Probability

what is probability?

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

In other words probability is a measure of likeliness of something happening.

Given 'n' observations of an event, it denotes the proportion of observations where a given event occurs

$$P(E) = \frac{\#\text{ of outcomes in which event occurs}}{\text{total possible }\#\text{ of outcomes}}$$

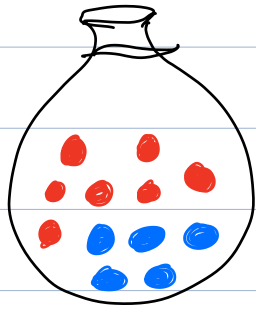
* The probability of an event is always between 0 and 1.

i.e $$\text{probability} = \frac{\text{Favorable outcomes}}{\text{Total outcomes}}$$

as favorable outcomes are always less than or equal to total outcomes so

$$0 \leq P(E \leq 1$$

ex:-



$P(\text{red}) = \dfrac{7}{12}$ ← Number of red marbles

← total no. of marbles (sample space)

$P(\text{blue}) = \dfrac{5}{12}$ ← Number of blue marbles

← total no. of marbles (sample space)

Probability - Terminology -

Experiment : An Experiment or trial is any procedure that can be infinitely repeated and has a well-defined set of possible outcomes, known as sample space.

Random Experiment : An Experiment is said to be random if it has more than one possible outcome and the outcome cannot be predicted.

* Experiment is deterministic if it has only one outcome.

Outcome : An outcome is a possible result of an experiment or trial.

Sample space : All the possible outcomes of an experiment is sample space.

Event : An event is a set of outcomes of an experiment whose probability has to be calculated.

- Simple Event : It has a single sample point.
- Compound Event : It has more than one sample point.
- Sure Event : It will happen for sure, i.e $P(E) = 1$
- Impossible Event : It will not happen i.e $P(E) = 0$

* Probability is a measure of how likely an event is. So, if it is 60% chance that it will rain tomorrow, the probability of outcome "it rained" for tomorrow is 0·6

Ex:-

i) Probability experiment : Roll a die

outcome : $\{3\}$

sample space : $\{1, 2, 3, 4, 5, 6\}$

Event : $\{$Die is Even$\} = \{2, 4, 6\}$

$$p(\text{Even}) = \frac{\text{len}\{2,4,6\}}{\text{len}\{1,2,3,4,5,6\}} = \frac{3}{6} = \frac{1}{2}$$

2) Probability Experiment: Deck of Cards

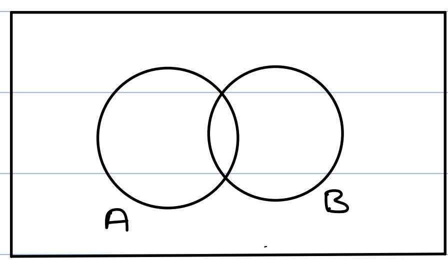
Sample Space = { . . . . . } (total 52 Cards)

Event : { card drawn is face card } (12 face cards)

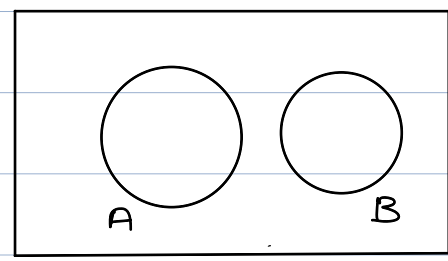
$$P(E = \text{face card}) = \frac{12}{52}$$

* probability can be thought in terms of sets and venn diagrams.

ex:-



Joint set

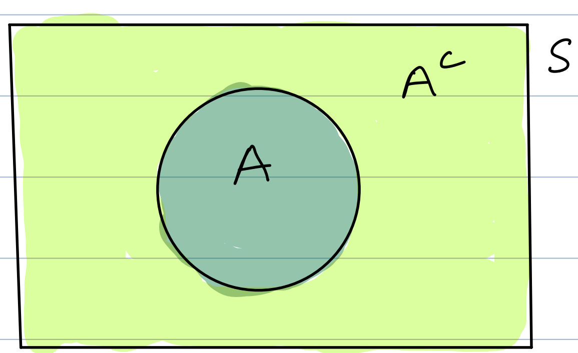


Dis-Joint set

We will see few of the Events:

1) Complimentary Event
2) Dependent Event
3) Independent Event
4) Mutually Exclusive Event
5) Exhaustive Event

Complimentary Event: The complement of an Event is the subset of outcomes in the sample space that are not in the event. A complement is itself an event.



Universal Set
↓
Sample space

set A
↓
Event

The complement of an event A is denoted as A' or $\bar{A}$ or A^c....

An even and its Complement are mutually exclusive and exhaustive.

ex:- A: It rained

A^c: It did not rain

$P(A^c) = 1 - P(A)$

or $P(A) + P(A^c) = 1$.

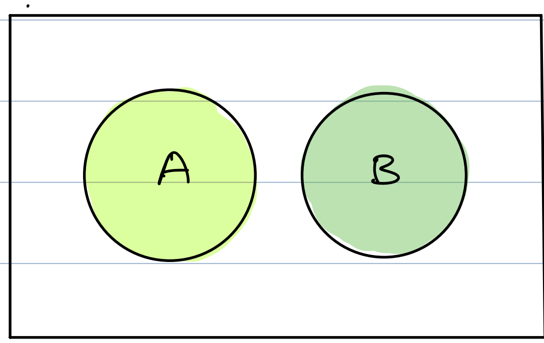
Dependent Events: Two are more events that depend on one another are known as dependent Events. If one event is by chance changed, then another is likely to differ.

ex:- Not paying power bill on time and having your power cut off.

Independent Events: Independent events are those events whose occurence is not dependent on any other event.

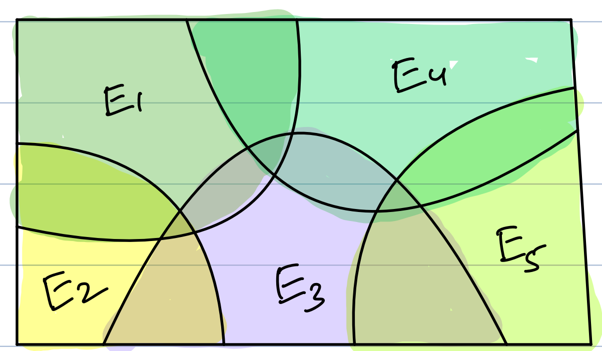
ex:- If we flip a coin in the air and get the outcome as Head, then again if we flip the coin but this time we get the outcome as Tail. In both cases the occurence of both events is independent of each other.

Mutually Exclusive Events: Two events are said to be mutually exclusive if they cannot occur at the same time or simaltaneously. In other words, mutually exclusive events are considered disjoint events, then the probability of both the events occuring at the same time will be zero.



$P(A \cap B) = 0.$

Exhaustive Events: A set of Events are called exhaustive events if atleast one of them necessarily occurs whenever the experiment is performed. Also the union of all these events constitutes the sample space of the experiment.



$E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5 = S \rightarrow$ sample space

ex:- In the toss of a coin event Heads and event Tail are mutually exhaustive events.

ODD's in Favour $\longrightarrow$ $\underset{\text{favourable outcomes}}{m} : \underset{\text{non-favourable outcomes}}{n}$

$$P(E) = \frac{m}{m+n}$$

ODD's Against $\longrightarrow$ $\underset{\text{non-favourable outcomes}}{m} : \underset{\text{favourable outcomes}}{n}$

$$P(E) = \frac{n}{m+n}$$

ex:- i) ODD's in favour of an event E is 3:7

$$P(E) = \frac{3}{10}$$

2) ODD's against an event E is 4:5

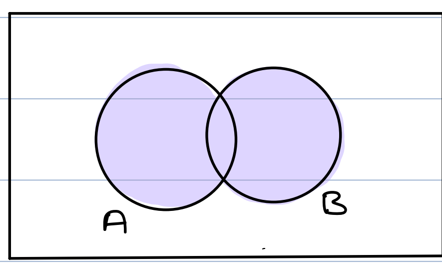
$$P(E) = \frac{5}{9}$$

Algebra of Events:

Consider two events A and B

- Intersection $(A \cap B)$: set of all elements common to A and B
- Union $(A \cup B)$: Set of all elements belonging to either A or B
- Difference $(A - B)$: Elements belonging to A but not to B.
- A and B are mutually exclusive or disjoint if $A \cap B = \phi$

Union $(A \cup B)$: The union of two sets A and B It is denoted by $A \cup B$.



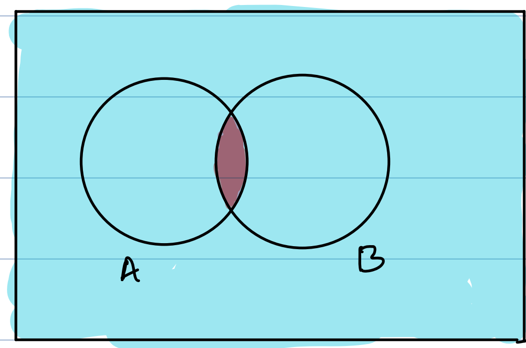
$A \cup B$
or
A or B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If A and B are mutually exclusive, then we have $P(A \cup B) = P(A) + P(B)$ as $P(A \cap B) = 0$.

Intersection ($A \cap B$) or (A and B):

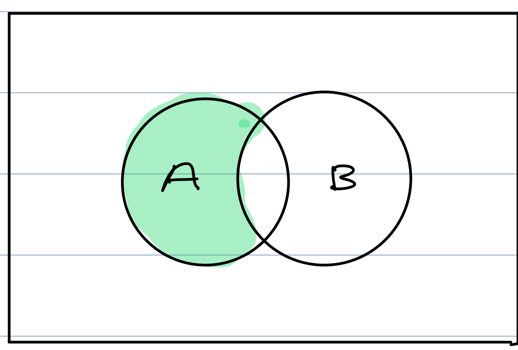
A intersection B is a set that contains elements that are common in both sets A and B.



$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

Difference (A - B):

The Difference between two events A and B is denoted by A-B and this contains the elements that are in A but not in B.



$$P(A-B) = P(A) - P(A \cap B)$$

Rules of probability:

1) The probability of an event always lies between 0 and 1.

$$0 \leq P(E) \leq 1$$

ex:- In a dice roll:

$$P(E=7) = \frac{0}{6} = 0$$

$$P(E \leq 6) = \frac{6}{6} = 1$$

2) Summation of Probabilities of all events in a sample space is 1. or probability of the univeral set is 1.

$$P(S) = 1$$

or

$$\sum_{i=1}^{n} P(E_i) = 1$$

ex:- In a toss of a coin $S = \{H, T\}$

$$P(H) = \frac{1}{2} \qquad P(T) = \frac{1}{2}$$

$$P(H) + P(T) = \frac{1}{2} + \frac{1}{2} = 1$$

3) If $E_1, E_2, E_3, \ldots$ are mutually exclusive then

$$P(E_1 \cup E_2 \cup E_3 \ldots) = P(E_1) + P(E_2) + P(E_3) + \ldots$$

ex:- In a toss of a coin:

$$P(H \cup T) = 1$$

$$P(H) + P(T) = \frac{1}{2} + \frac{1}{2} = 1$$

4) Probability of not getting the event is 1 minus probability of getting the event.

$$P(\text{not } E) = 1 - P(E)$$

ex:- In the toss of a coin

$$P(\text{not } H) = 1 - P(H)$$

$$= 1 - \frac{1}{2}$$

$$= \frac{1}{2}$$

In the roll of dice:

$$P(\text{not getting } 1) = P(2) + P(3) + P(4) + P(5) + P(6)$$

we know

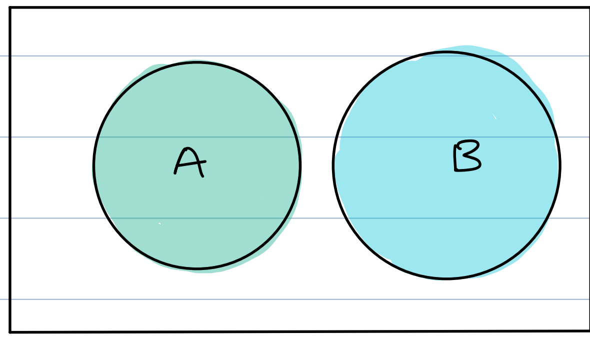
$$P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$$

$$\therefore P(2) + P(3) + P(4) + P(5) + P(6) = 1 - P(1)$$

$$\therefore P(\text{not getting } 1) = 1 - P(\text{getting } 1)$$

5) For dis-joint events A, B:

The probability of A or B is equal to probability of A plus probability of B.

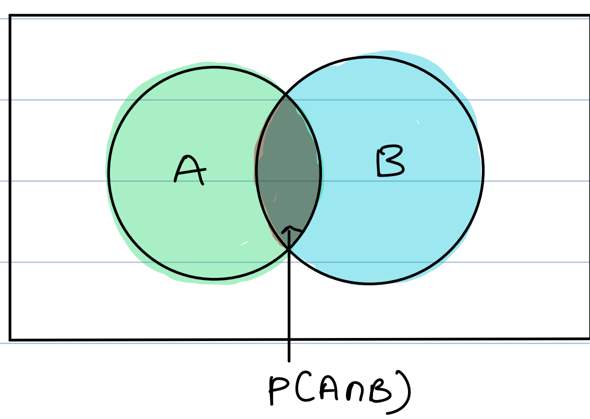


P(one or both events)

$$P(A \cup B) = P(A) + P(B)$$

6) For Joint events A, B:

The probability of A or B is equal to Probability of A plus probability B minus Probability of A and B.



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Types of probability:

1) Marginal Probability
2) Joint Probability
3) Conditional probability

Marginal probability: It is the probability of an event happening, such as $P(A)$, and it can be mentioned as unconditional probability. It does not depend on the occurence of another event.

ex:- The likelihood that a card is drawn from a deck of cards is black

$$P(\text{black}) = \frac{26}{52} = \frac{1}{2}$$

and the probability that a card drawn is 7

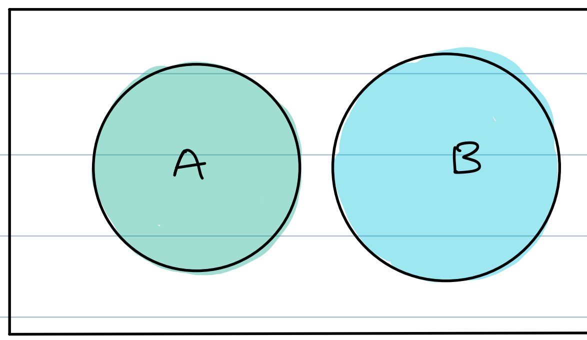
$$P(7) = \frac{4}{52} = \frac{1}{13}$$

both are independent events since the outcome of another event does not condition the result of one event.

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.

i.e 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.

With replacement



Joint probability
$P(A \cap B)$

ex:- Find the probability that the number three will occur twice when two dice are rolled at the same time.

$$S = \{1, 2, 3, 4, 5, 6\}$$

$$P(E = 3 \text{ on die } 1) = 1/6 = P(E_1)$$

$$P(E = 3 \text{ on die } 2) = 1/6 = P(E_2)$$

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$$

$$= \frac{1}{6} \times \frac{1}{6}$$

$$= \frac{1}{36}$$

ex:- The likelihood that a card is black and seven is equal to

$$P(\text{Black and seven}) = \frac{2}{52} = \frac{1}{26}$$

(There are 2 Black-7 in a deck of 52: the 7 of clubs and 7 of spades)

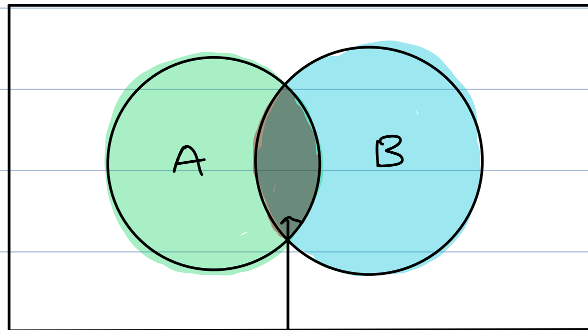
$$P(\text{Black}) = \frac{26}{52} = \frac{1}{2}$$

$$P(\text{seven}) = \frac{4}{52} = \frac{1}{13}$$

$$\therefore P(\text{Black}) \times P(\text{Seven}) = \frac{1}{2} \times \frac{1}{13} = \frac{1}{26}$$

$$\because P(\text{Black and seven}) = P(\text{Black}) \times P(\text{seven})$$

Joint Probability without replacement:



Joint probability
$P(A \cap B)$

$$P(A \cap B) = P(A) \times P(B|A)$$

probability of B happening given A has already happened

ex:- probability of drawing 2 cards from a deck one after another without replacement

- First spade card
- Second Face card.

Conditional probability:

Probability of a conditional event is conditional Probability.

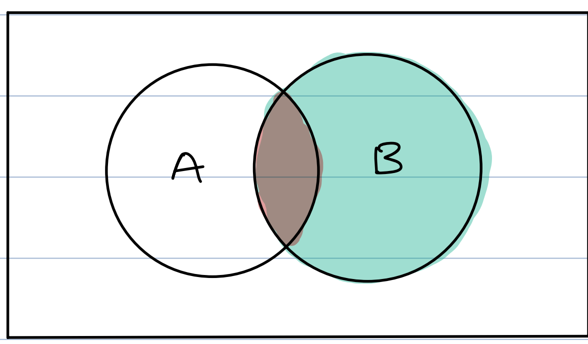
Conditional Event: Denoted by $(A|B)$ or $(B|A)$

$\left(\frac{A}{B}\right)$ – Event A has to happen when B has already happened.

$\left(\frac{B}{A}\right)$ – Event B has to happen when A has already happened.

P(A given B) or $P(A|B)$ – Probability of occurence of A given B has already happened.

so here the sample space becomes B



B has already happened so here

i.e we want to find probability of A happening within B which is nothin but

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

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

with regards to independent events

$$P(A|B) = P(A)$$

as $P(A)$ happenning has no dependence on $P(B)$ happenning or not.

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

$$\therefore\ P(A \cap B) = P(A) \times P(B)$$

for independent events

If A & B are independent, so are A^c & B, A & B^c, A^c & B^c.

i.e

$P(A^c \cap B) = (1 - P(A))\, P(B)$

$P(A^c \cap B^c) = (1 - P(A))(1 - P(B))$

if A & B are independent events

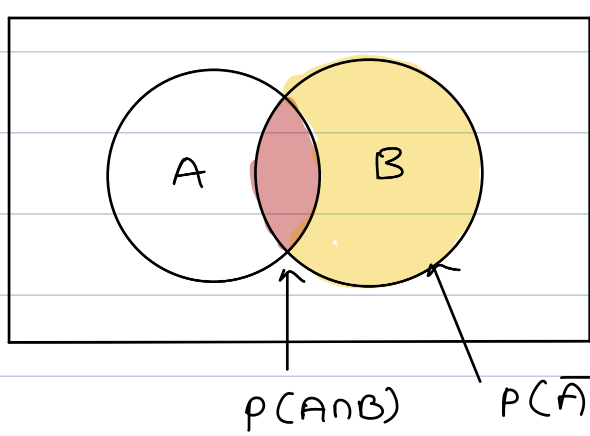
i.e if the existence of B does not matter to the existence of A then they are independent ex:- If two students are asked to solve a question. The probability of one solving the question is independent from the probability of the other solving it.

Q1) If A and B are mutually exclusive events then $\frac{A}{B}$ and $\frac{B}{A}$ are impossible events as

$$P\left(\frac{A}{B}\right) \; \& \; P\left(\frac{B}{A}\right) = 0 \quad \text{as} \quad P(A\cap B) = 0$$

Q2) $$P\left(\frac{A}{B}\right) + P\left(\frac{\bar{A}}{B}\right) = 1$$

$$\frac{P(A\cap B)}{P(B)} + \frac{P(\bar{A}\cap B)}{P(B)}$$



$$\therefore P(A\cap B) + P(\bar{A}\cap B) = P(B)$$

$$\therefore P\left(\frac{A}{B}\right) + P\left(\frac{\bar{A}}{B}\right) = \frac{P(B)}{P(B)}$$

$$= 1$$

Q3) A bag contains 5 Red balls, 4 blue balls and 3 green balls then find of Probability of drawing Red ball given a blue ball was already drawn without replacement.

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

$$= \frac{\frac{5}{11} \times \frac{4}{12}}{\frac{4}{12}} = \frac{5}{11}$$

Q4) If $P(E) = 0.40$, $P(F) = 0.35$ & $P(E \cup F) = 0.55$ find $P(E|F)$ & $P(F|E)$

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

$$\therefore P(E \cap F) = P(E) + P(F) - P(E \cup F)$$

$$= 0.40 + 0.35 - 0.55$$

$$= 0.20$$

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{0.2}{0.35} = \frac{4}{7}$$

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{0.2}{0.4} = \frac{1}{2}$$

Q5) A couple has 2 children. Find the Probability that both are boys, if it is known that

(1) one of them is a boy.

(2) the older child is a boy

$S = \{GG, GB, BG, BB\}$

(1) Here one needs to be boy so $S = \{GB, BG, BB\}$

$$\therefore \ P(\text{both boys}) = \frac{\text{len}\{BB\}}{\text{len}\{GB, BG, BB\}} = \frac{1}{3}$$

(2) Here the first child is already boy

so $S = \{BG, BB\}$

$$\therefore \ P(\text{both boys}) = \frac{\text{len}\{BB\}}{\text{len}\{BG, BB\}} = \frac{1}{2}$$

Q) If $P_1, P_2, P_3, \dots P_n$ are probabilities that Certain events happen, then Probability of

(1) all of them happening simaltaneously is →

$\because$ Independent events then

$P_1 \times P_2 \times P_3 \times \dots \times P_n$

(2) None of them happening is

$\because$ Independent events

$(1-p_1) \times (1-p_2) \times (1-p_3) \times \cdots \times (1-p_n)$

(3) Atleast one of them happening is : means
$1-$ none of them happening.
$1-((1-p_1)\times(1-p_2)\times(1-p_3)\times\cdots\times(1-p_n))$

Q) A & B are 2 independent events. The probability that both A & B occur is $1/6$ & probability that none of them occur is $1/3$. Find the probability of occurence of A.

if $P(A) = p$ & $P(B) = q$

Since independent $P(A \cap B) = P(A)P(B)$

$\therefore pq = \frac{1}{6}$ & $(1-p)(1-q) = \frac{1}{3}$

$$1 - p - q - pq = \frac{1}{3}$$
$$1 - p - q - \frac{1}{6} = \frac{1}{3}$$

$p+q = \frac{5}{6}$ since $pq = \frac{1}{6}$

Solution is either $p = \frac{1}{2}$ & $q = \frac{1}{3}$ or $p = \frac{1}{3}$ & $q = \frac{1}{2}$

Q) A problem of mathematics is given to 3 students whose chances of solving are $\frac{1}{2}, \frac{1}{3}$ & $\frac{1}{4}$. what is the probability that the problem is solved.

Since problem of solving is given then problem of not solving are

$$\left(1-\frac{1}{2}\right) \times \left(1-\frac{1}{3}\right) \times \left(1-\frac{1}{4}\right)$$

problem of atleast one solving is

1 − problem of not solving

$$= 1 - \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4}$$

$$= \frac{3}{4}$$

* if we want problem is solved exactly by 1.

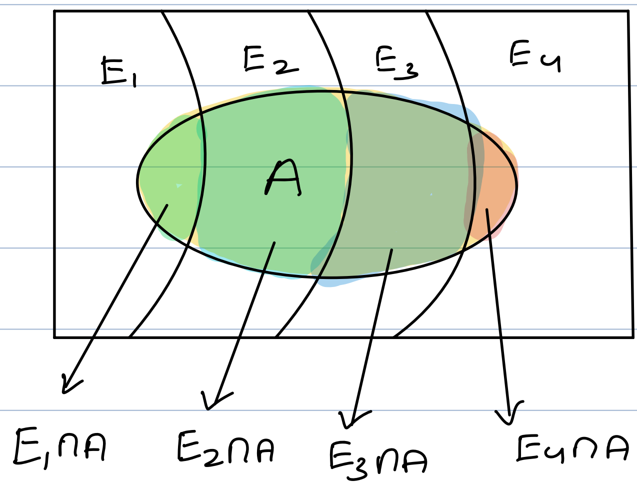
$$\frac{1}{2} \times \left(1-\frac{1}{3}\right) \times \left(1-\frac{1}{4}\right) + \left(1-\frac{1}{2}\right) \times \frac{1}{3} \times \left(1-\frac{1}{4}\right) \times$$

$$\left(1-\frac{1}{2}\right) \times \left(1-\frac{1}{3}\right) \times \frac{1}{4}$$

$$= \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4} + \frac{1}{2} \times \frac{1}{3} \times \frac{3}{4} + \frac{1}{2} \times \frac{2}{3} \times \frac{1}{4}$$

$$= \frac{11}{24}$$

Total Probability Theorem



If E_1, E_2, E_3, E_4 are 4 mutually Exclusive and Exhaustive events

then $p(A) = p(E_1 \cap A) + p(E_2 \cap A) + p(E_3 \cap A) + p(E_4 \cap A)$

if there are n Events which are mutually exclusive and exhaustive events.

$\therefore p(A) = p(E_1 \cap A) + p(E_2 \cap A) + \ldots\ldots + p(E_n \cap A)$

but $p(A | E_1) = \dfrac{P(A \cap E_1)}{P(E_1)}$

$\therefore P(A \cap E_1) = p(E_1) \times p(A | E_1)$

Substituting

Total probability theorem

$$P(A) = P(E_1) \times P(A|E_1) + P(E_2) \times P(A|E_2) + \cdots + P(E_n) \times P(A|E_n)$$

for $E_1, E_2, \cdots E_n$ which are mutually exclusive and exhaustive events.

Chain rule for events

Also known as product rule permits the calculation of any member of the joint distribution of a set of random variables using only conditional probabilites. For two random events A & B

$$P(A \cap B) = P(B|A) \cdot P(A)$$

Baye's theorem

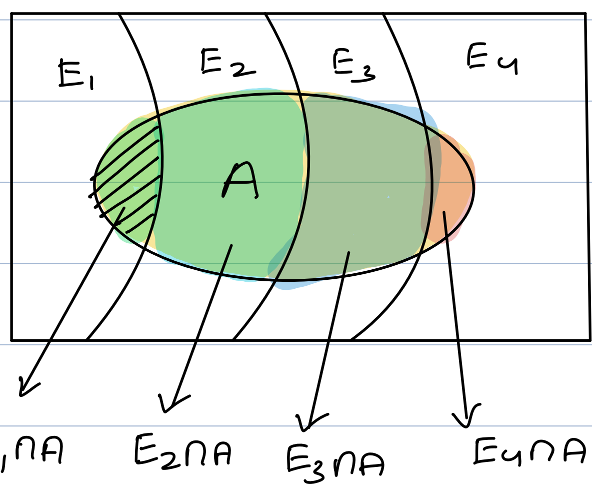
If we want to find the probability of event E_i given A has already happened then

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

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

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

another way to think :



If E_1, E_2, E_3, E_4 are 4 mutually Exclusive and Exhaustive events

then $$P(A|E_1) = \frac{P(E_1 \cap A)}{P(E_1 \cap A) + P(E_2 \cap A) + P(E_3 \cap A) + P(E_4 \cap A)}$$

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

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

where $P(A|B)$ is the probability of condition when event A is occuring while event B has already occured.

Bayes theorem Derivation

Bayes theorem can be derived for events and random variables using the definition of conditional probability.

from the definition of conditional probability.

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

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

Here, the joint probability $P(A \cap B)$ of both events A and B being true such that

$P(A\cap B) = P(B \cap A)$

$P(A\cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)$

$\therefore\ P(A|B) = \dfrac{P(B|A)\,P(A)}{P(B)}$ where $P(B) \neq 0$

Q1) <u>Balls & bags</u>

4R	3R	2R
5G	6G	7G
Bag 1	Bag 2	Bag 3

(1) you draw one ball find the probability it is Red.

p (drawing a red ball)

$= P(\text{selecting bag } 1) \times P(\text{red} \mid \text{bag } 1) +$

$\quad P(\text{selecting bag } 2) \times P(\text{red} \mid \text{bag } 2) +$

$\quad P(\text{Selecting bag } 3) \times P(\text{red} \mid \text{bag } 3)$

$= \frac{1}{3} \times \frac{4}{9} + \frac{1}{3} \times \frac{3}{9} + \frac{1}{3} \times \frac{2}{9} = \frac{1}{3}$

(2) You draw a red ball find the probability that it came from bag 2.

$p(\text{selected bag } 2 \mid \text{red ball})$

$$= \frac{P(\text{selecting bag 2}) \times P(\text{red} \mid \text{bag 2})}{P(\text{drawing a red ball})}$$ ← from total probability theorem

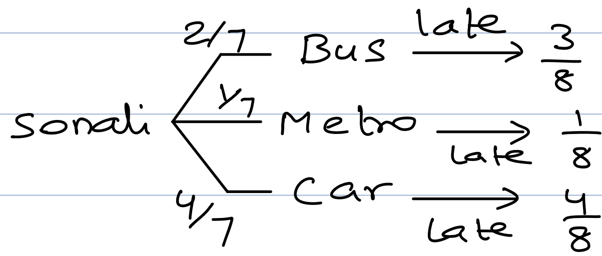
$$= \frac{\frac{1}{3} \times \frac{3}{9}}{\frac{1}{3} \times \frac{4}{9} + \frac{1}{3} \times \frac{3}{9} + \frac{1}{3} \times \frac{2}{9}}$$

$$= \frac{1/9}{1/3} = \frac{3}{9} = \frac{1}{3}$$

Q2) <u>modes of transport</u>

sonali goes by bus, metro & car with probabilities of $2/7$, $1/7$, $4/7$ respectively. If the probability that she goes by bus and reaches late to school is $3/8$. she goes by metro and reaches late is $\frac{1}{8}$ and she goes by car and reaches late is $\frac{4}{8}$.

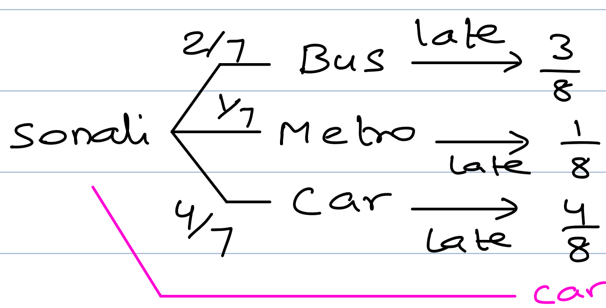
(1) Find the probability that sonali reached school late.



$$P(\text{Sonali late}) = \frac{2}{7} \times \frac{3}{8} + \frac{1}{7} \times \frac{1}{8} + \frac{4}{7} \times \frac{4}{8}$$

$$\begin{aligned} P(\text{Late}) &= P(\text{Bus}) \times P(\text{Late} \mid \text{bus}) + \\ &\quad P(\text{metro}) \times P(\text{late} \mid \text{metro}) + \\ &\quad P(\text{Car}) \times P(\text{late} \mid \text{car}) \\ &= \frac{2}{7} \times \frac{3}{8} + \frac{1}{7} \times \frac{1}{8} + \frac{4}{7} \times \frac{4}{8} \\ &= \frac{23}{56} \end{aligned}$$

(2) Sonali reached School late find the Probability that she travelled by car.



$$\text{probability}(\text{sonali car and late}) = \frac{4}{7} \times \frac{4}{8}$$

$$P(\text{sonali car} \mid \text{late}) = \frac{P(\text{sonali car and late})}{P(\text{late})}$$

$$p(\text{sonali car} \mid \text{late}) = \frac{\frac{4}{7} \times \frac{4}{8}}{\frac{2}{7} \times \frac{3}{8} + \frac{1}{7} \times \frac{1}{8} + \frac{4}{7} \times \frac{4}{8}}$$

$$= \frac{16/56}{23/56} = \frac{16}{23}$$

Q) <u>Truth or lie</u> :

Consider a student A, whose probability to tell the truth is $\frac{3}{8}$ and tell a lie is $\frac{5}{8}$

(1) A rolls a die find the probability that he reports a 6.

$$p(\text{reports } 6) = p(\text{truth}) \times p(\text{six}) + p(\text{lie}) \times p(\text{not six})$$

$$= \frac{3}{8} \times \frac{1}{6} + \frac{5}{8} \times \frac{5}{6}$$

$$= \frac{28}{48} = \frac{7}{12}$$

(2) He rolls a die and reports a '6', find the probability that it is actually '6'

$$p(\text{Truth} \mid 6) = \frac{p(\text{truth}) \times p(\text{six})}{p(\text{reports a six})}$$

$$= \frac{\frac{3}{8} \times \frac{1}{6}}{\frac{3}{8} \times \frac{1}{6} + \frac{5}{8} \times \frac{5}{6}}$$

$$= \frac{3}{28}$$

Q) <u>Biased coin</u>

1 Fair coin — $\{H, T\}$

1 Biased coin — $P(H) \to 3/5$, $P(T) \to 2/5$

(1) find the probability of getting a head.

$$P(\text{head}) = P(\text{fair coin}) \times P(\text{head} \mid \text{fair coin}) + P(\text{biased coin}) \times P(\text{head} \mid \text{biased coin})$$

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

$$= \frac{11}{20}$$

(2) find the probability that it was a biased coin given head.

$$P(\text{biased} \mid \text{head}) = \frac{P(\text{biased}) \times P(\text{head} \mid \text{biased})}{P(\text{head})}$$

$$= \frac{\frac{1}{2} \times \frac{3}{5}}{\frac{11}{20}}$$

$$= \frac{6}{11}$$

Q) In loan defaulters older people make only 1.4%. Now the probability that someone defaults on a loan is 0.184. Find the probability of default on loan knowing that he is an old person. older people make up only 0.8%.

$P(\text{older people} \mid \text{loan default}) = 1.4\%$
$= 0.014$

$P(\text{loan default}) = 0.184$

$P(\text{older people}) = 0.8\%$
$= 0.008$

$P(\text{loan default} \mid \text{older people}) = ??$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$ (Bayer's theorem)

$\therefore$ P(loan default | older people)

$$= \frac{P(\text{older people} \mid \text{loan default}) \cdot P(\text{loan default})}{P(\text{older people})}$$

$$= \frac{0.014 \times 0.184}{0.008}$$

$$= 0.322$$

Q) Spam Assassin works by having users train the system. It looks for patterns in the words in emails marked as spam by the user.
For example, it may have learned that word 'free' appears in 30% of the mails marked as spam, i.e $P(\text{Free} \mid \text{spam}) = 0.30$. Assuming 1% of non-spam mail includes the word 'free' and 50% of all mails received by the user are spam, find the probability that a mail is spam if the word 'free' appears in it.

$P(\text{Free} \mid \text{spam}) = 0.30$

$P(\text{Free} \mid \text{non-spam}) = 1\% = 0.01$

$p(\text{spam}) = 50\%$ $\quad \therefore \; p(\text{non-spam}) = 0.5$

$= 0.5$

$p(\text{spam} \mid \text{free}) = ??$

<u>method 1 to find the answer:</u>

$$p(\text{spam} \mid \text{free}) = \frac{p(\text{free} \mid \text{spam}) \cdot p(\text{spam})}{p(\text{free})}$$

Total emails : 100

spam emails : 50 , non-spam emails = 50

30% free (15 mails) 70% no free (35 mails)

1% free (0.5 mails) 99% no free (49.5 mails)

$\therefore$ total free emails = 15.5

$$p(\text{free}) = \frac{15.5}{100} = 0.155$$

$$p(\text{spam} \mid \text{free} = \frac{0.3 \times 0.5}{0.155} = 0.96774$$

method 2 to find answer:

formula:

$$P(spam|free)$$

$$= \frac{P(spam) * P(free|spam)}{P(free|spam) * P(spam) + P(free|no\text{-}spam) * P(no\text{-}spam)}$$

$$= \frac{0.5 \times 0.3}{0.3 \times 0.5 + 0.01 \times 0.5}$$

$$= \frac{0.15}{0.15 + 0.005}$$

$$= 0.96774$$

Frequency tables and contingency tables :

Contingency table - both categorical variables.

frequency table - one categorical variables.

Contingency tables : (Also called Cross-tab or two-way tables) summarize the relationship between several categorical variables. It is a special type of frequency distribution table, where two variables are shown simaltaneously.

ex:- Contingency table summarising 2 variables, Loan Default and Age. (Cross-tab)

		Age			
		Young	middle-Aged	old	Total
Loan Defaults	NO	5252	13684	130	19066
	Yes	1793	2426	60	4279
	Total	7045	16110	190	23345

Now if we want to convert them into probabilities we will divide each cell with total which is 23345

i.e $\frac{5252}{23345} = 0.225$ so on.

Probabilities

		Age			
		Young	middle-Aged	old	Total
Loan Defaults	NO	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Marginal Probability

Probability describing a single attribute

		Age			
		Young	middle-Aged	old	Total
Loan Defaults	NO	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Marginal Probability

ex:-

$P(yes) = 0.184$

$P(old) = 0.008$

Joint probability

Probability describing a combination of attributes

		Age			
		Young	middle-Aged	old	Total
Loan Defaults	NO	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Joint Probability

ex:-

P(Yes and Old) = 0.003

P(NO and Young) = 0.225

Union probability:

Probability of happening of the event A or B.

		Age			
		Young	middle-Aged	old	Total
Loan Defaults	NO	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

ex:-

$p(\text{yes or old}) = p(\text{yes}) + p(\text{old}) - p(\text{yes and old})$

$= 0.184 + 0.008 - 0.003$

$= 0.189$

Conditional Probability:

It is ratio between joint probability and marginal probability

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

		Age			
		Young	middle-Aged	old	Total
Loan	NO	0.225	0.586	0.005	0.816
Defaults	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

ex:- Probability that a person will not default on Loan given he/she is middle aged.

$$p(\text{NO} | \text{middle-aged}) = \frac{p(\text{NO and middle-aged})}{p(\text{middle aged})}$$

$$= \frac{0.586}{0.690} = 0.85$$

ex:- Considering 'titanic' data set in pandas.

- First do a crosstab between 'pclass' and 'survived'.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df = sns.load_dataset("titanic")
pd.crosstab(df['pclass'], df['survival'],
                 margins = True)
```

survived	0	1	All
pclass			
1	80	136	216
2	97	87	184
3	372	119	491
All	549	342	891

This is a contingency or frequency table

```
pclassSurv = pd.crosstab(df['pclass'], df[survival],
                 margins = True, margin_names='Total')

pclassSurv
```

Survived	0	1	Total
pclass			
1	80	136	216
2	97	87	184
3	372	119	491
Total	549	342	891

To create marginal probability table divide the variable pclassSurv by total ie 891.

pclassSurv_ProbTable = pclassSurv / 891

pclassSurv_ProbTable

↳

Survived	0	1	Total
pclass			
1	0.0897..	0.1526..	0.2424..
2	0.1088..	0.0976..	0.2065...
3	0.4175..	0.1335..	0.5510..
Total	0.6161..	0.3838..	1.0000..

From this table we can find marginal, Joint, union and conditional probability.

ex:- P(pclass1) = 0.2424

$P(\text{Pclass } 2 \text{ and } \underset{\substack{\downarrow \\ 1}}{\text{survived}}) = 0.0976..$

$P(\text{P class } 3 \text{ or not } \underset{\substack{\downarrow \\ 0}}{\text{Survived}})$

$= P(\text{Pclass } 3) + P(\text{not survived}) -$
$P(\text{Pclass } 3 \text{ and not survived})$

$= 0.5510.. + 0.6161... - 0.4175....$

$= 0.7496$

$P(\text{Pclass } 3 \mid \text{not survived})$

$$= \frac{P(\text{Pclass } 3 \text{ and not survived})}{P(\text{not survived})}$$

$$= \frac{0.4175...}{0.6161...}$$

$= 0.6776...$

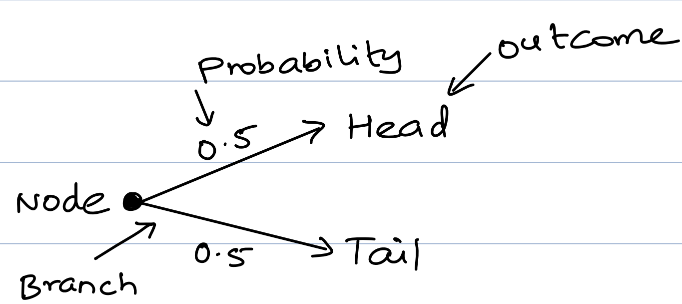
Probability Tree :

A probability tree is a picture indicating probabilities and some conditional probabilities for combinations of two or more events.

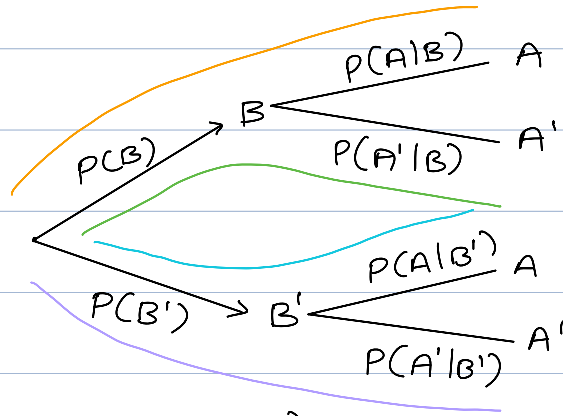
- One of the easiest ways to solve a probability problem is to construct the probability tree and read the answer.

The steps to draw a probability tree are as follows:

1) Draw branches of the first set of outcomes and write the individual probabilities along the branches.
2) Repeat this process for the remaining outcomes.
3) multiply the values of probabilities of connecting branches to get the likelihood of occurence of each outcome.



Generalized Probability Tree



$P(B \text{ and } A) = P(B) \times P(A|B)$

$P(B \text{ and } A') = P(B) \times P(A'|B)$

$P(B' \text{ and } A) = P(B') \times P(A|B')$

$P(B' \text{ and } A') = P(B') \times P(A'|B')$

$P(B') - P(\text{not } B)$

$P(A') - P(\text{not } A)$

Conditional probability, Bayes theorem

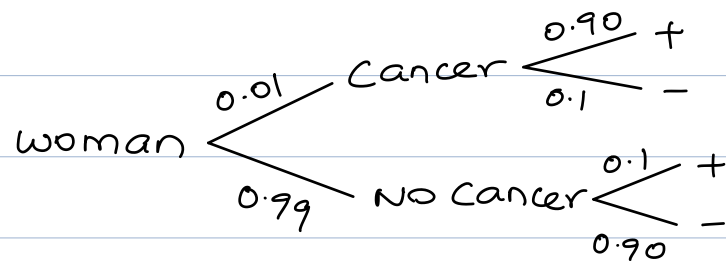
$$P(B|A) = \frac{P(B) \times P(A|B)}{P(A)}$$

$$= \frac{P(A|B) \times P(B)}{P(A|B) \times P(B) + P(A|B') \times P(B')}$$

Q) Approximately 1% of women aged 40-50 have breast cancer. A women with breast cancer has a 90% chance of a positive test from a mammogram, while a woman without has a 10% chance of a false positive result.

(1) what is the probability a woman has breast

cancer given that she had a positive result.



$$p(\text{cancer} \mid +) = ?$$

$$= \frac{p(+ \mid \text{cancer}) \times p(\text{cancer})}{p(+)}$$

$$= \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.99 \times 0.1}$$

$$= 0.083$$

<u>Note</u>: Since we are calculating cancer given test positive we are considering the values associated with test positive only.