Feb 17,25

: **Probability** : (as classification / identify unusual pattern /
uncertain at every trail

- Experiment / Random Experienment
  - more than one output
  - Repititions under same condition.
- Trial
- Outcome (result of experiment)
- Sample space (complete list of outcomes)
- Events (all subsets of sample space) n(A)
  - in which we intrested.
- → Counting Principles
- Laws.

Loop
- motite....

Single events
- one single outcome

Compound
- more than one outputs
  - with their own prob.

A~, B~ can be independent, Dependent, Mutually ~~exclusivity~~ exclusive, non-mutual ex
can be at the same time

- M·Excl : can't occur at the same at once / time
  ↳
- NonMutual : same ✓ can occur together.
  - Dicing two Dices.
- ~~D~~ Independent :
  Don't effect.
- Dependent

Types of Events	Can Occur together	Affects Probability
1. Independent	✓	✗
2. Dependent	✓	✓
3. M.Ex	✗	✗
4. N.M.Ex	✓	Depends

To calculate probability

→ Classical $1/n$

→ R.F approach → varies with each entry.

→ Subjective.

↳ personal judgement (within 100%)

$$P(A) = \frac{n(A)}{n(S)} = \frac{\text{no. of favourable outcomes}}{\text{total sample space.}}$$

$n(A)$ → event space

sum of probability = 1

To check whether the probability is valid or not?

if Discrete : in b/w (stairs)

$$0 < p(A_i) < 1$$

impossible $=0$ ↲ ↳ sure

random variable :

if Contineous : sum of all events = 1

$$\Sigma P(A_i) = 1 \qquad f(x) > 0$$

$$\int f(x) = 1$$

(slope) ↲

$$P(A) = \frac{\text{outcomes}}{\text{total outcomes}}.$$

For counting $n(S)$:

(no-selection of objects)

- Rule of multiplication → throwing a coin & a dice at a same time
- Permutations } for selection → total → selection → arrangement
- Combination. ↳ no order imp.

Laws:

→ Law of Complementation

$P(S) = 1 - P(F)$ / $P(A) = 1 - P(A^c)$

→ Law of Addition:

OR $P(A \cup B) = P(A) + P(B) - P(A \overset{\text{and}}{\cap} B)$ non-mutually

$P(A \cup B) = P(A) + P(B)$ ← mutually exclusive

→ Law of Multiplication: (A, B) are independent

AND $P(A \cap B) = P(A) . P(B)$ $\quad P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$

↳ condition:

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

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

if the events are

non independent: $P(A|B) \neq P(A)$

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

$$P(F|M_1 \cap M_2) = \frac{P(F \cap M_1 \cap M_2)}{P(M_1 \cap M_2)}$$

for mutually exclusive: $P(A \text{ and } B) = 0$ $\quad P(A \cap B) = 0$

1. $3 \times 2 \times 4 = 24$

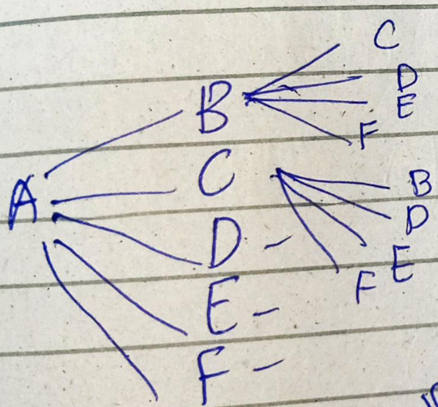
2. ${}^6C_3 = \frac{6!}{3!(6-3)!}$

$= \frac{6!}{3!\,3!}$

3. $P_6 = {}^6P_3 = \frac{n!}{(n-r)!}$

4. 2^3

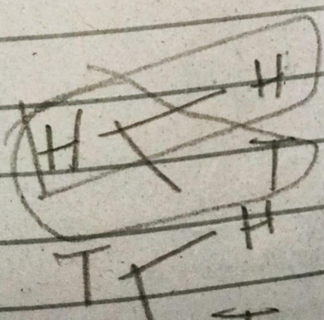
$\{(H,H,H), (T,T,T)\ (H,H,T)\ (H,T,H)\ (H,$



$$\begin{bmatrix} H & H & H \\ H & H & T \\ H & T & H \\ H & T & T \\ T & H & H \\ T & H & T \\ T & T & H \\ T & T & T \end{bmatrix}$$

$\boxed{ABC}$

~~CBA~~



A B C D E F

A B C	B C D
A B D	B C E
A B E	B C F
A B F	B D E
A C D	B D F
A C E	B E F
A C F	C D E
A D F	C D F
A D F	C E F
A E F	D E F

Poss

Design	Const	Total Time.
2	6	8
	7	9
	8	10
3	6	9
	7	10
	8	11
4	6	10
	7	11
	8	12

February 26, 2025 Wednesday

$$\underbrace{P(A_i|B)}_{\text{Posterior}} = \frac{\overbrace{P(A_i)}^{\text{prior}}\,\overbrace{P(B|A_i)}^{\text{conditional}}}{P(A)\,P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

Baye's Theorem

$$= \frac{P(A_i \cap B)}{\underbrace{\sum_{i=1}^{n} P(A_i \cap B)}_{\text{total probability}}}$$

$$= \frac{P(A_i)\,P(B|A_i)}{\sum_{i=1}^{n} P(A_i \cap B)} = \frac{P(A_i \cap B)}{P(B)}$$

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

Suppose that parts from two suppliers are used in the manufacturing process and that a machine breaksout bcz it attempts to process a bad part. Given the information is that the part is bad what is the probability that it came from supplier one, and what is the probability that it came from supplier 2.

$A_1 \sim S_1$

$A_2 \sim S_2$

$G \sim$ Good

$B \sim$ Bad

Step I (Supplier)

Step II (conditional)

$P(A_1)$ 0.65

$P(G|A_1) = 0.98$

$P(B|A_1) = 0.02$

1

$P(A_2)$ 0.35

$P(G|A_2) = 0.95$

$P(B|A_2) = 0.05$

1

1

$$P(A_1|B) = \frac{P(A_1)\,P(B|A_1)}{P(A)P(B|A_1) + P(A_2)\,P(B|A_2)}$$

$$P(A_2|B) = \frac{P(A_2)\,P(B|A_2)}{P(A)\,P(B|A_1) + P(A_2)P(B|A_2)}$$

Events A_i	Prior Prob. $P(A_i)$	Cond. Prob $P(B\|A_i)$	Joint Prob. $P(A_i \cap B)$
A_1	0.65	0.02	$0.65 \times 0.02 = 0.0130$
A_2	0.35	0.05	$0.35 \times 0.05 = 0.0175$
			sum : 0.0305

Posterior. Prob $P(A_i|B)$

$0.0130 / 0.0305$
$= 0.4262$

$0.0175 / 0.0305$
$= 0.5738$

Q:

A manufacturing firm employees 3 analytical plans for the design and development of a particular product for cost reasons, all three are used at varying time in fact plans 1, 2 and 3 are used for 30%, 20% and 50% of the products respect

The defect rate is different for the three procedures as follows

$P(D \mid P_1) = 0.01$
$P(D \mid P_2) = 0.03$
$P(D \mid P_3) = 0.02$

if a random product was observed and found to be defective, which plan was most likely used and thus responsible.

Events	Prior Prob.	Cond.	Joint	Posterior
P_1	0.30	0.01	$0.30 \times 0.01 = 0.003$	$0.003/0.019 = 0.1579$
P_2	0.20	0.03	0.20×0.03 0.006	$= 0.3158$
P_3	0.50	0.02	0.50×0.02 0.01	$= 0.5263$

sum = 0.019

2nd Ramzan

Baye's For Spam Filter

$$P(S|W) = \frac{P(W|S)\,P(S)}{P(W|S)P(S) + P(W|H)P(H)}$$

$P(S|W)$ is the prob that a msg is spam, knowing that "word" in it.

$P(S)$ is overall prob that any given msg is spam.

$P(H)$ is " " " is not spam (Ham)

$P(W|S)$ " " word appears in spam msgs.

$P(W|H)$ - - - - - Ham msgs.

For Equally Likely Events

$$P(S|W) = \frac{P(W|S) \times 0.5}{P(W|S)\,0.5 + P(W|H)\,0.5}$$

$$= \frac{P(W|S)}{P(W|S) + P(W|H)}$$

$$= \frac{\overset{(\text{success})}{P(\text{word})}}{P(\text{word}) + q(\text{word})}$$

↳ failure.

case :

When you recieve an email, there are two hypothesis.

H_1 : Email is Spam
H_2 : Email is not spam.

Observed Data :

D = email contains the word "CASH"

Objective :

Pick MAP hypothesis.
MAP : Maximum Posterior Prob / hypo

Suppose we have analyzed 1000 previous emails and found.

→ 700 emails were spam
→ 300 emails were ham.

Therefore, 70% of emails were spam & 30% of emails were not spam.

$$P(H_1) = 0.70$$
$$P(H_2) = 0.30$$

Furthermore, we also found that
350 of spam emails include the word "CASH"
100 of ham emails include the word "CASH"

Likelihoods:

$$P(D|H_1) = \frac{350}{100} = 0.5$$

$$P(D|H_2) = \frac{100}{300} = 0.33$$

By Baye's Theorem: ($A \cap B = A \text{ And } B$)

$$P(H_i|D) = \frac{P(H_i \cap D)}{P(D)}$$

$$= \frac{P(D|H_i)\,P(H_i)}{P(D|H_1)\,P(H_1) + P(D|H_2)\,P(H_2)}$$

$$P(H_1|D) = \frac{P(D|H_1)\,(P(H_1))}{P(D|H_1)(P(H_1)) + P(D|H_2)\,P(H_2)}$$

$$= \frac{0.5 \times 0.70}{(0.5)(0.70) + (0.33)(0.30)}$$

$$\boxed{P(H_1|D) = 0.78}$$

$$P(H_2|D) = \frac{P(D|H_2)\,(P(H_2))}{P(D|H_1)\,P(H_1) + P(D|H_2)\,P(H_2)}$$

$$= \frac{0.33 \times 0.30}{(0.5)(0.70) + (0.33)(0.30)}$$

$$\boxed{P(H_2|D) = 0.22}$$

Ques:

Suppose that we have found, the word 'Rolex' occurs in 250 of 2000 messages known to be spammed and 5 of 1000 messages known to be hamed. Estimate the probability that an incoming msg containing the word Rolex is spam. Assuming that it is equally likely that an incoming msg is spam or ham, if our threshold of rejecting a msg as spam is 0.9 will be reject that msg.

$$P(D|H1) = \frac{25\cancel{0}}{200\cancel{0}} = 0.125$$

$$P(D|H2) = \frac{5}{1000} = 0.005$$

$$r(\text{rolex}) = \frac{P(\text{rolex})}{P(\text{Rolex}) + q(\text{Rolex})}$$

$$= \frac{0.125}{0.125 + 0.005} = 0.962$$

will Reject

Two Word Problem

4 Ramzan

Ques:

Suppose that we train a baysean spam filter on a set of $\boxed{2000}$ spam messages and 1000 messages that are not spam. The word stock appears in 400 spam messages and 60 messages that are not spam And the word "Undervalued" appears in 200 spam messages & 25 ham messages.

Estimate the probability that an incoming message containting both the words "Stock" and "Undervalued" is spam. Assuming that we have no prior knowledge about whether it is spam, will we reject these messages as spam, when we set the threshold at 0.9?

$$r(w_1, w_2) = \frac{P(word_1)\,P(word_2)}{P(word_1)\,p(word_2) + q(word_1)\,q(w_2)}$$

$$P(S|W_{12}) = \frac{P(W_1|S)\ P(W_2|S)\ P(S)}{P(W_1|S)\ P(W_2|S)\ P(S) + P(W_1|H)\ P(W_2|H)\ P(H)}$$

$$P(H)|W_{12} = \frac{P(W_1|H)\,P(W_2|H)\,P(H)}{\quad\text{''}\quad}$$

Equally likely:

$$= \frac{(0.2)(0.1)}{(0.2)(0.1) + (0.06)(0.025)}$$

$$= 0.93$$

✱ Question:

Suppose we do spam filtering using an information

D_1 = Email includes the word Western,

D_2 = Email includes the word 'Union'

S → email is spam

H → email is ham

Suppose 10% email is spam:

Spam

20% includes 'western', 'union', "western union"

Non Spam

5% includes 'western'

5% " , union

4% " 'western union'

Task:

The email including 'western union' is spam or ham,

'Union' → spam/ham

Western → spam/ham.

(Western Union) $\rightarrow$ S = ?

$\rightarrow$ H = ?

(Western, Union) $\rightarrow$ S = ?

$\rightarrow$ H = ?

$P(S) = 0.1$
$P(H) = 0.9$

For western union:

$$P(S|W) = \frac{P(W|S)\,P(S)}{P(W|S)\,P(S) + P(W|H)\,P(H)}$$

$$= \frac{(0.2)(0.1)}{(0.2)(0.1) + (0.04)(0.9)}$$

$$\boxed{P(S|W) = 0.36}$$

36% of emails containing word "WesternUnion" are spam.

$$P(H|W) = \frac{P(W|H)\,P(H)}{P(W|H)\,P(H) + P(W|S)\,P(S)}$$

$$= \frac{(0.04)(0.9)}{(0.04)(0.9) + (0.2)(0.1)} = \boxed{0.64}$$

For 'western', 'union' in Spam:

$$= \frac{(0.2)(0.2)(0.1)}{(0.2)(0.2)(0.1) + (0.05)(0.05)(0.9)}$$

$$\boxed{P(S|W_{12}) = 0.64}$$

$$P(H|W_{12}) = 1 - 0.64$$
$$\boxed{= 0.36}$$

There are 1000 documents, out of which, 700 are sad documents and 300 happy. Out of happy docs. 100 contains the word "happy".

Out of sad docs. 100 contains "happy". Happy docs 50 contains the word "Shock". Sad docs. 350 contains "Shock".

Computer ~~has~~ is given a document which has both the words "happy" and "Shock"

Find max. posterior hypothesis.

Events	Prior	Cond.Prior	Joint
Sad	0.7	0.14+0.5	0.448
happy	0.3	0.33+0.17	~~0.448~~ 0.15
			0.598

Posterior =

0.75 sad.

0.25 hap