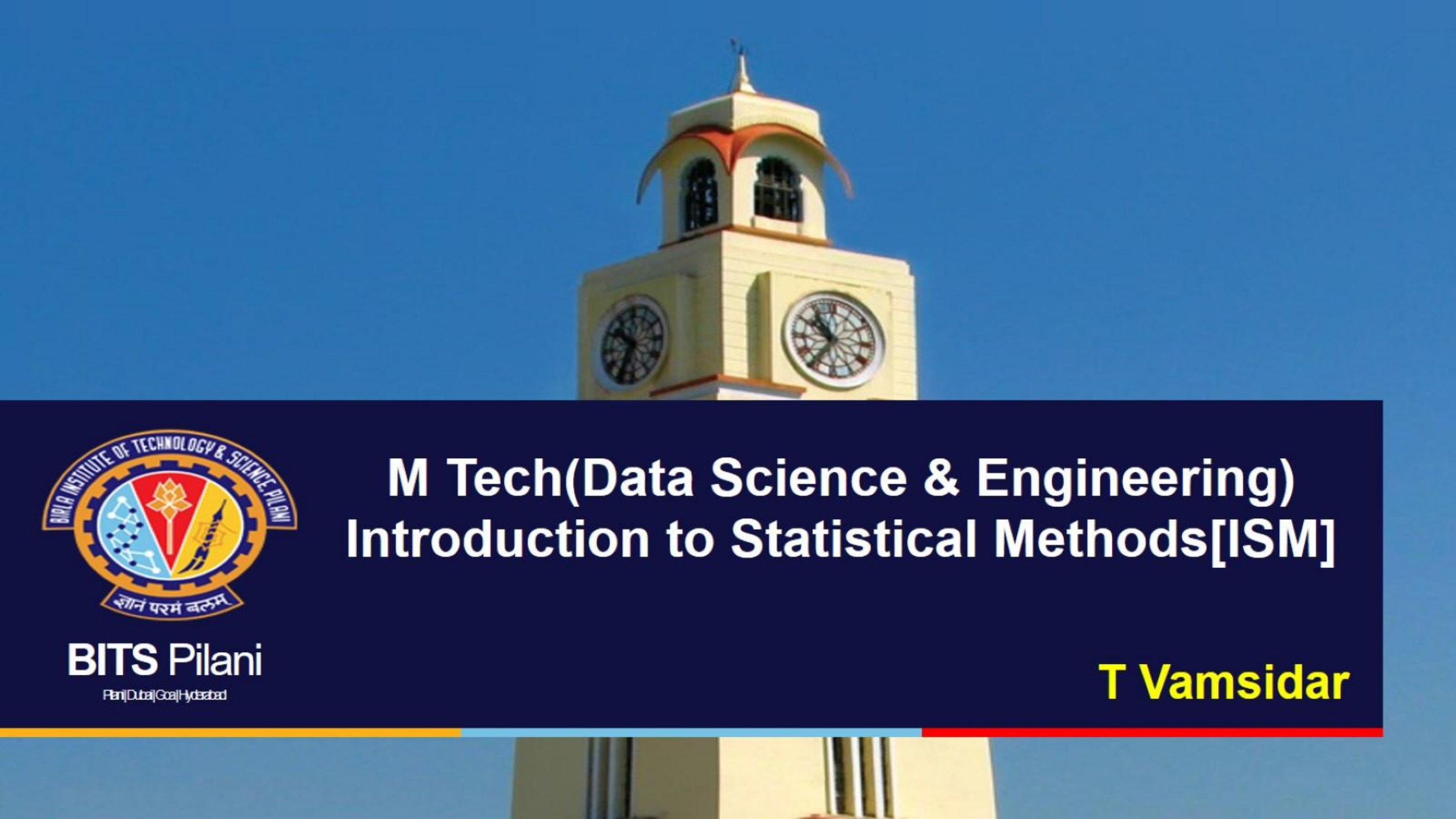


# M Tech(Data Science & Engineering)
# Introduction to Statistical Methods[ISM]

BITS Pilani
Pilani|Dubai|Goa|Hyderabad

T Vamsidar


# Webinar Session No - 1

31st May -2022

**Timimgs : 7.30 to 9.00 PM**

# Learning objectives



- Problems on Basic probability
- Conditional Probability
- Baye's theorem
- Random variables

# Problem - 1



Suppose that 55% of all adults regularly consume coffee, 45% regularly consume carbonated soda, and 70% regularly consume at least one of these two products.

a. What is the probability that a randomly selected adult regularly consumes both coffee and soda? → $P(A \cap B)$

b. What is the probability that a randomly selected adult doesn't regularly consume at least one of these two products?

→ $P(\overline{A \cup B})$

# Solution -1



Let **A** be the event that an adult consumes coffee.

Let **B** be the event that an adult consumes carbonated soda.

Given probabilities of the events.

$$P(A) = 0.55, P(B) = 0.45$$

$$and\ P(A \cup B) = 0.7$$

# Solution -1



## (a) Probability of consuming both coffee and soda

The case is intersection of event A and B,

$$\boldsymbol{P(A \cap B)} = P(A) + P(B) - P(A \cup B)$$
$$= 0.55 + 0.45 - 0.7$$
$$= \mathbf{0.3}$$

# Solution -1



**(b)** The given case represents compliment of the event when the adult consumes at least one of the drink.

$$P[(A \cup B)^c] = 1 - P(A \cup B)$$
$$= 1 - 0.7 = 0.3$$

# Problem - 2



Let $A$ denote the event that the next request for assistance from a statistical software consultant relates to the SPSS package, and let $B$ be the event that the next request is for help with SAS. Suppose that $P(A) = 0.30$ *and* $P(B) = 0.50$

a. Calculate $P(A')$ .

b. Calculate $P(A \cup B)$ .

c. Calculate $P(A' \cap B')$.

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

# Solution -2



Let A be the event that the next request for assistance from a statistical software consultant relates to the SPSS package.

Let B be the event that the next request is for help with SAS.

$P(A) = 0.30 \; and \; P(B) = 0.50$

i. $\boldsymbol{P(A')} = 1 - P(A)$

$= 1 - 0.30$

$= \mathbf{0.70}$

ii. $\boldsymbol{P(A \cup B)} = P(A) + P(B) - P(A \cap B)$

$= P(A) + P(B) - \{P(A) * P(B)\}$ [ A and B are independent events]

$= 0.3 + 0.5 - [0.3 * 0.5]$

$= \mathbf{0.65}$

iii. $P(\bar{A} \cap \bar{B}) = P[(A \cup B)^c]$

$= 1 - P(A \cup B)$

$= 1 - 0.65 = 0.35$

# Problem - 3



The route used by a certain motorist in commuting to work contains two intersections with traffic signals. The probability that he must stop at the first signal is .4, the Analogous probability for the second signal is .5, and the probability that he must stop at least one of the two signals is .6.

What is the probability that he must stop

a. At both signals? $\rightarrow P(A \cap B)$

b. At the first signal but not at the second one? $\rightarrow P(A \cap B^c)$

c. At exactly one signal?

# Solution -3



Let A be the event that he must stop at the first signal

Let **B** be the event that he must stop at the second signal

$P(A) = 0.40 \; and \; P(B) = 0.50$

And given that

$P(A \cup B) = 0.6$

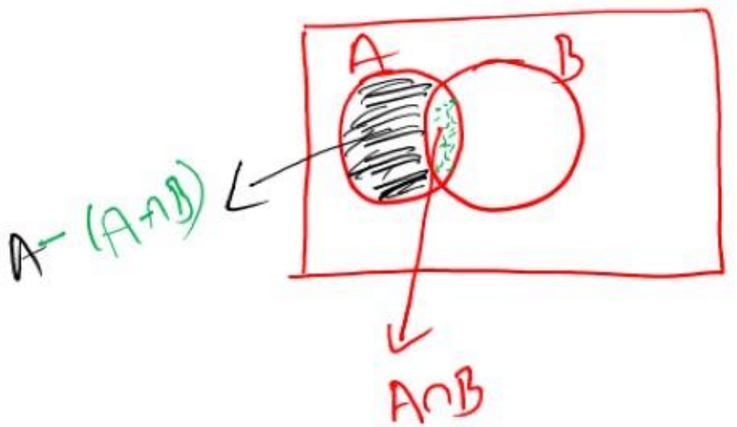
*i.* $P(A \cap B) = P(A) + P(B) - P(A \cup B)$

$= 0.4 + 0.5 - 0.6 = 0.3$

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

$= 0.4 - 0.3$

$= 0.1$



ii. $P(A \cap B') = P(A) - P(A \cap B)$

$= 0.4 - 0.3 = 0.1$

$P[(A \cap B^c) \cup (A^c \cap B)]$

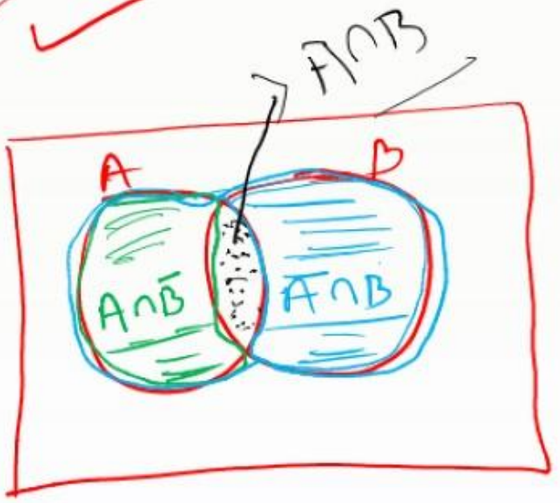
$= P(A \cap B^c) + P(A^c \cap B)$

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

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

iii. $P(\text{At exactly one signal}) = P(A \cup B) - P(A \cap B)$

$= 0.6 - 0.3 = 0.3$



# Problem - 4



**The odds that person X speaks the truth are 3:2 and the odds that person Y Speaks the truth are 5:3. In what percentage of cases are they likely to contradict each on an identical point?**

# Solution -4



Let A = X speaks the truth

$\bar{A}$ = X tell a lie

Let B = Y speaks the truth

$\bar{B}$= Y tell a lie

$$P(A) = \frac{3}{3+2}\ ;\ P(\bar{A}) = \frac{2}{3+2}$$

$$P(B) = \frac{5}{5+3}\ ;\ P(\bar{B}) = \frac{3}{5+3}$$

The event C that X and Y contradict each other on an-identical point.

That can happen in two ways

(i) X speaks the truth and Y tell a lie, i.e, $A \cap \bar{B}$

(ii) X tell a lie and Y speaks the truth, i.e, $\bar{A} \cap B$

# Solution -4



**the events $(A \cap \bar{B})$ and $(\bar{A} \cap B)$ mutually exclusive events**

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

$$= P(A) \,.\, P(\bar{B}) + P(\bar{A}) \,.\, P(B)$$

[ A and B are independent events]

$$= \frac{3}{5} \cdot \frac{3}{8} + \frac{2}{5} \cdot \frac{5}{8}$$

$$= 0.475$$

47.5 % **of cases are they likely to contradict each on an identical point.**

# Problem - 5



Consider the following information about travellers on vacation (based partly on a recent Travelocity poll): **40%** check work email, **30%** use a cell phone to stay connected to work, **25%** bring a laptop with them, **23%** both check work email and use a cell phone to stay connected, and **51%** neither check work email nor use a cell phone to stay connected nor bring a laptop. In addition, **88** out of every 100 who bring a laptop also check work email, and **70** out of every 100 who use a cell phone to stay connected also bring a laptop.

a. What is the probability that a randomly selected traveller who checks work email also uses a cell phone to stay connected?

b. What is the probability that someone who brings a laptop on vacation also uses a cell phone to stay connected?

# Solution - 5



Let $A_E = \{Check\ work\ mail\}$
Let $A_c = \{use\ a\ cell\ phone\ to\ stay\ connected\ to\ work\}$
Let $A_l = \{bring\ a\ laptop\}$
We are given the following probabilities
$P(A_E) = 0.4$
$P(A_c) = 0.3$ and $P(A_l) = 0.25$

$P(A_E \cap A_c) = 0.23$

$P(A'_E \cap A'_c \cap A'_l) = 0.51$
$P(A_E / A_l) = 0.88 = 88/100$

$P(A_l / A_c) = 0.7$

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

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

$A_C \& A_L$

(ii) $P(A_C/A_L) = \dfrac{P(A_C \cap A_L)}{P(A_L)}$

$= \dfrac{P(A_L/A_C)\, P(A_C)}{P(A_L)}$

a. We need to find the probability that a randomly selected traveller who checks work email also uses a cell phone to stay connected.

Which is a conditional probability of $A_c$ given $A_E$ i.e,

$$P\left(A_c / A_E\right) = \frac{P(A_c \cap A_E)}{A_E}$$

$$= \frac{0.23}{0.4} = 0.575$$

**b. The probability that someone who brings a laptop on vacation also uses a cell phone to stay connected.**

**Which is a conditional probability of $A_c$ given $A_l$ i.e,**

$$P\left(A_c/A_l\right) = \frac{P(A_c \cap A_l)}{A_l} = \frac{P\left(A_l/A_c\right).P(A_c)}{P(A_l)}$$

$$= \frac{(0.7)*(0.3)}{0.25} = 0.84$$

# Problem - 6



**At a certain gas station, 40% of the customers use regular gas ($A_1$), 35% use plus gas ($A_2$), and 25% use premium ($A_3$). Of those customers using regular gas, only 30% fill their tanks (event *B*). Of those customers using plus, 60% fill their tanks, whereas of those using premium, 50% fill their tanks.**

a. What is the probability that the next customer will request plus gas and fill the tank?

b. What is the probability that the next customer fills the tank? → $P(B)$

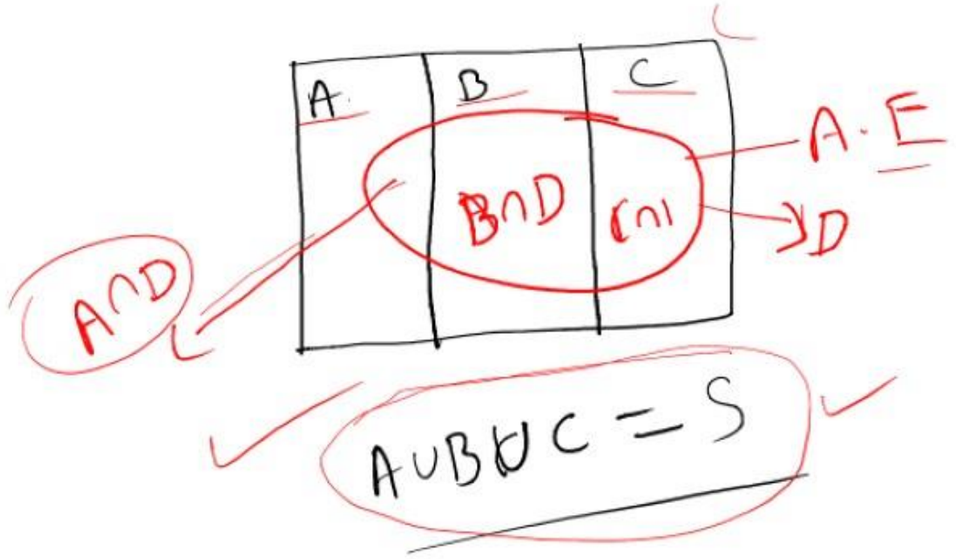
c. If the next customer fills the tank, what is the probability that regular gas is requested? Plus? Premium?

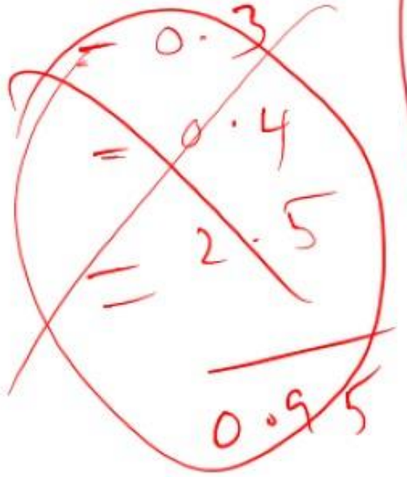
$P(A_1/B)$ $P(A_2/B)$ $P(A_3/B)$

$P(B/A_1) = 30\%$ $P(B/A_2) = 0.6$

$P(B/A_3) = 0.5$

A




$P(A_1) = 40\% = 0.4$

$P(A_2) = 35\% = 0.35$

$P(A_3) = 25\% = 0.25$

$= 1.0$

# Solution - 6



Probabilities of customers using regular gas

$$P(A_1) = 40\% = 0.4$$

Probabilities of customers using plus gas

$$P(A_2) = 35\% = 0.35$$

Probabilities of customers using premium gas

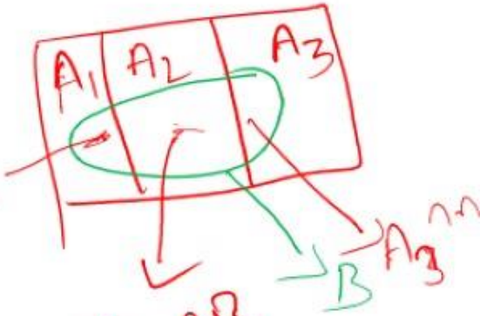
$$P(A_3) = 25\% = 0.25$$

Also given with conditional probabilities of full gas tank

$$P\left(B/A_1\right) = 30\% = 0.3$$

$$P\left(B/A_2\right) = 60\% = 0.6$$

$$P\left(B/A_3\right) = 50\% = 0.5$$




$P(B) = P(A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)$

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

(a) The probability that next customer will requires plus gas and fill the tank is $P(A_2 \cap B) = P(A_2) * P\left(B/A_2\right)$

$= 0.35 * 0.6$

$= 0.21$

$A_2, B$

$A, B \rightarrow$ Multiply

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

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

(b) The probability of next customer filling the tank is

$$P(B) = P(A_1)\,P\left(B/A_1\right) + P(A_2)P\left(B/A_2\right) + P(A_3)P\left(B/A_3\right)$$

$$= (0.4 \times 0.3) + (0.35 \times 0.6) + (0.25 \times 0.5)$$

$$= 0.455$$

(c) If the next customer fills the tank, probability of requesting regular gas is

$$P\left(A_1/B\right) = \frac{P(A_1)\,P\left(B/A_1\right)}{P(B)}$$

$$= \frac{0.4 * 0.3}{0.455} = 0.264$$

$\frac{P(A_1 \cap B)}{P(B)}$

If the next customer fills the tank, probability of requesting plus gas is

$$P\left(\frac{A_2}{B}\right) = \frac{P(A_2)\,P\left(\frac{B}{A_2}\right)}{P(B)}$$

$$= \frac{0.35 * 0.6}{0.455} = 0.462$$

If the next customer fills the tank, probability of requesting premium gas is

$$P\left(\frac{A_3}{B}\right) = \frac{P(A_3)\,P\left(\frac{B}{A_3}\right)}{P(B)}$$

$$= \frac{0.25 * 0.5}{0.455} = 0.274$$

# Problem - 7



"A company wants to find reasons for the dissatisfaction among employees because of which they are leaving the company. With this view they took the feedback among the employees and identified three reasons i.e. working conditions, pay hike, commuting. The probabilities of dissatisfaction with these three factors are 0.6, 0.3 and 0.1 respectively. And the probabilities that they are leaving the organization with these reasons are 0.3, 0.5and 0.8 respectively."

As a data scientist, use an appropriate statistical model / method to model this case and suggest the company the probable reasons on priority so that they can focus on it to retain the employees.

# Solution - 7



**Let W.C = with the working conditions employees are dissatisfaction.**

$P(W.C) = 0.6$

**Let P.H = with the pay hike employees are dissatisfaction.**

$P(P.H) = 0.3$

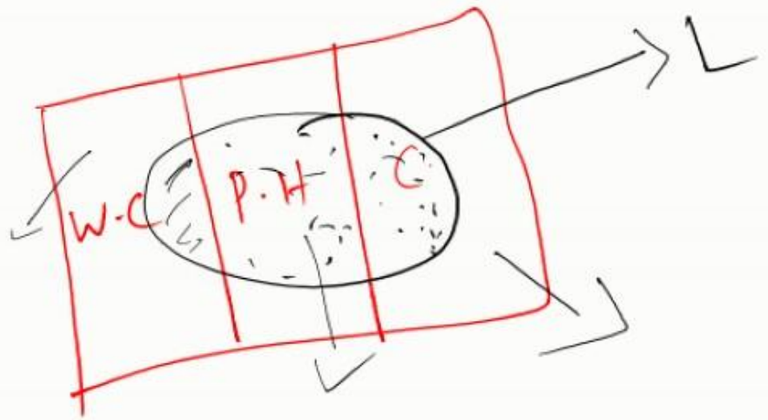
**Let CM = with the commuting employees are dissatisfaction.**

$P(CM) = 0.1$

1.0

**And**

**Let L be event that they are** leaving the company.



# Solution - 7



$$P(^{L}/_{W.C}) = 0.3\ , \quad P(^{L}/_{P.H}) = 0.5, \qquad P(^{L}/_{CM}) = 0.8$$

P[dissatisfaction among employees because of W.C and they are leaving the company]

$$P[(W.C) \cap L] = P(W.C)P(^{L}/_{W.C}) = (0.6)(0.3) = 0.18$$

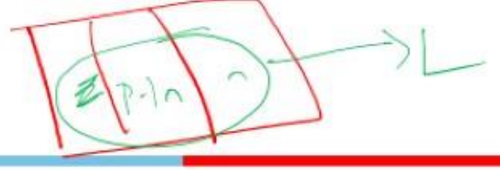
P[dissatisfaction among employees because of P.H and they are leaving the company]

$$P[(P.H) \cap L] = P(P.H)P(^{L}/_{P.H}) = (0.3)(0.5) = 0.15$$

P[dissatisfaction among employees because of CM and they are leaving the company]

$$P[(CM) \cap L] = P(CM)P(^{L}/_{CM}) = (0.1)(0.8) = 0.08$$

# Solution - 7





$$P[\text{leaving the company}] = P[L] = P[(W.C) \cap L] + P[(P.H) \cap L] + P[(CM) \cap L]$$

$$= P(W.C)P(^{L}/_{W.C}) + P(P.H)P(^{L}/_{P.H}) + P(CM)P(^{L}/_{CM})$$

$$= (0.6)(0.3) + (0.3)(0.5) + (0.1)(0.8)$$

$$= 0.41$$

$P$[they are leaving the organization, because of W.C]

$$= P(^{W.C}/_{L}) = \frac{P[(W.C) \cap D]}{P[D]} = \frac{0.18}{0.41} = 0.4390$$

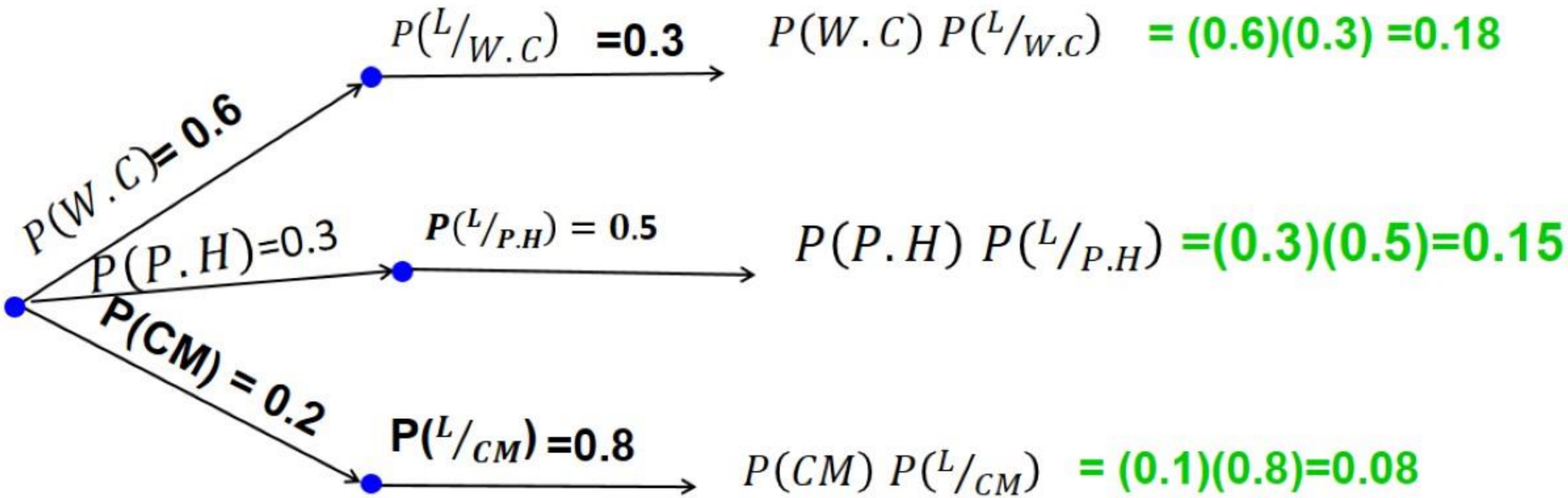
$$P(^{P.H}/_{L}) = \frac{P[(P.H) \cap D]}{P[D]} = \frac{0.15}{0.41} = 0.3658$$

$$P(^{CM}/_{L}) = \frac{P[(CM) \cap D]}{P[D]} = \frac{0.08}{0.41} = 0.1951$$

Most of the employees are leaving the organization with the reason working conditions.

# Probability Tree Diagram






$P(W.C)\ P(L/W.C)$ = (0.6)(0.3) =0.18

$P(P.H)\ P(L/P.H)$ =(0.3)(0.5)=0.15

$P(CM)\ P(L/CM)$ = (0.1)(0.8)=0.08

**P(L)** $= P(W.C)\ P(L/W.C) + P(P.H)\ P(L/P.H) + P(CM)\ P(L/CM)$ = 0.41

# Problem - 8



A company that manufactures video camera as produces a basic model and a deluxe model. Over the years 40% of the cameras sold have been the basic model. Of those buying the basic model 30% purchase an extended warranty, whereas 50% of all deluxe purchasers do so. What is the probability that a randomly selected customer has an extended warranty, how likely is it that he or she has a basic model?

# Solution - 8



Let B be the event of a basic model.

Let D be the event of a deluxe model.

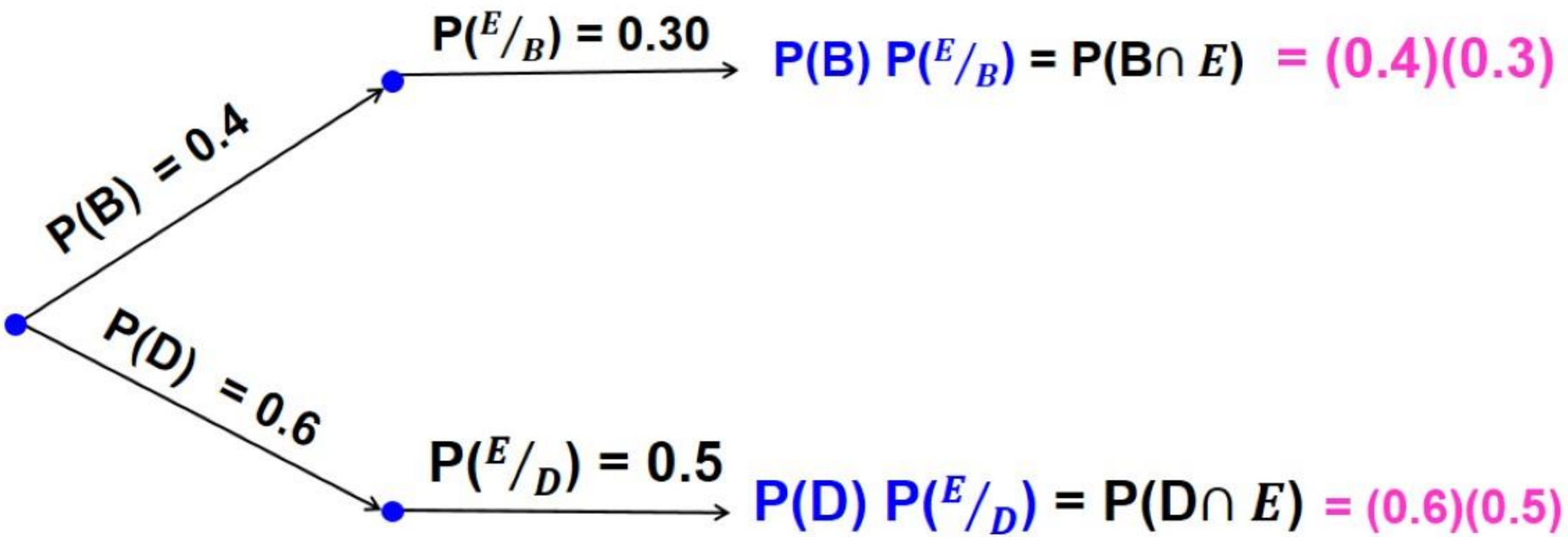
Let E be event of extended warranty.

P(B) = 0.40, P(D)= 0.60,

P(E / B)= 0.30, and P(E / D)= 0.50

# Probability Tree Diagram






P(B) P(E/B) = P(B∩ E) = (0.4)(0.3)

P(D) P(E/D) = P(D∩ E) = (0.6)(0.5)

P(E) = P(B) P(E/B) + P(D) P(E/D) = 0.42

# Solution - 8



**the probability that a randomly selected customer has an extended warranty, how likely is it that he or she has a basic model is**

$$P(B/E) = \frac{P(B \cap E)}{P(E)} = \frac{(0.4)(0.3)}{0.42} = 0.2857$$

# Problem - 9



The chance that doctor A will diagnose a disease X correctly is 60%. The chance that a patient will die by his treatment after correct diagnose is 40% and the chance of death by wrong diagnose is 70%. A patient of doctor A, who had disease X, died. What is the chance that his disease was correctly.

# Solution-9



Let $E_1$= event that disease X is diagnosed correctly by doctor A

Let $E_2$= event that a patient of doctor A who has disease X died

Then $P(E_1) = \frac{60}{100} = 0.6$ and $P(\bar{E}_1) = 1 - P(E_1) = 0.4$

$P(E_2/E_1) = \frac{40}{100} = 0.4$ and $P(E_2/\bar{E}_1) = \frac{70}{100} = 0.7$

# By Baye's theorem

$$P\left(\frac{E_1}{E_2}\right) = \frac{P(E_1)P\left(\frac{E_2}{E_1}\right)}{P(E_1)P\left(\frac{E_2}{E_1}\right)+P(\bar{E}_1)P\left(\frac{E_2}{\bar{E}_1}\right)}$$

$$= \frac{0.6 * 0.4}{0.6 * 0.4 + 0.4 * 0.7} = \frac{6}{13}$$

# Probability Mass function(Pmf)

## (Probability distribution function)

Let 'X' be a discrete random variable taking values $x_1, x_2, \ldots \ldots . x_n$ then the probability mass function $P(X = x)$ is defined under the following conditions

$(i)\ P(X = x) \geq 0$

$(ii)\ \sum_{\forall x} P(X = x) = 1$

# Continuous Probability Distributions



**A random variable is called continuous when it assumes values in a given interval.**

**The probability that a random variable X assumes different values x in a given interval, say [a, b], is denoted by f(x) = P(a ≤ X ≤ b), called probability density function (pdf).**

And

$$\mathrm{f(x)} = P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{dx}$$

# Properties of continuous probability distribution



A function f(x) to be a probability density function (pdf),
if it satisfies the following conditions

1. $f(X) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. For any $a \leq b$,

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

Let X be a continuous random variable.

Then $P(a \leq X \leq b) = P(a < X \leq b)$

$= P(a \leq X < b)$

$= P(a < X < b)$

**NOTE**: The Cumulative Distribution Function of continuous random variable X is denoted by F(X) and is defined as

$$F(X) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

and $\mathbf{f(X) = \frac{d}{dx}[F(X)]}$

# Problem - 10



Verify that $P(X) = \frac{x+3}{25}$ $for\ x = 1,2,3,4,5$

Serve as Probability mass function?

# Solution:

Given that $P(X) = \frac{x+3}{25}$ $for\ x = 1,2,3,4,5$

$P(X = 1) = \frac{1+3}{25} = \frac{4}{25}$,

$P(X = 2) = \frac{5}{25}$

$P(X = 3) = \frac{6}{25}$

$P(X = 4) = \frac{7}{25}$

$P(X = 5) = \frac{8}{25}$

(i) $0 \leq P(A) \leq 1$

$$\sum_{x=1}^{5} P(X = x) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

$$= \frac{4}{25} + \frac{5}{25} + \frac{6}{25} + \frac{7}{25} + \frac{8}{25}$$

$$= \frac{30}{25} > 1$$

Given P(X) is not a Probability mass function

[ we know that Total probability is 1 ]

# Problem - 11



For the discrete probability distribution

| X=x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| P(X=x) | 0 | k | 2k | 2k | 3k | $k^2$ | $2k^2$ | $7k^2 + k$ |

find (i) k (ii) $P(X < 6)$ (iii) $P(X \geq 6)$

(iv) Mean and Variance

# Solution - 11



(i) We know that the Total probability is 1

$$\sum_{i=0}^{7} P(X = x_i) = 1$$

$$0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$$

$$10k^2 + 9k - 1 = 0$$

$$k = -1 \text{ and } k = {}^1/_{10}$$

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

then $k = \frac{1}{10}$ [ since P(X) ≥ 0, so $k \neq -1$]

# Solution - 11



(ii) $P(X < 6) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$
$+P(X = 4) + P(X = 5)$
$= 0 + k + 2k + 2k + 3k + k^2$
$= k^2 + 8k$
$= 0.81$ [$k = \frac{1}{10} = 0.1$ ]

(iii) $P(X \geq 6) = P(X = 6) + P(X = 7)$

or

$P(X \geq 6) = 1 - P(X < 6) = 1 - 0.81$
$= 0.19$

# Solution - 11



Mean = $E[x] = \sum_{x=0}^{7} xP(X = x)$

= 0 + k + 4k + 6k + 12k + $5k^2$ + $12k^2$ + $49k^2$ + $7k$

= $66k^2 + 30k = 3.66$

| X=x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| P(X=x) | 0 | k | 2k | 2k | 3k | $k^2$ | $2k^2$ | $7k^2 + k$ |

$\boldsymbol{var(x) = E(x^2) - [E(x)]^2}$

$E(x^2) = \sum_{x=0}^{7} x^2 P(X = x)$

= 0 + k + 8k + 18k + 48k + $25k^2$ + $72k^2$ + $343k^2$ + $49k$

= $440k^2 + 75k = 16.8$

Then $var(x) = E(x^2) - [E(x)]^2 = 16.8 - (3.66)^2$

= 3.4044

# Thanks