LECTURE 8

Agenda:

1. Random variables
2. Probability mass function
3. Probability distribution function

RANDOM VARIABLES

Many times when we observe a random experiment, there is a numerical quantity associated with the experiment that we are interested in. For example,

(i) Experiment : Sample n people (WOR) from a population of N people
Quantity of interest : The average height

(ii) Experiment : A pandemic spreading through town
Quantity of interest : Percentage of people affected

(iii) Experiment : Toss a coin 3 times
Quantity of interest : # of heads

These quantities of interest associated with random experiments are called RANDOM VARIABLES, generally denoted by $X, Y, Z, \dots\dots$.

DEFINITION: A <u>random variable</u> is a real-valued function whose domain is the sample space.

Notationally, $X : S \longrightarrow \mathbb{R}$.

The set of possible values that a random variable X takes, or equivalently, the range of X is generally denoted by $\mathcal{X}$.

DEFINITION: A random variable is called <u>discrete</u> if the set of possible values that it takes is discrete.

PROBABILITY MASS FUNCTION

We focus on discrete random variables for a while. It is quite important to know the various chances with which the random variables takes various values. Mathematically, we need to know $P(X = x)$ for every $x \in \mathcal{X}$.

DEFINITION: The <u>probability mass function</u> of a random variable X (discrete), is given by

$$p_X(x) \triangleq P(X = x) \quad \text{for every } x \in \mathcal{X}.$$

Note that $p_X(x) \geq 0$ and $\sum_{x \in \mathcal{X}} p_X(x) = 1$.

Example 1: Sample n people (WOR) from a population of N people, with no preference to any specific person.

Quantity of interest (X) = The average height

S = All subsets of size n of the N people

Typical sample point $\underline{x} = \underbrace{(x_1, x_2, ---, x_n)}_{\text{unordered}}$.

Mathematically $X : S \longrightarrow \mathbb{R}$ with with

$$X(\underline{x}) = \frac{\sum_{i=1}^{n} x_i}{n} .$$

Suppose $N = 3$, $n = 2$. Hence, we have 3 people with heights say 50, 60, 70 inches. Hence,

$$S = \{ (M1, M2), (M2, M3), (M3, M1) \}$$

$$P((M1, M2)) = P((M2, M3)) = P((M3, M1)) = \frac{1}{3} .$$

$$\left.\begin{array}{l} X((M1, M2)) = 55 \\ X((M2, M3)) = 65 \\ X((M3, M1)) = 60 \end{array}\right\} \Rightarrow \mathcal{X} = \{55, 60, 65\}$$

Also, $P(X = 55) = P(X = 60) = P(X = 65) = \frac{1}{3}$.

Hence, the probability mass function is given by

$$p_X(55) = p_X(60) = p_X(65) = \frac{1}{3}.$$

<u>EXAMPLE 2:</u> Toss a fair coin 3 times.
$X =$ # of heads (Quantity of Interest)

$$S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$$

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

$$X(HHH) = 3$$

$$X(HHT) = X(HTH) = X(THH) = 2$$

$$X(TTH) = X(THT) = X(HTT) = 1$$

$$X(TTT) = 0$$

This gives $\mathcal{X} = \{0, 1, 2, 3\}$.

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

$P(X=3) = \frac{1}{8}$.

Hence, the probability mass function is given by

$$p_X(0) = \frac{1}{8},\ p_X(1) = \frac{3}{8},\ p_X(2) = \frac{3}{8},\ p_X(3) = \frac{1}{8}.$$

Sometimes it is important to study a random variable by looking at its cumulative properties, i.e., probabilities of the type $P(X \le b)$ for any real number b.

> DEFINITION: The distribution function F for a random variable X is defined by
>
> $$F_X(b) \triangleq P(X \le b) \quad \text{for all } b \text{ in } \mathbb{R} \text{ (real numbers)}$$

If X is a discrete random variable,

$$F_X(b) = P(X \le b) = \sum_{x \in \mathcal{X} : x \le b} p_X(x) .$$

Consider the random variable X in Example 1. Note that

$$p_X(55) = p_X(60) = p_X(65) = \frac{1}{3} .$$

Hence,

$$F_X(b) = \begin{cases} 0 & \text{if } b < 55, \\ \frac{1}{3} & \text{if } 55 \le b < 60, \\ \frac{2}{3} & \text{if } 60 \le b < 65, \\ 1 & \text{otherwise.} \end{cases}$$

Let us look at the plot of F_X