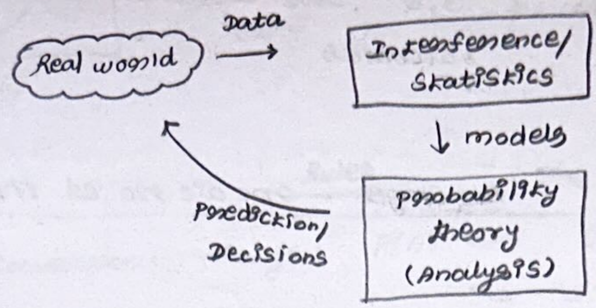
Probability & Random variables



Probability models & Axioms

Probabilistic model: quantitative description of a uncertain model situation

1) possible outcomes (Sample space)
2) Law → Assign probability for outcomes (whether one outcome is dominant?)

Axioms — properties that follow from the axioms

Sample space

* possible outcome
* Beliefs about likelihood of outcomes

} Two steps.

*** Set of possible outcome (Ω) → capital omega**

List (properties):

→ mutually exclusive → **At the end only one outcome.**

→ collectively exhaustive → **The result is the element of Ω.**

→ At the right granularity (Distinguishable – physically different outcomes of discrete elements)

Ω

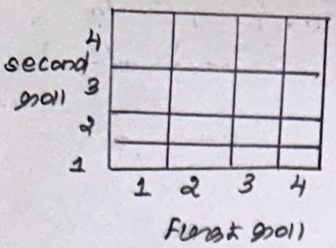
H	T

Exactly one o/p going to happen from Ω.

Sample spaces of some examples

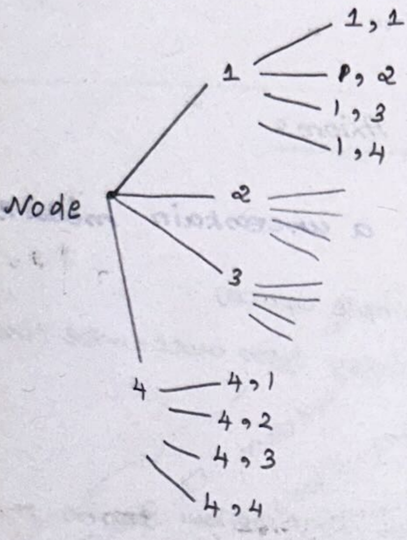
→ Tetrahedral die (Two rolls) → 4 faces

$$\Omega = \left\{ \begin{matrix} (1,1), & (1,2), & (1,3), & (1,4) \\ (2,2) & \cdots & & \\ (4,1) & \cdots & & (4,4) \end{matrix} \right\} \quad n(\Omega) = 16 \text{ elements}$$



2,3 & 3,2 are different outcomes

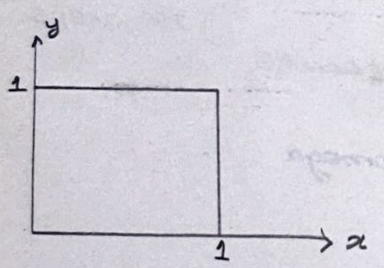
'Two stages' – one die rolled then other



16 possible outcomes

Sequential description

Continuous example



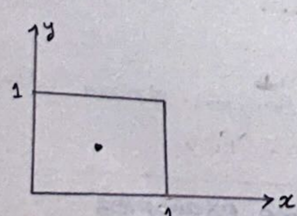
(x, y) such that $0 \leq x, y \leq 1$

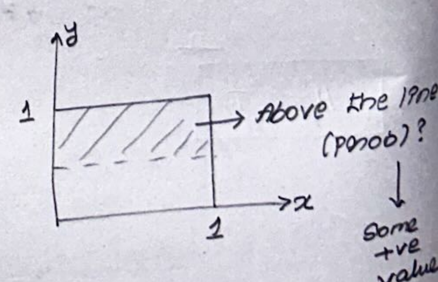
'Infinite elements in sample space'

Probability axioms

which are more likely to occur – Assign probability

Difficulties

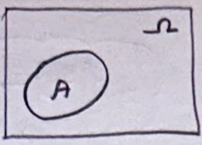




what's the probability of one individual point? (∞ elements)

$\frac{1}{\infty} = 0$ (Individual probability)

$\therefore$ we find (assign) probability for a subset.



$P(A)$

Event

* A subset of the sample space
- probability is assigned to events.

Rules

By convention: Range: $P(A) \longrightarrow 0 \le P(A) \le 1$

$P(A) = 0 \rightarrow$ won't happen

$P(A) = 1 \rightarrow$ certainly happen.

Axioms

① Non-negative axiom: $P(A) \ge 0$

② Normalization $P(\Omega) = 1$ [outcome is an element of Ω]

Absolute certainity

Intuitive (Feeling)

③ (Finite) Additivity

If, $A \cap B = \phi$, $P(A \cup B) = P(A) + P(B)$

[$A, B \rightarrow$ disjoint events]

$P(A \cap B) = 0$

$\phi \rightarrow$ Empty set.

Simple properties of probability

Counter points: (Intuitive)

$P(A) \ge 0 \longrightarrow P(A) \le 1$

$P(\Omega) = 1 \longrightarrow P(\phi) = 0$

$P(A) + P(\bar{A}) = 1$

↳ Doesn't happen

↳ happen

$A, B, C \rightarrow$ disjoint events

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

Disjoint events $(A, B, \dots)$

$A \cup A^c = \Omega$

$A \cap A^c = \phi$

$P(A \cup A^c) = 1$

$\boxed{P(A) + P(\bar{A}) = 1}$

$P(A) = 1 - P(\bar{A}) \quad \le 1$

$\therefore$ A and $\bar{A}$ are disjoint events

$P(\Omega) + P(\Omega^c) = 1$

$P(\Omega) = 1 - P(\Omega^c)$

$P(\Omega) = 1 - P(\phi)$

$P(\Omega) = 1$

Proof

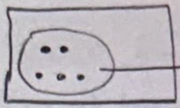
$P(A \cup B \cup C) = P(A) + P(B) + P(C)$

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

$= P(A) + P(B) + P(C)$

$\therefore$ Proved.

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^{n} P(A_i)$$



→ subset of individual elements

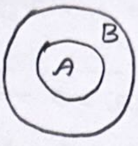
$\therefore \{s_1\} \rightarrow s_1$ is the only element.

$$P(\{s_1, s_2, \dots, s_k\}) = P(\{s_1\} \cup \{s_2\} \cup \dots \cup \{s_k\})$$

$$= P(\{s_1\}) + \dots + P(\{s_k\}) = P(s_1) + \dots + P(s_k)$$

more properties

* If $A \subset B$, then $P(A) \leq P(B)$



$\therefore$ A is the subset of B

$$\left.\begin{aligned} P(A \cup B) &= P(B) \\ P(A \cap B) &= P(A) \end{aligned}\right] \text{ A is the subset of B}$$

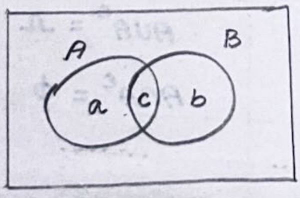
$$B = A \cup \underbrace{(B \cap A^c)}_{\text{outside } A}$$

$$P(B) = P(A) + P(B \cap A^c)$$

$$\boxed{\therefore \; P(A) + P(B \cap A^c) \geq P(A)}$$

Not disjoint (A, B)

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



$$a = P(A \cap B^c)$$
$$b = P(B \cap A^c)$$
$$c = P(A \cap B)$$

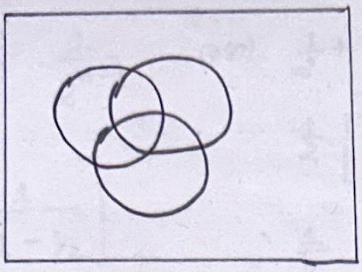
$$P(A \cup B) = P(A \cap B^c) + P(B \cap A^c) + P(A \cap B)$$
$$= a + b + c$$

$$\boxed{\begin{aligned} P(A \cup B) = P(A) + P(B) - P(A \cap B) &= (a + c) + (b + c) - c \\ &= a + b + c \end{aligned}}$$

consequence of axiom

$$\boxed{\therefore \; P(A \cup B) \leq P(A) + P(B)} \rightarrow \text{union bound}$$

$P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c + B^c \cap C)$



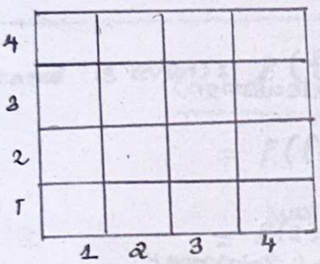
$A \cup B \cup C = A \cup (B \cap A^c) \cup (C \cap A^c \cap B^c)$

A: full A

$(B \cap A^c)$: B removed A (common)

$(C \cap A^c \cap B^c)$: C removed common pieces of A & B.

Discrete example - Fair die - Tetrahedral



y = II roll

X = I roll

$P(X=1) = 4\left(\frac{1}{16}\right) = \frac{1}{4}$ (1 in first roll)

$Z = \min(X, Y)$

$= \min(2, 3) = 2$

$P(Z=2) = \begin{pmatrix}\text{Both}\\ 2\end{pmatrix} + \begin{pmatrix}X=2\\ Y>2\end{pmatrix} + \begin{pmatrix}X>2\\ Y=2\end{pmatrix}$

$= \frac{1}{16} + 2\left(\frac{1}{16}\right) + 2\left(\frac{1}{16}\right)$

$= 5/16$

$P(Z=4) = \begin{pmatrix}\text{Both}\\ 4\end{pmatrix} + \begin{pmatrix}X=4\\ Y>4\end{pmatrix} + \begin{pmatrix}X>4\\ +\end{pmatrix}$

$\begin{pmatrix}X=4\\ Y>4\end{pmatrix}$: Not possible

$\begin{pmatrix}X>4\\ +\end{pmatrix}$: Not

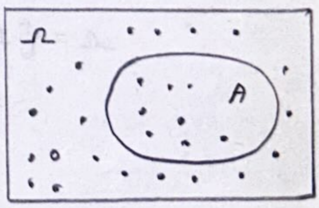
$= \frac{1}{16}$

Discrete uniform law

Assume Ω has n equally likely elements

Assume A has k elements

$$\boxed{P(A) = k\left(\frac{1}{n}\right) = \frac{k}{n}}$$

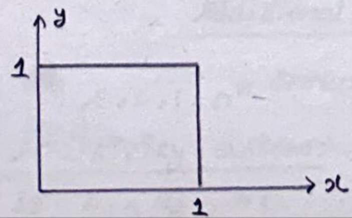


Prob $= \frac{1}{n}$

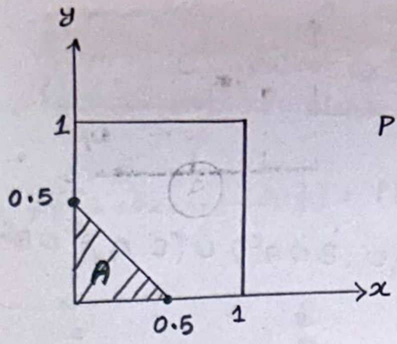
(For a particular element)

continuous uniform law

(x, y) show that $0 \le x, y \le 1$



* Assume uniform probability law (probability = Area)



$P(\{(x,y) \mid x+y \leq \frac{1}{2}\}) =$

$0 + 0.5 = 0.5$

$0.5 + 0 = 0.5$

$Area = \frac{1}{2} \times b \times h$

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

$P(A) = \frac{1}{8}$

ii) $P(\{(0.5, 0.3)\})$ = single point = Area = 0

Steps (Probability calculation)

* Specify sample space
* specify probability law
* Identify an event of interest
* Calculate..

(Pictures are immensely useful)

Idea: In continuous case: converting to unit area - what's the area of our event (subset)

---

countable additivity

Discrete & Infinite sample space:

$\Omega = \{1, 2, \dots\}$

'Head in n tosses' (For the 1st time)

1st time + 2nd time + ... + n time

P(Heads for the first time) = head in 1 + Tail in 1 Head in 2 + $T_1\ T_2\ H_3 + T_1\ T_2\ T_3\ H_4$

$+ \dots + T_1\ T_2 \dots T_{n-1}\ H_n$

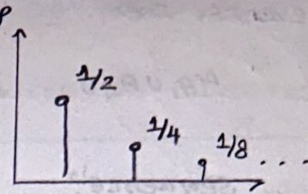
$= \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) + \dots + \left(\frac{1}{2}\right)^n$

$\boxed{\therefore P(\text{Head}) = \frac{1}{2} = P(\text{tail})}$

$n = 1, 2, 3, \dots$

check:

$$= \sum_{i=1}^{n} \frac{1}{2^i} \leq 1 \quad ?$$



$$= \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{2^{n-1}} \quad \text{(or)} \quad \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

$$\frac{1}{2}\left[1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots\right]$$

$$= \frac{1}{2}\left[\frac{1}{1 - 1/2}\right] \qquad \frac{1}{2}\sum_{i=0}^{\infty} \frac{1}{2^i}$$

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

$$\sum_{i=0}^{\infty} ar^k = \frac{a}{1-r}$$

---

p(outcome is even) $= p(\{2, 4, 6, 8, \ldots\})$

$= p(\{2\} \cup \{4\} \cup \cdots \cup \{\text{even numbers}\})$ → disjoint

$= P(2) + P(4) + \cdots$

'Getting Head in even attempts'

$$= T_1 H_2 + T_1 T_2 T_3 H_4 + \cdots + T_1 T_2 T_3 \cdots T_{n-1} H_n$$

n-even

$$= \frac{1}{2^2} + \frac{1}{2^4} + \cdots$$

$$= \frac{1}{4}\left(1 + \frac{1}{4} + \cdots\right)$$

$$= \frac{1}{4} \sum_{i=0}^{\infty} \left(\frac{1}{4}\right)^n$$

$$= \frac{1}{4}\left(\frac{1}{1 - \frac{1}{4}}\right)$$

$$= \frac{1}{4} \cdot \frac{1}{3/4}$$

$$= \frac{4}{4 \times 3} = 1/3$$

→ Is this correct?

(Is our additivity correct?)

$$P(\{2\} \cup \{4\} \cdots \cup \{\}) = P(2) + P(4) + \cdots$$

? (Is this allowed?)

<u>Additional axiom: Countable Additivity axiom</u>

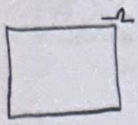
* Strengthens the finite additivity axiom.

<u>Countable Additivity axiom:</u>

If $A_1, A_2, A_3, \ldots$ is an infinite sequence of disjoint events, then

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots\dots$$

<u>why sequence?</u>



unit square

$$P(\Omega) = P(\text{Union}\{(x,y)\}) = \sum P(\{x,y\}) = \sum 0 = 0$$

unit square

↓ Union of all single element subsets

$$\boxed{\therefore P(\Omega) = 1} \qquad \therefore P(\Omega) = ? \sum P(\{(x,y)\})$$

∴ Additivity only valid for sequence.

* Additivity holds for countable sequence of elements
* The unit square (similarly the real line etc) is not countable (its elements can't be arranged in a sequence)

Infinite sets

↙ countable (Discrete)
* Integers

↘ uncountable
* unit square

(Confirming)

"Area is a legitimate probability law on the unit square, as long as we don't try to assign probabilities / areas to 'very strange' sets. [measure theory]

<u>Interpretation & uses of probabilities</u>

Narrow view: Branch of math

Axioms ⇒ Theorems.

Thm: 'Frequency' of event A is P(A)

* Are probabilities frequencies?
  * P(coin toss yields heads) = 1/2 → ∞ repitation can be done
  * P(President of ... will be reelected) = 0.7 → Election can't be done that many times

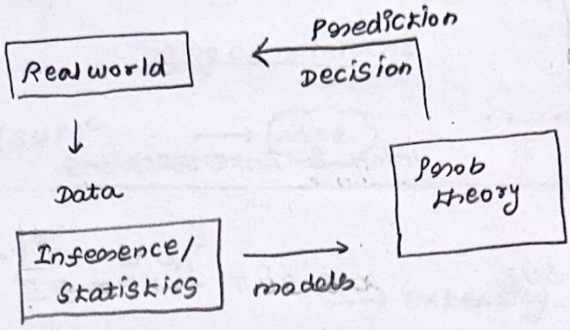
→ Description of beliefs (subjective)

→ Betting preferences

Role of probability theorem

A framework for analyzing phenomena with uncertain outcomes

- Rules for consistent reasoning
- Used for predictions & decisions.



Supplement : A mathematical Background

Sets

* collection of distinct elements
* $\{a, b, c, d\}$ → Finite

R = real numbers → Infinite.

*If x is an element of S $\quad x \in S$

x is not an element of S $\quad x \notin S$

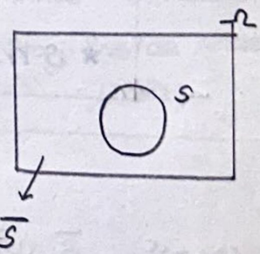
Specifying sets

$$\{x \in R \ : \cos(x) > 1/2\}$$

↳ condition

'elements of $x \in R$ satisfying condition alone'

Ω (collection of all possible outputs) = Universal set



$(\overline{\overline{S}}) = S$

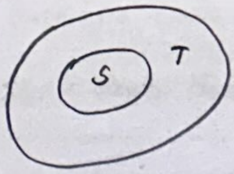
$\overline{S}$, $x \in \overline{S}$ if $x \in \Omega$, $x \notin S$

ϕ: Empty set (No elements)

$\Omega^c = \phi$

$\therefore$ Ω has all elements.

* commutative , Associative.

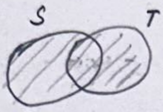


$S \subset T$ (subset)

$\boxed{\begin{matrix} x \in S \\ x \in T \end{matrix}}$

$\boxed{S \subseteq T \quad \text{(same as)} \quad S \subset T}$

Union & Intersections



$S \cup T \qquad x \in S \cup T \longleftrightarrow x \in S$ or $x \in T$

$S \cap T \qquad x \in S \cap T \longleftrightarrow x \in S$ and $x \in T$

"we can define union, intersection for infinite collection of sets"

$x \in \bigcup_n S_n$ iff $x \in S_n$, for some n

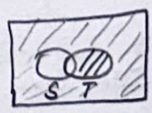
$x \in \bigcap_n S_n$ iff $x \in S_n$, for all n

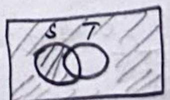
- $S \cup T = T \cup S$
- $S \cup (T \cup U) = (S \cup T) \cup U$
- $S \cap (T \cup U) = (S \cap T) \cup (S \cap U)$
- $(S^c)^c = S$
- $S \cup \Omega = \Omega$
- $S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$
- $S \cap S^c = \phi$
- $S \cap \Omega = S$

Demorgan's laws

* $(S \cap T)^c = \bar{S} \cup \bar{T}$

$x \in (S \cap T)^c \rightarrow x \notin S \cap T \rightarrow \begin{cases} x \notin S \rightarrow \\ \text{or} \\ x \notin T \rightarrow \end{cases}$





$\rightarrow \begin{cases} x \in S^c \\ \text{(or)} \\ x \in T^c \end{cases} \rightarrow x \in S^c \cup T^c$

I Law

Interchange   II Law

$S \to S^c$ $T \to T^c$

$S^c \to S$ $T^c \to T$

$(S^c \cap T^c)^c = S \cup T$

Taking complement

$S^c \cap T^c = (S \cup T)^c$ ⟶ Note.

$\left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c$

⟶ extending to n sets

$\left(\bigcup_n S_n\right)^c = \bigcap_n S_n^c$

Sequences & their limits

$a_1, a_2, a_3, \ldots$

$i \in N = \{1, 2, 3, \ldots\}$

sequence a_i, $\{a_i\}$

$\boxed{a_i \in S}$

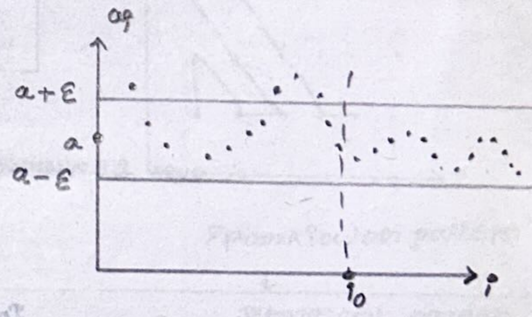
$S \to$ Real numbers (or) Euclidean space etc...

Formally: $f: N \to S$

$f(i) = a_i$ (gives i th element)

$a_i \to a$ (converges to some number a)

$$\boxed{\lim_{i \to \infty} a_i = a}$$



(After time it needs to stay inside that limit)

For any $\varepsilon > 0$,

After some time our function will be exists (i_0) – $i \geq i_0$ – our function stays b/w $a+\varepsilon$ & $a-\varepsilon$

then, $|a_i - a| < \varepsilon$

$a_i \to a$

$b_i \to b$

$a_i + b_i$ converges to $a+b$

$a_i \cdot b_i$ converges to ab

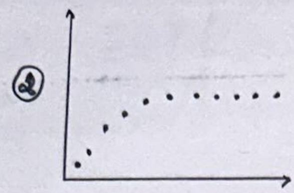
g: continuous

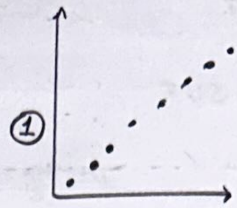
$g(a_i)$ converges to $g(a)$

when does a sequence converge? → Increase forever (Atleast don't go down) ↘ Decrease.

* If $a_i \leq a_{i+1}$, for all i, then either

① - the sequence 'converges' to ∞





(converges to ∞)

(settles to some value)

① → forever ↑

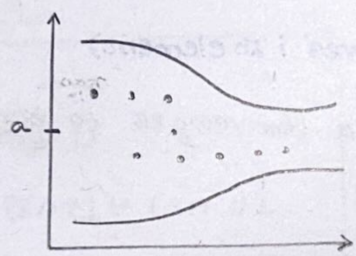
② → saturate.

② - The sequence converges to some real number a

* If $|a_i - a| \leq b_i$, for all i, $b_i \to 0$, then $a_i \to a$

Another way of establishing convergence is to derive some bound on the distance of our sequence from the number that we suspect to be the limit. If distance becomes smaller & smaller if we can manage to bound that distance by a certain number and that number goes down to zero.

sandwich argument:



→ 1st sequence converges to a

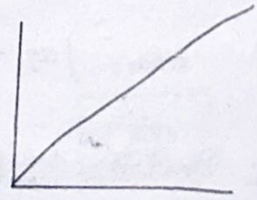
→ 2nd sequence converges to a

∴ our sequence will be in b/w the two sequences & converges to the same a

Infinite series

$$\boxed{\sum_{i=1}^{\infty} a_i = \lim_{n\to\infty} \sum_{i=1}^{n} a_i}$$ → valid as long as the limit exits.

* If $a_i \geq 0$: limit exits



* If terms a_i don't all have same sign

→ limit need not exists (series - not well defined)

→ limit may exist but be different if we sum in a different order.

monotonically ↑ 'Always exist'

Fact: limit exists but order of summation is independent if $\sum_{i=1}^{\infty} |a_i| < \infty$

(rearrange terms - different limit is obtained)

$\therefore$ $|a_i| \geq 0$ $\therefore$ By $a_i \geq 0 \to$ limit exists.

Geometric series

$$S = \sum_{i=0}^{\infty} \alpha^i = 1 + \alpha + \alpha^2 + \cdots \qquad |\alpha| < 1$$

$$(1-\alpha)(1+\alpha+\cdots+\alpha^n) = (1+\alpha+\cdots\alpha^n) - (\alpha+\alpha^2+\alpha^3+\cdots\alpha^{n+1})$$

$$(1-\alpha)\,S = 1-\alpha^{n+1}$$

$$S = \frac{1-\alpha^{n+1}}{1-\alpha} \qquad \text{As } n \to \infty$$

$$\boxed{S = \frac{1}{1-\alpha}}$$

Alternate method

$$S = 1 + \sum_{i=1}^{\infty} \alpha^i$$

$$= 1 + \alpha \sum_{i=0}^{\infty} \alpha^i$$

$$S = 1 + \alpha S$$

$$S(1-\alpha) = 1$$

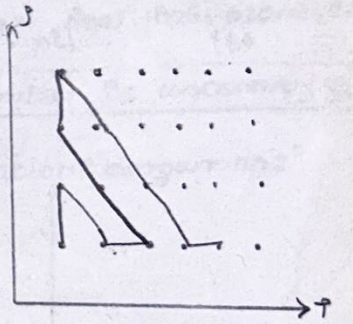
$$S = \frac{1}{1-\alpha}$$

$\therefore$ only possible $S < \infty$ (Finite)

$\therefore$ we are subtracting

order of summation in series with multiple indices

$$\sum_{i \geq 1, j \geq 1} a_{ij}$$



'particular pattern'

$\downarrow$

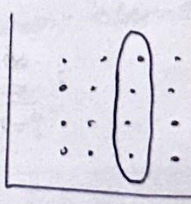
Different order gives different results.

As $\sum |a_{ij}| < \infty \to$ order doesn't matter.

or

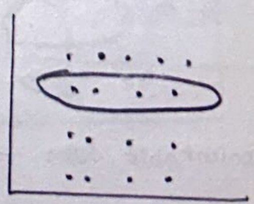
Fix i - change j

$$\sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} a_{ij} \right)$$



Fix j - change i

$$\sum_{j=1}^{\infty} \left(\sum_{i=1}^{\infty} a_{ij} \right)$$

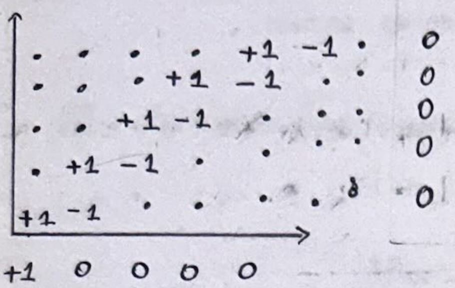


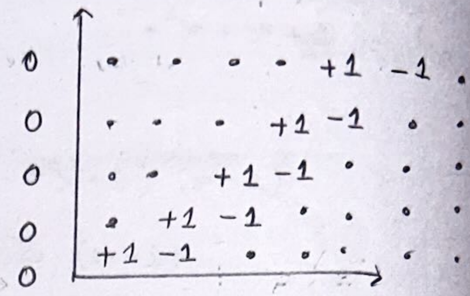
The results are going to be the same as long as

$$\boxed{\sum |a_{ij}| < \infty}$$

$$\sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} a_{ij}\right) = \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} a_{ij}\right)$$

$$1+0+0+\cdots = 1 \neq \sum 0 = 0$$





$1+0+0+\cdots = 1$ $\qquad$ $\sum = 0$

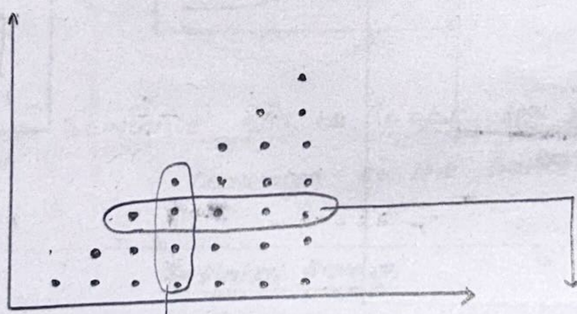
why?

order of summation may matter.

In this example sum of all absolute values is ∞ [∞ly many +1 & -1]

$\therefore \sum |a_{ij}|$ is not $< \infty$

$\sum_{(i,j):\, j \le i} a_{ij}$ → Limited range of indices ($j \le i$)



$a_{33} + a_{43} + a_{53} + \cdots$ (moving i starting from j)

oneway

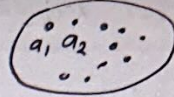
$$\sum_{i=1}^{\infty} \sum_{j=1}^{i} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} a_{ij}$$

(As long as $\sum_{(i,j):\, j \le i} |a_{ij}| < \infty$)

Countable & uncountable sets

countable set: can be put in 1-1 correspondence with positive integers.

- we are able to arrange all the elements of
Ω in a sequence.



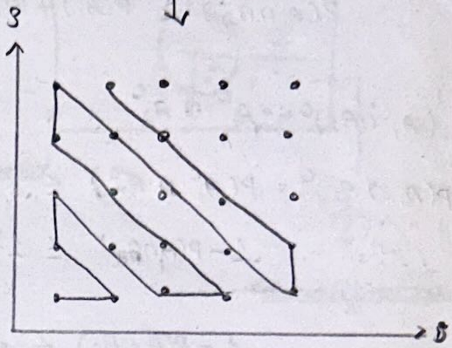
$\Omega = \{a_1, a_2, \dots\}$

* positive integers 1, 2, 3, ...
* Integers 0, 1, -1, 2, -2, 3, -3, ...
(Alternate order)
* pairs of positive integers
* Triples or quad " "
* Rational numbers r, with
$0 < r < 1$

$\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \dots$

we can leave $\frac{2}{4}$

'sequence'



'path' - we can cover all the elements in a sequence.

Uncountable

- The interval [0, 1]
- Real line, plane, space, ...

'unit interval is uncountable - Extended for Real line, plane, etc.'

Proof that a set of Real numbers is uncountable

The reals are uncountable : 'Cantor's diagonalization argument'

$\{x \in (0,1): \text{decimal expansion only } 3, 4\}$

Countable: $\{x_1, x_2, \dots, x_n\}$ → can't be written as a sequence.

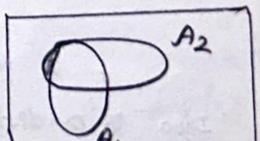
x_1: 0.343443...
x_2: 0.4443443...
x_3: 0.33434444..

$0.433\dots\dots = x$
$\neq x_i$

(so there will be a different element)

'contradiction - uncountable'

Bonferroni's inequality



Union bound:

* very few → smart students A_1
* very few → Beautiful A_2

very few students are students are smart or beautiful

$P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$

★ suppose:

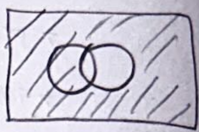
most — smart

most — Beautiful.

Then: most of the students are smart & beautiful.

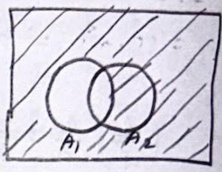
$$P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1$$

A_2^c



$$(A_1 \cap A_2)^c = A_1^c \cup A_2^c$$

A_1^c



$$P(A_1 \cap A_2)^c = P(A_1^c \cup A_2^c) \leq P(A_1^c) + P(A_2^c)$$

$$1 - P(A_1 \cap A_2) \leq 1 - P(A_1) + 1 - P(A_2)$$

$$1 - P(A_1 \cap A_2) \leq 2 - P(A_1) - P(A_2)$$

$P(A_1^c \cup A_2^c) = 1 - P(A \cap B)$

$$\boxed{P(A_1) + P(A_2) - 1 \leq P(A_1 \cap A_2)}$$

$$\boxed{P(A_1 \cap A_2 \cap \dots \cap A_n) \geq P(A_1) + \dots + P(A_n) - (n-1)}$$ → General form

∴ Big sets intersection will also be big.

Lecture: 2: Conditioning & Bayes rule

Conditional probability:

Took a person at random from the registry of residents

↓

Probability < 18 years (say: 0.25)

Now: we know the person chosen at random is married

★ still the probability is same? (No)

'modifying based on additional knowledge - making our belief to change'

∴ Revised probability - Conditional probability.

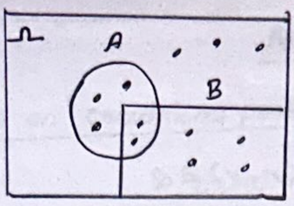
Bayes rule: Foundation of the field of inference.

(Guide how to process data) & make inference on data acquired.

Conditional Probability

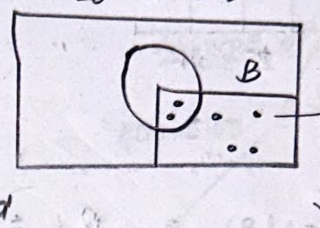
Idea of conditioning : use of new info to revise a model.

$P(A) = \frac{5}{12}$, $P(B) = \frac{6}{12}$



→ Equally occurring 1/12.

If B has occurred.



$\frac{1}{6}$

'Equally occurs'

'someone: Tells Event B has occurred'

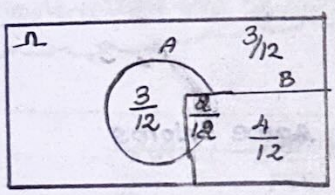
Zero prob: outside B

P(A|B) = conditional prob of A gn B has occurred.

$P(A|B) = \frac{2}{6}$, $P(B|B) = \frac{6}{6} = 1$

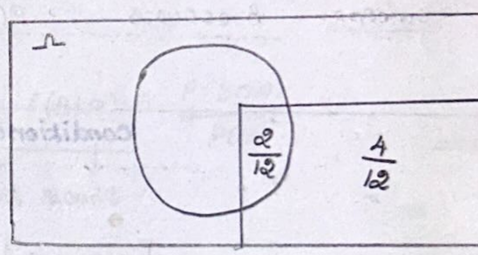
(Two points) in A∩B

'If not equally occurred'



'collective probability'

Event B has occurred



P(A|B) = Prob of A gn that B has occurred

'we know B has occurred – A has 1/3 % of occurring in B'

[A has two dots
B has 6 dots]

∴ $P(A|B) = \frac{1}{3}$, $P(B|B) = 1$ [B is certain to occur]

Formula:

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

$$= \frac{2/12}{6/12}$$

$$= 2/6$$

$$P(A|B) = 1/3$$

∴ B → has happened

what prob A happened in B

↓

A is a part of B (part)

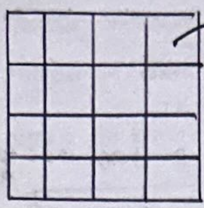
∴ $P(B) > 0$

If $P(B) = 0$, $P(A|B)$ remain undefined.

∴ P(A|B) → Not a theory – A definition.

A die roll example

Two rolls of 4-sided die:



$1/16$ (Equal probability)

I roll

II roll

* B – event $\min(x,y) = 2$

$(2,3), (2,4), (3,2), (4,2), (2,2)$

* M – $\max(x,y)$ $\quad (1,1)$ alone.

$$P(M=1 \mid B) = \frac{0}{5/16} = 0$$

maximum = 1

$M = 3$ $\{(1,3), \underline{(2,3)}, (3,3), (3,1), \underline{(3,2)}\}$

$$P(M=3 \mid B) = \frac{\frac{2}{16}}{5/16} = 2/5 = \frac{P(M=3 \text{ and } B)}{P(B)}$$

$$= 0.4$$

Concept: B occurs $\therefore$ $P(A|B) = \frac{2}{5} \rightarrow$ A has 2 shares out of 5 shares of B.

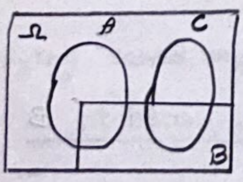
Conditional probability obey the same axioms

$\downarrow$

Share the properties of ordinary probabilities

* $P(A|B) \geq 0$ $\quad [P(B) > 0]$
* $P(\Omega|B) = 1$, $\frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$ [sample space has prob = 1 even under conditional]
* $P(B|B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

Additivity axiom



$A \cap C = \phi$, then $P(A \cup C \mid B) = P(A|B) + P(C|B)$

$\therefore$ $A \cap B$, $B \cap C$ are disjoint

Proof:

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

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

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

'Also true for finitely many - countably many events'

'conditional probability satisfies all the std. probability axioms'

Radar example

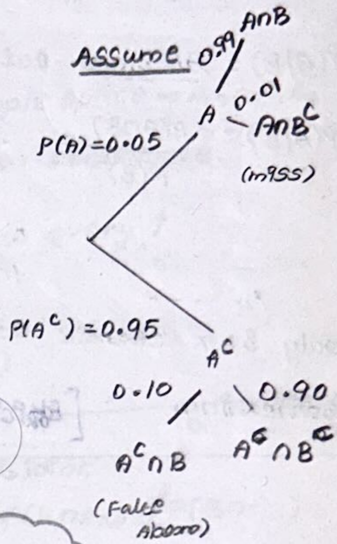
models based on conditional probabilities:

'Detection of an object by the radar'

A: Airplane is flying above.

B: Something registers on radar screen

Assume



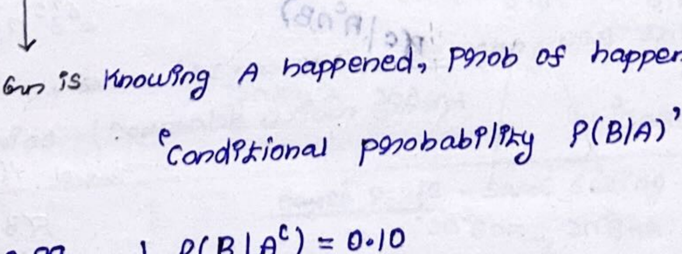
$P(A \cap B) \rightarrow$ Airplane is there & gets detected

$P(A \cap \bar{B}) \rightarrow$ Airplane flying but not detected

So Actually it is a conditional prob $\left\{ \begin{array}{l} P(B|A) = 0.99 \\ P(\bar{B}\,|\,A) = 0.01 \end{array} \right.$

$P(B|A) \rightarrow$ Detected by radar already plane is up there.

$P(A|B) = \frac{P(A \cap B)}{P(B)} \rightarrow$ conditioning event , $P(B|A) = \frac{P(B \cap A)}{P(A)}$



Gn is knowing A happened, Prob of happening B

'Conditional probability $P(B|A)$'

$P(B|A) = 0.99$ | $P(B|A^c) = 0.10$

$P(\bar{B}\,|\,A) = 0.01$ | $P(\bar{B}\,|\,\bar{A}) = 0.90$.

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

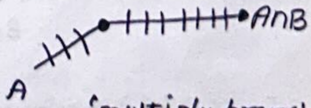
Probability of Both A and B occur:

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

(multiplying branches)

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

↳ we don't have $P(B|A)$



'multiply branches'

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

$= (0.05)(0.99) = 0.0495.$

$P(B) =$ Radar sees the plane $= (A \cap B) + (A^c \cap B)$

$= (0.05) \cdot (0.99) + (0.95)(0.10)$

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

$= 0.0495 + 0.095$

$= 0.1445$

$P(A|B) =$ Airplane out there, radar detects it

$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.0495}{0.1445}$

$= 0.3425 \rightarrow$ Not really convenient.

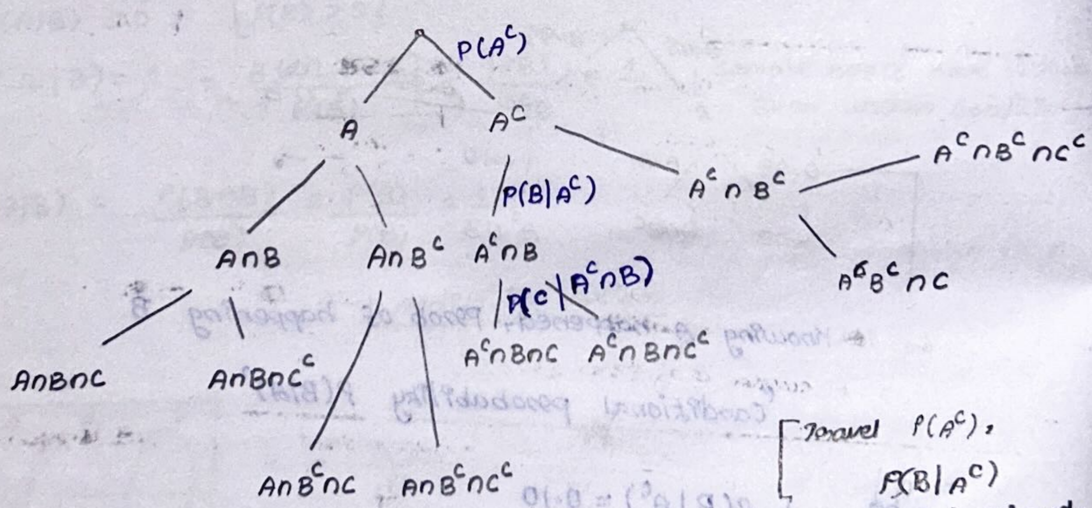
only 34% chance of aeroplane flying when the radar detects something. [Efficiency can be calculated]

Multiplication rule

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

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

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



[Travel $P(A^c)$,
$P(B|A^c)$ ↳ Already occ
$P(c^c|A^c \cap B)$]

$P(A^c \cap B \cap C^c) = P(A^c \cap B) \cdot P(C^c | A^c \cap B)$

$= P(A^c) \cdot P(B|A^c) \cdot P(C^c | A^c \cap B)$

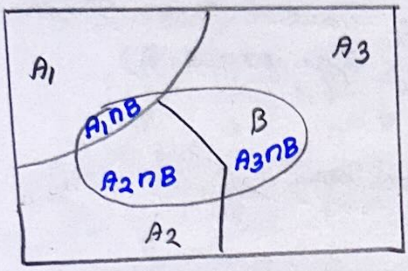
↓ Branch 1 ↓ Branch 2 ↓ Branch 3

we can find the probabilities of a particular leave - by multiplying the conditional probabilities of the branches.

$$P(A_1 \cap A_2 \cap A_3 \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \ldots \ldots P(A_n | A_1 \cap A_2 \cap A_3 \ldots A_{n-1})$$

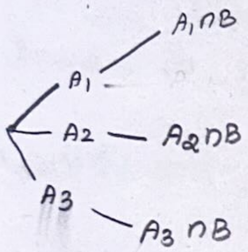
(General version)

Total probability theorem



* Partition of sample space → A_1, A_2, A_3
* Divide & conquer technique.
* Have $P(A_i)$ for every i

(A_1, A_2, A_3 — subspaces - partitions)



Disjoint

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)$$
$$= P(A_1) \cdot P(B|A_1) + P(A_2) P(B|A_2) + P(A_3) P(B|A_3)$$

$\sum P(A_i) = 1$ (All partitions)

$$\boxed{P(B) = \sum P(A_i) P(B|A_i)}$$

weighted average. of $P(B|A_i)$

The probability that an event occurs is the weighted average of the prob it has under each partitions. (weights - prob of B in each scenarios)

If $A_1, A_2, \ldots$, is a sequence — we can use this for infinite set scenarios (Countable axiom sum)

Bayes rule – same setting as total prob. theorem

* $P(A_i)$ → Initial beliefs [how likely they are]
* Have $P(B|A_i)$ for every i

'Event B - indeed occurs' – 'B going to occur for sure'

'So this may cause us to change our belief on $P(A_i)$'

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)\, P(B|A_i)}{P(B)} = \frac{P(A_i)\, P(B|A_i)}{\sum_{j=1} P(A_j)\, P(B|A_j)}$$

(Total prob theorem)

→ Thomas Bayes, presbyterian minister (c. 1701 - 1761)

→ Bayes theorem - published posthumously (After dead)

→ systematic approach for incorporating new evidence (learn from experience)

↓

Branch of math: Bayesian Inference.

Bayesian inference:

* Initial beliefs $P(A_i)$ on possible causes of an observed event B

* model of the world under each A_i $(P(B|A_i))$

$A_i \xrightarrow[P(B|A_i)]{\text{model}} B$ [how actually B occurs]

* Draw conclusions about causes

$B \xrightarrow[P(A_i/B)]{\text{inference}} A_i$ [causes of B] using B occurs in a particular scenario

↓

affects A_i

'we can view conditional prob going in other direction' through bayes.

↓

* how A_i's causing B

* we need: how A_i's happening on B.

After exp: B indeed occurs, so revise our beliefs on A_i s.

Lecture: 3: Independence

Independence of two events:

* A - occurred, A will generally change the occurance of B so we will use conditional probability.

* If conditional & undconditional prob are same - the occurance of A doesn't have any useful info on the occurance of B.

∴ such case: A & B are independent

↓

If any two events in a collection are independent 'pairwise independence'

↓

Independence of collection - different - involves more properties.

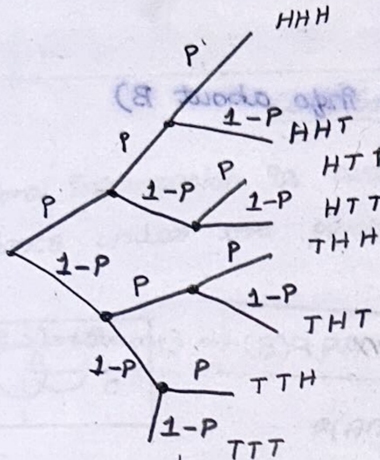
Coin toss

Total prob: Event B – weighted sum of all events

Bayes : Altering all events – B surely going to occur.

A model on conditional probabilities

* 3 tosses of a biased coin : $P(H) = P$, $P(T) = 1-P$



$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

$P(T \mid TT)$

$\Downarrow$

Two tails in 1 & 2 event.

mul rule: $P(THT) = P(T_1) \cdot P(H_2 \mid T_1) \cdot P(T_3 \mid T_1 H_2)$

$= (1-P) \cdot (P) \cdot (1-P)$

Total probability:

$P(1 \text{ head}) = P(HTT) + P(THT) + P(TTH)$

$= P(1-P)(1-P) + (1-P)P(1-P) + (1-P)(1-P)P$

$= 3P(1-P)^2$ [Equally likely]

Bayes rule:

$P(\text{First toss is H} \mid 1 \text{ head}) = \dfrac{P(\text{First head} \cap 1 \text{ head})}{P(1 \text{ head})}$

$\therefore P(H_1 T_2 T_3)$ happens in $\{P(HTT), P(THT), P(TTH)\}$

$= \dfrac{P(1-P)^2}{3P(1-P)^2} = \dfrac{1}{3}$

comment: $P(H_2 \mid H_1) = P = P(H_2 \mid T_1)$ → In this example

Result in 1 doesn't change the 2nd result.

$P(H_2) = P(H_1)\,P(H_2 | H_1) + P(T_1)\,P(H_2 | T_1)$

$$\boxed{P(H_2) = P(T_1 \cap H_2) + P(H_1 \cap H_2)}$$

Independence of two events

Previous: Result of 1 doesn't affect 2.

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

(occurrence of A provides of no new info about B)

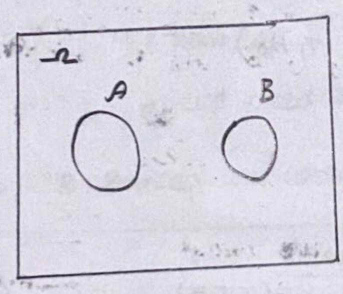
* $P(A \cap B) = P(A) \cdot P(B|A)$
  $= P(A) \cdot P(B)$

$\therefore$ Definition of independence: $\boxed{P(A \cap B) = P(A) \cdot P(B)}$ → symmetric

→ symmetric w.r.t A and B

→ Implies $P(A|B) = P(A)$

→ Applies even if $P(A) = 0$ [In conditional $P(B|A)$ $P(A) \neq 0$]



$P(A \cap B) = 0$ [disjoint]

$\therefore P(A) \cdot P(B) > 0$

$P(A) > 0$ $P(B) > 0$

'Disjoint'

'A and B are not independent'. If A occurs we can say B doesn't.

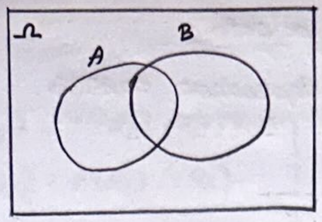
'Independence & disjoint are different'

↓

Two events A & B determined by two physically distinct & non-interacting processes.

Independence of event Compliments

If A and B are independent, then A and B^C are also independent.



$$A = (A \cap B) \cup (A \cap B^c)$$

$$P(A) = P(A \cap B) \cup P(A \cap B^c)$$

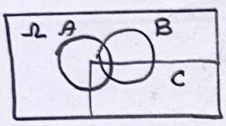
$$P(A) = P(A)\,P(B) + P(A \cap B^c)$$

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

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

Conditional Independence

→ Additional information is available, given C, is defined as independence under the probability law $P(\cdot \mid C)$

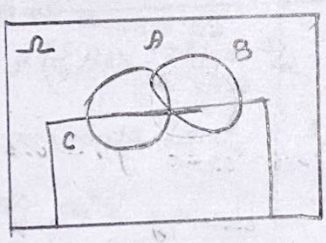


$$P(A \cap B) = P(A)\,P(B)$$

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

$$= P(A \mid C) \cdot P(B \mid C)$$

Assume A & B are independent:



If we are told C occurred, are A and B independent?

↓

No!

∴ If C already occurs, if A occurs - we know B doesn't occur.

A & B are not independent in the new world - C introduced

'Conditioning may affect independence'

In normal case: A & B are independent. But when we introduced conditional probability - It's no longer independent.

> 'Independence & unconditional independence are different'

example

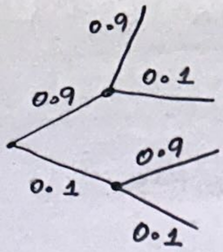
Two unfair coins, A and B

$$P(H \mid \text{coin } A) = 0.9, \quad P(H \mid \text{coin } B) = 0.1$$

Given a coin:

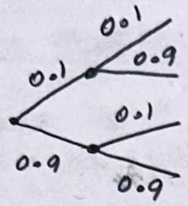
Independent tosses.

Coin A



→ universe 1

Coin B



→ universe 2.

coin is chose with 0.5 probability

★ Are coin tosses are independent?

unconditional	conditional
$P(toss\ 11 = H) = P(A \cap H) + P(B \cap H)$	$P(toss = 11 = H \mid first\ 10\ heads)$
$= P(A)\, P(H_{11} \mid A) + P(B)\, P(H_{11} \mid B)$	'10 heads in a row is almost unlikely'
A chosen heads + B chosed head.	'Info gn - we are most likely dealing with coin A'
$= (0.5)(0.9) + (0.5)(0.1)$	∴ we can't achieve with B.
$= 0.5$	$\approx P(H_{11} \mid A) = 0.9$

'Different' - Affects the belief.

---

we can't have independence b/w different events.

Independence of a collection of events

Information on some of the events doesn't change prob related to the relevant events.

$$P(A_3 \cap A_4^c) = P(\underbrace{A_3 \cap A_4^c} \mid \underbrace{A_1 \cup (A_2 \cap A_5)})$$

Note: Indices are different.

Events $A_1, A_2, \dots A_n$ are independent if: $P(A_i \cap A_j \cap \dots \cap A_m) = P(A_i)\, P(A_j) \dots P(A_m)$ for any distinct indices

$i, j \ldots m$

<u>$n=3$</u>

$$P(A_1 \cap A_2) = P(A_1)\,P(A_2)$$
$$P(A_1 \cap A_3) = P(A_1)\,P(A_3)$$
$$P(A_2 \cap A_3) = P(A_2)\,P(A_3)$$

pairwise Independent.

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$$

we can say,

$$P(A_3) = P(A_3 \mid A_1 \cap A_2) = P(A_3 \mid A_1 \cap A_2^c) = P(A_3 \mid A_1^c \cap A_2)$$

$\therefore A_1, A_2, A_3$ are Independent.

Independence vs pairwise Independence

→ Two independent fair coin tosses

H_1 : 1 toss H

H_2 : 2 toss H

$P(H_1) = P(H_2) = 1/2$

H_1	AH 1/4	HT 1/4
	TH 1/4	TT 1/4

$A_2 = 1/4 + 1/4 = 1/2$

C : Two tosses had the same result :

$$P(\{HH, TT\}) = 1/2$$

$P(H_1 \cap C)$ = Head corresponds head = 1/4

$$P(H_1 \cap C) = P(H_1) \cdot P(C)$$
$$= 1/2 \cdot 1/2$$
$$= 1/4$$

$\therefore H_1$ and C are Independent

$|||^{ly}$ H_2 and C are independent

$|||^{ly}$ H_1 and H_2 are independent

$P(H_1 \cap H_2 \cap C)$ = Heads in 1, Heads in 2 and the two tosses are same

$= 1/4$

Different.

$$P(H_1 \cap H_2 \cap C) = \frac{2}{4} \times \frac{2}{4} \times \frac{2}{4} = \frac{1}{8}$$

H_1, H_2 and C are pairwise independent but collectively not.

---

why?

$P(C|H_1)$ = only happens when HH = 1/4 (Tells about C)

$P(C \mid H_1) = P(H_2 \mid H_1) = P(H_2) = 1/2 = P(C)$

$\therefore$ In $P(C \mid H_1)$: C may be H_2 or T_2 (tail or heads)

In $P(C \mid H_1 \cap H_2)$ = C occurs and (H_1 and H_2 already occured)

= sure C

$= 1$

$H_1 \cap H_2 \rightarrow$ carry useful info on C.

"But in $P(C \mid H_1) \rightarrow$ still know H_1, C has $1/2$ chance of occuring"

Reliability

Independence: whenever we know: we can do simpler calculations.

Applications

n number of units (up or down)

[Independent]

P_i : Prob of being "up"

"Failure of some: doesn't affect others"

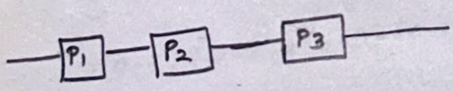
U_i : ith unit up

$u_1, u_2, \ldots u_n$ independent

F_i : ith unit down

$F_1, F_2, \ldots F_n$ independent

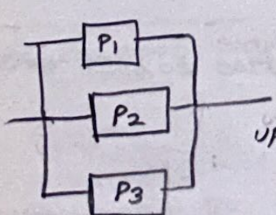
consequence



$P(\text{system up}) = P(u_1 \cap u_2 \cap u_3)$

$= P(u_1) \cdot P(u_2) \cdot P(u_3)$

$= P_1 \, P_2 \, P_3$



up - atleast one is up

$P(\text{system}) = P(u_1 \cup u_2 \cup u_3)$

[consequence]

(or)

All units fail

$= 1 - P(F_1 \cap F_2 \cap F_3)$

$= 1 - P(F_1) P(F_2) P(F_3)$

$= 1 - (1 - P_1)(1 - P_2)(1 - P_3)$

$= 1 - \overline{P_1} \, \overline{P_2} \, \overline{P_3}$

King's sibling

King comes from a family of two children. prob that his sibling is female.

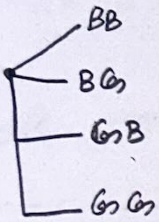
Soln:

King → men [Boys have precedence] → Even the girl is born first.

$P(boy) = P(girl) = 50\% = 1/2$ → Birth.

'Independent'

'King's gender is independent of sibling' → Either boy → girl.



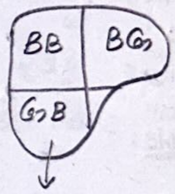
1/4	1/4
BB	BG
GB	GG
1/4	1/4

→ Not occurs.

1) There is a King → Atleast one boy

2) $P(Girl \mid King) = \frac{P(Girl \cap King)}{P(King)}$

$= \frac{2/4}{3/4} = 2/3$



Each of one has cond. prob 1/3

From fig $P(girl \mid King) = \frac{2}{3}$

<u>making more assumptions</u>

* Royal family: exactly 2 children.
* Atleast one boy ↓
* In that case girl sibling = 2/3

<u>Alternate assumption</u>

* Royal family: continue until a boy.

* 2 children: First girl then boy. , 2 boy (nd)

$P(G) = 1$ (sure)

<u>Alternate</u>

children until 2 boys. [2 children]

Exactly 2 boys: $P(G) = 0$

2/3 is the answer until our assumption: 2 children

2 children was a girl. (predetermined)

'Careful with assumptions'

Lec-4 - Counting

Discrete uniform law:

→ Assume Ω has n equally likely element

→ A has k elements

$P(A) = \frac{k}{n}$

Basic counting principle:

* 20 players available: 5 for starting, 7 - bench (5+7 - choosing)

App: permutation, combination, partitions, number of subsets, binomial probabilities.

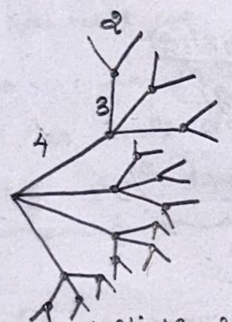
Counting principle

4 shirts

3 ties

2 jackets

Possible:



4 shirts, 3 ties, 2 jackets

'A combination is available'

12 combinations for 4 shirts, 3 ties.

24 comb 4 S, 3 t, 2 Jackets.

Generalize

* r stages

* n_i choices at i

$r = 3$

$n_1 = 4, n_2 = 3, n_3 = 2$

Irrespective of previous choice: the offered choices for a particular choice is the same.

examples:

No. of license plates with 2 letters followed by 3 digits.

$26 \times 26 \times n \times n \times n$ choices → repitation allowed

No repitation:

$26 \times 25 \times n \times (n-1) \times (n-2)$ choices