

Probability And Random Processes

→ Modules :

M1) Basics Of Probability

M2) Discrete Random Variables

M3) Continuous Random Variables

M4) Tail Bounds & Limit Theorems

M5) Random Processes

→ Module 1:

1) Approach to define probability

6) Counting

2) Probability Space

3) Continuity of Probability

4) Conditional Probability, Independence

5) Bayes Theorem & Total Probability Theorem

→ Classical Approach of Probability :-

- For an event E,

$$P(E) = \frac{\text{no. of outcomes favourable to } E}{\text{total no. of outcomes}}$$

- Issues :

- Fails when the outcomes are not equally likely.
- Fails when the no. of possible outcome is infinite.

→ Frequency Approach of Probability :-

- For an event E, where the experiment has been performed n times,

$$P(E) = \frac{n_E}{n}$$

n_E - no. of times E has occurred

$$\Rightarrow P(E) = \lim_{n \rightarrow \infty} \frac{n_E}{n}$$

- Issues :

- We cannot perform an experiment infinite times.
- The ratio may not converge.

→ Axiomatic Approach of Probability :-

- The approach depends on a set of axioms.
- Probability space is a set represented by 3 entities.

$$P.S = (\Omega, \mathcal{F}, P)$$



Sample Space Event Space Probability law

- Sample Space : Set of all outcomes of the random experiment.
- Event Space : Set of all favorable outcomes
- Probability Law : Function that defines probability.

• Set Theory :-

$$A \setminus B = \{x \in A \text{ st } x \notin B\}$$

$$A \cup B = \{x \in A \text{ or } x \in B\}$$

$$\text{Power} : \left(\bigcup_{i=1}^{\infty} A_i \right)^c = \bigcap_{i=1}^{\infty} A_i^c$$

$$\bigcup_{i=1}^{\infty} A_i = \{x \in A_i \text{ for some } i \in \mathbb{N}\} \rightarrow \text{By defn } \not\subseteq \cup$$

$$\Rightarrow \left(\bigcup_{i=1}^{\infty} A_i \right)^c = \{x \notin A_i \text{ } \forall i \in \mathbb{N}\} \quad \begin{matrix} \text{inverse of the} \\ \text{prev. statement} \end{matrix}$$

$$A_i^c = \{x \notin A_i\}$$

$$\bigcap_{i=1}^{\infty} A_i^c = \{x \in A_i \text{ } \forall i \in \mathbb{N}\} \rightarrow \text{By defn. of } n$$

$$\Rightarrow \bigcap_{i=1}^{\infty} A_i^c = \underline{\underline{\left(\bigcup_{i=1}^{\infty} A_i \right)^c}}$$

• Sample Space (Ω) :

- The elements of Ω , are all the possible outcomes of the random experiment.
- The elements of Ω must be mutually exclusive, (disjoint from each other) and collectively exhaustive (cover all possibility)
- Countably infinite sample space : Tossing a coin until we see tail.

Uncountably infinite sample space : Throwing a dart on a square, $\Omega = \{(x,y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$

• Event Space (\mathcal{F}) :

- An event is a subset of a sample space
- An event space is a set of subsets of Ω that form a

σ -field.

- Axioms of σ -field:

i) $\Omega \in \mathcal{F}$

ii) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (Closure under complements)

iii) $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (Closure under countable unions)

↳ if we take $A_i = \emptyset$ & $i \geq k$, then we can use

(this statement for finite union as well)

↳ Same can be said for \cap , using (ii) and De-Morgan's

Note: $A \Delta B = A \setminus B \cup B \setminus A$

iv) $A, B \in \mathcal{F} \Rightarrow A \Delta B \in \mathcal{F}$ → Not an axiom. Implied by (ii) - (iii)

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

$$A \in \mathcal{F}, B \in \mathcal{F} \Rightarrow A \cap B^c \in \mathcal{F}, \text{ (by } B \cap A^c \in \mathcal{F}) \\ \therefore A \Delta B \in \mathcal{F}$$

• Smallest σ -field with E : $\{\Omega, \emptyset, E, E^c\}$

" " " " $A, B : \{\Omega, \emptyset, A, B, A^c, B^c, A \cup B, A \cup B^c, A \cap B, A \cap B^c, A^c \cup B, A^c \cap B, A^c \cap B^c, A^c \cup B^c, A \Delta B, (A \Delta B)^c\}$

Example: Complete $\{\Omega, \emptyset, \{1\}, \{2, 3\}\}$ for $\Omega = \{1, 2, 3, 4\}$

$$\mathcal{F} = \{\Omega, \emptyset, \{1\}, \{2, 3\}, \{2, 3, 4\}, \{1, 4\}, \{1, 2, 3\}, \{4\}\}$$

- Define the smallest σ -set of X, Y, Z as $\sigma(X, Y, Z)$, then,

$$\sigma(\Omega, \emptyset, A, B) = \sigma(A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c)$$

Disjoint sets that collectively exhaust Ω .

Any union of the 4 sets, have their complements as a union of the other sets.

- Probability Law (P) :- $P : \mathcal{F} \rightarrow \mathbb{R}$ is a probability law if it follows the below axioms.

i) $P(\Omega) = 1$ (Normalization)

ii) $P(E) \geq 0 \quad \forall E \in \mathcal{F}$ (Non-negativity)

iii) If A_1, A_2, \dots are mutually exclusive (disjoint), then,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

Properties:

i) $P(\emptyset) = 0$.

Proof: Let $A_i = \emptyset \quad \forall i$

$$\begin{aligned} P\left(\bigcup_i A_i\right) &= \sum_i P(A_i) \\ &= P(\emptyset) = \sum P(\emptyset) \Rightarrow P(\emptyset) = 0 \end{aligned}$$

Example: Construct a probability law for rolling a die.

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

Defn

$$\left\{ \begin{array}{l} P(\{\text{i}\}) = p_i \quad \forall i \in \{1, 2, 3, 4, 5, 6\} \quad \text{st. } \sum_{i=1}^6 p_i = 1 \quad * p_i \geq 0 \\ P(A) = \sum_{i \in A} P(\{\text{i}\}), \quad A \subseteq \Omega \end{array} \right.$$

$$P(\Omega) = \sum_{i \in \Omega} P(\{\text{i}\}) = \sum_{i \in \Omega} p_i = 1 \quad (\text{Normalization})$$

$$p_i \geq 0 \Rightarrow \sum_{i \in A} p_i \geq 0 \quad \forall A \subseteq \Omega \Rightarrow P(A) \geq 0 \quad \forall A \subseteq \Omega$$

(Non-negativity)

By defn. additivity is satisfied.

ii) $P(A) \leq 1 \quad \forall A \subseteq \Omega$

iii) If $A \subseteq B$, then $P(A) \leq P(B)$

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

v) $P(A) + P(A^c) = 1$

Proof:

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

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

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

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

$$\Rightarrow P(B|A) = P(B) - P(A \cap B)$$

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

iii) If $A \subseteq B$, $P(A \cup B) = P(B)$

$$\Rightarrow P(B) = P(A) + P(B) - P(A \cap B)$$
$$= P(A) + P(B/A) + P(\cancel{A \cap B}) - \cancel{P(A \cap B)}$$

$$\Rightarrow P(B) = P(A) + P(B/A)$$

$$\Rightarrow \underline{\underline{P(B) \geq P(A)}}$$

• Continuity Of Probability :-

- A continuous function has a definite limit for all points in its domain.

- A function $f: \mathbb{R} \rightarrow \mathbb{R}$ if a sequence $x_n \rightarrow x^* \Rightarrow f(x_n) \rightarrow f(x^*)$ as $n \rightarrow \infty$ (Formal defn), ie,

$$\lim_{n \rightarrow \infty} f(x_n) = f\left(\lim_{n \rightarrow \infty} x_n\right)$$

- But, probability law is a function on sets. For such a function, continuity is defined as,

Let $A_1, A_2, A_3, \dots, n \in \mathbb{N}$ be a sequence of events, then

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) = P\left(\lim_{n \rightarrow \infty} \bigcup_{i=1}^n A_i\right)$$

Proof:

Claim 1: Consider 3 sets A_1, A_2, A_3 . Then

$$A_1 \cup A_2 \cup A_3 = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2))$$

As a union of disjoint sets B_1, B_2, B_3 .

$$\Rightarrow B_i = A_i \setminus \bigcup_{k=1}^{i-1} A_k \longrightarrow \bigcap_{i=1}^{\infty} B_i = \emptyset$$

Proof of Claim 1:

To prove, $B_i \cap B_j = \emptyset \quad \forall i \neq j$

WLOG let $i < j$. Let $x \in B_i$

$$\Rightarrow x \in A_i \setminus \bigcup_{k=1}^{i-1} A_k \Rightarrow x \notin \bigcup_{k=1}^{i-1} A_k \quad \& \quad x \in A_i$$

$$\begin{aligned} \text{Assume } x \in B_j &\Rightarrow x \in A_j \setminus \bigcup_{k=1}^{j-1} A_k \\ &\Rightarrow x \notin \bigcup_{k=1}^{j-1} A_k \end{aligned}$$

But we proved $x \in A_i \longrightarrow \text{Contradiction}$

$\therefore x \in B_i \Rightarrow x \notin B_j, \text{ if } i < j$

Similarly can be stated for $i > j$.

$\therefore x \in B_i \Rightarrow x \notin B_j \text{ if } i > j$

$$\Rightarrow \underline{B_i \cap B_j = \emptyset \text{ } \forall i \neq j}$$

$$\underline{\text{Claim 2: } \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i \text{ } \forall n \in \mathbb{N}} \quad \textcircled{1}$$

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i \quad \longrightarrow \textcircled{2}$$

Proof of Claim 2:

For $\textcircled{1}$, if $n = 1$, $B_1 = A_1 \rightarrow$ trivially proved

Assume true for $n = k$, ie, $\bigcup_{i=1}^k A_i = \bigcup_{i=1}^k B_k := C_k$

To prove for $n = k+1$,

$$\bigcup_{i=1}^{k+1} A_i = A_{k+1} \cup C_k$$

$$\begin{aligned} \bigcup_{i=1}^{k+1} B_i &= B_{k+1} \cup C_k = (A_{k+1} \bigcup_{i=1}^k A_i) \cup C_k \\ &= (A_{k+1} \setminus C_k) \cup C_k \end{aligned}$$

$$\Rightarrow \boxed{\bigcup_{i=1}^{k+1} B_i = \bigcup_{i=1}^{k+1} A_i} \Rightarrow \textcircled{1} \text{ is valid.}$$

To extend the equality till ∞ , we can prove that any element in LHS will belong to RHS. (Induction is not valid for $n \rightarrow \infty$)

Final Proof: To prove:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right)$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) \quad \text{As defined and proved above}$$

$$= \sum_{i=1}^{\infty} P(B_i) \quad [\text{Additivity Axiom}]$$

Applies only because summation is defined
of \mathbb{R} and limit is well
- defined for \mathbb{R} .

$$\begin{aligned} &= \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n B_i\right) \quad [\text{Additivity Axiom}] \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) \quad [\text{Claim 2}] \end{aligned}$$

$$\Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \underbrace{\lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right)}$$

Note: A sequence $\{x_n\}$ is said to converge to x^* iff
 $\forall \varepsilon > 0$, $\exists n_0 \in \mathbb{N}$ et $|x_n - x^*| < \varepsilon \quad \forall n > n_0$.

Corollary 1:

i) $A_i \subseteq A_{i+1} \quad \forall i \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n)$

ii) $A_i \supseteq A_{i+1} \quad \forall i \Rightarrow P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n)$

Corollary 2 : Union Bound for infinite events,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

→ For finite union, induction can be used for proof
for infinite, continuity.

→ Conditional Probability :-

• $P(A|B)$ is the probability of A given that B has occurred.

• $P(A|B) \propto P(A \cap B)$ If $P(A \cap B) = 0$, $P(A|B) = 0$

$$= P(A|B) = k P(A \cap B), \text{ if } A=B, P(B|B) = k P(B \cap B)$$
$$\Rightarrow l = k P(B)$$
$$\Rightarrow k = l/P(B)$$

$$\Rightarrow P(A|B) = P(A \cap B) / P(B)$$

→ Independence :

• 2 events A and B are independent if,

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

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

- 3 events A, B, C are independent if,

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

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

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

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

- For n events,

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i) \quad \# I \subseteq \{1, 2, 3, \dots, n\}$$

- A collection of sets is a partition of Ω if they are mutually exclusive and exhaustive.

$$(A_i \cap A_j = \emptyset \quad \forall i \neq j), \quad (\bigcup_{i=1}^n A_i = \Omega)$$

→ Total Probability Theorem :-

Let $\{A_1, A_2, A_3, \dots, A_n\}$ be a partition of Ω , and B be any event. Then,

$$P(B) = \sum_i P(B \cap A_i)$$

$$\Rightarrow P(B) = \sum_i P(B|A_i)P(A_i), \quad \text{if } P(A_i) \neq 0 \quad \forall i$$

Proof:

$$B = \bigcup_{i=1}^n (B \cap A_i)$$

$$\Rightarrow P(B) = P\left(\bigcup_{i=1}^n (B \cap A_i)\right)$$

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

\rightarrow Bayes' Theorem :-

$\{A_1, A_2, A_3, \dots, A_n\}$ be a partition of Ω . B be any event.

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

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

\rightarrow Multiplication Rule :

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

\rightarrow Conditional Independence :

2 events A and B are said to be independent given C if,

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

- If $P(C) = 1$, then A and B are truly independent, ie,

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

→ Counting Techniques :-

Suppose there are n objects

i) No. of k -length sequences : ${}^n P_k = \frac{n!}{(n-k)!}$

ii) No. of k -length sets : ${}^n C_k = \frac{n!}{k!(n-k)!}$

iii) No. of ways of choosing r sets of sizes $n_1, n_2, n_3, \dots, n_r$,

$$= {}^n C_{n_1} \cdot {}^{n-n_1} C_{n_2} \cdot {}^{n-n_1-n_2} C_{n_3} \cdots {}^{n-n_1-n_2-\cdots-n_{r-1}} C_{n_r}$$

$$= \frac{n!}{(n-n_1)! n_1!} \frac{(n-n_1)!}{(n-n_1-n_2)! n_2!} \cdots \cdots$$

$$= \frac{n!}{n_1! n_2! n_3! \cdots n_r!}$$

→ Random Variables :-

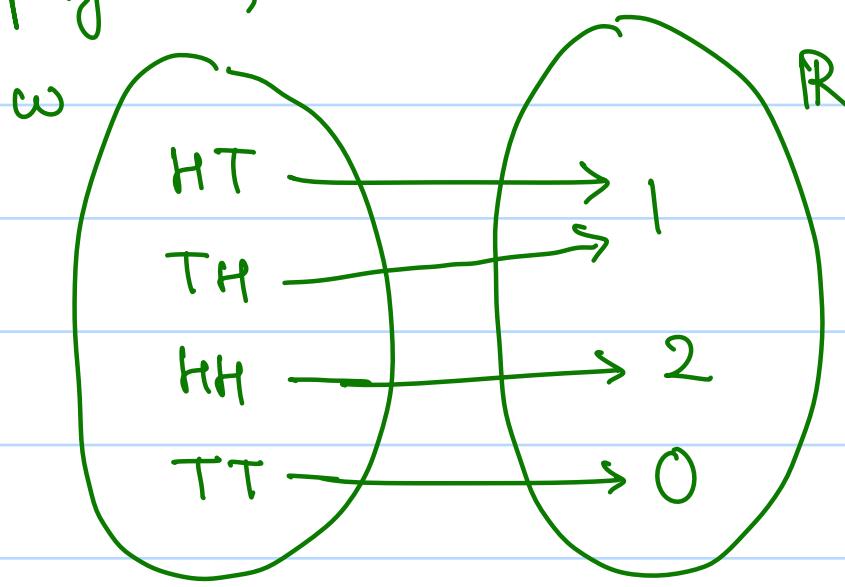
° A random variable is a function $X: \Omega \mapsto \mathbb{R}$ such that,

$$\{X \leq x\} \triangleq \{\omega \mid X(\omega) \leq x\} \in \mathcal{F}, \forall x \in \mathbb{R}$$

Example: $\Omega = \{\text{HT, TH, HH, TT}\}$

$X: \Omega \mapsto \mathbb{R}$ st $X(\omega) = \text{No. of heads}$

The mapping is,



$$\{x \leq x_0\} = \begin{cases} \emptyset, & x_0 < 0 \\ \{\text{TT}\}, & x_0 \in [0, 1) \\ \{\text{HT, TH, TT}\}, & x_0 \in [1, 2) \\ \Omega, & x_0 \geq 2 \end{cases}$$

Theorem: $X: \Omega \rightarrow \mathbb{R}$ on \mathcal{F} , then

i) $X^{-1}((-\infty, x]) \in \mathcal{F}$

ii) $X^{-1}([x_1, x_2]) \in \mathcal{F} \quad \forall x_1, x_2 \in \mathbb{R}$

iii) $X^{-1}(\{x\}) \in \mathcal{F} \quad \forall x \in \mathbb{R}$

Proof of (i):

wkt $X^{-1}(-\infty, x] \in \mathcal{F} \quad \forall x \in \mathbb{R}$

$$(-\infty, x) = \bigcup_{n=1}^{\infty} \left(-\infty, x - \frac{1}{n} \right]$$

$$X^{-1}\left(\left(-\infty, x - \frac{1}{n}\right]\right) \in \mathcal{F} \quad \forall n \Rightarrow X^{-1}\left(\bigcup_{n=1}^{\infty} \left(-\infty, x - \frac{1}{n} \right]\right) \in \mathcal{F}$$

Proof of (ii) :

$$\text{wkt } x^{-1}(-\infty, x_1] \in \mathcal{F}$$

$$x^{-1}(-\infty, x_2) \in \mathcal{F}$$

Since subtraction is closed in \mathcal{F} , $x^{-1}(-\infty, x_2) \in \mathcal{F}$

$$\Rightarrow x^{-1}([x_2, \infty)) \in \mathcal{F}$$

$$\Rightarrow x^{-1}(-\infty, x_1] \cap x^{-1}([x_2, \infty)) \in \mathcal{F}$$

$$\Rightarrow x^{-1}([x_2, x_1]) \in \mathcal{F}$$

iv) $x^{-1}((x_1, x_2)) \in \mathcal{F}$

$$(x_1, x_2) = (x_1, \infty) \cap (-\infty, x_2)$$

$$x^{-1}(-\infty, x_1] \in \mathcal{F} \Rightarrow x^{-1}(x_1, \infty) \in \mathcal{F}$$

$$x^{-1}([x_2, \infty)) \in \mathcal{F} \Rightarrow x^{-1}(-\infty, x_2) \in \mathcal{F}$$

$$\Rightarrow x^{-1}((x_1, \infty) \cap (-\infty, x_2)) \in \mathcal{F}$$

$$\Rightarrow x^{-1}((x_1, x_2)) \in \mathcal{F}$$

• Borel σ -Algebra :

Smallest σ -field on \mathbb{R} containing sets of the form $(-\infty, x]$ $\forall x \in \mathbb{R}$, ie,

$$B(\mathbb{R}) = \sigma((-\infty, x] \forall x \in \mathbb{R})$$

- Contains all possible subsets of \mathbb{R} .

• Cumulative Distribution Function (CDF) :-

- A CDF is a function $F_X : \mathbb{R} \mapsto [0,1]$ such that,

$$F_X(x) = P(\{X < x\})$$

Example: For,

$$\{X \leq x\} = \begin{cases} \emptyset, & x < 0 \\ \{\text{TT}\}, & x \in [0,1) \\ \{\text{TT, HT, TH}\}, & x \in [1,2) \\ \Omega, & x \geq 2 \end{cases}$$

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & x \in [0,1) \\ \frac{3}{4}, & x \in [1,2) \\ 1, & x \geq 2 \end{cases}$$

Example: Let $X(\omega) = c$, $c \in \mathbb{R}$. Then,

$$\{X \leq x\} = \begin{cases} \emptyset, & x < c \\ \Omega, & x \geq c \end{cases}$$

$$\Rightarrow F_X(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}$$

• Indicator R.V :-

Consider an $A \in \mathcal{F}$. Then $I_A: \Omega \rightarrow \mathbb{R}$ such that,

$$I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

$$\Rightarrow \{I_A < x\} = \begin{cases} \emptyset, & x < 0 \\ A^c, & x \in [0,1) \\ \Omega, & x \geq 1 \end{cases}$$

$\{I_A < x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}$. So I_A is a valid r.v.

I_A is an indicator r.v for the event A .

• $B_1 \cap B_2 = \emptyset$, then,

$$I_{B_1 \cup B_2}(\omega) = I_{B_1}(\omega) + I_{B_2}(\omega) \quad \forall \omega \in \Omega$$

B_1, B_2, B_3, \dots form a partition of Ω , then,

$$I_{\bigcup_{i=1}^n B_i} = \sum_{i=1}^n I_{B_i} = 1$$

• Theorems of CDF :-

i) $\lim_{x \rightarrow \infty} F_X(x) = 1, \quad \lim_{x \rightarrow -\infty} F_X(x) = 0$

$$\lim_{x \rightarrow \infty} F_X(x) = \lim_{x \rightarrow \infty} P_X(X \leq x) = P(\Omega) = 1$$

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{x \rightarrow -\infty} P_X(X \leq x) = P(\emptyset) = 0$$

P_X - Probability law of X

$$2) x < y \Rightarrow F_X(x) \leq F_X(y)$$

3) $F_X(x)$ is always right continuous, i.e.,

$$\lim_{\epsilon \rightarrow 0^+} F_X(x + \epsilon) = F_X(x) \quad \forall x \in \mathbb{R}$$

Proof:

$$\lim_{n \rightarrow \infty} F_X\left(x + \frac{1}{n}\right) = \lim_{n \rightarrow \infty} P\left(X \leq x + \frac{1}{n}\right)$$

Let $\{X \leq x + \frac{1}{n}\} = B_n \Rightarrow B_n$ is a decreasing sequence

By continuity of probability $P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} P(B_n)$

$$\Rightarrow \lim_{n \rightarrow \infty} F_X\left(x + \frac{1}{n}\right) = P\left(\bigcap_{n=1}^{\infty} \{X \leq x + \frac{1}{n}\}\right)$$

$$= P(\{X \leq x\}) = F_X(x)$$

$$\therefore \lim_{n \rightarrow \infty} F_X\left(x + \frac{1}{n}\right) = F_X(x)$$

$$\Rightarrow \underset{\longrightarrow}{\text{RHL}} = F_X(x)$$

$$4) F_X(x) - \lim_{\epsilon \rightarrow 0^+} F_X(x - \epsilon) = P(\{X = x\})$$

$$5) P_X(x_1 \leq X \leq x_2) = P(\omega : x_1 \leq X(\omega) \leq x_2) \\ = P(X$$

\rightarrow Discrete Random Variable :-

A random variable is discrete if Range(X) $\subseteq \mathbb{R}$ is finite / countably infinite.

• Probability Mass Function :

$P_X : X \mapsto [0,1]$, given by

$$P_X(x) = P(X = x) = P(\{\omega : X(\omega) = x\})$$

- Lemma: For an rv st $x \in \{x_1, x_2, x_3, \dots\}$

$$i). \sum_{i=1}^{\infty} P_X(x_i) = \sum_{i=1}^{\infty} P(X = x_i)$$

$$= P\left(\bigcup_{i=1}^{\infty} \{X = x_i\}\right) \quad [\text{Additivity}]$$

$$= P(\{X \in \{x_1, x_2, x_3, \dots\}\}) = P(\Omega)$$

$$\therefore \sum_{i=1}^{\infty} P_X(x_i) = \underline{\underline{1}}$$

• CDF of a Discrete R.V :-

$$F_X(x) = P(\{X \leq x\}) = P\left(\bigcup_{i \in \mathbb{N}: x_i \leq x} \{X = x_i\}\right)$$

$$\Rightarrow F_X(x) = \sum_{i \in \mathbb{N}: x_i \leq x} P_X(x_i)$$

• Functions on R.V :-

Let $X: \Omega \rightarrow \mathbb{R}$ and $Y: \mathbb{R} \rightarrow \mathbb{R}$, ie
a function on X . Y is also a r.v.

$$\text{Proof: } Y^{-1}(B) = \{y : Y(y) \in B\} \quad \forall B \in \mathcal{B}(\mathbb{R})$$

$$= \{X(\omega) : Y(X(\omega)) \in B\}$$

$$X^{-1}(Y^{-1}(B)) = \{x : Y(X(x)) \in B\}$$

$$= \{\omega : X(\omega) \in Y^{-1}(B)\}$$

Since $X^{-1}(s) \in \mathcal{F}$ $\forall s \in \mathcal{B}(\mathbb{R})$

$$Y^{-1}(B) \in \mathcal{X}(\mathcal{F}) = \mathcal{F}' \quad \forall B \in \mathcal{B}(\mathbb{R})$$

Better proof
needed.

$\Rightarrow Y$ is a valid r.v

• PMF of y : $P_y(Y = y) = \sum_{Y(\omega)=y} P_X(X = \omega)$

• Expectation :

$$E[X] = \sum_{x \in \mathbb{R}} x P_X(x)$$

• If $y = Y(x)$,

Law of the Unconscious Statistician : $E[Y] = \sum_{x \in \mathbb{R}} y(x) P_X(x)$

Proof: $E[Y] = \sum_{y \in Y} y P_Y(y) = \sum_{y \in Y} y \sum_{Y(\omega)=y} P_X(\omega)$

$$= \sum_{y \in Y} \sum_{Y(\omega)=y} y(\omega) P_X(\omega) = \sum_{x \in \mathbb{R}} y(x) P_X(x)$$

All possible x ←

◦ Variance:

$$\text{Var}[X] = E[(X - E[X])^2]$$

$$\Rightarrow \text{Var}[X] = E[X^2] - E[X]^2$$

◦ n^{th} Moment of an RV = $E[X^n]$. (defn)

◦ Examples of RV :-

1) Bernoulli RV: Binary opp. RV. (ex: Coin Toss)

$$P(\{H\}) = p, P(\{T\}) = 1-p$$

$$X(H) = 1, X(T) = 0 \longrightarrow X \text{ is a Bernoulli R.V}$$

$$E[X] = p, \text{Var}[X] = p - p^2$$

2) Binomial RV: (ex: Coin Tossed n times)

Any event ω : Sequence of H's and T's of length n .

$$P(\{H\}) = p, P(\{T\}) = 1-p. X(\omega) = \text{No. of heads in } \omega$$

$$\Rightarrow P_X(k) = {}^n C_k p^k (1-p)^{n-k} \quad \forall k \in \mathbb{N}$$

$$E[X] = np \quad \text{Var}[X] = np(1-p)$$

3) Geometric RV: (ex: Toss a coin till heads)

$X(\omega)$ = No. of coin tosses in ω , to get a head ;

$$P_X(k) = p(1-p)^{k-1} \quad E[X] = \frac{1}{p} \quad \text{Var}[X] = \frac{1-p}{p^2}$$

4) Poisson RV: $X \in \{0, 1, 2, \dots\}$

$$P_X(k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots \quad \lambda \in \mathbb{R}$$

$$E[X] = \text{Var}[X] = \lambda.$$

• Let $Y \sim \text{Binomial}(n, p)$, As $n \rightarrow \infty$ st $np = \lambda$, a constant, we have,

$$\lim_{n \rightarrow \infty} P_Y(k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, 3, \dots$$

Proof:

$$\lim_{n \rightarrow \infty} P_Y(k) = \lim_{n \rightarrow \infty} {}^n C_k p^k (1-p)^{n-k}$$

$$= \lim_{n \rightarrow \infty} \frac{n!}{(n-k)! k!} p^k (1-p)^{n-k}$$

$$= \lim_{n \rightarrow \infty} \frac{n!}{(n-k)! k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2) \cdots (n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left(\frac{n}{n} \right) \left(\frac{n-1}{n} \right) \left(\frac{n-2}{n} \right) \dots \left(\frac{n-k+1}{n} \right)^1$$

$$\left(1 - \frac{\lambda}{n} \right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n} \right)^{n-k}$$

$$= \frac{\lambda^k}{k!} e^{\lim_{n \rightarrow \infty} \left(\frac{-\lambda}{n} \right) (n-k)} \quad (1 + f(n))^{g(n)}$$

$$= \frac{\lambda^k}{k!} e^{\lim_{n \rightarrow \infty} -\lambda + \frac{\lambda k}{n}} \quad f(n) \rightarrow 0$$

$$g(n) \rightarrow \infty$$

as $n \rightarrow \infty$,

then

$$\Rightarrow \lim_{n \rightarrow \infty} P_Y(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$e^{\lim_{n \rightarrow \infty} g(n) \ln(1 + f(n))}$$

$$= e^{\lim_{n \rightarrow \infty} g(n) f(n)}$$

→ Multiple Random Variables :-

- (X, Y) are said to be jointly discrete if (x, y) takes values in some countable subset of \mathbb{R}^2

$$\text{Joint PMF } P_{X,Y}(x,y) = P(X=x \cap Y=y)$$

$$P_X(x) = \sum_{y \in Y} P_{X,Y}(x,y)$$

$$\begin{aligned} \text{Proof: } P_X(x) &= P(X=x) \\ &= P(X=x \cap \bigcup_{y \in Y} (Y=y)) \\ &= P\left(\bigcup_y (X=x \cap Y=y)\right) \\ &= \sum_{y \in Y} P(X=x \cap Y=y) = \sum_{y \in Y} P_{X,Y}(x,y) \end{aligned}$$

- $P((X,Y) \in B) = \sum_{(x,y) \in B} P_{XY}(x,y) \quad \forall B \subseteq \mathbb{R}^2$

- Functions on 2 Rv's :-

If $X: \Omega \rightarrow \mathbb{R}$, $Y: \Omega \rightarrow \mathbb{R}$ Rv's, $Z = g(X, Y)$,
 $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ is also a Rv.

$$P_Z(z) = \sum_{x,y : g(x,y) = z} P_{XY}(x,y)$$

- Independence :-

2 Rv's X and Y are said to be independent if

$$P_{XY}(x,y) = P_X(x) P_Y(y) \quad \forall x, y \in X, Y$$

- If $x, y \in \{0,1\}$, then X, Y are independent.

- If X, Y are independent, then $E[XY] = E[X]E[Y]$

- Independent \Leftrightarrow Uncorrelated

- n Rv's $X_1, X_2, X_3 \dots X_n$ are said to be independent if,

$$P_{X[1:n]}(x_{[1:n]}) = \prod_{i \in [1:n]} P_{X_i}(x_i)$$

- If X and Y are independent, $h(X)$ and $g(Y)$ are independent

• Let $Z = g(X, Y)$ be a r.v, then,

$$E[Z] = \sum_{x,y} g(x,y) P_{XY}(x,y)$$

→ Properties of $E[X]$ and $\text{Var}[X]$:-

$$1) Y = aX + b \Rightarrow E[Y] = aE[X] + b$$

$$\text{Var}[Y] = a^2 \text{Var}[X]$$

$$2) Y = X_1 + X_2 \Rightarrow E[Y] = E[X_1] + E[X_2] \rightarrow \text{Prove by LOTUS}$$

$$\Rightarrow E[\sum X_i] = \sum E[X_i]$$

$$\Rightarrow \text{Var}[Y] = \text{Var}[X_1] + \text{Var}[X_2] + 2 \text{Cov}[X_1, X_2]$$

Where,

$$\text{Cov}[X_1, X_2] = E[X_1 \cdot X_2] - E[X_1]E[X_2]$$

Definition: Correlation Coefficient = $\rho(X_1, X_2)$

$$= \frac{\text{Cov}[X_1, X_2]}{\sqrt{\text{Var}[X_1]} \sqrt{\text{Var}[X_2]}}$$

Theorem: $|\rho(X_1, X_2)| \leq 1$. Equality iff $X_1 = aX_2 + b$, $a, b \in \mathbb{R}$
ie X_1 and X_2 are linearly dependent.

Proof: By Cauchy-Schwarz Inequality,

$$E[XY] \leq \sqrt{E[X^2]E[Y^2]}$$

Proof of Cauchy-Schwarz,

$$E[(x - \alpha y)^2] \geq 0 \quad (\text{Equality if } x = \alpha y \text{ (lin. dep)})$$

$$\rightarrow E[x^2] - 2\alpha E[xy] + \alpha^2 E[y^2] \geq 0 \rightarrow \text{To a Quadratic in } \alpha$$

$$\rightarrow \sqrt{(2E[xy])^2 - 4(E[y^2])(E[x^2])} \leq 0 \rightarrow \text{Discriminant} < 0$$

$$\Rightarrow 4E[xy]^2 - 4E[y^2]E[x^2] \leq 0$$

$$\Rightarrow E[xy] \leq \sqrt{E[y^2]E[x^2]}$$

To Prove the theorem,

$$x \longrightarrow (x - E[x]) \quad y \longrightarrow (y - E[y])$$

$$\rightarrow E[(x - E[x])(y - E[y])] \leq \sqrt{E[(x - E[x])^2]E[(y - E[y])^2]}$$

$$\Rightarrow |\text{Cov}[x, y]| \leq \sqrt{\text{Var}[x]\text{Var}[y]}$$

$$\Rightarrow \rho[x, y]^2 \leq 1 \Rightarrow |\rho[x, y]| \leq 1$$

$$3) \quad X, Y, \text{ st } Z = X + Y$$

$$P_Z(z) = \sum_{\substack{x, y \\ \text{st } x+y \in Z}} P_{XY}(x, y) = \sum_x P_{XY}(x, z-x)$$

$$= \sum_y P_{XY}(z-y, y)$$

$$\Rightarrow \underline{P_X * P_Y} \quad \text{If } X, Y \text{ are independent}$$

Example: X, Y are Geometric RVs. Find $P_Z(z)$ where $Z = X + Y$
if X, Y are independent.

$$\begin{aligned}
 P_Z(z) &= P_X(x) * P_Y(y) \\
 &= \sum_x P_X(x) P_Y(z-x) \\
 &= \sum_x p(1-p)^{x-1} q(1-q)^{z-x-1} \\
 &= pq \frac{(1-q)^{z-1}}{(1-p)} \sum_x \left(\frac{1-p}{1-q}\right)^x
 \end{aligned}$$

If $p = q$,

$$P_Z(z) = (z-1) p^2 (1-p)^{z-2}$$

→ Conditioning :-

Conditioning on RV $X: \Omega \rightarrow \mathbb{R}$ on an event $A \subseteq \Omega$

$$P_{X|A}(x) \triangleq P(X=x|A)$$

$$P_{X|A}(x) = \frac{P(X=x, A)}{P(A)}$$

° Theorem: If A_1, A_2, \dots, A_n forms a partition in Ω , then

$$P_X(x) = \sum_i P_{X|A_i}(x) P(A_i)$$

- Conditioning on Rv X on another Rv Y

$$P_{X|Y}(x) = P(X=x | Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

$$\Rightarrow P_{X|Y}(x) = \frac{P_{XY}(x,y)}{P_Y(y)}$$

- Conditional Expectance :-

$$E[X|Y=y] = \sum_{x \in X} x P_{X|Y}(x|y)$$

$$E[X|Y] = \sum_{y \in Y} \sum_{x \in X} x P_{X|Y}(x|y)$$

$$E[X|A] = \sum_{x \in X} x P_{X|A}(x) \quad \# A \subseteq \Omega$$

$$E[g(x)|A] = \sum_{x \in X} g(x) P_{X|A}(x) \quad \# A \subseteq \Omega$$

- Total Expectation Theorem :-

If A_1, A_2, \dots, A_n form a partition of Ω , with $P(A_i) > 0 \ \forall i$, then

$$E[X] = \sum_{i=1}^n E[X|A_i] P(A_i)$$

Proof: $\sum_{i=1}^n E[X|A_i] P(A_i)$

$$= \sum_{i=1}^n \sum_{x \in X} x P_{X|A_i}(x) P(A_i)$$

$$= \sum_{x \in X} x \sum_{i=1}^n P_{X|A_i}(x) P(A_i)$$

$$= \sum_{x \in X} x \sum_{i=1}^n \frac{P(X=x \cap A_i)}{P(A_i)}$$

$$= \sum_{x \in X} x \sum_{i=1}^n P(X=x \cap A_i)$$

$$= \sum_{x \in X} x P(X=x)$$

$$= \sum_{x \in X} x P_X(x)$$

$$= \underline{\underline{E[X]}}$$

III by $E[X] = \sum_y E[X|y=y] P_Y(y)$

- Conditional Expectance as a RV :-

$$\phi(y) \triangleq E[X|y=y] = \sum_{x \in X} x P_{X|Y}(x|y)$$

$$\phi : \text{Range}(y) \mapsto \mathbb{R}$$

$\phi(y)$ is a function of an RV. Therefore it is also an RV.

- Theorem: $E[\phi(y)] = E[X]$

$$E[\phi(y)] = \sum_{y \in Y} \phi(y) P_Y(y)$$

$$= \sum_{y \in Y} E[X|Y=y] P_Y(y)$$

$$= E[X] \quad (\text{Total Expectation Theorem})$$

- Conditional Independence of Rvs :-

X and Y are conditionally independent given A with $P(A) > 0$

if

$$P_{X,Y|A}(x,y) = P_{X|A}(x) P_{Y|A}(y)$$

- Conditional Variance :

$$\text{Var}[X|Y=y] = E[(X - E[X|Y=y])^2 | Y=y]$$

Let $\Psi(y) = \text{Var}[X|Y=y]$, $\Psi(y)$ is an Rv.

$$\text{Var}[X|Y=y] = E[X^2|Y=y] - (E[X|Y=y])^2 \rightarrow \begin{matrix} \text{Same Proof as} \\ \text{in Non-Condition} \\ - \text{al Case} \end{matrix}$$

- Law Of Total Variance :-

$$\text{Var}[X] = E[\Psi(Y)] \neq \text{Var}[\Phi(Y)]$$

$$E[\Psi(Y)] = \sum_{y \in Y} \Psi(y) P_Y(y) \quad (\text{LOTUS})$$

$$= \sum_{y \in Y} (E[X^2|Y=y] - E[X|Y=y]^2) P_Y(y)$$

$$= \sum_{y \in Y} E[x^2 | y=y] P_Y(y) - \sum_{y \in Y} E[x | y=y]^2 P_Y(y)$$

$$\sum_{y \in Y} E[x^2 | y=y] P_Y(y) = E[x^2] \quad (\text{LOTUS})$$

$$\sum_{y \in Y} E[x | y=y]^2 P_Y(y) =$$

\rightarrow Memoless Property of a Geometric RV :-

$$P_X(k) = (1-p)^{k-1} p, \quad k \in \mathbb{N}$$

$$\begin{aligned} P(X > n) &= \sum_{k=n+1}^{\infty} (1-p)^{k-1} p \\ &= p \frac{(1-p)^{n+1}}{1-(1-p)} = \underline{(1-p)^n} \end{aligned}$$

Property: $P(X > m+n | X > m) = P(X > n)$

$$P(X > m+n | X > m) = \frac{P(X > m+n \cap X > m)}{P(X > m)}$$

$$= \frac{P(X > m+n)}{P(X > m)}$$

$$= \frac{(1-p)^{m+n}}{(1-p)^m}$$

$$= \underline{(1-p)^n}$$

• CDF of an RV is given by,

$$F_X(x) = P(X < x) = \sum_{x_i \leq x} P_X(x_i)$$

$$\Rightarrow P_X(x) = F_X(x) - F_X(x-1)$$

Example: $X = \max\{X_1, X_2, X_3\}$, X_1, X_2, X_3 are independent.

$$P_{X_i}(k) = \frac{1}{100} \quad \forall i \in \{1, 2, 3\}, \quad \forall k \in [1:10]$$

$$P_X(k) = P(X = k) = P(\max\{X_1, X_2, X_3\} = k)$$

$$F_X(k) = P(X \leq k) = P(\max\{X_1, X_2, X_3\} \leq k)$$

