Random variables

i) Random variable :-

Random variable

Def (i) : Random variable is a real valued function defined on each and every outcome of the sample space connected with the random experiment.

On the other hand a function which can take real numbers like 0, 1, 2, 3 ..... with certain probability is also called as Random variable.

Def (ii) : For a mathematical definition of the random variable.

Let (S, B, P) be the probability space.

Where S : sample space of the random experiment

B : Borel σ-field. It is the set of all subsets in S.

P : probability function on B.

Def (iii) : A Random variable is a function $x(\omega)$ with domain S and range $(-\infty ; +\infty)$ such that for every real number a, the event $\{\omega : x(\omega) \leq a\} \in B$

where B is borel σ-field

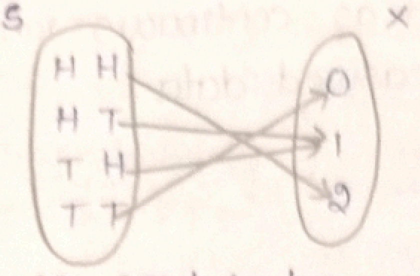
Example : consider a random experiment of two tosses of a coin.

Then the sample space $S = \{HH, HT, TH, TT\}$

Let x be number of heads, it can be explained below.

$x(HH) = 2 \quad x(HT) = 1 \quad x(TH) = 1 \quad x(TT) = 0$

Then the random variable x takes the values 0, 1, 2. It can be shown in the following mapping from



Here 'x' is said to be valued function defined on sample space 'S'. $R_x = \{0, 1, 2\}$

Types of Random variable :-

Random variable is of two types they are

1) Discrete Random variable
2) continuous Random variable

Discrete Random variable :-

If a random variable x is finite (or countably infinite) then x is said to be discrete random variable. (or)

A random variable x defined over a discrete sample space be called a discrete random variable. the discrete random variable represent the count data.

Example :-

1) The number of heads in tossing of a coin.
2) the number of defectives in a sample of k items.
3) The number of accidents per year.
4) the number of points on the dice in its thrown.

continuous Random variable :-

If a random variable takes all the possible values between certain limits is called a continuous random variable. (or)

A function which can take large number of values from an infinite interval is called as continuous random variable.

In other words a function which takes +ve, -ve, integer, decimal values from an infinite interval is also called as continuous random variable

A CRV represent measured data.

Examples:

i) Heights
ii) Weights
iii) Distance
iv) Area
v) volume
vi) the temperature recorded in a city
vii) Age of a group of persons

2M/5M

Probability function :-

A function which can satisfy certain axioms or conditions related to probability is called a probability function. It is of two types. They are

A) Discrete probability function
B) continuous probability function

A) Discrete probability function (or) probability mass function (P.M.F) :-

Let us consider x be a discrete random variable which assumes the values $x_1, x_2, \ldots x_i \ldots x_n$ with the respective probabilities $P_1, P_2, \ldots P_i \ldots P_n$.

In $P_i = P(x = x_i)$ or $p(x_i)$ is known as probability mass function and is denoted by P.M.F. If it satisfied the following properties.

i) $P(x_i) \geq 0 \ \forall \ i$
ii) $0 \leq p(x_i) \leq 1$
iii) $\sum_{i=1}^{n} p(x_i) = 1$

Ex:-

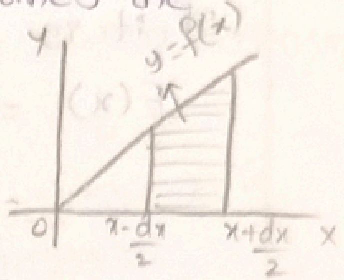
x_i :	0	1
$p(x_i)$:	$\frac{1}{2}$	$\frac{1}{2}$

i) $p(x_i) \geq 0$
ii) $0 \leq p(x_i) \leq 1$
iii) $\sum p(x_i) = 1$ Given problem is P.M.F

the set $(x_i, P(x_i))$ is called the probability distribution of a discrete random variable.

B) Continuous probability function (or) probability density function (p.d.f):-

Let 'x' is a continuous random variable. $x_1, x_2, \dots x_n$ be the values of continuous variable 'x'. $f(x_1), f(x_2) \dots f(x_i) \dots f(x_n)$ becomes the corresponding continuous probability. Now if any one of these continuous probability i.e $f(x_i)\ \forall\ i = 1, 2, \dots n$



$$f(x) = P\left(x - \frac{dx}{2} \le x \le x + \frac{dx}{2}\right)\ \forall\ x = x_1, x_2 \dots x_n$$

the function f(x) is called probability density function (p.d.f).

If it satisfies the following properties

i) $f(x) \ge 0\ \forall\ x,\ -\infty < x < \infty$

ii) $\int_{-\infty}^{\infty} f(x)dx = 1$, iv) $0 \le f(x) \le 1$

iii) the probability of an event in the interval (a, b) is calculated by

$$P(a \le X \le b) = \int_a^b f(x)dx = F(b) - F(a)$$

Example: $f(x) = \frac{x}{2}\quad 0 < x < 2$

i) $\int_0^2 \frac{x}{2}dx = \left(\frac{x^2}{4}\right)_0^2 = 1$

ii) $f(x) \ge 0$

$\therefore f(x)$ is p.d.f.

V Imp

Define Distribution function (or) Cumulative Distribution function and describe its properties :-

Let X is a random variable, $x_1, x_2, \dots x_n$ be the values of random variable X. $P(x_1), P(x_2) \dots P(x_n)$ becomes the corresponding probability. Now the distribution function of random variable x can be defined as

$$F(x) = p(X \le x) \quad \forall\ x = x_1, x_2, \dots x_n$$

If x is a discrete random variable then its distribution function can be defined as

$$F(x) = p(X \le x)$$
$$= \sum_{0}^{x} p(x)$$

If it is a continuous random variable then

$$F(x) = p(X \le x)$$
$$= \int_{-\infty}^{\infty} f(x)\,dx$$

Properties (or) characteristics of Distribution function :-

1) Like probability function, distribution function always becomes positive.
i.e $F(x) \ge 0$ like $p(x) \ge 0$

2) Like the probability function, distribution function always lies in between 0 and 1.
i.e $0 \le F(x) \le 1$ like $0 \le p(x) \le 1$

3) If we differentiate distribution then we get probability density function
i.e $F'(x) = \frac{d}{dx} F(x) = f(x)$

4) If integrate probability density function then we get distribution function.
i.e $\int_{-\infty}^{x} f(x)dx = F(x)$

5) If $a<b$ then $F(a)<F(b)$

6) $F(+\infty) = 1$

proof: By def $F(x) = p(X \le x) = \int_{-\infty}^{x} f(x)\,dx$

$F(+\infty) = P(X \le +\infty) = \int_{-\infty}^{\infty} f(x)\,dx$ = Total probability

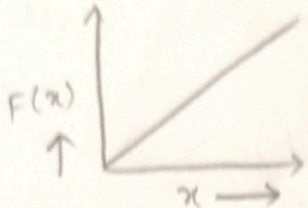
7) $F(-\infty) = 0$

proof: By definition

$$F(x) = p(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

similarly $F(-\infty) = p(X \le -\infty) = \int_{-\infty}^{-\infty} f(x)\,dx = 0$

8) Distribution function always becomes a continuous increasing function towards right side or non decreasing function or step function.



9) If X is a continuous variable then $p(a \le x \le b) = F(b) - F(a)$

proof: $$P(a \le x \le b) = \int_{a}^{b} f(x)\,dx = [F(x)]_{a}^{b} = F(b) - F(a)$$

10) If $F(x)$ is the distribution function of a random variable X, then $F(x) \le F(y)$ for $(x < y)$

proof: For $x < y$ from property $p(a < x < b) = F(b) - F(a)$

$$P(x < X < y) = F(y) - F(x)$$

since probability always ≥ 0

$$P(x < X < y) \ge 0$$

$$F(y) - F(x) \ge 0$$

$$F(y) \ge F(x)$$

$$F(x) \le F(y) \text{ for } x < y.$$

Various formulae on Discrete Random Variable:-

1) Mean $= \sum x \cdot p(x) = \mu_1'$

2) variance $= \sum x^2 p(x) - (\text{mean})^2$ (or)

$$\mu_2' = \sum x^2 p(x)$$

variance $= \mu_2 = \mu_2' - \mu_1'^2$

3) $F(x) = P(X \le x) = \sum^{x} P(X = x)$

Various Formulae on continuous Random Variable:

1) mean $= \int x f(x)dx = \mu_1'$

2) variance $= \int x^2 f(x)dx - (\text{mean})^2$

(or)

$\mu_2' = \int x^2 f(x)dx$

Variance $= \mu_2 = \mu_2' - \mu_1'^2$

3) Harmonic mean 'H' is given by

$$\frac{1}{H} = \int \frac{1}{x} f(x)\,dx$$

4) Geometric mean 'G' is given by

$$\log G = \int \log x \; f(x)dx$$

5) Median 'M' is given by (in the range a,b)

$$\int_a^M f(x)dx = \int_M^b f(x)\,dx = \frac{1}{2} \quad \text{i.e} \quad \int_a^M f(x)dx = \frac{1}{2} \text{ (or) } \int_M^b f(x)dx$$

6) Mean deviation about mean

$$MD = \int |x - \text{mean}|\, f(x)\,dx$$

7) r^{th} moment about origin $\mu_r' = \int x^r f(x)\,dx$

8) r^{th} moment about mean $\mu_r = \int (x - \text{mean})^r f(x)dx$

Problems:-

1) A Random variable x has the following probability function:

$X=x$	-2	-1	0	1	2	3
$P(x)$	0.1	K	0.2	2K	0.3	K

mean variance

Find i) K ii) mean and variance

iii) Construct distribution function and draw its graph

Sol:- i) Since Total probability $= 1$

$$\Sigma p(x) = 1$$

$$0.1 + K + 0.2 + 2K + 0.3 + K = 1$$

$$0.6 + 4K = 1$$

$$4K = 1 - 0.6$$

$$4K = 0.4$$

$K = \frac{0.4}{4} = 0.1$

x	-2	-1	0	1	2	3
$P(x)$	0.1	0.1	0.2	0.2	0.3	0.1

ii) $\mu_1' = \text{mean} = \Sigma x\, p(x) = -2\times0.1 + -1\times0.1 + 0\times0.2 + 1\times0.2 + 2\times0.3 + 3\times0.1$

$= -0.2 - 0.1 + 0 + 0.2 + 0.6 + 0.3$

$\mu_1' = \text{mean} = -0.3 + 1.1 = 0.8$

$\mu_2' = \Sigma x^2 p(x) = (-2)^2\times0.1 + (-1)^2\times0.1 + 0^2\times0.2 + 1^2 + 0.2 + (2)^2\times0.3 + 3^2\times0.1$

$\mu_2' = 4\times0.1 + 1\times0.1 + 0 + 0.2 + 4\times0.3 + 9\times0.1$

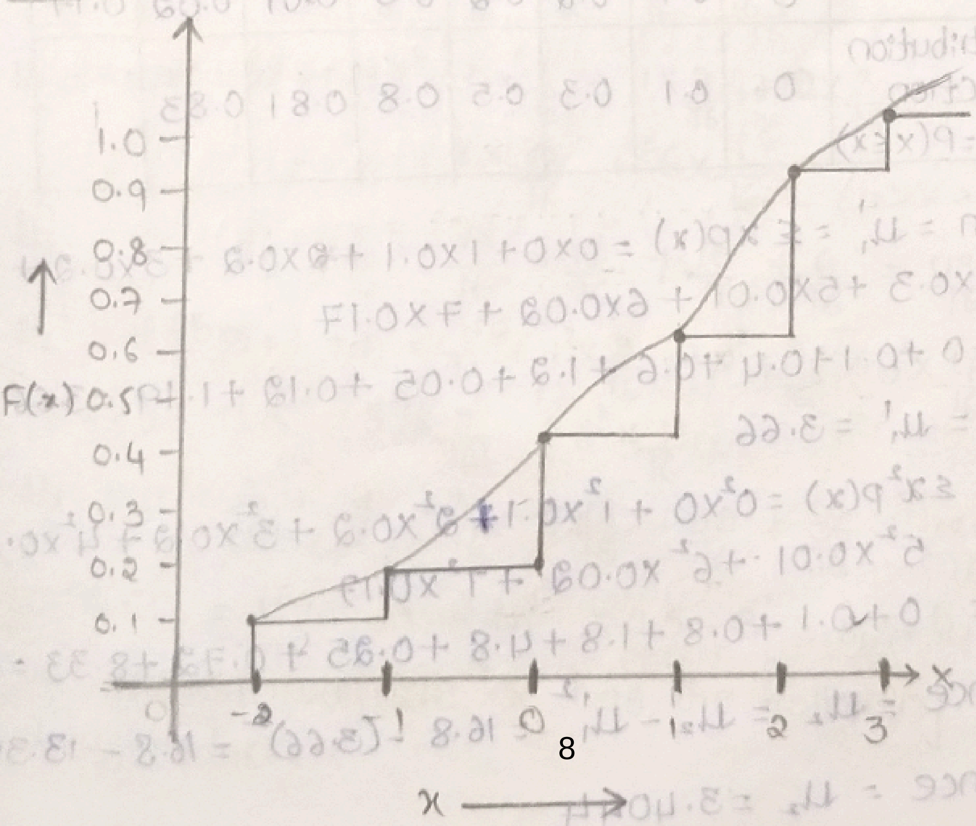
$\mu_2' = 0.4 + 0.1 + 0 + 0.2 + 1.2 + 0.9 = 2.8$

variance $= \mu_2 = \mu_2' - {\mu_1'}^2 = 2.8 - (0.8)^2 = 2.8 - 0.64 = 2.16$

iii) Distribution function

x	-2	-1	0	1	2	3
$P(x)$	0.1	0.1	0.2	0.2	0.3	0.1
$F(x)$	0.1	0.2	0.4	0.6	0.9	1

Graph:-



Bivariate Random variables :-

Bivariate Random variable : If two random variables are studied simultaneously in respect of their distribution then it is known as bivariate studies.

If two random variables are defined on same probability space. they are called joint distributed random variables. Here both of the random variable are discrete or continuous.

Ex:- Heights and weights of a group of persons.

Bivariate discrete Random variable :-

A random variable (x, y) is said to be a two dimensional discrete random variable if it can take only a countable number of points (x, y) in a two dimensional space. It is also called joint discrete random variable.

Ex:- Let us consider an experiment of rolling two dies simultaneously take the random variable x as the number of points on the first dice and y as the number of points on the second dice. (x, y) is a Bivariate discrete Random variable whose range $(x, y) = \{(1,1) (1,2) \ldots\ldots (6,6)\}$

Bivariate continuous random variable :-

A random variable (x, y) is said to be two dimensional continuous random variable. If it can take all possible points between certain lines (limit) of two dimensional space. It is also called joint continuous random variable.

Ex:- Heights and weights of a group of persons.

Joint probability mass function (or) Bivariate discrete probability distribution :-

Let x and y be two discrete random variable taking the values $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_n$ respectively.

The function P_{ij} defined on $X = x_i$ and $Y = y_i$ is called joint probability mass function of (X, Y) is denoted by $P(X = x_i, Y = y_i)$ or $P(X = x_i \cap Y = y_i)$ or $P(x_i, y_i)$ and is represented in the following bivariate table.

X \ Y	y_1	y_2	...	y_j	...	y_n	Marginal of X
x_1	P_{11}	P_{12}		P_{1j}		P_{1n}	$P_{1 \cdot}$
x_2	P_{21}	P_{22}		P_{2j}		P_{2n}	$P_{2 \cdot}$
⋮							
x_i	P_{i1}	P_{i2}		P_{ij}		P_{in}	$P_{i \cdot}$
⋮							
x_m	P_{m1}	P_{m2}		P_{mj}		P_{mn}	$P_{m \cdot}$
Marginal of Y	$P_{\cdot 1}$	$P_{\cdot 2}$		$P_{\cdot j}$		$P_{\cdot n}$	$P_{\cdot \cdot} = 1$

The joint probability mass function of X and Y is denoted by P_{ij} or $P(x_i, y_j)$

$$P(x_i, y_j) = P(X = x_i \cap Y = y_j) = P_{ij}$$

it should satisfy the following conditions

i) $P(x_i, y_j) \geq 0 \ \forall \ i, j$ iii) $0 \leq P_{ij} \leq$

ii) $\sum_{i=1}^{m} \sum_{j=1}^{n} P(x_i, y_i) = 1$

Example:-

X \ Y	1	2	3	Total
1	0.1	0.2	0.1	0.4
2	0.05	0.15	0.4	0.6
Total	0.15	0.35	0.5	1

i) $P(x_i, y_j) \geq 0$

ii) $\sum \sum P(x_i, y_j) = 1$

∴ It is joint probability mass function.

Marginal probability function (marginal probability distribution)

Marginal probability is the sum of the probabilities of two or more events. since they are shown on the margin or bottom of the rows and columns. they are called marginal probabilities.

X	x_1	x_2	x_i	x_m	$\sum_{i=1}^{m} P_{i\cdot} = 1$
$p(x_i)$	$P_{1\cdot}$	$P_{2\cdot}$	$P_{i\cdot}$	$P_{m\cdot}$	1

this is called marginal probability distribution of x.

Y	y_1	y_2	y_j	y_n	$\sum_{j=1}^{n} P_{\cdot j} = 1$
$P(y_j)$	$P_{\cdot 1}$	$P_{\cdot 2}$	$P_{\cdot j}$	$P_{\cdot n}$	1

This is called marginal probability distribution of y.

Marginal probability mass function:-

If $p(x_i, y_j)$ is the joint probability mass function of (x, y) then the marginal probability mass function of x is denoted by $p(x_i)$ or $p(X = x_i)$ or p_i and is defined as

$$P(X=x_i) = P(X=x_i, Y=y_1) + P(X=x_i, Y=y_2 + \cdots +$$
$$P(X=x_i, Y=y_n)$$

$$P(X=x_i) = P_{i1} + P_{i2} + \cdots + P_{in}$$

$$P(X=x_i) = \sum_{j=1}^{n} P_{ij} \text{ or } \left(\sum_{j=1}^{n} P(x_i, y_j)\right) \quad i = 1, 2, \ldots m$$

Not necessarily

$$P(x_i) = \sum_{j=1}^{n} P(x_i, y_j) \quad \forall \; i = 1, 2 \ldots m$$

Conditions:- i) $p(x_i) \geq 0$ (or) $P(X) \geq 0$

ii) $\sum_{i=1}^{m} p(x_i) = 1$ $\quad \leq P(X) = 1$

similarly the marginal probability mass function of y is $P(y_j) = \sum_{i=1}^{m} P(x_i, y_j) \quad \forall \; j = 1, 2, \ldots n$

conditions :- i) $P(y_j) \geq 0$ (or) $P(Y) \geq 0$

ii) $\sum_{j=1}^{n} P(y_j) = 1$ (or) $P(Y) = 1$

Note :-

MPMF of $X = P(x) = \sum_{y} p(x,y)$

MPMF of $Y = P(y) = \sum_{x} p(x,y)$

Conditional probability mass function

There are two conditional probability functions.

Conditional probability function of Y given X is denoted by $P(Y=y_j / X=x_i)$ or $P(Y/X)$

$$P(Y=y_j / X=x_i) = \frac{P(X=x_i \cap Y=y_j)}{P(X=x_i)} \text{ or } P(Y/X) = \frac{P(x,y)}{P(x)}$$

Properties:-

i) $P(Y=y_j / X=x_i) \geq 0 \ \forall j$

ii) $\sum_{j=1}^{n} P(Y=y_j / X=x_i) = 1$

The conditional probability function of X given Y is denoted by $P(X=x_i/Y=y_j)$ or $P(X/Y)$

$$P(X=x_i \mid Y=y_j) = \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)} \text{ (or) } P(x/y) = \frac{P(x,y)}{P(y)} \ [P(y) > 0]$$

Properties:-

i) $P(X=x_i \mid Y=y_j) \geq 0 \ \forall i$

ii) $\sum_{i=1}^{m} P(X=x_i / Y=y_j) = 1$

Bivariate Distribution : continuous Random variables

Joint probability density function

It is the probability function of a two dimensional continuous random variable. The function $f(x,y)$ is said to joint probability density function if it satisfy the following properties.

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

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

iii) $P(a<x<b, c<y<d) = \int_a^b \left(\int_c^d f(x,y)\,dy \right) dx$

Marginal probability density function :-

these are the probability functions of x and y individually.

the marginal probability density function of x is denoted by $f(x)$ or $f_x(x)$ and is defined as

$$\text{Mpdf of } x = f(x) = \int_y f(x,y)\,dy$$

properties : i) $f(x) \geq 0 \ \forall x$

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

The marginal probability density function of y is denoted by $f(y)$ or $f_y(y)$ and is defined as

$$f(y) = \int_x f(x,y)\,dx$$

properties :- i) $f(y) \geq 0 \ \forall y$

ii) $\int_{-\infty}^{\infty} f(y)\,dy = 1$

Conditional probability density function :-

suppose x and y are jointly distributed continuous random variable with j.p.d.f $f(x,y)$ and let the marginal probability density function of x and y be $f(x)$ and $f(y)$.

the conditional probability density function of y given x is denoted by $f(y/x)$ and is defined as $f(y/x) = \dfrac{f(x,y)}{f(x)}$ $\quad f(x)>0$

properties :-

i) $f(y/x) \geq 0 \ \forall y$

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

The conditional probability density function of x given y is denoted by $f(x/y)$ and is defined as

$$f(x/y) = \frac{f(x,y)}{f(y)} \quad f(y) > 0$$

properties:-

i) $f(x/y) \geq 0 \ \forall x$

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

no need

Distribution function of Bivariate Random variables:

Joint Distribution function or Joint probability Distribution function.

If X, Y are jointly distributed random variables. the joint distribution function of x and y is denoted by $F(x,y)$ and is defined as

$$F(x,y) = P(X \leq x,\ Y \leq y)$$

For discrete Random variables

$$F(x,y) = \sum_{i=-\infty}^{x} \sum_{j=-\infty}^{y} P(x_i, y_j)$$

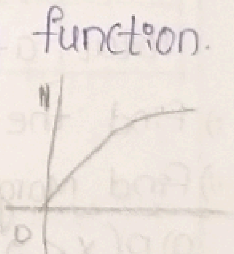
For continuous Random variables

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

properties of joint probability distribution function:-

i) $P(a_1 \leq X \leq b_1,\ a_2 \leq Y \leq b_2) = F(b_1, b_2) + F(a_1, a_2) - F(a_1, b_2) - F(b_1, a_2)$

ii) $F(x,y)$ is monotonically non-decreasing function. (or) increasing



iii) $0 \leq F(x,y) \leq 1$

iv) $F(-\infty, -\infty) = 0$

v) $F(+\infty, +\infty) = 1$

vi) We know that if the density function $f(x,y)$ is continuous then $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\,dy\,dx$

vii) $f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\, \partial y}$

Independence of Random variables :-

Let us consider x, y be two random variables with the joint p.d.f $p(x,y)$ or $f(x,y)$ and marginals $p(x), p(y)$ or $f(x), f(y)$ and continuous $p(x/y), p(y/x)$ or $f(x/y), f(y/x)$ respectively for discrete and continuous random variables. then the random variables x and y are said to be stochastically independent or independent if

$$p(x) = p(x/y)$$
$$p(y) = p(y/x)$$
$$f(x) = f(x/y)$$
$$f(y) = f(y/x)$$

Not necessary

the joint p.d.f can be expressed as

$$p(x,y) = p(x)\,p(y)$$
$$f(x,y) = f(x)\,f(y)$$