

Distributions

Distribution

- A mathematical function that provides the **probabilities of occurrence** of different **possible outcomes** in an experiment.
- The probability distribution is a description of a random phenomenon in terms of the probabilities of events

Distribution Parameters

- Probability Mass Function (PMF) - Discrete /
Probability Density Function (PDF) - Continuous
 - Function that gives a probability that a random variable is equal to some value
- Cumulative Distribution Function (CDF)
 - Function evaluating the probability that X will take a value less than or equal to x

Simulation of a Tossing a coin

- Consider that a fair coin is tossed 6 times.
- The goal is to count the number of HEADS

```
1  # Importing basic libraries
2  import numpy as np
3  import pandas as pd
4  import matplotlib.pyplot as plt
5  import seaborn as sns
6  from tqdm import tqdm
7  sns.set()
```

Simualting the coin toss for 10 Lakhs time

In [30]:

```
1 no_trials=10_00_000
2 data=[]
3 for _ in tqdm(range(no_trials)):
4     data.append(np.sum(np.random.randint(2,size=6)))
5 print(len(data))
```

100%|██| 10000000/10000000 [00:26.0<]

10000000

In [31]:

```
1 data[:10]
```

Out[31]: [4, 5, 2, 4, 2, 2, 4, 3, 2, 5]

Finding the frequency of each HEAD's possibility

```
In [32]: 1 np.unique(data,return_counts=True)
```

```
out[32]: (array([0, 1, 2, 3, 4, 5, 6]),  
          array([ 15516,  94613, 233949, 311535, 234917,  93874,  15596],  
            dtype=int64))
```

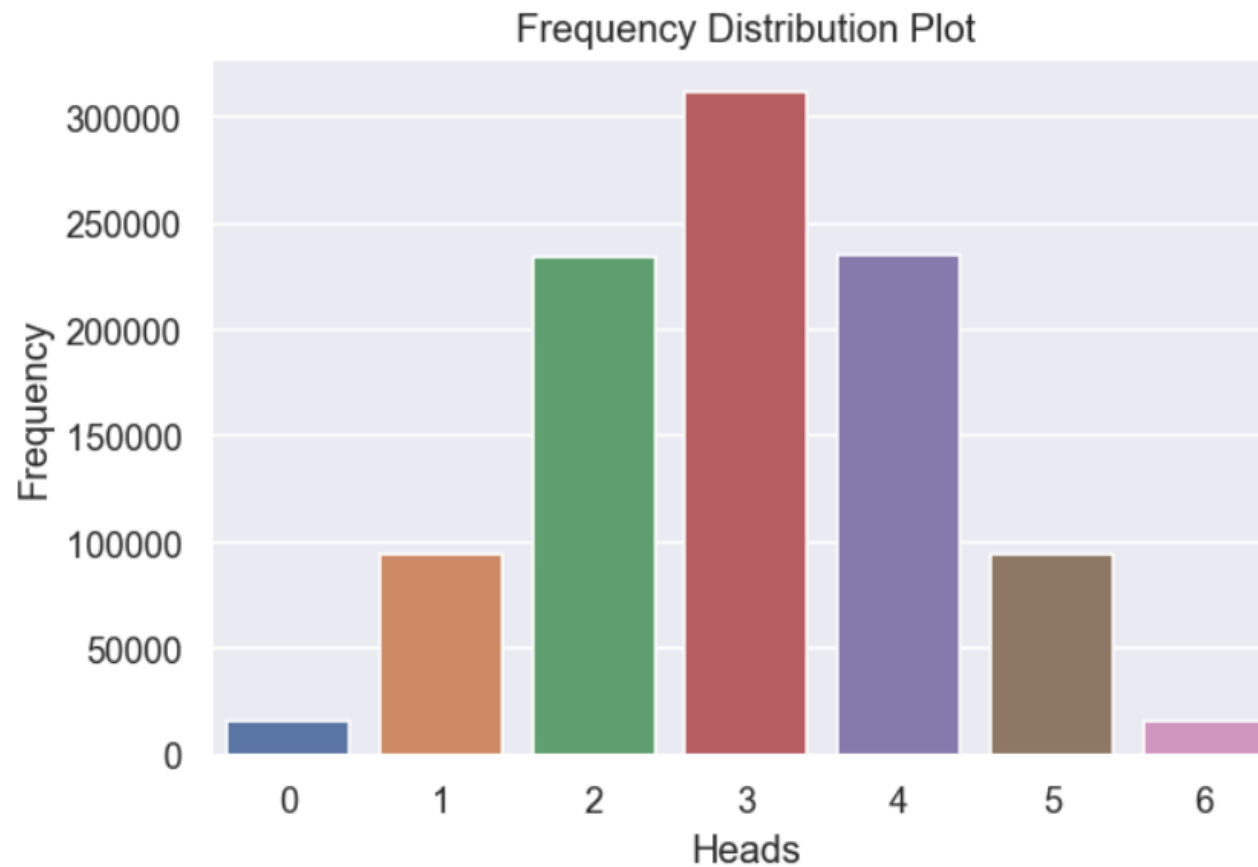
```
In [33]: 1 Heads,counts=np.unique(data,return_counts=True)  
        2 df=pd.DataFrame({"Heads":Heads,"Frequency":counts})  
        3 df
```

```
out[33]:
```

	Heads	Frequency
0	0	15516
1	1	94613
2	2	233949
3	3	311535
4	4	234917
5	5	93874
6	6	15596

Plot - Frequency Distribution plot

```
1 plt.figure(dpi=120)
2 sns.barplot(x="Heads",y="Frequency",data=df)
3 plt.title("Frequency Distribution Plot")
4 plt.show()
```



Calculating the probabilities for each outcomes

In [42]:

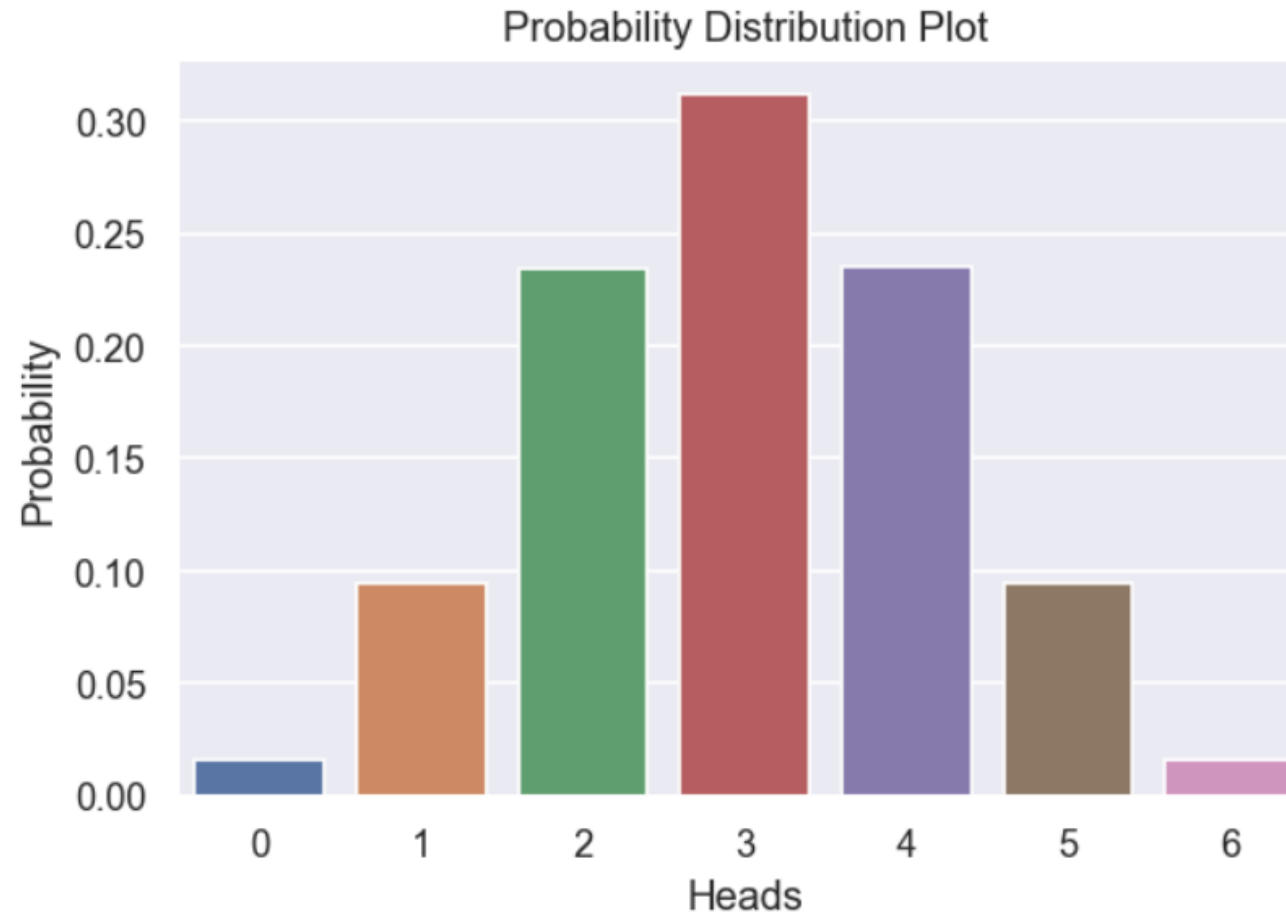
```
1 #Calculating the propbabilities
2 df["Probability"]=df["Frequency"]/no_trials
3 df
```

Out[42]:

	Heads	Frequency	Probability
0	0	15516	0.015516
1	1	94613	0.094613
2	2	233949	0.233949
3	3	311535	0.311535
4	4	234917	0.234917
5	5	93874	0.093874
6	6	15596	0.015596

Probability Distribution Plot

```
1 plt.figure(dpi=120)
2 sns.barplot(x="Heads",y="Probability",data=df)
3 plt.title("Probability Distribution Plot")
4 plt.show()
```



Types of Distribution

- **Discrete Distribution**

- Binomial Distribution
- Poisson Distribution

- **Continuous Distribution**

- Normal Distribution