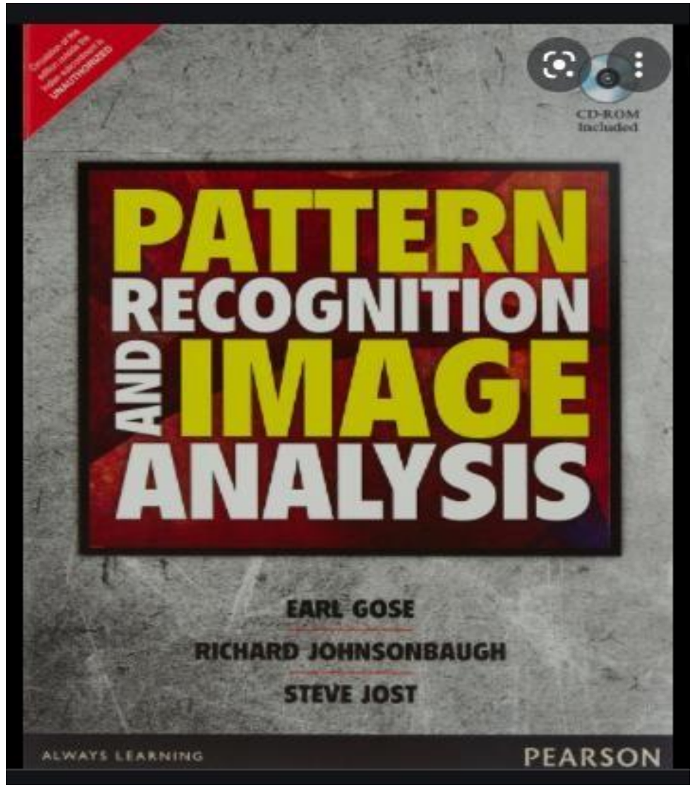
Text Books:



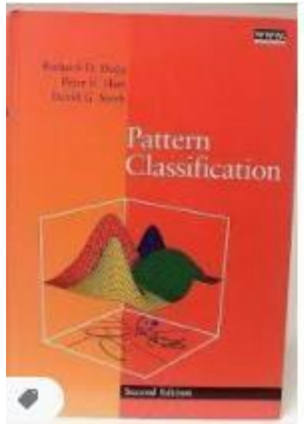
Text Books:

Sl. No.	Author/s	Title	Publisher Details
1	Earl Gose, Richard Johnsonbaugh, Steve Jost.	Pattern recognition and Image analysis	Pearson 2015

Reference Books and Web sources

Reference Books:

Sl. No.	Author/s	Title	Publisher Details
1	Richard O.Duda, Peter E.Hart, David G. Stork	Pattern Classification	John Wiley publication, 2nd edition, 2001.
2	A.K.Jain, R.Bolle, S.Pankanti	Biometric: Personal Identification in network society	Kluwer academic publishers, 1999.
3	Robert Schalkoff	Pattern Recognition: Statistical, Structural and Neural Approaches	John Wiley & Sons, Inc.1992.
4	Christopher M. Bishop	Pattern Recognition and Machine Learning	Springer publication, 2006



Web Resources:

Sl. No.	Web link
1	https://nptel.ac.in/courses/106/108/106108057/
2	https://nptel.ac.in/courses/106/106/106106046/

UNIT - 1

- Introduction: Applications of Pattern Recognition, Statistical Decision Theory and Analysis. Probability: Introduction to probability, Probabilities of Events, Random Variables, Joint Distributions and Densities, Moments of Random Variables, Estimation of Parameters from samples

Introduction: Definition

- Pattern recognition is the theory or algorithm concerned with the automatic detection (recognition) and later classification of objects or events using a machine/computer.

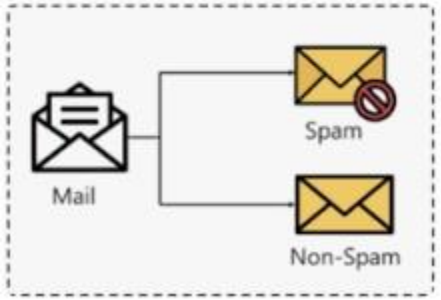
- Applications of Pattern Recognition
- Some examples of the problems to which pattern recognition techniques have been applied are:
  - Automatic inspection of parts of an assembly line
  - Human speech recognition
  - Character recognition
  - Automatic grading of plywood, steel, and other sheet material
  - Identification of people from
    - finger prints,
    - hand shape and size,
    - Retinal scans
    - voice characteristics,
    - Typing patterns and
    - handwriting
  - Automatic inspection of printed circuits and printed characters
  - Automatic analysis of satellite picture to determine the type and condition of agricultural crops, weather conditions, snow and water reserves and mineral prospects.
  - Classification and analysis in medical images. : to detect a disease

Features and classes

- Properties or attributes used to classify the objects are called **features.**
- A collection of “similar” (not necessarily same) objects are grouped together as one “class”.
- For example:

- All the above are classified as character T
- Classes are identified by a label.
- Most of the pattern recognition tasks are first done by humans and automated later.
- Automating the classification of objects using the same features as those used by the people can be difficult.
- Some times features that would be impossible or difficult for humans to estimate are useful in automated system. For example satellite images use wavelengths of light that are invisible to humans.

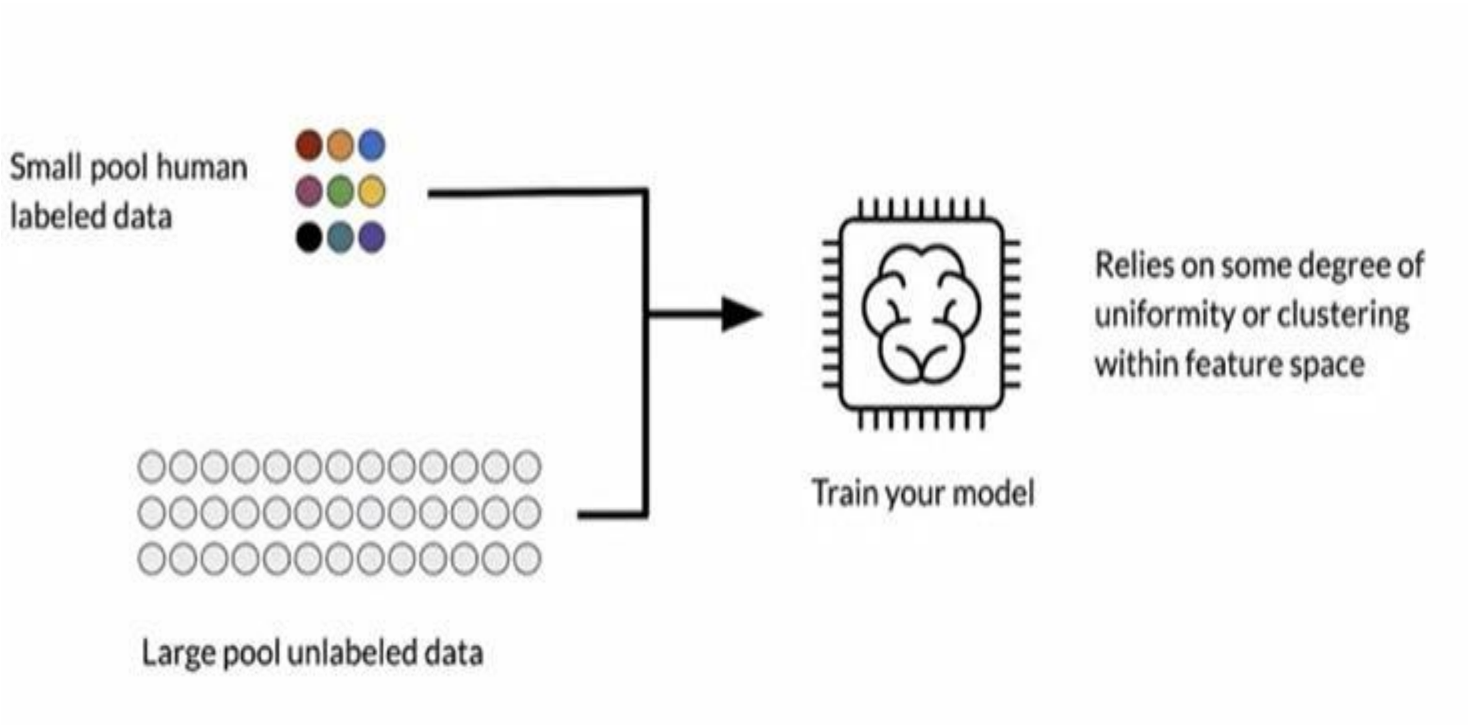
Two broad types of classification



- Supervised classification
  - Guided by the humans
  - It is called supervised learning because **the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process**.
  - We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher.
  - Classify the mails as span or non span based on redecided parameters.
- Unsupervised classification
  - Not guided by the humans.
  - Unsupervised Classification is called **clustering.**

Another classifier : Semi supervised learning

It makes use of a small number of labeled data and a large number of unlabeled data to learn



Samples or patterns

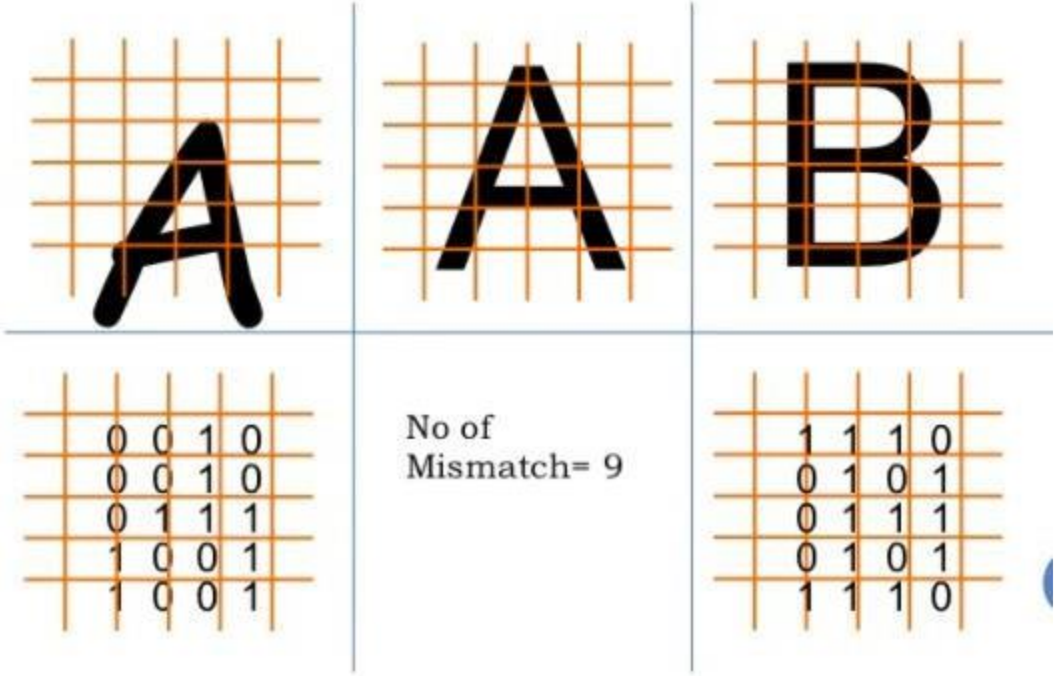
- The individual items or objects or situations to be classified will be referred as **samples or patterns or data.**
- **The set of data is called "Data Set".**

Training and Testing data

- Two types of data set in supervised classifier.
  - Training set : 70 to 80% of the available data will be used for training the system.
  - In Supervised classification Training data is **the data you use to train an algorithm or machine learning model to predict the outcome you design your model to predict**.
  - Testing set : around 20-30% will be used for testing the system. Test data is used to measure the performance, such as accuracy or efficiency, of the algorithm you are using to train the machine.
  - Testing is the measure of quality of your algorithm.
  - Many a times even after 80% testing, failures can be see during testing, reason being not good representation of the test data in the training set.
- **Unsupervised classifier does not use training data**

Statistical Decision Theory

- Decision theory, in statistics, **a set of quantitative methods for reaching optimal decisions**.



Example for Statistical Decision Theory

- Consider Hypothetical Basket ball Association:
- The prediction could be based on the difference between the home team's average number of points per game (apg) and the visiting team's 'apg' for previous games.
- The training set consists of scores of previously played games, with each home team is classified as winner or loser
- Now the prediction problem is : given a game to be played, predict the home team to be a winner or loser using the feature 'dapg',
- Where dapg = Home team apg – Visiting team apg

Game	*dapg*	Home Team	Game	*dapg*	Home Team
1	1.3	Won	16	-3.1	Won
2	-2.7	Lost	17	1.7	Won
3	-0.5	Won	18	2.8	Won
4	-3.2	Lost	19	4.6	Won
5	2.3	Won	20	3.0	Won
6	5.1	Won	21	0.7	Lost
7	-5.4	Lost	22	10.1	Won
8	8.2	Won	23	2.5	Won
9	-10.8	Lost	24	0.8	Won
10	-0.4	Won	25	-5.0	Lost
11	10.5	Won	26	8.1	Won
12	-1.1	Lost	27	-7.1	Lost
13	2.5	Won	28	2.7	Won
14	-4.2	Won	29	-10.0	Lost
15	-3.4	Lost	30	-6.5	Won

Data set of games showing outcomes, differences between average numbers of points scored and differences between winning percentages for the participating teams in previous games

- The figure shown in the previous slide, lists 30 games and gives the value of dapg for each game and tells whether the home team won or lost.
- Notice that in this data set the team with the larger apg usually wins.

- For example in the 9^{th} game the home team on average, scored 10.8 fewer points in previous games than the visiting team, on average and also the home team lost.

- When the teams have about the same apg's the outcome is less certain. For example, in the 10^{th} game , the home team on average scored 0.4 fewer points than the visiting team, on average, but the home team won the match.

- Similarly 12^{th} game, the home team had an apg 1.1. less than the visiting team on average and the team lost.

Histogram of dapg

- Histogram is a convenient way to describe the data.
- To form a histogram, the data from a single class are grouped into intervals.
- Over each interval rectangle is drawn, with height proportional to number of data points falling in that interval. In the example interval is chosen to have width of two units.
- General observation is that, the prediction is not accurate with single feature ‘dgpa’

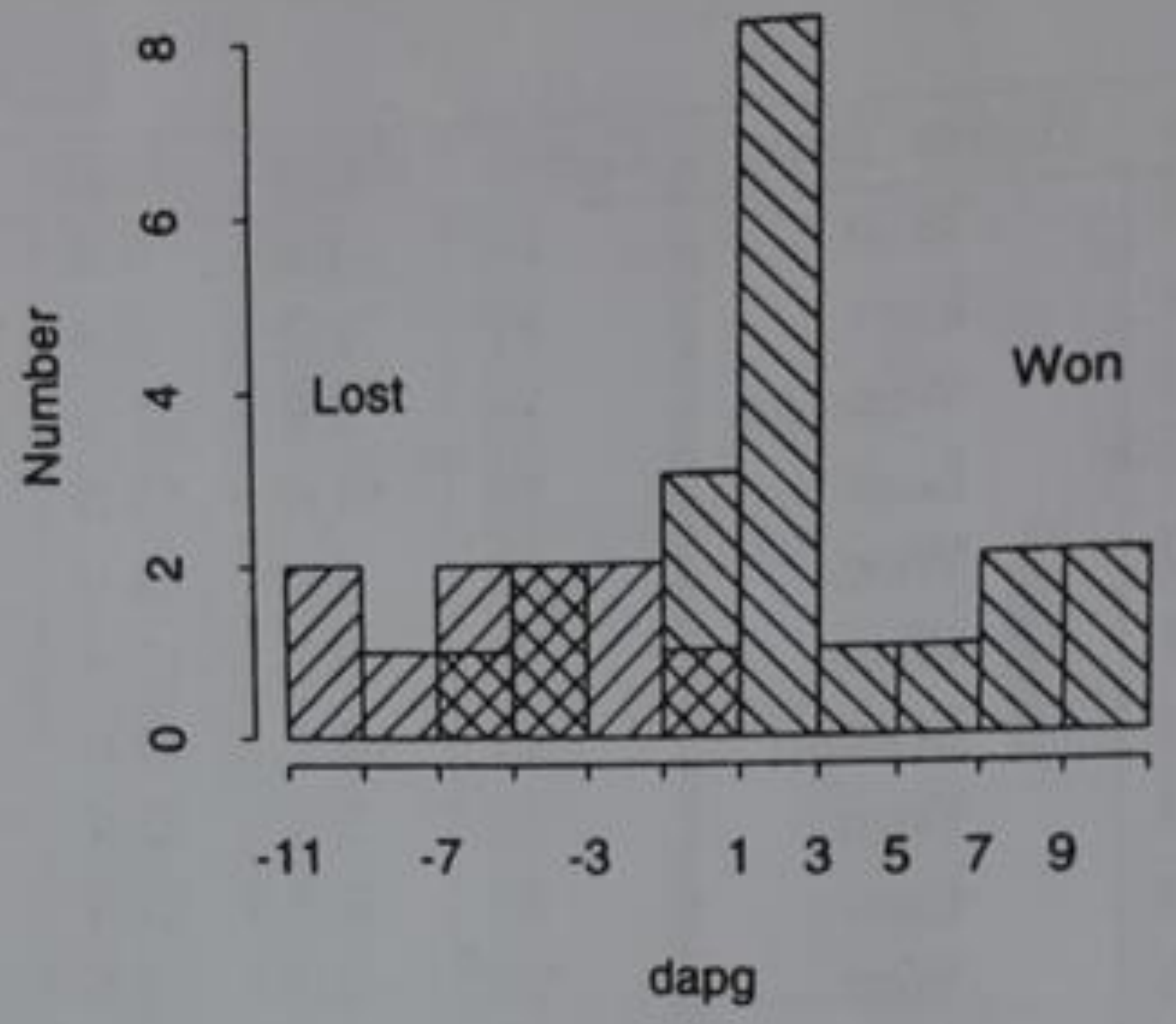


Figure 1.2: Histogram of *dapg*.

Prediction

- To predict normally a threshold value T is used.
- ‘dgpa’ > T consider to be won
- ‘dgpa’ < T consider to be lost

- T is called decision boundary or threshold.
- If T=-1, four samples in the original data are misclassified.
  - Here 3 winners are called losers and one loser is called winner.
- If T=0.8, results in no samples from the loser class being misclassified as winner, but 5 samples from the winner class would be misclassified as loser.
- IF T=-6.5, results no samples from the winner class being misclassified as losers, but 7 samples from the loser would be misclassified as winners.
- By inspection, we see that when a decision boundary is used to classify the samples the minimum number of samples that are misclassified is four.
- In the above observations, the minimum number of samples misclassified is 4 when T=-1

- To make it more accurate let us consider two features.
- Additional features often increases the accuracy of classification.
- Along with ‘dapg’ another feature ‘dwp’ is considered.

- wp= winning percentage of a team in previous games
- dwp = difference in winning percentage between teams
- dwp = Home team wp – visiting team wp

Game	*dapg*	*dwp*	Home Team	Game	*dapg*	*dwp*	Home Team
1	1.3	25.0	Won	16	-3.1	9.4	Won
2	-2.7	-16.9	Lost	17	-1.7	6.8	Won
3	-0.5	5.3	Won	18	2.8	17.0	Won
4	-3.2	-27.5	Lost	19	4.6	13.3	Won
5	2.3	-18.0	Won	20	3.0	-24.0	Won
6	5.1	31.2	Won	21	0.7	-17.8	Lost
7	-5.4	5.8	Lost	22	10.1	44.6	Won
8	8.2	34.3	Won	23	2.5	-22.4	Won
9	-10.8	-56.3	Lost	24	0.8	12.3	Won
10	-0.4	13.3	Won	25	-5.0	-3.8	Lost
11	10.5	16.3	Won	26	8.1	36.0	Won
12	-1.1	-17.6	Lost	27	-7.1	-20.6	Lost
13	2.5	5.7	Won	28	2.7	23.2	Won
14	-4.2	16.0	Won	29	-10.0	-46.9	Lost
15	-3.4	-3.4	Lost	30	-6.5	19.7	Won

Data set of games showing outcomes, differences between average number of points scored and differences between winning percentages for the participating teams in previous games

- Now observe the results on a scatterplot

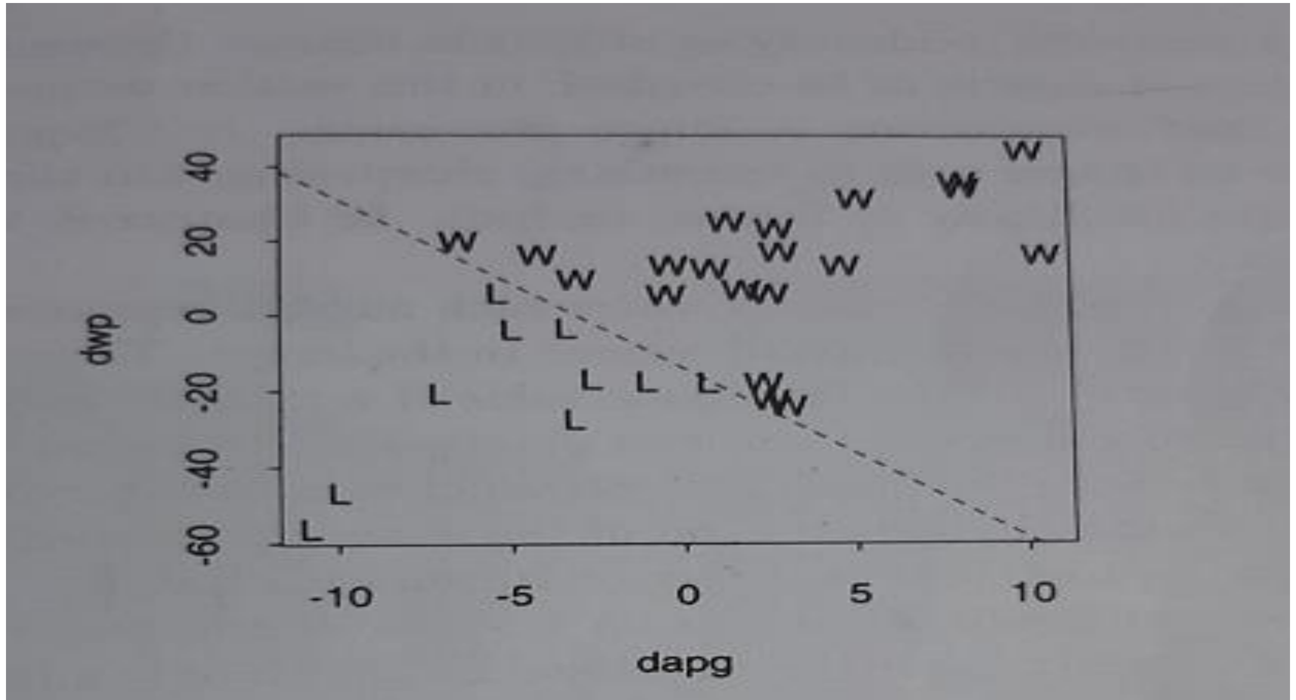


Figure 1.4: A scatterplot of *dapg* versus *dwp*.

- Each sample has a corresponding feature vector (dapg, dwp), which determines its position in the plot.
- Note that the feature space can be classified into two decision regions by a straight line, called a linear decision boundary. (refer line equation). Prediction of this line is logistic regression.
- If the sample lies above the decision boundary, the home team would be classified as the winner and it is below the decision boundary it is classified as loser.

Prediction with two parameters.

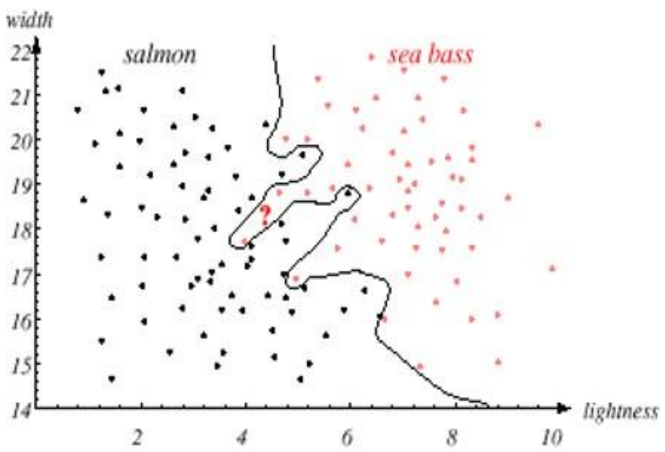
- Consider the following : springfield (Home team)

$$
\begin{aligned}
\text{Springfield's } apg &= 98.3 \\
\text{Centerville's } apg &= 102.9 \\
\text{Springfield's } wp &= 21.4 \\
\text{Centerville's } wp &= 58.1.
\end{aligned}
$$

- dapg= home team apg – visiting team apg = 98.3-102.9 = -4.6
- dwp = Home team wp – visiting team wp = -21.4-58.1 = -36.7

- Since the point (dapg, dwp) = (-4.6,-36.7) lies below the decision boundary, we predict that the home team will lose the game.

- If the feature space cannot be perfectly separated by a straight line, a more complex boundary might be used. (non-linear)



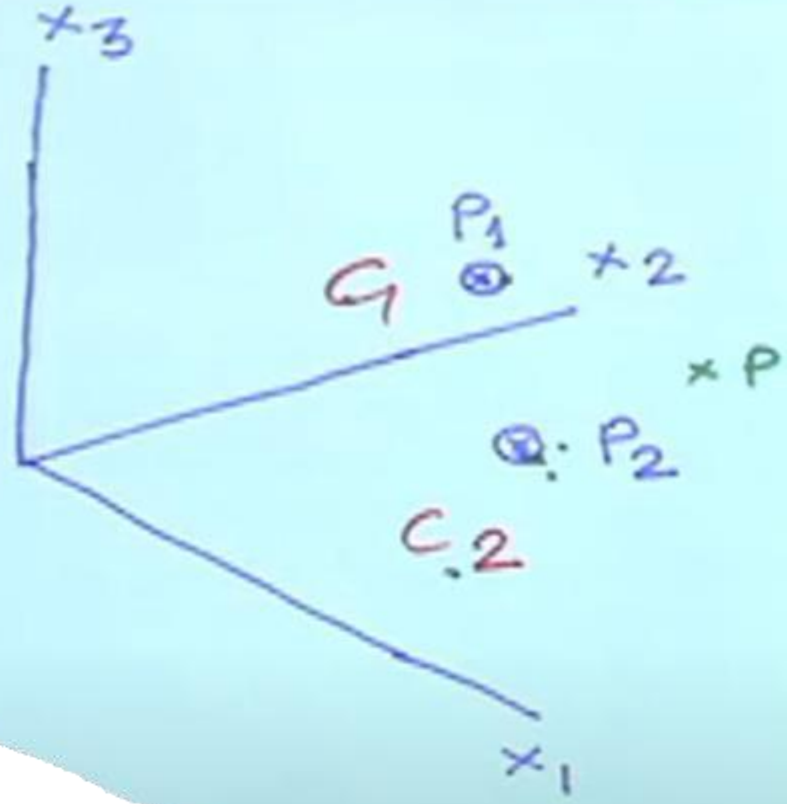
- Alternatively a simple decision boundary such as straight line might be used even if it did not perfectly separate the classes, provided that the error rates were acceptably low.

Simple illustration of Pattern Classification

- A pattern/object can be identified by set of features.
- Collection of features for a pattern forms feature vector.

- Example : (in next slide)
- P1 and P2 are two patterns with 3 features, so 3 Dimensional feature vector.
- There are two classes C1 and C2.
- P1 belongs to C1 and P2 belongs to C2
- Given P, a new pattern with feature vector, it has to be classified into one of the class based on the similarity value.
- If d1 is the distance between (p and p1) and d2 is the distance between (p and p2) then p will be classified into the class having least difference.

$M = 3.$



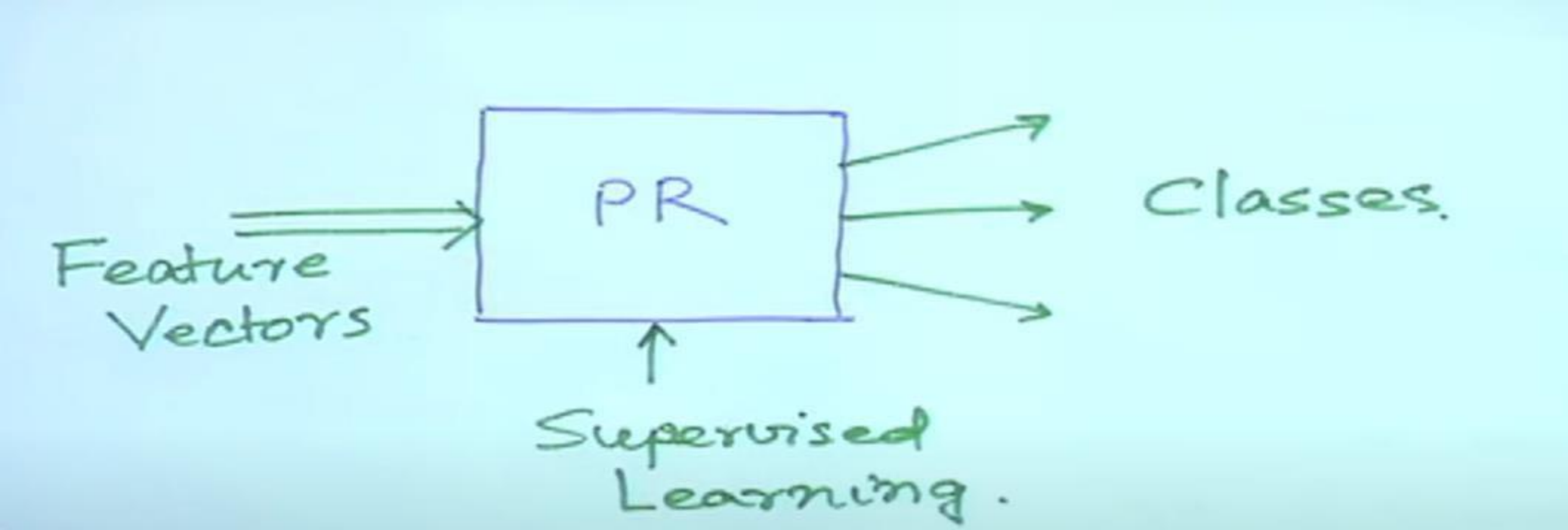
$$P_1 \approx \langle 3, 5, 1 \rangle$$

$$P_2 \approx \langle f_1, f_2, f_3 \rangle$$

$$P \approx \langle x_1, x_2, x_3 \rangle$$

$$d(P_1, P) > d(P_2, P)$$

Block diagram of Pattern recognition and classification



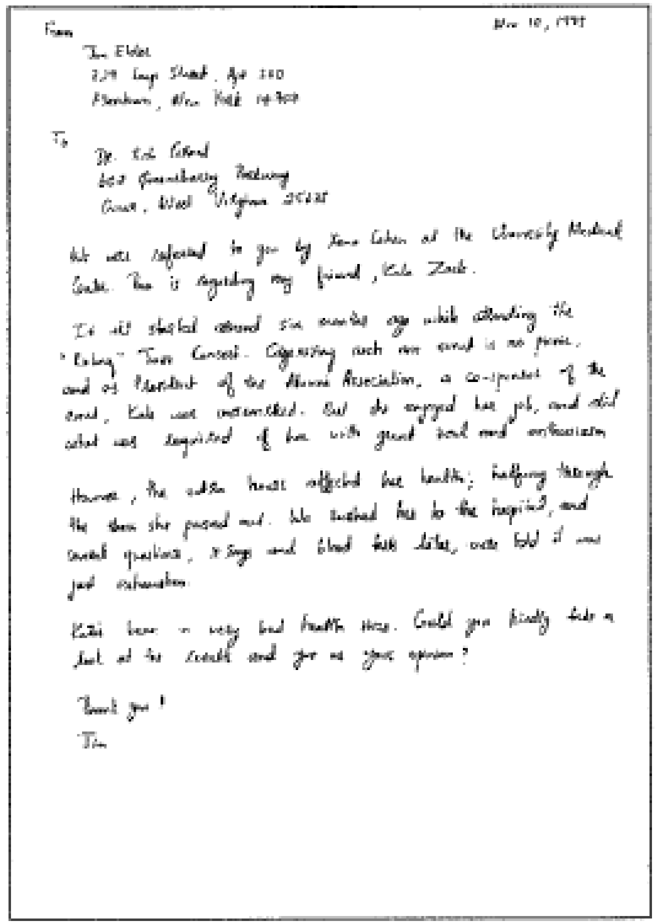
Input to our pattern recognition system will be feature vectors and output will be decision about selecting the classes

- Having the model shown in previous slide, we can use it for any type of recognition and classification.
- It can be
    - speaker recognition
    - Speech recognition
    - Image classification
    - Video recognition and so on…

- It is now very important to learn:
  - Different techniques to extract the features
  - Then in the second stage, different methods to recognize the pattern and classify
    - Some of them use statistical approach
    - Few uses probabilistic model using mean and variance etc.
    - Other methods are - neural network, deep neural networks
    - Hyper box classifier
    - Fuzzy measure
    - And mixture of some of the above

Examples for pattern recognition and classification

Handwriting Recognition



From
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

Nov 10, 1999

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

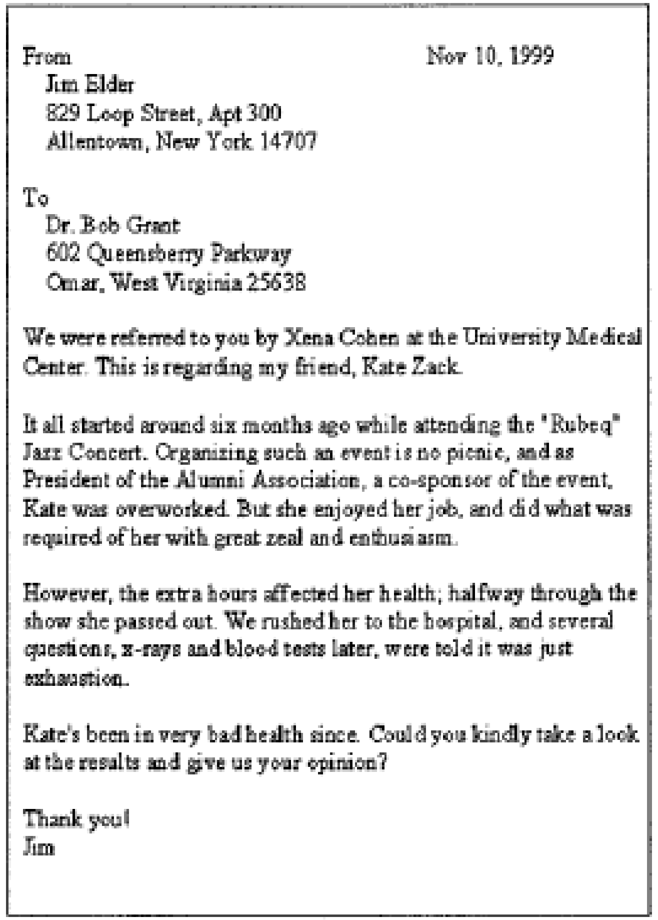
We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim



From
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

Nov 10, 1999

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

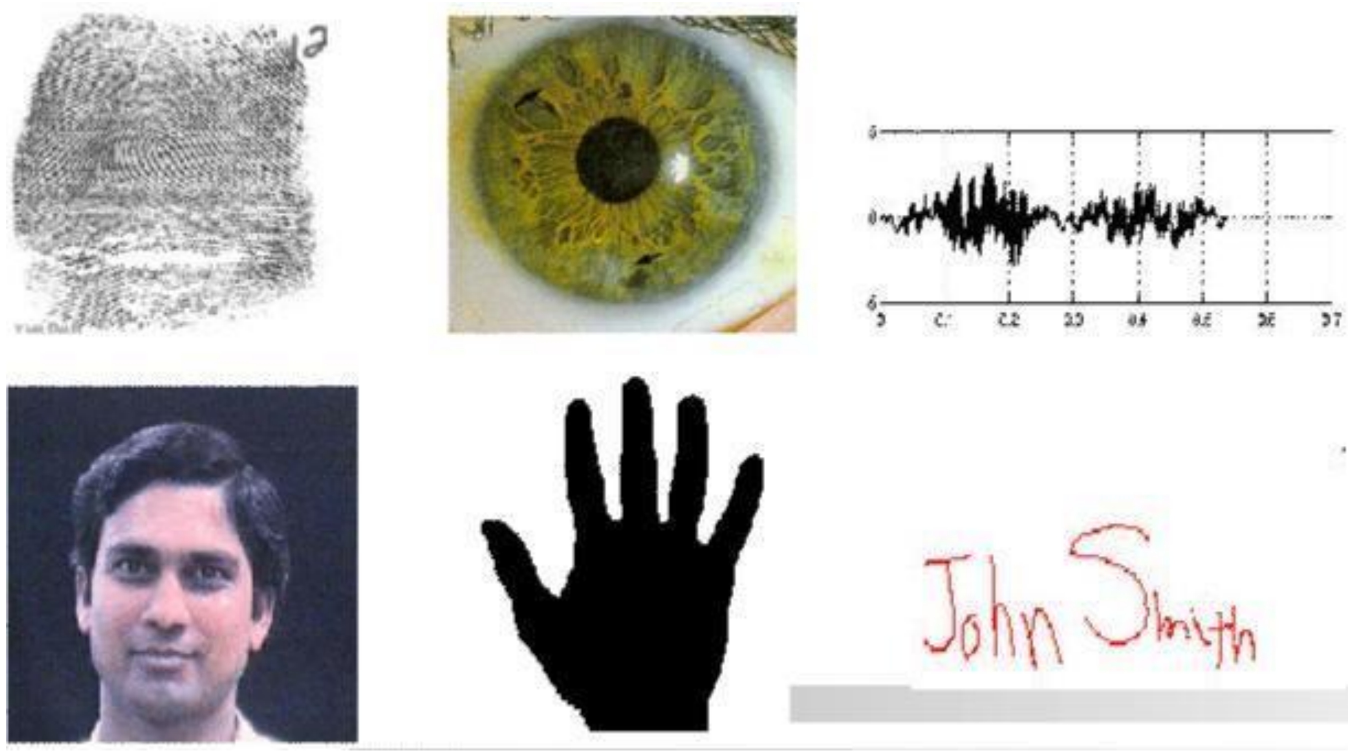
However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim

License Plate Recognition

Biometric Recognition



Face Detection/Recognition

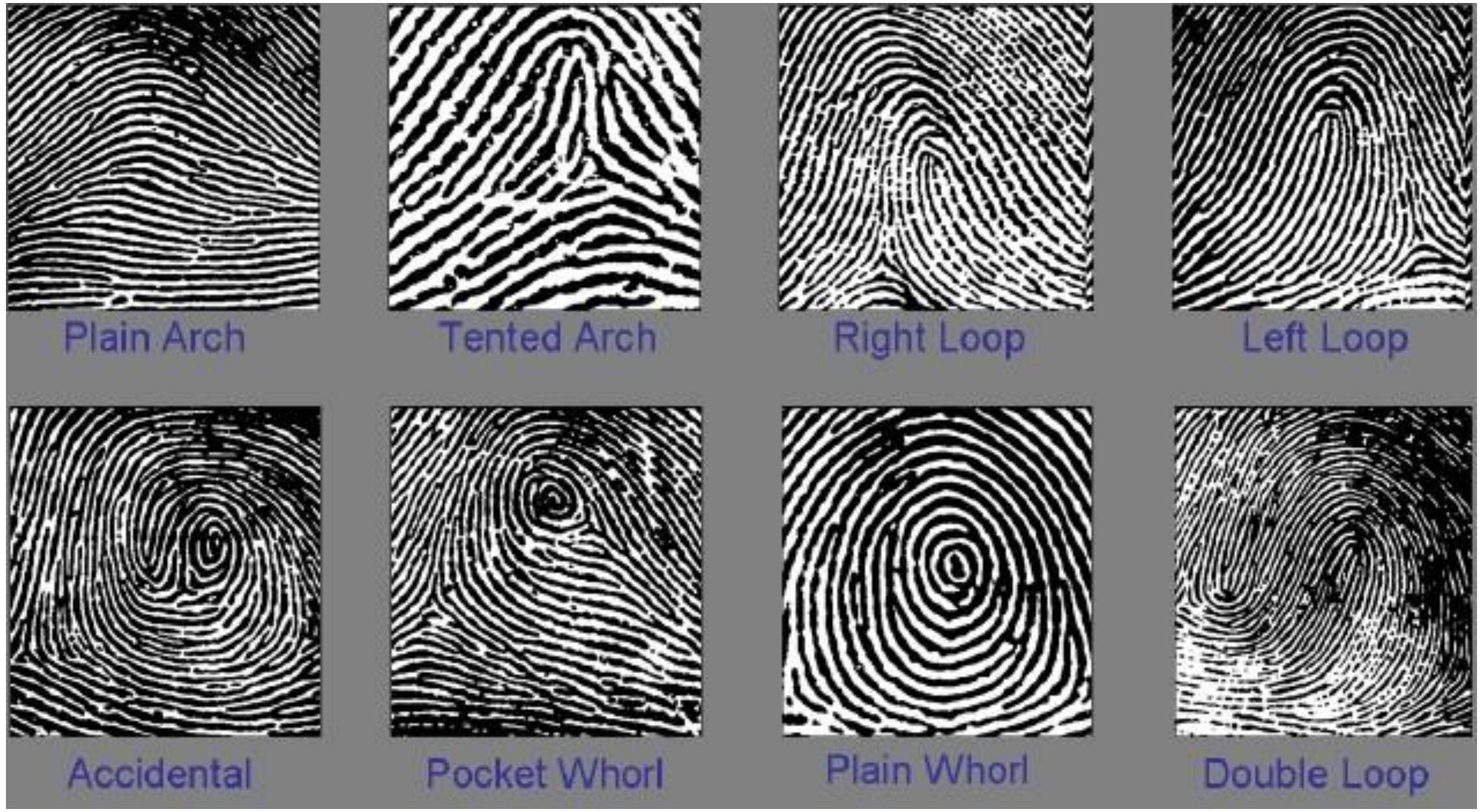
Detection

Matching

Recognition

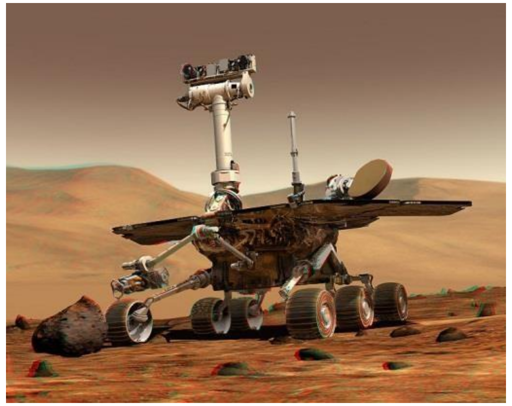
Fingerprint Classification

Important step for speeding up identification



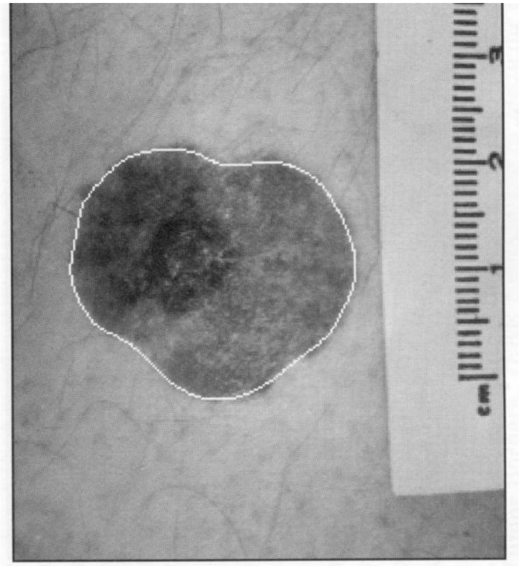
Autonomous Systems

Obstacle detection and avoidance
Object recognition



Medical Applications

Skin Cancer Detection



Breast Cancer Detection

Land Cover Classification

(using aerial or satellite images)

Many applications including “precision” agriculture.

Probability:
Introduction to probability
Probabilities of Events

What is covered?

- Basics of Probability
- Combination
- Permutation
- Examples for the above
- Union
- Intersection
- Complement

What is a probability

- **Probability** is the branch of mathematics concerning numerical descriptions of how likely an event is to occur

- The probability of an event is a number between 0 and 1, where, roughly speaking, 0 indicates that the event is not going to happen and 1 indicates event happens all the time.

$$\text{Probability of an event} = \frac{\text{Chance favouring the event}}{\text{Total possible events}}$$

Experiment

- The term experiment is used in probability theory to describe a process for which the outcome is not known with certainty.

  Example of experiments are:

    Rolling a fair six sided die.

    Randomly choosing 5 apples from a lot of 100 apples.

Event

- An event is an outcome of an experiment. It is denoted by capital letter. Say E1,E2… or A,B….and so on

- For example toss a coin, H and T are two events.

- The event consisting of all possible outcomes of a statistical experiment is called the "Sample Space". Ex: { E1,E2…}

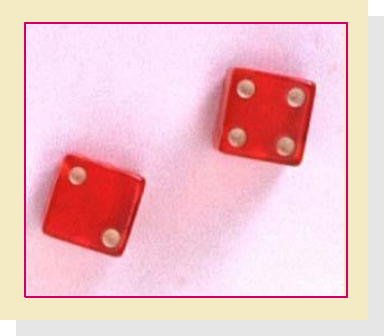
Examples

Sample Space of Tossing a coin = {H,T}

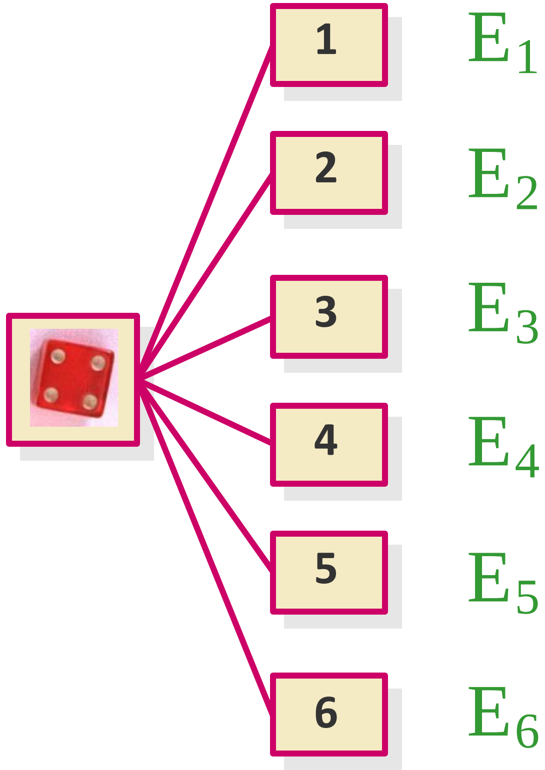
Tossing 2 Coins = {HH,HT,TH,TT}

Example

- **The die toss:**

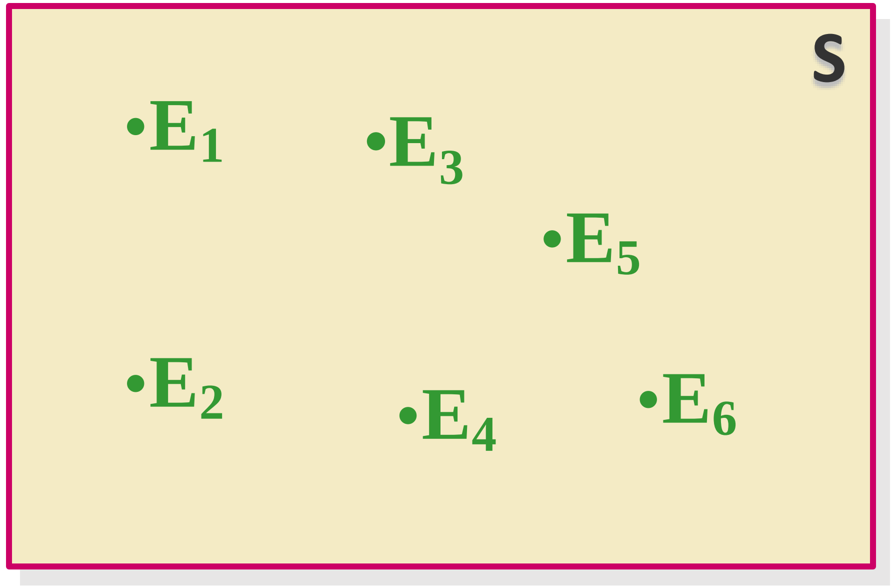


- Simple events:



Sample space:

$$S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$$



The Probability of an Event P(A)

- The probability of an event A measures "how often" A will occur. We write **P(A).**

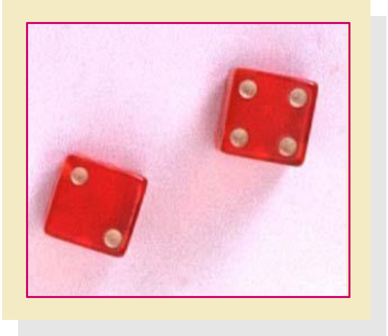
- Suppose that an experiment is performed n times. The relative frequency for an event A is

$$\frac{\text{Number of times A occurs}}{n} = \frac{f}{n}$$

- If we let n get infinitely large,

$$\boxed{P(A) = \lim_{n\to\infty} \frac{f}{n}}$$

The Probability of an Event



- P(A) must be between 0 and 1.
  - If event A can never occur, P(A) = 0. If event A always occurs when the experiment is performed, P(A) =1.

  - **Then P(A) + P(not A) = 1.**
  - So P(not A) = 1-P(A)

- The sum of the probabilities for all simple events in S equals 1.

Example 1

Toss a fair coin twice. What is the probability of observing at least one head?

1st Coin	2nd Coin	E_i	$P(E_i)$
H	H	HH	1/4
	T	HT	1/4
T	H	TH	1/4
	T	TT	1/4

P(at least 1 head)

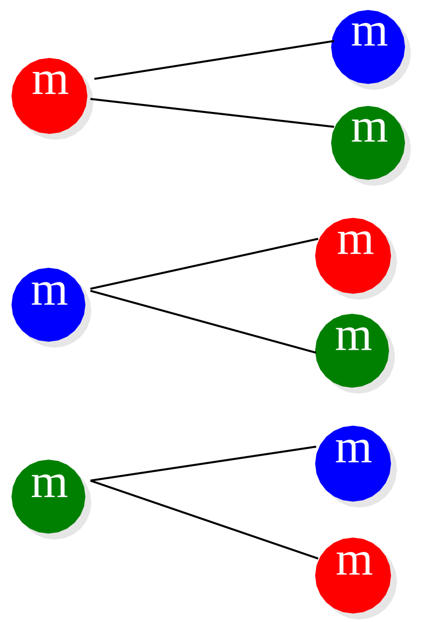
$= P(E_1) + P(E_2) + P(E_3)$

$= 1/4 + 1/4 + 1/4 = 3/4$

Example 2

A bowl contains three colour Ms®, one red, one blue and one green. A child selects two M&Ms at random. What is the probability that at least one is red?

1st M&M 2nd M&M



E_i	$P(E_i)$
RB	1/6
RG	1/6
BR	1/6
BG	1/6
GB	1/6
GR	1/6

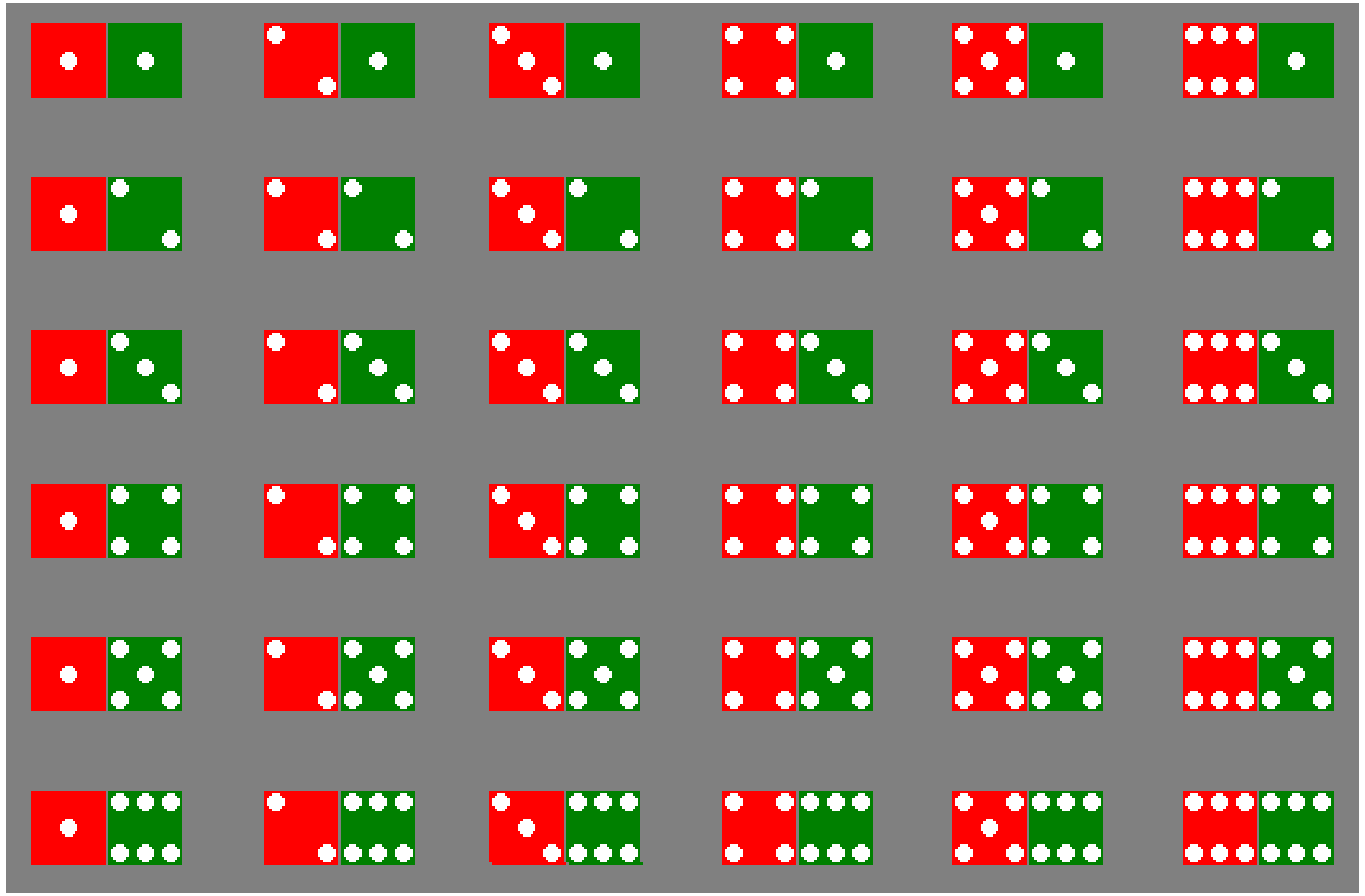
P(at least 1 red)

= P(RB) + P(BR)+ P(RG) + P(GR)

= 4/6 = 2/3

Example 3

The sample space of throwing a pair of dice is



Example 3

Event	Simple events	Probability
Dice add to 3	(1,2),(2,1)	2/36
Dice add to 6	(1,5),(2,4),(3,3), (4,2),(5,1)	5/36
Red die show 1	(1,1),(1,2),(1,3), (1,4),(1,5),(1,6)	6/36
Green die show 1	(1,1),(2,1),(3,1), (4,1),(5,1),(6,1)	6/36

Permutations

- The number of ways you can arrange ***n*** distinct objects, taking them ***r*** at a time is

$$P_r^n = \frac{n!}{(n-r)!}$$

where $n! = n(n-1)(n-2)\ldots(2)(1)$ and $0! \equiv 1$.

Example: How many 3-digit lock combinations can we make from the numbers 1, 2, 3, and 4?

The order of the choice is important!

$$P_3^4 = \frac{4!}{1!} = 4(3)(2) = 24$$

Examples

Example: A lock consists of five parts and can be assembled in any order. A quality control engineer wants to test each order for efficiency of assembly. How many orders are there?

The order of the choice is important!

$$P_5^5 = \frac{5!}{0!} = 5(4)(3)(2)(1) = 120$$

Combinations

- The number of distinct combinations of ***n*** distinct objects that can be formed, taking them ***r*** at a time is

$$C_r^n = \frac{n!}{r!(n-r)!}$$

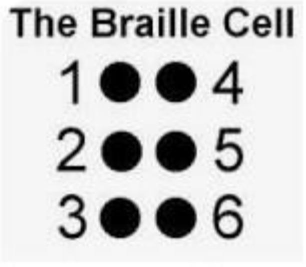
Example: Three members of a 5-person committee must be chosen to form a subcommittee. How many different subcommittees could be formed?

The order of the choice is not important!

__________ _____________ _____

Is it combination or permutation?

- Having 6 dots in a braille cell, how many different character can be made?



- It is a problem of combination
- C6,0+C 6,1 + C6,2 + C6,3+ C6,4+C6,5+ C6,6=1+6+15+20+15+6+1 = 64
- (Why combination is used not permutation? : reason each dots is of same nature)
- 64 different characters can be made.
- Where N is from 0 to 6. (It is the summation of combinations..)

$$\sum_i C(N,i) = 2^N$$

Having 4 characters, how may 2 character words can be formed:

Permutation : P6,2= 12

Combination: C6,2 = 6

Remember Permutation is larger than combination

Summary:

- So formula for Permutation is : (order is relevant)

$$P_r^n = \frac{n!}{(n-r)!}$$

- Formula for Combination is: (Order is not relevant)

$$C_r^n = \frac{n!}{r!(n-r)!}$$

Event Relations

Special Events

The Null Event, is also called as empty event represented by - ϕ

$\phi = \{ \} =$ the event that contains no outcomes

The Entire Event, The Sample Space - S

$S =$ the event that contains all outcomes

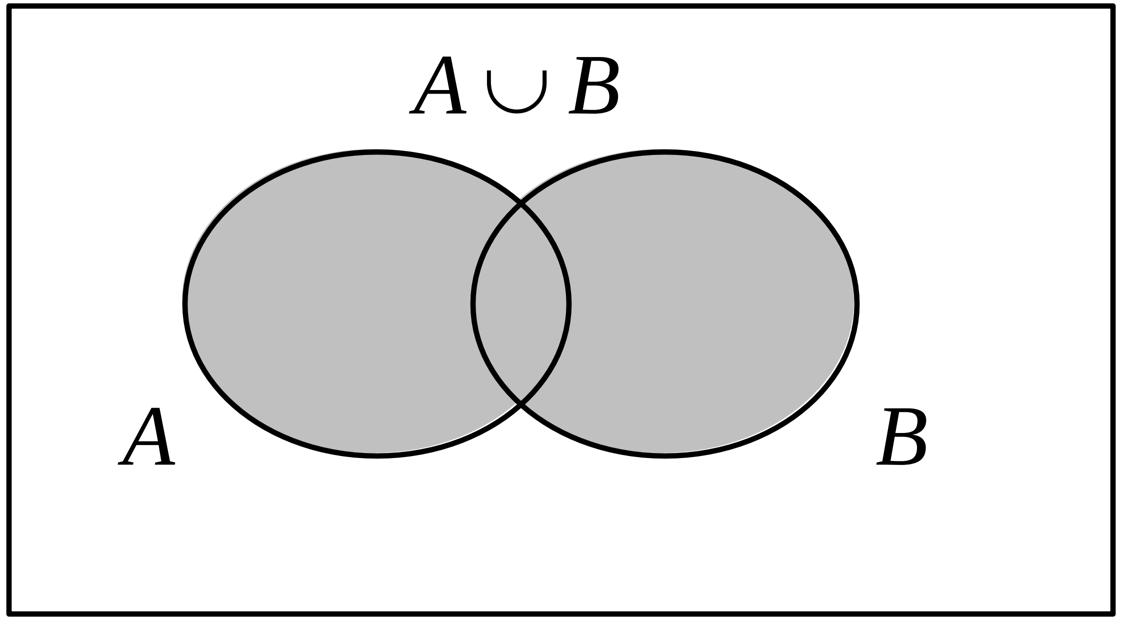
3 Basic Event relations

1. **Union** if you see the word **or**,
2. **Intersection** if you see the word **and**,
3. **Complement** if you see the word **not**.

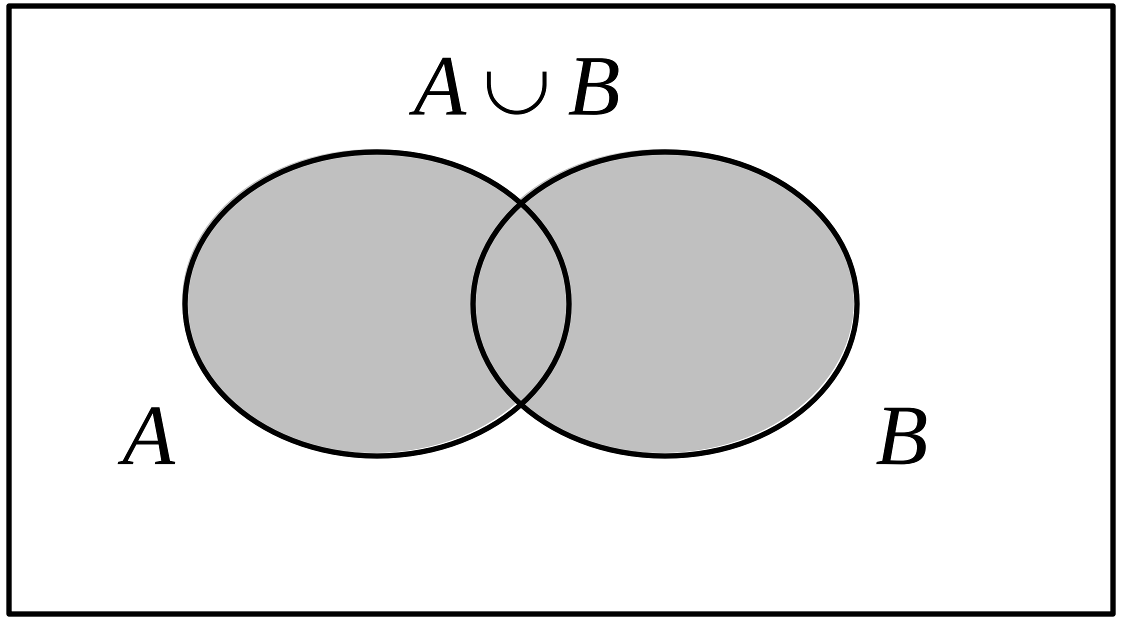
Union

Let A and B be two events, then the **union** of A and B is the event (denoted by $A \cup B$) defined by:

$$A \cup B = \{e \mid e \text{ belongs to } A \textbf{ or } e \text{ belongs to } B\}$$



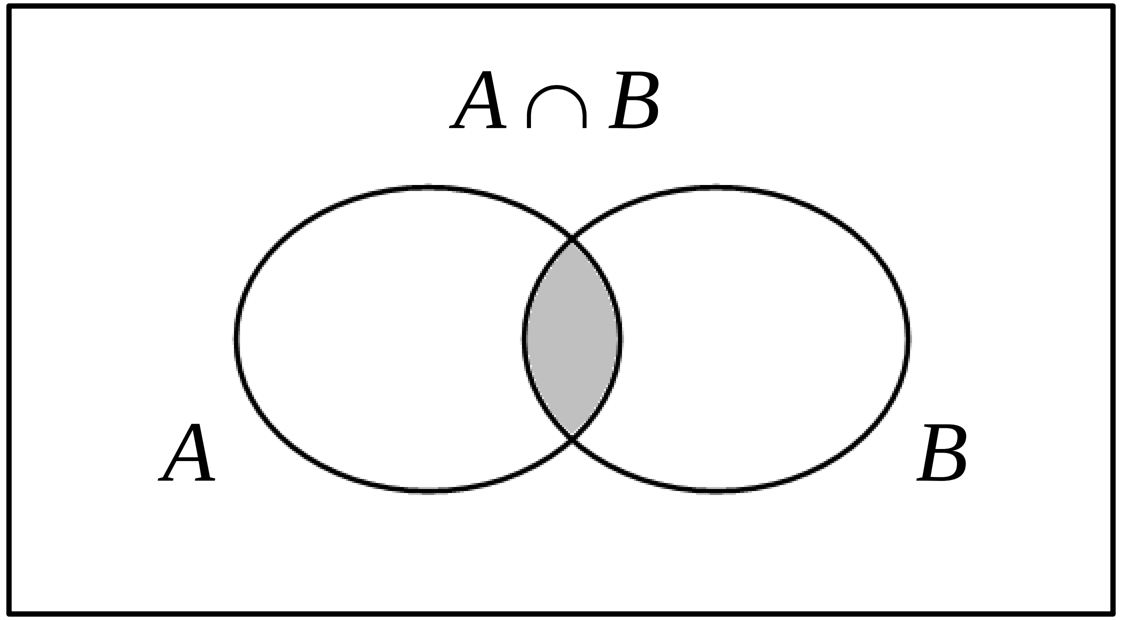
The event $A \cup B$ **occurs** if the event A **occurs or** the event and B **occurs or both occurs.**



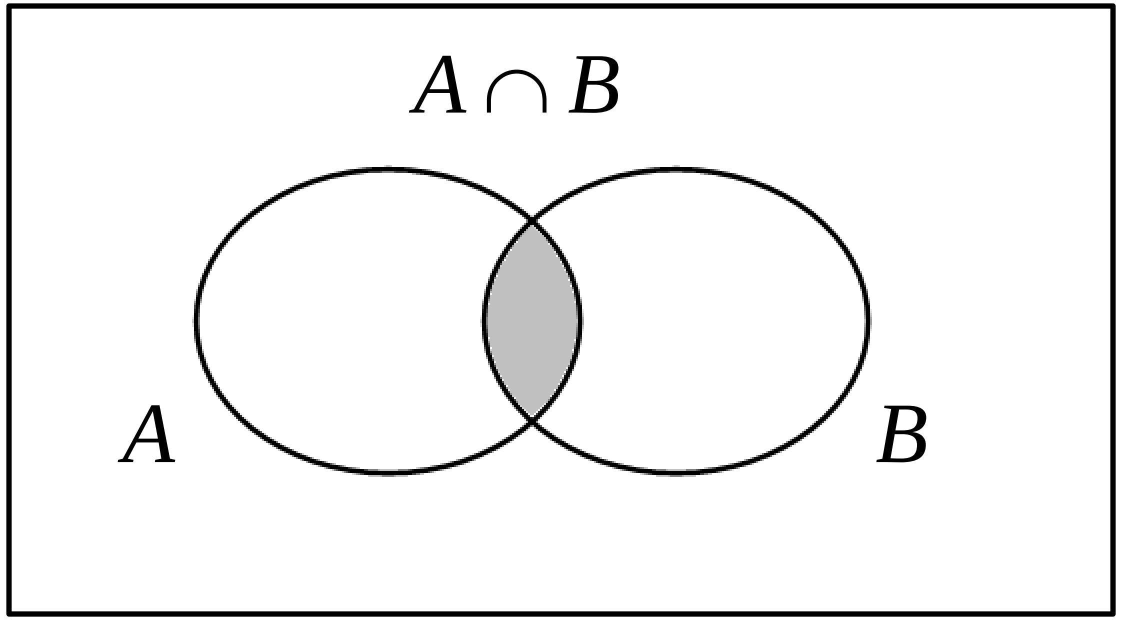
Intersection

Let A and B be two events, then the **intersection** of A and B is the event (denoted by $A \cap B$) defined by:

$$A \cap B = \{e |\ e \text{ belongs to } A \textbf{ and } e \text{ belongs to } B\}$$



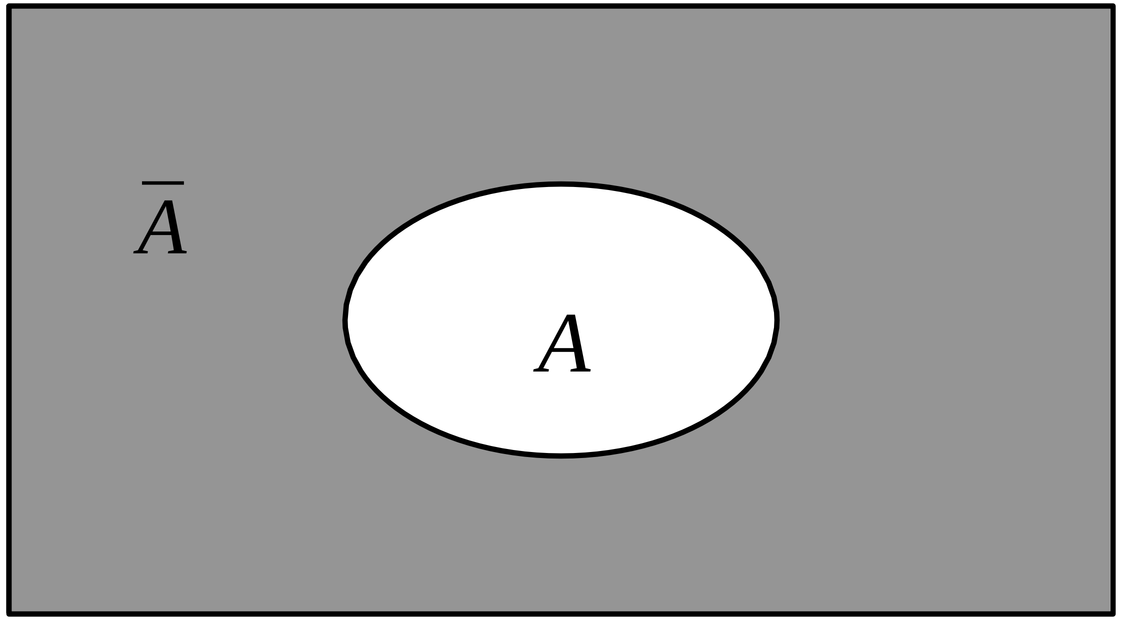
The event $A \cap B$ **occurs** if the event A **occurs and** the event and B **occurs** .



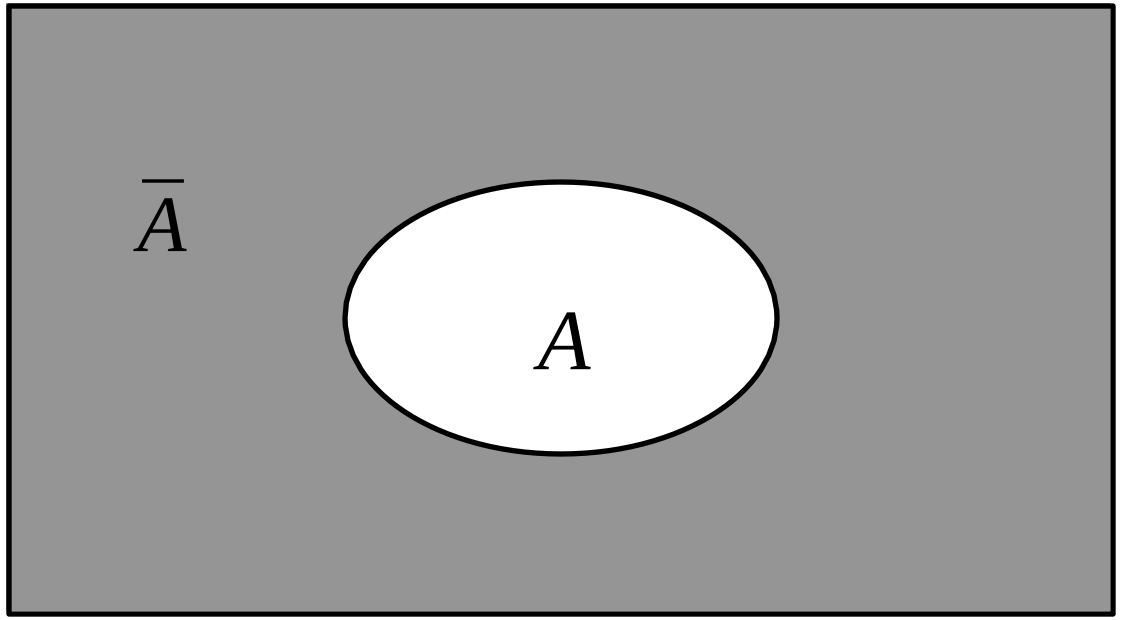
Complement

Let A be any event, then the **complement** of A (denoted by $\bar{A}$) defined by:

$$\bar{A} = \{e \mid e \textbf{ does not } \text{belongs to } A\}$$



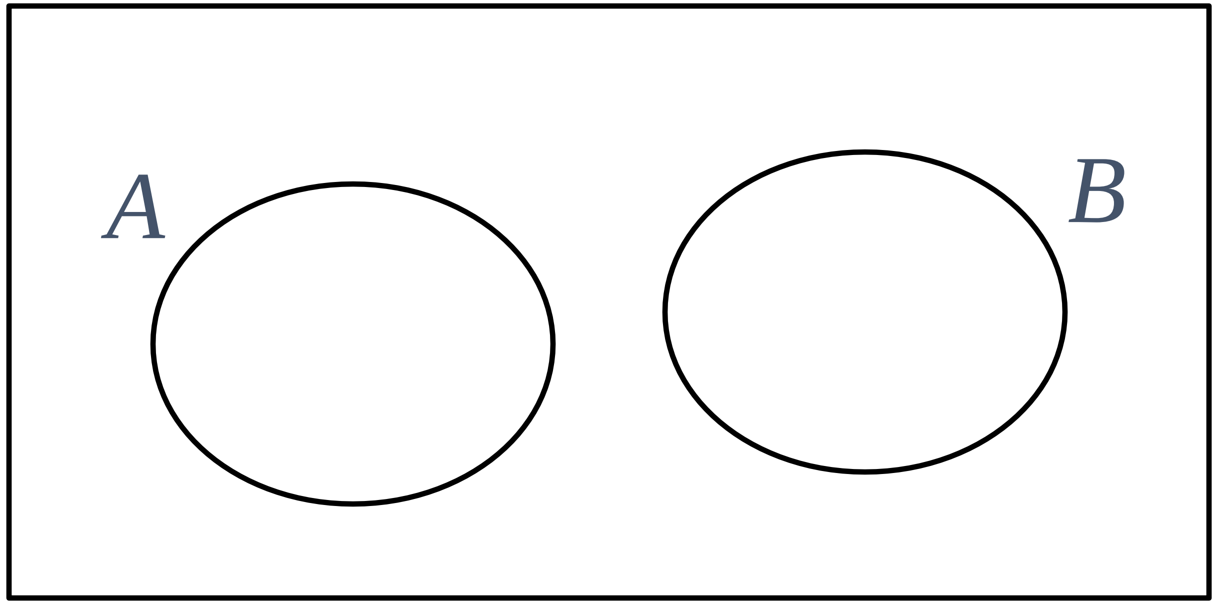
The event $\bar{A}$ **occurs** if the event A **does not occur**



Mutually Exclusive

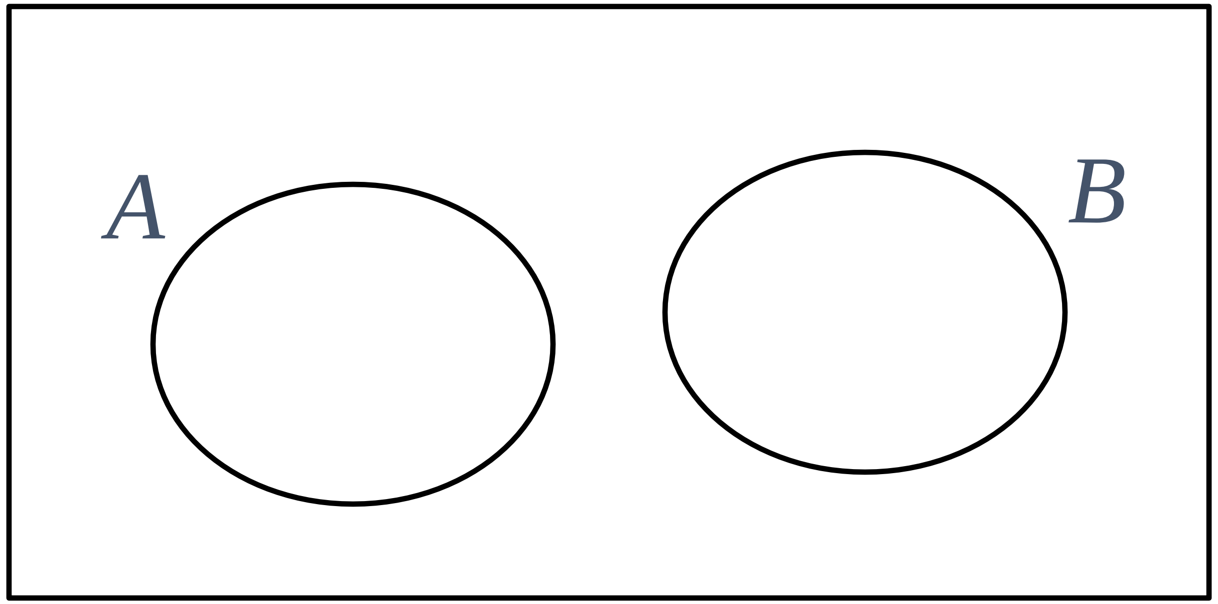
Two events A and B are called **mutually exclusive** if:

$$A \cap B = \phi$$



If two events A and B are **mutually exclusive** then:

1. They have no outcomes in common. They can't occur at the same time. The outcome of the random experiment can not belong to both A and B.



Rules of Probability

Additive Rule

Rule for complements

Probability of an Event *E*.

(revisiting … discussed in earlier slides)

Suppose that the sample space $S = \{o_1, o_2, o_3, \ldots o_N\}$ has a finite number, N, of outcomes.

Also each of the outcomes is equally likely (because of symmetry).

Then for any event E

$$P[E] = \frac{n(E)}{n(S)} = \frac{n(E)}{N} = \frac{\text{no. of outcomes in } E}{\text{total no. of outcomes}}$$

$$\textbf{Note}: \text{ the symbol } n(A) = \text{ no. of elements of } A$$

Additive rule
(In general)

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

or

$$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B]$$

The additive rule (Mutually exclusive events) if $A \cap B = \phi$

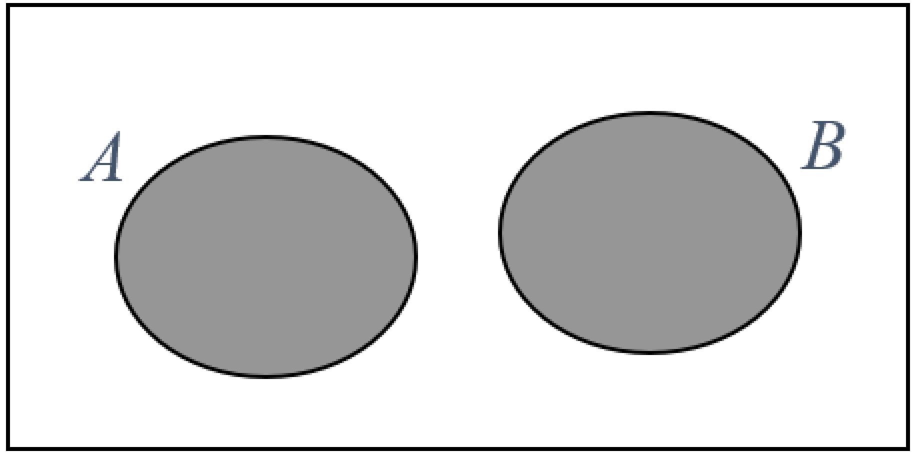
$$P[A \cup B] = P[A] + P[B]$$

i.e.

$$P[A \text{ or } B] = P[A] + P[B]$$

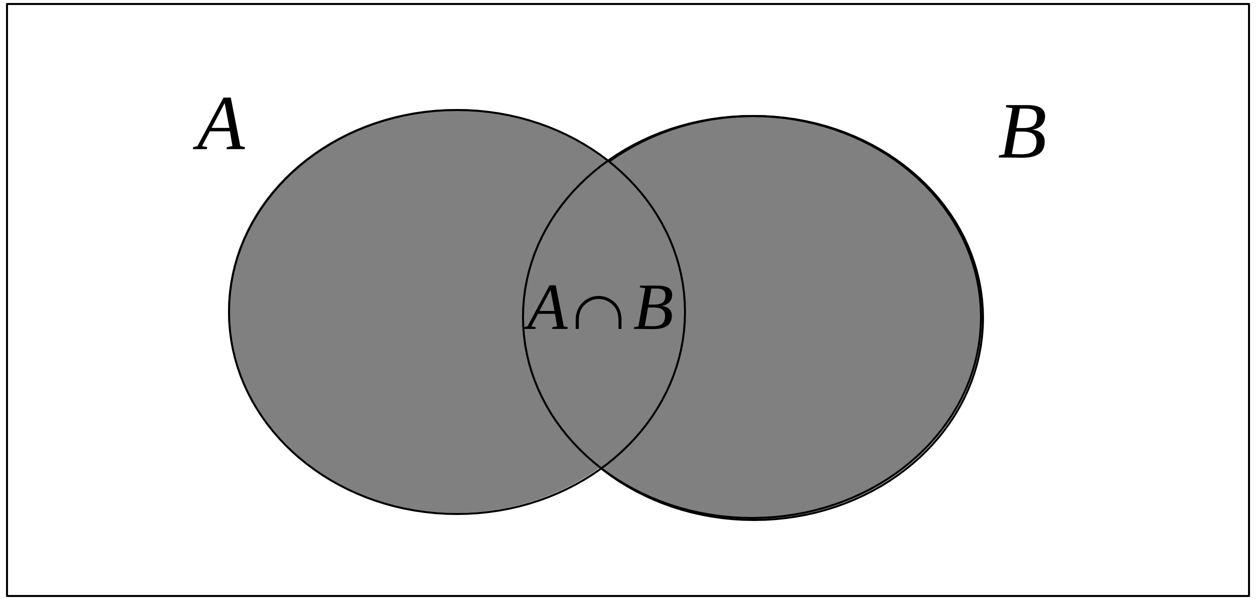
if $A \cap B = \phi$

(A and B mutually exclusive)



Logic

$A \cup B$



When P[A] is added to P[B] the outcome in A $\cap$ B are counted twice

hence

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

$$\boxed{P[A \cup B] = P[A] + P[B] - P[A \cap B]}$$

Example:

Bangalore and Mohali are two of the cities competing for the National university games. (There are also many others).

The organizers are narrowing the competition to the final 5 cities.

There is a 20% chance that Bangalore will be amongst the final 5.

There is a 35% chance that Mohali will be amongst the final 5 and

an 8% chance that both Bangalore and Mohali will be amongst the final 5.

What is the probability that Bangalore or Mohali will be amongst the final 5.

Solution:

Let A = the event that Bangalore is amongst the **final 5**.

Let B = the event that Mohali is amongst the **final 5**.

Given $P[A] = 0.20$, $P[B] = 0.35$, and $P[A \cap B] = 0.08$

What is $P[A \cup B]$?

Note: "and" $\equiv \cap$, "or" $\equiv \cup$.

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

$$= 0.20 + 0.35 - 0.08 = 0.47$$

Find the probability of drawing an ace or a spade from a deck of cards.

There are 52 cards in a deck; 13 are spades, 4 are aces.

Probability of a single card being spade is: 13/52 = 1/4.
Probability of drawing an Ace is : 4/52 = 1/13.
Probability of a single card being both Spade and Ace = 1/52.

Let A = Event of drawing a spade .
Let B = Event drawing Ace.

Given $P[A]$ =1/4, $P[B]$ =1/13, and $P[A \cap B]$ = 1/52

$$P\left[A \cup B\right] = P\left[A\right] + P\left[B\right] - P\left[A \cap B\right]$$

$$P[A \cup B] = 1/4 + 1/13 - 1/52$$

Rule for complements

Rule for complements

The Complement Rule states that the **sum of the probabilities of an event and its complement must equal 1**, or for the event A, P(A) + P(A') = 1.

$$P[A'] = 1 - P[A]$$

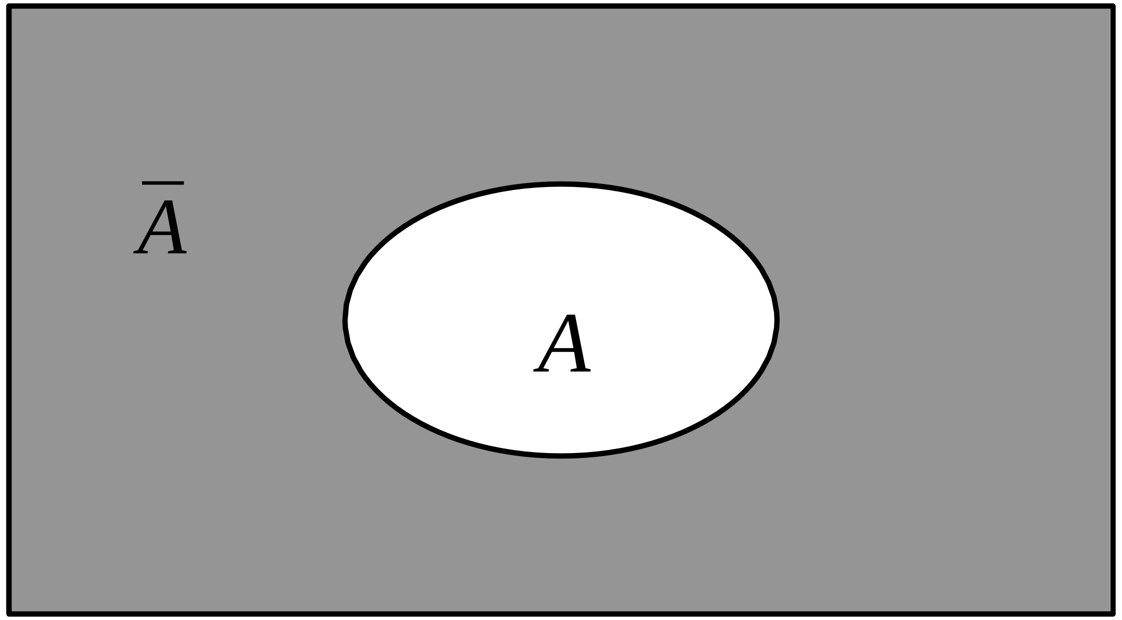
or

$$P[\text{not } A] = 1 - P[A]$$

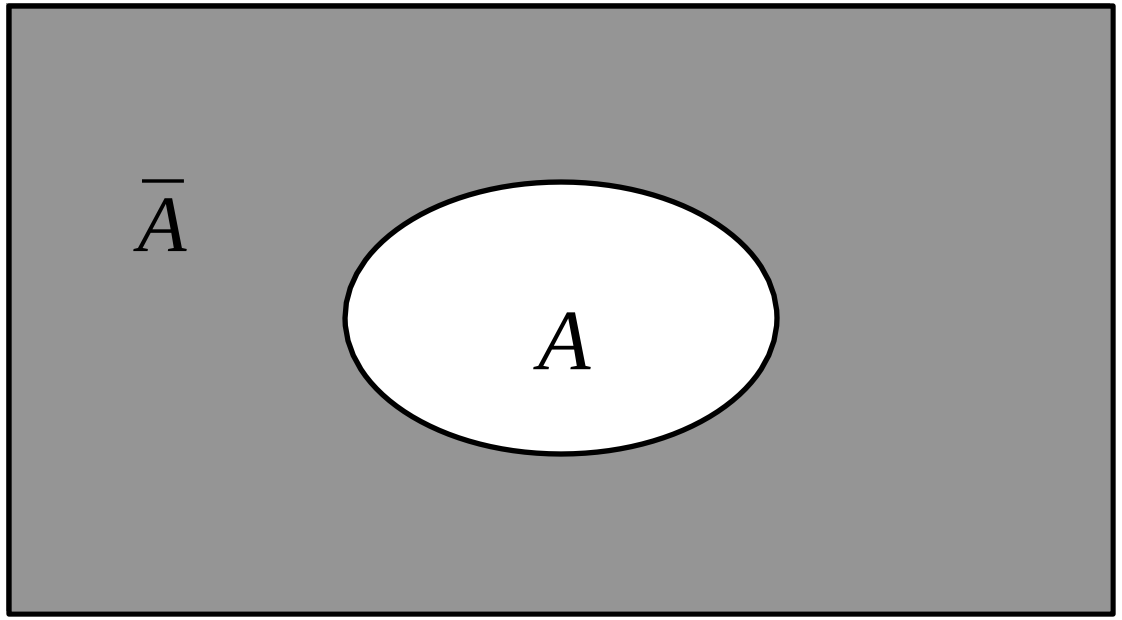
Complement

Let A be any event, then the **complement** of A (denoted by $\bar{A}$) defined by:

$\bar{A}$ = {e| e **does not** belongs to A}



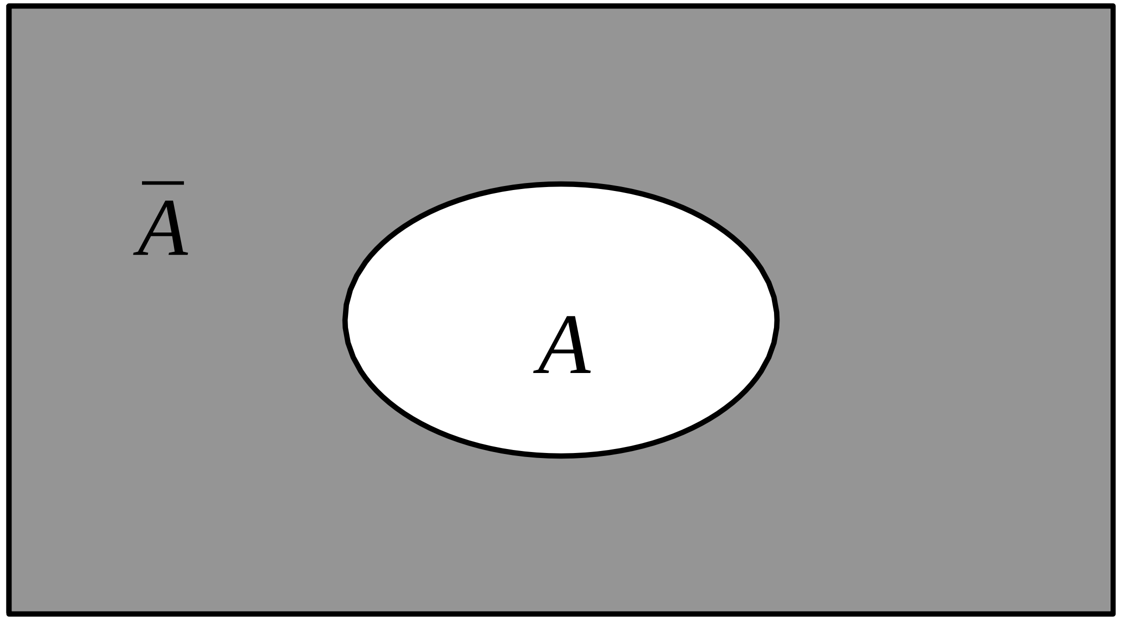
The event $\bar{A}$ **occurs** if the event A **does not occur**



Logic:

$\bar{A}$ and A are **mutually exclusive**.

and $S = A \cup \bar{A}$



thus $1 = P\left[S\right] = P\left[A\right] + P\left[\bar{A}\right]$

and $P\left[\bar{A}\right] = 1 - P\left[A\right]$

What Is Conditional Probability?

- 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.
- Bayes' theorem is a mathematical formula used in calculating conditional probability.

Definition

Suppose that we are interested in computing the probability of event *A* and we have been told event *B* has occurred.

Then the conditional probability of *A* given *B* is defined to be:

$$P\left[A|B\right] = \frac{P\left[A \cap B\right]}{P\left[B\right]} \qquad \text{if} \quad P\left[B\right] \neq 0$$

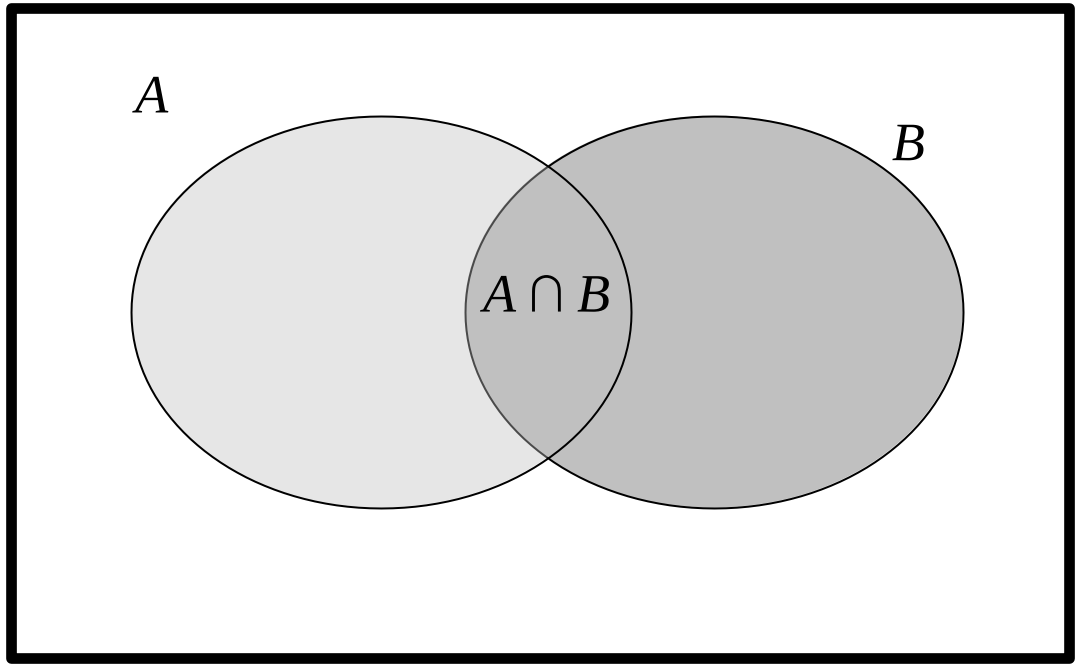
Illustrates that probability of A, given(|) probability of B occurring

Rationale:

If we're told that event *B* has occurred then the sample space is restricted to *B.*

The event *A* can now only occur if the outcome is in of $A \cap B$. Hence the new probability of *A in B*is:

$$P\left[A \middle| B\right] = \frac{P\left[A \cap B\right]}{P\left[B\right]}$$



An Example

Twenty – 20 World cup started:

For a specific married couple the probability that the husband watches the match is 80%,

the probability that his wife watches the match is 65%,

while the probability that they both watch the match is 60%.

If the husband is watching the match, what is the probability that his wife is also watching the match

Solution:

Let B = the event that the husband watches the match

$P[B]$= 0.80

Let A = the event that his wife watches the match

$P[A]$= 0.65 and

$P[A \cap B]$= 0.60

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{0.60}{0.80} = 0.75$$

Another example

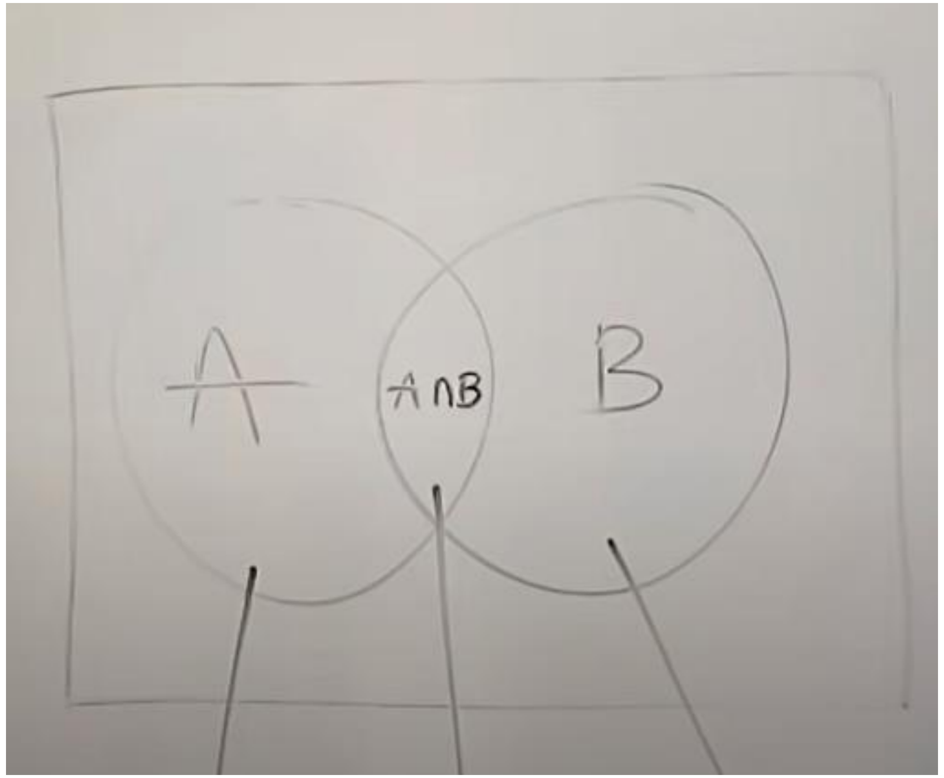
- There are 100 Students in a class.
- 40 Students likes Apple
  - Consider this event as A, So probability of occurrence of A is 40/100 = 0.4
- 30 Students likes Orange.
  - Consider this event as B, So probability of occurrence of B is 30/100=0.3

- 20 Students likes Both Apple and Orange, So probability of Both A and B occurring is = A intersect B = 20/100 = 0.2

- Remaining Students does not like either Apple nor Orange

- What is the probability of A in B, means what is the probability that A is occurring given B :

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



40 20 30

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

P(A|B) = 0.2/0.3 = 0.67

P(A|B) indicates that A occurring in the sample space of B.

Here we are not considering the entire sample space of 100 students, but only 30 students.

More Example Problem for Conditional Probability

Example : Calculating the conditional probability of rain given that the biometric pressure is high.

Weather record shows that high barometric pressure (defined as being over 760 mm of mercury) occurred on 160 of the 200 days in a data set, and it rained on 20 of the 160 days with high barometric pressure. If we let R denote the event "rain occurred" and H the event " High barometric pressure occurred" and use the frequentist approach to define probabilities.

$$P(H) = 160/200 = 0.8$$

and $P(R \text{ and } H) = 20/200 = 0.10$ (rain and high barometric pressure intersection)

We can obtain the probability of rain given high pressure, directly from the data.

$$P(R|H) = 20/160 = 0.10/0.80 = 0.125$$

Representing in conditional probability

$$P(R|H) = P(R \text{ and } H)/P(H) = 0.10/0.8 = 0.125.$$

In my town, it's rainy one third (1/3) of the days.

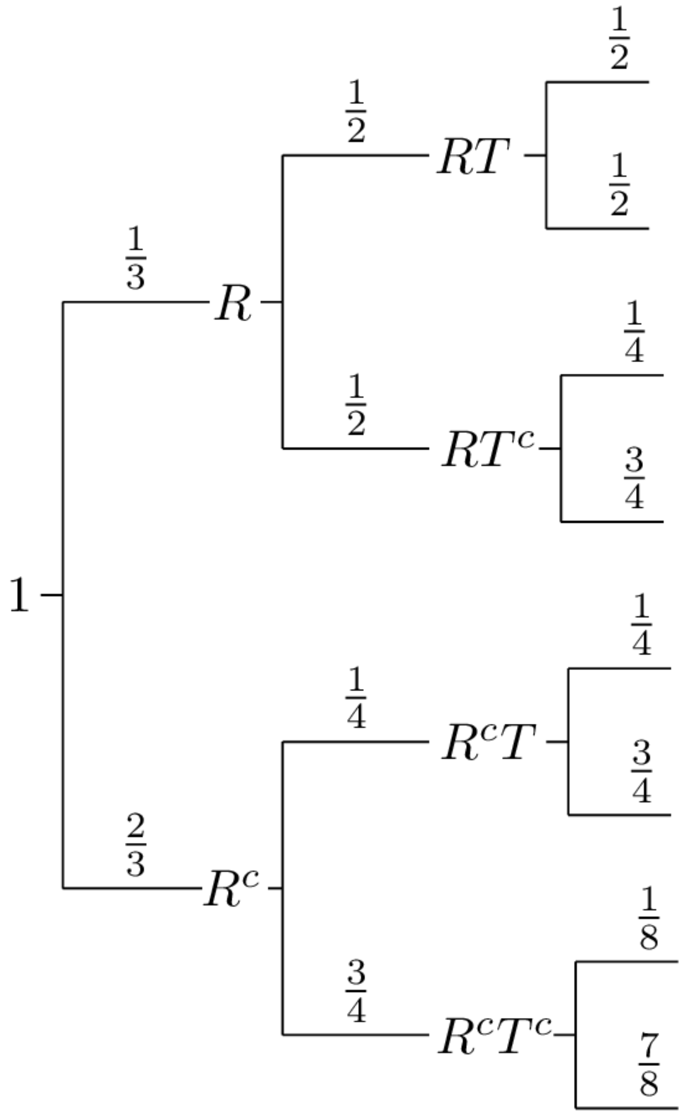
Given that it is rainy, there will be heavy traffic with probability 1/2, and given that it is not rainy, there will be heavy traffic with probability 1/4.

If it's rainy and there is heavy traffic, I arrive late for work with probability 1/2.

On the other hand, the probability of being late is reduced to 1/8 if it is not rainy and there is no heavy traffic.

In other situations (rainy and no traffic, not rainy and traffic) the probability of being late is 0.25. You pick a random day.

- What is the probability that it's not raining and there is heavy traffic and I am not late?
- What is the probability that I am late?
- Given that I arrived late at work, what is the probability that it rained that day?



$$RTL \to P(RTL) = \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{12}$$

$$RTL^c \to P(RTL^c) = \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{12}$$

$$RT^cL \to P(RT^cL) = \frac{1}{3} \times \frac{1}{2} \times \frac{1}{4} = \frac{1}{24}$$

$$RT^cL^c \to P(RT^cL^c) = \frac{1}{3} \times \frac{1}{2} \times \frac{3}{4} = \frac{3}{24} = \frac{1}{8}$$

$$R^cTL \to P(R^cTL) = \frac{2}{3} \times \frac{1}{4} \times \frac{1}{4} = \frac{1}{24}$$

$$R^cTL^c \to P(R^cTL^c) = \frac{2}{3} \times \frac{1}{4} \times \frac{3}{4} = \frac{1}{8}$$

$$R^cT^cL \to P(R^cT^cL) = \frac{2}{3} \times \frac{3}{4} \times \frac{1}{8} = \frac{1}{16}$$

$$R^cT^cL^c \to P(R^cT^cL^c) = \frac{2}{3} \times \frac{3}{4} \times \frac{7}{8} = \frac{7}{16}$$

Let **R** be the event that it's rainy, **T** be the event that there is heavy traffic, and **L** be the event that I am late for work. As it is seen from the problem statement, we are given conditional probabilities in a chain format. Thus, it is useful to draw a tree diagram for this problem. In this figure, each leaf in the tree corresponds to a single outcome in the sample space. We can calculate the probabilities of each outcome in the sample space by multiplying the probabilities on the edges of the tree that lead to the corresponding outcome.

a. The probability that it's not raining and there is heavy traffic and I am not late can be found using the tree diagram which is in fact applying the chain rule:

$$
\begin{aligned}
P(R^c \cap T \cap L^c) &= P(R^c)P(T|R^c)P(L^c|R^c \cap T) \\
&= 2/3 \cdot 1/4 \cdot 3/4 \\
&= 1/8.
\end{aligned}
$$

b. The probability that I am late can be found from the tree. All we need to do is sum the probabilities of the outcomes that correspond to me being late. In fact, we are using the law of total probability here.

$P(L) = P(R \text{ and } T \text{ and } L) + P(R \text{ and } T^c \text{ and } L) + P(R^c \text{ and } T \text{ and } L) + P(R^c \text{ and } T^c \text{ and } L)$

$= 1/12 + 1/24 + 1/24 + 1/16$

$= 11/48.$

c. We can find $P(R|L)$ using

$P(R|L) = P(R \cap L) P(L) P(R|L) = P(R \cap L) P(L).$

We have already found $P(L) = 11/48$ and we can find $P(R \cap L)$ similarly by adding the probabilities of the outcomes that belong to $R \cap L$.

Random Variables

Random variable takes a random value, which is real and can be finite or infinite and it is generated out of random experiment.

The random value is generated out of a function.

Example: Let us consider an experiment of tossing two coins.

Then sample space is S= { HH, HT, TH, TT}

Given X as random variable with condition: number of heads.

X(HH) =2

X(HT) =1

X(TH) =1

X(TT) = 0

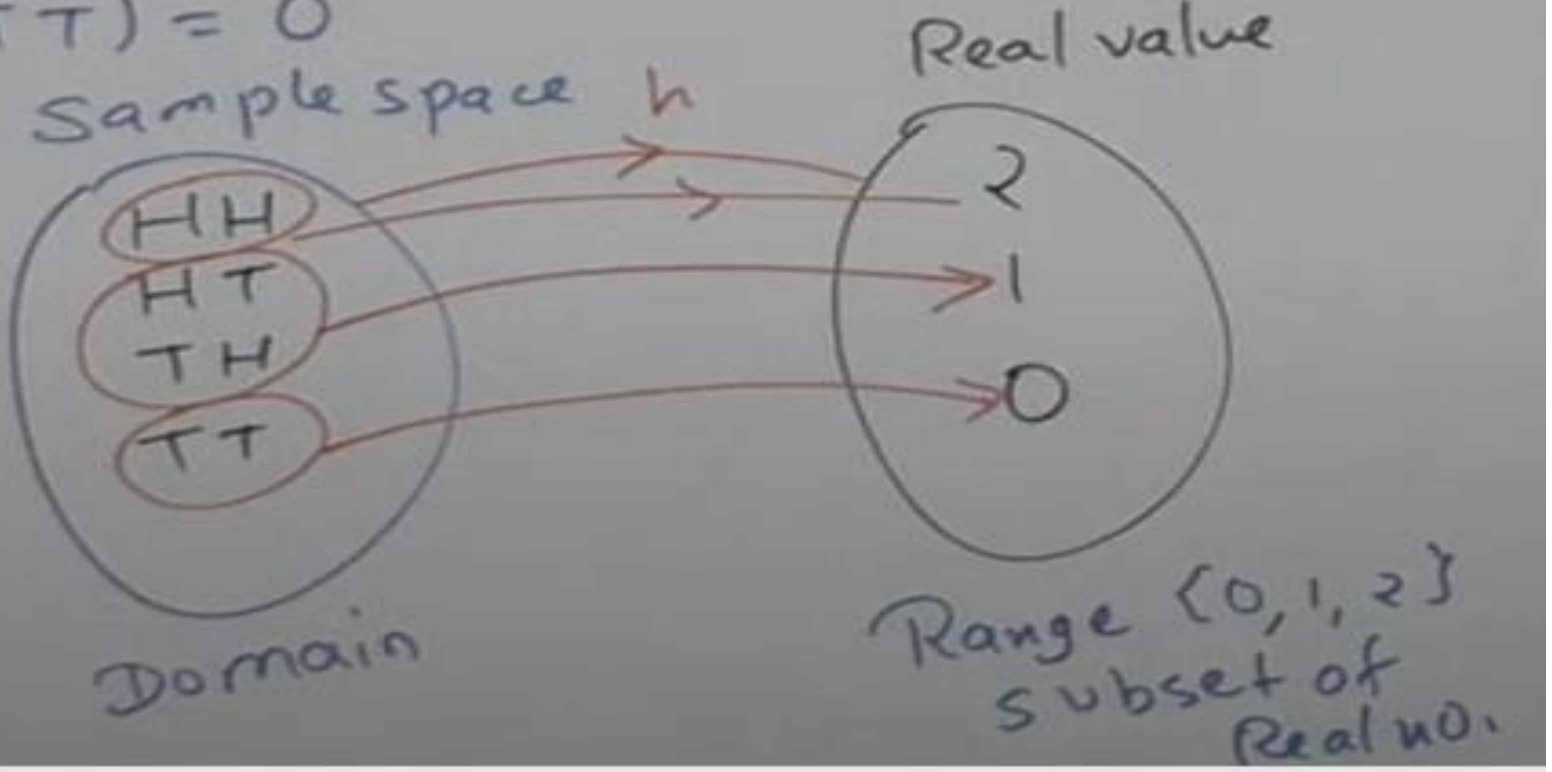
$X =$ no. of heads

$X(HH) = 2$

$X(HT) = 1$

$X(TH) = 1$

$X(TT) = 0$



- Two types of random variables
  - **Discrete random variables**
  - **Continuous random variable**

Discrete random variables

- If the variable value is finite or infinite but countable, then it is called discrete random variable.

- Example of tossing two coins and to get the count of number of heads is an example for discrete random variable.

- Sample space of real values is fixed.

Continuous Random Variable

- If the random variable values lies between two certain fixed numbers then it is called continuous random variable. The result can be finite or infinite.

- Sample space of real values is not fixed, but it is in a range.

- If X is the random value and it’s values lies between a and b then,

It is represented by : $a <= X <= b$

Example: Temperature, age, weight, height…etc. ranges between specific range.

Here the values for the sample space will be infinite

Probability distribution

- Frequency distribution is a listing of the observed frequencies of all the output of an experiment that actually occurred when experiment was done.
- Where as a probability distribution is a listing of the probabilities of all possible outcomes that could result if the experiment were done. (distribution with expectations).

Broad classification of Probability distribution

- Discrete probability distribution
  - Binomial distribution
  - Poisson distribution

- Continuous Probability distribution
  - Normal distribution

Discrete Probability Distribution: Binomial Distribution

- A binomial distribution can be thought of as simply **the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times**. (When we have only two possible outcomes)

- Example, a coin toss has only two possible outcomes: heads or tails and taking a test could have two possible outcomes: pass or fail.

Assumptions of Binomial distribution (It is also called as Bernoulli's Distribution)

- Assumptions:
  - Random experiment is performed repeatedly with a fixed and finite number of trials. The number is denoted by 'n'
  - There are two mutually exclusive possible outcome on each trial, which are know as "Success" and "Failure". Success is denoted by 'p' and failure is denoted by 'q'. and p+q=1 or q=1-p.
  - The outcome of any give trail does not affect the outcomes of the subsequent trail. That means all trials are independent.
  - The probability of success and failure (p&q) remains constant for all trials. If it does not remain constant then it is not binomial distribution. For example tossing a coin the probability of getting head or getting a red ball from a pool of colored balls, here every time after the ball is taken out it is again replaced to the pool.
  - With this assumption let see the formula

Formula for Binomial Distribution

$$P_x = \binom{n}{x} p^x q^{n-x}$$

OR

$$P(X=r) = {}^{n}C_{r}\, p^{r} q^{n-r}$$

Where P is success and

q is failure

n = number of trials

$$\binom{n}{X} p^{X} (1-p)^{n-X}$$

X = # successes out of n trials

p = probability of success

$1\text{-}p$ = probability of failure

Binomial Distribution: Illustration with example

- Consider a pen manufacturing company
- 10% of the pens are defective

- (i)Find the probability that exactly 2 pens are defective in a box of 12
- So n=12,
- p=10% = 10/100 = 1/10
- q= (1-q) =90/100 = 9/10
- X=2

$$P(X=r) = {}^{n}C_{r}\, p^{r}\, q^{n-r}$$

$$P[X=2] = {}^{n}C_{x}\, p^{x}\, q^{n-x}$$

$$= {}^{12}C_{2} \left(1/10\right)^{2} \left(\frac{9}{10}\right)^{10}$$

- Consider a pen manufacturing company
- 10% of the pens are defective

- (i)Find the probability that at least 2 pens are defective in a box of 12
- So n=12,
- p=10% = 10/100 = 1/10
- q= (1-q) =90/100 = 9/10
- X>=2
- P(X>=2) = 1- [P(X<2)]
- = 1-[P(X=0) +P(X=1)]

Binomial distribution: Another example

- If I toss a coin 20 times, what's the probability of getting exactly 10 heads?

$$\binom{20}{10}(.5)^{10}(.5)^{10} = .176$$

The Binomial Distribution: another example

- Say 40% of the class is female.
- What is the probability that 6 of the first 10 students walking in will be female?

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

$$= \binom{10}{6} (.4^6)(.6^{10-6})$$

$$= 210(.004096)(.1296)$$

$$= .1115$$

Real-Life Examples of the Binomial Distribution

Example 1: Number of Side Effects from Medications

Medical professionals use the binomial distribution to model the probability that a certain number of patients will experience side effects as a result of taking new medications.

For example, suppose it is known that 5% of adults who take a certain medication experience negative side effects. We can use a Binomial Distribution Calculator to find the probability that more than a certain number of patients in a random sample of 100 will experience negative side effects.

- P(X > 5 patients experience side effects) = **0.38400**
- P(X > 10 patients experience side effects) = **0.01147**
- P(X > 15 patients experience side effects) = **0.0004**

And so on.

Example 2: Number of Fraudulent Transactions

Banks use the binomial distribution to model the probability that a certain number of credit card transactions are fraudulent.

For example, suppose it is known that 2% of all credit card transactions in a certain region are fraudulent. If there are 50 transactions per day in a certain region, we can use a Binomial Distribution Calculator to find the probability that more than a certain number of fraudulent transactions occur in a given day:

- P(X > 1 fraudulent transaction) = **0.26423**
- P(X > 2 fraudulent transactions) = **0.07843**
- P(X > 3 fraudulent transactions) = **0.01776**

And so on.

This gives banks an idea of how likely it is that more than a certain number of fraudulent transactions will occur in a given day.

Example 3: Number of Spam Emails per Day

Email companies use the binomial distribution to model the probability that a certain number of spam emails land in an inbox per day.

For example, suppose it is known that 4% of all emails are spam. If an account receives 20 emails in a given day, we can use a Binomial Distribution Calculator to find the probability that a certain number of those emails are spam:

- P(X = 0 spam emails) = **0.44200**
- P(X = 1 spam email) = **0.36834**
- P(X = 2 spam emails) = **0.14580**

Example 4: Shopping Returns per Week

Retail stores use the binomial distribution to model the probability that they receive a certain number of shopping returns each week.

For example, suppose it is known that 10% of all orders get returned at a certain store each week.

If there are 50 orders that week, we can use a [Binomial Distribution Calculator]() to find the probability that the store receives more than a certain number of returns that week:

- P(X > 5 returns) = **0.18492**
- P(X > 10 returns) = **0.00935**
- P(X > 15 returns) = **0.00002**

And so on.

This gives the store an idea of how many customer service reps they need to have in the store that week to handle returns.

Poisson Distribution

The **Poisson distribution** is the discrete probability distribution of the number of events occurring in a given time period, given the average number of times the event occurs over that time period.
Formula

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

where

λ is the expected value(average of x),
k is number of events observed over a time period and
e is Euler's number.

A Poisson distribution is a discrete probability distribution, meaning that it gives the probability of a discrete (i.e., countable) outcome. For Poisson distributions, the discrete outcome is the number of times an event occurs, represented by k.

You can use a Poisson distribution to predict or explain the number of events occurring within a given interval of time or space.

You can use a Poisson distribution if:

1. Individual events happen at random and independently. That is, the probability of one event doesn’t affect the probability of another event.
2. You know the mean number of events occurring within a given interval of time or space. This number is called λ (lambda), and it is assumed to be constant.

When events follow a Poisson distribution, λ is the only thing you need to know to calculate the probability of an event occurring a certain number of times.

Examples of Poisson distributions

Horse kick deaths

One of the first applications of the Poisson distribution was by statistician Ladislaus Bortkiewicz. In the late 1800s, he investigated accidental deaths by horse kick of soldiers in the Prussian army. He analyzed 20 years of data for 10 army corps, equivalent to 200 years of observations of one corps.

He found that a mean of 0.61 soldiers per corps died from horse kicks each year. However, most years, no soldiers died from horse kicks. On the other end of the spectrum, one tragic year there were four soldiers in the same corps who died from horse kicks.

Using modern terminology:

•A death by horse kick is an "event."

•The time interval is one year.

•The mean number of events per time interval, λ, is 0.61.

•The number of deaths by horse kick in a specific year is k.

Other examples of Poisson distributions

For example, a Poisson distribution could be used to explain or predict:

•Text messages per hour

•Machine malfunctions per year

•Website visitors per month

•Influenza cases per year

Continuous Probability Distributions

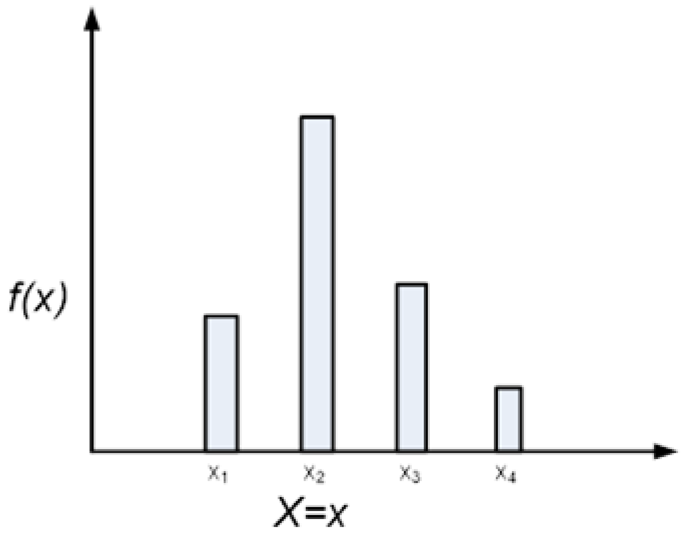
- When the random variable of interest can take any value in an interval, it is called continuous random variable.
  - Every continuous random variable has an infinite, uncountable number of possible values (i.e., any value in an interval).

- **Examples**
  - Temperature on a given day, Length, height, intensity of light falling on a given region.
  - The length of time it takes a truck driver to go from New York City to Miami.
  - The depth of drilling to find oil.
  - The weight of a truck in a truck-weighing station.
  - The amount of water in a 12-ounce bottle.

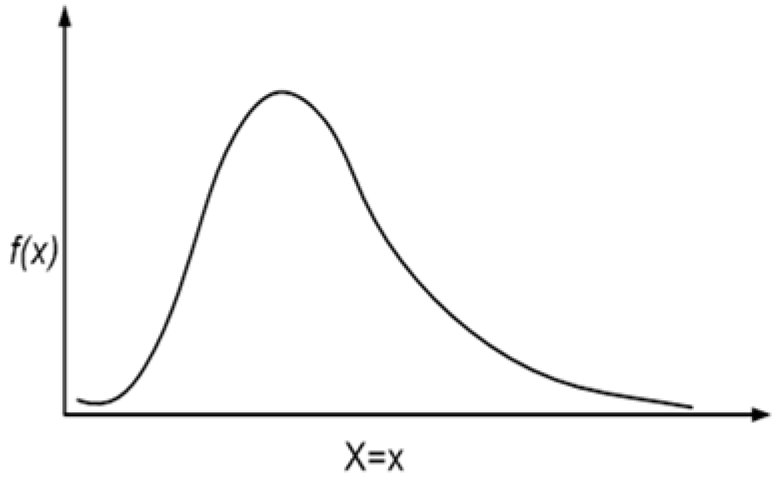
For each of these, if the variable is X, then x>0 and less than some maximum value possible, but it can take on any value within this range

- Continuous random variable differs from discrete random variable. Discrete random variables can take on only a finite number of values or at most a countable infinity of values.

- A continuous random variable is described by Probability density function. This function is used to obtain the probability that the value of a continuous random variable is in the given interval.



Discrete Probability distribution



Continuous Probability Distribution

NOTE : Density Functions are sometimes also called distributions

1. Distributions of discrete random variables must sum up to 1. $\sum_{i=1}^{n} P(xi) = 1$

2. Densities of continuous random variables must integrate to 1, because the probability is 1 that x lies between -∞ to + ∞

1. $f(x) \geq 0$, for all $x \in R$

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

3. $P(a \leq X \leq b) = \int_{a}^{b} f(x)\, dx$

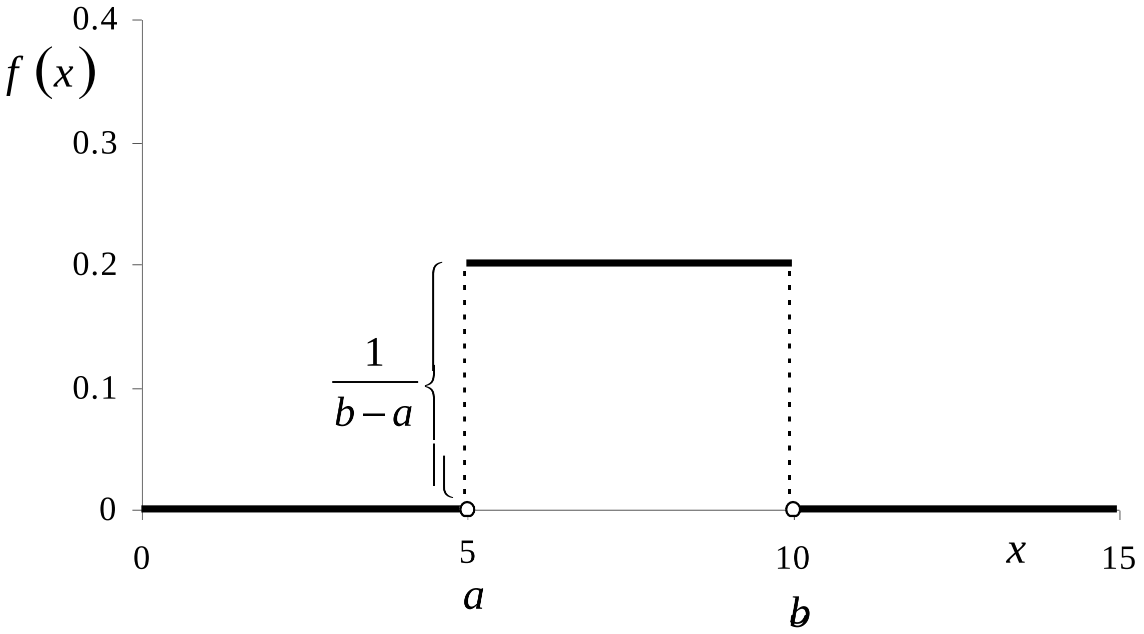
Continuous Uniform Distribution

- For Uniform distribution, f(x) is constant over the possible value of x.
- Area looks like a rectangle.
- For the area in continuous distribution we need to do integration of the function.
- However in this case it is the area of rectangle.
- Example to time taken to wash the cloths in a washing machine. (for a standard condition)

Continuous Distributions

The Uniform distribution from a to b

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



NORMAL DISTRIBUTION

- The most often used continuous probability distribution is the normal distribution; it is also known as Gaussian distribution.
- Its graph called the normal curve is the bell-shaped curve.

- Such a curve approximately describes many phenomenon occur in nature, industry and research.
  - Physical measurement in areas such as meteorological experiments, rainfall studies and measurement of manufacturing parts are often more than adequately explained with normal distribution.

NORMAL DISTRIBUTION Applications:

The normal (or Gaussian) distribution, is a very commonly used (occurring) function in the fields of probability theory, and has wide applications in the fields of:

- Pattern Recognition;
- Machine Learning;
- Artificial Neural Networks and Soft computing;
- Digital Signal (image, sound , video etc.) processing
- Vibrations, Graphics etc.

The probability distribution of the normal variable depends upon the two parameters μ and σ

- **The parameter μ is called the mean or expectation of the distribution.**
- **The parameter σ is the standard deviation; and variance is thus σ^2.**
- **Few terms:**
  - **Mode: Repeated terms**
  - **Median : middle data (if there are 9 data, the 5th one is the median)**
  - **Mean : is the average of all the data points**
  - **SD- standard Deviation, indicates how much the data is deviated from the mean.**
    - **Low SD indicates that all data points are placed close by**
    - **High SD indicates that the data points are distributed and are not close by.**
  - **SD given by the formula (S)**
  - **Where S is sample SD**

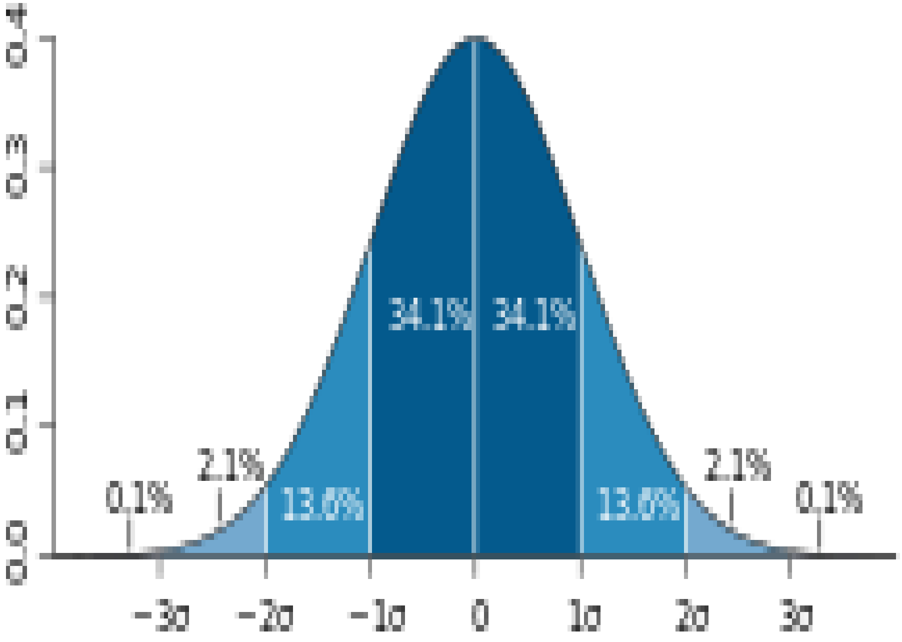
  - **If you want population SD, represented by**

σ^{2} and then divide by N not N-1

The Normal distribution

(mean μ, standard deviation σ)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



A plot of normal distribution (or bell-shaped curve) where each band has a width of 1 standard deviation – See also: 68–95–99.7 rule.

Standard Normal Distribution : In the above equation probability is computed for particular value of x. If you want a range then it has to be integrated.

For Standard Normal distribution:

- For standard normal distribution, the area under the given range is given by:

$$\int_a^b p(x)\,dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$$

$$= C\left(\frac{b-\mu}{\sigma}\right) - C\left(\frac{a-\mu}{\sigma}\right).$$

z

Problem: Normal distribution

- Consider an electrical circuit in which the voltage is normally distributed with mean 120 and standard deviation of 3. What is the probability that the next reading will be between 119 and 121 volts?
-

between 119 and 121

The z-transformation is

$$z = (x - 120)/3,$$

so we obtain from Figure 2.11 or 2.12

$$\begin{aligned} P(119 \leq x \leq 121) &= C\left(\frac{121-120}{3}\right) - C\left(\frac{119-120}{3}\right) \\ &= C(0.333) - C(-0.333) \\ &= 0.631 - (1 - 0.631) = 0.631 - 0.369 = 0.262. \end{aligned}$$

The value 0.631 was obtained by linear interpolation between $C(0.33)$ and $C(0.34)$

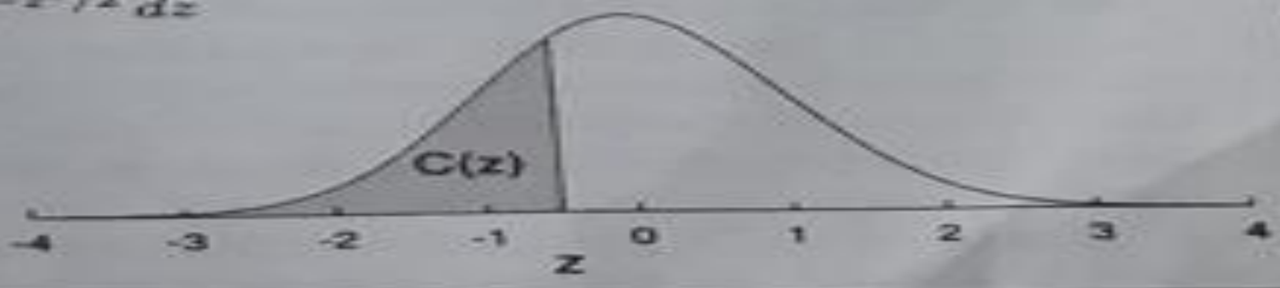
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986
3.0	0.998650								
3.5	0.9998674								
4.0	0.99996833								
4.5	0.999996602								
5.0	0.9999997133								
5.5	0.99999998101								
6.0	0.999999999013								
6.5	0.9999999999588								
7.0	0.99999999999872								

$$C(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

Figure 2.11: Areas under the st

z	0.00	−0.01	−0.02	−0.03	−0.04	−0.05	−0.06	−0.07	−0.08	−0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−3.0	1.350×10^{-3}									
−3.5	2.326×10^{-4}									
−4.0	3.167×10^{-5}									
−4.5	3.398×10^{-6}									
−5.0	2.867×10^{-7}									
−5.5	1.800×10^{-8}									
−6.0	9.866×10^{-10}									
−6.5	4.016×10^{-11}									
−7.0	1.280×10^{-12}									

$$C(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$



igure 2.12: Areas under the standard normal curve from $-\infty$ to z, where $z < 0$. For

1. Most graduate schools of business require applicants for admission to take the Graduate Management Admission Council's GMAT examination. Scores on the GMAT are roughly normally distributed with a mean of 527 and a standard deviation of 112. What is the probability of an individual scoring above 500 on the GMAT?

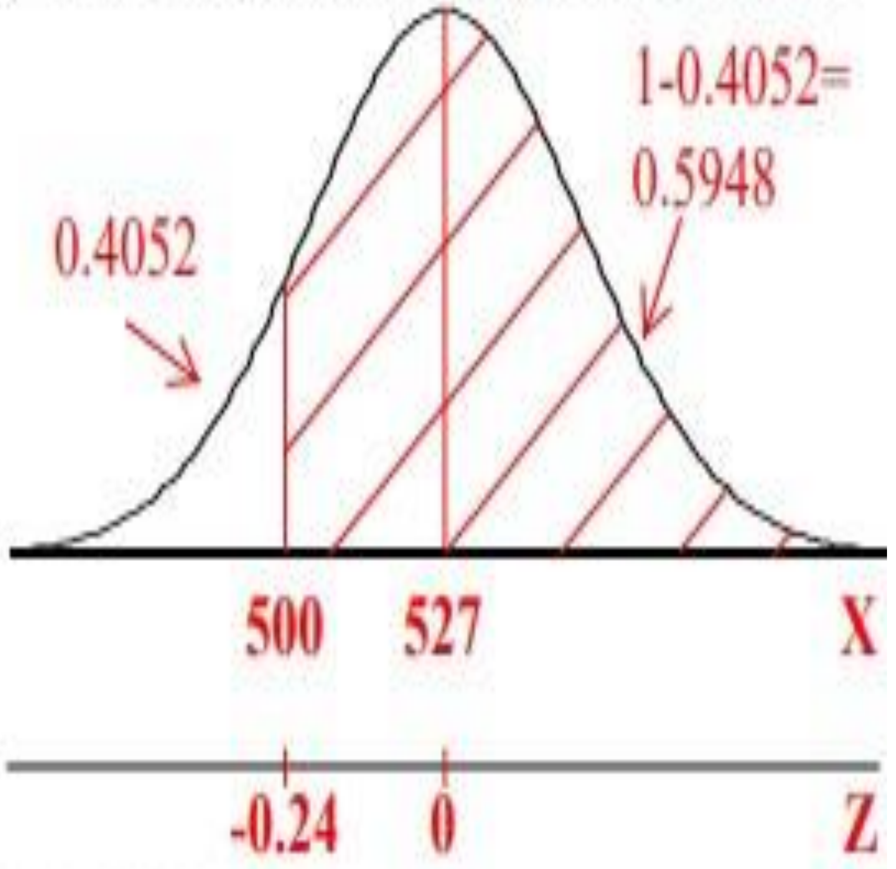
Normal Distribution

$$Z = \frac{500-527}{112} = -0.24107$$

$\mu = 527$

$\sigma = 112$

$\Pr\{X > 500\} = \Pr\{Z > -0.24\} = 1 - 0.4052 = \boxed{0.5948}$



Another problem

5. The average number of acres burned by forest and range fires in a large New Mexico county is 4,300 acres per year, with a standard deviation of 750 acres. The distribution of the number of acres burned is normal. What is the probability that between 2,500 and 4,200 acres will be burned in any given year?

Normal Distribution

$$Z = \frac{2500 - 4300}{750} = -2.40$$

$\mu = 4300$

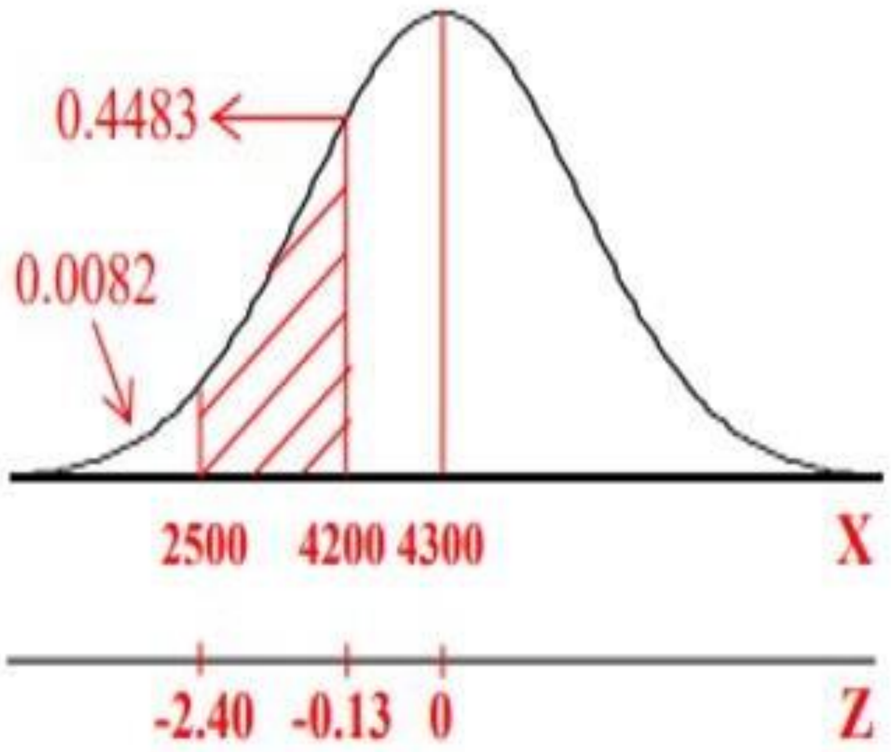
$$Z = \frac{4200 - 4300}{750} = -0.13333$$

$\sigma = 750$

$P(2500 < X < 4200) = P(-2.40 < Z < -0.13)$

$P(-2.40 < Z < -0.13) = P(Z < -0.13) - P(Z < -2.40)$

$P(-2.40 < Z < -0.13) = 0.4483 - 0.0082 = \boxed{0.4401}$



Joint Distributions and Densities

- The joint random variables (x,y) signifies that , simultaneously, the first feature has the value x and the second feature has the value y.

- If the random variables x and y are discrete, the joint distribution function of the joint random variable (x,y) is the probability of P(x,y) that both x and y occur.
- Thus, the joint distribution gives the probability of every possible combination of outcomes of the random variables that makeup the joint random variable.

Example 2.9 *The joint distribution function of the outcomes of flipping biased coins.*

A biased coin A has $P(head) = 0.6$ and coin B has $P(head) = 0.3$. Suppose that we flip coin A, and if the outcome is *head*, we flip coin B, but if the result of the first flip is *tail*, we flip coin A again. Since the probability of *head* on the second flip depends on the outcome of the first flip, the two events are not independent. Let x be the outcome of the first flip and let y be the outcome of the second flip. The joint distribution $P(x, y)$ is

(x, y)	$P(x, y)$
$(head, head)$	$(0.6)(0.3) = 0.18$
$(head, tail)$	$(0.6)(0.7) = 0.42$
$(tail, head)$	$(0.4)(0.6) = 0.24$
$(tail, tail)$	$(0.4)(0.4) = 0.16$

Joint distribution in continuous random variable

- If x and y are continuous, then the probability density function is used over the region R, where x and y is applied is used.
- It is given by:

$$P((x, y) \text{ is in } R) = \iint_R p(x, y)\, dx\, dy,$$

- Where the integral is taken over the region R. This integral represents a volume in the xyp plane.

Moments of Random Variables

Moments are very useful in statistics because they tell us much about our data.

Moments describe how the probability mass of a random variable is distributed.

- In mathematics, the **moments** of a function are quantitative measures related to the shape of the function's graph.
- It gives information about the spread of data, skewedness and kurtosis.

- If the function is a probability distribution, then there are four commonly used moments in statistics

    The first moment is the expected value - measure of center of the data

    The second central moment is the variance - spread of our data about the mean

    The third standardized moment is the skewness - the shape of the distribution

    The fourth standardized moment is the kurtosis - measures the peakedness or flatness of the distribution.

Computing Moments for population

Higher Order Moments

CENTRED

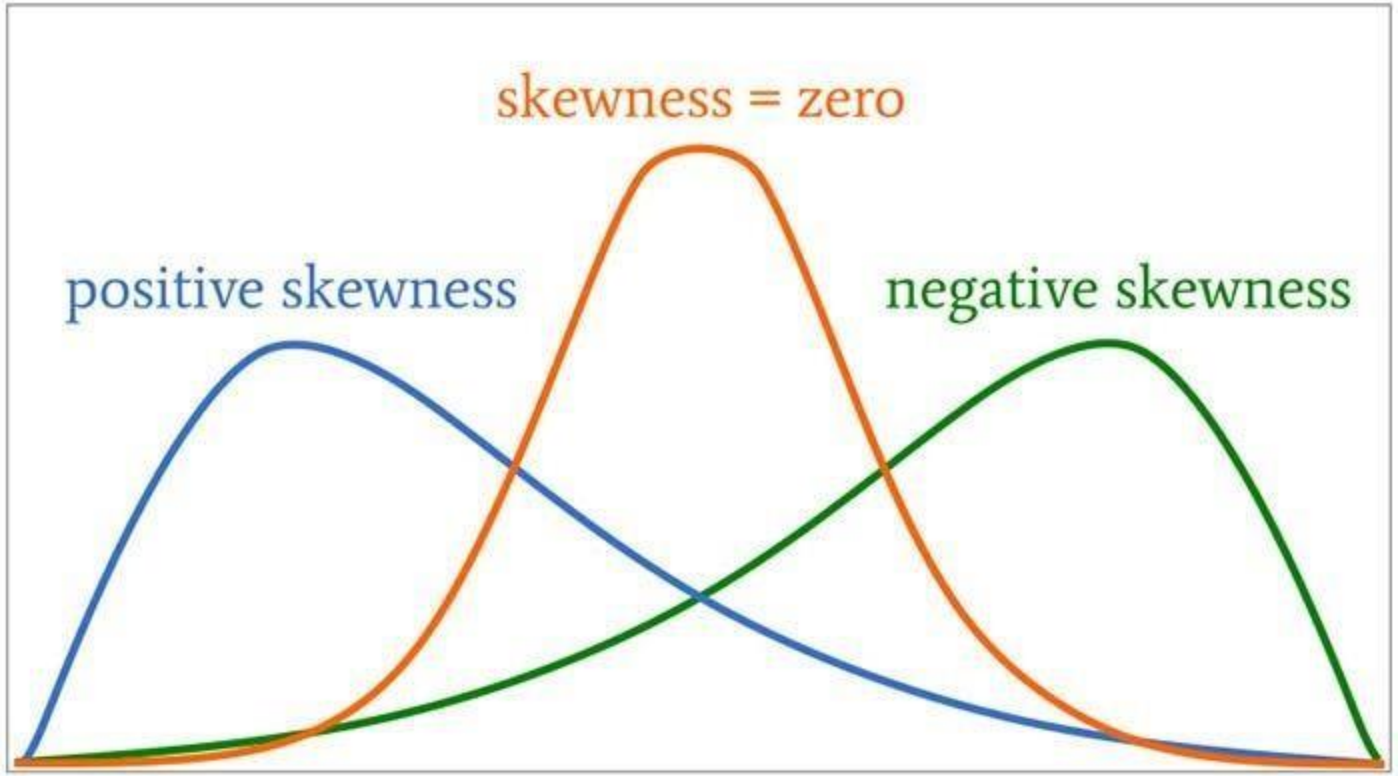
1. $\dfrac{\sum x}{n}$

2. $\dfrac{\sum x^2}{n}$ $\quad$ $\dfrac{\sum (x-\mu)^2}{n}$

3. $\dfrac{\sum x^3}{n}$ $\quad$ $\dfrac{\sum (x-\mu)^3}{n}$

4. $\dfrac{\sum x^4}{n}$ $\quad$ $\dfrac{\sum (x-\mu)^4}{n}$

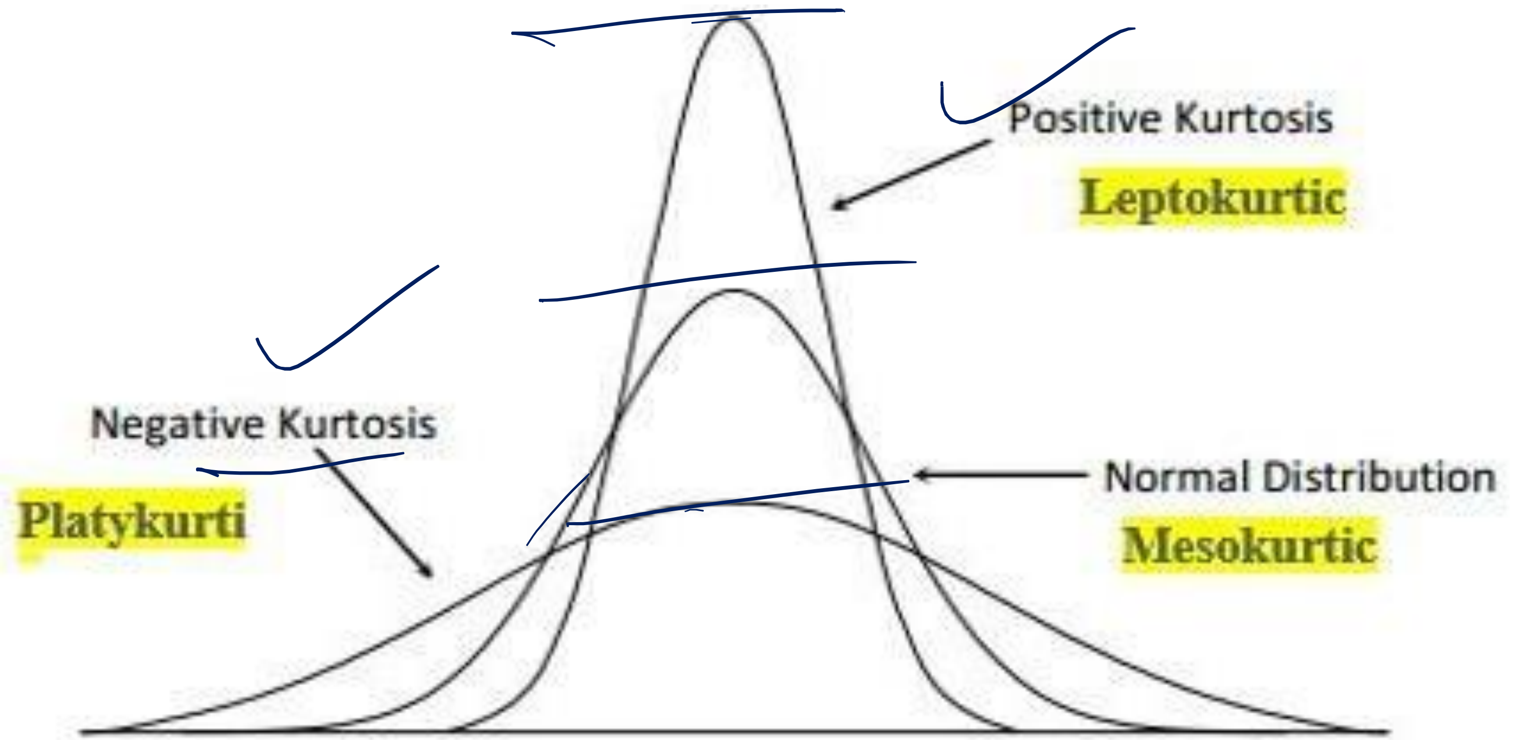
Moment 3: To know the Skewness

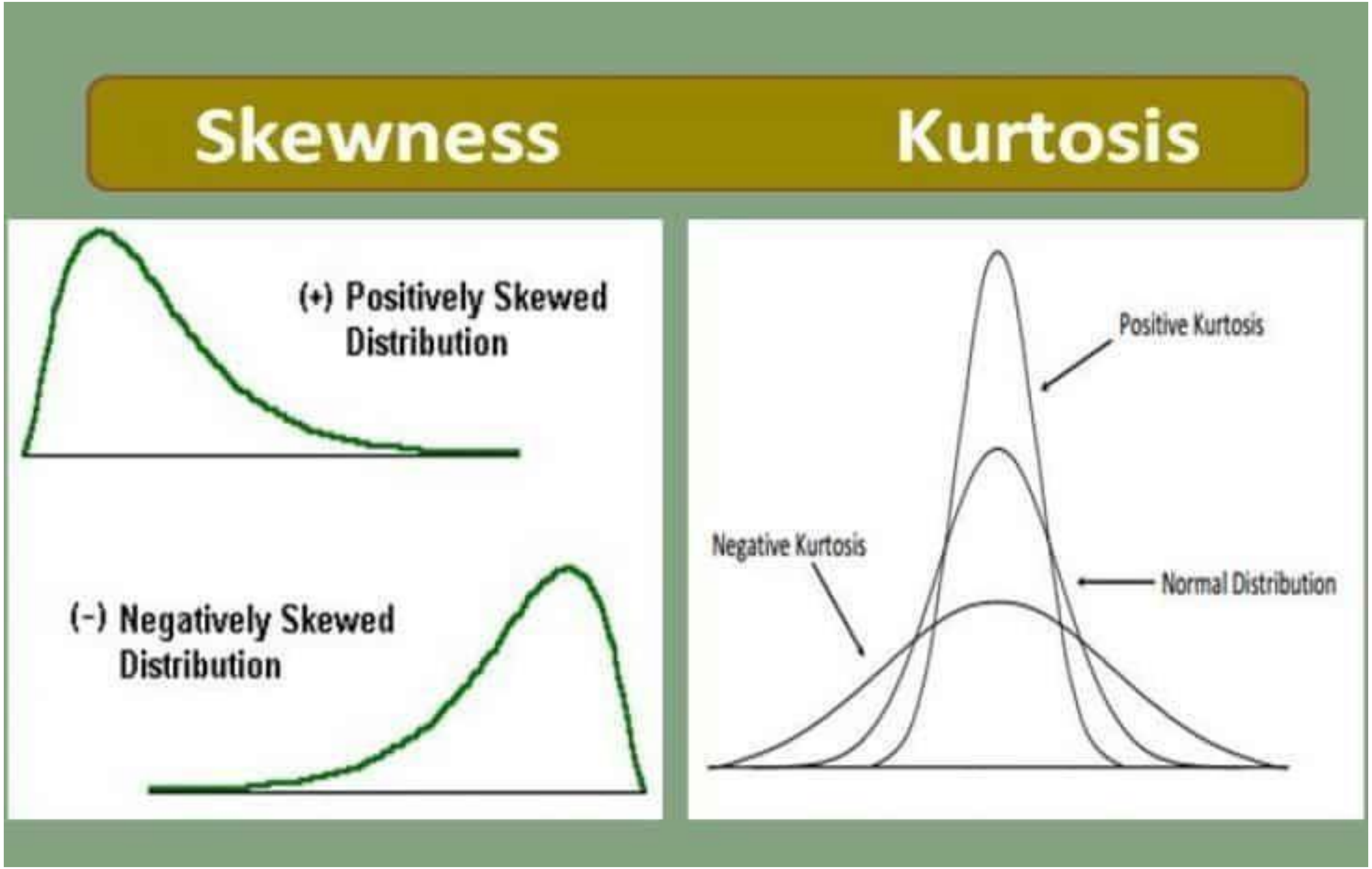


In positive Skewness,
Mean is > median
and
Median>mode

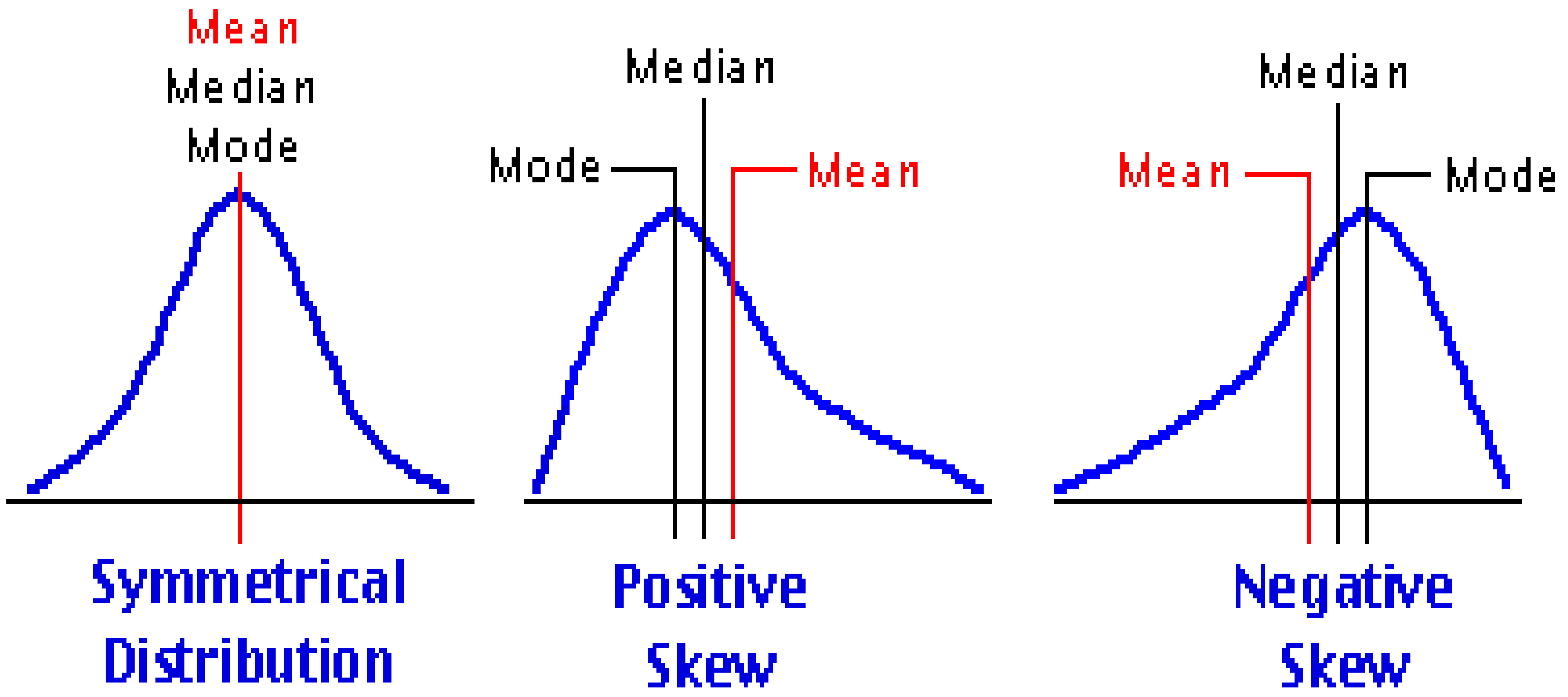
And it is reverse in case of –ve skewness

Moment 4 : To know the Kurtosis




Skewness
Kurtosis
(+) Positively Skewed Distribution
(-) Negatively Skewed Distribution
Positive Kurtosis
Negative Kurtosis
Normal Distribution

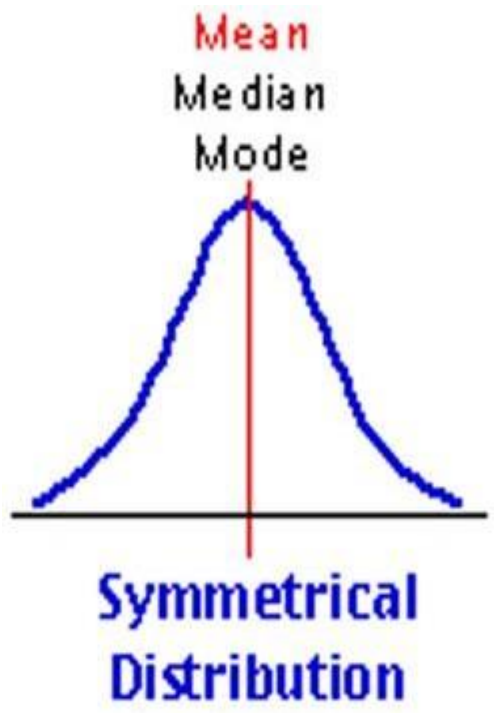
D


Mean
Median
Mode
Symmetrical
Distribution
Median
Mode
Mean
Positive
Skew
Median
Mean
Mode
Negative
Skew

Normal Distribution

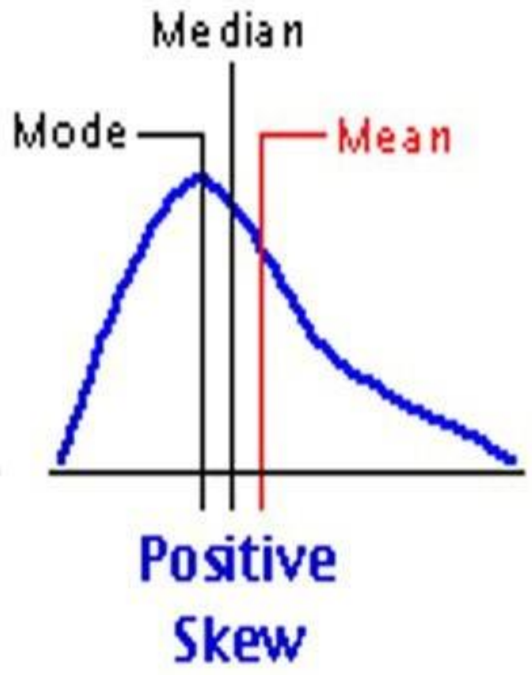
- Consider an example of x values:
- 4,5,5,6,6,6,7,7,8
- Mode, Median and mean all will be equal

- = Mode is 6
- = Median is 6
- = Mean is also 6

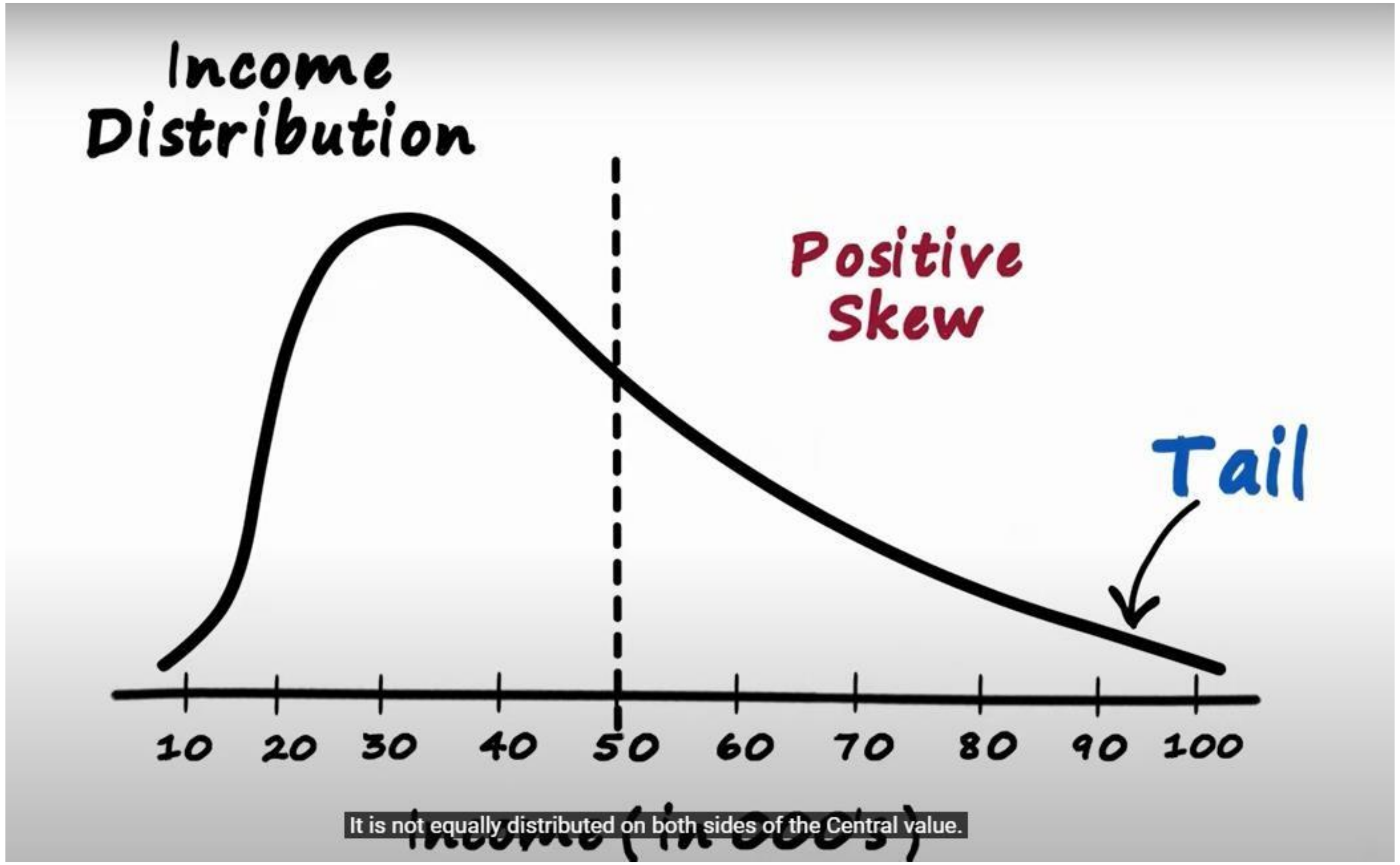


Positive Skew

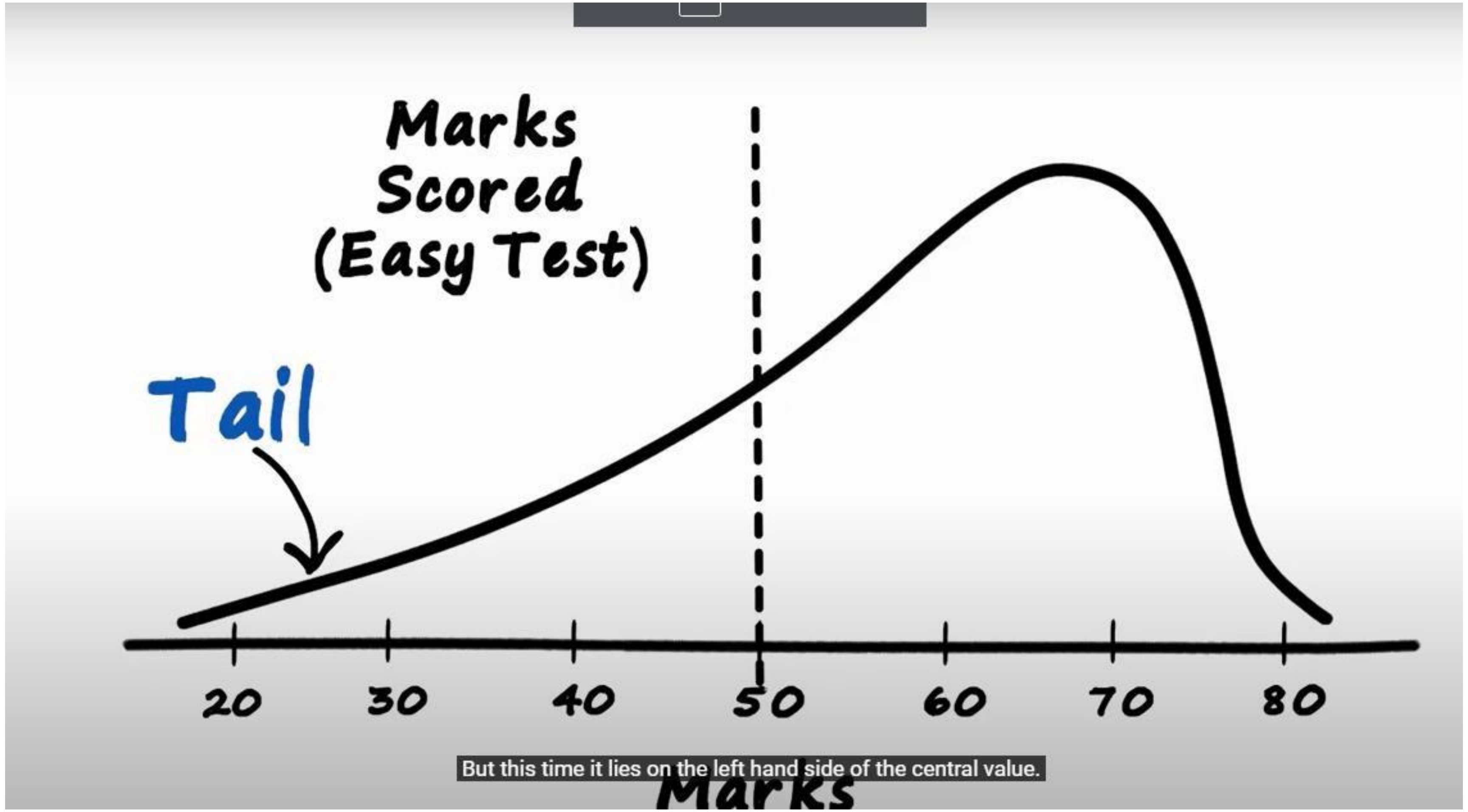
- Consider an example of x values:
- 5,5,5,6,6,7,8,9,10
- (It is an example for Normal Distribution)
- = Mode is 5
- = Median is 6
- = Mean is also 6.8



+ve skew


Income
Distribution
Positive
Skew
Tail
10
20
30
40
50
60
70
80
90
100
It is not equally distributed on both sides of the Central value.

-ve skew


Marks
Scored
(Easy Test)
Tail
20
30
40
50
60
70
80
But this time it lies on the left hand side of the central value.
Marks

Difference between PDF and PMF

1. The full form of PDF is Probability Density Function whereas the full form of PMF is Probability Mass Function
2. PMF is used when there is a need to find a solution in a range of discrete random variables whereas PDF is used when there is a need to find a solution in a range of continuous random variables.
3. PDF uses continuous random variables whereas PMF uses discrete random variables.
4. Pdf formula is $F(x)= P(a < x < b)= \int_a^b f(x)dx>0$ whereas pmf formula is $p(x)= P(X=x)$
5. The solutions of PDF falls in the radius of continuous random variables whereas the solutions of PMF falls in the radius

Moments for random variable:

- The “moments” of a random variable (or of its distribution) are expected values of powers or related functions of the random variable.

Formula for Computing Kth Central moment of Random variable

The k^{th} **central moment** of X

$$\mu_k^0 = E\left[\left(X-\mu\right)^k\right]$$

$$= \begin{cases} \sum_x \left(x-\mu\right)^k p\left(x\right) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} \left(x-\mu\right)^k f\left(x\right)dx & \text{if } X \text{ is continuous} \end{cases}$$

Let X be a discrete random variable having support $x = \langle 1, 2 \rangle$ and the pmf is

$$p_X(x) = \begin{cases} 3/4 & \text{if } x = 1 \\ 1/4 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases}$$

Using this compute mean (first order moment)

First order moment is the mean.

Solution:

$$\begin{aligned} \mu_X(1) &= \mathrm{E}[X] \\ &= \sum_{x \in R_X} p_X(x)x \\ &= \frac{3}{4} \cdot 1 + \frac{1}{4} \cdot 2 \\ &= \frac{5}{4} \end{aligned}$$

For example computation of 3rd Order moment

- The third central moment of can be computed as follows:
- Here X value is 1 and 2 and Probability is ¾ and ¼ respectively. Consider Mean is 5/4

$$
\begin{aligned}
\mu_X(3) &= \mathrm{E}\left[(X - \mathrm{E}[X])^3\right] \\
&= \sum_{x \in R_X} p_X(x)\left(x - \frac{5}{4}\right)^3 \\
&= \frac{3}{4} \cdot \left(1 - \frac{5}{4}\right)^3 + \frac{1}{4} \cdot \left(2 - \frac{5}{4}\right)^3 \\
&= \frac{3}{4} \cdot \left(-\frac{1}{4}\right)^3 + \frac{1}{4} \cdot \left(\frac{3}{4}\right)^3 \\
&= -\frac{3}{4^4} + \frac{27}{4^4} \\
&= \frac{24}{256} = \frac{3}{32}
\end{aligned}
$$

Estimation of Parameters from Samples

- There are 3 kinds of estimates for these parameters:
  - Method of moments estimates
  - Maximum likelihood estimates
  - Unbiased estimates.

End of Unit 1