In [1]:
```
from  scipy.stats import binom,poisson,expon
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import warnings
import math
warnings.filterwarnings("ignore")
```

Binomial Distribution in python

A company manufactures LED bulbs with a faulty rate of 30%. If I randomly select 6 chosen LEDs, what is the probability of having 2 faulty LEDs in my sample? Calculate the average value of this process. Also evaluate the standard deviation associated with it ?

P=0.3 ; Q=1-P=0.7

n=total number of trials=6

k=number of trail that will be successed=2

z=Total number of random samples =500

why we take random samples(z) ?

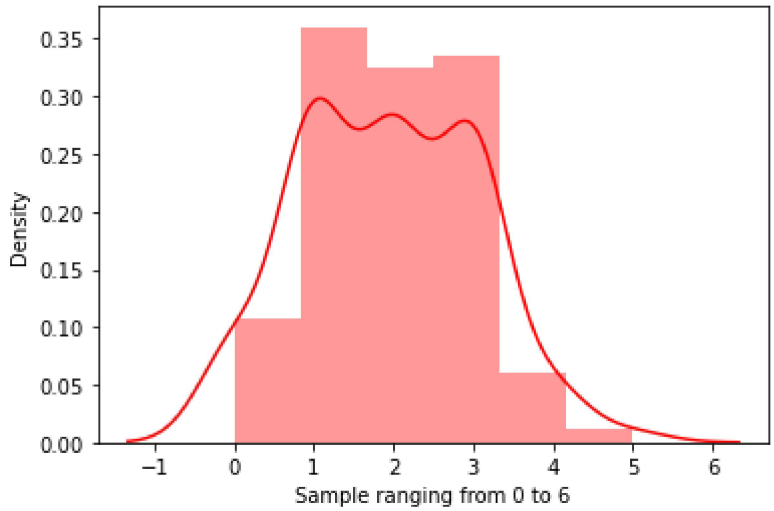
When we analyse data as a M.L engineer we must ensure how the uncertainty introduced by random samples affects our datasets .We also try to evaluate how data would be affected by random error.

In [2]:
```
### This function will generate 100 random sample ranging from 0 to 6 distri
binomial_data=binom.rvs(n=6,p=0.3,size=100)
binomial_data
```

Out[2]:
```
array([0, 2, 2, 2, 3, 1, 3, 1, 2, 3, 3, 2, 4, 1, 1, 2, 2, 0, 0, 2, 1, 2,
       1, 1, 1, 2, 3, 0, 1, 2, 0, 1, 1, 3, 0, 4, 4, 1, 1, 1, 3, 3, 3, 3,
       1, 2, 3, 2, 1, 3, 3, 2, 2, 3, 1, 2, 3, 3, 1, 2, 3, 0, 3, 2, 2, 2,
       1, 3, 5, 1, 2, 1, 0, 1, 3, 2, 0, 1, 1, 2, 2, 1, 1, 2, 1, 3, 1, 2,
       3, 3, 4, 3, 3, 2, 3, 1, 3, 4, 1, 3])
```

In [3]:
```
### Lets plot the histogram for the same
sns.distplot(binomial_data,hist=True,kde=1,color="red")
plt.xlabel("Sample ranging from 0 to 6")
```

Out[3]: Text(0.5, 0, 'Sample ranging from 0 to 6')



In [4]:
```
#Probability of getting faulty out of 6 trials
probab=binom.pmf(k=2,n=6,p=0.3)
print("Probability will be :",probab)
cdf=binom.cdf(k=2,n=6,p=0.3)
print("CDF will be :",cdf)
mean,var=binom.stats(n=6,p=0.3)
print("mean := ",mean)
print("standard deviation :=",math.sqrt(var))
```

```
Probability will be : 0.3241349999999999
CDF will be : 0.74431
mean :=  1.7999999999999998
standard deviation := 1.1224972160321822
```

Poission Distribution in python

Its Probability Mass Function is given by this formula:

$$f(k;\lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where

- k is the number of successes (the number of times a desired even happening)
- λ is the given rate
- e is Euler's number: $e = 2.71828\ldots$
- $k!$ is the factorial of k

Customers arrive at a rate of 72 per hour to my shop. What is the probability of k customers arriving in 4 minutes? a) 5 customers, b) not more than 3 customers.

a) 5 Customer ---> lambda = (72/60)*4 = 4.8

In [5]:

```
print(poisson.pmf(k=5,mu=4.8))
```

```
0.17474768364388296
```

b) not more then 3 customer

In [6]:

```
## takig cumulative probability distribution as we want to sun till 3
print(poisson.cdf(k=3,mu=4.8))
```

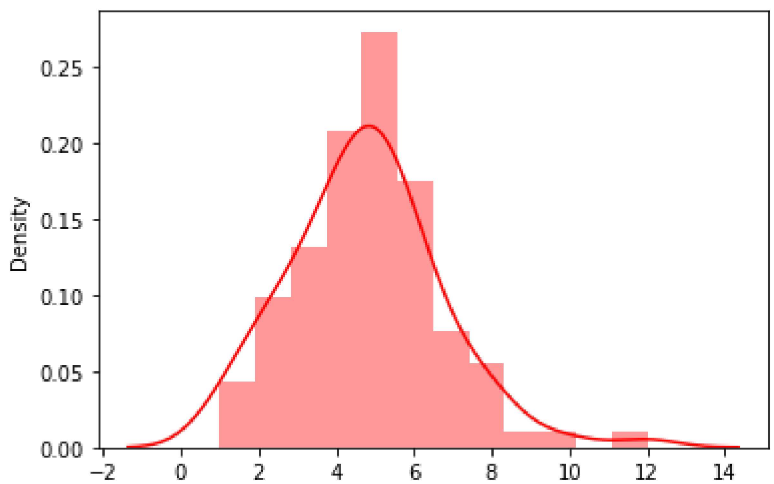
```
0.2942299164965642
```

In [7]:
```
### Sample distribution
poisson_data=poisson.rvs(mu=4.8,size=100)
poisson_data
```

Out[7]:
```
array([ 4,  4,  3,  7,  6,  4,  5,  5,  5,  4,  3,  2,  1,  7,  4,  6,  4,
        5,  4,  3,  1,  3,  8,  8,  4,  4,  4,  5,  5,  5,  1,  6,  3,  6,
        5,  6,  3,  7,  5,  6,  5, 12,  5,  2,  2,  4,  6,  3,  5,  7,  6,
        7,  3,  6,  6,  2,  6,  2,  3,  4,  4,  6,  5,  2,  5,  5,  6,  8,
        3,  4,  1,  5,  6,  5,  6,  2,  9,  3,  8,  5,  4,  7,  4,  4,  5,
        5,  2,  3,  5, 10,  4,  6,  5,  5,  5,  2,  8,  4,  7,  5])
```

In [8]:
```
### Distribution plot
sns.distplot(poisson_data,hist=True,kde=True,color="red")
```

Out[8]: <AxesSubplot:ylabel='Density'>



Exponential Distribution

How do you know when to use exponential distribution? The exponential distribution concerns the amount of time until a particular event occurs.

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

Formula 1 — Probability density function of exponential distribution (Image by author).

The formula takes two arguments as <Lambda> and <x>. The value lambda represents the mean number of events that occur in an interval. The x represents the moment that the event will occur. Thereby, when the average occurrence of the events is lambda, f(x, lambda) gives the probability of occurrence of the event at the moment x. Because moments in time can't be negative the function returns to zero if x is less than zero. The probability changes based on the value lambda and the x value

Suppose there is a coffee shop where customers order coffee on an average of 15 times per hour. The question would be: "What is the probability that the next coffee order will arrive after 5 minutes?"

lambda = 15/60 = 1/4

x = 5

In [9]:
```
# Order exactly at 5. minute
expon.pdf(5 ,scale=4)
```

Out[9]: 0.07162619921504752

In [10]:
```
# Order after 5 minutes
1 - expon.cdf(5 ,scale=4)
```

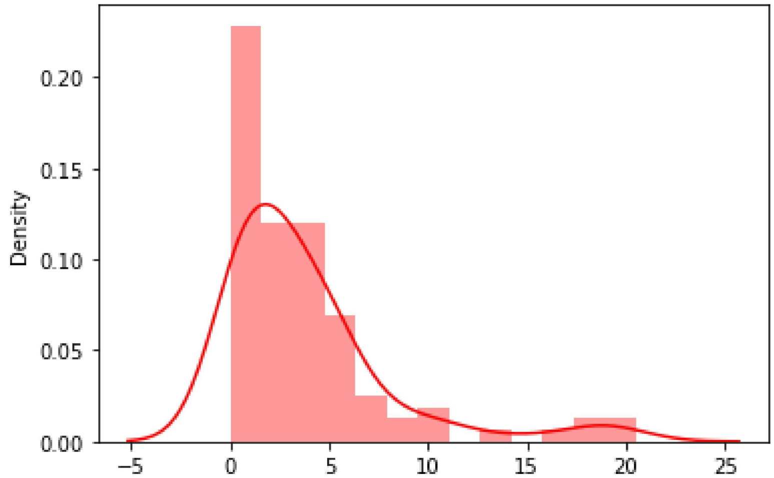
Out[10]: 0.28650479686019015

In [14]:
```
### lets look how the distribution looks like
expon_data = expon.rvs(scale=4,size= 100)

```

In [11]:
```
sns.distplot(expon_data ,hist=True,kde=True,color="red")
```

Out[11]: <AxesSubplot:ylabel='Density'>



Connection between Exponential and Poisson distribution

If the Poisson distribution deals with the number of occurrences in a fixed period of time, the exponential distribution deals with the time between occurrences of successive events as time flows by continuously

Connection between Bernoulli and Binomial distribution

Bernoulli deals with the outcome of the single trial of the event, whereas Binomial deals with the outcome of the multiple trials of the single event. Bernoulli is used when the outcome of an event is required for only one time, whereas the Binomial is used when the outcome of an event is required multiple times.