Course: Artificial Intelligence and Machine Learning

Code: 20CS51I, WEEK - 4

SESSION – 2 Probability

- Basic concepts
- Conditional and Joint probability
- Bayes' Theorem

Probability Distributions

- Discrete Probability
- Continuous
- Central Limit Theorem

What is Probability?

Probability is a branch of mathematics that deals with calculating the likelihood of a given event's occurrence. It is expressed as a number between 0 and 1.

An event with a probability of 1 can be considered a certainty. An event with a probability of 0 can be considered an impossibility. The higher the probability of an event, the more certain we are that the event will occur.

What is Statistics?

Statistics is a branch of mathematics concerned with the collection, classification, analysis, and interpretation of numerical facts, for drawing inferences on the basis of their quantifiable likelihood (probability).

Statistics can interpret aggregates of data too large to be intelligible by ordinary observation because such data (unlike individual quantities) tend to behave in a regular and predictable manner.

Understanding a chance

Chance is a possibility of something happening.

Examples:

Is it possible that we observe a black colour sun?

Answer: It is not possible to observe a black colour sun. The chance of this happening is zero.

Is it possible that if today is a Monday, tomorrow is a Tuesday?

Answer: Definitely, it is possible. There is a 100 % chance that tomorrow is a Tuesday if it is a Monday today.

Is it possible that we have a pizza for dinner?

Answer: It may or may not be possible, there is some chance. We may not be absolutely sure about having a pizza for dinner.

Observation:

Example 1 and example 2 had a clear possibility of occurrence. However, example 3 had an unclear possibility of occurrence.

When a chance takes us to a situation where the outcome cannot be predicted exactly (i.e. there is an unclear possibility of occurrence), Probability is used to quantify the variability in the outcome.

What is Probability?

By Probability, we measure the likelihood of an event in which we are interested.

What is the Likelihood of an event?

The Likelihood of an event is the frequency with which the event may occur.

Conditional Probability is used to determine the likelihood of an event when a partial information about the event is known.

Conditional probability is defined as the likelihood of an event or outcome occurring, based on the occurrence of a previous event or outcome. Conditional probability is calculated by multiplying the probability of the preceding event by the updated probability of the succeeding, or conditional, event.

Examples:

As an example, suppose you are drawing three marbles—red, blue, and green—from a bag. Each marble has an equal chance of being drawn. What is the conditional probability of drawing the red marble after already drawing the blue one?

First, the probability of drawing a blue marble is about 33% because it is one possible outcome out of three. Assuming this first event occurs, there will be two marbles remaining, with each having a 50% chance of being drawn.

As another example to provide further insight into this concept, consider that a fair die has been rolled and you are asked to give the probability that it was a five. There are six equally likely outcomes, so your answer is 1/6.

But imagine if before you answer, you get extra information that the number rolled was odd. Since there are only three odd numbers that are possible, one of which is five, you would certainly revise your estimate for the likelihood that a five was rolled from 1/6 to 1/3.

Another Example of Conditional Probability

As another example, suppose a student is applying for admission to a university and hopes to receive an academic scholarship. The school to which they are applying accepts 100 of every 1,000 applicants (10%) and awards academic scholarships to 10 of every 500 students who are accepted (2%).

Of the scholarship recipients, 50% of them also receive university stipends for books, meals, and housing. For the students, the chance of them being accepted and then receiving a scholarship is .2% (.1 x .02). The chance of them being accepted, receiving the scholarship, then also receiving a stipend for books, etc. is .1% (.1 x .02 x .5).

In an experiment of rolling a die, the die is rolled twice. It was observed that the sum of the outcomes was even. What is the chance that the first outcome was 2?

In an experiment where a word is to be guessed. It was observed that the first letter of the word is ‘A’. What is the chance that the second letter is ‘b’?

Observation: In the above examples, what we have available is the partial information about the outcome of an experiment. In example 1, the sum is given as even and in example 2, the first letter is given.

Scenario: In an experiment of rolling a die, what is the likelihood that the outcome is 2, given that the outcome is even?

Consider a fair roll of a die i.e. all the outcomes of an experiment have an equal probability of occurrence From the previous example :

$$P(\underbrace{\text{Outcome being 2}}_{\text{Event A}} \mid \underbrace{\text{given that outcome is even}}_{\text{Event B}})$$

It is given that the outcome is even (Event B)– so possible values are 2,4,6. Therefore the total number of elements for Event B =3

Event A can take any ONE value amongst 2,4,6. Therefore the total number of elements for Event A=1 So,

Conditional probability is undefined if the conditioning event has zero probability i.e. P(B)>0 is mandatory.= 1/6.

From the previous example :

It is given that the outcome is even (Event B)– so possible values are 2,4,6. Therefore the total number of elements for Event B =3

Event A can take any ONE value amongst 2,4,6. Therefore the total number of elements for Event A=1 So,

$$P(A|B) = \frac{\text{Number of elements of } A \cap B}{\text{Number of elements of } B} = \frac{1}{3}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3}$$

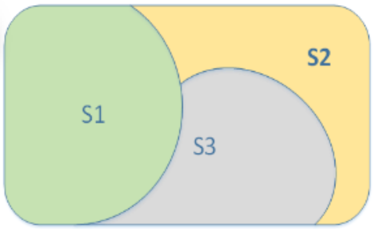
Conditional probability is undefined if the conditioning event has zero probability i.e. P(B)>0 is mandatory.

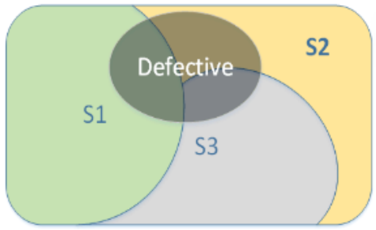
Consider a retailer procures bread from 3 different suppliers 40% from S1, 35% from S2 and 25% from S3.

It was found that 5% of bread supplied by S1, 2% supplied by S2 and 3% supplied by S3 are defective.

No other supplier is supplying bread - exhaustive

All suppliers are different – mutually exclusive

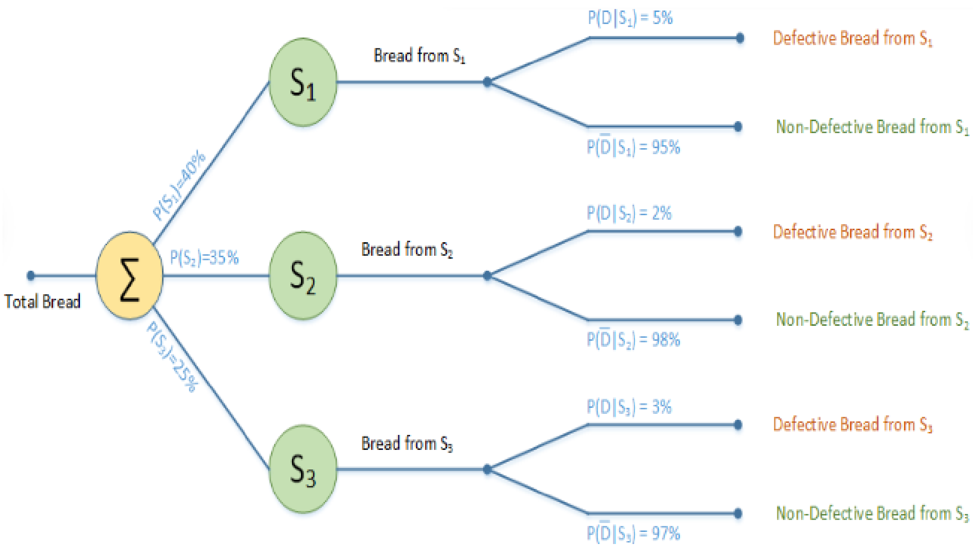




What is the possibility of a customer picking a defective bread?

What is the possibility that a selected defective bread is supplied by S2?

To visualize,



If we have an event let's say D, in which a customer picks a defective bread, what would be the probability of finding a defective bread?

P(D) = P(S1)P(D|S1)+P(S2)P(D|S2)+P(S3)P(D|S3)

=(0.4×0.05)+(0.35×0.02)+(0.25×0.03)

= 0.0345

In general for ‘n’ suppliers

$$P(D) = \sum_{i=1}^{n} P(S_i)P(D|S_i)$$

How would we calculate the probability of the event where a selected defective bread is supplied by S2?

$$P(S_2|D) = \frac{P(S_2)P(D|S_2)}{P(S_1)P(B|S_1) + P(S_2)P(B|S_2) + P(S_3)P(B|S_3)}$$

In general, the probability that the defective bread supplied by supplier ‘k’ (among ‘n’ suppliers, $1 \leq k \leq n$) is

$$P(S_k|D) = \frac{P(S_k)P(D|S_k)}{\sum_{i=1}^{n} P(S_i)P(D|S_i)}$$

Joint Probability

A statistical measure that calculates the likelihood of two events occurring together and at the same point in time is called Joint probability.

Let A and B be the two events, joint probability is the probability of event B occurring at the same time that event A occurs.

Formula for Joint Probability

The following formula represents the joint probability of events with intersection.

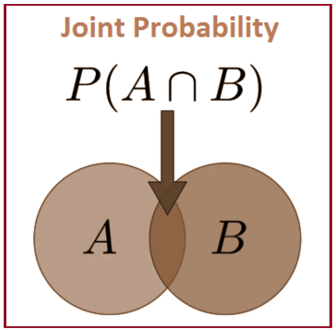
P (A⋂B)

where,

A, B= Two events

P(A and B),P(AB)=The joint probability of A and B

The symbol "∩" in a joint probability is called an intersection. The probability of event A and event B happening is the same thing as the point where A and B intersect. Hence, the joint probability is also called the intersection of two or more events. We can represent this relation using a Venn diagram as shown below.



Problem Statement 1:

In a class, it was observed out of the total strength there are 50 boys and 40 girls. Their teacher wants to select 1 boy and 1 girl for representing the class in a function. So, what would be the total number of ways that the teacher can make this selection?

Problem Statement 2:

In a university, there is a student who has 3 library tickets and there are 8 books of his interest in the library. Out of these 8, he does not want to borrow a book named Java Advance, unless Java Beginners is also borrowed. In how many ways can he select the three books to be borrowed?

Problem Statement 3:

If we have a total of 6 different colour flags. What would be the total number of signals that can be sent taking one or more at a time?

Statistics

The health and welfare council wants to survey your city to understand the lifestyle of the residents and if need be, improve the facilities provided based on the conclusions drawn from the data collected.

Some of the data being collected for this analysis is listed below:

- Personal information such as name, age, income of household and educational qualifications.
- Quality of water supply across the city.

- Medical facilities such as availability of hospitals, doctors etc.

The collection, analysis, interpretation, presentation and organization of data is termed as statistics.

Statistics is used to study a population/process (data).

In the example cited earlier, the data is personal information of residents, water being supplied and information about the medical facilities.

We can leverage the analysis of this data to answer a few questions which could be but not limited to:

- What is the average income of households in the city?
- Which is the most common disease in a given area?
- How much water does a household use in a month?
- How good is the water being supplied?
- Are there enough medical facilities being provided?

When all data required for observation/analysis is collected and studied, the data is referred to as the population.

On the other hand, when limited data is being collected/analyzed, this data is referred to as sample and is used as an indicative of the entire population.

Example:

Population: Data is being collected from all the residents of the city about their age, educational qualification and medical history.

Sample: On the other hand to analyze the water supply, a few litres of water are being collected from different water supply lines that run in the city.

Types of statistics

Statistics can be broadly classified into descriptive and inferential statistics.

Descriptive statistics

Summarization of data to describe the main features of the sample.

For example, in the survey cited earlier, the knowledge of descriptive statistics can be leveraged to answer some questions like:

- What is the average income of households in the city?

- Which is the most common disease in a given area?

Inferential statistics

When working with samples of the population the techniques and processes we use to draw conclusions come under inferential statistics let's understand it as we progress through this course.

For example, in the survey cited earlier, inferential statistics can be used to answer the following questions:

- How good is the water being supplied?
- Are there enough medical facilities being provided?

Bayes' theorem

Bayes' theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of conditional probability. Bayes theorem is also known as the formula for the probability of "causes". For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Bayes Theorem Statement

Let E1, E2,…, En be a set of events associated with a sample space S, where all the events E1, E2,…, En have nonzero probability of occurrence and they form a partition of S. Let A be any event associated with S, then according to Bayes theorem,

$$P(E_i \mid A) = \frac{P(E_i)P(A \mid E_i)}{\sum_{k=1}^{n} P(E_k)P(A|E_k)}$$

for any k = 1, 2, 3, …., n

Hypotheses: The events E1, E2,… En is called the hypotheses

Priori Probability: The probability P(Ei) is considered as the priori probability of hypothesis Ei

Posteriori Probability: The probability P(Ei|A) is considered as the posteriori probability of hypothesis Ei

Bayes' theorem is also called the formula for the probability of "causes". Since the Ei's are a partition of the sample space S, one and only one of the events Ei occurs (i.e. one of the events Ei must occur and the only one can occur). Hence, the above formula gives us the probability of a particular Ei (i.e. a "Cause"), given that the event A has occurred.

Bayes Theorem Formula

If A and B are two events, then the formula for the Bayes theorem is given by:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad where \;\; P(B) \neq 0$$

Where P(A|B) is the probability of condition when event A is occurring while event B has already occurred.

Examples and Solutions

Some illustrations will improve the understanding of the concept.

Example 1:

A bag I contains 4 white and 6 black balls while another Bag II contains 4 white and 3 black balls. One ball is drawn at random from one of the bags, and it is found to be black. Find the probability that it was drawn from Bag I.

Solution:

Let E1 be the event of choosing bag I, E2 the event of choosing bag II, and A be the event of drawing a black ball.

Then,

$$P(E_1) = P(E_2) = \frac{1}{2}$$

Also, P(A|E1) = P(drawing a black ball from Bag I) = 6/10 = 3/5

P(A|E2) = P(drawing a black ball from Bag II) = 3/7

By using Bayes' theorem, the probability of drawing a black ball from bag I out of two bags,

$$P(E_1|A) = \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)}$$

$$= \frac{\frac{1}{2} \times \frac{3}{5}}{\frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{3}{7}}$$

$$= \frac{7}{12}$$

Example 2:

A man is known to speak the truth 2 out of 3 times. He throws a die and reports that the number obtained is a four. Find the probability that the number obtained is actually a four.

Solution:

Let A be the event that the man reports that number four is obtained.

Let E1 be the event that four is obtained and E2 be its complementary event.

Then, P(E1) = Probability that four occurs = 1/6.

P(E2) = Probability that four does not occur = 1- P(E1) = 1 – (1/6) = 5/6.

Also, P(A|E1)= Probability that man reports four and it is actually a four = 2/3

P(A|E2) = Probability that man reports four and it is not a four = 1/3.

By using Bayes' theorem, probability that number obtained is actually a four, P(E1|A)

$$= \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)} = \frac{\frac{1}{6} \times \frac{2}{3}}{\frac{1}{6} \times \frac{2}{3} + \frac{5}{6} \times \frac{1}{3}}$$

$$= \frac{2}{7}$$

Bayes Theorem Applications

One of the many applications of Bayes' theorem is Bayesian inference, a particular approach to statistical inference. Bayesian inference has found application in various activities, including medicine, science, philosophy, engineering, sports, law, etc. For example, we can use Bayes' theorem to define the accuracy of medical test results by considering how likely any given person is to have a disease and the test's overall accuracy. Bayes' theorem relies on consolidating prior probability distributions to generate posterior probabilities. In Bayesian statistical inference, prior probability is the probability of an event before new data is collected.

Practice Problems

Solve the following problems using Bayes Theorem.

A bag contains 5 red and 5 black balls. A ball is drawn at random, its colour is noted, and again the ball is returned to the bag. Also, 2 additional balls of the colour drawn are put in the bag. After that, the ball is drawn at random from the bag. What is the probability that the second ball drawn from the bag is red?

Of the students in the college, 60% of the students reside in the hostel and 40% of the students are day scholars. Previous year results report that 30% of all students who stay in the hostel scored A Grade and 20% of day scholars scored A grade. At the end of the year, one student is chosen at random and found that he/she has an A grade. What is the probability that the student is a hosteler?

From the pack of 52 cards, one card is lost. From the remaining cards of a pack, two cards are drawn and both are found to be diamond cards. What is the probability that the lost card is a diamond?

Probability Distributions

What is a continuous distribution?

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.

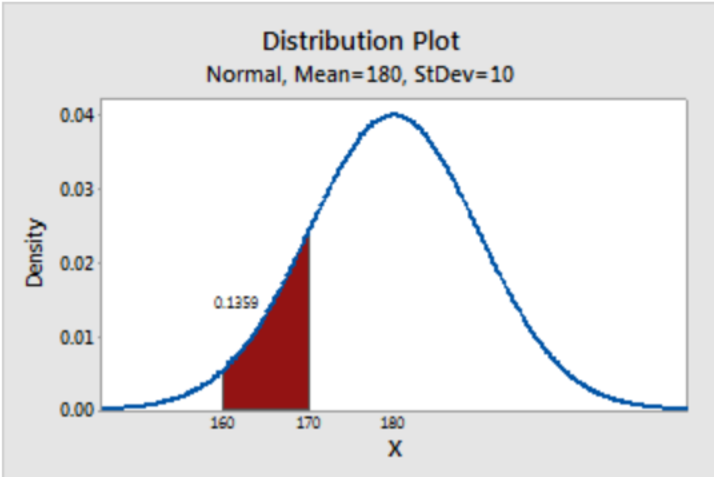
Probabilities of continuous random variables (X) are defined as the area under the curve of its PDF. Thus, only ranges of values can have a nonzero probability. The probability that a continuous random variable equals some value is always zero.

Example of the distribution of weights

The continuous normal distribution can describe the distribution of weight of adult males. For example, you can calculate the probability that a man weighs between 160 and 170 pounds.

Distribution plot of the weight of adult males

The shaded region under the curve in this example represents the range from 160 and 170 pounds. The area of this range is 0.136; therefore, the probability that a randomly selected man weighs between 160 and 170 pounds is 13.6%. The entire area under the curve equals 1.0.



However, the probability that X is exactly equal to some value is always zero because the area under the curve at a single point, which has no width, is zero. For example, the probability that a man weighs exactly 190 pounds to infinite precision is zero. You could calculate a nonzero probability that a man weighs more than 190 pounds, or less than 190 pounds, or between 189.9 and 190.1 pounds, but the probability that he weighs exactly 190 pounds is zero.

Since quantile is the inverse of CDF, given the probability, we can compute the weight that the ice-cream scoop might take from the interval [101.7, 124.3], as follows:

```
#Continuous Uniform Distribution: Quantile

from scipy.stats import uniform

x= 105.0 #change the values to any value in the interval [101.7, 124.3] and observe the output

low= 101.7 #lower limit of distribution

diff= 22.6 #upper limit = low + diff

#The probability that the weight of a single scoop of ice-cream takes the value 105 from

#the interval [101.7 grams ,124.3 grams] , can be computed as follows:

punif = uniform.cdf(x,loc = low,scale = diff)

#Since quantile is the inverse of CDF, given the probability,

#we can compute the weight that the ice-cream scoop might take from the interval [101.7, 124.3], as follows:
```

qunif = uniform.ppf(punif,loc = low,scale = diff)

qunif

Output: 105.0

We can generate random values analogous to observations of a continuous uniform experiment as follows:

#Continuous Uniform Distribution: Random Generation

from scipy.stats import uniform

low= 101.7 #lower limit of distribution

diff= 22.6 #upper limit = low + diff

#We can generate random values analogous to observations of a continuous uniform experiment as follows:

#The example below shows the random generation of 50 values which lie in the interval [101.7, 124.3].

runif = uniform.rvs(loc = low,scale = diff, size = 50)

runif

Output:

```
array([111.74043187, 102.59191934, 116.05724232, 110.69327256,
       104.1636733 , 117.76719557, 117.81060091, 117.63881238,
       119.46625942, 121.78518631, 117.84494571, 114.82481422,
       113.96164858, 113.71617025, 122.35187651, 103.27691912,
       118.45678217, 111.02593295, 102.83009163, 113.48155871,
       104.89733703, 112.79926966, 107.38620618, 118.01092993,
       107.61155234, 123.42617192, 102.97664222, 111.73862345,
       106.45575318, 104.10070942, 116.47178257, 103.19835684,
       114.7895256 , 122.3025134 , 102.69016823, 120.57348655,
       122.00286771, 104.58981127, 103.75017103, 105.12221767,
       114.31901729, 109.7265801 , 117.18544862, 122.7356993 ,
       112.02770252, 105.81218868, 117.57683009, 108.16432034,
       105.65527171, 108.33135846])
```

The example above shows the random generation of 50 values which lie in the interval [101.7, 124.3].

What is a discrete distribution?

A discrete distribution describes the probability of occurrence of each value of a discrete random variable. A discrete random variable is a random variable that has countable values, such as a list of non-negative integers.

With a discrete probability distribution, each possible value of the discrete random variable can be associated with a non-zero probability. Thus, a discrete probability distribution is often presented in tabular form.

Example of the number of customer complaints

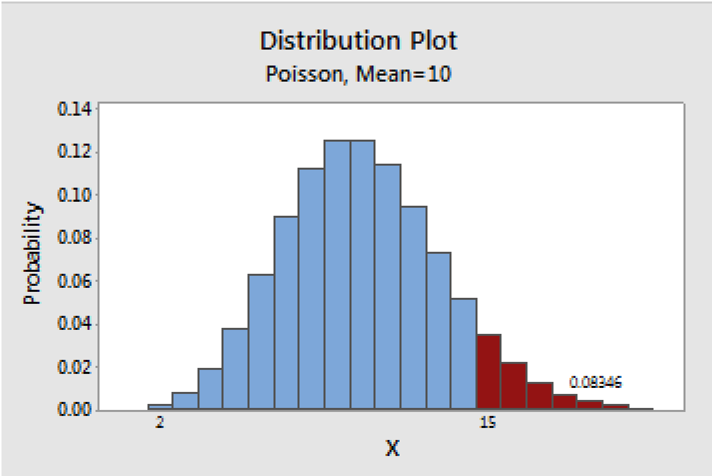
With a discrete distribution, unlike with a continuous distribution, you can calculate the probability that X is exactly equal to some value. For example, you can use the discrete Poisson distribution to describe the number of customer complaints within a day. Suppose the average number of complaints per day is 10 and you want to know the probability of receiving 5, 10, and 15 customer complaints in a day.

x	P (X = x)
5	0.037833
10	0.125110
15	0.034718

You can also view a discrete distribution on a distribution plot to see the probabilities between ranges.

Distribution plot of the number of customer complaints

The shaded bars in this example represents the number of occurrences when the daily customer complaints is 15 or more. The height of the bars sums to 0.08346; therefore, the probability that the number of calls per day is 15 or more is 8.35%.



The Discrete Uniform Distribution

#Discrete Uniform Distribution

from scipy.stats import randint

x = 1 #change the values to 2,3,4,5,6 and check

low = 1

high = 6 #total no. of possible outcomes + 1

#The flavor of ice-cream selected by a customer is equally likely to be any one of the 5 available flavors.

#The cummulative distribution can be computed as follows:

prandint = randint.cdf(x, low, high)

print("When x=1, CDF is", prandint)

Output:

```
When x=1, CDF is 0.2
```

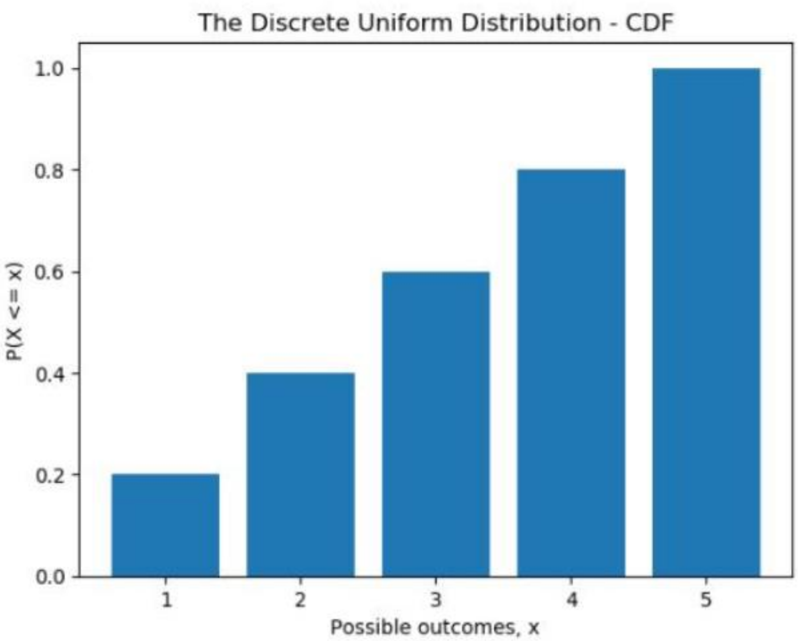
The CDF graph for discrete uniform distribution can be plotted as follows:

import matplotlib.pyplot as plt; plt.rcdefaults()

import numpy as np

import matplotlib.pyplot as plt

```
low = 1

high = 6

x = np.arange(1, 6, 1)

#Plot CDF

y2 = randint.cdf(x, low, high)

plt.bar(x, y2)

plt.xlabel('Possible outcomes, x')

plt.ylabel('P(X <= x)')

plt.title('The Discrete Uniform Distribution - CDF')

 plt.show()
```



Central limit theorem

Central limit theorem is a statistical theory which states that when the large sample size has a finite variance, the samples will be normally distributed and the mean of samples will be approximately equal to the mean of the whole population.

In other words, the central limit theorem states that for any population with mean and standard deviation, the distribution of the sample mean for sample size N has mean μ and standard deviation σ / √n .

As the sample size gets bigger and bigger, the mean of the sample will get closer to the actual population mean. If the sample size is small, the actual distribution of the data may or may not be normal, but as the sample size gets bigger, it can be approximated by a normal distribution. This statistical theory is useful in simplifying analysis while dealing with stock indexes and many more.

Central Limit Theorem Statement

The central limit theorem states that whenever a random sample of size n is taken from any distribution with mean and variance, then the sample mean will be approximately normally distributed with mean and variance. The larger the value of the sample size, the better the approximation to the normal.

Assumptions of Central Limit Theorem

- The sample should be drawn randomly following the condition of randomization.
- The samples drawn should be independent of each other. They should not influence the other samples.
- When the sampling is done without replacement, the sample size shouldn't exceed 10% of the total population.
- The sample size should be sufficiently large.

Formula

Central Limit Theorem for Sample Means,

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

Steps

The steps used to solve the problem of the central limit theorem that are either involving '>' '<' or "between" are as follows:

1) The information about the mean, population size, standard deviation, sample size and a number that is associated with "greater than", "less than", or two numbers associated with both values for a range of "between" is identified from the problem.

2) A graph with a centre as mean is drawn.

3)

The formula $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is used to find the z-score.

4) The z-table is referred to find the 'z' value obtained in the previous step.

5) Case 1: Central limit theorem involving ">".

Subtract the z-score value from 0.5.

Case 2: Central limit theorem involving "<".

Add 0.5 to the z-score value.

Case 3: Central limit theorem involving "between".

Step 3 is executed.

6) The z-value is found along with x bar.

The last step is common to all three cases, that is to convert the decimal obtained into a percentage.