```
import scipy
import numpy  as  np
from scipy.stats import binom
```

Binomial Distribution

What is the expected value (mean)for one biased coin flip with 30% probability of sucess?

```
binom.stats(n=1,p=0.3,moments='mvsk')
```

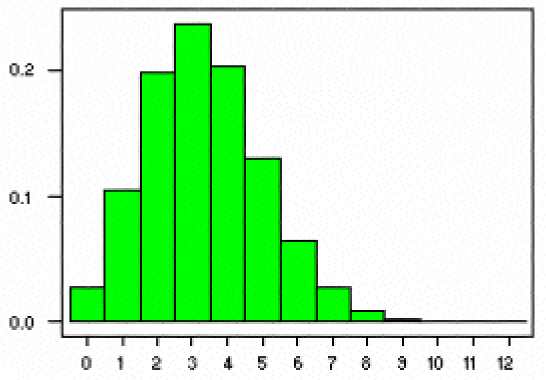
```
(array(0.3), array(0.21), array(0.87287156), array(-1.23809524))
```

```
binom.stats(n=1,p=0.3)
```

```
(array(0.3), array(0.21))
```

```
binom.std(n=1,p=0.3)
```

```
0.458257569495584
```



A survey found that 65% of all financial consumers were very satisfied with their primary financial institution. Suppose that 25 financial consumers are sampled and if the Gallup survey result still holds true today, what is the probability that exactly 19 are very satisfied with their primary financial institution?

The value of p is .65 (very satisfied), the value of $q = 1 - p = 1 - .65 = .35$ (not very satisfied), $n = 25$, and $x = 19$. The binomial formula yields the final answer.

$$_{25}C_{19}(.65)^{19}(.35)^{6} = (177{,}100)(.00027884)(.00183827) = .0908$$

```
print(binom.pmf(k=19,n=25,p=0.65))
```

```
0.09077799859322791
```

```
binom.cdf(k=19,n=25,p=0.65)
```

```
1-binom.cdf()
```

```
r=binom.rvs(n,p,size=10)
```

According to the U.S. Census Bureau, approximately 6% of all workers in Jackson, Mississippi, are unemployed. In conducting a random telephone survey in Jackson, what is the probability of getting two or fewer unemployed workers in a sample of 20?

$$\begin{array}{ccccccc} x=0 & & x=1 & & x=2 & & \\ {}_{20}C_0(.06)^0(.94)^{20} & + & {}_{20}C_1(.06)^1(.94)^{19} & + & {}_{20}C_2(.06)^2(.94)^{18} & = & \\ .2901 & + & .3703 & + & .2246 & = & .8850 \end{array}$$

```
binom.cdf(2,20,0.06)
```

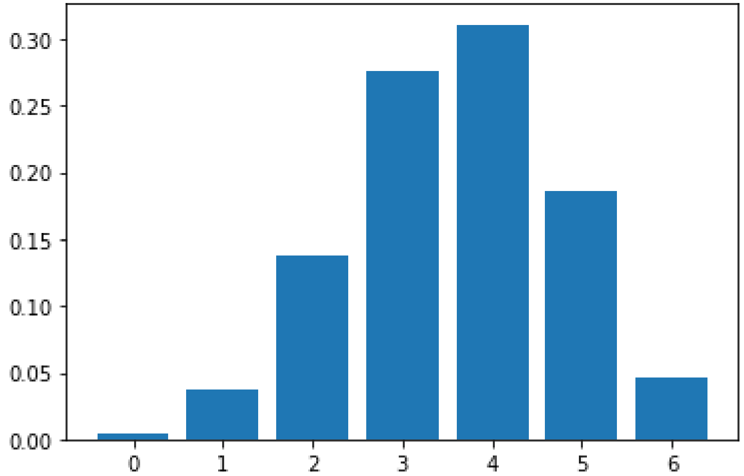
```
0.8850275957378545
```

```
from scipy.stats import binom
import matplotlib.pyplot as plt
# setting the values  of n and p
n = 6
p = 0.6
# defining list of r values
r_values = list(range(n + 1))
print(r_values)
# list of pmf values
dist = [binom.pmf(r, n, p) for r in r_values ]    # for every r.v r we will plot p(r)  (r w.r.t p(r))
# plotting the graph

plt.bar(r_values, dist)

plt.show()
```

```
[0, 1, 2, 3, 4, 5, 6]
```



Poisson Problems

```
from scipy.stats import poisson
```

Suppose bank customers arrive randomly on weekday afternoons at an average of 3.2 customers every 4 minutes. What is the probability of exactly 5 customers arriving in a 4-minute interval on a weekday afternoon?

```
poisson.pmf(5,3.2)
```

```
0.11397938346351824
```

Bank customers arrive randomly on weekday afternoons at an average of 3.2 customers every 4 minutes. What is the probability of having more than 7 customers in a 4-minute interval on a weekday afternoon?

```
p=poisson.cdf(7,3.2)

prob_more=1-p
prob_more
```

```
0.01682984174895752
```

A bank has an average random arrival rate of 3.2 customers every 4 minutes. What is the probability of getting exactly 10 customers during an 8-minute interval?

```
poisson.pmf(10,6.4)
```

```
0.052790043854115495
```

```
from scipy.stats import poisson
import matplotlib.pyplot as plt
```

```
mu = 0.6 #lambda
```

```
mean, var, skew, kurt = poisson.stats(mu, moments='mvsk')

print(mean, var, skew, kurt)
```

```
0.6 0.6 1.2909944487358056 1.6666666666666667
```

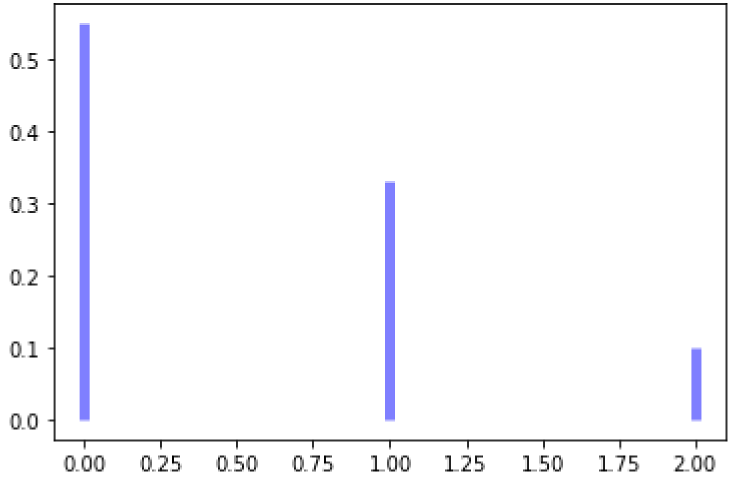
```
from scipy.stats import poisson
import matplotlib.pyplot as plt
fig, ax = plt.subplots(1, 1)

mu = 0.6
mean, var, skew, kurt = poisson.stats(mu, moments='mvsk')

x = np.arange(poisson.ppf(0.01, mu),
              poisson.ppf(0.99, mu))

ax.vlines(x, 0, poisson.pmf(x, mu), colors='b', lw=5, alpha=0.5)
```

```
<matplotlib.collections.LineCollection at 0x7fe4cea5d6d0>
```



Hypergeometric Distribution

Suppose 18 major computer companies operate in the United States and that 12 are located in California's Silicon Valley. If three computer companies are selected randomly from the entire list, what is the probability that one or more of the selected companies are located in the Silicon Valley?

$$N = 18,\ n = 3,\ A = 12,\text{ and } x \geq 1$$

This problem is actually three problems in one: $x = 1$, $x = 2$, and $x = 3$. Sampling is being done without replacement, and the sample size is 16.6% of the population. Hence this problem is a candidate for the hypergeometric distribution. The solution follows.

$$\begin{array}{ccccccc} x = 1 & & x = 2 & & x = 3 & & \\ \dfrac{{}_{12}C_1 \cdot {}_6C_2}{{}_{18}C_3} & + & \dfrac{{}_{12}C_2 \cdot {}_6C_1}{{}_{18}C_3} & + & \dfrac{{}_{12}C_3 \cdot {}_6C_0}{{}_{18}C_3} & = & \\ .2206 & + & .4853 & + & .2696 & = & .9755 \end{array}$$

```
from scipy.stats import hypergeom
p=hypergeom.sf(0,18,3,12)  #hypergeom.sf(x-1,N,n,A)     # sf=1-cdf
p
```

```
0.9754901960784306
```

Continuous Distribution

Uniform Distribution

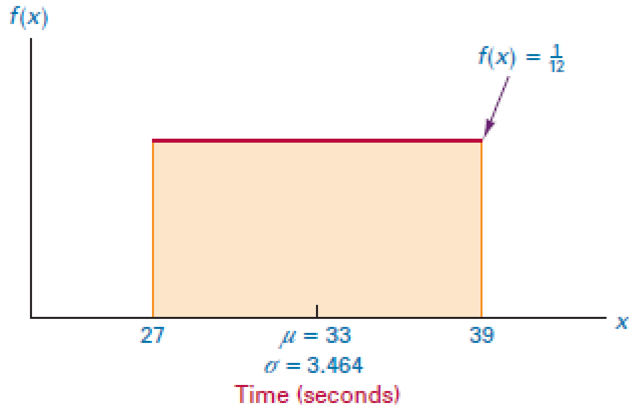
Suppose the amount of time it takes to assemble a plastic module ranges from 27 to 39 seconds and that assembly times are uniformly distributed. Describe the distribution. What is the probability that a given assembly will take between 30 and 35 seconds?

$$f(x) = \frac{1}{39 - 27} = \frac{1}{12}$$

$$\mu = \frac{a+b}{2} = \frac{39+27}{2} = 33$$

$$\sigma = \frac{b-a}{\sqrt{12}} = \frac{39-27}{\sqrt{12}} = \frac{12}{\sqrt{12}} = 3.464$$

The height of the distribution is 1/12. The mean time is 33 seconds with a standard deviation of 3.464 seconds.



$$P(30 \le x \le 35) = \frac{35-30}{39-27} = \frac{5}{12} = .4167$$

There is a .4167 probability that it will take between 30 and 35 seconds to assemble the module.

```
from scipy.stats import uniform
uniform.mean(loc=27,scale=12)
```

```
33.0
```

```
uniform.cdf(np.arange(30,36,1),loc=27,scale=12)
```

```
array([0.25      , 0.33333333, 0.41666667, 0.5       , 0.58333333,
       0.66666667])
```

```
prob=0.66666667-0.25
prob
```

```
0.41666667
```

According to the National Association of Insurance Commissioners, the average annual cost for automobile insurance in the United States in a recent year was $691. Suppose automobile insurance costs are uniformly distributed in the United States with a range of from$ 200 to \$1,182. What is the standard deviation of this uniform distribution?

```
uniform.mean(loc=200,scale=982)
```

```
691.0
```

```
uniform.std(loc=200,scale=982)
```

```
283.4789821721062
```

Normal Distribution

```
from scipy.stats import norm
```

cdf(x<val)

```
val,m,s=68,65.5,2.5
print(norm.cdf(val,m,s))
```

```
0.8413447460685429
```

cdf(x>val)

```
print(1-norm.cdf(val,m,s))
```

```
0.15865525393145707
```

cdf(val1<x<val2)

```
print(norm.cdf(val,m,s)-(norm.cdf(63,m,s)))
```

```
0.6826894921370859
```

What is the probability of obtaining a score greater than 700 on a GMAT test that has a mean of 494 and a standard deviation of 100? Assume GMAT scores are normally distributed.

P(x>700|m=494,100)

```
print(1-norm.cdf(700,494,100))
```

```
0.019699270409376912
```

For the same GMAT examination, what is the probability of randomly drawing a score that is 550 or less?

```
print(norm.cdf(550,494,100))
```

```
0.712260281150973
```

What is the probability of randomly obtaining a score between 300 and 600 on the GMAT exam?

```
print(norm.cdf(600,494,100)-(norm.cdf(300,494,100)))
```

    0.2550348541262666

w.r.t area

```
norm.ppf(0.95)
```

    1.6448536269514722

```
norm.ppf(1-.6772)
```

    -0.45988328292440145

Exponential distribution

A manufacturing firm has been involved in statistical quality control for several years. As part of the production process, parts are randomly selected and tested. From the records of these tests, it has been established that a defective part occurs in a pattern that is Poisson distributed on the average of 1.38 defects every 20 minutes during production runs. Use this information to determine the probability that less than 15 minutes will elapse between any two defects.

The value of λ is 1.38 defects per 20-minute interval. The value of μ can be determined by

$$\mu = \frac{1}{\lambda} = \frac{1}{1.38} = .7246$$

On the average, it is .7246 of the interval, or (.7246)(20 minutes) = 14.49 minutes,

```
mu1=1/1.38
mu1
```

```
0.7246376811594204
```

In this problem, the probability question involves 15 minutes and the interval is 20 minutes. Thus x_0 is 15/20, or .75 of an interval. The question here is to determine the probability of there being less than 15 minutes between defects. The probability formula always yields the right tail of the distribution—in this case, the probability of there being 15 minutes or more between arrivals. By using the value of x_0 and the value of λ, the probability of there being 15 minutes or more between defects can be determined.

$$P(x \geq x_0) = P(x \geq .75) = e^{-\lambda x_0} = e^{(-1.38)(.75)} = e^{-1.035} = .3552$$

The probability of .3552 is the probability that at least 15 minutes will elapse between defects. To determine the probability of there being less than 15 minutes between defects, compute $1 - P(x)$. In this case, $1 - .3552 = .6448$. There is a probability of .6448 that less than 15 minutes will elapse between two defects when there is an average of 1.38 defects per 20-minute interval or an average of 14.49 minutes between defects.

```
from scipy.stats import expon
expon.cdf(0.75,0,(1/1.38))
```

```
0.6447736190750485
```