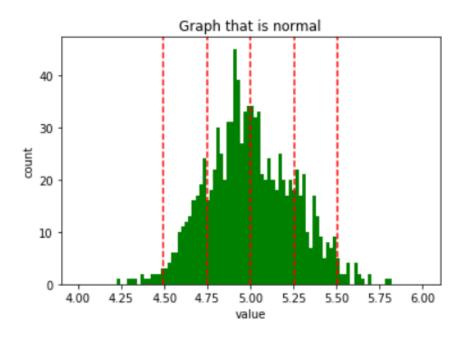
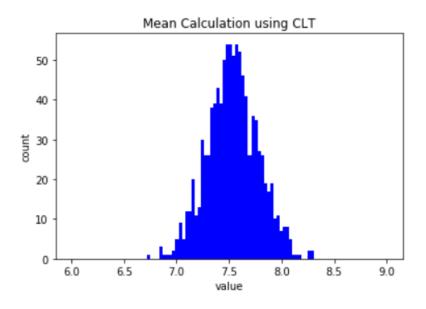


Fig 2.20: Gamma graph uses CLT and becomes a normal graph

### **LogNormal Distribution:**

This is the part where we check the application of Central Limit Theorem with a lognormal function of mu being at 2 and sigma at 0.2. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.21. That is what Central Limit Theorem states.



From the given data, we could find the theoretical data mean and variances. Then, we could also compute mean and variance from the normal data generated.

```
LogNormal Data Mean is: 7.54
LogNormal Data Variance is: 2.3191
Experimental Normal Data Mean is: 7.53
```

Fig 2.22 and 2.23: Shows results between lognormal and experimental normal values

Experimental Normal Data Variance is: 0.0579

We would get an lognormal graph from the data we generated, if we did average it out as follows in Fig 2.24

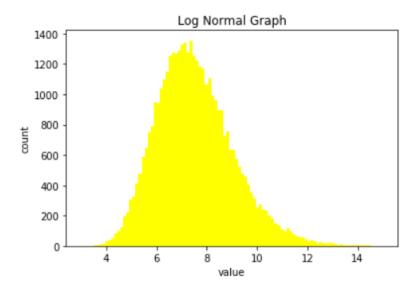


Fig 2.24: LogNormal Graph for 40,000 values

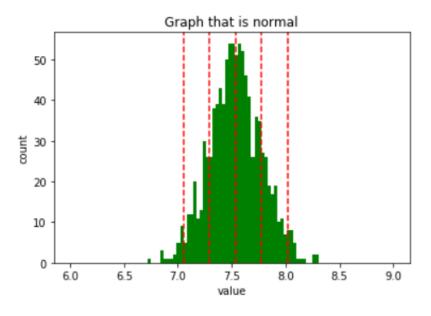


Fig 2.25: LogNormal graph uses CLT and becomes a normal graph

#### **Binomial Distribution:**

This is the part where we check the application of Central Limit Theorem with a binomial function of n being at 10 and p at 0.4. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.26 . That is what Central Limit Theorem states.

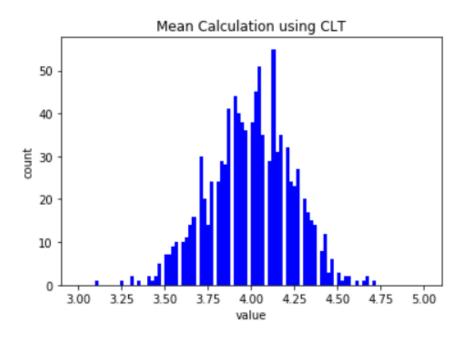


Fig 2.26: Plotting the graph of values from Binomial Distribution

```
Binomial Data Mean is: 4.0
Binomial Data Variance is: 2.4

Experimental Normal Data Mean is: 4.0

Experimental Normal Data Variance is: 0.0561
```

Fig 2.27 and 2.28: Shows results between binomial and experimental normal values

We would get an binomial graph from the data we generated, if we did average it out as follows in Fig 2.29

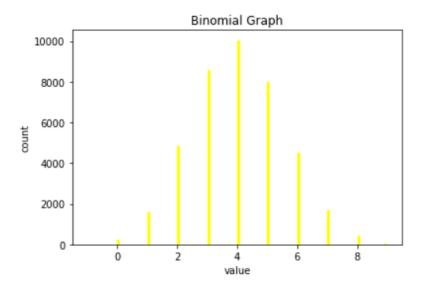


Fig 2.29: Binomial Graph for 40,000 values

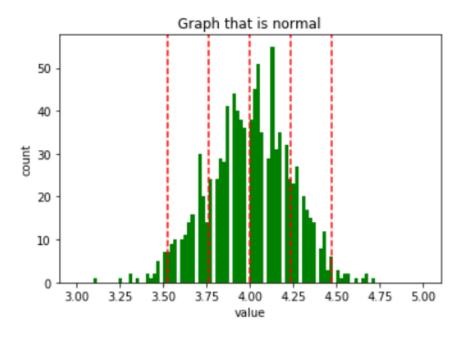


Fig 2.30: Binomial graph uses CLT and becomes a normal graph

#### **Geometric Distribution:**

This is the part where we check the application of Central Limit Theorem with a geometric function of p at 0.4. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.31. That is what Central Limit Theorem states.

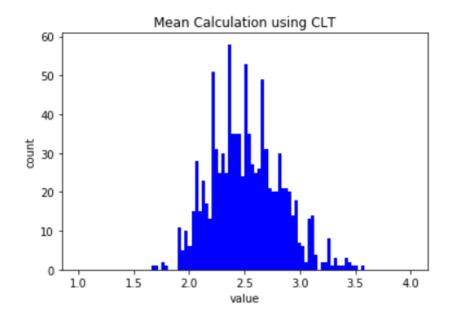


Fig 2.31: Plotting the graph of values from Geometric Distribution

```
Geometric Data Mean is: 2.5
Geometric Data Variance is: 3.75

Experimental Normal Data Mean is: 2.51

Experimental Normal Data Variance is: 0.0997
```

Fig 2.32 and 2.33: Shows results between geometric and experimental normal values

We would get an geometric graph from the data we generated, if we did average it out as follows in Fig 2.34

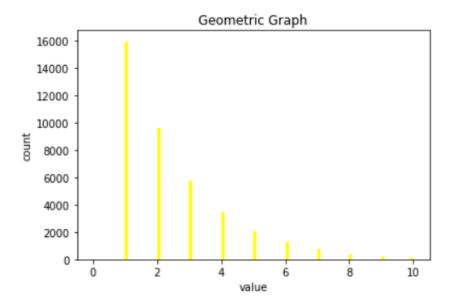


Fig 2.34: Geometric Graph for 40,000 values

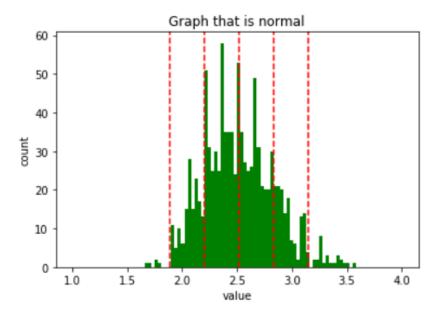


Fig 2.35: Geometric graph uses CLT and becomes a normal graph

#### **Poisson Distribution:**

This is the part where we check the application of Central Limit Theorem with a poisson function of lambda at 2. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.36 . That is what Central Limit Theorem states.

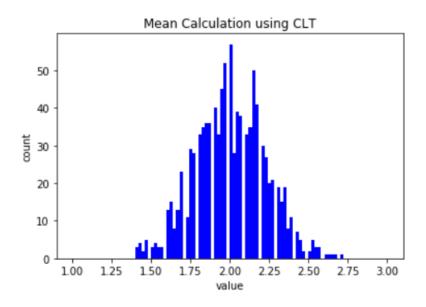


Fig 2.36: Plotting the graph of values from Poisson Distribution

```
Poisson Data Mean is: 2
Poisson Data Variance is: 2

Experimental Normal Data Mean is: 2.0

Experimental Normal Data Variance is: 0.0505
```

Fig 2.37 and 2.38: Shows results between poisson and experimental normal values

We would get a poisson graph from the data we generated, if we did average it out as follows in Fig 2.39

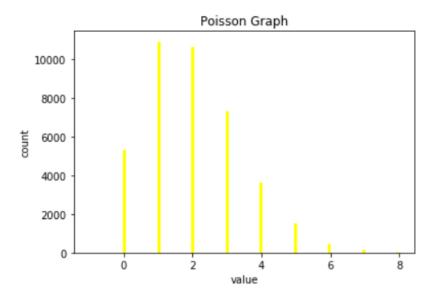


Fig 2.39: Poisson Graph for 40,000 values

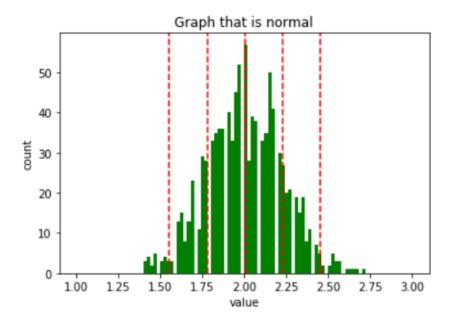


Fig 2.40: Poisson graph uses CLT and becomes a normal graph

### **Negative Binomial Distribution:**

This is the part where we check the application of Central Limit Theorem with a negative binomial function of n at 10 and p at 0.4. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.41 . That is what Central Limit Theorem states.

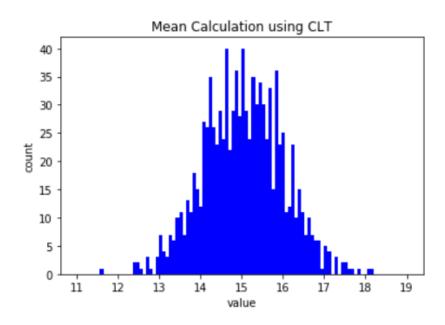


Fig 2.41: Plotting the graph of values from Negative Binomial Distribution

```
Negative Binomial Data Mean is: 6.67
Negative Binomial Data Variance is: 11.1111
Experimental Normal Data Mean is: 15.04
Experimental Normal Data Variance is: 0.9108
```

Fig 2.42 and 2.43: Shows results between negative binomial and experimental normal values

We would get a negative binomial graph from the data we generated, if we did average it out as follows in Fig 2.44

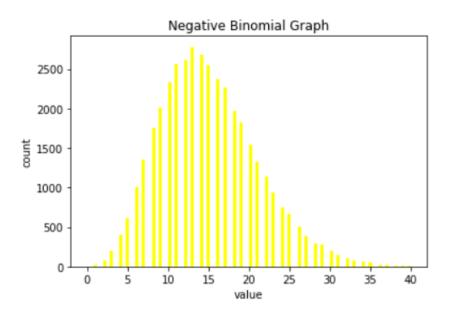


Fig 2.44: Negative Binomial Graph for 40,000 values

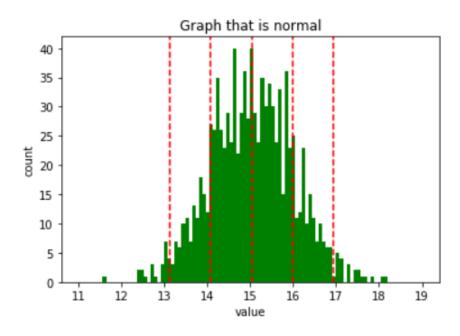


Fig 2.45: Negative Binomial graph uses CLT and becomes a normal graph

#### **Discrete Uniform Distribution:**

This is the part where we check the application of Central Limit Theorem with a discrete uniform function of a at 2 and b at 6. Taking a set of 40 values from the distribution and considering it as a set. Then, taking 1000 such sets to compute the graph with respect to the number of responses, which always becomes a Normal Distribution as shown in Fig 2.46. That is what Central Limit Theorem states.

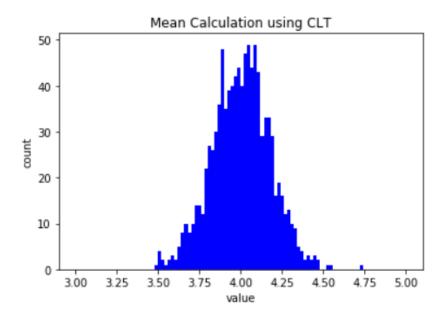


Fig 2.46: Plotting the graph of values from Discrete Uniform Distribution

```
Discrete Uniform Data Mean is: 4.0
Discrete Uniform Data Variance is: 2.0

Experimental Normal Data Mean is: 4.0

Experimental Normal Data Variance is: 0.0308
```

Fig 2.47 and 2.48: Shows results between discrete uniform and experimental normal values

We would get a discrete uniform graph from the data we generated, if we did average it out as follows in Fig 2.49

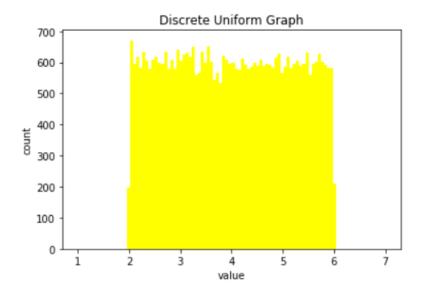


Fig 2.49: Discrete Uniform Graph for 40,000 values

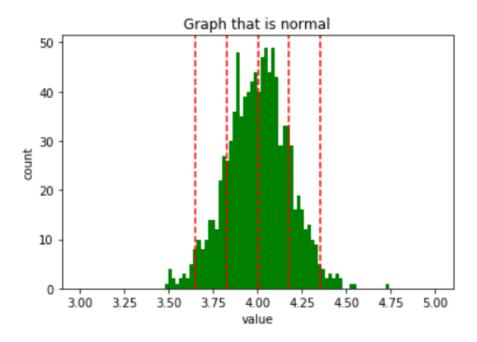


Fig 2.50: Discrete uniform graph uses CLT and becomes a normal graph