Lecture 5

7 Random Variable

The **sample space**, defined as the set of all possible outcomes of an experiment, represents the *set of uncertainty*.

Depending upon the underlying random experiment, we can have different types of sample spaces:

- S is finite.
- S is countably infinite.
- S is uncountable.

Examples

1. **Number of games to win the first match**:
   - Minimum: 1
   - Maximum: Unknown upper limit
2. **Number of bikers wearing helmets**:
   - Minimum: 0
   - Maximum: Unknown
3. **Number of patients visiting a doctor**:
   - Minimum: 0
   - Maximum: Unknown
4. **Lifetime / Survival time**:
   - Uncountable
   - Represented in terms of an interval (a, b)

Example: Random Variable from Sample Space

Consider 5 coins and 5 matches won. Let $X \in \{0, 1, 2, 3, 5\}$ represent the number of matches won.

Elements from the sample space are mapped to real numbers via a **mathematical function**, which is defined as a **random variable**:

$$X : S \to \mathbb{R}$$

Example: Tossing a Coin Twice

Sample space:

$$S = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$$

Let X = number of heads:

$$X \in \{0, 1, 2\}$$

Support of a random variable: The set of values of the random variable, corresponding to which we have non-zero probability.

Can $X = 3$?

Yes, but with probability $P = 0$.

Probability Distribution of X

X	0	1	2
P_X	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Here,

$$P(X = 0) = P(\mathrm{TT}) = P(\text{1st toss T} \cap \text{2nd toss T})$$

Since the tosses are independent:

$$P(\mathrm{TT}) = P(T) \cdot P(T)$$

In general:

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

Remark

If we map elements from S to $\mathbb{R}$, the resulting function is a **random variable**.

7.1 Types of Sample Spaces and Random Variables

- S binary: $X \in \{0, 1\}$
- S finite: $X \in \{0, 1, \ldots, n\}$
- S infinite countable: $X \in \{0, 1, 2, \ldots\}$
- S uncountable: $X \in (a, b)$

Sample spaces can be:

- Finite
- Countably infinite

- Uncountable

In such cases, if X takes discrete values, we have a **Discrete Random Variable**.

7.2 Discrete Random Variables

A **Discrete Random Variable** X answers the question "How many?" or represents counting/categorizing outcomes. The values of X are taken from the set of real numbers (may be ranked/nominal).

Let:

$$X : \{x_1, x_2, \ldots, x_n\}, \quad P(X = x_i) = p_i$$

Here, p_i is given by:

$$p_i = P(X = x_i) \quad \text{(Probability mass function, PMF)}$$

Conditions:

1. $p_i \geq 0, \quad \forall i$
2. $\displaystyle\sum_{i=1}^{n} p_i = 1$

7.3 Defining a Random Variable and Probability Distribution

For defining any random variable:

$$x : \quad -1,\ 0,\ 15,\ 20 \quad \text{(any real values)}$$

For probability distribution:

$$p : \quad a,\ b,\ c,\ d$$

where $a > 0, b > 0, c > 0, d > 0$ and $a + b + c + d = 1$.
For the same random variable, we can have different probability distributions. Example:

$$\left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\}, \quad \left\{\frac{1}{8}, \frac{3}{8}, \frac{1}{4}, \frac{1}{4}\right\}, \quad \left\{\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}\right\}$$

where $\sum p_i = 1$.

7.4 Distribution: Capturing Behaviour of Uncertainty

A **distribution** captures the behaviour or information of uncertainty.

- Every uncertainty can be explained with an underlying distribution.

- The task is to write the random variable and the underlying probability distribution.
- Values of a random variable can be:
  - Categories
  - Ordinal (e.g., 0–1)
  - Ranked
- We can have a probability distribution for these values.

Example

India vs England – 5 matches. Let X = number of wins by India:

$$X \in \{0, 1, 2, 3, 4, 5\}$$

The probability P_X can be obtained from historical/statistical data.

7.5 Experiments and Data Collection

Consider an experiment: "How many successes in n trials?" We can determine probabilities:

- From **historical data** (statistics)
- From **simulation**
- By processing **experimental observations**

In practice, we have:

$$\begin{aligned} X &: \quad x_1, x_2, \ldots, x_n \quad \text{(observations)} \\ P &: \quad p_1, p_2, \ldots, p_n \quad \text{(probabilities)} \end{aligned}$$

7.6 Cumulative Distribution Function (CDF) / Probability Law

The CDF accumulates probabilities:

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

Example: Let $X \in \{0, 1, 2, 3\}$ with:

$$P(X = 0) = \frac{1}{8}, \quad P(X = 1) = \frac{3}{8}, \quad P(X = 2) = \frac{3}{8}, \quad P(X = 3) = \frac{1}{8}$$

Then:

$$F_X(1) = P(X \leq 1) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$$
$$F_X(-5) = P(X \leq -5) = 0$$
$$F_X(0.2) = P(X \leq 0.2) = \frac{1}{8}$$
$$F_X(2) = P(X \leq 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

The CDF can be defined for *every* real number.

Graphical Representation

The CDF is a step function with jumps at each point x_i where $P(X = x_i) > 0$.

- Jumps correspond to the probability mass function (PMF).
- Maximum jump size = 1
- Minimum jump size = 0

7.7 Properties of CDFs

1. $\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to +\infty} F_X(x) = 1$
2. $F_X(x)$ is non-decreasing.
3. $F_X(x)$ is **right-continuous**:
$$F_X(x) = F_X(x^+)$$
It may have left-hand limits (LHL) but is not necessarily left-continuous (for discrete RVs, it has jumps).

7.8 CDF Defines a Distribution

If $f_X(x)$ is given by:

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

We check if it satisfies CDF properties:

- Non-decreasing ✓
- $F_X(-\infty) = 0, \quad F_X(+\infty) = 1 \quad \checkmark$
- Right-continuity at all points ✓ except at $x = 0$ (must be verified)

If a function is not right-continuous at some point, it is **not** a valid CDF.

8 Relationship Between CDF and Probability Distribution

Points of Discontinuity

For a discrete random variable X, the **points of discontinuity** of the cumulative distribution function (CDF) correspond to the values in the support of X.

Example:

$$\text{Points of discontinuity: } 0, 1, 2$$

$$P_X(0) = \frac{1}{2}, \quad P_X(1) = \frac{1}{4}, \quad P_X(2) = \frac{1}{4}$$

CDF $\leftrightarrow$ Probability Distribution

Given the CDF $F_X(x)$, the probability mass function (PMF) for a discrete random variable can be obtained as:

$$p_X(x) = F_X(x^+) - F_X(x^-)$$

where:

- $F_X(x^+) = \lim_{t \to x^+} F_X(t)$ (right-hand limit)
- $F_X(x^-) = \lim_{t \to x^-} F_X(t)$ (left-hand limit)

These jumps in the CDF correspond to the probabilities $P(X = x)$.

8.1 From Sample Space to Random Variable

The sample space S (set of uncertainties) can be mapped to a discrete random variable X. Once X is defined, we can specify its probability distribution without explicitly working with the full probability space $(S, \mathcal{F}, P)$ in detail.