

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

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

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

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

• 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.

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

→ Total Probability Theorem :-

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

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

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

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 :-

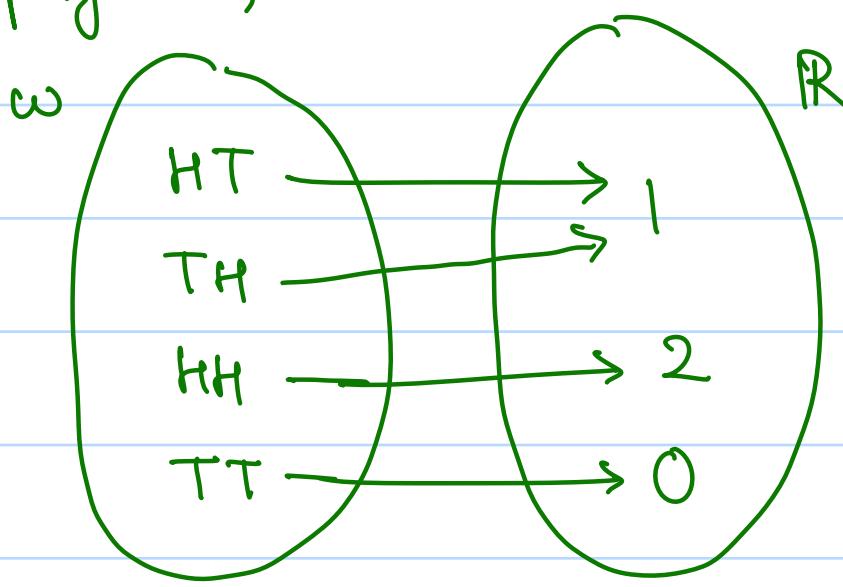
° A random variable is a function $X: \Omega \mapsto \mathbb{R}$ such that,

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

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

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

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

Proof: $P_X(x) = P(X=x)$

$$= P(X=x \cap \bigcup_{y \in Y} (Y=y))$$

$$= P(\bigcup_y (X=x \cap Y=y))$$

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

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

- 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 rv, 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,

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

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

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

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

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

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.

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

- 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.

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

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

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

→ Application of Discrete RVs :-

- Minimum Mean Square Error Estimation :-

$$X \sim P_X \longrightarrow P_{Y|X} \longrightarrow Y \xrightarrow{f} \hat{X} = f(Y)$$

- Theorem: The function $\phi(y) = E[X|Y=y]$, gives the min. value of $E[(X-f(y))^2]$.

- Proof:

$$E[(X-f(y))^2] = E[E[(X-f(y))^2|Y]]$$

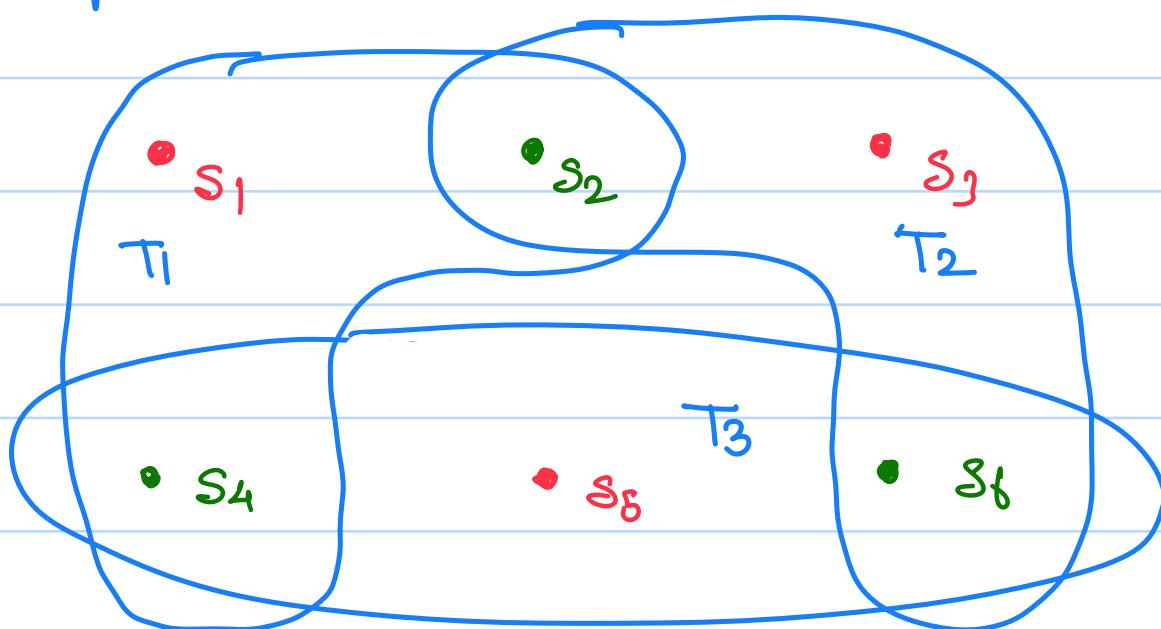
$$\begin{aligned}
 &= \sum_y P_y(y) E[(X - f(y))^2 | Y = y] \\
 &= \sum_y P_y(y) E[x^2 + f(y)^2 - 2xf(y) | Y = y] \\
 &= \sum_y P_y(y) (E[x^2 | Y = y] + f(y)^2 - 2E[X | Y = y]f(y))
 \end{aligned}$$

Each of the terms in the summation is minimized when,

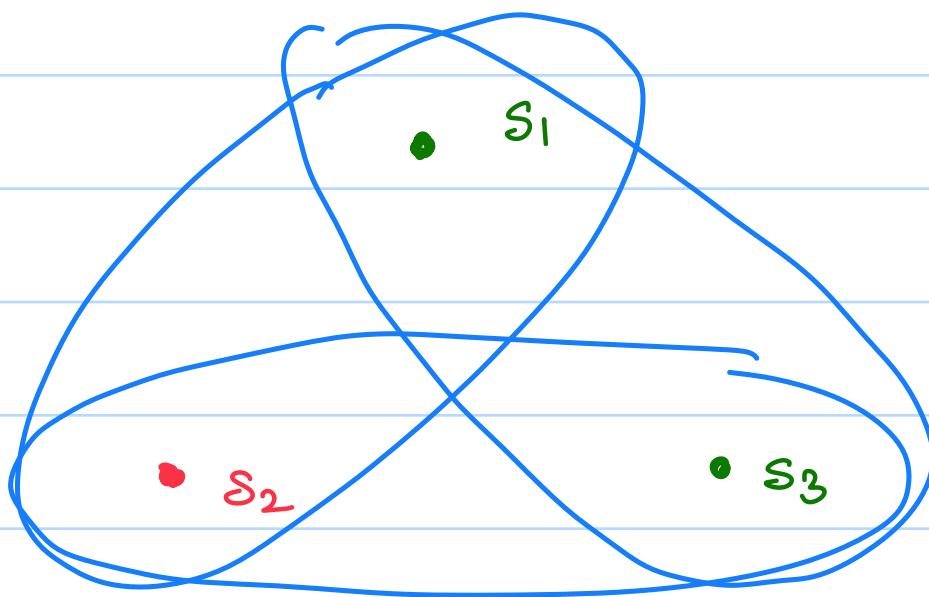
$$f(y) = \phi(y) = E[X | Y = y]$$

o Combinatorics and Graph Theory :-

Let S be a set of some elements and $T_1, T_2, T_3 \dots, T_m \subseteq S$ st. $n(T_i) = l + i \in [1:m]$, where $l \in \mathbb{N}$, then 2-coloring of S is coloring each element of S such that $T_i \neq$ Monochromatic $\forall i \in [1:m]$, i.e., T_i has elements of both colors $\forall i \in [1:m]$.



In the above S , 2-coloring is possible.



In the above S , 2-coloring is not possible.

- Theorem:

If $m < 2^{d-1}$, then \exists a valid 2 coloring of the set S .

- Proof: Let $S = \{x_1, x_2, x_3 \dots, x_n\}$. Randomly color each element of S black or white, independently and identically distributed, each with probability $\frac{1}{2}$.

Let E_i be the event that T_i is monochromatic.

$\Omega = \{W, B\}^n$, i.e., each element of Ω, w , is an n -length sequence of W and B.

$$P(\{w\}^n) = P(\{B\}^n) = \frac{1}{2^n}$$

$$\Rightarrow P(E_i) = \frac{1}{2^d} \times 2 = \frac{1}{2^{d-1}}$$

$$P(\exists \text{ monochromatic } T_i) = P(\bigcup_{i=1}^m E_i) = \sum_{i=1}^m P(E_i)$$

$$= \sum_{i=1}^m \frac{1}{2^{d-1}} = \frac{m}{2^{d-1}}$$

$$\frac{m}{2^{d-1}} < 1 \Rightarrow m < 2^{d-1}$$

$$P(\text{no monochromatic } Ti) = 1 - \frac{m}{2^{d-1}} > 0 \text{ if } m < 2^{d-1}$$

$P(\text{no monochromatic } Ti) > 0 \Rightarrow \exists \omega \in \Omega \text{ st the associated coloring is a valid 2-coloring.}$

[Let $A = \{\omega \in \Omega : \omega \text{ satisfies } E\}$. If $P(A) > 0$, then $\exists \omega \in \Omega \text{ st } \omega \text{ satisfies } E$] \rightarrow Probabilistic Method

- The converse of the above theorem is false.

- Entropy (Uncertainty) :-

- Consider an RV with,

$$P_X(0) = P_X(1) = \frac{1}{2} (X_1)$$

and another RV with,

$$P_X(0) = 0.9, P_X(1) = 0.1 (X_2)$$

Intuitively, we see that the uncertainty of realization in X_1 is higher than X_2 . This uncertainty is measured by the Entropy of X .

$$H(X) = \sum_{x \in X} P_X(x) \log \frac{1}{P_X(x)}$$

- If $X \sim \text{Binomial RV}$,

$$H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} := \underline{h(p)}$$

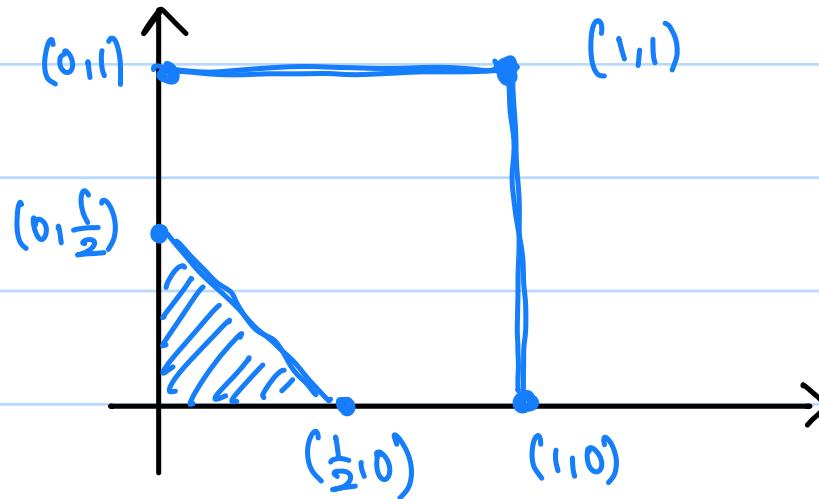
→ Continuous Random Variable :-

- We know that an RV is a map from (Ω, \mathcal{F}, P) to $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ such that,

$$\{ \omega : X(\omega) \leq x \} \in \mathcal{F} \quad \forall x \in \mathbb{R}.$$

- Now, if Ω is an uncountably infinite set,

Take $\Omega = [0,1]^2$, ie, Ω is a unit square.



Let $P(A) = \text{area}(A) \quad \forall A \subseteq [0,1]^2$

$$P((x,y) : x+y \leq \frac{1}{2}) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

$$P((0.4, 0.5)) = 0, \quad P(\Omega) = 1$$

$$\text{But } \Omega = \bigcup_{x,y \in [0,1]} \{(x,y)\}$$

$$\Rightarrow P(\Omega) = P\left(\bigcup_{x,y \in [0,1]} \{(x,y)\}\right)$$

$$= \sum_{x,y \in [0,1]} P(\{(x,y)\}) = \underline{0} \neq 1$$

\therefore Additivity only holds for countable number of disjoint sets.

- Definition:

A random variable X is said to be continuous if its CDF can be expressed as,

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad \forall x \in \mathbb{R}$$

for some integrable function $f_X(x) : \mathbb{R} \mapsto [0, \infty)$ called the probability density function (PDF) of X .

$$f_X(x) = F_X'(x)$$

$$P(x \in [a,b]) = \int_a^b f_X(u) du$$

} [Fundamental Theorem of Calculus]

$f_X(x)$ directly does not give us probability. We can think of $f_X(x) \Delta x$ as the probability of $(x, x + \Delta x)$, which becomes probability of x for small Δx .

- Theorem: If a continuous Rv X has a PDF f_x , then

$$a) \int_{-\infty}^{\infty} f_x(x) dx = 1$$

$$b) P(X=x) = 0 \quad \forall x \in \mathbb{R}$$

- Proof:

$$a) \int_{-\infty}^{\infty} f_x(x) dx = P(x \in (-\infty, \infty)) = P(\Omega) = \underline{\underline{1}}$$

$$b) P(X=x) = \int_x^x f_x(x) dx = \underline{\underline{0}}$$

- Expectation:

The expectation of a continuous Rv X with PDF f_x
is given by,

$$E[X] = \int_{-\infty}^{\infty} x f_x(x) dx$$

- If X is a continuous Rv, then $g(X)$ can be either a discrete or a continuous Rv, depending on the definition of g .

- Theorem:

If X and $g(X)$ are continuous Rv's, then,

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_x(x) dx \quad [\text{LOTUS}]$$

• Proof:

Lemma: For a non-negative continuous RV Y , (ie $f_Y(y) = 0 \forall y < 0$)

$$E[Y] = \int_0^\infty P(Y > y) dy$$

$$\begin{aligned} \text{Proof: } \int_0^\infty P(Y > y) dy &= \int_0^\infty \int_y^\infty f_Y(t) dt dy \\ &= \int_{t=0}^\infty \int_{y=0}^t f_Y(t) dy dt \\ &= \int_{t=0}^\infty f_Y(t) \left(\int_{y=0}^t dy \right) dt \\ &= \int_{t=0}^\infty t f_Y(t) dt \\ &= \underline{\underline{E[Y]}} \end{aligned}$$

Proving the theorem assuming $g(x)$ is non-negative,

$$\begin{aligned} E[g(x)] &= \int_0^\infty P(g(x) > y) dy \\ &= \int_0^\infty P(X \in \{x : g(x) > y\}) dy \\ &= \int_{y=0}^\infty \int_{x: g(x) > y} f_X(x) dx dy \\ &= \int_{x: g(x) > y} f_X(x) \int_{y=0}^{g(x)} dy dx \end{aligned}$$

$$= \int_{x: g(x) > y} g(x) f_x(x) dx$$

$$= \int_{-\infty}^{\infty} g(x) f_x(x) dx$$

For any general $g(x)$,

$$\text{Prove by showing that } E[Y] = \int_0^\infty P(Y > y) dy - \int_0^\infty P(Y < -y) dy$$

$$\int_0^\infty P(Y > y) dy - \int_0^\infty P(Y < -y) dy$$

$$= \int_0^\infty \int_y^\infty f_y(t) dt dy - \int_0^\infty \int_{-\infty}^{-y} f_y(t) dt dy$$

$$= \int_0^\infty f_y(+t) \int_0^t dy dt - \int_{-\infty}^0 f_y(+t) \int_0^{-t} dy dt$$

$$= \int_0^\infty f_y(t) t dt - \int_{-\infty}^0 f_y(t) (-t) dt$$

$$= \int_0^\infty f_y(t) \cdot t dt + \int_{-\infty}^0 f_y(t) \cdot t dt = \int_{-\infty}^\infty t f_y(t) dt$$

$$= E[Y]$$

$$\therefore E[Y] = \int_0^\infty P(Y > y) dy - \int_0^\infty P(Y < -y) dy$$

$$E[g(x)] = \int_0^\infty P(g(x) > x) dx - \int_0^\infty P(g(x) < -x) dx$$

$$= \int_0^\infty \int_{t: g(t) > x} f_x(t) dt dx - \int_0^\infty \int_{t: g(t) < -x} f_x(t) dt dx$$

$$= \int_{-\infty}^\infty \int_0^{g^+(t)} f_x(t) dx dt + \int_{-\infty}^\infty \int_0^{g^-(t)} f_x(t) dx dt$$

$$= \int_{-\infty}^\infty f_x(t) \int_0^{g^+(t)} dx dt + \int_{-\infty}^\infty f_x(t) \int_0^{g^-(t)} dx dt$$

$$= \int_{-\infty}^\infty g^+(t) f_x(t) dt + \int_{-\infty}^\infty g^-(t) f_x(t) dt$$

$$g^+(t) = \max\{0, g(t)\}$$

$$g^-(t) = \max\{0, -g(t)\}$$

$$= \int_{-\infty}^\infty (g^+(t) + g^-(t)) f_x(t) dt$$

$$= \int_{-\infty}^\infty g(t) f_x(t) dt$$

$$= \underline{\mathbb{E}[g(x)]}$$

o Variance:

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

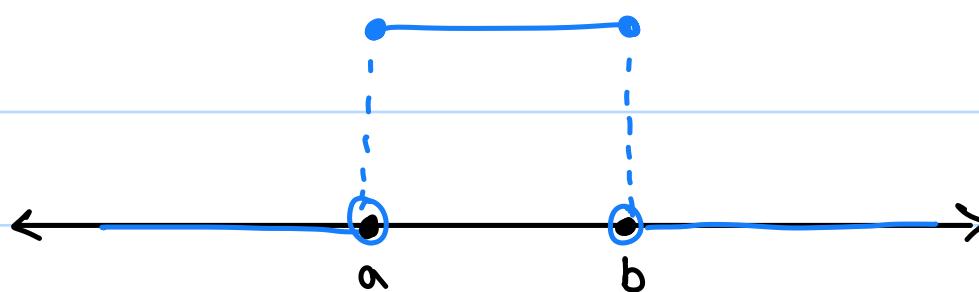
$$= \int_{-\infty}^\infty (x - \mathbb{E}[x]^2) f_x(x) dx$$

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

• Examples of Continuous Rvs :-

i) Uniform Continuous Rv :-

$$f_x(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{otherwise} \end{cases}$$



$$E[X] = \frac{a+b}{2}$$

$$Var[X] = \frac{(b-a)^2}{12}$$

2) Exponential Rv:

$$f_x(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad \lambda \in \mathbb{R}_+$$

Good model for the amount of time taken until a certain event happens.

$$E[X] = \int_{-\infty}^{\infty} x f_x(x) dx = \int_{-\infty}^{\infty} x \lambda e^{-\lambda x} dx$$

$$t = \lambda x, \quad dt = \lambda dx$$

$$\Rightarrow E[X] = \frac{1}{\lambda} \int_0^{\infty} t e^{-t} dt$$

$$= \frac{1}{\lambda} \left[t \int e^{-t} dt - \int \left(\int e^{-t} dt \right) dt \right]_0^\infty$$

$$= \frac{1}{\lambda} \left[-te^{-t} + \int e^{-t} dt \right]_0^\infty$$

$$= \frac{1}{\lambda} \left[-te^{-t} - e^{-t} \right]_0^\infty$$

$$= \frac{1}{\lambda} \left[-e^{-t}(t+1) \right]_0^\infty$$

$$= \frac{1}{\lambda} (0 - (-1))$$

$$\Rightarrow \underline{E[X] = \frac{1}{\lambda}}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx \quad [\text{LOTUS}]$$

$$= \int_0^{\infty} x^2 x e^{-\lambda x} dx$$

$$= \lambda \left[x^2 \int e^{-\lambda x} dx - \int 2x \int e^{-\lambda x} dx dx \right]_0^\infty$$

$$= \lambda \left[\left(\frac{-1}{\lambda} \right) x^2 e^{-\lambda x} - 2 \left(\frac{-1}{\lambda} \right) \int x e^{-\lambda x} dx \right]_0^\infty$$

$$= \left[-x^2 e^{-\lambda x} + 2 \left(x \int e^{-\lambda x} dx - \int \left(\int e^{-\lambda x} dx \right) dx \right) \right]_0^\infty$$

$$= \left[-x^2 e^{-\lambda x} + 2 \left(-\frac{x e^{-\lambda x}}{\lambda} + \frac{1}{\lambda} \int e^{-\lambda x} dx \right) \right]_0^\infty$$

$$= \left[-x^2 e^{-\lambda x} - \frac{2x e^{-\lambda x}}{\lambda} - \frac{2 e^{-\lambda x}}{\lambda^2} \right]_0^\infty$$

$$= \left[-\left(-\frac{2}{\lambda^2} \right) \right]$$

$$\Rightarrow E[X^2] = \frac{2}{\lambda^2}$$

$$\Rightarrow \text{Var}[X] = E[X^2] - E[X]^2 \\ = \frac{2}{\lambda^2} - \frac{1}{\lambda^2}$$

$$\Rightarrow \underline{\text{Var}[X]} = \frac{1}{\lambda^2}$$

3) Gaussian RV:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x) dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \end{aligned}$$

$$\text{Let } \frac{x-\mu}{\sigma} = t, dt = dx\left(\frac{1}{\sigma}\right)$$

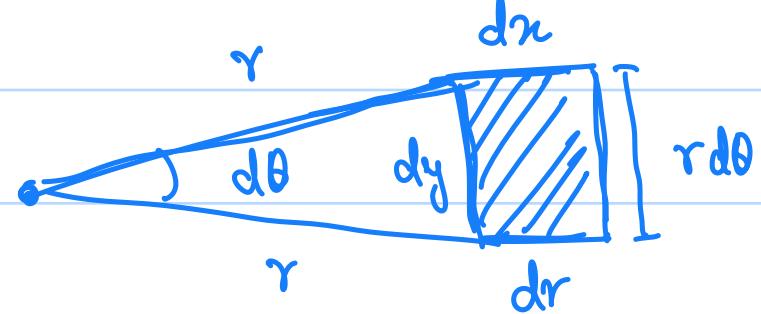
$$\begin{aligned} \Rightarrow \int_{-\infty}^{\infty} f_X(x) dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt := I \end{aligned}$$

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{(x^2+y^2)}{2}\right) dx dy$$

Let $x = r \cos \theta, y = r \sin \theta$

$$dx dy = \text{Area of a small element} = r dr d\theta$$



$$\Rightarrow I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_{r=0}^{\infty} \exp\left(-\frac{r^2}{2}\right) r dr d\theta$$

$$= \frac{1}{2\pi} \int_{r=0}^{\infty} r \exp\left(-\frac{r^2}{2}\right) \int_0^{2\pi} d\theta dr$$

$$= \int_{r=0}^{\infty} r \exp\left(-\frac{r^2}{2}\right) dr$$

$$\frac{d}{dr} \exp\left(-\frac{r^2}{2}\right) = -r \exp\left(-\frac{r^2}{2}\right)$$

$$\Rightarrow \int_{r=0}^{\infty} r \exp\left(-\frac{r^2}{2}\right) dr = \left[-\exp\left(-\frac{r^2}{2}\right) \right]_0^{\infty} = \underline{\underline{1}}$$

$$\Rightarrow I^2 = 1 \Rightarrow I = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} f_x(x) dx = 1 \Rightarrow f_x(x) \text{ is a valid PDF}$$

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$\text{Let } t = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma t + \mu \Rightarrow dx = \sigma dt$$

$$E[X] = \int_{-\infty}^{\infty} \frac{\sigma t + \mu}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2}\right) (\sigma dt)$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \exp\left(-\frac{t^2}{2}\right) dt + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{d}{dt} \left(\exp\left(-\frac{t^2}{2}\right) \right) dt + \mu$$

$$= \frac{\sigma}{\sqrt{2\pi}} \left[\exp\left(-\frac{t^2}{2}\right) \right]_{-\infty}^{\infty} + \mu$$

$$= \underline{\mu}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$\text{Let } t = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma t + \mu, dx = \sigma dt$$

$$\Rightarrow E[X^2] = \int_{-\infty}^{\infty} (\sigma t + \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2}\right) \sigma dt$$

$$= \int_{-\infty}^{\infty} (\sigma^2 t^2 + 2\sigma\mu t + \mu^2) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \exp\left(-\frac{t^2}{2}\right) dt + \frac{2\sigma\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \exp\left(-\frac{t^2}{2}\right) dt$$

$$+ \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \exp\left(-\frac{t^2}{2}\right) dt + \mu^2$$

$$\int_{-\infty}^{\infty} t^2 \exp\left(-\frac{t^2}{2}\right) dt = \left[t \int t \exp\left(-\frac{t^2}{2}\right) dt - \int \int t \exp\left(-\frac{t^2}{2}\right) dt dt \right]_{-\infty}^{\infty}$$

$$= \left[-t \exp\left(-\frac{t^2}{2}\right) + \int \exp\left(-\frac{t^2}{2}\right) dt \right]$$

$$= [0 + \sqrt{2\pi}] = \sqrt{2\pi}$$

$$\Rightarrow E[X^2] = \sigma^2 + \mu^2$$

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

• If $\mu = 0$ and $\sigma = 1$, then the RV is referred to as a "Standard Gaussian RV".

To transform any Gaussian RV X into a standard Gaussian RV T ,

$$T = \frac{X - \mu}{\sigma}$$

CDF of a standard gaussian RV,

$$\phi(x) = P(X \leq x)$$

$$\Rightarrow \phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

Lemma: $\phi(-x) = 1 - \phi(x)$, $\forall x \in \mathbb{R}$

$$\begin{aligned} \phi(-x) &= \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx - \int_{-x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx - \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s^2}{2}\right) ds \quad (s = -x) \\ &= 1 - \phi(x) \end{aligned}$$

$$\Rightarrow \underline{\phi(x) = 1 - \phi(-x)}$$

i.e., $P(Z \leq -x) = P(Z \geq x)$, $\forall x \in \mathbb{R}$, \forall std. Gaussian

Rv Z.

→ Joint CDF :-

The joint CDF of 2 RVS X and Y is defined as,

$$F_{XY}(x,y) = P(X \leq x, Y \leq y)$$

° Properties :-

1) $\lim_{x \rightarrow \infty} F_{XY}(x,y) = F_Y(y)$

$$\lim_{n \rightarrow \infty} F_{XY}(x,y) = \lim_{n \rightarrow \infty} P(X \leq n, Y \leq y)$$

$$= P\left(\bigcup_{n=1}^{\infty} \{(X \leq n) \cap (Y \leq y)\}\right)$$

$$= P(\{Y \leq y\})$$

$$= \underline{F_Y(y)}$$

2) $\lim_{x \rightarrow -\infty} F_{XY}(x,y) = 0$ (III proof as above)

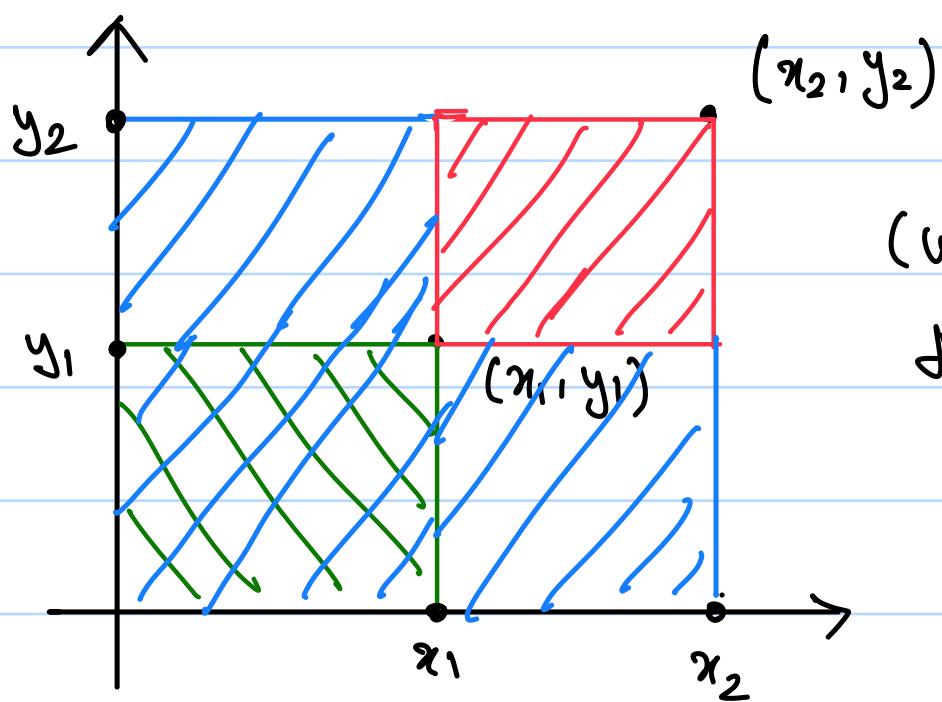
3) $\lim_{\delta \rightarrow 0^+} \lim_{\epsilon \rightarrow 0^+} F_{XY}(x+\delta, y+\epsilon) = F_{XY}(x,y)$ [Right Continuity]

4) $P(X \in [x_1, x_2], Y \in [y_1, y_2]) = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1)$

$$P(X \in [x_1, x_2], Y \in [y_1, y_2]) = P(X \leq x_2, Y \leq y_2)$$

$$= P(X \leq x_2, Y \leq y_1) - P(X \leq x_1, Y \leq x_2)$$

$$+ P(X \leq x_1, Y \leq y_1)$$



$$\Rightarrow P(X \in [x_1, x_2], Y \in [y_1, y_2]) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

• If X, Y are jointly continuous,

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) du dv$$

- For any region $B \subseteq \mathbb{R}^2$

$$P((x, y) \in B) = \iint_B f_{X,Y}(x, y) dy dx$$

Using the intuition that any general region can be approx. as a sum of rectangles and using the below formula

$$P(X \in [x_1, x_2], Y \in [y_1, y_2]) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{X,Y}(x, y) dy dx$$

$$- \iint_{-\infty}^{\infty} f_{X,Y}(x, y) dy dx = \lim_{n \rightarrow \infty} \lim_{y \rightarrow \infty} F_{XY}(x, y) = 1$$

- If X, Y are jointly discrete Rvs,

$$i) F_{XY}(x_1, y) = \sum_{d \leq x} \sum_{k \leq y} P_{XY}(d, k)$$

$$ii) P_{XY}(x_1, y) = F_{XY}(x_1, y) - F_{XY}(x_1, y-1) - F_{XY}(x_1, y) + F_{XY}(x_1, y-1)$$

- If X, Y are jointly continuous and given $F_{XY}(x_1, y)$,

$$f_{XY}(x_1, y) = \frac{S}{S_x} \frac{S}{S_y} F_{XY}(x_1, y) = \frac{S^2 F_{XY}(x_1, y)}{S_x S_y}$$

- Theorem: If X, Y are jointly continuous, then they are individually continuous.

Proof:

$$F_{XY}(x_1, y) = \int_{-\infty}^{y_1} \int_{-\infty}^x f_{XY}(u, v) du dv$$

$$\lim_{y \rightarrow \infty} F_{XY}(x_1, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_1} f_{XY}(u, v) du dv$$

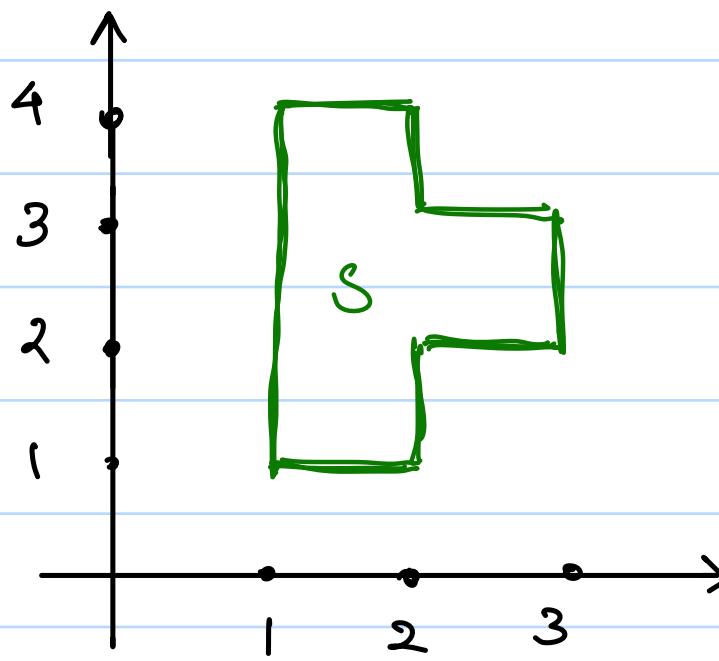
$$F_X(x) = \int_{-\infty}^x \int_{-\infty}^{\infty} f_{XY}(u, v) dv du$$

$$\Rightarrow F_X(x) = \int_{-\infty}^x f_X(v) dv \quad \left. \right\} x, y \text{ are continuous.}$$

$$\text{Hence } F_Y(y) = \int_{-\infty}^y f_X(v) dv$$

- $f_{XY}(x_1, y) = \frac{S^2 F_{XY}(x_1, y)}{S_x S_y}$. Since $S_x S_y$ represents a small area, we can say that f_{XY} is probability per unit area. ($S_x S_y = 1$).

Example:



$$f_{XY}(x_1, y_1) = \begin{cases} \frac{1}{4}, & (x_1, y_1) \in S \\ 0, & \text{ow.} \end{cases} \quad \text{Find } f_X(x), f_Y(y)$$

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x_1, y_1) dy \\ &= \int_1^4 \frac{1}{4} dy = \frac{1}{4}(3) = \frac{3}{4} \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x_1, y_1) dx \\ &= \int_1^2 \frac{1}{4} dx = \frac{1}{4}(2) = \frac{1}{2} \end{aligned}$$

• LOTUS of Joint Rvs :-

$$E[g(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, y_1) f_{XY}(x_1, y_1) dx dy$$

• Independence of Cont. Rvs :-

independent, then,

If x_1, x_2, \dots, x_n are said to be

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

Example: $Z = X + Y$, X, Y are jointly continuous. Find $f_Z(z)$.

$$f_{XY}(x, y) = f_{XY}(x, z-x)$$

$$\Rightarrow f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x) dx$$

$$F_Z(z) = P(Z \leq z) = P(X+Y \leq z)$$

$X + Y \leq z \rightarrow$ Half plane in \mathbb{R}^2 , let it be B .

$$\begin{aligned} P(X+Y \leq z) &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_{XY}(x, y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^z f_{XY}(x, v-x) dv dx . \quad [y=v-x] \\ &= \int_{-\infty}^z \int_{-\infty}^{\infty} f_{XY}(x, v-x) dx dv \end{aligned}$$

$$F_Z(z) = \int_{-\infty}^z f_Y(v-x) dv$$

Example: $f_{XY}(x, y) = 2e^{-x-y}$. Find if X, Y are indep. if X, Y are tne, and $0 < x < y < \infty$

$$f_X(x) = \int_0^y 2e^{-x-y} dy = 2e^{-x} \int_0^y e^{-y} dy$$

- Conditioning of Continuous Rvs :-

- Let X be a continuous RV and $A = \{x \in B\}$

$$F_{X|A}(x) = \int_{-\infty}^x f_{X|A}(u) du$$

$$F_{X|A}(x) = \frac{P(X \leq x, X \in B)}{P(X \in B)}$$

$$= \int_{(-\infty, x] \cap B} f_X(u) du \cdot \frac{1}{P(X \in B)}$$

$$= \int_{-\infty}^x \left(\frac{f_X(u)}{P(X \in B)} \mathbb{I}_B(u) \right) du \quad [\mathbb{I}_B : \text{indicator RV of } B]$$

$$\Rightarrow f_{X|A}(x) = \frac{f_X(x) \mathbb{I}_B(x)}{P(X \in B)} = \begin{cases} \frac{f_X(x)}{P(A)}, & x \in B \\ 0, & x \notin B \end{cases} \quad \text{The event } A$$

- Theorem: Let A_1, A_2, \dots, A_n be a partition of the sample space st $P(A_i) > 0 \ \forall i \in [1:n]$, then

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

Proof:

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

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

[Total Probability Theorem
on probability space]

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

$$\frac{dF_X(x)}{dx} = \sum_{i=1}^n P(A_i) \frac{dF_{X|A_i}(x)}{dx}$$

$$\Rightarrow f_X(x) = \sum_{i=1}^n f_{X|A_i}(x) P(A_i)$$

- Conditioning on Another RV :-

$$F_{X|A} = P(X \leq x | A) \quad [\text{By defn}]$$

Let $A = \{Y = y\}$, where Y is a continuous RV. In discrete case this can be used for conditioning over an RV, but,

$P(Y = y) = 0 \rightarrow$ Not allowed for conditioning.

$$\therefore \text{Let } A = \{Y \in (y, y+dy)\}$$

$$F_{X|\{Y \in (y, y+dy)\}} = \frac{P(X \leq x, Y \in (y, y+dy))}{P(Y \in (y, y+dy))}$$

$$= \frac{\int_y^{y+dy} \int_{-\infty}^x f_{XY}(u, v) du dv}{\int_y^{y+dy} f_Y(v) dv}$$

$$\Rightarrow f_{x|y}(x|y) = \frac{f_{xy}(x,y)}{f_y(y)}$$

[Differentiate and let
 $\Delta y \rightarrow 0$]

$$f_{xy}(x|y) = \lim_{\Delta y \rightarrow 0} f_x|_{\{y \in [y, y + \Delta y]\}}(x)$$

$$\frac{P(x \leq x, y \in [y, y + \Delta y])}{P(y \in [y, y + \Delta y])} = \frac{F_{xy}(x, y + \Delta y) - F_{xy}(x, y)}{F_y(y + \Delta y) - F_y(y)}$$

$$\lim_{\Delta y \rightarrow 0} \frac{F_{xy}(x, y + \Delta y) - F_{xy}(x, y)}{F_y(y + \Delta y) - F_y(y)} = \lim_{\Delta y \rightarrow 0} \frac{\frac{F_{xy}(x, y + \Delta y) - F_{xy}(x, y)}{\Delta y}}{\frac{F_y(y + \Delta y) - F_y(y)}{\Delta y}}$$

$$= \frac{\frac{\partial}{\partial y} F_{xy}(x, y)}{\frac{\partial}{\partial y} F_y(y)} = \frac{\frac{\partial F_{xy}(x, y)}{\partial y}}{f_y(y)}$$

$$\frac{\partial}{\partial x} \left(\frac{\frac{\partial F_{xy}(x, y)}{\partial y}}{f_y(y)} \right) = \frac{f_{xy}(x, y)}{f_y(y)}$$

$$\Rightarrow \frac{\partial}{\partial x} \left(F_x|_{y \in [y, y + \Delta y]}(x) \right) = \frac{f_{xy}(x, y)}{f_y(y)}$$

$$\therefore f_{x|y}(x, y) = \frac{f_{xy}(x, y)}{f_y(y)}$$

$$\cdot E[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x) dx$$

$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

$$E[X|Y] = \phi(Y), \quad \phi(y) = E[X|Y=y]$$

• Total Expectation Theorem :-

- Let A_1, A_2, \dots, A_n be the partition of a sample space, then

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

Proof:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} x \sum_{i=1}^n f_{X|A_i}(x) P(A_i) dx \\ &= \sum_{i=1}^n P(A_i) \int_{-\infty}^{\infty} x f_{X|A_i}(x) dx \\ &= \sum_{i=1}^n E[X|A_i] P(A_i) \end{aligned}$$

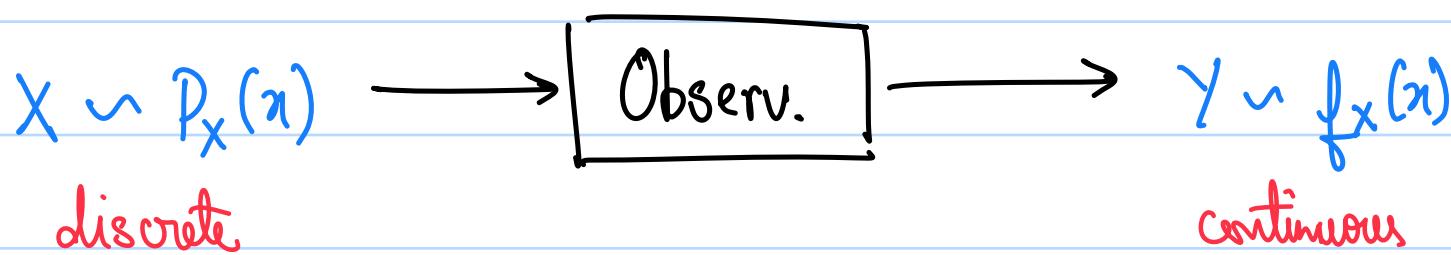
$$- E[X] = \int_{-\infty}^{\infty} E[X|Y=y] f_Y(y) dy = E[E[X|Y]]$$

$$\begin{aligned} \text{Proof: } E[X|Y=y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \\ &= \int_{-\infty}^{\infty} x \frac{f_{XY}(x,y)}{f_Y(y)} dx \end{aligned}$$

$$\Rightarrow E[X|Y=y] f_Y(y) = \int_{-\infty}^{\infty} x f_{XY}(x,y) dx$$

$$\begin{aligned} \int_{-\infty}^{\infty} E[X|Y=y] f_Y(y) dy &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x,y) dx dy \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{XY}(x,y) dy dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx = \underline{\underline{E[X]}}. \end{aligned}$$

→ Conditioning b/w a Discrete and Continuous Rvs :-



$$P(X=x|Y=y) = \frac{P_X(x) f_{Y|X}(y|x)}{f_Y(y)}$$

III to Bayes Theorem.

• Proof:

$$P(A|Y=y) = \lim_{\Delta y \rightarrow 0} P(A | y - \Delta y \leq Y \leq y + \Delta y)$$

$$= \frac{P(A) P(y - \Delta y \leq Y \leq y + \Delta y | A)}{P(y - \Delta y \leq Y \leq y + \Delta y)}$$

$$= \frac{P(A) \left(\frac{F(y + \Delta y | A) - F(y - \Delta y | A)}{\Delta y} \right)}{\left(\frac{F(y + \Delta y) - F(y - \Delta y)}{\Delta y} \right)}$$

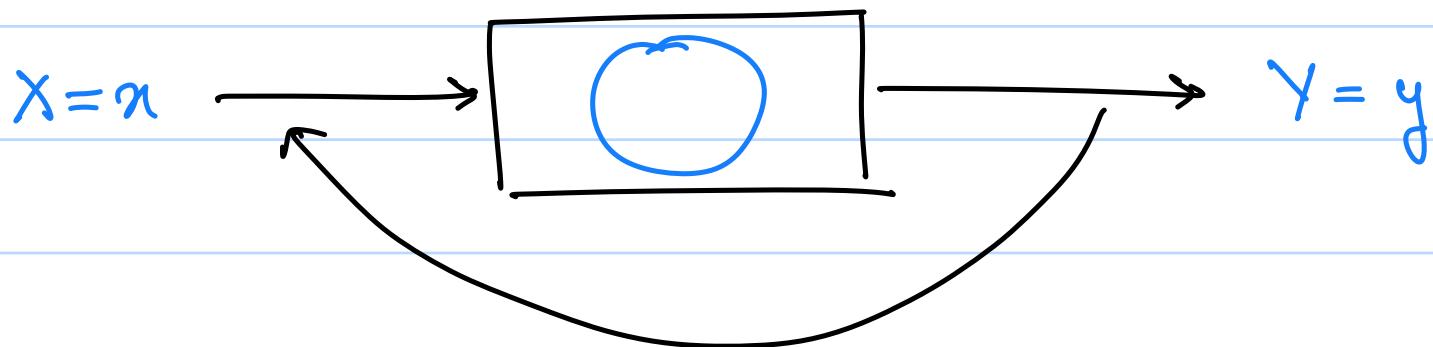
(multiply both sides by $1/\Delta y$)

$$= \frac{P(A) f_{Y|A}(y)}{f_Y(y)}$$

$$\Rightarrow P(A|Y=y) \triangleq \frac{P(A) f_{Y|A}(y)}{f_Y(y)}$$

$$\therefore P(X=x | Y=y) \triangleq \frac{P_X(x) f_{Y|X}(y|x)}{f_Y(y)}$$

- This procedure is also termed as **inferring** of X by Y .
- $P_X(x)$ - Poisson distribution of X
 $P_{X|Y}(x|y)$ - Posterior distribution of X on observing Y .
- Objective:- On observing $Y=y$, we have to find the highest probable value for $X=x$, where Y is continuous and X is discrete.



- Decision Rule:

$$\hat{x}_{MAP} =$$

