9. Probability Distribution

- A random variable is a real-valued function whose domain is the sample space of a random experiment.
- The probability distribution of a random variable X is the system of numbers:

X:	x_1	x_2	…	x_n
P(X):	p_1	p_2	…	p_n

Where, $P_i > 0 = \sum_{i=1}^{n} p_i = 1, i = 1, 2, ..n$

Here, the real numbers $x_1, x_2, \ldots, x_n$ are the possible values of the random variable X and p_i ($i = 1, 2, \ldots, n$) is the probability of the random variable X taking the value of x_i i.e., $P(X = x_i) = p_i$

Cumulative Probability

Cumulative probability is the probability of observing less than or equal to a given number of successes.

Cumulative distribution function :

Let X be a discrete random variable. Suppose x1, x2, ... xn is the range set of X and p1, p2 ..... pn are the respective probabilities of values of X. The cumulative distribution function (c.d.f) of X at some fixed value x is denoted by F(x) and is defined as

Fx = P X ≤ x , x∈R

i.e. Fxi = P X ≤ xi = p1 + p2 + ... + pi ; i = 1,2, ... n

Cumulative Frequency Plot
It is a way of displaying a cumulative frequency graphically. It shows the number, percentage, or proportion of observations in a a data set that are less than or equal to particular values. c.d.f. of a discrete random variable is a step function.

- **Mean/expectation of a random variable:** Let X be a random variable whose possible values $x_1, x_2, x_3 \ldots x_n$ occur with probabilities $p_1, p_2, p_3 \ldots p_n$ respectively. The mean of X (denoted by$m$) or the expectation of X (denoted by E(X)) is the number $\sum_{i=1}^{n} x_i p_i$.

  That is, $E(X) = \mu = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots x_n p_n$

- **Variance of a random variable:** Let X be a random variable whose possible values $x_1, x_2 \ldots x_n$ occur with probabilities $p(x_1), p(x_2) \ldots p(x_n)$ respectively. Let m = E(X) be the mean of X. The variance of X denoted by Var (X) or ${\sigma_x}^2$ is calculated by any of the following formulae:
  - ${\sigma_x}^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$
  - ${\sigma_x}^2 = E(X - \mu)^2$
  - ${\sigma_x}^2 = \sum_{i=1}^{n} x_i^2 p(x_i) - \left[\sum_{i=1}^{n} x_i p(x_i)\right]^2$

- $\sigma_x^2 = E(X^2) - [E(X)]^2$ where $[E(X)]^2 \sum_{i=1}^{n} x_i^2 p(x_i)$

  It is advisable to students to use the fourth formula.

Probability Density Function

Let X be a continuos random variable. The function $f(x)$ is said to be a probability density function of X if

(i) $f(x) \geq 0$

(ii) ∫-∞∞fxdx=1

Let X take values in the interval (a, b) and let x1, x2 be an interval contained in (a, b). Then

1. $P(a < X < b)$ =∫abfxdx=1

Geometrically, this represents the area bounded by the curve $y = f(x)$, x-axis, $x = a$ and $x = b$. The area bounded by the curve is 1.

2. $P(x_1 < X < x_2) = P(x_1 \leq X < x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X \leq x_2)$ = ∫x1x2fxdx as $P(x = x_1) = P(x = x_2) = 0$

Geometrically, this represents the area bounded by the curve $y = f(x)$, x-axis, $x = x_1$ and $x = x_2$.

Distribution Function

Let X be a continuos random variable with probability density function $f(x)$. For every real number x_i, the distribution function $F(x_i)$ is defined as

Fxi=PX≤xi=∫-∞xifxdx

1. $P[X > x] = 1 - P[X \leq x] = 1 - F(x)$

2. $P(x_1 < X < x_2) = P(x_1 \leq X < x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$, for (x_1, x_2) contained in the range of X.