Ic-252

INTERPRETATION OF PROBABILITY

1. Frequency Interpretation

→ Estimate of prob. is completely dependent on samples.

2. Classical Interpretation

→ Estimate of prob. based on concepts of likelihood, based on knowledge of physical world. (without doing an experiment)

PROBABILITY THEORY

⇒ Sample Space :- Set consists of all possible outcomes for an Experiment.

Eg :- Roll of a dice = $\{1, 6, 2, 3, 4, 5\} = S$

⇒ Event :- A subset of Sample Space.

Events can be represented as sets A, B, C
⇒ this enables us to define operations on the sets.

Eg: →

$C = A \cup B$: Event C is made up of outcomes of either A or B or both.

$C = A \cap B$: Event C is made up of outcomes from both A and B.

$C = A^c$: Set of outcomes which are not in A.

$A \subset B$: Set of outcomes in A which are also in B.

Axioms Of Probability

* Axiom-1

$$\boxed{0 \leq P(E) \leq 1}$$

* Axiom-2

$$\boxed{P(S) = 1}$$

** Axiom-3

If two events E & F are "Mutually exclusive" then,

$$\boxed{P(E \cup F) = P(E) + P(F)}$$

Axiom-3 is also true for n mutually exclusive events. i.e.

$$\boxed{P(E_1 \cup E_2 \ldots\ldots E_n) = \sum_{j=1}^{n} P(E_j)}$$

PROPERTIES

1. $P(E^c) = 1 - P(E)$

Proof :→

we know,

$S = E^c + E$

[∵ We know that Complement are mutually exclusive]

so, $P(S) = P(E^c + E)$

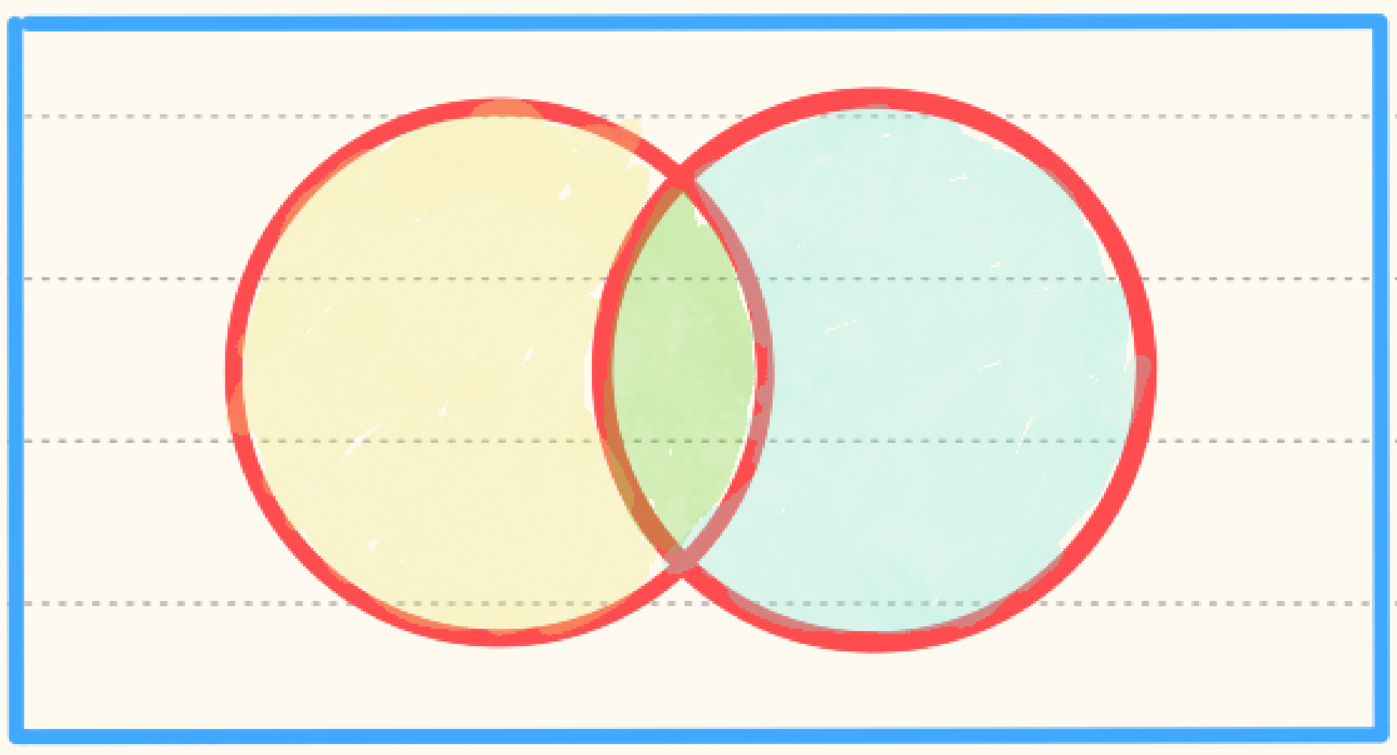
$P(S) = P(E^c) + P(E)$ [Axiom-3]

$1 = P(E^c) + P(E)$ [Axiom-2]

$\Rightarrow P(E^c) = 1 - P(E)$

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

Proof :→



$$(E \cup F) = [E - EF] \cup [EF] \cup [F - EF] \quad \text{— (I)}$$

$$P(E \cup F) = P(E - EF) \cup P(EF) \cup P(F - EF)$$

$$= P(E - EF) + P(EF) + P(F - EF) \quad \text{[Axiom-3]} \quad \text{— (II)}$$

$$E = (E - EF) \cup (EF)$$

$$P(E) = P(E - EF) + P(EF) \quad \text{— (III)}$$

$$P(F) = P(F - EF) + P(EF) \quad \text{— (IV)}$$

So, $P(E \cup F) = P(E - EF) + P(EF) + P(F - EF)$

$$= P(E - EF) + P(EF) + P(F - EF) + P(EF) - P(EF)$$

$$= P(E) + P(F) - P(EF)$$

So,

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

NOTE:→ In Proof, $P(E \cap F)$ is written as $P(EF)$.

Bayes Rule

$$\boxed{P(E/F) = \frac{P(E \cap F)}{P(F)}}$$

$P(E/F)$ = Probability of event E if event F has happened.

$$P(E/F) = \frac{P(EF)}{P(F)}$$

$$\Rightarrow P(E/F) \cdot P(F) = P(EF) \quad \text{— (I)}$$

$$\Rightarrow P(F/E) \cdot P(E) = P(EF) \quad \text{— (II)}$$

$$P(E/F) \cdot P(F) = P(F/E) \cdot P(E)$$

$$\boxed{P(E/F) = \frac{P(F/E) \cdot P(E)}{P(F)}}$$

Bayes Rule

finding probability of event E given some measurement F.

$P(F/E)$ is used because F can be noisy. Hence, we should use another probability measure which quantifies the uncertainity in F.

Chances of labelling C_i for measurement F.

$$P(C_i/F) = \frac{P(C_i F)}{P(F)} = \frac{P(F/C_i)\,P(C_i)}{P(F)}$$

$$S = C_1 \cup C_2 \quad \Big(\text{Also Expressed as } \underbrace{\textstyle\bigcup_i C_i}_{\text{Union of Mutually exclusive events}}\Big)$$

$$F = \bigcup_i FC_i$$

$$P(F) = \sum P(FC_i)$$

$$P(F) = \sum_i P(C_i)\,P(F/C_i) = \sum_{j=1}^{k} P(F \cap B_j)$$

Total probability of F over the sample space S.

Likelihood : $P(f/C_i)$ assuming $f \rightarrow C_i$
what are chances of this corresponding

Prior: chances of that class actually being present.

Theorem :→ Multiplication rule of C.P.

*

$$\boxed{P(A_1 \cap A_2 \cap \cdots A_n) = P(A_1)\, P(A_2/A_1)\, P(A_3/A_2 \cap A_1) \cdots\cdots P(A_n/A_1 \cap A_2 \cdots A_{n-1})}$$

Proof :→

$$RHS = \cancel{P(A_1)} \cdot \frac{\cancel{P(A_2 \cap A_1)}}{\cancel{P(A_1)}} \cdot \frac{P(A_3 \cap \cancel{A_2 \cap A_1})}{\cancel{P(A_2 \cap A_1)}} \cdots \frac{P(A_1 \cap A_2 \cdots A_n)}{\cancel{P(A_1 \cap A_2 \cdots A_{n-1})}}$$

$$= P(A_1 \cap A_2 \cap A_3 \cdots \cap A_n)$$

Assumption → $P(A_1 \cdots \cap A_{n-1}) > 0$

This would make all the terms in denominator > 0.

Conditional Version of Law of Total Probability

$$\boxed{P(A/C) = \sum_{j=1}^{k} P(B_j/C)\, P(A/B_j \cap C)}$$

Proof:→

$$RHS = \sum_{j=1}^{k} \frac{\cancel{P(B_j \cap C)}\, P(A \cap B_j \cap C)}{P(C)\, \cancel{P(B_j \cap C)}}$$

$$= \sum_{j=1}^{k} \frac{P(A \cap B_j \cap C)}{P(C)}$$

$$= \sum_{j=1}^{k} P(A \cap B_j / C)$$

$$= \sum_{j=1}^{k} P(A/C, B_j)$$

$$= P(A/C)$$

[Related to Concept of Total Prob.]

INDEPENDENCE

If:

$$\boxed{P(A/B) = P(A)} \quad \text{(I)}$$

i.e. A is not dependent on B.

$$\boxed{P(A \cap B) = P(A)\, P(B)} \quad \text{(II)}$$

Practical Example in DS: $\underline{P(f/c_i)}$

Likelihood of one attribute over a class.

A case of multiple attributes:

$$\underline{P(f_1, f_2 \cdots f_n / c)}$$

Likelihood of many attributes over one class.

If $f_1, f_2 \cdots f_n$ are independent then,

$$P(f_1 \cap f_2 \cap f_3 \cdots f_n / c) = P(f_1/c)\, P(f_2/c) \cdots P(f_n/c)$$

Decomposition of high-dimensional likelihood in low-dimensional likelihood.

INDEPENCE OF COMPLEMENT

If A and B are independent then A and B^c are independent.

$$\Rightarrow \boxed{P(AB^c) = P(A)\,P(B^c)}$$

Proof:→ $A = AB \cup AB^c$

$$\Rightarrow P(A) = P(AB) + P(AB^c)$$

$$\Rightarrow P(A) = P(A)P(B) + P(AB^c)$$

$$\Rightarrow P(AB^c) = P(A)\big(1 - P(B)\big)$$

$$\Rightarrow P(AB^c) = P(A)\,P(B^c)$$

* NOTE:→
Mutually exclusiveness and Independence are two completely different things.

INDEPENCE OF MULTIPLE EVENTS

Example of 3 sets:

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

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

& $P(B \cap C) = P(B)\,P(C)$

& $P(C \cap A) = P(C)\,P(A)$

Then A, B, C are Independent.

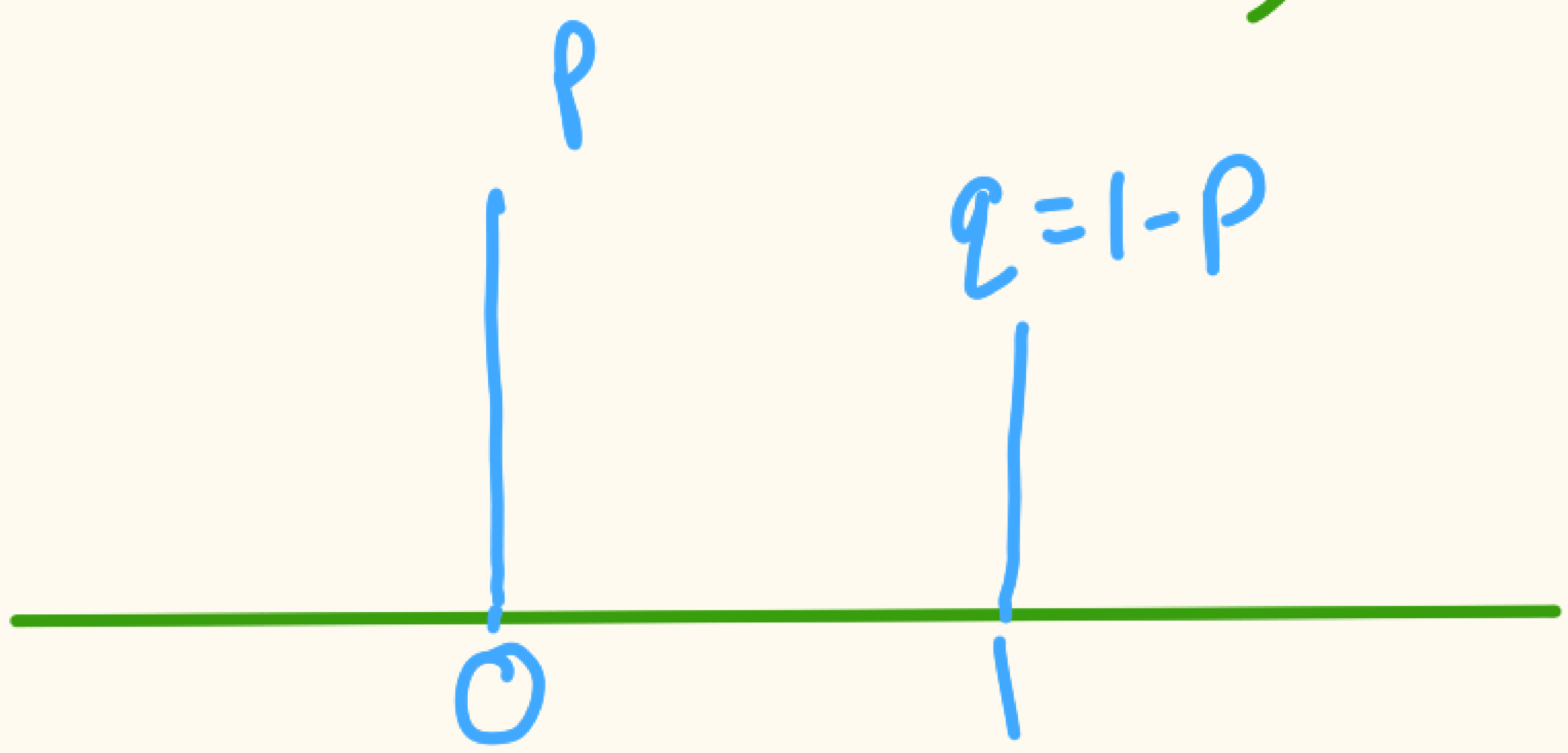
For n general sets $(A_1, A_2 \cdots A_n)$

If Independence of all subsets of $\{A_1, A_2, A_3 \cdots A_n\}$ follows then $\{A_1, A_2 \cdots A_n\}$ are independent.

DISTRIBUTIONS

Independence $\longrightarrow$ a distribution of two independent events (Yes/No)

$$\left.\begin{aligned} P(A=0) &= p \\ P(A=1) &= q = 1-p \end{aligned}\right\} \text{Bernaulli's Distribution}$$



↳ Repeat this experiment (which has 2 outcomes) multiple times.

↳ One way to consider probability in such a scenario is given by Binomial Distribution.

Binomial Distribution

If an exp. with 2 outcomes with probability p and q (1-p) is repeated k times, prob. of k success in n exp. is given by Binomial Dis.
Assumption: Each trial is Independent.

Birthday Pairing

Total No. of possible states = 365

No. of person = k

Total no. of states = $(365)^k$
→ Total no. of outcomes.

To Calculate no. of cases where atleast 2 share a birthday.

$= 1 - \left[\text{No. of Cases in which no 2 people share a birthday.}\right]$

${}^{365}P_K$ is no. of cases where no two people share birthdays, So, Probability that atleast 2 people share a birthday $= \left[1 - \frac{{}^{365}P_k}{(365)^k}\right]$

Assumption $k < 365$.

BINOMIAL DISTRIBUTION

Eg :→ out of 10 trials, prob. of one instance could be as follows,

$$P(n, k, p) = \underbrace{p \times p \times p \times p}_{4\text{times}} \times \underbrace{(1-p)(1-p) \cdots - (1-p)}_{6\text{times}}$$

other scenarios :→ Different ways of selecting k out of n.

So, Eg :→ 1) $p(1-p)(p) \cdot - \cdot$

2) $p(p)(1-p) \cdot - -$

etc..

So,

overall Probability $(P(p, n, k))$

So, $P(p, n, k) = {}^{n}C_{k}\, p^{k} (1-p)^{n-k}$

Assigning probabilities to numbers or range of numbers rather than in terms of outcomes, events . . . .

Mapping between the space of events, outcomes etc. to the space of numbers.

$$\underset{\text{(Sample space)}}{\Omega} \longrightarrow R^{n}$$

[Mapping between sample space & real line]

a function $\Rightarrow$ Random Variable

Eg:→ R.V is the sum of the face values on the two dice. (Discrete Values)

→ R.V. Recorded temperature on a sensor.

we could also plot the events on a real line to get R.V.

Eg.

Eg:→ Prop of continuous R.V.

Like, $P(a \le X \le b)$, $P(X \ge c)$, $P(X < c)$ } Assigning prob. to different intervals on R.

Distribution Function (F) or Cummutative distribution fuction (CDF)

$$\boxed{F_X(x) = P\{X \le x\}}$$

Random Variabl

Defines Interval

Properties of Distribution functions

1. $F(+\infty) = 1$ & $F(-\infty) = 0$

Proof:→ $F(+\infty) = P(x \leq +\infty) = 1$

$F(-\infty) = P(x \leq -\infty) = 0$

2. It is a non decreasing function of x:

$x_1 < x_2$ then $F(x_1) \leq F(x_2)$.

Proof:→ Since, $x < x_1$ is subset of $x < x_2$ so, probability of $x \leq x_2$ would be either equal to or greater than probability of $x \leq x_1$.

3. $P(x_1 < x \leq x_2) = F(x_2) - F(x_1)$

Proof:→ Since, Both are mutually exclusive

$\{x \leq x_2\} = \{x < x_1\} \cup \{x_1 < x \leq x_2\}$

So, $P(x \leq x_2) = P(x \leq x_1) + P(x_1 < x \leq x_2)$

So, $F(x_2) = F(x_1) + P(x_1 < x \leq x_2)$

Hence, Proved

Continuous Random Variable

$$P(x \in B) = \int_B f(x)\, dx$$

$$\boxed{P(a \le x \le b) = \int_a^b f(x)\, dx}$$

$$P(X = a) = \int_a^a f(x)\, dx = 0$$

$$F(a) = P(X \in (-\infty, a]) = \int_{-\infty}^{a} f(x)\, dx$$

From fundamental theorem of calculus,

$$\boxed{\frac{d}{da}\left(F(a)\right) = f(a)}$$

Described $f(x)$ in terms of $F(x)$, This $f(x)$ is Probability density function. (P.d.f)

** $P\{x_1 < x(\xi) \leq x_2\} = F_x(x_2) - F_x(x_1)$

$$P\{x_1 < x(\xi) \leq x_2\} = \int_{x_1}^{x_2} f_x(x)\,dx$$

** $P(x \leq x \leq x + \Delta x) \simeq f(x)\,\Delta x$

If Δx is sufficiently small,

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x \leq x \leq x + \Delta x)}{\Delta x}$$

$$\boxed{\begin{gathered} P(x = a) \neq f(a) \\ f(a) = \lim_{\Delta a \to 0} \frac{P(a \leq x \leq a + \Delta a)}{\Delta a} \end{gathered}}$$