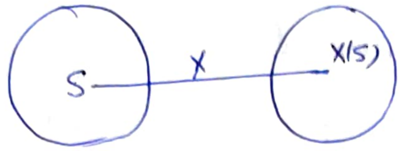
Probability Distribution:– Probability distribution is the theoretical counterpart of frequency distribution, and plays an important role in the theoretical study of populations. A probability model can be developed, for a given idealized conditions in a game of chance by incorporating all the factors that have a bearing on this game. In building such model, the empirical data of frequency distribution, A.M., variance etc are to be taken into account. In the discrete case we consider discrete uniform distribution, Binomial, Hypergeometric, Poisson distributions. The continuous Probability distributions we study are uniform distribution, normal distribution, Exponential, gamma, Weibull distributions which are of great practical importance.

Recall that in a random experiment, the outcomes (or results) are governed by chance mechanism and the sample space S of which such a random experiment consists of all outcomes of the experiment. When the elements (outcomes / events) of the sample space are non-numeric, they can be quantified by assigning a real number to every event of the sample space. The assignment rule, known as the random variable (R.V.) provides the power of abstraction and thus discards unimportant finest-grain description of the sample space.

A random variable X on a sample space S is a function $X: S \to R$ from S to the set of real ~~nu~~ numbers R, which assigns a real number $X(s)$ to each sample points s of S.



S: Sample Space R_X: Possible values of X

Range space R_X is the set of all possible values of X is a subset of real numbers R.

Although X is called a random 'variable' note that it is infact a "single-valued function".

<u>Notation</u>:- If R.V. is denoted by X, then x (corresponding small letter) denotes one of its values.

<u>Discrete</u>:- A R.V. X is said to be discrete R.V. if its set of possible outcomes, the sample space S, is countable (finite or an unending sequence with as many elements as there are whole numbers)

<u>Continuous</u>:- A R.V. X is said to be continuous R.V. if S contains infinite numbers equal to the number of points on a line segment.

Probability Distributions:-

Discrete Probability Distribution:- Each event in a sample space has certain probability (or chance) of occurence (or happening). A formula representing all these probabilities which a discrete R.V. assumes is known as the discrete probability distribution.

Ex Let X denote the discrete R.V. which denotes the minimum of the two numbers that appear in a single throw of a pair of fair dice. Then X is a function from the sample space S consisting of 36 ordered pair {(1,1)(1,2) ,- - (6,6)} to a subset of real numbers {1,2,3,4,5,6}

The Event minimum 5 can appear in the following cases (occurence) (5,5) (5,6), (6,5). Thus R.V. X assigns to this event of the sample space a real number 3. The probability of such an event happening is $\frac{3}{36}$ since there are 36 exhaustive cases. This is represented as

$$P(X=x_i) = P_i = f(x_i) = P(X=5) = f(5) = \frac{3}{36}$$

Calculating in a similar way the other probabilities the distribution of probabilities of this discrete R.V. is denoted by the discrete probability distribution as follows:

$X = x_i$	1	2	3	4	5	6
$P(X=x_i) = f_i = P_i$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

Discrete Probability distribution, probability function or probability mass function of a discrete R.V. X is the function $f(x)$ satisfying the following conditions:

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

(ii) $\Sigma f(x) = 1$

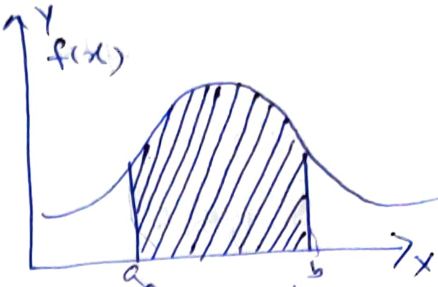
(iii) $P(X=x) = f(x)$

Thus Probability distribution is the set of ordered pairs $(x, f(x))$ ie, outcome x and its probability (chance) $f(x)$.

Continuous Probability Distribution:- For a continuous R.V. X, the function $f(x)$ satisfying the following is known as the probability density function (P.D.F) or simply density function

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

(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$



(iii) $P(a < X < b) = \int_a^b f(x)\,dx$ = area under $f(x)$ between ordinates $x = a$ and $x = b$

Note 1:- $P(a < X < b) = P(a \leq X < b)$

$= P(a < X \leq b) = P(a \leq X \leq b)$

ie inclusion or non-inclusion of end points, does not change the probability, which is not the case in the discrete distributions.

Note 2: Probability at a point

$$P(X = a) = \int_{a-\Delta x}^{a+\Delta x} f(x)\,dx$$

Theoretical Probability distribution:- Generally, frequency distribution are formed from the observed or experimental data. However, frequency distribution of certain populations can be deduced mathematically by fitting a theoretical probability distributions under certain assumption.

Examples:- The shoes-industry should know the 'sizes' of foot of the population, the food industry the 'tastes' (Menu) of the population etc.

Three such important theoretical distribution in order of their discovery are.

(i) Binomial (due to James Bernoulli, 1700)

(ii) Normal (due to De-Moivre 1733) also credited to Laplace (1774), Gauss (1809)

(iii) Poisson (due to S.D. Poisson 1837)

Discrete Probability distributions:- Binomial, Poisson, geometric, negative binomial, hypergeometric, multinomial, multivariate hypergeometric distributions.
Continuous Probability distributions:- Uniform (rectangular), normal, Gamma, Exponential, χ^2, Beta, bivariate normal, 't', 'F' distributions.

Discrete Probability Distribution:-

Ex1. Prove that (a) $E(KX) = KE(X)$

(b) $E(X+K) = E(X) + K$

(c) $E(X+Y) = E(X) + E(Y)$

Soln (a) $E(KX) = \dfrac{\Sigma K f_i x_i}{\Sigma f_i} = K \dfrac{\Sigma f_i x_i}{\Sigma f_i} = KE(X)$

(b) $E(X+K) = \dfrac{\Sigma f_i (x_i + K)}{\Sigma f_i} = \dfrac{\Sigma f_i x_i}{\Sigma f_i} + K \dfrac{\Sigma f_i}{\Sigma f_i}$

$= E(X) + K$

(c) $E(X+Y) = \dfrac{\Sigma f_i (x_i + y_i)}{\Sigma f_i} = \dfrac{\Sigma f_i x_i}{\Sigma f_i} + \dfrac{\Sigma f_i y_i}{\Sigma f_i}$

$= E(X) + E(Y)$

Note1:- Above results can be proved for Continuous Case by 'replacing' Σ by $\int_{-\infty}^{\infty}$

Note 2:- Above results are rewritten in 'μ' notation as (a) $\mu_{KX} = K\mu_X$ (b) $\mu_{X+K} = \mu_X + K$ (c) $\mu_{X+Y} = \mu_X + \mu_Y$

Q.2 Prove that (a) $Var(X+K) = Var(X)$

(b) $Var(KX) = K^2 Var(X)$

Hence $\sigma_{X+K} = \sigma_X$ and $\sigma_{KX} = |K|\sigma_X$.

Soln $Var(X+K) = \Sigma(x_i + K)^2 f(x_i) - \mu_{X+K}^2$

by using the result $Var(X) = E(X^2) - \mu_X^2$

$= \Sigma(x_i^2 + K^2 + 2Kx_i) f(x_i) - (\mu_X + K)^2$

$= \Sigma x_i^2 f_i + K^2 \Sigma f_i + 2K \Sigma x_i f_i - (\mu_X^2 + K^2 + 2\mu_X K)$

$= \Sigma x_i^2 f_i + K^2 + 2K\mu_X - \mu_X^2 - 2K\mu_X - K^2$

$= [Var(X) + \mu_X^2] - \mu_X^2 = Var(X)$

(b) $Var(KX) = \Sigma(Kx_i)^2 f_i - \mu_{KX}^2$

$= K^2 \Sigma x_i^2 f_i - (K\mu_X)^2 = K^2(\Sigma x_i^2 f_i - \mu_X^2)$

$= K^2 Var(X)$.

Q.2 Determine the discrete probability distribution, Expectation, variance, S.D of a discrete random variable (D.R.V.) X which denotes the minimum of the two numbers that appear when a pair of fair dice is thrown once.

Soln The total no. of cases are $6 \times 6 = 36$

The minimum number could be 1, 2, 3, 4, 5, 6, ie,

$X(S) = X(a,b) = \min\{a,b\}$. The number 6 will appear only in one case (6,6), so

$f(6) = P(X=6) = P(\{6,6\}) = \frac{1}{36}$

For minimum 5, favourable cases are (5,5), (5,6) (6,5) so $f(5) = P(X=5) = \frac{3}{36}$

Q:3 A player tosses a fair coins. He wins Rs 500 if 3 heads occur, Rs 300 if 2 heads occur, Rs 100 if one head occurs. On the other hand, he losses Rs. 1500 if 3 tails occur. Find the value of the game to the player. Is it favourable?

Solⁿ Let X = D.R.V = number of heads occuring in 3 tosses of a fair coin. The sample space S is

$$S = \{H, T\} \times \{H, T\} \times \{H, T\}$$

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

Probability of all 3 heads $= P(X=3) = \frac{1}{8}$

Probability of all 3 tails $= P(X=0) = \frac{1}{8}$

Probability of 2 Heads $= P(X=2) = \frac{3}{8}$

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

Discrete probability distibution is

$X = x_i$	0	1	2	3
$P(X=x_i) = f(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Expected value of the game

$$= 500 \times \frac{1}{8} + 300 \times \frac{3}{8} + 100 \times \frac{3}{8} - 1500 \times \frac{1}{8}$$

$$= \frac{200}{8} = 25 \text{ rupees.}$$

Game is favourable to the player since $E > 0$

For minimum 4, favourable cases are (4,4) (4,5) (4,6, 5,4) so

$$f(4) = P(X=4) = \frac{5}{36}$$

For minimum 3, favourable cases are (3,3) (3,4), (3,5) (3,6) ,(6,3) (5,3) (4,3) so

$$f(3) = P(X=3) = \frac{7}{36}$$

For minimum 2, favourable cases are (2,2) (3,3) (2,4) (2,5) (2,6) (6,2) (5,2) (4,2) (3,2) so

$$f(2) = P(X=2) = \frac{9}{36}$$

Similarly

$$f(1) = P(X=1) = \frac{11}{36}$$

Thus the required discrete probability distribution

$X = x_i$	1	2	3	4	5	6
$P(X=x_i) = f(x_i) = f_i$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

Mean = Expectation = $E(X) = \Sigma x_i f_i$

$$E(X) = 1 \cdot \frac{11}{36} + 2 \cdot \frac{9}{36} + 3 \cdot \frac{7}{36} + 4 \cdot \frac{5}{36} + 5 \cdot \frac{3}{36} + 6 \cdot \frac{1}{36}$$

$$= 2.5$$

$$Var(X) = \Sigma x^2 f_i - \mu^2$$

$$= 1 \cdot \frac{11}{36} + 4 \cdot \frac{9}{36} + 9 \cdot \frac{7}{36} + 16 \cdot \frac{5}{36} + 25 \cdot \frac{3}{36} + 36 \cdot \frac{1}{36} - (2.5)^2$$

$\sigma^2 = 1.9745$, so $\sigma = S.D = 1.4$