GROUP 17 ASSIGNMENT 1

TEAM MEMBERS:

EDURU SUMASREE (S20180010052)
KOMAKULA SAI SRI THANYA (S20180010083)
GUDAPATI SAI DIVYA (S20180010059)

GROUP-17

Discrete Distributions:

BINOMIAL DISTRIBUTION:

A binomial random variable is the number of successes ‘x’ in ‘n’ repeated trials of a binomial experiment. The probability distribution of a binomial random variable is called a binomial distribution.

$b(x; n, P) = {}_nC_x * P^x * (1 - P)^{n-x}$

Example:flip a coin 2 times and count the number of times the coin lands on heads.

FIRST PART:

Parameters of distribution:

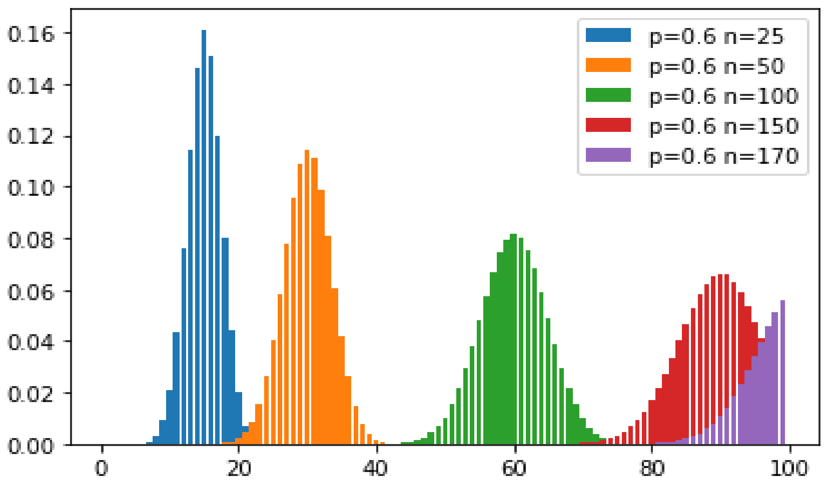
Parameters are n,p where n is number of trials and p is probability of success for each trial.

Distribution for different values of parameters:

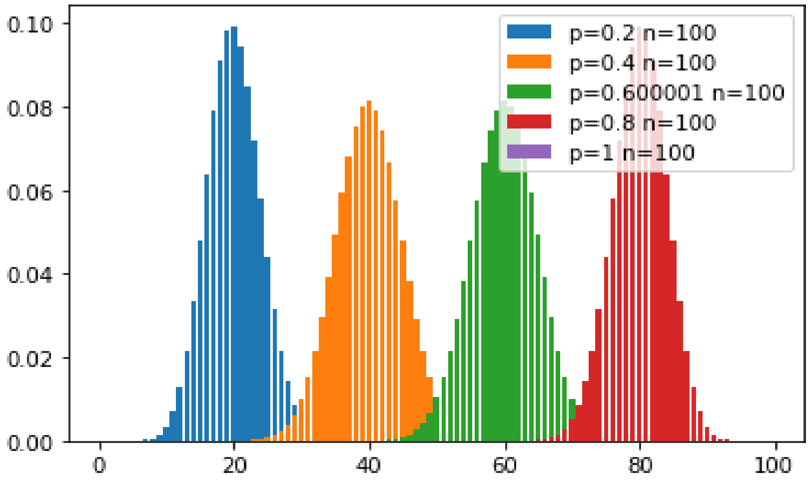
Considering n=[25,50,100,150,170] for 5 different samples and p= [0.2,0.4,0.600001,0.8,1] respectively

The distribution is as follows:

When n is varying and p is constant(i.e, p=0.6) :



When p is varying and n is constant(i.e, n= 100) :

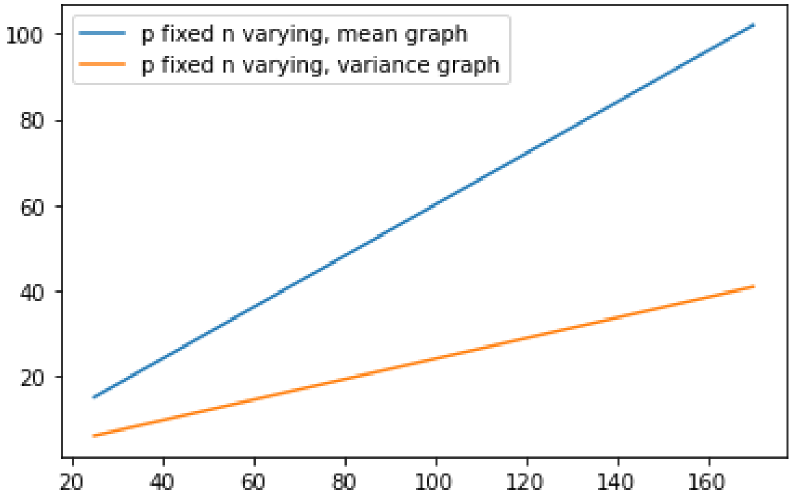


Mean and variance changes for different parameters:

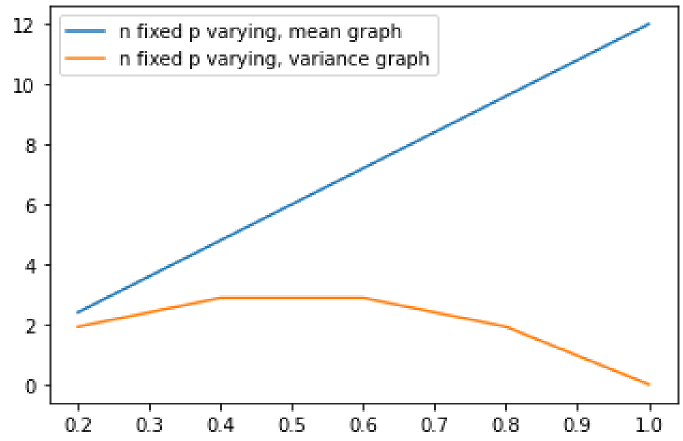
Mean= np

Variance= np(1-p)

When p is fixed and n is varying:



When n is fixed and p is varying:



SECOND PART:

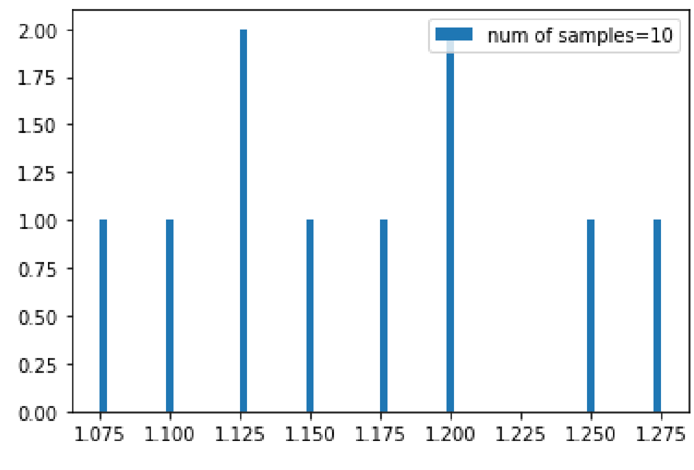
Mean of sample mean: 1.2000275000000002

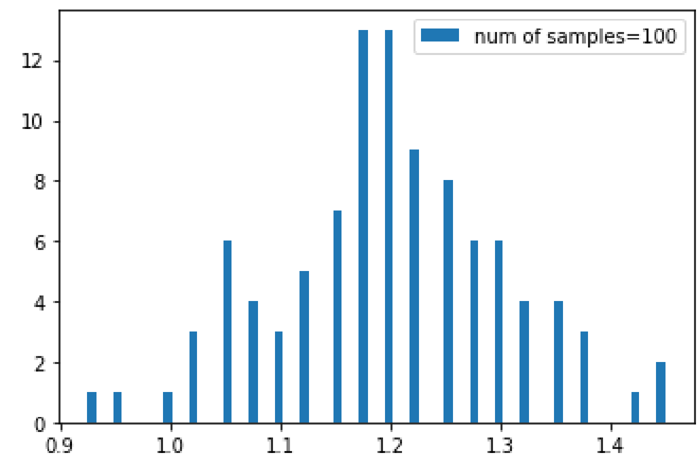
Variance of sample mean: 0.011927186743750002

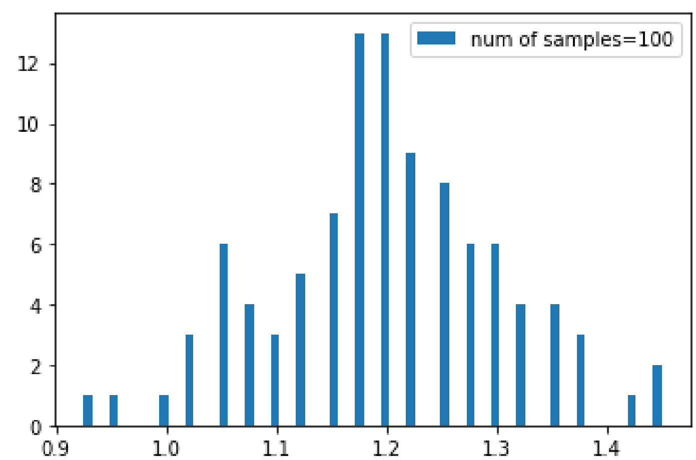
If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

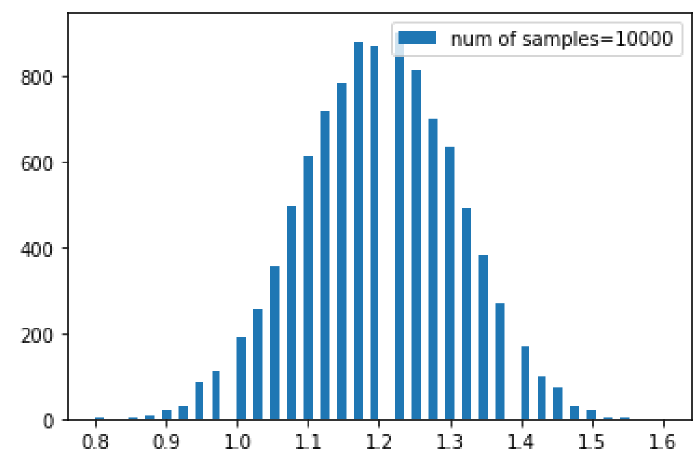
To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=10000,i.e 4th plot.,the curve is in the shape of bell.









GEOMETRIC DISTRIBUTION:

The geometric distribution is the probability distribution of the number of failures we get by repeating a Bernoulli experiment until we obtain the first success.
If X is a discrete random variable following geometric distribution with parameter p then
f(x) = ((1 − p)^(x-1))*(p) this is probability of success

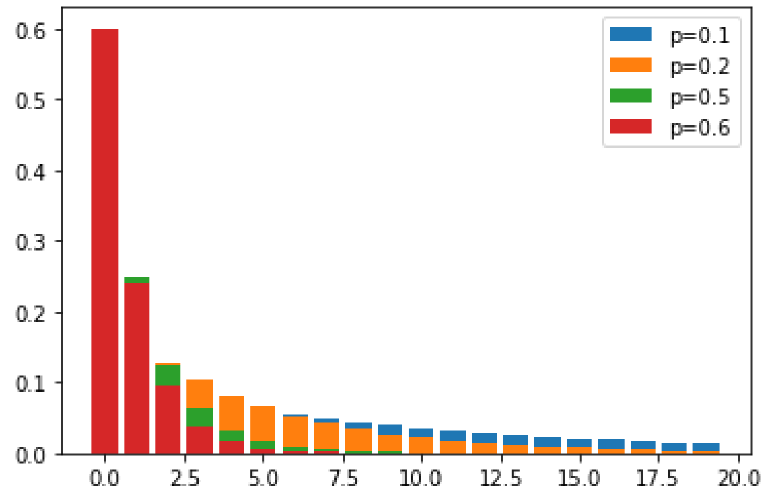
Example: If we toss a coin until we obtain a head, the number of tails before the first head has a geometric distribution.

FIRST PART:

Parameters of distribution:
p is the parameter which represents the probability of success .

Distribution for different values of parameters:
Considering 4 different samples with p= [0.1,0.2,0.5,0.6] respectively

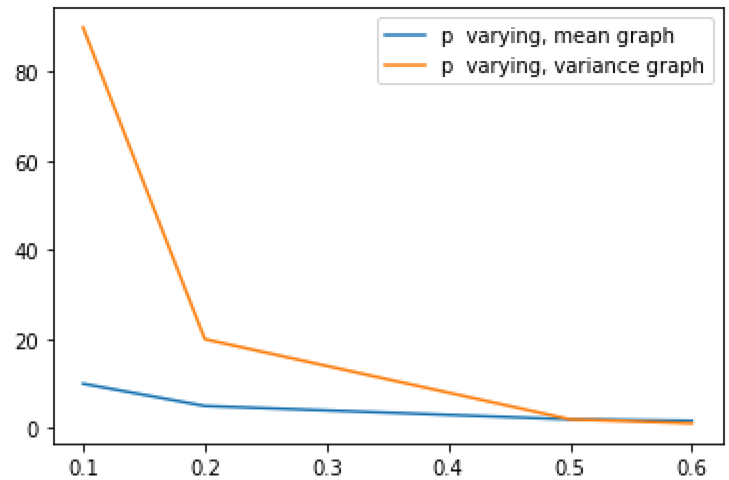
The distribution is as follows:



Mean and variance changes for different parameters:

Mean = 1/p

Variance = (1-p)/p*p



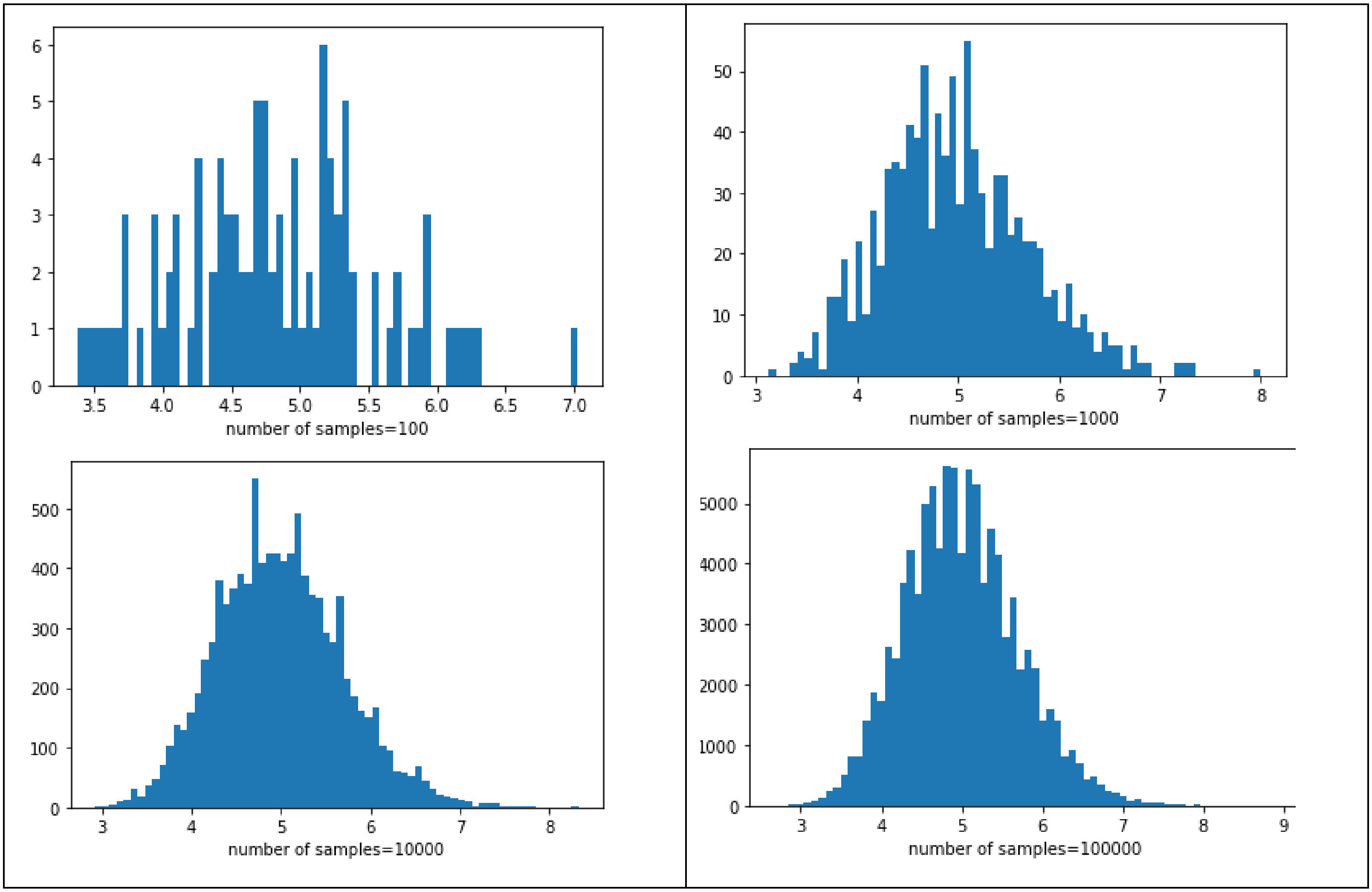
SECOND PART:

Mean of sample mean: 4.99326
Variance of sample mean: 0.49927032239999997

If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=100000,i.e 4th plot ,the curve is in the shape of bell.



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.

f(x) =($e^{-\lambda}$ λ^x)/x!

Where , X is poisson random variable has parameter λ is average rate of value

Example: A video store averages 400 customers every Friday night. We can find the probability that 600 customers will come in on any given Friday night by poisson distribution.

FIRST PART:

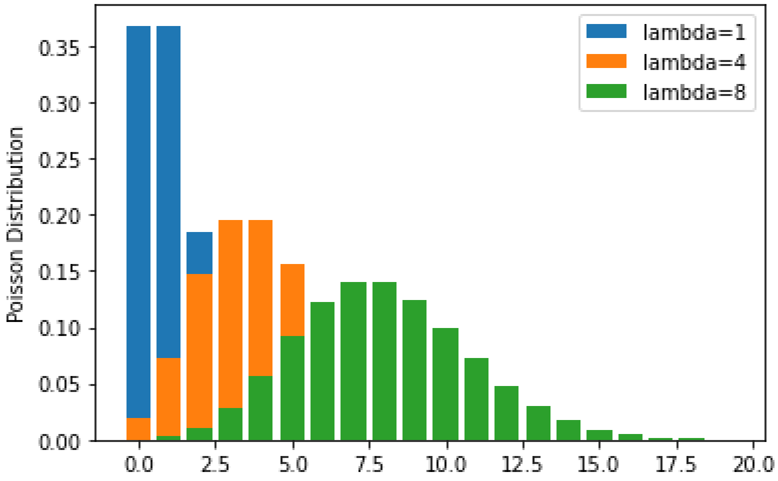
Parameters of this distribution:

- λ, which is the expected number of events in the interval (events/interval * interval length)

Distribution for different values of parameters:

Considering 4 different samples with λ= [1,3,4,8] respectively

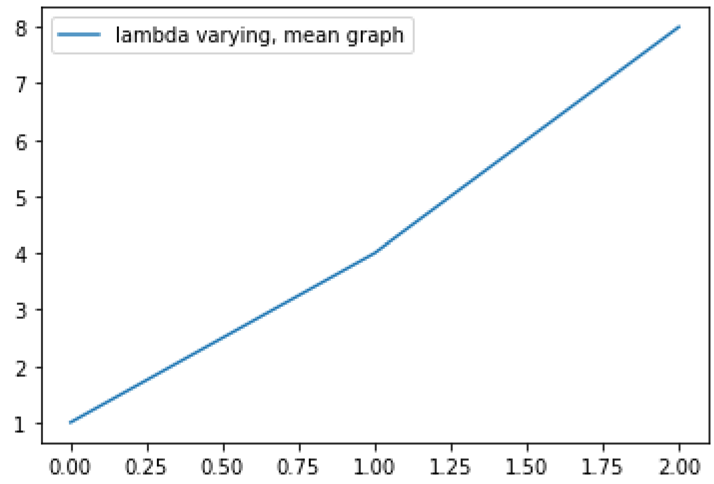
The distribution is as follows:



Mean and variance changes for different parameters:

Mean = λ

Variance = λ



SECOND PART:

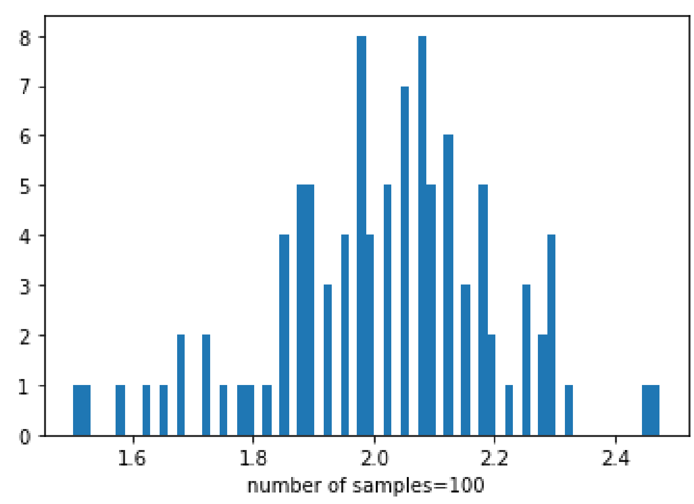
Mean of sample mean: 1.99862

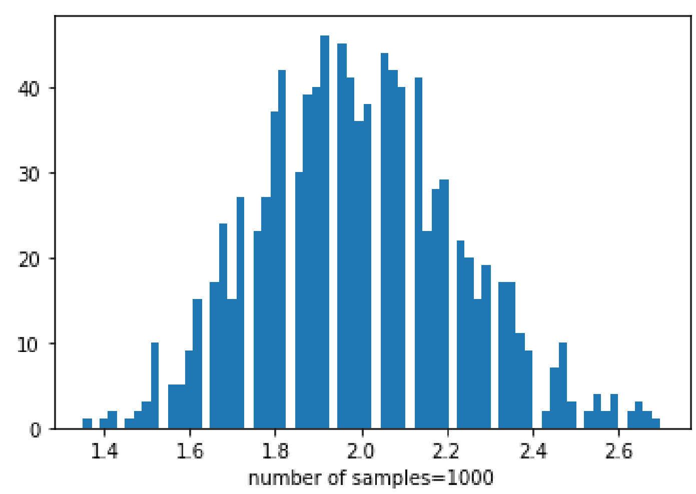
Variance of sample mean: 0.0503848456

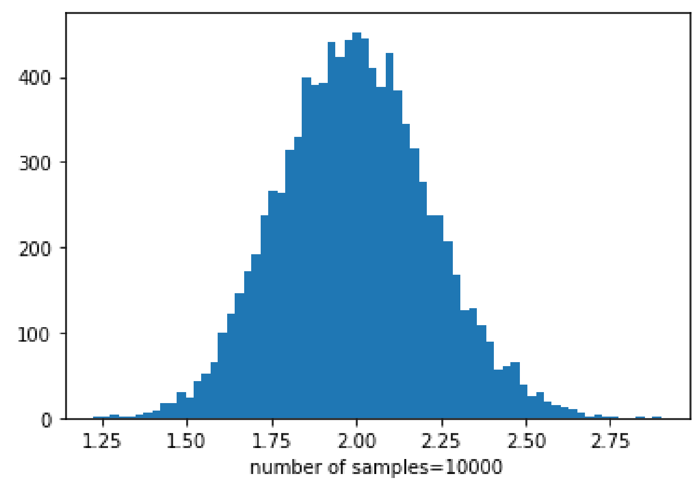
If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

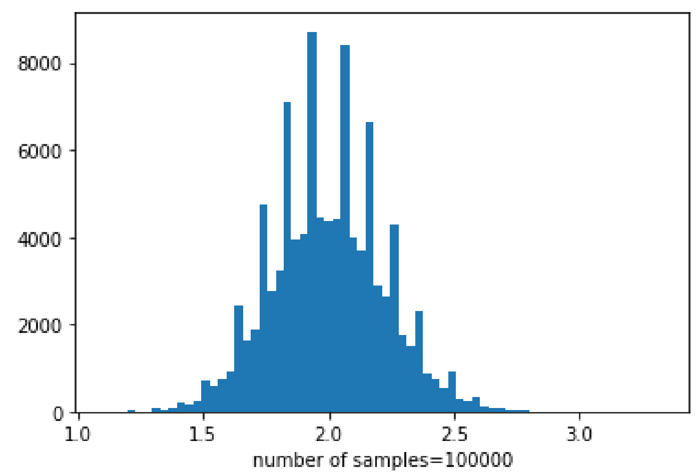
To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=100000,i.e 4th plot.The curve is in the shape of bell.









NEGATIVE BINOMIAL DISTRIBUTION:

The negative binomial experiment is almost the same as a binomial experiment with one difference: a binomial experiment has a fixed number of trials.

The random variable is the number of repeated trials, X, that produce a certain number of successes, r. In other words, it's the number of failures before a success.

$$nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \quad x = 0, 1, 2, \ldots$$

Example:

Take a standard deck of cards, shuffle them, and choose a card. Replace the card and repeat until you have drawn two aces. Y is the number of draws needed to draw two aces. As the number of trials isn't fixed (i.e. you stop when you draw the second ace), this makes it a negative binomial distribution.

FIRST PART:

Parameters of distribution:

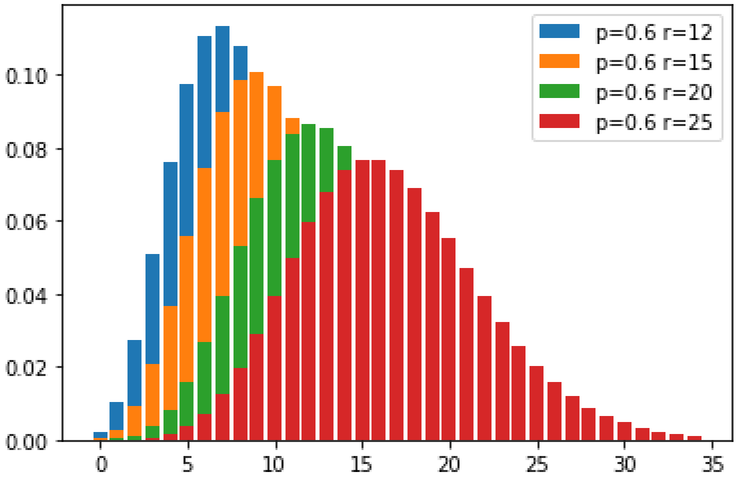
Parameters are r, p where r represents the number of failures before experiment is stopped and p represents success probability .

Distribution for different values of parameters:

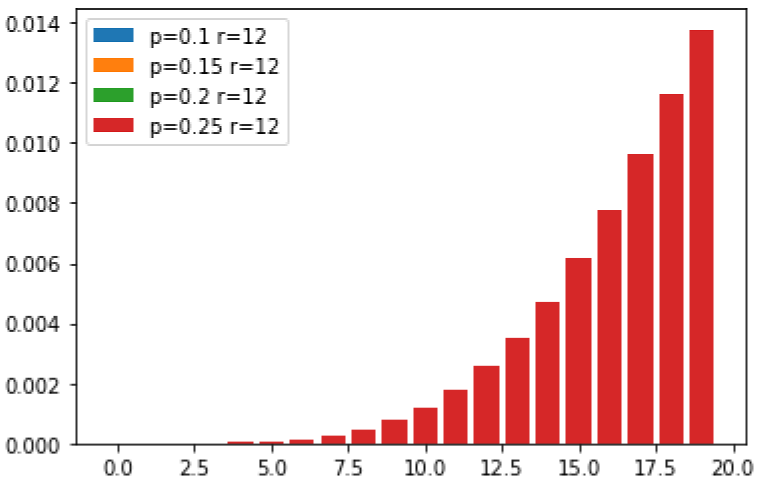
Considering p =[0.1,0.15,0.2,0.25] and r=[12,15,20,25] respectively.

The distribution is as follows:

When r is varying and p is constant(i.e, p=0.6):



When p is varying and r is constant(i.e, r=12):

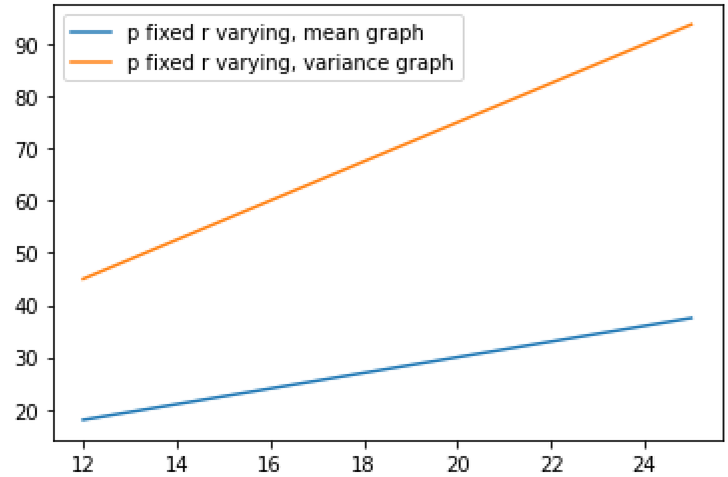


Mean and variance changes for different parameters:

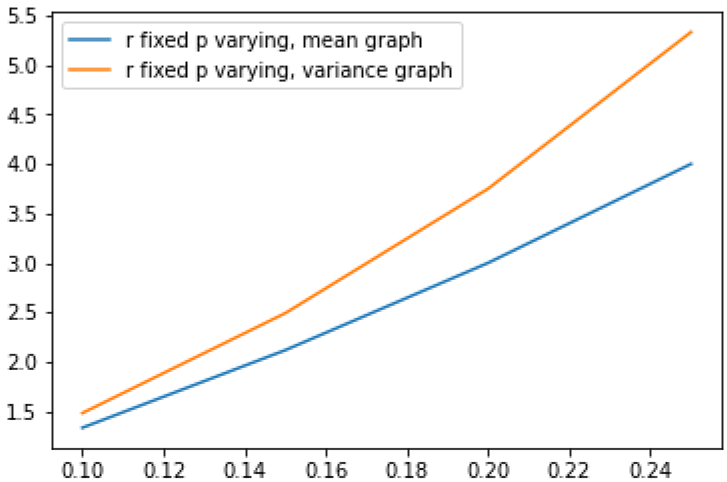
Mean = (p*r)/(1-p)

Variance = (p*r)/((1-p)*(1-p))

When p is fixed and r is varying:



When r is fixed and p is varying:



SECOND PART:

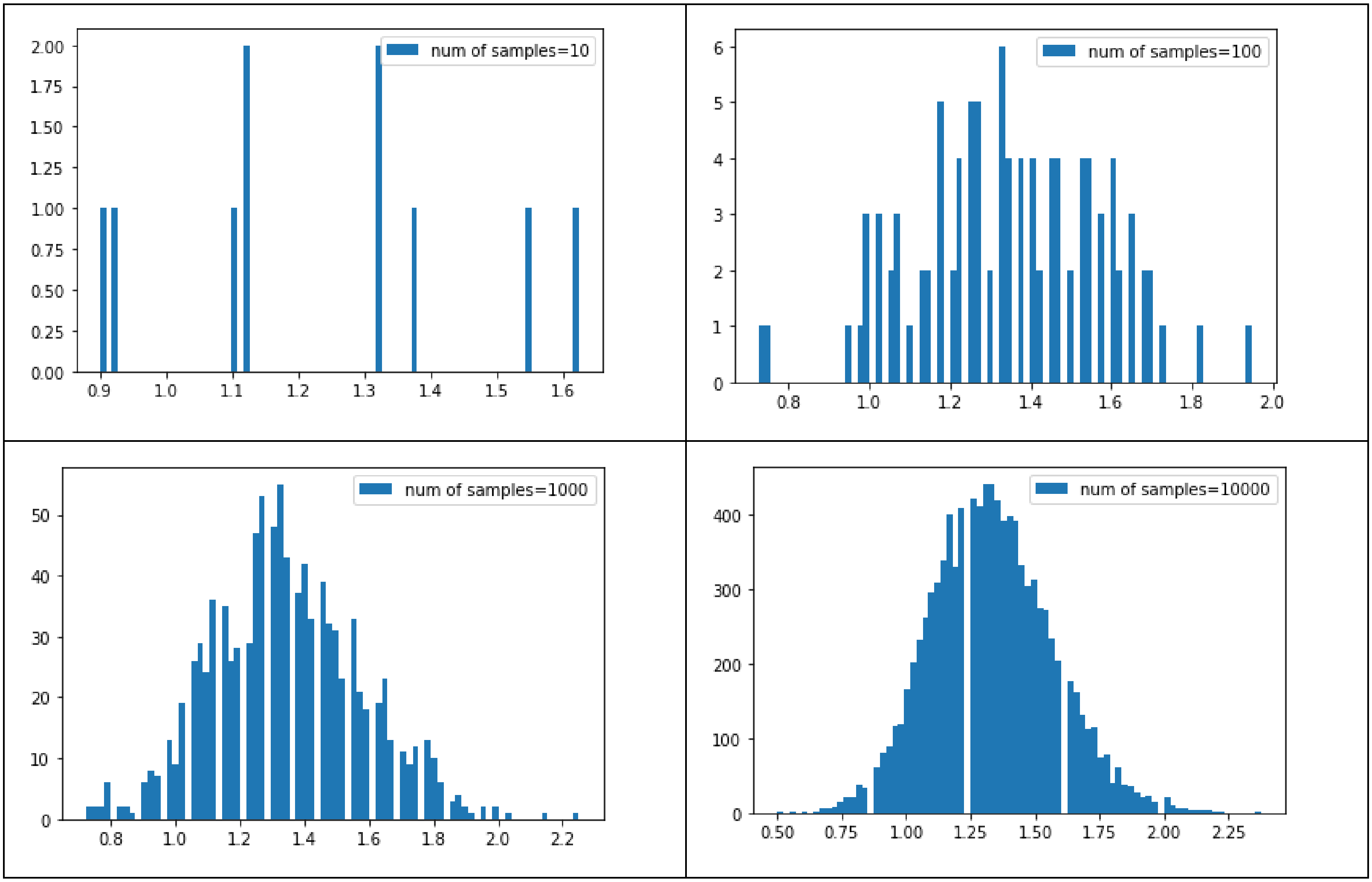
Mean of sample mean: 1.3295749999999997

Variance of sample mean: 0.056370944375

If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=10000,i.e 4th plot.The curve is in the shape of bell.



DISCRETE UNIFORM DISTRIBUTION:

A random variable X has a discrete uniform distribution if each of the n values in its range, say x1, x2, . . . , xn, has equal probability. Then, f(xi) = 1/n where f(x) represents the probability mass function.
Suppose X is a discrete uniform random variable on the consecutive integers a, a + 1, a + 2, . . . , b for a ≤ b:

f(x) = 1/(b-a)

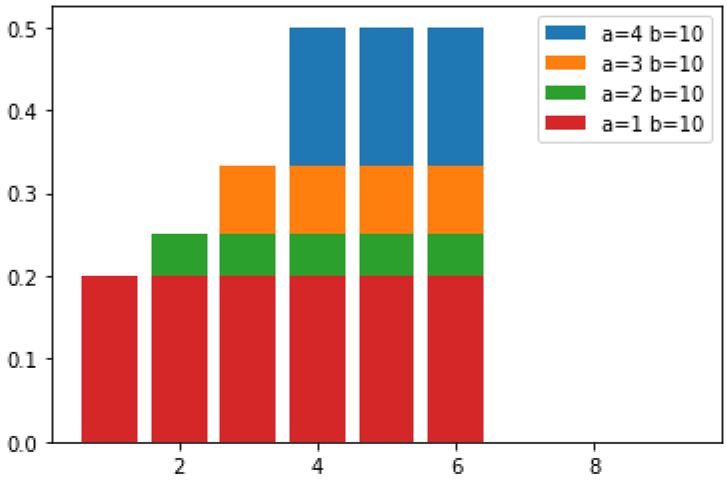
Example: Let X represent a random variable taking on the possible values of {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, and each possible value has equal probability

FIRST PART:
Parameters of the distribution: here a and b are the parameters

Distribution for different values of the parameters:

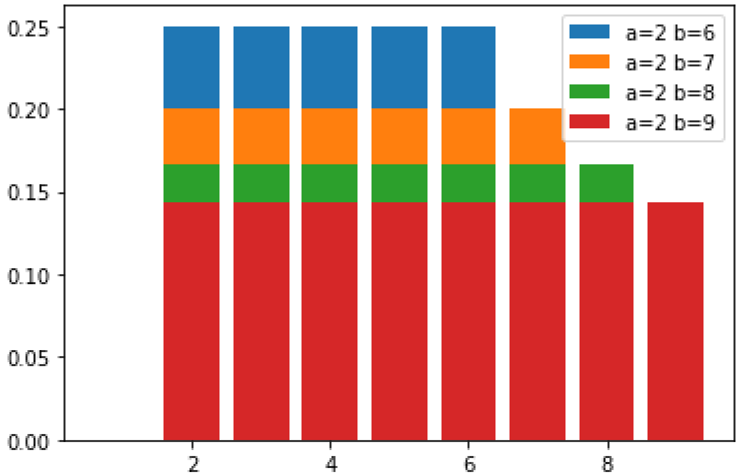
Considering different values for parameter a=[4,3,2,1]
Fixing the parameter b value: b=6 (we also have a condition : a<=b)

The distribution is as follows:



Considering different values for parameter b=[6,7,8,9]
Fixing the parameter “a” value as a=2

The distribution is as follows:



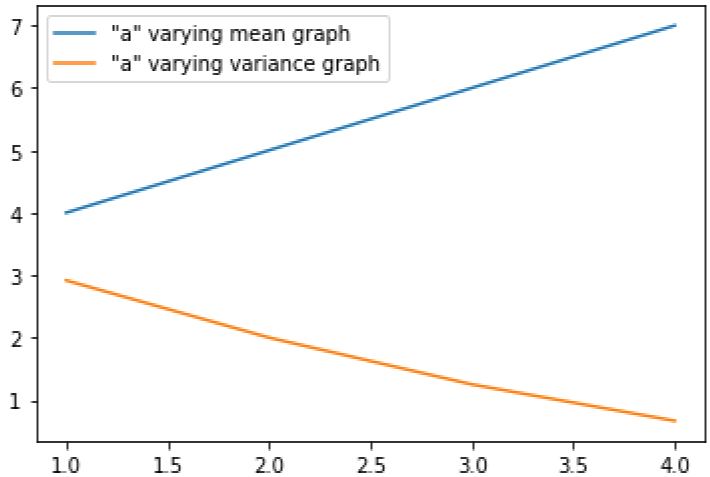
Mean and variance changes for different values of parameters:

Mean = μ = E(X) = b+a/2

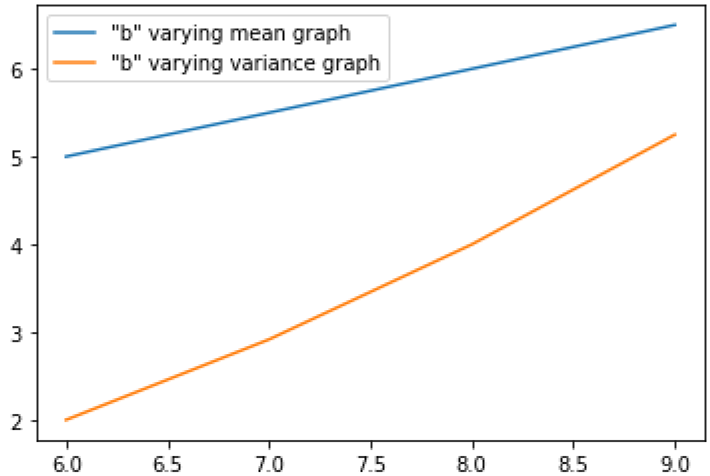
Variance = σ^2 =((b−a+1)^2−1)/12

Considering different values for parameter a=[4,3,2,1]
Fixing the parameter b value:
b=6 (we also have a condition : a<=b)

The graph is as follows:



Considering different values for parameter b=[6,7,8,9]
Fixing the parameter “a” value as a=2
The graph is as follows:



SECOND PART:

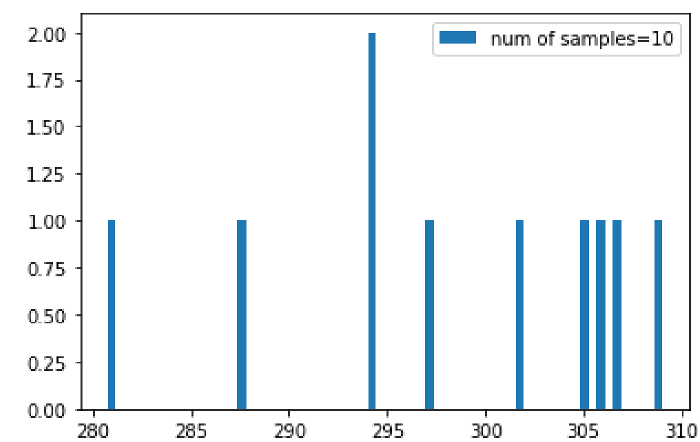
Mean of sample mean: 299.54300957565084

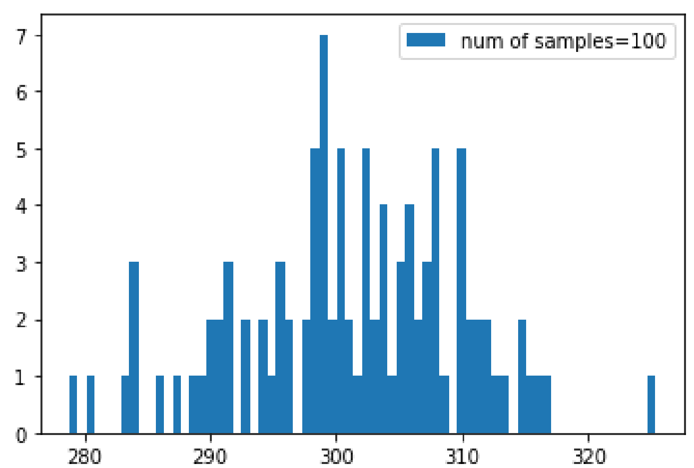
Variance of sample mean: 85.04151799402193

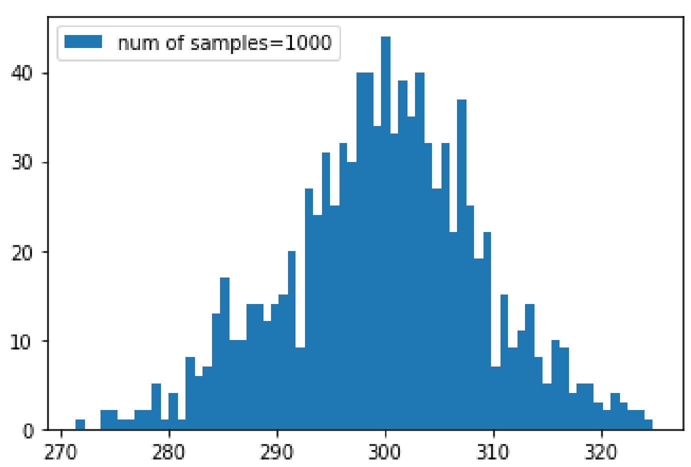
If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

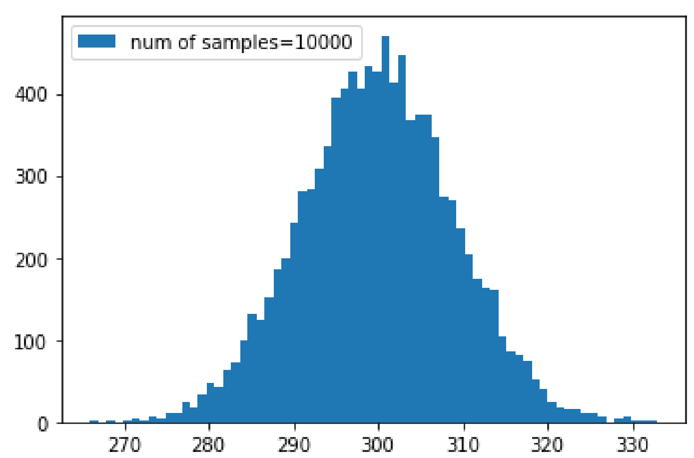
To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=10000,i.e 4th plot.The curve is in the shape of bell.









Continuous Distributions:

NORMAL DISTRIBUTION:

A normal (or Gaussian) distribution is a type of continuous probability distribution for a real valued random variable

Y = { 1/[σ * sqrt(2π)] } * e-(x - μ)2/2σ2

Example: Height of the population

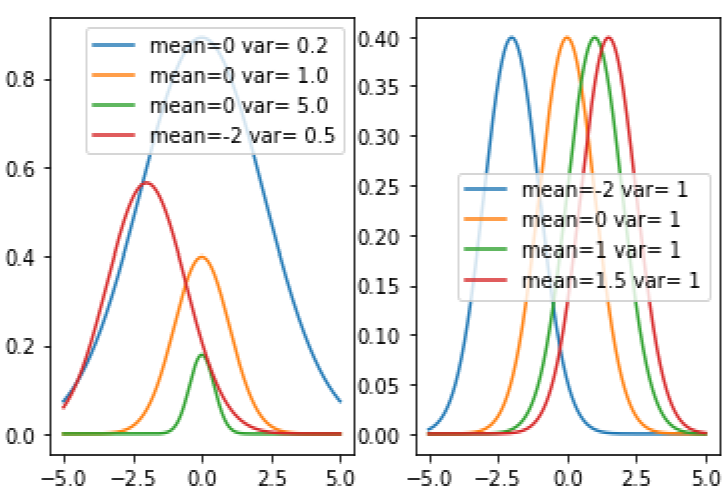
FIRST PART:

Parameters of the distribution: Parameters are Mean and variance

Distribution for different values of parameters:

Considering 4 different samples with mean= [-2,0,1,1.5] and variance =[0.2,1.0,5.0,0.5] respectively
The distribution is as follows:

Graphs when mean is varying and variance is constant (i.e, var=1) and graph when variance is varying and mean is constant (i.e, mean=0)



Mean and variance changes for different values of parameters:

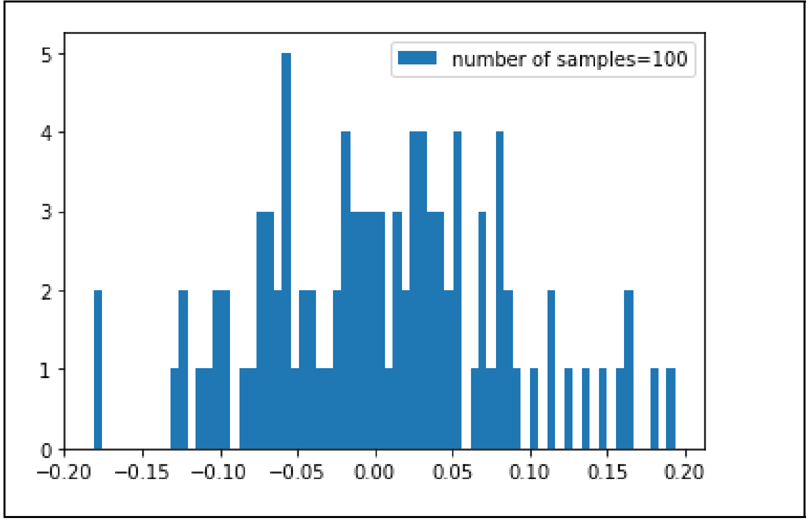
Here mean and variance are the parameters itself ,so mean and variance change will not be there.

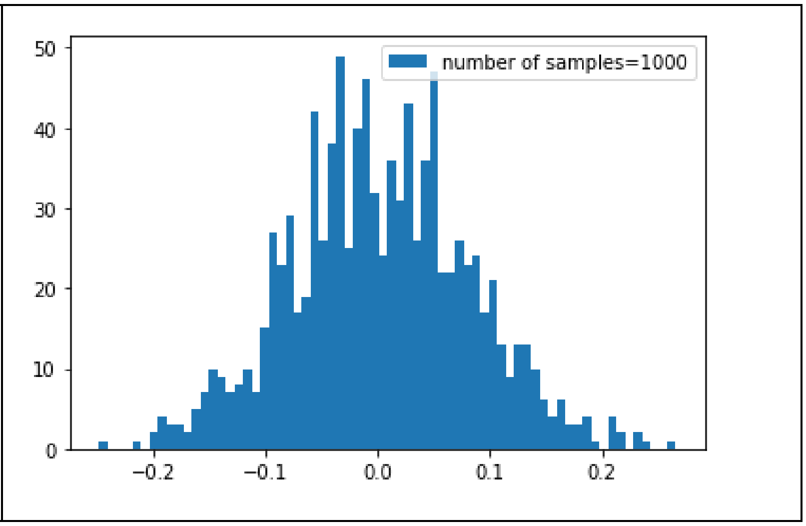
SECOND PART:
Mean of sample mean: 0.19964615013468887
Variance of sample mean: 0.0009925562649594072

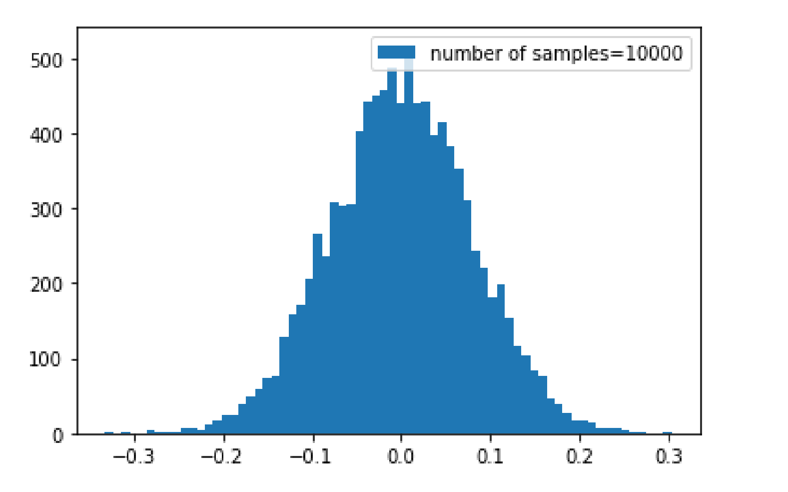
If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

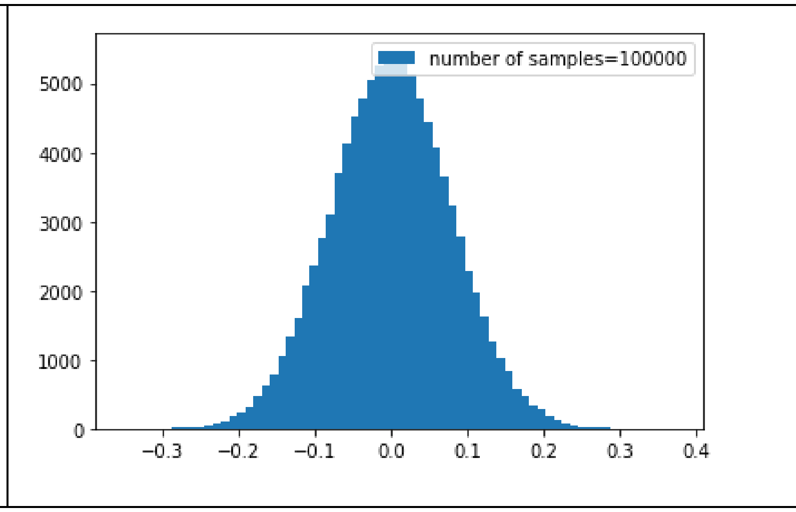
To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=10000,i.e 4th plot.The curve is in the shape of bell.









EXPONENTIAL DISTRIBUTION:

The exponential distribution is a continuous probability distribution used to model the time we need to wait before a given event occurs.

$f(x) = \lambda e^{-\lambda x}$

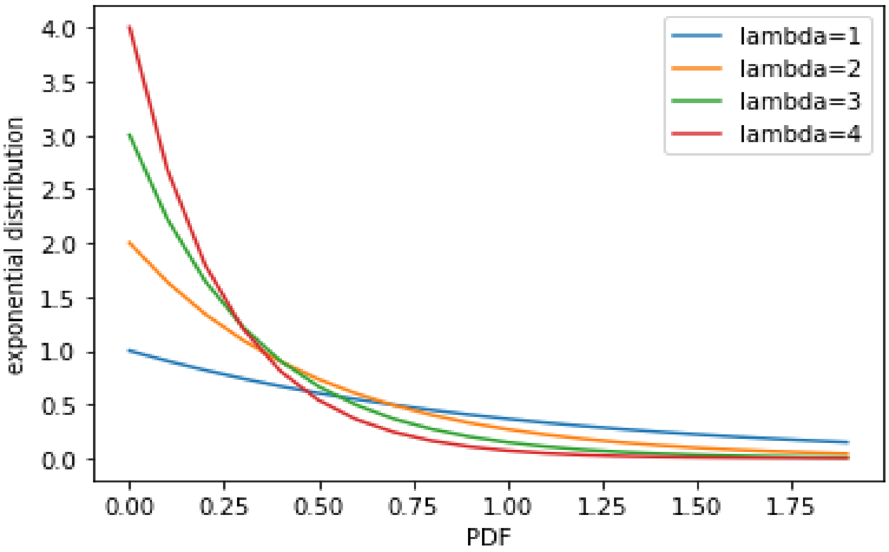
Where X is Random Variable with parameter λ

Example: The amount of time (beginning now) until an earthquake occurs has an exponential distribution.

FIRST PART:
Parameters of this distribution:
$\lambda > 0$ is the parameter of the distribution, often called the rate parameter.

Distribution for different values of parameters:
Considering 4 different samples with λ= [1,2,3,4] respectively

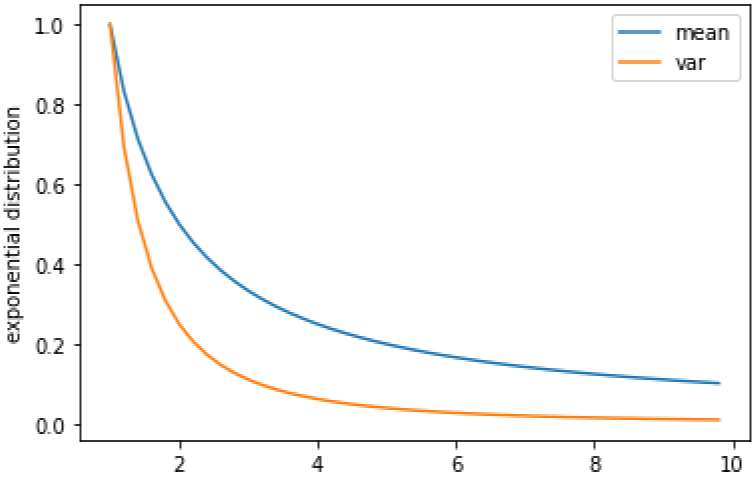
The distribution is as follows:



Mean and variance changes for different parameters:

Mean = 1/λ

Variance = 1/(λ^2)



SECOND PART:

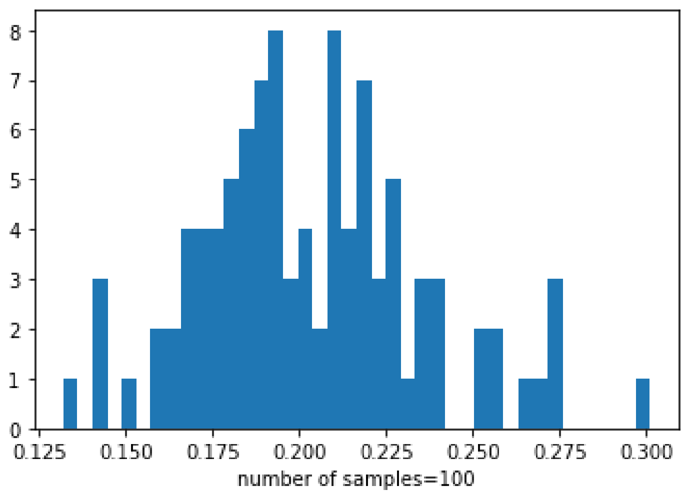
Mean of sample mean: 0.19964615013468887

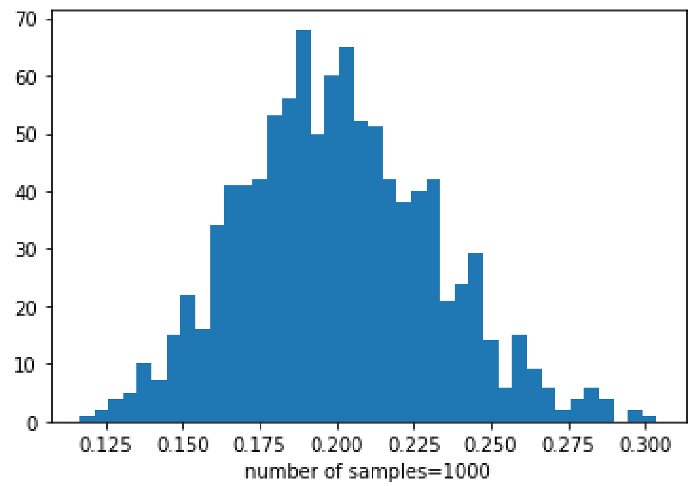
Variance of sample mean: 0.0009925562649594072

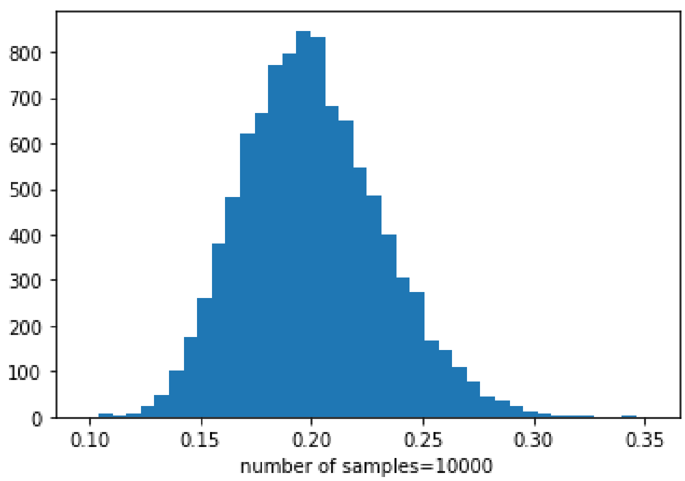
If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

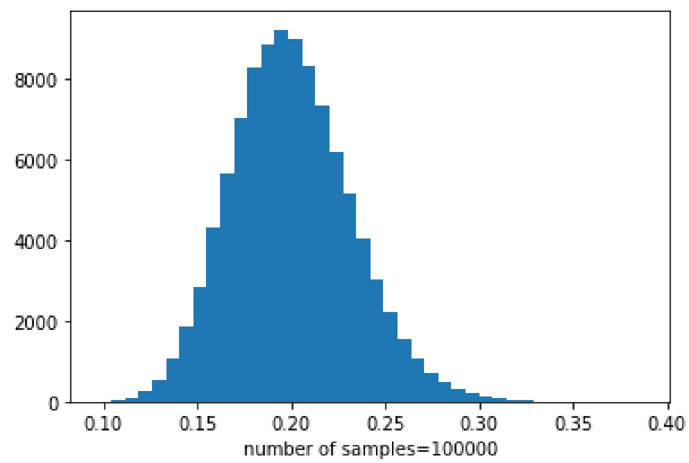
To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=100000,i.e 4th plot.The curve is in the shape of bell.









BETA DISTRIBUTION:

The Beta distribution is a type of probability distribution which represents all the possible values of probability.The most common use of this distribution is to model the uncertainty about the probability of success of a random experiment.

$$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$$

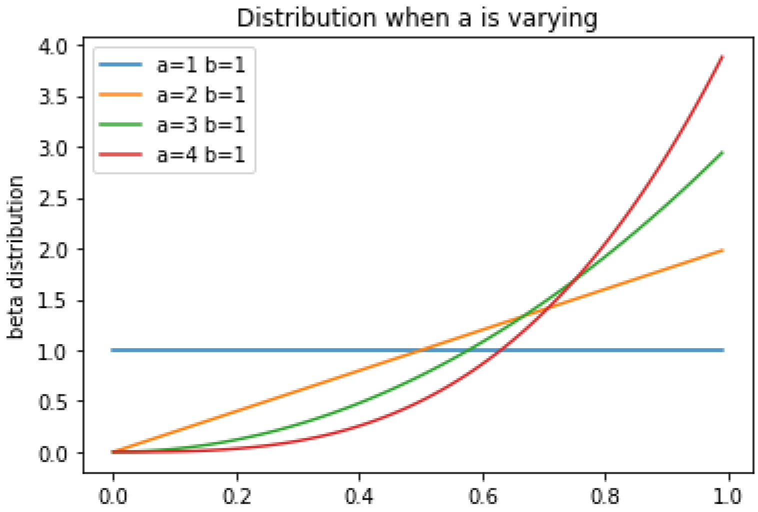
Example: Suppose, if in a basket there are balls which are defective with a Beta distribution of alpha=5 and beta=2. For finding the probability of defective balls in the basket from 20% to 30% this distribution can be used.

FIRST PART:

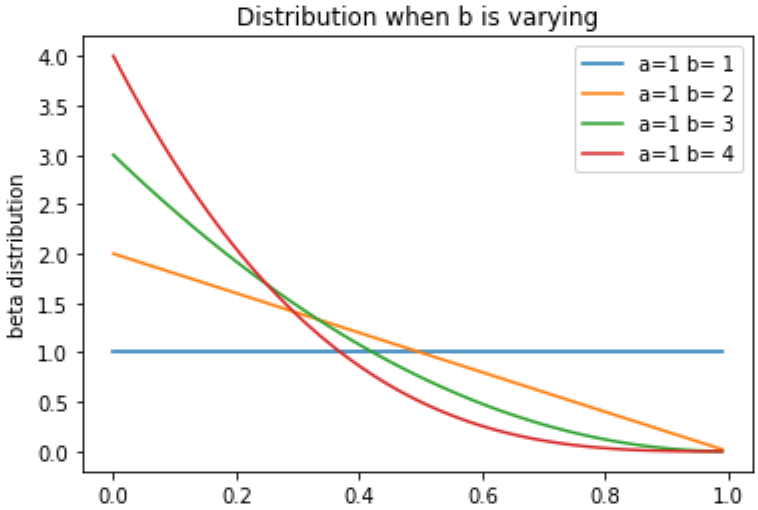
Parameters of distribution: Parameters are alpha and beta where alpha is shape parameter and beta is scale parameter.

Distribution for different values of parameters:
Considering alpha=[1,2,3,4] and beta=[1,2,3,4] respectively

The distribution is as follows:
When alpha is varying and beta is constant(i.e, beta=1):



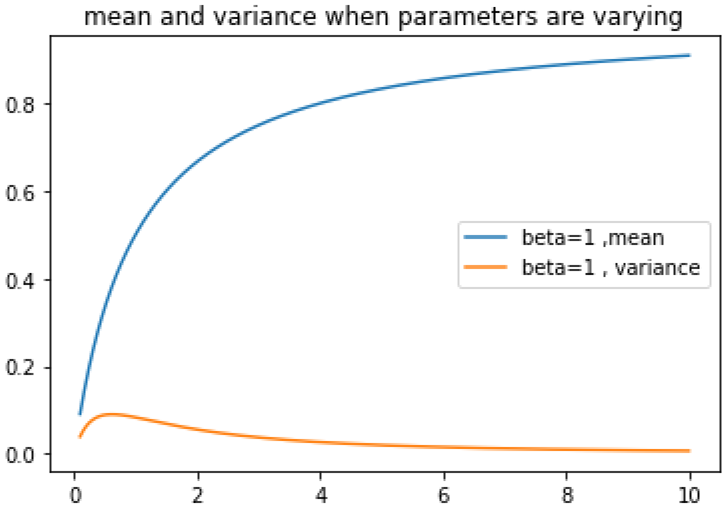
When beta is varying and alpha is constant(i.e, a=1):



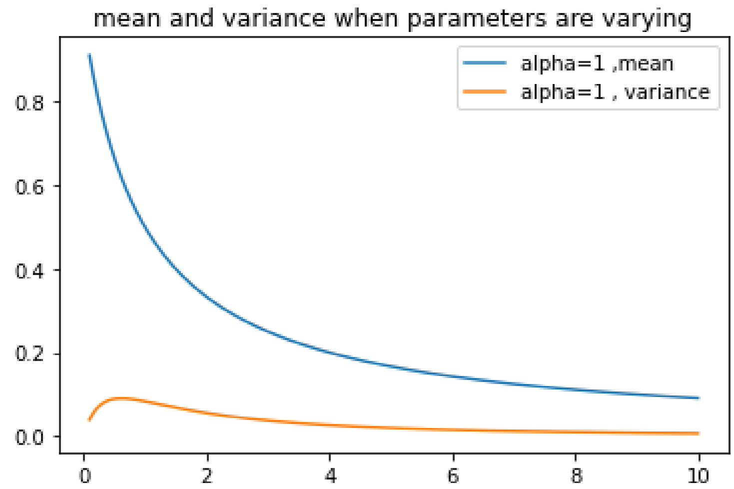
Mean and variance changes for different parameters:

Mean = alpha/(alpha+beta)
Variance= (alpha*beta)/((alpha+beta)2) * (alpha+beta+1)**

When alpha is varying and beta is constant(i.e, beta=1):



When beta is varying and alpha is constant:



SECOND PART:

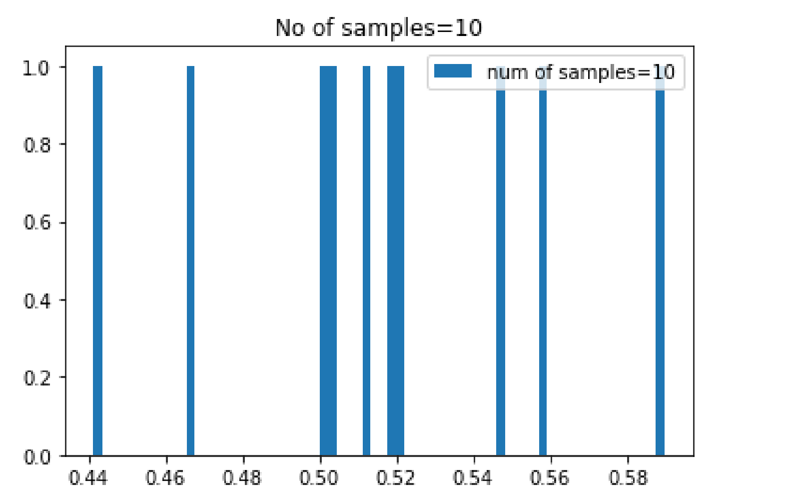
Mean of sample mean: 1.0008893537850958

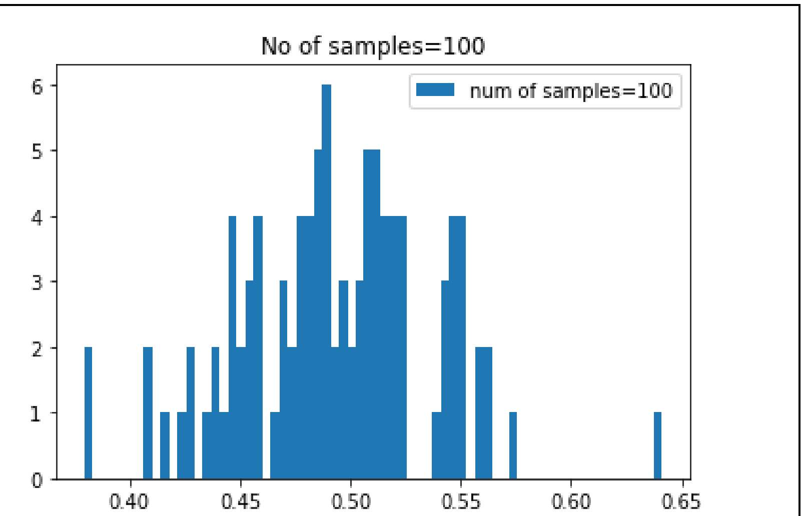
Variance of sample mean: 0.010058793093934395

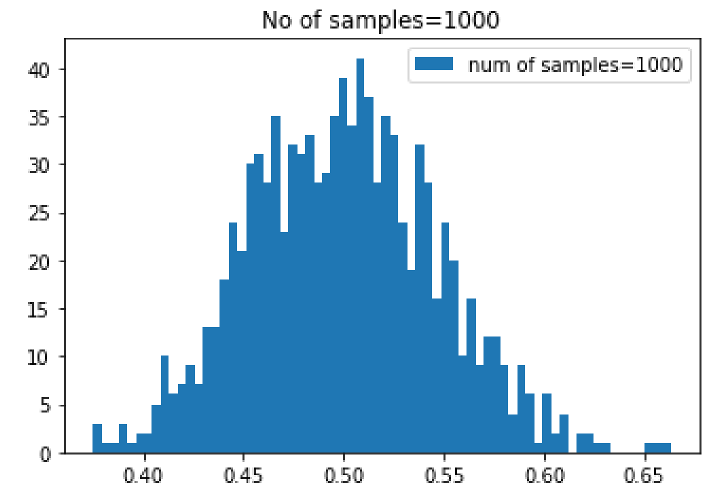
If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

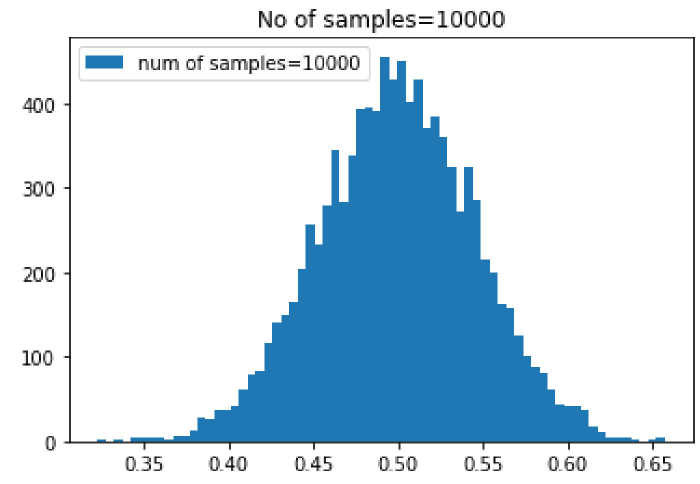
To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=10000,i.e 4th plot.The curve is in the shape of bell.









GAMMA DISTRIBUTION:

The Gamma distribution describes the distribution of waiting times until a specific number of independent events (typically deaths) have occurred.

$$P(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}.$$

Example:if the average mortality rate is one individual per five days (rate=1/5 or scale=5), then a Gamma distribution could be used to describe the distribution of expected waiting time before 10 individuals were dead.

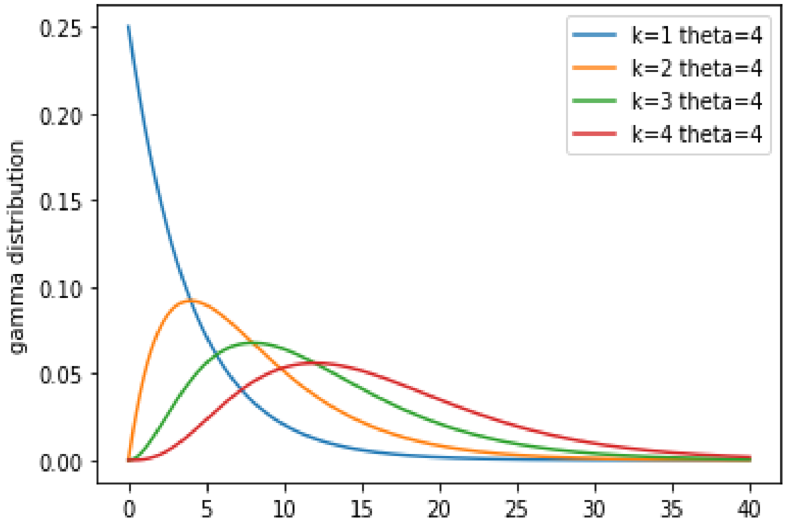
FIRST PART:

Parameters of this distribution:

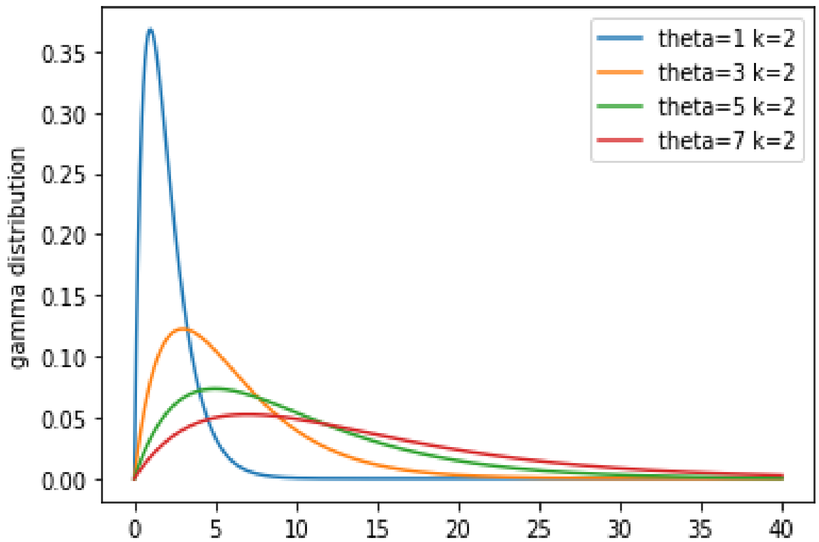
- θ : The rate of events happening which follows the Poisson process.
- k : The number of events for which you are waiting to occur.

Distribution for different values of parameters:
Considering k =[1,2,3,4] and θ=[1,3,5,7] respectively.
The distribution is as follows:

When k is varying and θ is constant(i.e, θ=4):



When θ is varying and k is constant(i.e, k=2):

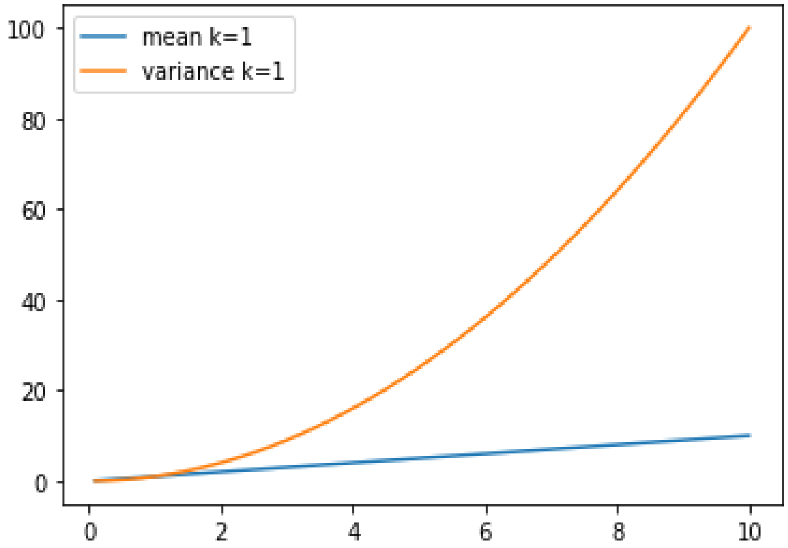


Mean and variance changes for different parameters:

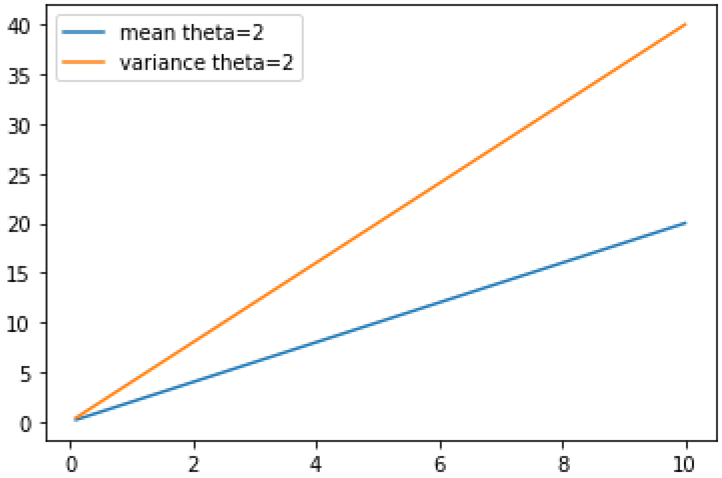
Mean = k*θ

Variance = k*θ*θ

When θ is varying and k is fixed i.e(k=1):



When k is varying and θ is fixed i.e(θ=2):

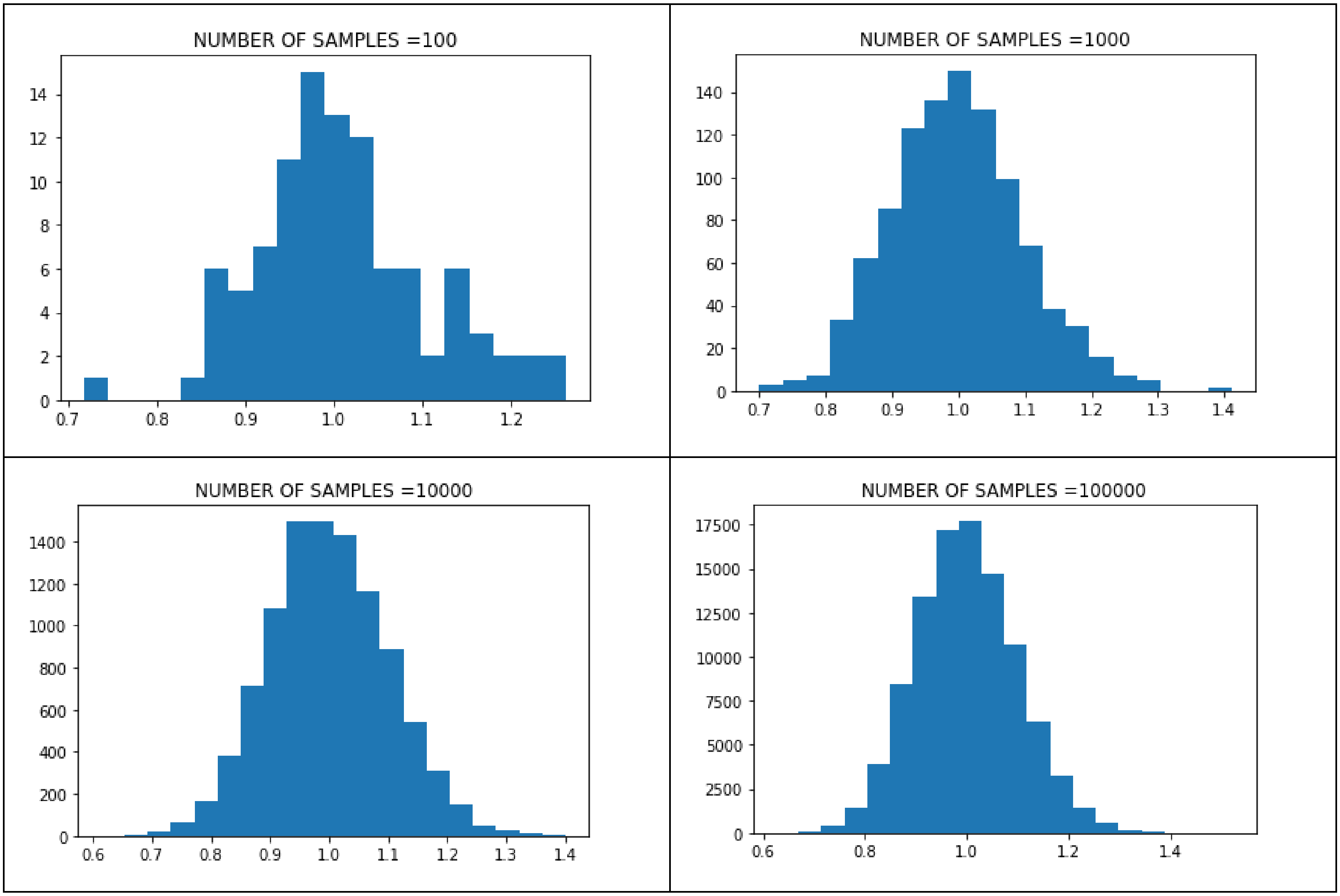


SECOND PART:
Mean of sample mean: 0.9981473035550558
Variance of sample mean: 0.009875503021356773

If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=100000,i.e 4th plot.The curve is in the shape of bell.



LOGNORMAL DISTRIBUTION:

Log-normal (or lognormal) distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then Y = ln(X) has a normal distribution.

$$\mathcal{N}(\ln x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \quad x > 0.$$

Example:growth rates or chemical concentrations might naturally operate on logarithmic or exponential scales.Consequently, when such data are collected on a linear scale, they might be expected to follow a log-normal distribution.

FIRST PART:

Parameters of this distribution:

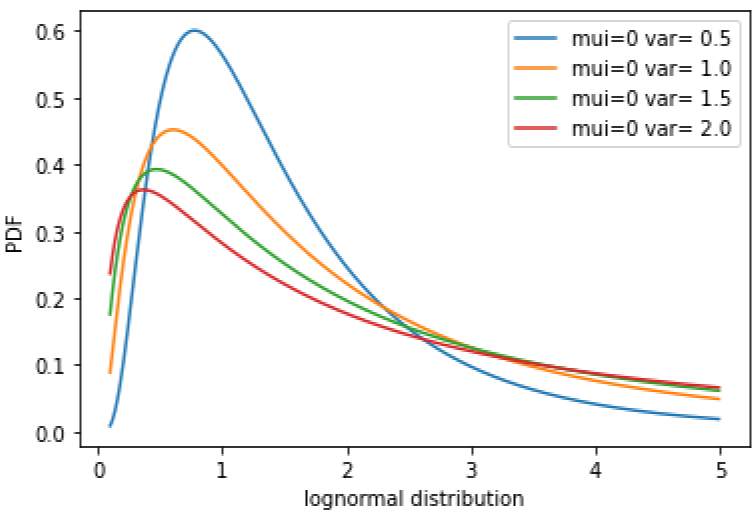
- μ - the mean. This defines the center of the distribution, the location of the peak.
- σ^2- the variance (or σ, the standard deviation) which defines the variability or spread of values around the mean.

Distribution for different values of parameters:

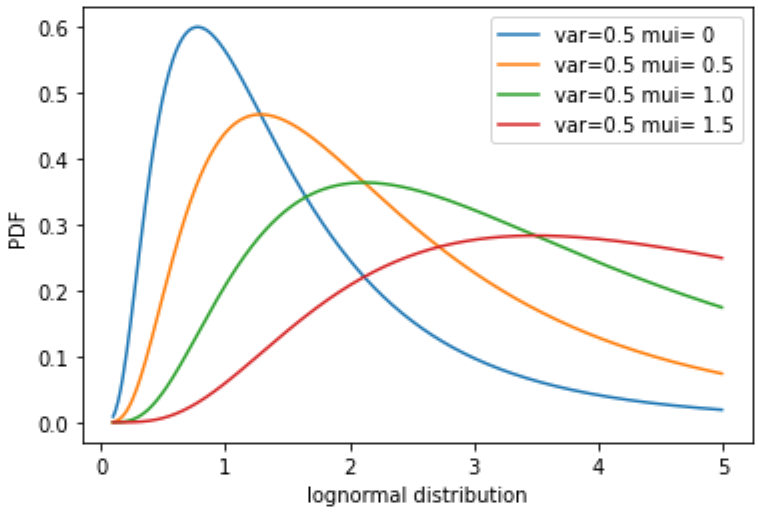
Considering μ =[0,0.5,1.0,1.5] and σ^2=[0.5,1.0,1.5,2.0] respectively.

The distribution is as follows:

When σ^2 is varying and μ is constant(i.e, μ=0):



When μ is varying and σ^2 is constant(i.e, σ^2=0.5):

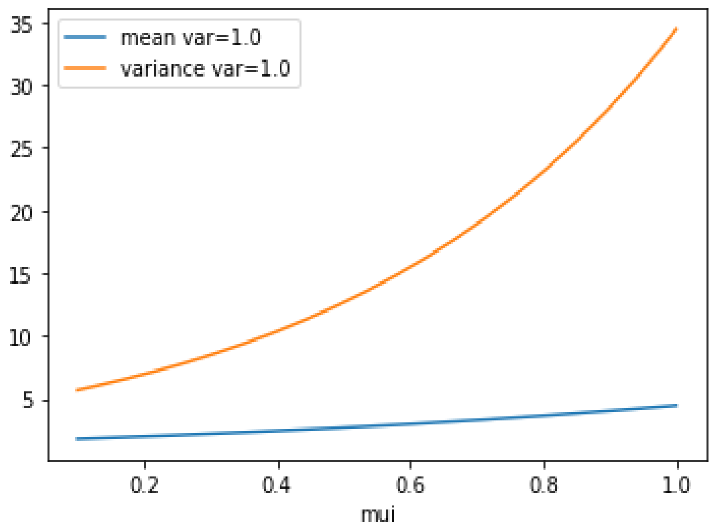


Mean and variance changes for different parameters:

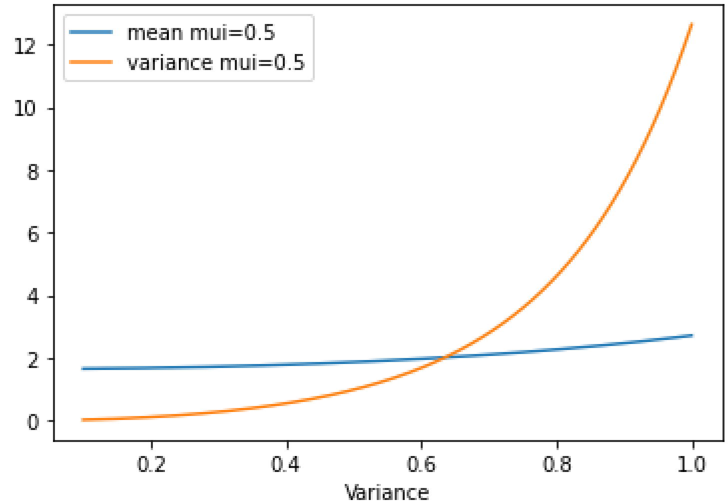
Mean = exp(μ + σ^2/2)

Variance = (exp(σ^2)-1)*exp(2μ + σ^2)

When μ is varying and σ^2 is fixed i.e(σ^2=1.0):



When σ^2 is varying and μ is fixed i.e(μ=0.5):



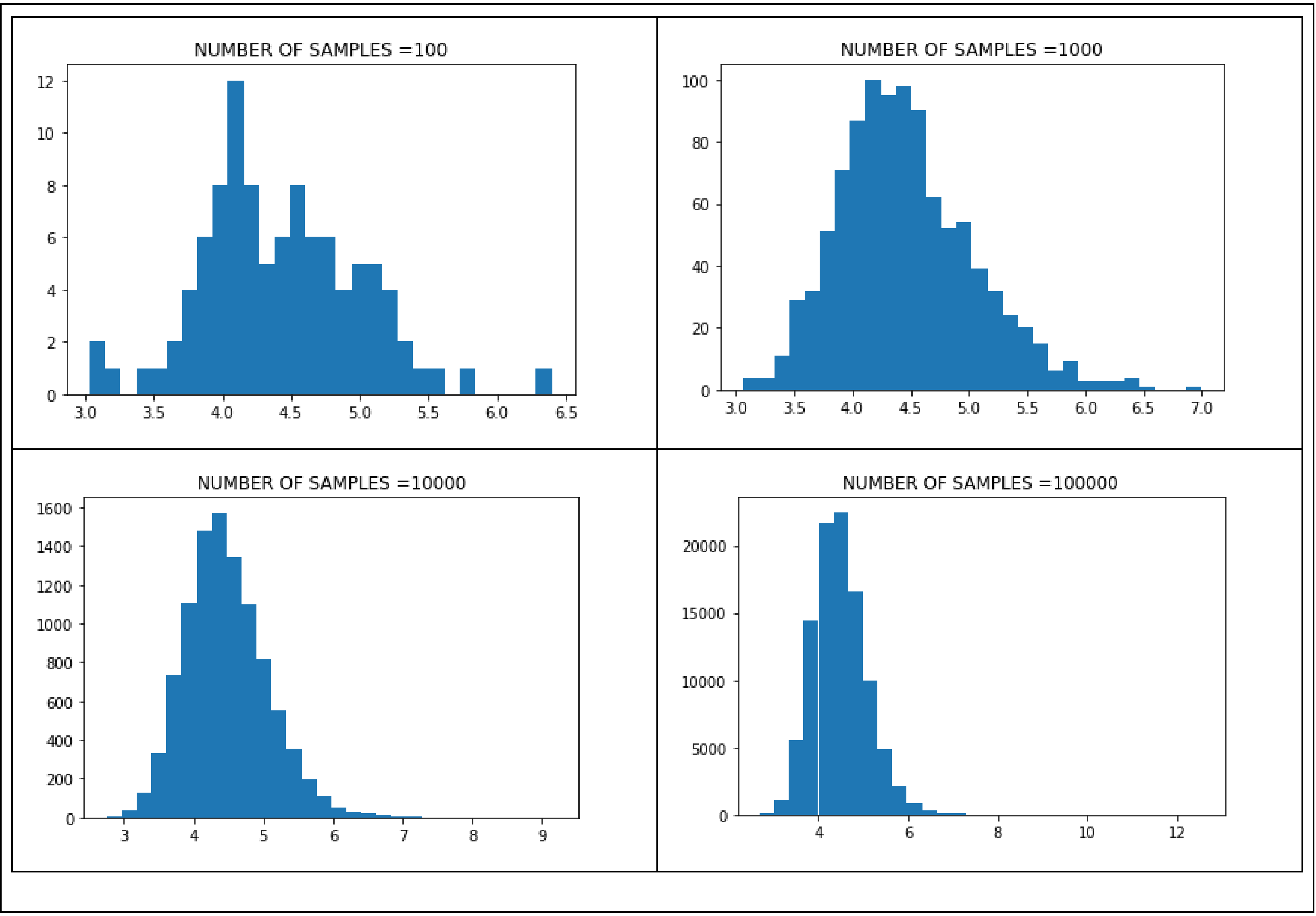
SECOND PART:

Mean of sample mean: 4.507422955127459
Variance of sample mean: 0.365882629909888

If the sample is random and sample size is large then the sample mean would be a good estimate of the population mean.

To prove CLT:

Here on plotting samples mean histograms by varying the number of samples every time, we can see that the graph slowly reaches and follows normal distribution . We can see that observation when n=100000,i.e 4th plot.The curve is in the shape of bell.


NUMBER OF SAMPLES =100
NUMBER OF SAMPLES =1000
NUMBER OF SAMPLES =10000
NUMBER OF SAMPLES =100000