Random = uncertain = non-deterministic

Date 10/10/24.

PROBABILTY:- (SELF NOTES)

How to determine if a high corelation b/w two variables is statistically significant?
↳ is it genuine.
↳ or random variation.

↳ Probability can help answer this question.
↳ in it we perform experiments
↳ from which we note the frequency of our outcomes

Probability provides mathematical tools/models to reason about uncertainity/randomness.

Examples given in slides.

Sample space:- All possible outcomes from a certain experiment.

Event:- the description of an outcome.
(which outcome is possible & which is not?)
↓
E

It is the result of a probability ∴ a subset of a sample space.
↘ Ω

An event can have the following description:-

- Zero outcome
- 1 outcome.
- Several Outcome
- All outcome.

} depends upon the definition of event.

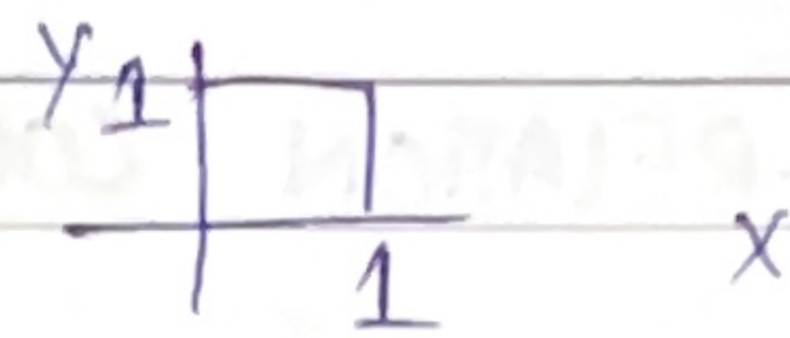
$E_1 \cup E_2$ ⇒ Either/all event 1 & event 2 occur.

$E_1 \cap E_2$ ⇒ common from event 1 & Event 2.

E_1' ⇒ All except event 1.

$E_1 - E_2$ ⇒ All of E_1 except E_2.

continuous sample space : $\Omega = \{ (x,y) \mid 0 \le x, y \le 1 \}$



discrete sample

A strong claim of probability (of a certain outcome) from an event A can only be made if multiple experiments are considered.

$$P(A) = \lim_{N \to \infty} \frac{\text{No. of times A occurs.}}{N}$$

indicates that we can do many exp.

$\boxed{\text{Remember}}$:-

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2)$
$- P(E_2 \cap E_3)$ $P(E_1 \cap E_3) + P(A_1 \cap B_2 \cap E_3)$.

Discrete ~~continuous~~ finite events.
Discrete " infinite " .
~~Continuous~~ events.

$P(E_1 \cup E_2 \cup E_3)$.

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

E_1 : even no.s
E_2 : greater no. than 3
E_3 : less than 5

$P(E_1 \cap E_2) \Rightarrow \{4, 6\} \Rightarrow \frac{2}{6}$

$P(E_1) \Rightarrow \frac{3}{6}$, $P(E_2) \Rightarrow \frac{3}{6}$, $P(E_3) = \frac{4}{6}$

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$
$3/6 + 3/6 - 2/6 \Rightarrow 0.667$

Date

USING CO-RELATION COEFF TO PREDICT.

Predictor.

$(\hat{y}_i)^P = a\hat{x}_i + b$. 'a' and 'b' are unknown.

The error is calculated as $U_i = \hat{y}_i - (\hat{y}_i)^P$.
Usually we want $\text{mean}\{(U)\} = 0$. — (a)

$\text{mean}\{U\} = \text{mean}\{(\hat{y}_i) - (\hat{y}_i)^P\}$

$= \text{mean}\{(\hat{y}_i)\} - \text{mean}\{(a\hat{x}_i + b)\}$

$= 0 - a\,\text{mean}\{\hat{x}_i\} + b$.

if $\text{mean}\{U\}$

$0 = b$

$\therefore$ to achieve (a) $b = 0$.

Similarly now to think about what should be the value of 'a' we take variance$\{U\}$?

$= \text{Var}(\{\hat{y} - \hat{y}p\}$

$= \text{mean}(\{y - \hat{a}x)^2\} \longrightarrow (?)$

$\Rightarrow \text{mean}(\{(\hat{y})^2 - 2(a)(\hat{x}(\hat{y}) + a^2(\hat{x})^2$

(?) — $\Rightarrow$ $\text{mean}(\hat{y}^2) - 2a\,\text{mean}(\hat{x}\hat{y}) + a^2\text{mean}(\hat{x})^2$

$1 \qquad\qquad\qquad\qquad 1$

$\Rightarrow 1 - 2ar + a^2$.

minimize $\quad d\text{var}\{u_i\}/da = 0 - 2r + 2a$

Date

$\therefore \ (\hat{y_i})^p = r\hat{x_i}$

—— x ——

PROCEDURE OF PREDICTING VALUES USING CORRELATION.

step ① $\hat{x_i} = \frac{x_i - mean(\{x_i\})}{std\{x_i\}}$, step ② $\hat{y_i} = \frac{y_i - mean(y_i)}{std(y_i)}$

correlation coeff $\Rightarrow \frac{1}{N} \sum_{i=1}^{N} (\hat{x_i})(\hat{y_i}) = r$

Predicted value points $\Rightarrow$ ———— (a)

$(\hat{y_i})^p = r\hat{x_i}$

Non standardizing (a) would be by the following formula.

$(y_p)^p = (std)_y \cdot r \cdot \hat{x_i} + mean(y)$

PROBABILITY (SELF NOTES!)

Calculation of Probability:-

(i) Discrite Countable finite Event. ↗ outcomes are atomic (can not be further simplified!)

(ii) Discrete Countable infinite Event ↓ Can be categorized

(iii) Continuous Event

eg tossing the coin until head.

Tossing (a) coin twice;

sample space = {HH, HT, TH, TT} ⇒ 4.

E_1 ⇒ Same outcomes arrive ⇒ 2 → No. of outcomes of this event.

$P(E_1) = \frac{2}{4}$ Ans!.

—— x ——

| QUESTION |

$$\frac{\binom{12}{8}}{\binom{14}{8}}$$ ⇒ Probability of Selecting HODs not including from CS.

$$\therefore 1 - \frac{\binom{12}{8}}{\binom{14}{8}}$$ ⇒ Probability of Selecting HODs from CS.

Date

Combinations formula;

let $\binom{14}{8} \Rightarrow \frac{N!}{k!\,(N-k)!}$ $\binom{N}{k} \Rightarrow \frac{14!}{8!\,(14-8)!}$

— x —

Conditional probability:

Probability of an event ? given that some event has occured.

e.g what is the probability of a CS student tested positive to have/have a diese?

what is know. is that the disease has occered. COVID is tested positive

what is to find is the P(D)?

∴ By conditional Probability.

$$P(D|T) \Rightarrow \frac{P(D \cap T)}{P(T)} \qquad P(T) \neq 0)$$

Date

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

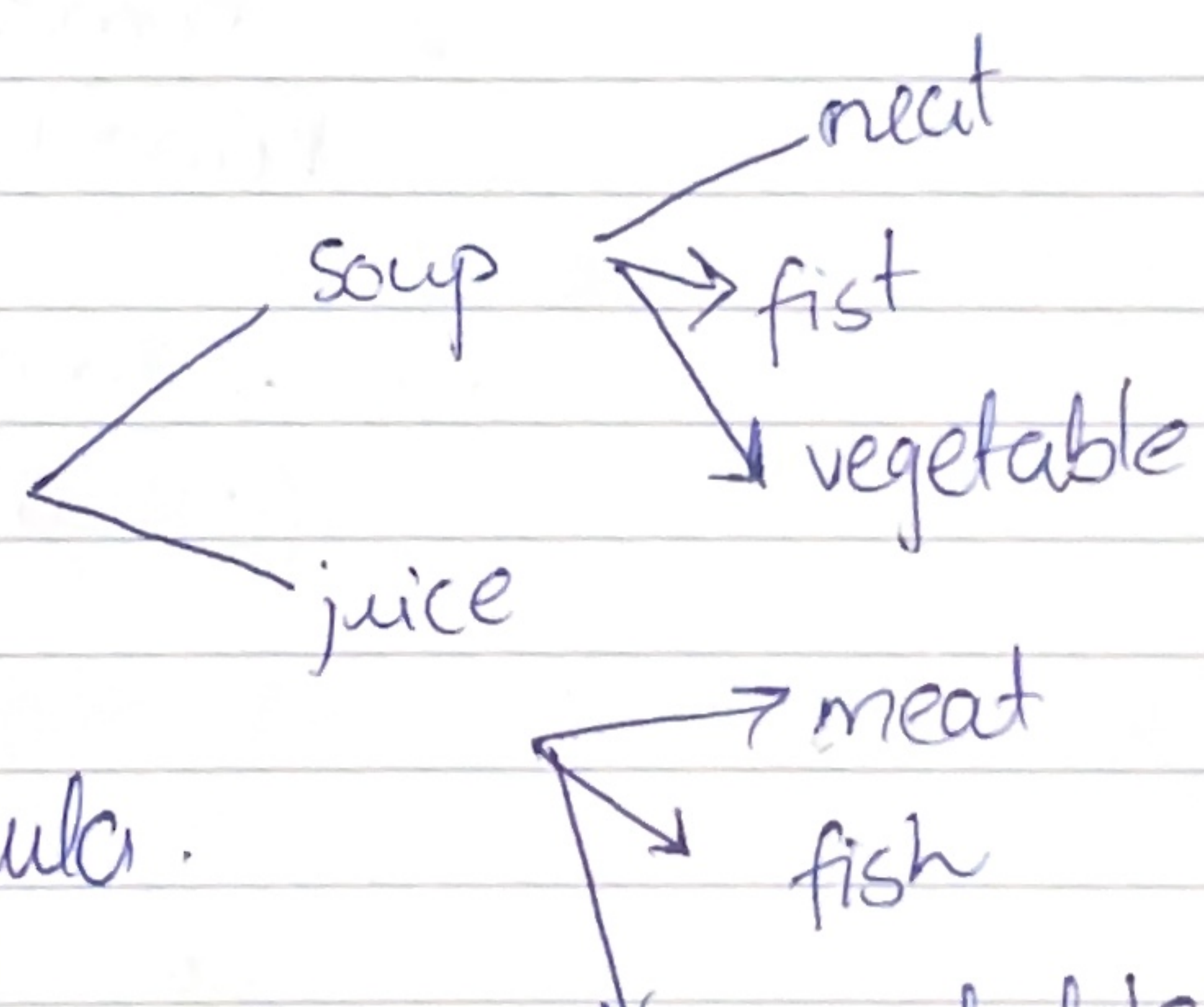
$P(B \cap A) = P(B|A)\,P(A).$

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

Theory of total Probability :-

A method of calculating probability of an event based on partioning the ~~even~~ sample space into disjoint subsets.

$P(\text{meat}) = P(\text{meat} \cap \text{soup}) +$
$P(\text{meat} \cap \text{juice})$



Now By joint Probability formula.

$P(\text{meat}|\text{soup}) \times P(\text{soup}) + P(\text{meat}|\text{soup}) \times P(\text{juice}).$