SDA Simulation Assignment -1

Group Number - 33

Team Details :

M Sai Amulya	-	S20170010099
G Sreeja	-	S20170010047
P Deepika Sowmya	-	S20170010110
B Gayatri Shivani	-	S20170010029

Discrete Distributions:

A discrete distribution describes the probability of occurrence of each value of a discrete random variable. With a discrete probability distribution, each possible value of the discrete random variable can be associated with a non-zero probability.

Binomial Distribution

The binomial distribution is the discrete probability distribution of the number of successes in a sequence of n independent experiments, each with its own Boolean-valued outcome.

Pmf of Distribution:

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

Parameters:

n - Independent Experiments,

p - Probability that an experiment is true,

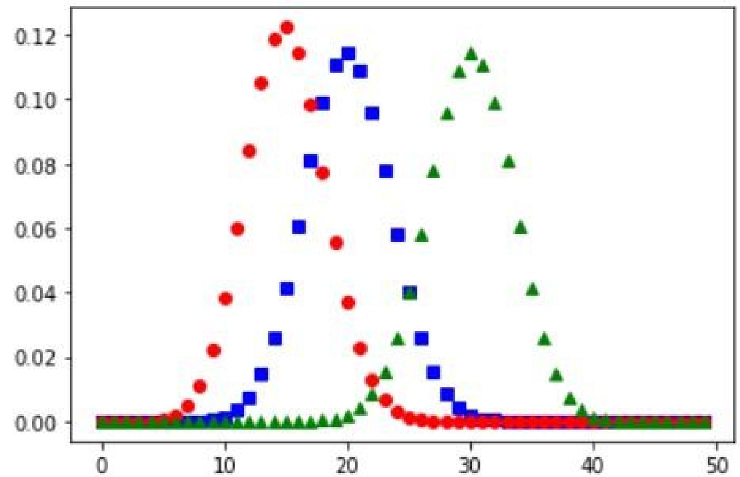
q - Probability that an experiment is false.

Mean : **E[x] =** $n * p$ **variance: Var[X] =** n*p*(1-p)

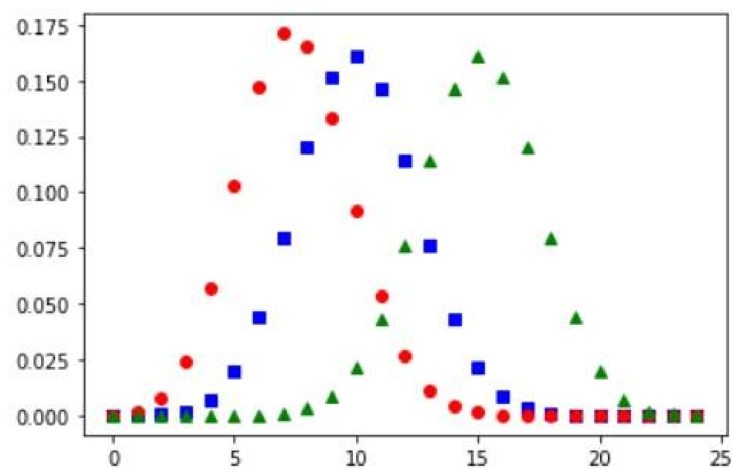
Part 1:

Distribution for different values of parameters :

- We have used scipy.stats.binom.rvs(n,p,size) to generate the binomial distribution.
- We have plotted graphs for different p and n values.
- The below graph is for n = 50 and with different p and q values.
- Graph with color blue has p = 0.4 , q=0.6, red has p = 0.3, q= 0.7 and green has p=0.6, q=0.4.



- The below graph is for n = 25 and with different p and q values.
- Graph with color blue has p = 0.4 , q=0.6, red has p = 0.3, q= 0.7 and green has p=0.6, q=0.4.



Mean and Variance for different values of parameters :

- Here we considered p values ranging from 0.1 to 0.9 with n value as 50.
- The table below shows the theoretical and population mean and variance, values are calculated using the formulas

	p	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	0.1	5.076	4.586	5.0	4.5
1	0.2	9.991	7.285	10.0	8.0
2	0.3	15.161	10.971	15.0	10.5
3	0.4	20.134	12.084	20.0	12.0
4	0.5	25.015	13.037	25.0	12.5
5	0.6	30.109	11.621	30.0	12.0
6	0.7	34.892	11.054	35.0	10.5
7	0.8	39.986	8.636	40.0	8.0
8	0.9	44.893	4.439	45.0	4.5

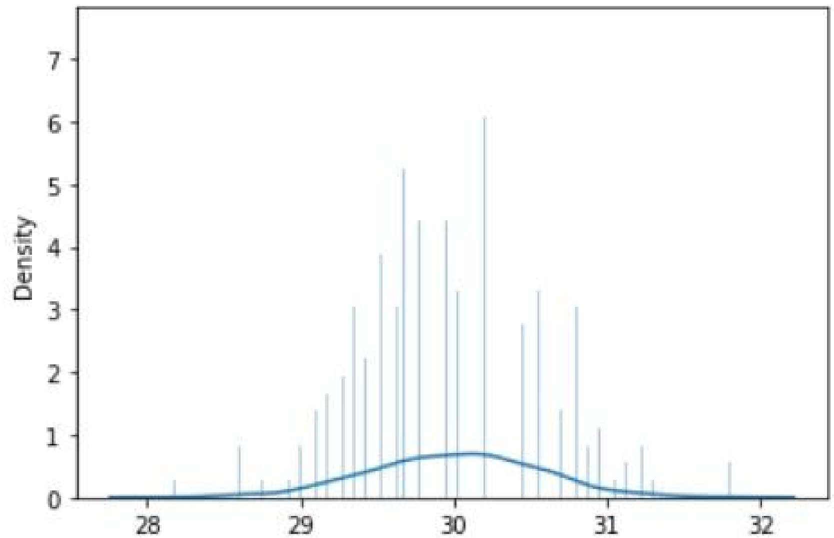
- Here we considered p values ranging from 0.1 to 0.9 with n value as 25.
- The table below shows the theoretical and population mean and variance, values are calculated using the formulas

	p	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	0.1	2.495	2.278	2.5	2.25
1	0.2	5.120	4.124	5.0	4.00
2	0.3	7.474	5.378	7.5	5.25
3	0.4	9.789	5.749	10.0	6.00
4	0.5	12.445	6.522	12.5	6.25
5	0.6	14.924	5.706	15.0	6.00
6	0.7	17.479	5.202	17.5	5.25
7	0.8	19.919	3.933	20.0	4.00
8	0.9	22.476	2.364	22.5	2.25

Part 2:

- A binomial distribution with n =50 and p=0.6 is considered.
- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean - 29.98 , population variance - 11.947600000000001
- For the sample mean distribution:

- Mean : 30.0039
- Variance : 1.95476513170617



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.

$$p_X(x) = \begin{cases} (1-p)^x p & \text{if } x \in R_X \\ 0 & \text{if } x \notin R_X \end{cases}$$

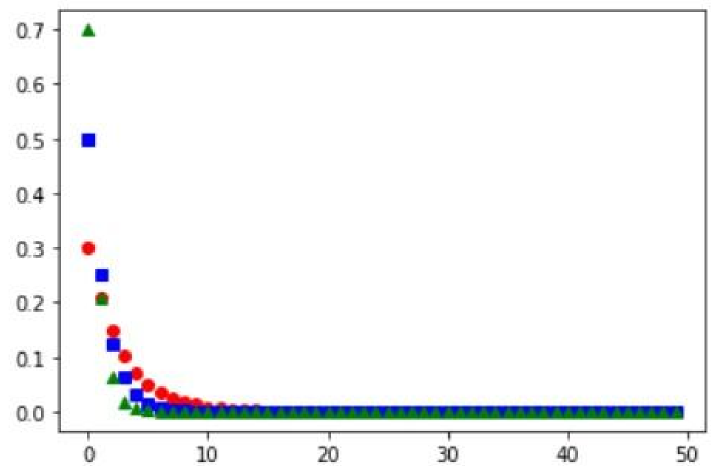
Parameters: p is the success probability

Mean : **E[x] =** $\frac{1-p}{p}$ **variance: Var[X] =** $\frac{1-p}{p^2}$

Part 1:

Distribution for different values of parameters :

- We have used np.random.geometric(p,size) to generate distribution.
- We have plotted graphs for different p values.
- Graph with color blue has p = 0.5, red has p = 0.3 and green has p=0.7.



Mean and Variance for different values of parameters :

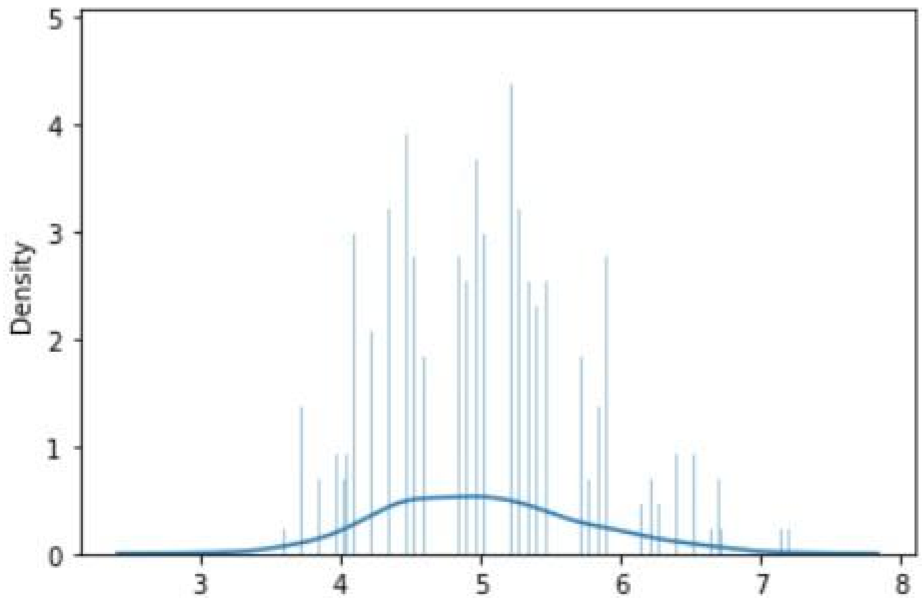
- Here we considered p values ranging from 0.1 to 0.9 with n value as 50.
- The table below shows the theoretical and population mean and variance, values are calculated using the formulas.

	p	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	0.1	9.9809	86.999500	9.000000	90.000000
1	0.2	5.0266	21.399400	4.000000	20.000000
2	0.3	3.3581	9.157011	2.333333	7.777778
3	0.4	2.4958	4.613200	1.500000	3.750000
4	0.5	2.0216	3.163400	1.000000	2.000000
5	0.6	1.6623	2.110544	0.666667	1.111111
6	0.7	1.4269	1.624116	0.428571	0.612245
7	0.8	1.2538	1.324400	0.250000	0.312500
8	0.9	1.1088	1.115746	0.111111	0.123457

Part 2:

- A Geometric distribution with p=0.2 is considered.
- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean 5.0024 , population variance 19.98459

- For the sample mean distribution:
    - Mean : 5.0228
    - Variance : 3.22995



Poisson Distribution

Poisson distribution expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event.

$$P(X) = \frac{e^{-\mu}\mu^{x}}{x!}$$

Parameters: μ

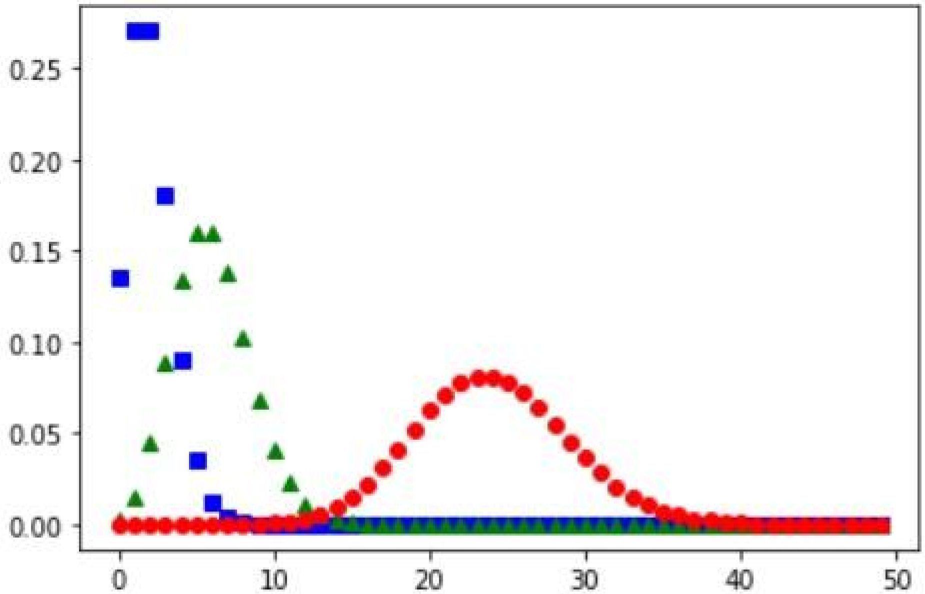
Mean : **E[x] =** μ **variance: Var[X] =** μ

Part 1:

Distribution for different values of parameters :

- We have used scipy.stats.poisson.rvs(lambda,size) to generate Poisson distribution.
- We have plotted graphs for different λ values.

- Graph with color blue has λ = 2, red has λ= 4 and green has λ=6.



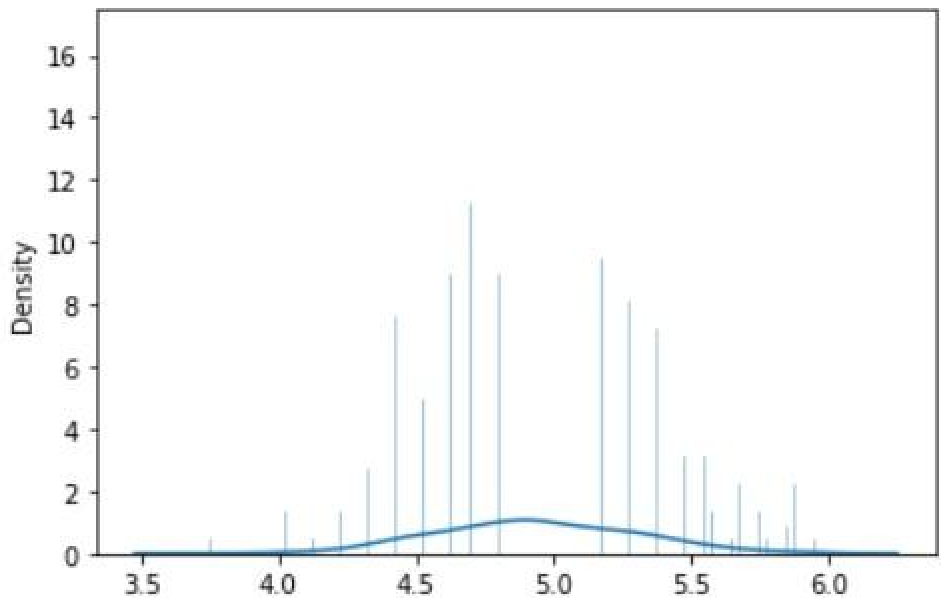
Mean and Variance for different values of parameters :

- Here we considered λ values ranging from 1 to 9 .
- The table below shows the theoretical and population mean and variance, values are calculated using the formulas.

	p	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	1	0.998	0.960	1	1.0
1	2	1.979	1.837	2	2.0
2	3	3.024	2.942	3	3.0
3	4	3.981	4.073	4	4.0
4	5	4.923	4.937	5	5.0
5	6	5.968	5.884	6	6.0
6	7	6.979	7.145	7	7.0
7	8	7.919	8.075	8	8.0
8	9	8.957	8.741	9	9.0

Part 2:

- A Poisson distribution with λ= 5 is considered.
- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean 4.935 , population variance 5.410775
- For the sample mean distribution:
    - Mean : 4.9385
    - Variance : 0.84831



Negative Binomial Distribution

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.

$$f(x) = \left(\begin{array}{c} x-1 \\ r-1 \end{array} \right) (1-p)^{x-r} p^r \quad \text{for} \quad x = r, r+1, r+2, \ldots$$

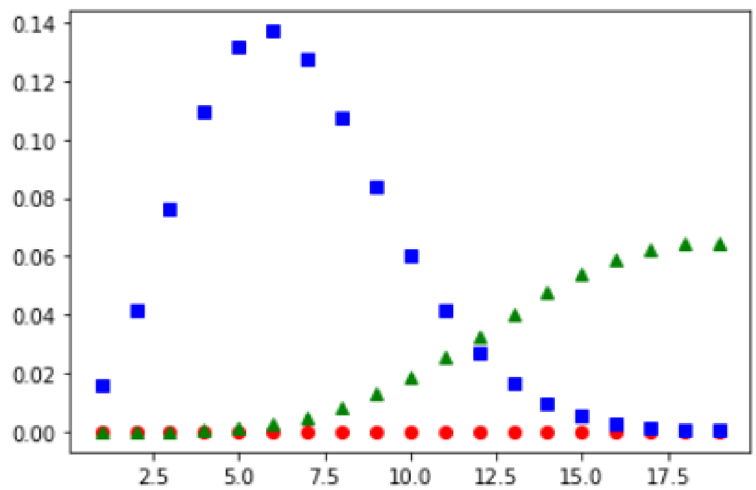
Parameters:
r= No.of failures until the experiment is stopped
p=Success probability

Mean : **E[x] =** $\frac{r}{p}$ **variance: Var[X] =** $\frac{r(1-p)}{p^2}$

Part 1:

Distribution for different values of parameters :

- We have used scipy.stats.nbinom.rvs(r,p,size) to generate distribution.
- We have plotted graphs for different parameters.



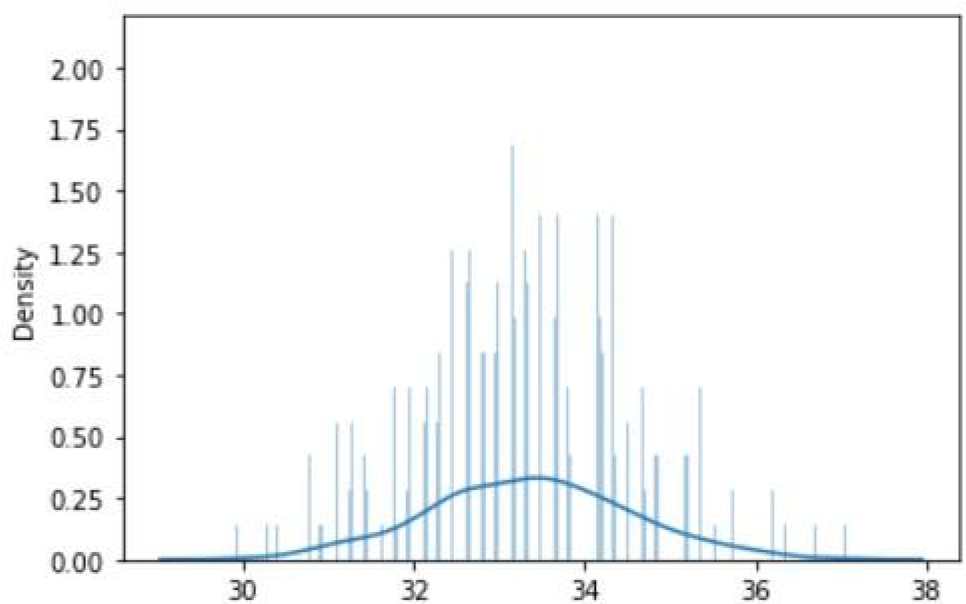
Mean and Variance for different values of parameters :

- Here we considered p values ranging from 0.1 to 0.9 and n value as 20 .
- The table below shows the theoretical and experimental mean and variance, values are calculated using the formulas.

	p	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	0.1	180.755	1603.671000	180.0	1800.000000
1	0.2	80.434	401.822000	80.0	400.000000
2	0.3	46.535	160.972111	46.66666666666667	155.555556
3	0.4	29.950	77.798000	30.0	75.000000
4	0.5	20.238	41.630000	20.0	40.000000
5	0.6	13.422	23.797778	13.333333333333336	22.222222
6	0.7	8.665	12.479531	8.571428571428573	12.244898
7	0.8	4.956	5.920000	5.0	6.250000
8	0.9	2.183	2.389049	2.2222222222222223	2.469136

Part 2:

- A Poisson distribution with n=50,p=0.6 is considered.
- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean 33.341 Population variance 55.084719
- For the sample mean distribution:
  - Mean : 33.362175
  - Variance : 8.960044288454833



Discrete Uniform Distribution:

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*.

Pmf: $\frac{1}{b-a}$

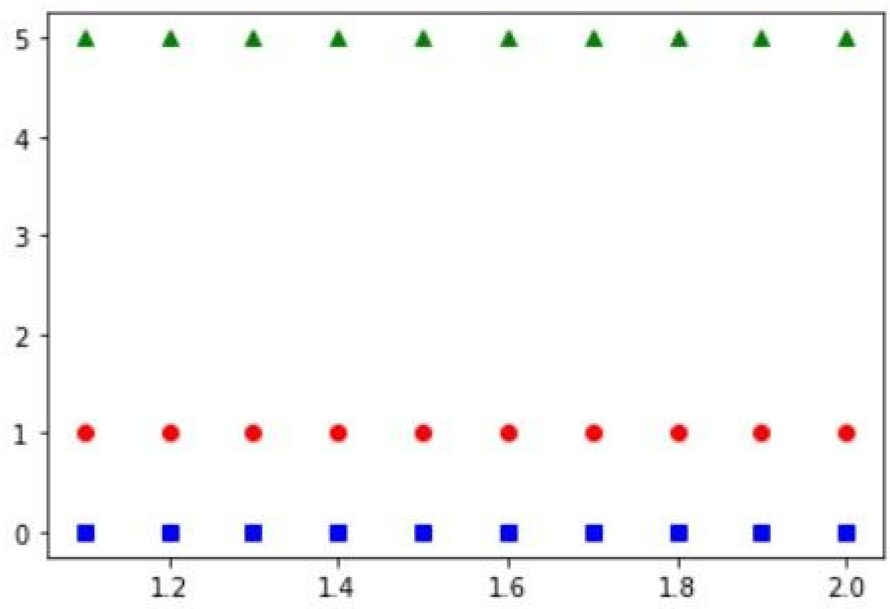
Parameters: a and b

Mean : **E[x] =** $\frac{b+a}{2}$ **variance: Var[X] =** $\frac{(b-a+1)^2-1}{12}$

Part 1:

Distribution for different values of parameters :

- We have used scipy.stats.uniform.rvs(a,b,size) to generate the Discrete Uniform distribution.
- The below graph is for different a and b values.
- Graph with color blue has a =1.2 , b=1.4, red has a =2, b= 2.6 and green has a=0.4, b=1.6.



Mean and Variance for different values of parameters :

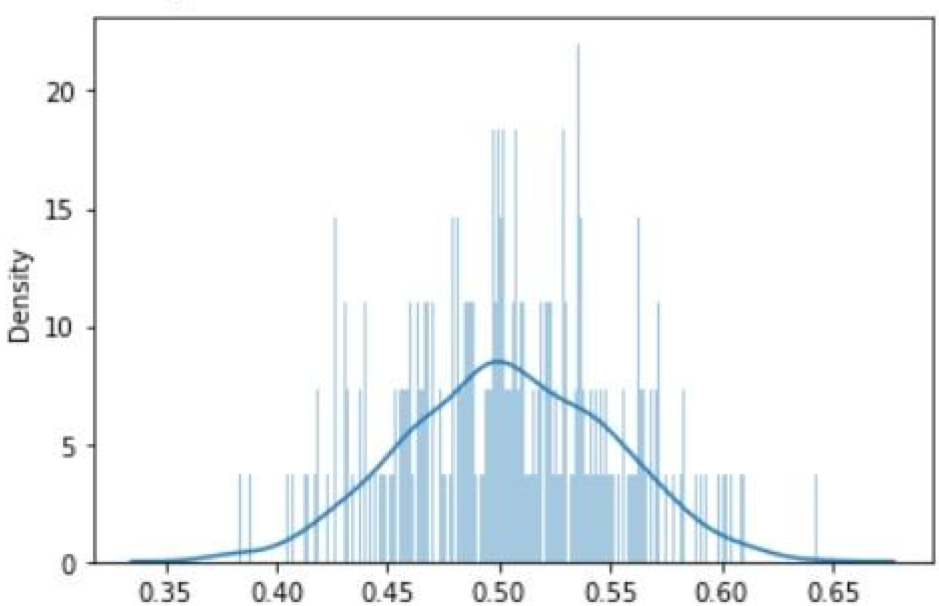
- For different values of a and b uniform distribution is considered.

- The table below shows the theoretical and population mean and variance, values are calculated using the formulas.

	a	b	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	2.0	2.6	0.519769	3.249620	2.3	0.030000
1	0.4	1.6	0.510549	0.322088	1.0	0.120000
2	1.2	1.4	0.500833	0.723460	1.3	0.003333

Part 2:

- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean 0.50392 and population variance 0.085842
- For the sample mean distribution:
    - Mean : 0.50429
    - Variance : 0.013533



Continuous Distributions:

A continuous distribution describes the probabilities of the possible values of a continuous random variable. A continuous random variable is a random variable with a set of possible values (known as the range) that is infinite and uncountable.

Gaussian Distribution:

The normal distribution is a probability function that describes how the values of a variable are distributed. It is a symmetric distribution where most of the observations cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions.

Pdf of distribution :

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

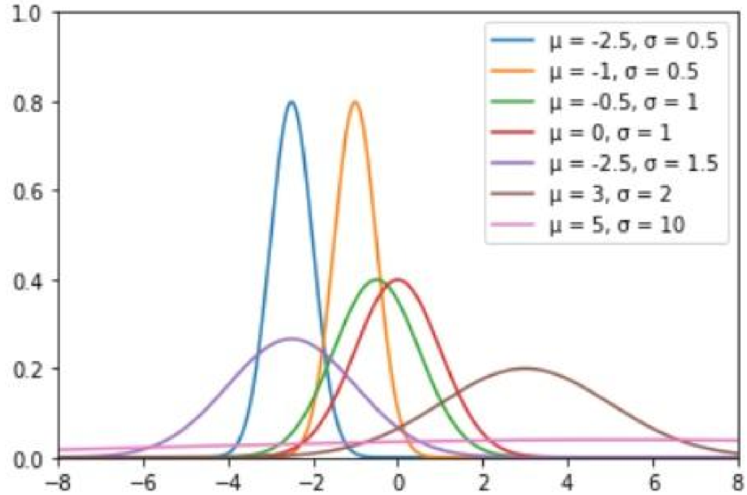
Parameters: μ,σ

Mean : E[x] = μ **variance:** Var[X] = σ^2

Part 1:

Distribution for different values of parameters :

- We have used scipy.stats.norm.rvs(μ, σ,size) to generate the Normal distribution.
- We have plotted graphs for different $\mu$and σ values as shown in the graph



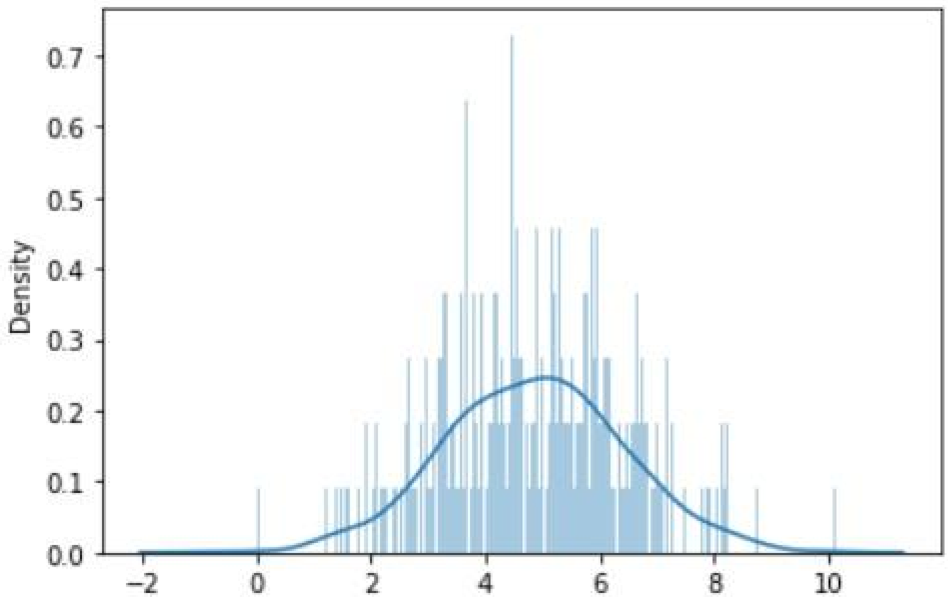
Mean and Variance for different values of parameters :

- Different values of $\mu$and σ are considered.
- The table below shows the theoretical and experimental mean and variance, values are calculated using the formulas.

	Param1(μ)	Param2(σ)	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	-2.5	0.5	-2.497985	0.250590	-2.5	0.25
1	-1.0	0.5	-1.001845	0.249700	-1.0	0.25
2	-0.5	1.0	-0.492714	1.004790	-0.5	1.00
3	0.0	1.0	-0.000348	0.989119	0.0	1.00
4	-2.5	1.5	-2.519313	2.227429	-2.5	2.25
5	3.0	2.0	3.002980	4.017649	3.0	4.00
6	5.0	10.0	5.052939	100.764728	5.0	100.00

Part 2:

- A Normal distribution with μ =5 and σ=10 is considered.
- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean 4.92886 , population variance 98.5664
- For the sample mean distribution:
  - Mean : 4.8607
  - Variance : 15.4904



Exponential Distribution:

The exponential distribution is the probability distribution of the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate.

Pdf of distribution :

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

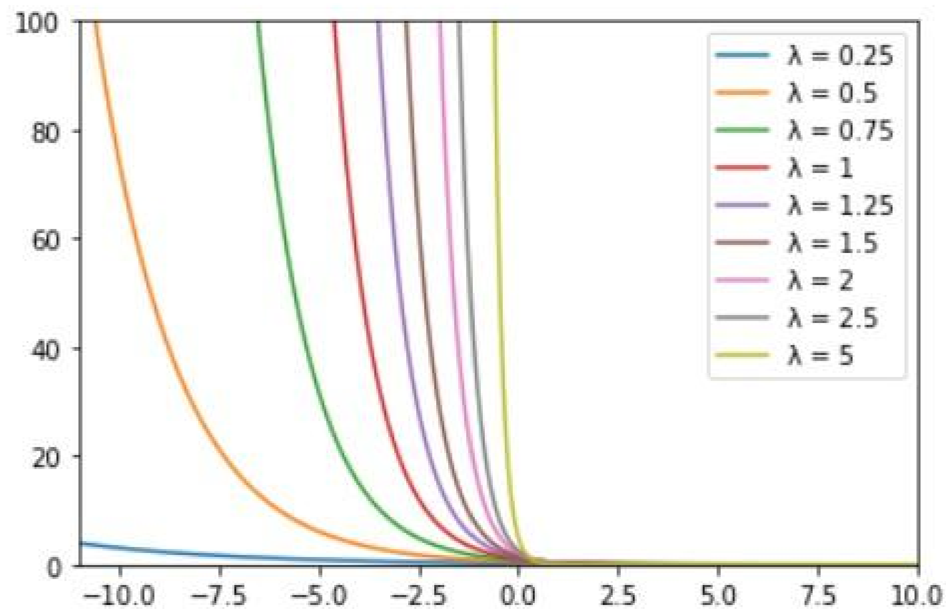
Parameters : λ

Mean : **E[x] =** $\frac{1}{\lambda}$ **variance: Var[X] =** $\frac{1}{\lambda^2}$

Part 1:

Distribution for different values of parameters :

- We have used scipy.stats.expon.rvs(λ,size) to generate the Exponential distribution.
- We have plotted graphs for different λ values as shown in the graph



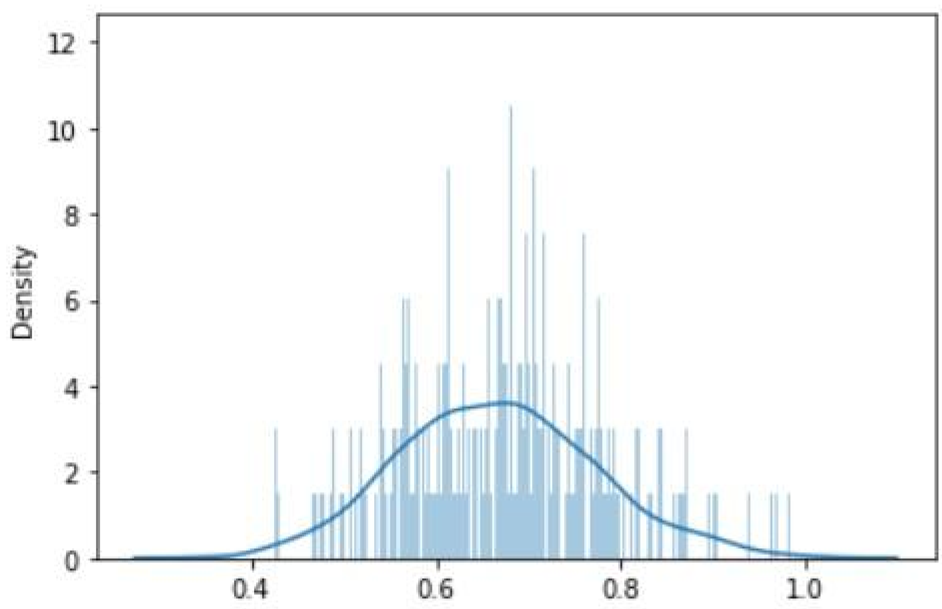
Mean and Variance for different values of parameters :

- Here we considered λ values ranging from 0.25 to 5.0 .
- The table below shows the theoretical and experimental mean and variance, values are calculated using the formulas

	Param1(λ)	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	0.25	4.046782	16.493264	4.000000	16.000000
1	0.50	2.018300	3.982406	2.000000	4.000000
2	0.75	1.323106	1.749813	1.333333	1.777778
3	1.00	0.991463	0.970321	1.000000	1.000000
4	1.25	0.792819	0.626030	0.800000	0.640000
5	1.50	0.664552	0.435423	0.666667	0.444444
6	2.00	0.502478	0.256474	0.500000	0.250000
7	2.50	0.397013	0.160400	0.400000	0.160000
8	5.00	0.200787	0.040972	0.200000	0.040000

Part 2:

- A Exponential distribution with λ =0.6666 .
- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean 0.66475 , population variance 0.44352
- For the sample mean distribution:
  - Mean : 0.66646
  - Variance : 0.070754



Beta Distribution:

The beta distribution is a family of continuous probability distributions defined on the interval [0, 1] parameterized by two positive shape parameters, denoted by α and β, that appear as exponents of the random variable and control the shape of the distribution

Pdf of distribution :

$$f_X(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & if\, x \in R_X \\ 0 & if\, x \notin R_X \end{cases}$$

Parameters :

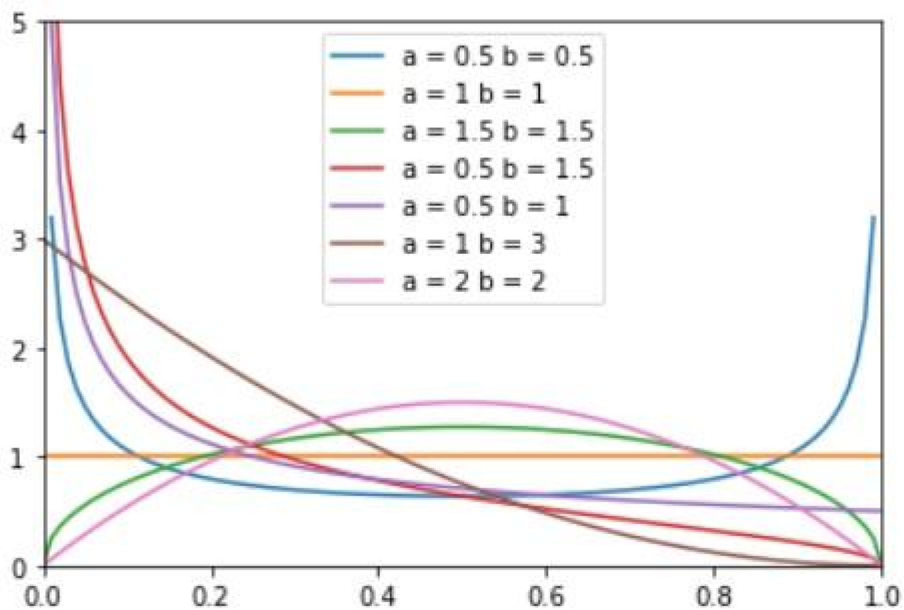
$\alpha > 0$ shape (real); $\beta > 0$ shape (real)

Mean : **E[x] =** $\frac{\alpha}{\alpha+\beta}$ **variance: Var[X] =** $\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$

Part 1:

Distribution for different values of parameters :

- We have used scipy.stats.beta.rvs(α, β,size) to generate the Beta distribution.
- We have plotted graphs for different αand β values as shown in the graph



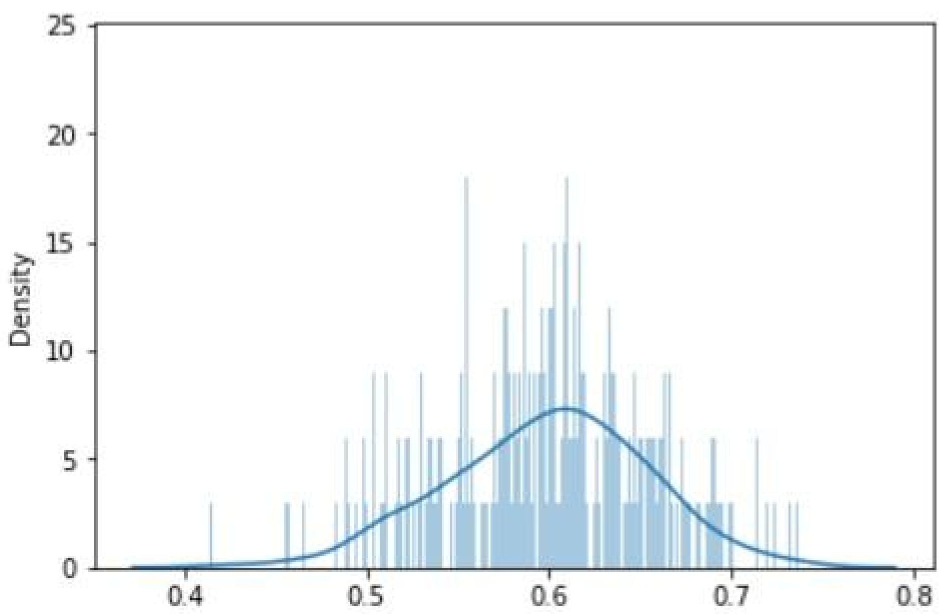
Mean and Variance for different values of parameters :

- Here we considered α values ranging from 0.5 to 2.0 and β from 0.5 to 2.0 .
- The table below shows the theoretical and experimental mean and variance, values are calculated using the formulas

	Param1(Alpha)	Param2(Beta)	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	0.5	0.5	0.505805	0.125520	0.500000	0.125000
1	1.0	1.0	0.497537	0.083193	0.500000	0.083333
2	1.5	1.5	0.498650	0.062357	0.500000	0.062500
3	0.5	1.5	0.251993	0.062940	0.250000	0.062500
4	0.5	1.0	0.336352	0.089617	0.333333	0.088889
5	1.0	3.0	0.250384	0.037210	0.250000	0.037500
6	2.0	2.0	0.499751	0.050057	0.500000	0.050000

Part 2:

- A Beta distribution with α =0.6 .
- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean 0.59907 , population variance 0.11976
- For the sample mean distribution:
  - Mean : 0.5989
  - Variance : 0.01935



Gamma Distribution:

Gamma distribution is a two-parameter family of continuous probability distributions. The exponential distribution, Erlang distribution, and chi-squared distribution are special cases of the gamma distribution.

Pdf of distribution :

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \quad \text{for} \quad x,\alpha,\beta > 0,$$

$$\Gamma(\alpha) = \int_0^{\infty} e^{-t} t^{\alpha-1} \mathrm{d}t.$$

Parameters :

shape parameter a(α)

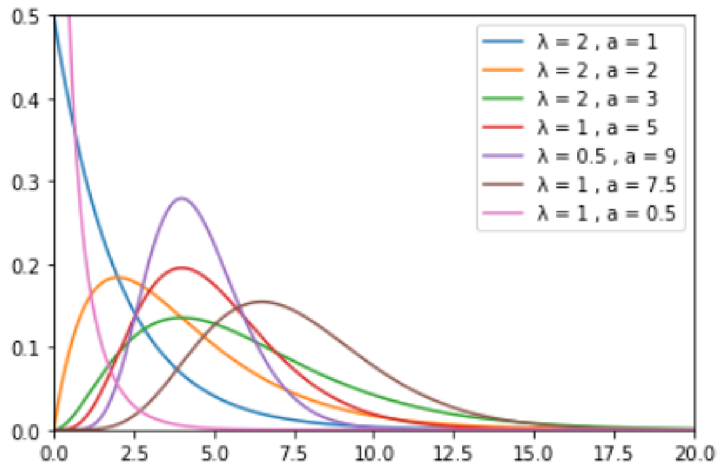
scale parameter λ.(β)

Mean : a/λ **variance:** a/λ^2

Part 1:

Distribution for different values of parameters :

- We have used scipy.stats.gamma.rvs(a, λ,size) to generate the Gamma distribution.
- We have plotted graphs for different $a$and λ values as shown in the graph



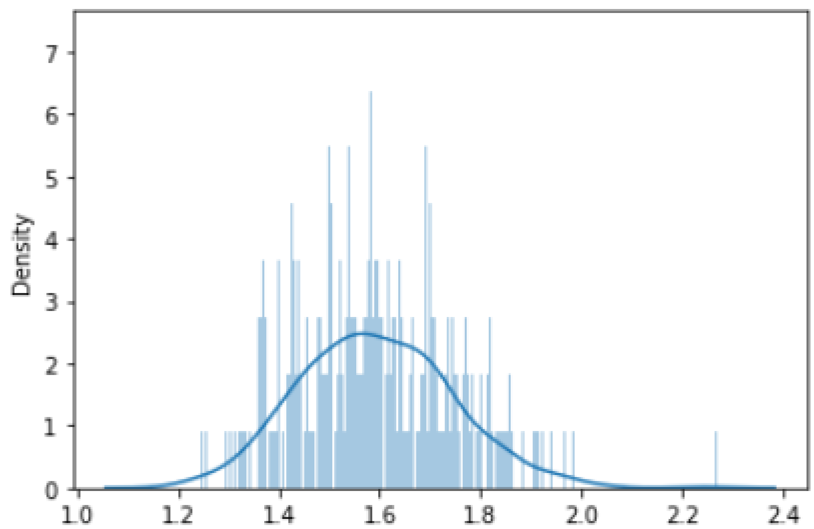
Mean and Variance for different values of parameters :

- Here we considered a values ranging from 1.0 to 0.5 and λ from 2.0 to 1.0 .
- The table below shows the theoretical and experimental mean and variance, values are calculated using the formulas

	Param1(a)	Param1(λ)	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	1.0	2.0	0.997270	1.013934	0.5	0.25
1	2.0	2.0	1.997531	1.963756	1.0	0.50
2	3.0	2.0	2.996328	2.999959	1.5	0.75
3	5.0	1.0	4.995472	4.980092	5.0	5.00
4	9.0	0.5	8.992353	8.969021	18.0	36.00
5	7.5	1.0	7.507723	7.423145	7.5	7.50
6	0.5	1.0	0.499208	0.485211	0.5	0.50

Part 2:

- A Gamma distribution with a =1 and λ= 0.6 .
- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean 1.59941 , population variance 0.96338
- For the sample mean distribution:
    - Mean : 1.59379
    - Variance : 0.15363



Lognormal Distribution:

log-normal 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.

Pdf of distribution :

$$P(x;\boldsymbol{\mu_N},\boldsymbol{\sigma_N}) = \frac{1}{x\sqrt{2\pi\log\left(1+\sigma_N^2/\mu_N^2\right)}}\exp\left(\frac{-\left[\log(x)-\log\left(\frac{\mu_N}{\sqrt{1+\sigma_N^2/\mu_N^2}}\right)\right]^2}{2\log\left(1+\sigma_N^2/\mu_N^2\right)}\right)$$

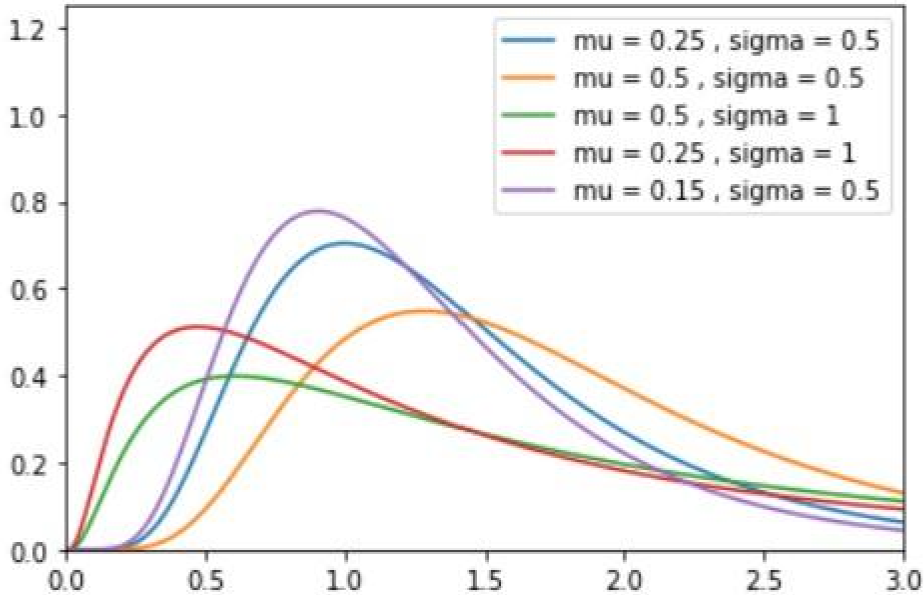
Parameters: $\mu \;\; and \; \sigma$

Mean : **E[x] =** $\exp(\mu + \frac{1}{2}\sigma^2)$ **variance: Var[X] =** $\exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$

Part 1:

Distribution for different values of parameters :

- We have used np.random.lognormal(μ, σ^2,size) to generate the Lognormal distribution.
- We have plotted graphs for different $\mu$and σ values as shown in the graph



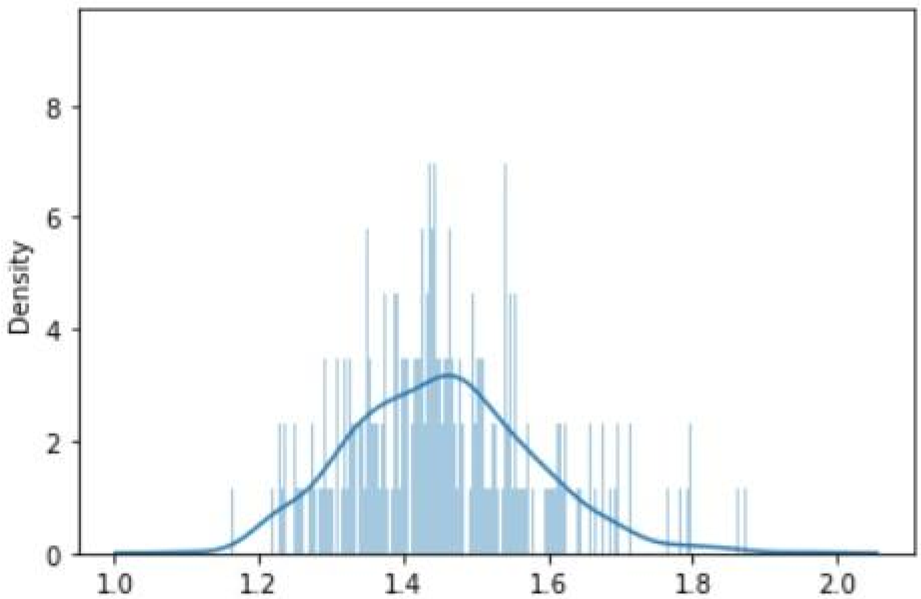
Mean and Variance for different values of parameters :

- Here we considered μ values ranging from 0.25 to 0.15 and σ^2 from 0.5 and 1.0.
- The table below shows the theoretical and experimental mean and variance, values are calculated using the formulas

	Param1(Mu)	Param2(Sigma)	Population Mean	Population Variance	Theoritical Mean	Theoritical Variance
0	0.25	0.5	1.445926	0.580367	1.454991	0.601282
1	0.50	0.5	1.875056	0.995771	1.868246	0.991346
2	0.50	1.0	2.762716	15.075362	2.718282	12.696481
3	0.25	1.0	2.109053	7.628773	2.117000	7.700805
4	0.15	0.5	1.316418	0.475213	1.316531	0.492288

Part 2:

- A Lognormal distribution with μ =0.25 and σ = 0.5 .
- Sample Mean distribution of 1000 number of samples with a sample size 40 is taken
- Population mean 1.45165 , population variance 0.60869
- For the sample mean distribution:
- Mean : 1.4529
- Variance : 0.99417


Density
0
2
4
6
8
1.0
1.2
1.4
1.6
1.8
2.0