Probability Distribution

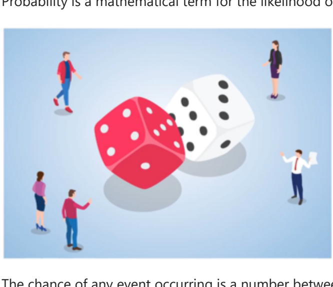
Agenda

In this session, we will cover the following concepts with the help of a business use case:

- Probability
- Different types of Probability Distribution:
  - Binomial Distribution
  - Poisson Distribution
  - Normal Distribution
  - Uniform Distribution
- Probability Density Function and Mass Function
- Cumulative Distribution Function
- Central Limit Theoram
- Estimation Theory

Probability and Its Importance

Probability is a mathematical term for the likelihood of any event happening.



The chance of any event occurring is a number between 0 and 1.

Uses of Probability

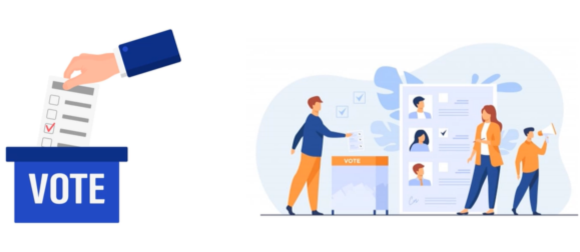
1. It is used to anticipate risks and identify ways to manage such risks.
2. It is used to make predictions about future events based on their likelihood.
3. It is a powerful tool used to incorporate uncertainty in planning and decision making.
4. Probability is used in a number of areas such as weather forecasting, scientific research, and healthcare.

Real-Life Examples:

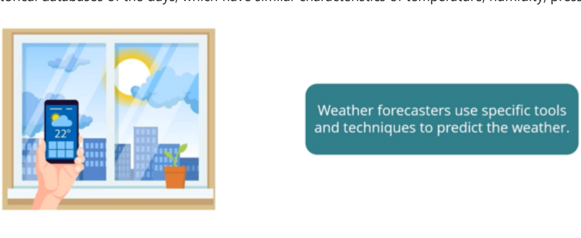
1. Flipping of a coin: It is essential to decide who will bat or bowl first before the beginning of any cricket match.This is determined by flipping a coin. Both head and tail of the coin have a 1 out of 2, i.e. a 50% chance of occurring. Hence, the probability of getting the desired outcome is 0.5.

2. Election forecasts: Many political analysts use the results of exit polls to predict whether a certain political party will come into power.

Probability plays a key role in the prediction of election results.



3. Weather forecasting: Before planning for an outing or a picnic, we always check the weather forecast to find out the probability whether rain may occur. Weather forecasters use specific tools and techniques to predict the weather. They do this by looking at all the other historical databases of the days, which have similar characteristics of temperature, humidity, pressure, etc.



Probability Distribution

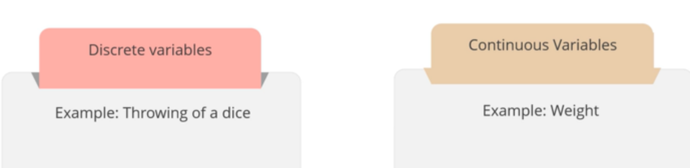
Probability distributions are statistical functions that describe the likelihood of obtaining possible values that a random variable can take.

The notation used by statisticians to describe probabilities is as given below:

p(x) = the likelihood that a random variable takes a specific value of x.

Categories of Variables

There are two types of variables for which probability distributions are defined:



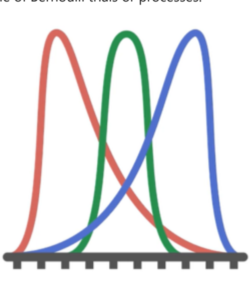
Similarly, for a single random variable, statisticians divide distributions into the following two types:

The probability distribution indicates how the total probability 1 or 100% is distributed across all values. Each value taken by a random variable clearly constitutes an event.

Consider the outcome of throwing a fair dice. Its probability distribution is:

Value	Probability
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
Total	1

Binomial Distribution and the Poisson Distribution are the two commonly used discrete probability distributions. These are based on the principle of Bernoulli trials or processes.

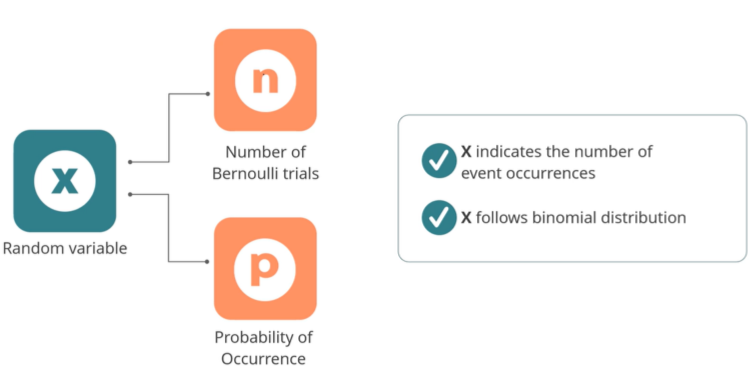


Probability Distribution: Binomial Distribution

A binomial distribution indicates the probability of a SUCCESS or FAILURE of a survey that is repeated multiple times. The binomial is a type of distribution that has two possible outcomes.

Binomial Distribution Example:

Let X denote a random variable (r.v.) that indicates the number of times an event occurs in 'n' Bernoulli trials with probability p of occurrence in trial. Then, X is said to follow a binomial distribution with parameters n and p where p is referred to as probability parameter.



If p is the probability of occurrence in a trial, the probability of non-occurrence is q= 1-p which is also a constant. So, when X follows the binomial with parameters n and p, X can take values 0, 1....r...n.

Then,

P{ X = 0}= q*q....q..q $=q^n$ as outcomes are independent.

This follows multiplication theorem of probability.

Similarly,

Probability of Pr

Pr {X =n} = p^n

Probability of Pr

Pr {X= r} = ^nC_r * p^r* q^{n-r}

where

r can vary from **0** to **n**

Note: The probability values can be obtained from statistical tables or using software.

Example to illustrate a real-world problem of a spare parts manufacturer:

If the probability of a machine producing a defective piece is 0.05, What is the probability of finding an inspection sample of size n has four or more defectives?

Solution:Let X denote the random variable indicating the number of defectives in the inspection sample of size 10.

Assuming that the process is Bernoulli, X follows a Binomial with parameters n= 10 and p = 0.1

The probabilities of X taking different values are as follows:

Probabilities of X taking different values:

R	0	1	2	3	4	5	6	7	8	9	10	Total
Pr{X=r}	0.0282	0.1211	0.2335	0.2668	0.2001	0.1029	0.0368	0.0090	0.0014	0.0001	0.0000	1

Pr {X ≥ 4} = 1 - Pr{ X≤ 3}
= 1 - 0.6496
= 0.3504

Probability that the inspection sample has four or more defectives is equal to Probability of X is greater than or equal to 4

Pr {X ≥ 4} = 1 Probability that the inspection sample has three or less defectives which is equal to Probability of X is less than or equal 3

Pr{ X≤ 3} = 1- 0.6496 = 0.3504

Note: The result justifies the assumption that the process is Bernoulli.

Binomial Distribution Using Python

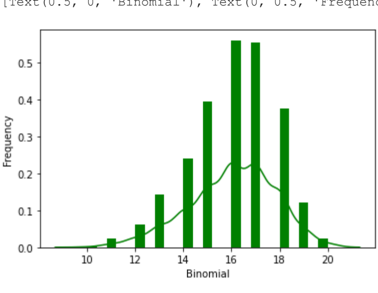
In [1]:

```
from scipy.stats import binom
import seaborn as sb

binom.rvs(size=10,n=20,p=0.8)

data_binom = binom.rvs(n=20,p=0.8,loc=0,size=1000)
ax = sb.distplot(data_binom,
                 kde=True,
                 color='green',
                 hist_kws={"linewidth": 22,'alpha':1})
ax.set(xlabel='Binomial', ylabel='Frequency')
```

```
/usr/local/lib/python3.7/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[1]:

```
[Text(0.5, 0, 'Binomial'), Text(0, 0.5, 'Frequency')]
```



Use Case: Determine the Probability of r Successes

Problem Statement:

Consider a random experiment of tossing a biased coin 6 times where the probability of getting a head is 0.6. If 'getting a head' is considered as 'success', then the binomial distribution table will contain the probability of r successes for each possible value of r.

r	0	1	2	3	4	5	6
P (r)	0.004096	0.036864	0.138240	0.276480	0.311040	0.186624	0.046656

Calculate a binomial distribution using the values of n and p to check if the distribution is normal.

Solution:

This distribution has a mean equal to np and variance of np(1-p).

Here, scipy.stats module contains various functions for statistical calculations and tests. The stats() function of the scipy.stats.binom module can be used to calculate a binomial distribution using the values of n and p.

In [2]:

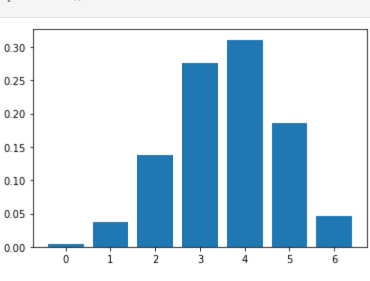
```
from scipy.stats import binom
# Setting the values of n and p
n = 6
p = 0.6
# Defining the list of r values
r_values = list(range(n + 1))
# Obtaining the mean and variance
mean, var = binom.stats(n, p)
# List of pmf values
dist = [binom.pmf(r, n, p) for r in r_values ]
# Printing the table
print("r\tp(r)")
for i in range(n + 1):
    print(str(r_values[i]) + "\t" + str(dist[i]))
# Printing mean and variance
print("mean = "+str(mean))
print("variance = "+str(var))
```

```
r       p(r)
0       0.0040960000000000015
1       0.03686400000000002
2       0.1382400000000001
3       0.2764800000000001
4       0.3110400000000001
5       0.1866240000000001
6       0.04665599999999999
mean = 3.5999999999999996
variance = 1.44
```

scipy.stats.binom.pmf function is used to obtain the probability mass function for a certain value of r, n, and p. We can obtain the distribution by passing all possible values of r(0 to n).

In [3]:

```
from scipy.stats import binom
import matplotlib.pyplot as plt
# Setting the values of n and p
n = 6
p = 0.6
# Defining the list of r values
r_values = list(range(n + 1))
# List of pmf values
dist = [binom.pmf(r, n, p) for r in r_values ]
# Plotting the graph
plt.bar(r_values, dist)
plt.show()
```

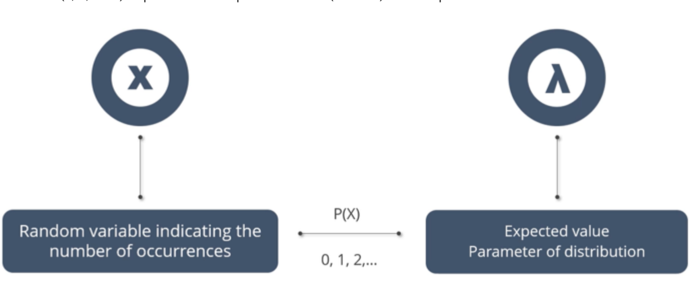


As we can see that the distribution is normal.

Probability Distribution: Poisson Distribution

Poisson distribution measures the probability of an event(s) occurring over a specified period.It can be used only when the opportunities for the occurrence of the outcome is a very large value and the probability of occurrence of the outcome during an opportunity is very low.

If X denotes the random variable indicating the number of occurrences which follows a Poisson distribution, the probabilities of X taking a given value (0, 1, 2 .....) depends on its expected value λ (lambda). λ is the parameter of the distribution.



The probabilities can be obtained from tables or by using a software.

For a given value of λ, the probabilities can be obtained from tables or software.

When X follows a Poisson distribution with parameter λ, the following scenarios are true:

E(X) = λ

Var(X) = λ

S.D(X) = √λ

Consider a real-world problem:

The number of spares required for any component in a machine follows a Poisson distribution with λ = 2. An enterprise using this machine must arrive at the number of spares of this component needed so that the probability of a stock out is at most 0.06. How can we determine this number?

Solution: Let the random variable X denote the number of spares required. X follows a Poisson distribution with parameter λ = 2.

K	0	1	2	3	4	5	6
P(X=k)	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.012
Cumulative	0.1353	0.406	0.6767	0.8571	0.9473	0.9834	0.9954
Stock out probability	0.8647	0.594	0.3233	0.1429	0.0527	0.0166	0.0046

K	0	1	2	3	4	5	6
P(X=k)	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.012
Cumulative	0.1353	0.406	0.6767	0.8571	0.9473	0.9834	0.9954
Stock out probability	0.8647	0.594	0.3233	0.1429	0.0527	0.0166	0.0046

K	0	1	2	3	4	5	6
P(X=k)	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.012
Cumulative	0.1353	0.406	0.6767	0.8571	0.9473	0.9834	0.9954
Stock out probability	0.8647	0.594	0.3233	0.1429	0.0527	0.0166	0.0046

Stockout probability for a given value = **1 – Cumulative value**

K	0	1	2	3	4	5	6
P(X=k)	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.012
Cumulative	0.1353	0.406	0.6767	0.8571	0.9473	0.9834	0.9954
Stock out probability	0.8647	0.594	0.3233	0.1429	0.0527	0.0166	0.0046

Stockout probability for a given value = **1 – Cumulative value**

Note:The stock out probability for a given value = 1 – Cumulative value

As we can see that when fewer spare units are stocked, the stock out probabilities are high.

when 3 spares are stocked, the stock out probability is 0.1429. The stock out probability when 4 spare parts are stocked is 0.0527. Since the stock out probability should not exceed 0.06, the firm should stock 4 units.

The binomial can sometimes be approximated by a Poisson theorem.

Binomial as a Poisson

If
- n is very large
- p is very small

Can be approximated as a Poisson with λ = n*p

This is called Poisson distribution.

Poisson Distribution Using Python

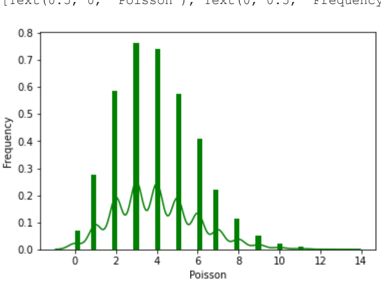
In [3]:

```
from scipy.stats import poisson
import seaborn as sb

data_binom = poisson.rvs(mu=4, size=10000)
ax = sb.distplot(data_binom,
                 kde=True,
                 color='green',
                 hist_kws={"linewidth": 25,'alpha':1})
ax.set(xlabel='Poisson', ylabel='Frequency')
```

```
C:\Users\alpika.gupta\Anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[3]:

```
[Text(0.5, 0, 'Poisson'), Text(0, 0.5, 'Frequency')]
```



Use Case: To Determine the Probability of Restaurants

Problem Statement:

Suppose you are going for a long drive. The rate of occurrences of good restaurants in a range of 10 miles is 2. In other words, the mean number of occurrences of restaurants in a range of 10 miles is 2. What is the probability that 0, 1, 2, 3, 4, or 5 restaurants will show up in the next 10 miles.

Solution:

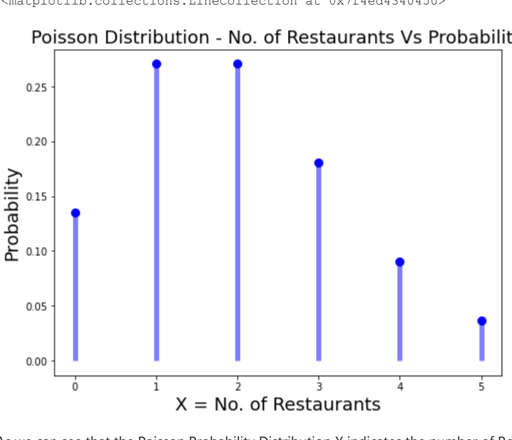
The probability of the different number of restaurants ranging from 0 to 5 that one could find within 10 miles given the mean number of occurrences of the restaurant in 10 miles is 2.

In [2]:

```
# Scipy.stats poisson class is used along with pmf method to calculate the value of probabilities
from scipy.stats import poisson
import matplotlib.pyplot as plt
#
# Random variable representing number of restaurants
# Mean number of occurrences of restaurants in 10 miles is 2
#
X = [0, 1, 2, 3, 4, 5]
lmbda = 2

# Probability values
#
poisson_pd = poisson.pmf(X, lmbda)
#
# Plot the probability distribution
#
fig, ax = plt.subplots(1, 1, figsize=(8, 6))
ax.plot(X, poisson_pd, 'bo', ms=8, label='poisson pmf')
plt.ylabel("Probability", fontsize="18")
plt.xlabel("X = No. of Restaurants", fontsize="18")
plt.title("Poisson Distribution - No. of Restaurants Vs Probability", fontsize="18")
ax.vlines(X, 0, poisson_pd, colors='b', lw=5, alpha=0.5)
```

Out[2]:

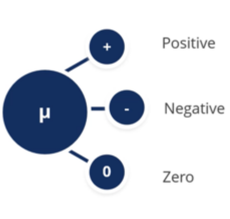
```
<matplotlib.collections.LineCollection at 0x7f4ed4340450>
```



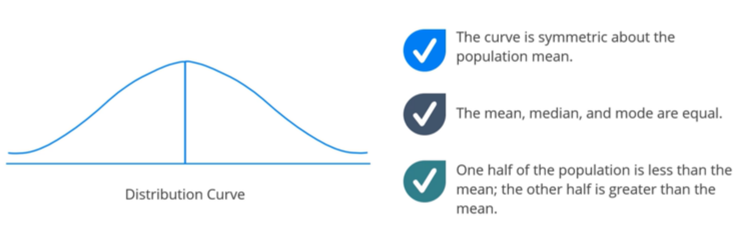
As we can see that the Poisson Probability Distribution X indicates the number of Restaurants in 10 miles.

Probability Distribution: Normal Distribution

Normal distribution extends between (minus infinity to plus infinity) -∞ to +∞. The distribution is completely specified by its mean μ and standard deviation σ. The value of population mean μ can be positive, negative, or zero.

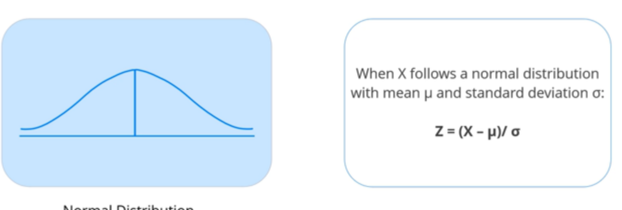


The properties of the Normal Distribution are:



Standard Normal Distribution

The normal distribution with mean 0 and standard deviation 1 is called the standard normal distribution.



Normal Distribution

Statistical tables are used to get probabilities for Z i.e., the normal distribution with zero mean and standard deviation unity. For other values of mean and standard deviation, probabilities are obtained using the transformation.

$$Z= (X - \mu)/ \sigma$$

Example:

A Factory Outlet uses a filling machine for low-pressure oxygen shells. It has been established that the weights of filled cylinders in grams follow a normal distribution with Population Mean μ = 1.433 and Standard Deviation σ = 0.033. The specification limits are 1.460+- 0.085.

How can we determine the probability that a cylinder will meet specifications?

Solution: Let the random variable X denote the weight of a filled in cylinder (grams).

Specification limits = (1.460 - 0.085, 1.460 + 0.085) or (1.375, 1.545)

X follows a normal distribution

$\mu = 1.433$ | $\sigma = 0.033$

Probability that a cylinder meets specification limits is shown below:

$P\{1.375 < X < 1.545\}$

$P\{(1.375 - 1.433)/0.033 < Z = (X - \mu)/\sigma < (1.545 - 1.433)/0.033\}$

$P\{-1.7575 < Z < 3.3939\}$

$= P\{-1.7575 < Z < 0\} + P\{0 < Z < 3.3939\}$

$= 0.4608 + 0.5 = 0.9608$

Pr {a cylinder meets specification limits}
= 0.9608

Pr {a cylinder does not meet specification limits}
= 1- 0.9608 = 0.0392

Percentage of cylinders that fall outside the specification limits:

3.92%

Probability Distribution: Uniform Distribution

A random variable X is said to follow a Uniform Distribution in the range [a, b] if the probability of lying in the range [c, d] is directly proportional to the length of the interval.

Hence, Pr {c < X < d} = (d-c)/(b-a)

Probability of c less than X is less than d is equal to d minus c divided by b minus a

The values a, b, c and d are all finite. But, they can be positive, negative, or zero.

Consider a large set of random numbers drawn from a random number table or from a computer (for instance, EXCEL). If a decimal is positioned at the start, the random numbers constitute a random sample from a uniform distribution with parameters 0 and 1.

Thus, the proportion of values lying between c and d where 0 is less than or equal to c which is less than or equal to d which is less than or equal to 1 ($0 \le c \le d \le 1$) would be approximately d-c. These values can be used to obtain random numbers from other distributions besides the uniform distribution. Random numbers are useful in several applications.

To calculate probabilities related to the uniform distribution in Python we can use the scipy.stats.uniform() function, which uses the following basic syntax:

scipy.stats.uniform(x, loc, scale)

where:

```
x:The value of the uniform distribution
loc:The minimum possible value
loc + scale: The maximum possible value
```

Example:

Suppose a bus arrives at a stop every 20 minutes. What is the probability that the bus will arrive in 8 minutes or less if you arrive at the bus stop?

In [1]:
```
from scipy.stats import uniform

# Calculate uniform probability
uniform.cdf(x=8, loc=0, scale=20) - uniform.cdf(x=0, loc=0, scale=20)
```

Out[1]:
```
0.4
```

The probability that the bus arrives in 8 minutes or less is 0.4.

Probability Distribution: Bernoulli Distribution

Bernoulli distribution is a discrete probability distribution which takes up only two distinct values: 1 and 0

1 indicates success

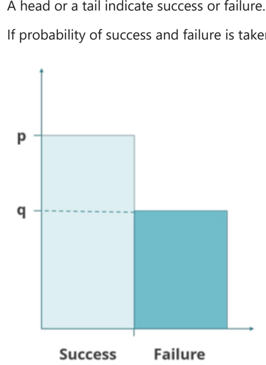
0 indicates failure

Example 1:

A coin toss is a variable that follows Bernoulli distribution. It can be head or tail.

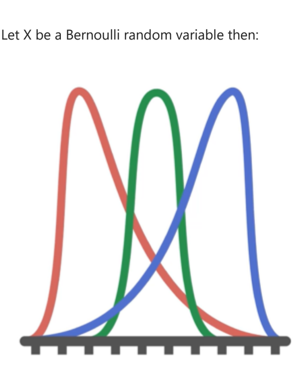
A head or a tail indicate success or failure. As a result, both of these probability range from 0 to 1.

If probability of success and failure is taken as p and q respectively then:



p → Probability of success
q → Probability of failure
p= 0<p<1
q= 1-p
0<q<1

Let X be a Bernoulli random variable then:



$Pr(X=1) = p$

$Pr(X=0) = 1 - p = q$

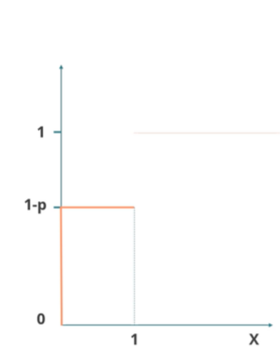
Let probability mass function be f and and possible outcomes for the distribution be k, then:

$f(k;p) = p^k (1-p)^{1-k}, k \{0,1\}$

Next,Cumulative distribution function is:

Cumulative Distribution Function

F(X) = 0, if X < 0
= 1-p, if 0 X ≤ 1
= 1, if X ≥ 1



F(X) is equal to 0 if x is less than 0; 1 minus p if x lies between 0 and 1 and 1 if x is greater than or equal to 1.

Mean of X

E(X) = Pr(X=1)*1 + Pr(X=0)*0
= p*1+ (1-p)*0

The mean value of this Bernoulli random variable is:

Variance is:

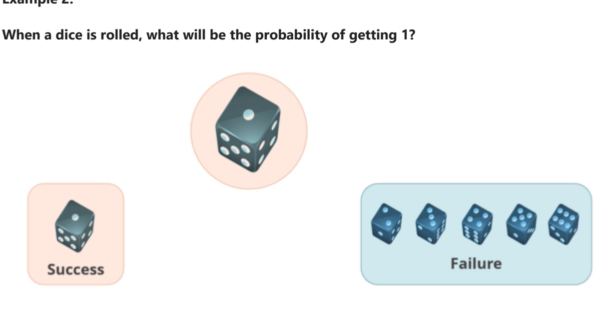
Variance of X

$E(X^2) = Pr(X=1)*1^2 + Pr(X=0)*0^2$
$= p$

$Var(X) = E(X^2) - E(X)^2$
$= p - p^2$
$= p (1-p)$
$= pq$

Example 2:

When a dice is rolled, what will be the probability of getting 1?



In this example, getting 1 is success and all other outcomes are failure.

The probability of getting 1 is ⅙. So p=⅙

Probability of failure q=⅚

So as explained above, mean of this distribution is p= ⅙

And variance is pq = ⅙* ⅚ = 5/36

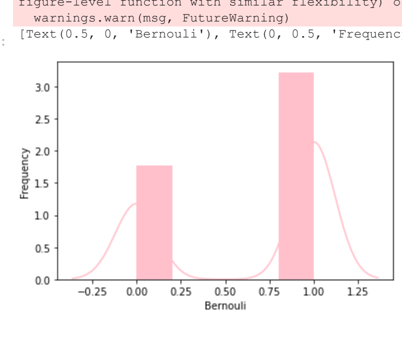
Note: scipy.stats.bernoulli() is a Bernoulli discrete random variable. It is inherited from the of generic methods as an instance of the rv_discrete class. It completes the methods with details specific for this particular distribution.

Python code to implement histograms to obtain a plot for the probability distribution curve

In [4]:
```
from scipy.stats import bernoulli
import seaborn as sb

data_bern = bernoulli.rvs(size=1000,p=0.6)
ax = sb.distplot(data_bern,
                 kde=True,
                 color='pink',
                 hist_kws={"linewidth": 22,'alpha':1})
ax.set(xlabel='Bernouli', ylabel='Frequency')
```

```
C:\Users\alpika.gupta\Anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[4]:
```
[Text(0.5, 0, 'Bernouli'), Text(0, 0.5, 'Frequency')]
```



Probability Density Function and Mass Function

The probability density function is used to specify the probability that the random variable falls within a range of values as opposed to taking on one particular value.

Since there are infinite possible values for a continuous random variable, the absolute likelihood that it will take up a particular value is 0.

But if we take two samples, we can determine how likely it is to fall in one sample compared to the other.

The probability is given by the integral of the probability density function over that range. It is the area under the density function above the horizontal axis between the lowest and highest values of that range.

The probability density function is non-negative everywhere and its integral over the entire space is equal to 1.

But when the random variables take only discrete values, the same function is called Probability Mass Function.

So formally, for a continuous random variable X with probability density function f(x):

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

The Probability Density Function is non-negative for all values of x.

So, $f(x) \geq 0$ for all x.

The area under the density curve and above the horizontal axis over the entire space is equal to 1. So,

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Example:

If X is a continuous random variable with the following density function:

$f(x) = x^2 / 9$ if $0 < x < 3 = 0$

To calculate P(1 < X < 2) as per definition

P(1<X<2) = $\int_1^2 x^2/9dx$ = (2^3/27)= (1^3/27)= 7/27

Cumulative Distribution Function

Probability Mass Function can't be defined for continuous random variables, Cumulative Distribution Function is used to describe their distribution. It can be used to describe:

The cumulative distribution function of a real valued random variable X evaluated at x is the probability that X will take a value less than or equal to x.

So formally, the Cumulative Distribution Function of a random variable X is defined as:

$F_x(x) = P(X \le x)$, for all x∈R

The CDF is monotone increasing. It means that

if $x_1 \leq x_2$, then $F_x(x_1) \leq F_x(x_2)$

Also, the CDF of a continuous random variable can be expressed as an integral of its probability density function F_x as follows:

$FX(x) = \int_{-\infty}^{\infty} f_x\,(t)dt$

CDF Properties:

Each CDF is right-continuous and non-decreasing. Furthermore,

$\lim_{x \to -\infty} f_x(x)$ = 0 and $\lim_{x \to +\infty} f_x(x)$ = 1

Example 1: Suppose X is uniformly distributed on the unit interval [0,1] then its CDF is given by:

$F_x(x) = 0$ if $x < 0$

$= x$ if $0 \le x \le 1$

$= 1$ if $x > 1$

Example 2: Suppose X takes only the discrete values 0 and 1 with equal probability, then its CDF is given by:

$F_x(x) = 0$ if $x < 0$

$= 1/2$ if $0 \le x < 1$

$= 1$ if $x \ge 1$

Central Limit Theorem

Scenario 1:

Consider a math's test score of a school of 1000 students.If we test 100 of them, the results won't significantly deviate from the results of the entire student population. According to the central limit theorem, the average test result for these 100 students will typically be the same as the average test result for the population of 1000 students.

Conversely, assume test scores for 1000 students but conduct the test for 100 students.We can reasonably conclude that the average test score for these 100 students reflects the population mean.

Again, suppose we have data for a particular sample and a population. The central limit theorem helps us to calculate the probability that a particular sample was drawn from a given population. If that probability is low, we can conclude confidently that the sample is not from that population.

Scenario 2

Consider the two given samples. We can infer whether they were likely drawn from the same population.

The central limit theorem states that given a sufficiently large sample size from a population with a finite variance level, the mean of all sampled variables from the same population will be approximately equal to the mean of the whole population. These samples approximate a normal distribution with their variances being approximately equal to the variance of the population, as the sample size grows.

Central Limit Theorem states that if X1, X2, X3,......, Xn are independent random variables that are identically distributed and have finite mean μ and variance σ^2.

then, if $S_n = X_1 + X_2 + X_3 + \ldots\ldots + X_n$ (n= 1,2,......),

$$\lim_{n\to\infty} P(a \le (S_n - n\mu)/\sigma\sqrt{n} \le b) = (1/\sqrt{2\Pi}) \int_a^b e^{(-u2/2)}\,du$$

That is, the random variable $(S_n - n\mu)/\sigma\sqrt{n}$, which is the standardized variable corresponding to S_n is asymptotically normal.

The theorem is also true under more general conditions.

For example:

When $X_1, X_2, X_3, \ldots\ldots, X_n$ are independent random variables with the same mean, same variance but not necessarily identically distributed.

Bayes' Theorem

Bayes' theorem describes how the conditional probability of each of a set of possible causes for a given observed outcome is computed by knowing the probability of each cause and the conditional probability of the outcome of each cause.

Bayes' theorem can be expressed as a mathematical equation and used to calculate the probability of one event based on its connection with another event. It is also known as Bayes' law or Bayes' rule.

Applications of Bayes' Theorem

Mathematically, Bayes' theorem is expressed as:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

where A and B are events

P(A|B) is the probability of A happening in case B happens.

P(B|A) is the probability of B happening in case A happens.

P(A) is the independent probability of A.

P(B) is the independent probability of B.

Example to illustrate how Bayes' theorem tries to predict one event in case the other is true, as shown below:

Rainy days bring us showers and lightning. Sometimes, there are showers with lightning, and sometimes showers without lightning. Similarly, sometimes there is lightning, but no showers. The Bayes' theorem helps us find how often there is lightning, when there are showers. The solution is expressed as: P(Lightning | Shower).

P(Lightning|Shower) = P(Shower|Lightning) x P(Lightning) / P(Shower)

Here, P(Lightning) is the probability of lightning, and P(Shower) is the probability of showers.

P(Shower|Lightning) is the probability of showers when there is lightning.

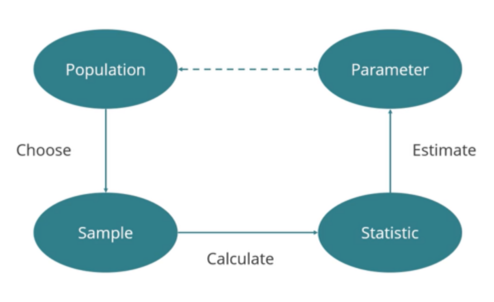
IT is required to know P(Shower|Lightning), which is the probability of showers when there is lightning. This is referred to as 'backwards' of what we want to predict, while what we want to predict is 'forwards'. Therefore, the formula predicts the "forward" event P(Lightning|Shower) by knowing the "backward" event, which is P(Shower|Lightning).

The problem is seasonal, and the various events are related. The Bayes' theorem helps us predict one when the other is known.

Estimation Theory

Estimation theory is the science of guessing or estimating the properties of a population from which data is collected. The guesswork happens via determination of the approximate value of a population parameter based on a sample statistic.

The estimator is a rule or formula that can be used to calculate the estimate based on the sample.



Good estimators:

1) Good estimators are unbiased, that is, the average value of the estimator equals the parameter to be estimated.

2) It also have minimum variance, that is, among all the unbiased estimators, the best one has a sampling distribution with the smallest standard error.

Point estimators and Interval estimators are the two kinds of estimators.

What is Point estimator?

1. A point estimator is a function of the data that is used to deduce the value of an unknown parameter in a statistical model.

1. A point estimate is one of many possible values for the point estimator. For example, the mean income of a class of graduate trainees is $800 per week; this is a point estimate.

1. A point estimator is assessed on three criteria:

- **Unbiasedness (mean):** It is a measurement of whether the mean of this estimator is close to the actual parameter.
- **Efficiency (variance):** It denotes whether the standard deviation of this estimator is close to the actual parameter.
- **Consistency (size):** It indicates whether the probability distribution of the estimator is concentrated on the parameter with an increase in the sample size.

Interval Estimator

An interval estimator of a population parameter under random sampling consists of two random variables, called the upper and lower limits, whose values decide intervals that expect to contain the parameter estimated.

- Interval estimates are all the ranges that an interval estimator can assume.
- There is a range within which a population parameter probably lies, and the interval estimate captures that.
- The mean income of a class of graduate trainees between **$775 and $950** per week is an example of an interval estimate.

An interval estimator can be assessed through its:

1) Accuracy or confidence level

2) Precision or margin of error

The design of an interval estimator consists of evaluating an unbiased point estimator and designating an interval of logical width around it.

Use Case: Finding a Point Estimate

Problem Statement:

An economics researcher is collecting data about grocery store employees in a country. The data represent a random sample of the number of hours worked by 40 employees from several grocery stores in the country. Find a point estimate of the population mean.

Solution:

In [1]:
```
# Random data of forty employees
data = [30, 26, 33, 26, 26, 33, 31, 31, 21, 37,
        27, 20, 34, 35, 30, 24, 38, 34, 39, 31,
        22, 30, 23, 23, 31, 44, 31, 33, 33, 26,
        27, 28, 25, 35, 23, 32, 29, 31, 25, 27]

n = len(data)
sample_mean =  sum(data) / n
print('sample mean: ', sample_mean)
```

```
sample mean:  29.6
```

So, the point estimate for the mean number of hours worked by grocery store employees in this country is 29.6 hours.