Textbooks:

1) Papoulis
2) Bertsekas
3) Grimmett

Prof. Gowtham Kurri

Probability And Random Processes

Grading:

- Assignments 15 %
- Quiz 1 & 2 15 % + 15 %
- In-Class Quiz 5 %
- Midsem 20 %
- Endsem 30 %

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
2) Probability Space
3) Continuity of Probability
4) Conditional Probability, Independence
5) Bayes Theorem & Total Probability Theorem
6) Counting

→ 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 outcomes 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

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

- 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.

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

Ω → Sample Space, $\mathcal{F}$ → Event Space, P → Probability law

- Sample Space : Set of all outcomes of the random experiment.

- Event Space : Set of all favourable outcomes

- Probability Law : Function that defines probability.

- Set Theory :-

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

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

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

$\bigcup_{i=1}^{\infty} A_i = \{ x \in A_i \text{ for some } i \in \mathbb{N} \}$ → By defn of $\cup$

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

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

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

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

- 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 possibilities)

- Countably infinite sample space : Tossing a coin until we see tails.
  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) $A_1, A_2, \ldots\ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (Closure under countable unions)

↪ if we take $A_i = \phi \ \forall \ i \geq$ some k, then we can use this statement for finite unions as well

(↪ Same can be said for $\cap$, using (ii) and De-Morgans)

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 (i) - (iii)

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

$A \in \mathcal{F}, B \in \mathcal{F} \Rightarrow A \cap B^c \in \mathcal{F}$. Ill'ly $B \cap A^c \in \mathcal{F}$

$$\therefore A \Delta B \in \mathcal{F}$$

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

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

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

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

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

$$\sigma(\Omega, \phi, A, B) = \sigma(\underbrace{A\cap B,\ A\cap B^c,\ A^c\cap B,\ A^c\cap B^c}_{\text{Disjoint sets that collectively exhaust } \Omega.})$$

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

- Probability Law (P) :- $P : \mathcal{F} \longmapsto \mathbb{R}$ 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 $A_1, A_2 \ldots$ are mutually exclusive / disjoint, then,

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

Properties :

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

Proof: Let $A_i = \phi \ \forall i$

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

$$= P(\phi) = \sum P(\phi) \Rightarrow P(\phi) = 0$$

Example: Construct a probability law for rolling a die.

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

Defn $\begin{cases} P(\{i\}) = p_i \quad \forall \, i \in \{1,2,3,4,5,6\} \text{ st } \sum_{i=1}^{6} p_i = 1 \text{ \& } p_i \geq 0 \\ P(A) = \sum_{i \in A} P(\{i\}), \quad A \subseteq \Omega \end{cases}$

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

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

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

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

$$\begin{aligned} P(B) &= P((A \cap B) \cup (B \setminus A)) \\ &= P(A \cap B) + P(B \setminus A) \end{aligned}$$

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

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

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

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

- Continuity Of Probability :-

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

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

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

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

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

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

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.

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

Proof of Claim 1:

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

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

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

Assume $x \in B_j \Rightarrow x \in A_j \setminus \bigcup_{k=1}^{j-1} A_k$

$$\Rightarrow x \notin \bigcup_{k=1}^{j-1} A_k$$

But we proved $x \in A_i \longrightarrow$ Contradiction

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

Ully can be stated for $i > j$.

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

$\Rightarrow \quad B_i \cap B_j = \phi \quad \forall \; i \neq j$

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

$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$ ——→ ②

Proof of Claim 2:

For ①, if $n = 1$, $B_1 = A_1 \longrightarrow$ 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 $n = k+1$,

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

$$\bigcup_{i=1}^{k+1} B_i = B_{k+1} \cup C_k = \left(A_{k+1} \setminus \bigcup_{i=1}^{k} A_i\right) \cup C_k$$

$$= (A_{k+1} \setminus C_k) \cup C_k$$

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

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

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

Final Proof: To prove:

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

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right)$$ 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 $\mathbb{R}$.

$$= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i)$$

$$= \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} B_i\right) \quad [\text{Additivity Axiom}]$$

$$= \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} A_i\right) \quad [\text{Claim 2}]$$

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

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

Corollary 1:

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

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

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

$= \quad P(A|B) = kP(A\cap B)$, if $A = B$, $P(B|B) = kP(B\cap B)$

$\Rightarrow \quad 1 = kP(B)$

$\Rightarrow \quad k = 1/P(B)$

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

→ **Independence** :

- 2 events A and B are independent if,

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

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

- 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,

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

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

$$\left(A_i \cap A_j = \phi \;\; \forall\, i \neq j\right), \;\left(\bigcup_{i=1}^{n} A_i = \Omega\right)$$

→ Total Probability Theorem :- Let $\{A_1, A_2, A_3 \ldots A_n\}$ be a partition of Ω, and B be any event. Then,

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

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

Proof:

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

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

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

→ **Bayes' Theorem** :-

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

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

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

→ **Multiplication Rule** :

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

→ **Conditional Independence** :

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

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

- 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 : ${}^nP_k = \dfrac{n!}{(n-k)!}$

ii) No. of k-length sets : ${}^nC_k = \dfrac{n!}{k!\,(n-k)!}$

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

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

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

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

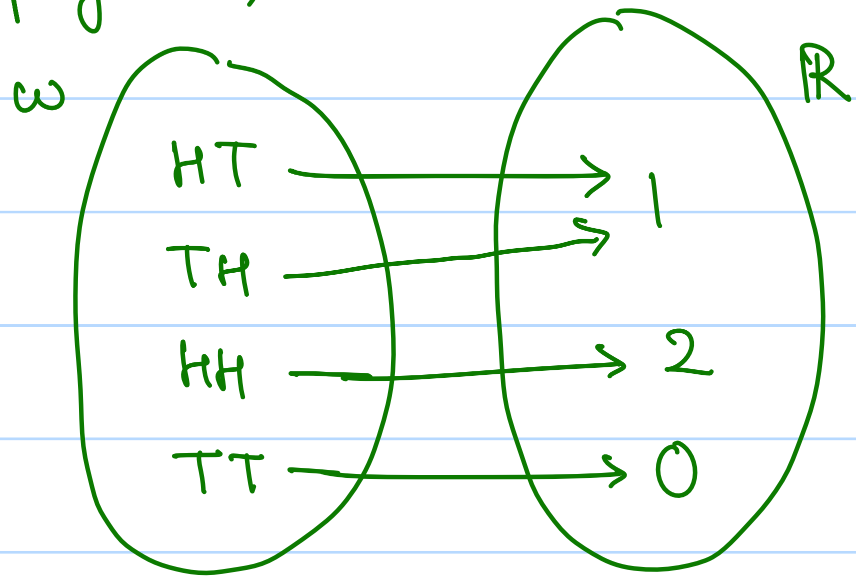
→ Random Variables :-

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

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

The mapping is,



$$\{x \le x_0\} = \begin{cases} \phi, & x_0 < 0 \\ \{TT\}, & x_0 \in [0,1) \\ \{HT, TH, TT\}, & x_0 \in [1,2) \\ \Omega, & x_0 \ge 2 \end{cases}$$

Theorem: $X : \Omega \mapsto \mathbb{R}$ 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) $X^{-1}(\{x_1\}) \in \mathcal{F} \quad \forall\, x_1 \in \mathbb{R}$

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) :

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

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

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

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

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

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

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,

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

<u>Example</u>: For,

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

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

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

$$\{X \leqslant x\} = \begin{cases} \phi, & x < c \\ \Omega, & x \geqslant c \end{cases}$$

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

- **Indicator R.V :-**

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

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

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

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

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

- $B_1 \cap B_2 = \phi$, then,

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

$B_1, B_2, B_3 \ldots$ 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) $\lim\limits_{x \to \infty} F_X(x) = 1$, $\lim\limits_{x \to -\infty} F_X(x) = 0$

$$\lim_{x \to \infty} F_X(x) = \lim_{x \to \infty} P_X(X \leq x) = P(\mathbb{R}) = \underline{1}$$

$$\lim_{x \to -\infty} F_X(x) = \lim_{x \to -\infty} P_X(X \leq x) = P(\phi) = \underline{0}$$

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, ie,

$$\lim_{\varepsilon \to 0^+} F_X(x+\varepsilon) = F_X(x) \quad \forall\, x \in \mathbb{R}$$

<u>Proof:</u>

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

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

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

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

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

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

$$\Rightarrow \underline{RHL = F_X(x)}$$

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

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

$= P(X$

→ Discrete Random Variable :-

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

- **Probability Mass Function** :

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

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

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

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

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

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

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

- **CDF of a Discrete R.V** :-

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

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

- **Functions on R.V :-**

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

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

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

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

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

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

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

$\Rightarrow$ Y is a valid r.v

Better proof needed.

- PMF of y : $\boxed{P_Y(Y = y) = \sum_{Y(x) = y} P_X(X = x)}$

- **Expectation:** $\boxed{\mathbb{E}[X] = \sum_{x \in \mathbb{R}} x P_X(x)}$

- If $y = Y(x)$,

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

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

$$= \underbrace{\sum_{y \in \mathcal{Y}} \sum_{Y(x) = y}}_{\text{All possible } x} Y(x) P_X(x) = \sum_{x \in \mathcal{X}} Y(x) P_X(x)$$

- **Variance:**

$$\boxed{Var[X] = E\left[(X - E[X])^2\right]}$$

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

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

- **Examples of RV:-**

1) **Bernoulli RV:** Binary o/p Rv. (ex: Coin Toss)

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

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

$$E[X] = p, \quad Var[X] = p - p^2$$

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

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

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

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

$$E[X] = np \qquad Var[X] = np(1-p)$$

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

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

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

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

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

$$E[X] = Var[X] = \lambda.$$

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

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

Proof:

$$\lim_{n\to\infty} P_Y(k) = \lim_{n\to\infty} {}^nC_k\, p^k (1-p)^{n-k}$$

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

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

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

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

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

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

$$= \frac{\lambda^k}{k!} e^{\lim_{n\to\infty} -\lambda + \frac{\lambda k}{n}}$$

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

$\longrightarrow$ $(1+f(x))^{g(x)}$

$f(x) \longrightarrow 0$

$g(x) \longrightarrow \infty$

as $x \to \infty$,

then

$e^{g(x)\ln(1+f(x))}$

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

→ Multiple Random Variables :-

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

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:
$$P_X(x) = P(X = x)$$
$$= P\left(X = x \cap \bigcup_{y \in Y} (Y = y)\right)$$
$$= P\left(\bigcup_{y} (X = x \cap Y = y)\right)$$
$$= \sum_{y \in Y} P(X = x \cap Y = y) = \sum_{y \in Y} P_{XY}(x,y$$

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

- <u>Functions on 2 Rv's</u> :-

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

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

- <u>Independence</u> :-

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 $\stackrel{\wedge}{=}$ 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 an RV, then,

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

→ <u>Properties of $E[X]$ and $Var[X]$</u> :-

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

$$Var[Y] = a^2 Var[X]$$

2) $Y = X_1 + X_2 \implies E[Y] = E[X_1] + E[X_2]$ → Prove by LOTUS

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

$\implies Var[Y] = Var[X_1] + Var[X_2] + 2\,Cov[X_1, X_2]$

Where,

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

<u>Definition</u>: Correlation Coefficient $= \rho(X_1, X_2)$

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

<u>Theorem</u>: $|\rho(X_1, X_2)| \leq 1$. Equality iff $X_1 = aX_2 + b$, $a, b \in \mathbb{R}$

i.e. X_1 and X_2 are linearly dependent.

<u>Proof</u>: By Cauchy-Schwartz Inequality,

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

Proof of Cauchy-Schwartz,

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

$$\Rightarrow E[X^2] - 2\alpha E[XY] + \alpha^2 E[Y^2] \geq 0 \rightarrow \text{Is a Quadratic wrt } \alpha$$

$$\Rightarrow \sqrt{(2E[XY])^2 - 4(E[Y^2])(E[X^2])} \leq 0 \rightarrow \text{Discriminant} < 0$$

$$\Rightarrow 4E[XY]^2 - 4E[Y^2]E[X^2] \leq 0$$

$$\Rightarrow E[XY] \leq E[Y^2]E[X^2]$$

To Prove the theorem,

$$X \longrightarrow (X - E[X]) \qquad Y \longrightarrow (Y - E[Y])$$

$$\Rightarrow E[(X-E[X])(Y-E[Y])]^2 \leq E[(X-E[X])^2]E[(Y-E[Y])^2]$$

$$\Rightarrow |\mathrm{Cov}[X,Y]|^2 \leq \sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}$$

$$\Rightarrow \rho[X,Y]^2 \leq 1 \Rightarrow |\rho[X,Y]| \leq 1$$

3) X, Y, et $Z = X + Y$

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

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

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

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

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

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

→ **Conditioning** :-

Conditioning on RV $X : \Omega \mapsto \mathbb{R}$ 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 an RV X on another RV Y

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

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

- Conditional Expectance :-

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

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

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

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

- Total Expectation Theorem :-

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

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

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

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

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

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

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

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

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

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

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

- Conditional Expectance as a Rv :-

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

$\phi : \text{Range}(Y) \longmapsto \mathbb{R}$

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

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

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

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

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

- <u>Conditional Independence of RVs :-</u>

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

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

- <u>Conditional Variance :</u>

$$\mathrm{Var}[X \mid Y = y] = E\left[(X - E[X \mid Y = y])^2 \mid Y = y \right]$$

Let $\psi(y) = \mathrm{Var}[X \mid Y = y]$, $\psi(y)$ is an RV.

$$\mathrm{Var}[X \mid Y = y] = E[X^2 \mid Y = y] - (E[X \mid Y = y])^2 \longrightarrow$$ Same Proof as in Non - Condition - al Case

- <u>Law Of Total Variance :-</u>

$$\mathrm{Var}[X] = E[\psi(Y)] + \mathrm{Var}[\phi(Y)]$$

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

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

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

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

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

→ **Memoryless Property of a Geometric RV :-**

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

$$P(X > n) = \sum_{k=n+1}^{\infty} (1-p)^{k-1} p$$

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

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

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

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

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

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

- CDF of an RV is given by,

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

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

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

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

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

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

→ 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:

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

$$= \sum_y P_Y(y)\, E\left[(X - f(y))^2 \mid Y = y\right]$$

$$= \sum_y P_Y(y)\, E\left[X^2 + f(y)^2 - 2Xf(y) \mid Y = y\right]$$

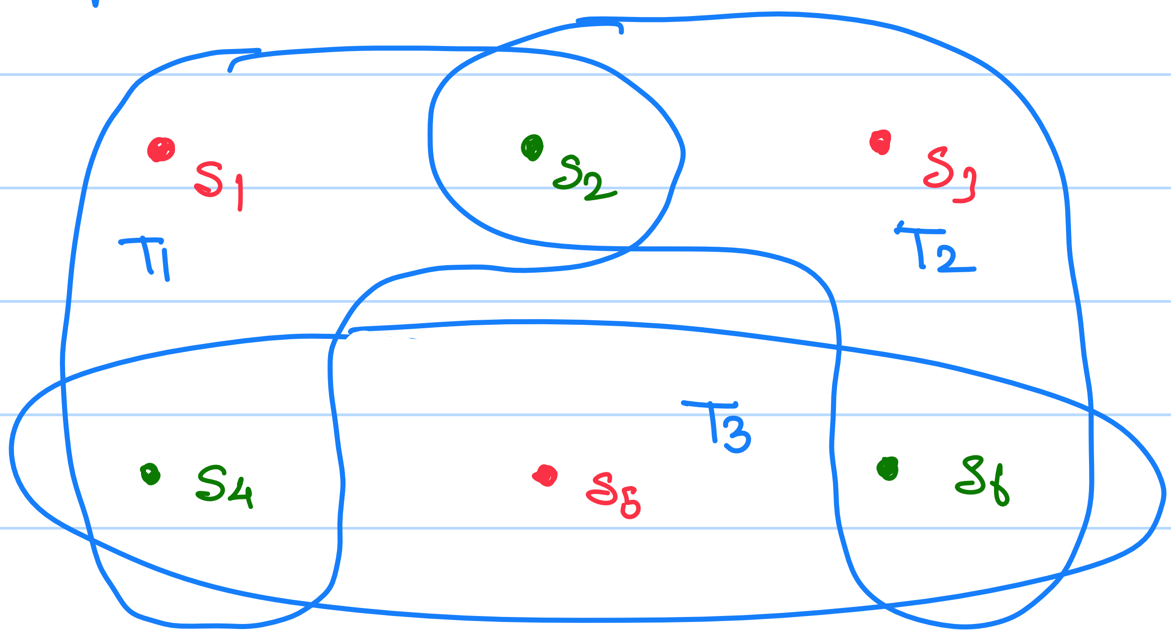
$$= \sum_y P_Y(y) \left(E[X^2 \mid Y = y] + f(y)^2 - 2E[X \mid Y = y] f(y) \right)$$

Each of the terms in the summation is minimized when,

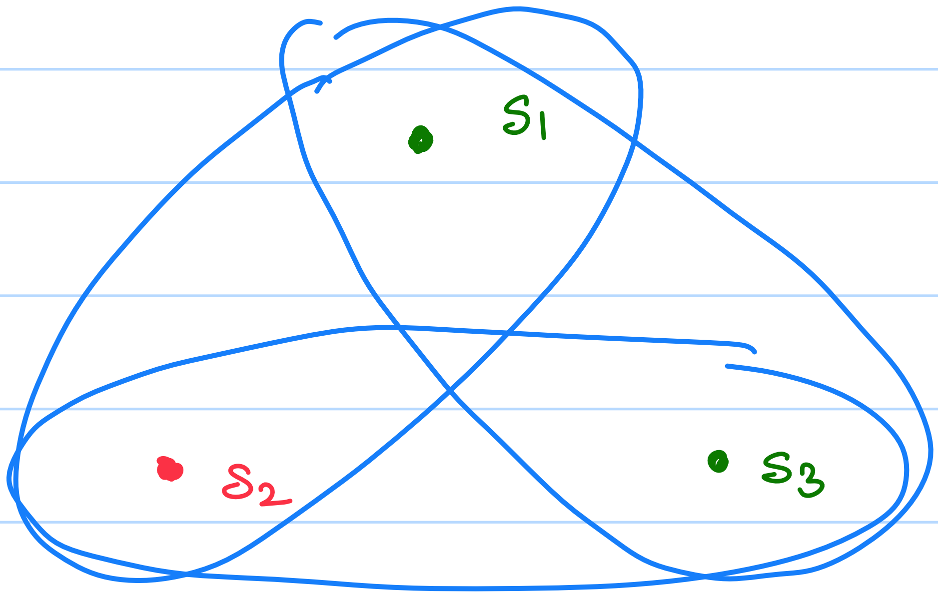
$$f(y) = \phi(y) = E[X \mid Y = y]$$

- Combinatorics and Graph Theory :-

Let S be a set of some elements and $T_1, T_2, T_3 \dots, T_m \subseteq S$ s.t. $n(T_i) = l$ $\forall$ $i \in [1:m]$, 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.

- <u>Theorem</u>:

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

- <u>Proof</u>: Let $S = \{x_1, x_2, x_3 \ldots, x_n\}$. Randomly color each element of S black or white, independentally and identically distributed, each with probability $1/2$.

Let E_i be the event that T_i is monochromatic.

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

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

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

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

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

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

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

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

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

- The converse of the above theorem is false.

- Entropy (Uncertainity) :-

- Consider an RV with,

$$P_X(0) = P_X(1) = 1/2 \quad (X_1)$$

and another RV with,

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

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

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

- If $X \sim$ Binomial RV,

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

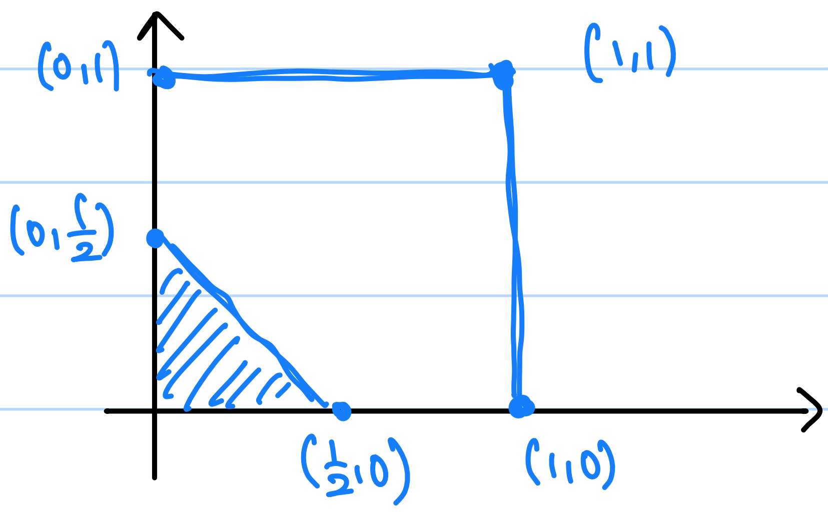
→ <u>Continuous Random Variable</u> :-

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

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

- Now, if Ω is an uncountably infinite set,

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



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

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

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

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

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

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

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

- <u>Definition</u>:

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.

$$\left.\begin{aligned} f_X(x) &= F_X'(x) \\ P(x \in [a,b]) &= \int_a^b f_X(u)\, du \end{aligned}\right\} \text{[Fundamental Theorem of Calculus]}$$

$f_X(x)$ 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) $P(X = x) = 0 \;\forall\; x \in \mathbb{R}$

- **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,

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

• **Proof:**

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

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

Proof:
$$\int_0^\infty P(Y>y)\, dy = \int_0^\infty \int_y^\infty f_Y(t)\, dt\, dy$$

$$= \int_{t=0}^\infty \int_{y=0}^t f_Y(t)\, dy\, dt$$

$$= \int_{t=0}^\infty f_Y(t) \left(\int_{y=0}^t dy \right) dt$$

$$= \int_{t=0}^\infty t\, f_Y(t)\, dt$$

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

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

$$E[g(X)] = \int_0^\infty P(g(X) > y)\, dy$$

$$= \int_0^\infty P(X \in \{x : g(x) > y\})\, dy$$

$$= \int_{y=0}^\infty \int_{x : g(x) > y} f_X(x)\, dx\, dy$$

$$= \int_{x : g(x) > y} f_X(x) \int_{y=0}^{g(x)} dy\, dx$$

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

$$= \underline{\underline{\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx}}$$

For any general $g(X)$,

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

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

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

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

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

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

$$= E[Y]$$

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

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

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

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

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

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

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

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

$$= \int_{-\infty}^{\infty} \left(g^+(t) + g^-(t)\right) f_X(t)\, dt$$

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

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

- <u>Variance</u>:

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

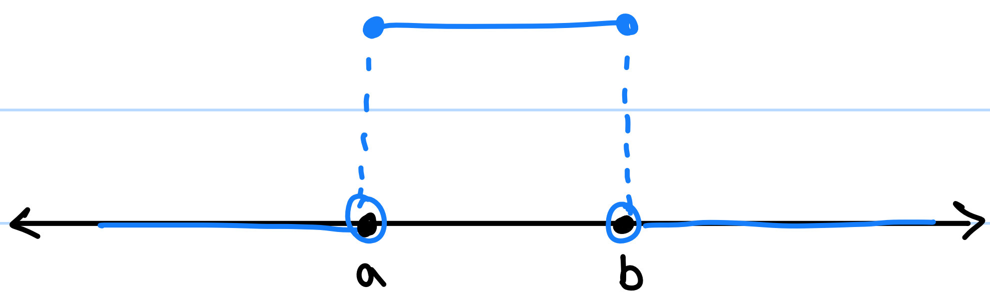
$$= \int_{-\infty}^{\infty} (x - E[X]^2) f_X(x)\, dx$$

$$= E[X^2] - (E[X])^2$$

- Examples of Continuous RVs :-

1) Uniform Continuous Rv :-

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



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

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

2) Exponential Rv:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2}$$

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

3) Gaussian RV:

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

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

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

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

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

$$\Rightarrow \int_{-\infty}^{\infty} f_X(x)\, dx = \frac{\sigma}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

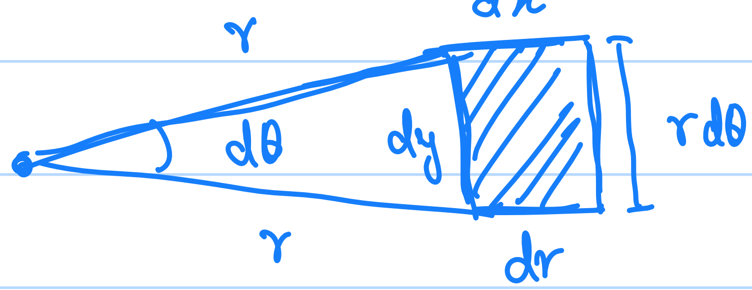
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt := I$$

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

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

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

$dx\, dy$ = Area of a small element = $r\, dr\, d\theta$



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

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

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

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

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

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

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

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

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

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

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

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

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

$$= \underline{\underline{\mu}}$$

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

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

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

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

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

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

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

$$\int_{-\infty}^{\infty} t^2 \exp\left(\frac{-t^2}{2}\right) dt = \left[t \int t \exp\left(\frac{-t^2}{2}\right) dt - \int 1 \int t \exp\left(\frac{-t^2}{2}\right) dt \, dt \right]_{-\infty}^{\infty}$$

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

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

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

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

- 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,

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

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

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

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

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

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

$$= 1 - \phi(x)$$

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

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

→ **Joint CDF :-**

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

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

- **Properties :-**

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

$$\lim_{n \to \infty} F_{XY}(n,y) = \lim_{n \to \infty} P(X \le n, Y \le y)$$

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

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

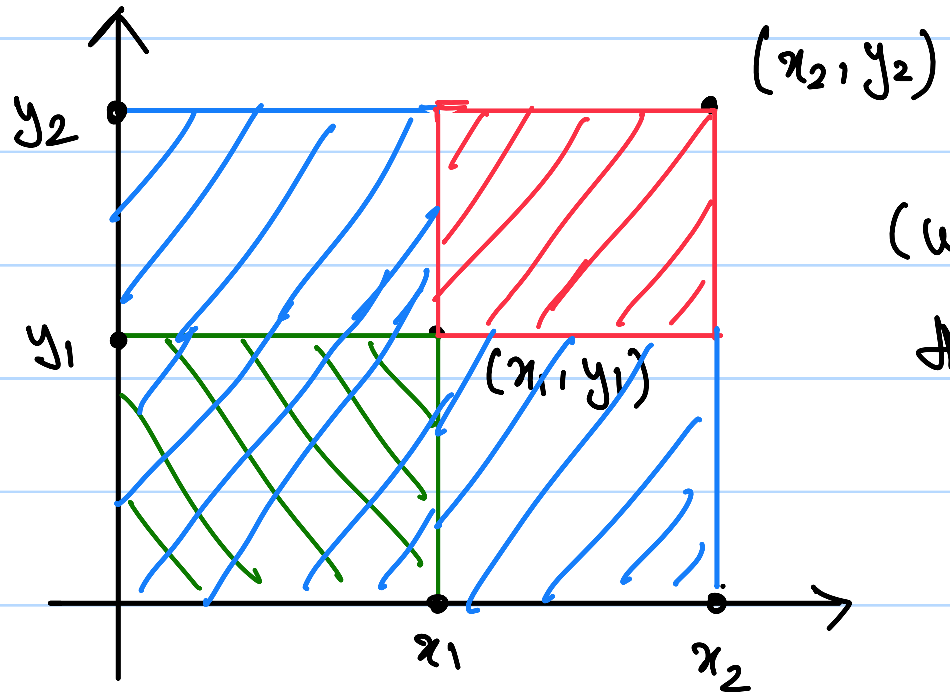
$$= \underline{\underline{F_Y(y)}}$$

2) $$\lim_{x \to -\infty} F_{XY}(x,y) = 0 \quad (||| \text{ proof as above})$$

3) $$\lim_{\delta \to 0^+} \lim_{\varepsilon \to 0^+} F_{XY}(x+\delta, y+\varepsilon) = F_{XY}(x,y) \quad [\text{Right Continuity}]$$

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

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



(We are removing the blue area)

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

- If X, Y are jointly continuous,

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

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

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

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

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

- $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx = \lim_{x \to \infty} \lim_{y \to \infty} F_{XY}(x, y) = 1$$

- If X, Y are jointly discrete RVs,

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

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

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

$$f_{XY}(x,y) = \frac{\delta}{\delta x}\frac{\delta}{\delta y} F_{XY}(x,y) = \frac{\delta^2 F_{XY}(x,y)}{\delta x\, \delta y}$$

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

Proof:

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

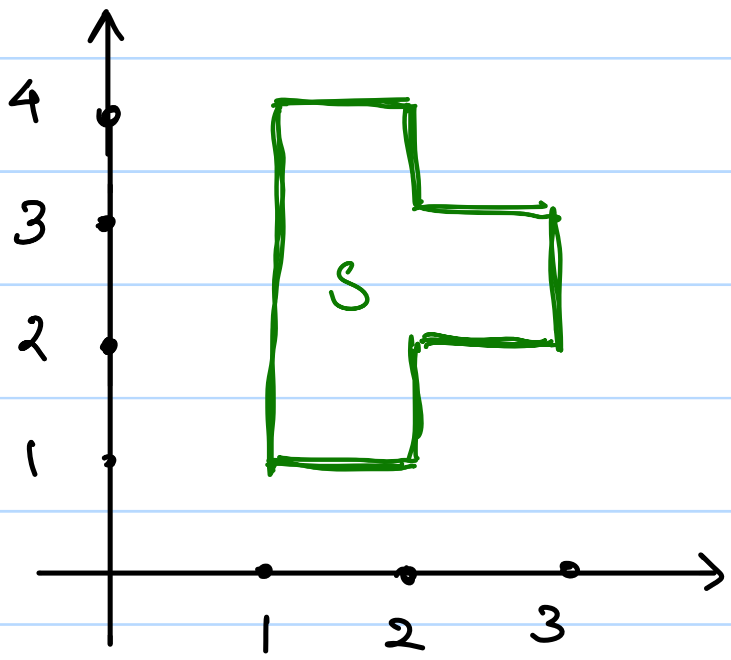
$$\lim_{y\to\infty} F_{XY}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{XY}(u,v)\,du\,dv$$

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

$$\left.\begin{aligned} \Rightarrow F_X(x) &= \int_{-\infty}^{x} f_X(u)\,du \\ \text{III}^{\text{ly}} \quad F_Y(y) &= \int_{-\infty}^{y} f_X(v)\,dv \end{aligned}\right\} X, Y \text{ are continuous.}$$

- $f_{XY}(x,y) = \dfrac{\delta^2 F_{XY}(x,y)}{\delta x\,\delta y}$. Since $\delta x\,\delta y$ represents a small area, we can say that f_{XY} is probability per unit area. ($\delta x \delta y = 1$).

Example:



$$f_{XY}(x,y) = \begin{cases} 1/4, & (x,y) \in S \\ 0, & \text{ow.} \end{cases}$$

Find $f_X(x)$, $f_Y(y)$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy$$

$$= \int_{1}^{4} 1/4 \, dy = 1/4 \,(3) = 3/4$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx$$

$$= \int_{1}^{3} 1/4 \, dx = 1/4 \,(2) = 1/2$$

- LOTUS of Joint RVs :-

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

- Independence of Cont. RVs :-

If $X_1, X_2 \ldots X_n$ are said to be independent, then,

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

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

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

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

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

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

$$P(X+Y \leqslant z) = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_{XY}(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_{XY}(x, v-x)\, dv\, dx \quad . \quad [y = v - x]$$

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

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

Example: $f_{XY}(x,y) = 2e^{-x-y}$. Find if X, Y are indp. if X, Y are the, 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 $A = \{x \in B\}$

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

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

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

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

The event A

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

- **Theorem:** Let $A_1, A_2 \ldots A_n$ 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:

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

$$= \sum_{i=1}^{n} P(X \leq x \mid A_i) P(A_i)$$ [Total Probability Theorem on probability space]

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

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

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

- <u>Conditioning on Another RV</u> :-

$$F_{X|A} = P(X \leq x \mid A)$$ [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,

$$P(Y = y) = 0 \longrightarrow \text{Not allowed for conditioning.}$$

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

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

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

$$\Rightarrow \boxed{f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}}$$

[Differentiate and let $dy \to 0$]

$$f_{XY}(x|y) = \lim_{\Delta y \to 0} f_{X|\{Y\in[y,y+\Delta y]\}}(x)$$

$$\frac{P(X \leq x,\, Y \in [y, y+\Delta y])}{P(Y \in [y, y+\Delta y])} = \frac{F_{XY}(x, y+\Delta y) - F_{XY}(x,y)}{F_Y(y+\Delta y) - F_Y(y)}$$

$$\lim_{\Delta y \to 0} \frac{F_{XY}(x,y+\Delta y) - F_{XY}(x,y)}{F_Y(y+\Delta y) - F_Y(y)} = \lim_{\Delta y \to 0} \frac{\dfrac{F_{XY}(x,y+\Delta y) - F_{XY}(x,y)}{\Delta y}}{\dfrac{F_Y(y+\Delta y) - F_Y(y)}{\Delta y}}$$

$$= \frac{\frac{\delta}{\delta y} F_{XY}(x,y)}{\frac{d}{dy} F_Y(y)} = \frac{\frac{\delta F_{XY}(x,y)}{\delta y}}{f_Y(y)}$$

$$\frac{\delta}{\delta x}\left(\frac{\frac{\delta F_{XY}(x,y)}{\delta y}}{f_Y(y)}\right) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

$$\Rightarrow \frac{\delta}{\delta x}\left(F_{X|Y\in\{y,y+\Delta y\}}(x)\right) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

$$\therefore \quad f_{X|Y}(x,y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

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

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

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

- <u>Total Expectation Theorem</u> :-

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

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

<u>Proof</u>:

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

$$= \sum_{i=1}^{n} P(A_i) \int_{-\infty}^{\infty} x f_{X|A_i}(x)\, dx$$

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

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

<u>Proof</u>: $E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$

$$= \int_{-\infty}^{\infty} \frac{x f_{XY}(x,y)}{f_Y(y)}\, dx$$

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

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

$$= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \underline{\underline{E[X]}}.$$

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

$$X \sim P_X(x) \longrightarrow \boxed{\text{Observ.}} \longrightarrow Y \sim f_X(x)$$

discrete   continuous

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

||| to Bayes Theorem.

• Proof:

$$P(A|Y=y) = \lim_{\Delta y \to 0} P(A|\, y-\Delta y \le Y \le y+\Delta y)$$

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

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

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

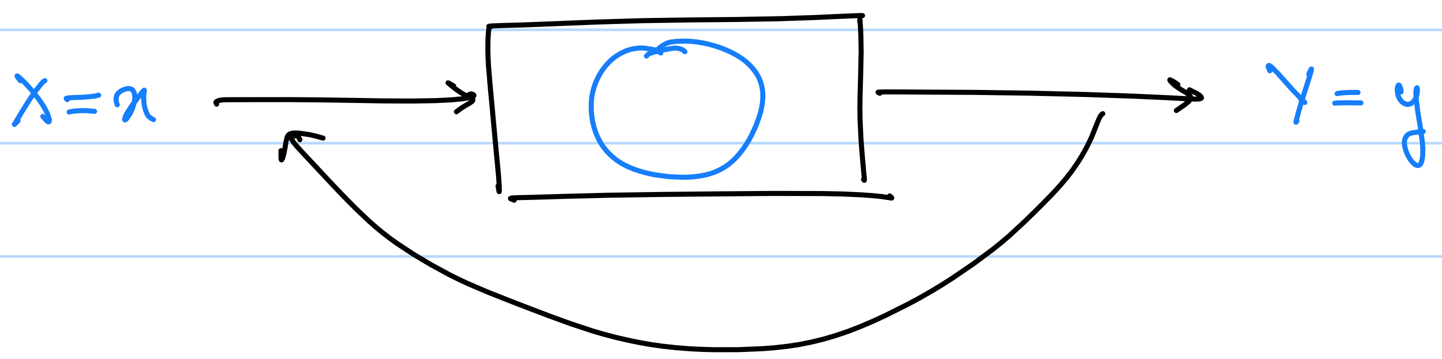
$$= \frac{P(A)\, f_{Y|A}(y)}{f_Y(y)}$$

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

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

- This procedure is also termed as inferring of X by Y.
- $P_X(x)$ – Prior distribution of X

  $P_{X|Y}(x|y)$ – Posterior distribution of X on observing Y.
- <u>Objective</u> :- On observing $Y=y$, we have to find the highest probable value for $X=x$, where Y is continuous and X is discrete.



- <u>Decision Rule</u>:

$$\hat{x}_{MAP} = \underset{x \in \{0,1,\dots M-1\}}{\arg\max}\; P_{X|Y}(x|y)$$

(Maximum Posterior Probability (MAP) rule)

- As an application of this decision rule, we can look at an abstraction of the digital communication system, based on a binary MAP.

- Binary MAP :- $X \in \{b, -b\}$

$$P_X(b) = P_1 \quad , \quad P_X(-b) = P_0$$

Let $Y = X + Z$, where $Z \sim \mathcal{N}(0, \sigma^2)$ [Gaussian RV]
and X, Z are independent.

$$F_{Y|\{X=b\}}(y) = P(Y \leq y \mid X = b)$$

$$= \frac{P(Y \leq y, X = b)}{P(X = b)}$$

$$= \frac{P(X + Z \leq y, X = b)}{P(X = b)}$$

$$= \frac{P(b + Z \leq y)\, P(X = b)}{P(X = b)}$$

$$= F_Z(y - b)$$

$$\Rightarrow \quad f_{Y|X}(y|b) = f_Z(y-b)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(y-b)^2}{2\sigma^2}}$$

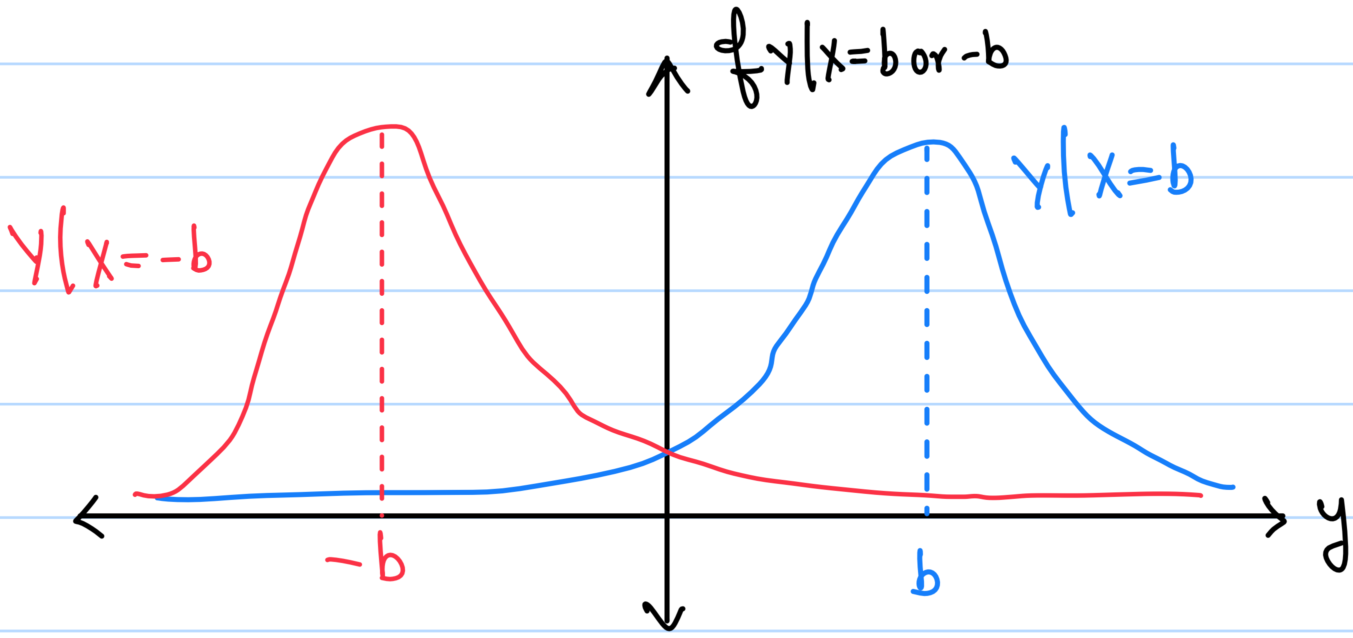
$$\Rightarrow \quad Y|\{X=b\} \sim \mathcal{N}(b, \sigma^2)$$

|||ly $f_{Y|X}(y|-b) = f_Z(y+b)$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+b)^2}{2\sigma^2}}$$

$\Rightarrow \quad Y|\{X=-b\} \sim \mathcal{N}(-b, \sigma^2)$

$\Rightarrow$



$$\hat{x} = \arg\max_{x \in \{b,-b\}} P_{X|Y}(x|y)$$

$$= \arg\max_{x \in \{b,-b\}} \frac{f_{Y|X}(y|x)\, P_X(x)}{f_Y(y)}$$

$$= \arg\max_{x \in \{b,-b\}} f_{Y|X}(y|x)\, P_X(x)$$

Let $\Lambda(y) = \dfrac{f_{Y|X}(y|b)}{f_{Y|X}(y|-b)}$

$$\Rightarrow \hat{x} = \arg\max f_{Y|X}(y|x)\, P_X(x) = \begin{cases} b, & \Lambda(y) > P_0/P_1 \\ -b, & \Lambda(y) < P_0/P_1 \end{cases}$$

$$\frac{f_{Y|X}(y|b)}{f_{Y|X}(y|-b)} = \frac{\exp\left(-\frac{(y-b)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(y+b)^2}{2\sigma^2}\right)}$$

$$= \exp\left(\frac{(y+b)^2-(y-b)^2}{2\sigma^2}\right)$$

$$= \exp\left(\frac{2by}{\sigma^2}\right)$$

$$\hat{x} = \begin{cases} b, & \exp\left(\frac{2by}{\sigma^2}\right) \geq \frac{P_0}{P_1} \\ -b, & \exp\left(\frac{2by}{\sigma^2}\right) < \frac{P_0}{P_1} \end{cases}$$

$$\Rightarrow \hat{x} = \begin{cases} b, & y \geq \frac{\sigma^2}{2b}\log\frac{P_0}{P_1} \\ -b, & y < \frac{\sigma^2}{2b}\log\frac{P_0}{P_1} \end{cases}$$

If $b = 1$ and $P_0 = P_1$

$$\Rightarrow \hat{x} = \begin{cases} 1, & y \geq 0 \\ -1, & y < 0 \end{cases}$$

$\rightarrow$ The basis of any digital comm. system.

Assuming $P_0 = P_1$, the probability of error will be,

$$P(\hat{X}_{map}(y) \neq x) = P(\hat{x} = -b \mid x = b) + P(\hat{x} = b \mid x = -b)$$

$$P(\text{Error} \mid x = b) = P(\hat{x} = -b \mid x = b) = P(Y < 0 \mid x = b)$$

$$= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-b)^2/2\sigma^2}\, dy$$

Let $\frac{y-b}{\sigma} = u$, $\Rightarrow$ $y = \sigma u + b$

$y = 0 \Rightarrow u = \frac{-b}{\sigma}$, and $y \to -\infty$, $u \to -\infty$

$$\Rightarrow P(\text{Error} \mid X = b) = \int_{-\infty}^{-b/\sigma} e^{-u^2/2}\, du \, \frac{1}{\sqrt{2\pi}}$$

$$= \phi\left(\frac{-b}{\sigma}\right) \quad (\text{CDF of a Standard Gaussian})$$

$$= 1 - \phi\left(\frac{b}{\sigma}\right)$$

$$P(\text{Error} \mid X = -b) = P(\hat{x} = b \mid X = -b)$$

$$= P(Y > 0 \mid X = -b)$$

$$= 1 - \phi\left(\frac{b}{\sigma}\right)$$

$$\Rightarrow \boxed{P(\text{Error}) = 2 - 2\phi\left(\frac{b}{\sigma}\right)}$$

→ Functions of Random Variables :-

Let $X : \Omega \mapsto \mathbb{R}$ and $Y = g(X)$.

- If X is discrete,

$$\boxed{P_Y(y) = \sum_{x : g(x) = y} P_X(x)}$$

- To find the PDF of Y, given the PDF of X, if X is continuous,

1) Calculate the CDF of Y in terms of $F_X(x)$.

$$F_Y(y) = P(g(X) < y)$$

2) Differentiate $F_Y(y)$ to get $f_Y(y)$.

- <u>Theorem</u>: Let X and $Y = g(X)$ be continuous RVs. Suppose $\exists$ a partition of $\mathbb{R}$ into intervals $I_1, I_2 \ldots I_n$, such that $g(x)$ is strictly monotone and differentiable in each I_i, then

$$f_Y(y) = \sum_{i=1}^{n} f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right|$$

$g_i(x)$ is the function in each interval I_i.

- <u>Functions on Multiple RVs</u> :-

$$Z = g(X, Y), \text{ ie}$$

$$Z(\omega) = g(X(\omega), Y(\omega)) \quad \forall \ \omega \in \Omega$$

1) <u>Sum of Independent RVs</u> :-

$$Z = X + Y \quad , \ X, Y \text{ are independent}$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx$$

If X is discrete and Y is continuous,

$$F_Z(z) = P(X+Y \leq z)$$
$$= \sum_{x \in X} P(X+Y \leq z \mid X = x)\; P(X=x)$$
$$= \sum_{x \in X} P(Y \leq z-x)\, P_X(x)$$
$$= \sum_{x \in X} F_Y(z-x)\, P_X(x)$$

$$\Rightarrow \quad f_Z(z) = \sum_{x \in X} f_Y(z-x)\, P_X(x)$$

- Two Functions of Two Rvs :-

Let X and Y be jointly continuous with joint PDF $f_{XY}(x,y)$, and

$$Z = g(X,Y)$$
$$W = h(X,Y)$$

To find $f_{Z,W}(z,\omega)$,

1) Compute $F_{Z,W}(z,\omega)$

$$F_{Z,W}(z,\omega) = P(Z \leq z, W \leq \omega)$$
$$= P(g(X,Y) \leq z, h(X,Y) \leq \omega)$$
$$= P((X,Y) \in B_{z,\omega})$$

$$\Rightarrow \quad F_{Z,W}(z,\omega) = \int_{x \in B_{z,\omega}} f_{XY}(x,y)\, dx\, dy$$

$$B_{z,\omega} = \{(x,y) : g(x,y) \leq z \text{ and } h(x,y) \leq \omega\}$$

2) Take double derivative,

$$f_{ZW}(z,\omega) = \frac{\delta^2 F_{ZW}(z,\omega)}{\delta z\, \delta \omega}$$

- <u>Jacobian</u>:

Let $x = g(t)$, then,

$$\int f(x)\, dx = \int f(g(t)) \frac{dx}{dt}\, dt$$
$$= \int \frac{f(g(t))}{\left.\frac{dt}{dx}\right|_{x=g(t)}}\, dt$$

Using this result, we can state the following theorem,

- **Theorem**: If $g_1(x,y)$ and $g_2(x,y)$ are continuous and differentiable and the mapping $(g_1, g_2)(x,y) \longmapsto (z,\omega)$ is one-one, then let (h_1, h_2) be the inverse mappings of (g_1, g_2), ie $x = h_1(z,\omega)$ and $y = h_2(z,\omega)$. If $(g_1, g_2): A \longmapsto B$, then

$$\int_{(x,y)\in A} f(x,y)\,dx\,dy = \int_{(z,\omega)\in B} \frac{f(h_1(z,\omega), h_2(z,\omega))}{|J(x,y)|_{\substack{x = h_1(z,\omega)\\ y = h_2(z,\omega)}}}\,dz\,d\omega$$

Where,

$$J(x_i, y_i) = \begin{vmatrix} \frac{\delta g_1(x,y)}{\delta x} & \frac{\delta g_1(x,y)}{\delta y} \\ \frac{\delta g_2(x,y)}{\delta x} & \frac{\delta g_2(x,y)}{\delta y} \end{vmatrix}_{\substack{x = x_i\\ y = y_i}}$$

In the context of probability,

- **Theorem**: Let X and Y be 2 jointly continuous Rvs and let $Z = g_1(X,Y)$ and $W = g_2(X,Y)$, where g_1 and g_2 are continuous and differentiable function, then,

$$f_{ZW}(z,\omega) = \sum_{i=1}^{n} \frac{f_{XY}(x_i, y_i)}{|J(x_i, y_i)|}$$

Where (x_i, y_i), $i \in [1:n]$ are the n solutions of $g_1(x,y) = z$ and $g_2(x,y) = \omega$.

Example: Let $X, Y \sim f_{XY}$, and $Z = aX + bY$, $W = cX + dY$

Assume $ad - bc = 0$. Find $f_{ZW}(z, \omega)$.

$$ax + by = z$$
$$cx + dy = \omega$$

$$ax = z - by \Rightarrow x = \frac{z - by}{a}$$

$$\Rightarrow \frac{cz - cby}{a} + dy = \omega$$

$$cz - cby + ady = a\omega$$

$$cz - y(cb - ad) = a\omega$$

$$y = \frac{a\omega - cz}{ad - bc}$$

$$x = \frac{z - b\left(\frac{a\omega - cz}{ad - bc}\right)}{a}$$

$$= \frac{zad - \cancel{zbc} - ba\omega + \cancel{zbc}}{\cancel{a}(ad - bc)}$$

$$x = \frac{zd - b\omega}{ad - bc}$$

$$J = \begin{vmatrix} \frac{\delta g_1(x,y)}{\delta x} & \frac{\delta g_1(x,y)}{\delta y} \\ \frac{\delta g_2(x,y)}{\delta x} & \frac{\delta g_2(x,y)}{\delta y} \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\therefore f_{ZW}(z, \omega) = \frac{f_{XY}\left(\frac{zd - b\omega}{ad - bc}, \frac{a\omega - cz}{ad - bc}\right)}{|ad - bc|}$$

(Only solns are $(x_1, y_1) = \left(\frac{zd - b\omega}{ad - bc}, \frac{a\omega - cz}{ad - bc}\right)$)

→ **Moment Generating Function :-**

- The moment generating function (MGF) associated with a RV X is given as $M_X : \mathbb{R} \longmapsto [0, \infty)$ st,

$$\boxed{M_X(s) = E[e^{sX}]}$$

The ROC of $M_X(s)$ is $\{ s \in \mathbb{R} : M_X(s) < \infty \}$

- Discrete X : $M_X(s) = \sum_{x \in X} e^{xs} P_X(x)$

Continuous X : $M_X(s) = \int_{-\infty}^{\infty} f_X(x) e^{xs} \, dx$

- **Application of MGFs :-**

1) Convenient computation of moments

2) Can be used to solve problems that involve the summation of RVs.

- **Theorem:** Suppose $M_X(s)$ is finite $\forall$ $s \in [-\varepsilon, \varepsilon]$ for some $\varepsilon > 0$, then,

$$\frac{d}{ds} M_X(s) \Big|_{s=0} = E[X]$$

$$\frac{d^n}{ds^n} M_X(s) \Big|_{s=0} = E[X^n] \longrightarrow n^{th} \text{ moment of } X$$

Proof:

$$\frac{d}{ds} M_X(s) = \frac{d}{ds} E[e^{sX}] = E[Xe^{sX}]$$

$$\Rightarrow \left.\frac{d}{ds} M_X(s)\right|_{s=0} = E[X]$$

$$\text{III ly} \quad \left.\frac{d^n}{ds^n} E[e^{sX}]\right|_{s=0} = \left.E[X^n e^{sX}]\right|_{s=0} = E[X^n]$$

- If $Y = aX + b$, $M_Y(y) = e^{bs} M_X(as)$
- If $Z = X + Y$, $M_Z(z) = M_X(x) M_Y(y)$, if X and Y are indp.
- Not all RVs have a converging MGF. ex: Cauchy distribution, X st $f_X(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$.

→ Characteristic Function :-

- The characteristic function of an RV X is given by,

$$\phi_X(t) = E[e^{itx}], \quad i = \sqrt{-1}$$

$$\Rightarrow \phi_X(t) = \int_{-\infty}^{\infty} e^{itx} \cdot f_X(x)\, dx \quad \text{[LOTUS]}$$

MGF, ϕ_X ~ Laplace and Fourier respectively.

- <u>Inversion Theorem</u> :-

If X is a continuous RV with PDF f_X and characteristic function $\phi_X(t)$, then

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\,\phi_X(t)\,dt$$ [Analogous to inv. FT]

if f_X is differentiable at x.

- <u>Properties of Characteristic Function</u> :-

1) $\phi_X(0) = 1$

2) $|\phi_X(t)| \leqslant 1$

3) $\left.\frac{d^n}{dt^n}\phi_X(t)\right|_{t=0} = i^n E[X^n]$

4) $Y = aX + b$, $\phi_Y(t) = e^{ibt}\phi_X(at)$

5) $Z = X + Y$, X, Y are independent $\Rightarrow$ $\phi_Z(t) = \phi_X(t)\,\phi_Y(t)$

→ <u>Gaussian Random Vectors</u> :-

Two RVs X_1, X_2 are said to be jointly Gaussian if their joint PDF is,

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2|K|}}\exp\left(-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2\end{bmatrix} K^{-1}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}\right)$$

Where $\mu_i = E[X_i]$, $K_{ij} = Cov(X_i, X_j)$ K - Covariance Matrix

Note: $Cov(X,Y) = E[XY] - E[X]E[Y]$

$Cov(X,X) = E[X^2] - (E[X])^2 = Var(X) = \sigma^2$

- **Theorem**: Uncorrelated jointly Gaussian RVs are independent

Proof: X_1, X_2 are uncorrelated $\Rightarrow K = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$

$$K^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2} \begin{bmatrix} \sigma_2^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix}$$

$$\begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_1 \end{bmatrix} \begin{bmatrix} \sigma_2^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \frac{1}{\sigma_1^2 \sigma_2^2}$$

$$= \frac{\sigma_2^2 (x_1 - \mu_1)^2 + \sigma_1^2 (x_2 - \mu_2)^2}{\sigma_2^2 \sigma_1^2}$$

$$= \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}$$

$$\Rightarrow f_{X_1, X_2}(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2 (\sigma_1^2 \sigma_2^2)}} \exp\left(\frac{-(x_1 - \mu_1)^2}{2\sigma_1^2} - \frac{(x_2 - \mu_2)^2}{2\sigma_2^2} \right)$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2} \right) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(\frac{-(x_2 - \mu_2)^2}{2\sigma_2^2} \right)$$

$$= f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

$\Rightarrow$ X_1 and X_2 are independent.

- $X_1, X_2, X_3 \dots X_n$ are jointly Gaussian if

$$f_{X_1, X_2 \dots X_n}(\bar{x}) = \frac{1}{\sqrt{(2\pi)^n |K|}} \exp\left(-\frac{1}{2}(\bar{x}-\bar{\mu}) K^{-1} (\bar{x}-\bar{\mu})^T\right)$$

- If $X_1, X_2 \dots X_n$ are jointly Gaussian, then the individual Rvs are also Gaussian. The converse need not be true.

- A linear combination of jointly Gaussian Rvs is also Gaussian.

★ Tail Bounds and Limit Theorems :-

- If we need $P(X \geq a)$, sometimes it would be better if we have the bounds of the above probability instead of the actual value, say when the distribution of X is complex.

- If we have the bounds on the mean and variance of the distribution then it is possible to find the exact bounds on $P(X \geq a)$.

→ Markov's Inequality :-

If X is a non-negative Rv with $E[X] < \infty$, then,

$$P(X \geq a) \leq \frac{E[X]}{a}, \quad \forall\, a > 0$$

Proof: $X = x\left[\mathbb{1}\{x \geq a\} + \mathbb{1}\{x < a\}\right]$

$$\Rightarrow E[X] = E[x\mathbb{1}\{x \geq a\}] + E[x\mathbb{1}\{x<a\}]$$
$$\geq E[x\mathbb{1}\{x \geq a\}]$$
$$\geq a\,P(X \geq a)$$
$$\Rightarrow P(X \geq a) \leq \frac{E[X]}{a}$$

Intuitively, as a increases, $P(X \geq a)$ should decrease, as per the expectance.

→ **Chebyshev's Inequality** :-

If X is an RV with finite mean μ and finite variance σ^2, then,

$$P(|x-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Proof: Let $Y = |x-\mu|$

$$P(Y \geq c) = P(Y^2 \geq c^2)$$
$$\leq \frac{E[Y^2]}{c^2} \quad \text{(By Markov)}$$

$$\frac{E[Y^2]}{c^2} = \frac{E[|x-\mu|^2]}{c^2} = \frac{\sigma^2}{c^2} \quad \text{(By defn. of variance)}$$

$$\Rightarrow P(|x-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

An alternate form, $c = k\sigma$,

$$P(|x-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

→ **Chernoff Bounds** :-

Let X be a random RV with $M_X(s) = E[e^{sX}]$, for $s \in [-\varepsilon, \varepsilon]$, $\varepsilon > 0$, then,

$$P(X \geq a) \leq \inf_{s>0} \frac{E[e^{sX}]}{e^{as}}$$

(inf – Infimum, ie minimum of the set)

$$P(X \leq a) \leq \inf_{s<0} \frac{E[e^{sX}]}{e^{as}}$$

Proof: $P(X \geq a) = P(e^{sX} \geq e^{sa})$ (for $s > 0$)

$$\leq \frac{E[e^{sX}]}{e^{sa}}$$

$$\Rightarrow P(X \geq a) \leq \inf_{s>0} \frac{E[e^{sX}]}{e^{sa}}$$

|||ly $P(X \leq a) = P(e^{sX} \geq e^{sa})$ (for $s < 0$)

$$\Rightarrow P(X \leq a) \leq \inf_{s<0} \frac{E[e^{sX}]}{e^{sa}}$$

- In order of strictness, generally,

	Markov	<	Chebyshev	<	Chernoff
Info Used :	Expectance		Expectance, Variance		MGF

→ <u>Convergence of Random Variables</u> :-

- Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables on some probability space $(\Omega, \mathcal{F}, P)$, we say X_n converges to an RV X in probability if

$$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0 \quad \forall \; \varepsilon > 0$$

- <u>Weak Law of Large Numbers</u> :-

Let $X_1, X_2 \ldots$, be a sequence of IID RVs, with mean μ and variance $\sigma^2 < \infty$, we have $\forall \; \varepsilon > 0$,

$$\lim_{n \to \infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right) = 0$$

or, $\dfrac{\sum_{i=1}^{n} X_i}{n}$ converges to μ in probability.

<u>Proof</u>: Let $M_n = \dfrac{\sum_{i=1}^{n} X_i}{n}$,

$$E[M_n] = \mu, \quad Var(M_n) = \frac{\sigma^2}{n}$$

By Chebyshev, $P(|M_n - \mu| > \varepsilon) \leq \dfrac{Var(M_n)}{\varepsilon}$

$$\Rightarrow P(|M_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon}$$

$$\Rightarrow \lim_{n \to \infty} P(|M_n - \mu| > \varepsilon) \leq \lim_{n \to \infty} \frac{\sigma^2}{n\varepsilon} = 0$$

By Sandwich theorem,

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right) = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{n} X_i}{n} \text{ converges to } \mu .$$

The above proof is true only for finite variance. However the theorem holds true for infinite variance as well.

- **Convergence in Distribution** :- We say a sequence of RVs $\{X_n\}$ converges in distribution to X if,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

$\forall$ x at which $F_X(x) = P(X \leq x)$ is continuous.

- **Central Limit Theorem** :- Let $\{X_n\}$ be a sequence of i.i.d RVs with mean $\mu < \infty$ and variance $\sigma^2 < \infty$, then the RV

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to the Std. Gaussian RV $\mathcal{N}(0,1)$ as $n \to \infty$, ie,

$$\lim_{n\to\infty} P(Z_n \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt$$

Proof: Let $Y_i = \dfrac{X_i - n\mu}{\sigma}$, Y_i has mean 0 and variance 1

WLOG, assume X_i has mean 0 and variance 1,

To show,

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,1)$$

Assume $M_X(s) < \infty \ \forall\ s \in (-\varepsilon, \varepsilon)$, for some $\varepsilon > 0$.

$$\begin{aligned}
M_{Z_n}(s) &= E\left[e^{sZ_n}\right] \\
&= E\left[\exp\left(\frac{s\sum_{i=1}^{n} X_i}{\sqrt{n}}\right)\right] \\
&= E\left[\exp\left(\frac{sX_1}{\sqrt{n}}\right)\exp\left(\frac{sX_2}{\sqrt{n}}\right)\exp\left(\frac{sX_3}{\sqrt{n}}\right)\cdots\right] \\
&= M_{X_1}\left(\frac{s}{\sqrt{n}}\right) M_{X_2}\left(\frac{s}{\sqrt{n}}\right)\cdots \\
&= \left[M_X\left(\frac{s}{\sqrt{n}}\right)\right]^n
\end{aligned}$$

Let $g(s) = \left[M_X\left(\frac{s}{\sqrt{n}}\right)\right]^n$

$$\log g(s) = n \log M_X\left(\frac{s}{\sqrt{n}}\right)$$

$$\lim_{n\to\infty} \log g(s) = \lim_{n\to\infty} n \log M_x\left(\frac{s}{\sqrt{n}}\right)$$

$$= \lim_{t\to 0} \frac{\log M_x(st)}{t^2}$$

$$= \lim_{t\to 0} \frac{\frac{1}{M_x(st)} M_x'(st) . s}{2t}$$

$$= \lim_{t\to 0} \frac{s}{2} \left(M_x'(st) \left(\frac{-1}{M_x(st)^2} M_x'(st) s \right) + \frac{1}{M_x(st)} \left(M_x''(st) . s \right) \right)$$

$$= \lim_{t\to 0} \frac{s}{2} \left(-\frac{M_x'(st)^2 . s}{M_x(st)^2} + \frac{M_x''(st) . s}{M_x(st)} \right)$$

$$= \lim_{t\to 0} \frac{-M_x'(st)^2 s^2 + M_x(st) M_x''(st) s^2}{2 M_x(st)^2}$$

$$= s^2 \left(\frac{-M_x'(0)^2 + M_x(0) M_x''(0)}{2 M_x(0)^2} \right)$$

$$= s^2 \left(\frac{-E[X]^2 + E[X^2]}{2} \right)$$

$$= \frac{s^2 (\mathrm{Var}(X))}{2}$$

$$\Rightarrow \lim_{n\to\infty} \log g(s) = \frac{s^2}{2} \rightarrow \lim_{n\to\infty} g(s) = e^{s^2/2}$$

Mgf of $\mathcal{N}(0,1)$, $M_{\mathcal{N}}(s) = E[e^{s\mathcal{N}}]$

$$E\left[e^{sN}\right] = \int_{x\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) \exp(sx)\, dx$$

$$= \int_{x\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2+sx}{2}\right) dx$$

$$= \int_{x\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2+sx-\left(\frac{s}{2}\right)^2}{2}\right) \exp\left(\frac{s^2}{2}\right) dx$$

$$= e^{s^2/2} \int_{x\in\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\left(x-\frac{s}{2}\right)^2}{2}\right) dx$$

$$= e^{s^2/2} = \lim_{n\to\infty} g(s)$$

$$\Rightarrow \lim_{n\to\infty} g(s) = M_N(s)$$

Since the MGF transformation is one-one,

$$\Rightarrow \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} = \mathcal{N}(0,1)$$

$$\Rightarrow \lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i - n\mu}{\sqrt{n}\,\sigma} = \mathcal{N}(0,1)$$

The above proof only works if $M_X(s) < \infty \ \forall\, s \in (-\varepsilon, \varepsilon)$, for some ε. If $\nexists$ such an ε, the above proof is invalid. But CLT still holds and can be shown via an alternate proof involving complex analysis.

- <u>Normal Approximation based on CLT</u> :-

Let $S_n = \sum_{i=1}^{n} X_i$, where X_i are IIDs with mean μ and variance σ^2.

To compute $P(S_n \leq x)$, if n is large, we can use the following approximation,

$$P(S_n \leq x) = P(S_n - n\mu \leq x - n\mu)$$
$$= P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{x - n\mu}{\sigma\sqrt{n}}\right)$$
$$\approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

Where $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx$.

- <u>Convergence in Mean Square Sense</u> :-

We say that a sequence of RVs $\{X_n\}$ converges to X in the mean square sense if.

$$\lim_{n\to\infty} E[(X_n - X)^2] = 0$$

- <u>Theorem</u>: If $X_n \to X$ in mean square sense, then $X_n \to X$ in probability as well.

<u>Proof</u>: $X_n \longrightarrow X$ in MSS, then $\lim_{n\to\infty} E[(X_n - X)^2] = 0$

$$P(X_n - X > \varepsilon) = P((X_n - X)^2 > \varepsilon^2)$$
$$\leq P((X_n - X)^2 \geq \varepsilon^2)$$
$$\leq \frac{E[(X_n - X)^2]}{\varepsilon^2} \quad \text{(Markov's inequality)}$$

$$\Rightarrow \lim_{n\to\infty} P(X_n - X < \varepsilon) \leq \lim_{n\to\infty} \frac{E[(X_n-X)]^2}{\varepsilon^2} = 0$$

$$\Rightarrow \lim_{n\to\infty} P(X_n - X < \varepsilon) = 0$$

$\therefore X_n \longrightarrow X$ in MSS $\Rightarrow X_n \longrightarrow X$ in probability.

- **Almost Sure Converge :-**

A sequence of RVs $\{X_n\}$ is said to converge almost surely if,

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

- **Strong Law of Large Numbers :-**

Similar to WLLN, SLLN also deals with the convergence of the sample mean to the true mean. However SLLN uses almost sure convergence, which is stronger that convergence in probability.

- Let $\{X_n\}$ be a sequence of IID RVs with mean μ, then,

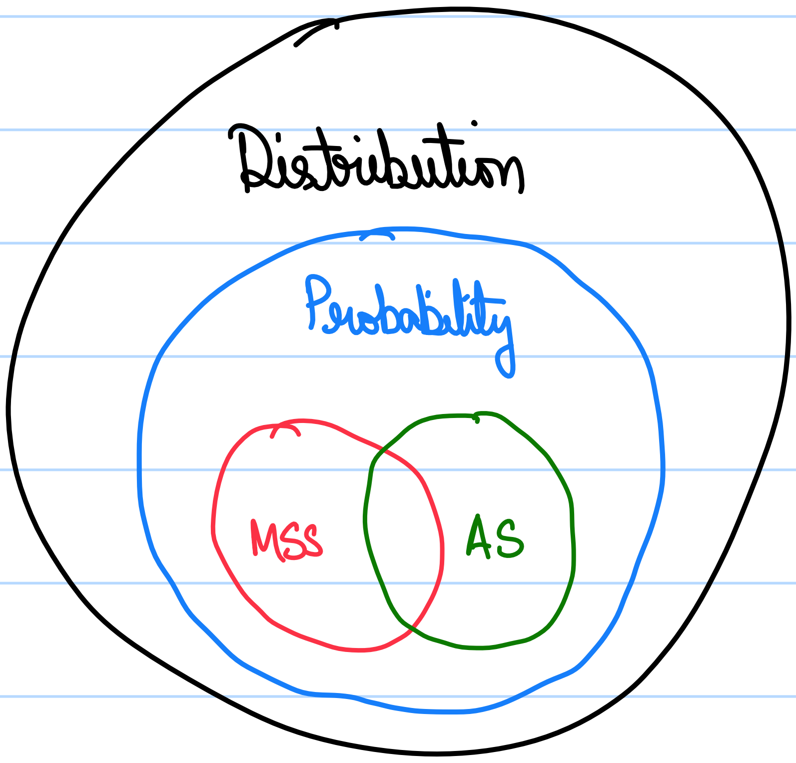
$$\frac{\sum_{i=1}^{n} X_i}{n} \longrightarrow \mu \text{ almost surely, ie,}$$

$$P\left(\left\{\omega : \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i(\omega)}{n} = \mu\right\}\right) = 1$$

- Hierarchy of Convergence Notions :-

- Theorem : The following implications hold,

$$\left.\begin{matrix} X_n \xrightarrow{as} X \Rightarrow \\ X_n \xrightarrow{MSS} X \Rightarrow \end{matrix}\right. \quad X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$$



Proof:

We already know that MSS $\Rightarrow$ Probability

To Prove Probability $\Rightarrow$ Distribution,

If $X_n \xrightarrow{P} X$, ie $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(X_n \leq x, X \leq x+\varepsilon) + P(X_n \leq x, X > x+\varepsilon) \\ &\leq F_X(x+\varepsilon) + P(|X_n - X| < \varepsilon) \end{aligned}$$

IIIly $F_X(x-\varepsilon) = P(X \leqslant x-\varepsilon)$

$$= P(X \leqslant x-\varepsilon, X_n \leqslant x) + P(X \leqslant x-\varepsilon, X_n > x)$$

$$\leqslant F_{X_n}(x) + P(|X_n - X| < \varepsilon)$$

$$\Rightarrow F_X(x-\varepsilon) - P(|X_n - X| < \varepsilon) \leqslant F_{X_n}(x) \leqslant F_X(x+\varepsilon) + P(|X_n - X| < \varepsilon)$$

Applying $\lim_{n\to\infty}$,

$$F_X(x-\varepsilon) \leqslant \lim_{n\to\infty} F_{X_n}(x) \leqslant F_X(x+\varepsilon) \quad \forall\ \varepsilon > 0$$

As $\varepsilon \to 0$, at x where $F_X(x)$ is continuous, by Sandwich theorem,

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{\varepsilon\to 0} F_X(x+\varepsilon) = F_X(x)$$

$$\Rightarrow \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

$\therefore$ Probability $\Rightarrow$ Distribution

To Prove Almost Sure $\Rightarrow$ Probability,

If $X_n \xrightarrow{A.S} X$, then $P(\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$

For an $\varepsilon > 0$, define a sequence of events A_n by,

$$A_n = \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\}$$

Define another sequence of events B_n by,

$$B_n = \{\omega : |X_k(\omega) - X(\omega)| < \varepsilon \ \forall\ k \geq n\}$$

We can see that $B_n \subseteq A_n$ and $B_1 \subseteq B_2 \subseteq B_3 \subseteq \cdots$, so,

$$\lim_{n\to\infty} P(B_n) = P\left(\bigcup_{n=1}^{\infty} B_n\right)$$

Also, $\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} \subseteq \bigcup_{n=1}^{\infty} B_n$, since $\exists\ n_0 \in \mathbb{N}$ st.

$$|X_n(\omega) - X(\omega)| < \varepsilon, \ \forall\ n \geq n_0$$ (Conventional defn. of convergence)

$$\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon \ \forall\ n \geq n_0\}$$
$$= B_{n_0} \subseteq \bigcup_{n=1}^{\infty} B_n$$

$$P\left(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1$$
$$\Rightarrow \quad P(B_{n_0}) = 1$$
$$\Rightarrow \quad P\left(\bigcup_{n=1}^{\infty} B_n\right) = 1$$
$$\Rightarrow \quad \lim_{n\to\infty} P(B_n) = 1$$

Since $B_n \subseteq A_n$, $P(B_n) \leq P(A_n) \leq 1$, So by Sandwich theorem,

$$\lim_{n\to\infty} P(A_n) = 1$$

$$\Rightarrow \quad \lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$$

$$\Rightarrow \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

$\therefore$ Almost Sure $\Rightarrow$ Probability.

★ **Random Processes** :-

A collection of Rvs usually indexed by time.

- **Discrete RP** : $(X_n \,;\, n \in \mathbb{Z})$
- **Continuous RP** : $(X_t \,;\, t \in \mathbb{R})$

For each $t \in \mathbb{R}$, X_t is a random variable.

$$\omega \longmapsto X_1(\omega), X_2(\omega), \ldots \quad \text{Discrete Rv}$$

$$\omega \longmapsto X_t(\omega) \;\forall\; t \in \mathbb{R} \quad \text{Continuous Rv}$$

- For a fixed ω, $(X_t(\omega), t \in T)$ is called the sample path at ω.
- **First-Order Distribution** : $F_{X_t}(x) = P(X_t \leq x)$
- **Second-Order Distribution** : $F_{X_{t_1}, X_{t_2}}(x_1, x_2) = P(X_{t_1} \leq x_1, X_{t_2} \leq x_2)$
- **nth Order Distribution** : $F_{X_{t_1}, X_{t_2} \cdots X_{t_n}}(x_1, x_2 \ldots, x_n) = P(X_{t_1} \leq x_1, \ldots X_{t_n} \leq x_n)$

→ Mean Function :-

The mean function of a RP $(X_t, t \in T)$ is defined as

$$\mu_X(t) = E[X_t]$$

- It gives us the expected value of the process at time t.

→ Correlation Function :-

$$R_X(t_1, t_2) = E[X_{t_1} X_{t_2}]$$

→ Covariance Function :-

$$C_X(t_1, t_2) = Cov(X_{t_1}, X_{t_2})$$
$$= E[X_{t_1} X_{t_2}] - E[X_{t_1}]E[X_{t_2}]$$
$$C_X(t_1, t_2) = R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$$

→ Bernoulli Process :-

$X_1, X_2, \ldots,$ are iid Bernoulli RVs with,

$$P(X_i = 1) = p, \quad P(X_i = 0) = 1 - p, \quad p \in [0,1]$$

- The random process $X = (X_n ; n \in \mathbb{N})$ can be visualized as a sequence of coin tosses in time.

- It is used to model systems involving arrivals of customers to a business. Arrival $\Rightarrow P(X_i = 1)$, No arrival $\Rightarrow P(X_i = 0)$, in the i^{th} period.

- The no. of successes S in n trials follows a Binomial (n, p) distribution

$$P_S(k) = {}^nC_k\, p^k (1-p)^{n-k}, \quad k \in \{0, 1, 2 \dots\}$$
$$E[S] = np$$
$$Var(S) = np(1-p)$$

- No. of trials for the first success is Geometric (p)

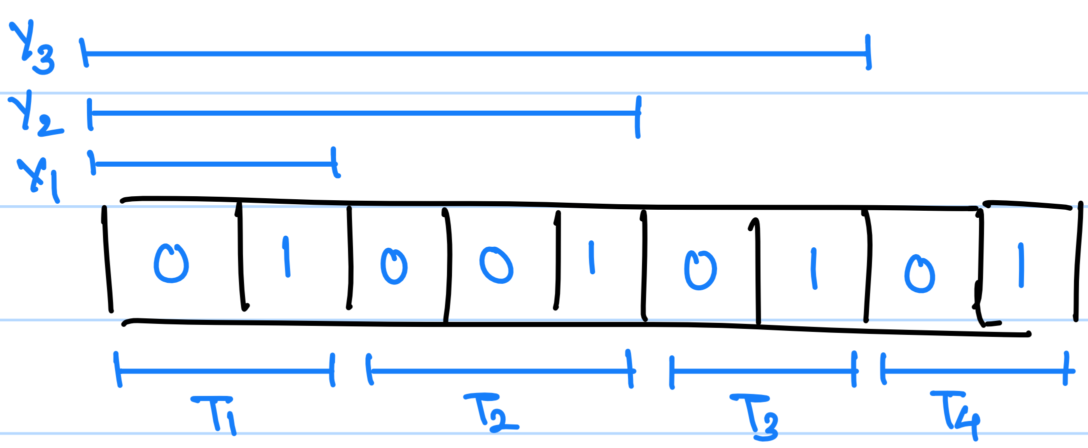
$$P_T(t) = (1-p)^{t-1} p, \quad t \in \{0, 1, 2, \dots\}$$
$$E[T] = 1/p$$
$$Var(T) = (1-p)/p^2$$

- <u>Fresh Start Property</u> :-

- For any $n \in \mathbb{N}$, $\hat{X}_i = X_{n+i}$; $i \in \mathbb{N}$ is also a Bernoulli process, independent of the past RVs.

- <u>Arrival Time</u>: Let Y_k denote the time of the k^{th} success, ie, k^{th} arrival time.

- <u>Inter-Arrival Time</u>: Let T_k represent the no. of trials following the $(k-1)^{th}$ success until the next success.



$T_1, T_2, T_3 \ldots$ are independent Rvs with the same geometric distribution

$$Y_k = \sum_{i=1}^{k} T_i \ , \quad E[Y_k] = \frac{k}{p} \ , \quad Var(Y_k) = \frac{k(1-p)}{p^2}$$