Random Variables

Random Variable:

Random variable is a function that assigns real number (value) to the sample space of random experiment.

OR

Random variable is a real-valued function defined on a sample space. It is also called chance variable, stochastic variable or simply a variate and it is abbreviated as r.v. It is denoted by Latin Letters X, Y, Z.

e.g.

Two coin toss.

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

Let

X = number of heads appear

$X(HH) = 2$

$X(HT) = 1$

$X(TH) = 1$

$X(TT) = 0$

$X = S_1 = \{0, 1, 2\}$ (Set)

or $S_2 = \{2, 1, 1, 0\}$

Domain		Range
HH	$\longrightarrow$	2
HT	$\longrightarrow$	1
TH	$\longrightarrow$	1
TT	$\longrightarrow$	0
(Total output)(Sample Space)		(Real Number)

Mathematically:

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

"X" is a function from sample space to real numbers.

Example:-

Red ball $\rightarrow$ 10/-

Blue ball $\rightarrow$ 50/-

Green ball $\rightarrow$ 100/-

$S = \{Red, Blue, Green\}$

Let X = amount of Prize won

$X(Red) = 10$

$X(Blue) = 50$

$X(Green) = 100$

$X = \{10, 50, 100\}$ (Possible set)

∵ real number

$-\infty$ to $+\infty$

Scanned with CamScanner

Domain Range

$R \rightarrow 10$

$B \rightarrow 50$

$G \rightarrow 100$

Example :-

A pair of dice thrown. What is the Probability of getting a total of ~~8~~ 2, 12.

$n(S) = 6^2 = 36$

Domain Range

$\{1,1\} \rightarrow 2$

$\{6,6\} \rightarrow 12$

OR

Let $X =$ Sum of dots appear on the dice.

Domain Range

$\{1,1\} \longrightarrow 2$

$\{1,2\}, \{2,1\} \longrightarrow 3$

$\{1,3\}, \{2,2\}, \{3,1\} \longrightarrow 4$

$\vdots \qquad \vdots$

$\{6,6\} \longrightarrow 12$

Types of Random Variable

i) Discrete random variable

ii) Continuous random variable

Discrete random variable

A random variable "X" is defined to be discrete if it can assume values which are finite or countably infinite.

e.g.

i) Number of heads obtained in coin tossing

ii) Number of defective items

iii) Number of fatal accidents

Continuous random variable

A random variable "X" is defined to be continuous if it can assume values which are infinite or uncountable.

e.g.

i) Number of stars in the sky

ii) Real numbers b/w $0 \leq X \leq 3$

0 — 1

infinite

∴ Infinite real numbers exist b/w two integers

Distribution Function

Let X be a random variable. The function F defined for all real x by $F(x) = P(X \le x) = P\{\omega : X(\omega) \le x\}$, $-\infty < x < \infty$ is called distribution function of the random variable "X".

A distribution function (d.f) is also called cumulative distribution function (cdf). It is also denoted by $F_X(x)$.

e.g.

Three coin are toss. $n(S) = 2^3 = 8$

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

Let

X = number of heads

$X = 0, 1, 2, 3$

$P(X=0) = P(TTT) = 1/8$

$P(X=1) = P(HTT, THT, TTH) = 3/8$

$P(X=2) = P(HHT, HTH, THH) = 3/8$

$P(X=3) = P(HHH) = 1/8$

X	0	1	2	3
$P(X=x)$	1/8	3/8	3/8	1/8

$F(0) = P(X \le 0) = P(X=0) = 1/8$

$F(1) = P(X \le 1) = P(X=0) + P(X=1) = 1/8 + 3/8 = 4/8 = 1/2$

$F(2) = P(X \le 2) = P(X=0) + P(X=1) + P(X=2) = 1/8 + 3/8 + 3/8 = 7/8$

$F(3) = P(X \le 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$= 1/8 + 3/8 + 3/8 + 1/8$

$= \frac{8}{8} = 1$

∵ We can calculate value of function at infinite real values.

$F(4) = P(X \le 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = 1$

$F(5) = 1$

$F(6) = 1$

Domain of d.f $= (-\infty, \infty)$

Range of d.f $= [0, 1]$

∵ When probability equal to zero in d.f?

e.g.

$F(-1) = P(X \le -1)$

$= 0$

$F(-2) = P(X \le -2)$

$= 0$

Let

$$F(2.5) = P(X \leq 2.5)$$
$$= P(X=0) + P(X=1) + P(X=2)$$
$$= 1/8 + 3/8 + 3/8 = 7/8$$

$$F(0.5) = P(X \leq 0.5)$$
$$= P(X=0)$$
$$= 1/8$$

$$F_X(x) = \begin{cases} 0 \;; & x < 0 \\ 1/8 \;; & 0 \leq x \leq 1 \\ 4/8 \;; & 1 \leq x < 2 \\ 7/8 \;; & 2 \leq x < 3 \\ 1 \;; & x \geq 3 \end{cases}$$

Properties of d.f

A d.f has following properties.

i) $F(-\infty) = 0$, $F(+\infty) = 1$; $[0, 1]$

$0 \leq F(x) \leq 1$ $\quad -\infty < x < \infty$

ii) $F(x)$ is increasing function, or non-decreasing function. $F(x_1) \leq F(x_2)$ if $x_1 \leq x_2$

iii) If F is the d.f of the r.v "X" and if $a < b$, then $P(a < X \leq b) = F(b) - F(a)$

Proof

The event "$a < X \leq b$" and "$X \leq a$" are disjoint and their union is the event "$X \leq b$".

Let

$X = 0, 1, 2, 3$ (Previous example)

$2 < X \leq 3 \Rightarrow X = 3$

$X \leq 2 \Rightarrow X = 0, 1, 2$

mutually exclusive exhastive events.

its union is

$X \leq 3 \Rightarrow X = 0, 1, 2, 3$

$a < X \leq b$

$X \leq a$

$X \leq b$

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

$P[(a < X \leq b) \cup (X \leq a)] = P(X \leq b)$

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

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

$P(a < X \leq b) = F(b) - F(a)$ **Proved**

Discrete Random Variable and its Probability Distribution (Pmf, cdf, mean, Variance, S.D)

Random Variable → Discrete → (Pmf)

Random Variable → Continuous → (Pdf)

(X) Random Variable
P(X) Probability dis
E(X) Mean
Var(X) Variance

★ Discrete Random Variable
Probability Distribution of discrete r.v
Probability Mass Function (Pmf)

X = a discrete r.v

$(X = x_i)$	x_1	x_2	----	x_n
$P(X = x_i)$	P_1	P_2	.....	P_n

Let X be a discrete r.v taking on distinct values $x_1, x_2 \dots x_n, \dots$ Then the function, denoted by $p(x)$ or $f(x)$, and defined by

$$f(x_i) = P(X = x_i) \quad \text{for } i = 1, 2, \dots n, \dots$$
$$= 0 \quad \text{for } x \neq x_i$$

is called the probability function (Pf) or (Pmf).

Properties of Probability Distribution

i) $P_1 + P_2 + \dots P_n = 1$

$\sum_{i=1}^{n} P_i = 1$

ii) $0 \leq P_i \leq 1$

Example:

Toss two coins.

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

r.v $\leftarrow$ X = number of heads

$X(HH) = 2$

$X(HT) = 1$

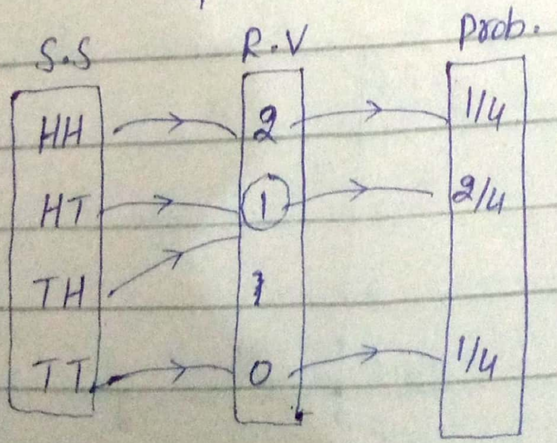
$X(TH) = 1$

$X(TT) = 0$

Its distribution of Probability is:

$(X = x_i)$	0	1	2
$P(X = x_i)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

} Discrete Prob. disⁿ.



Cumulative Distribution Function (cdf)

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

$F(X) = 0 \qquad x < 0$

$F(X) = \frac{1}{4} \qquad 0 \leq x < 1$

$F(X) = \frac{2}{4} + \frac{1}{4} = \frac{3}{4} \qquad 1 \leq x < 2$

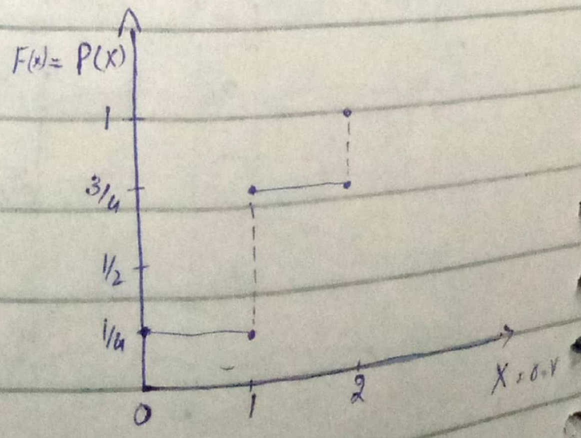
$F(X) = \frac{1}{4} + \frac{2}{4} + \frac{1}{4} = 1 \qquad 2 \leq x$ or $x \geq 2$

$$F_X(x) = \begin{cases} 0 & ; \quad x < 0 \\ 1/4 & ; \quad 0 \leq x < 1 \\ 3/4 & ; \quad 1 \leq x < 2 \\ 1 & ; \quad x \geq 2 \end{cases}$$

Graph of cdf

$F(x)$ in case of discrete r.v is a step function.

$F(-\infty) = 0 \quad ; \quad F(+\infty) = 1$



Mean of Discrete r.v

$E(X) = x_1P_1 + x_2P_2 + \cdots x_nP_n$

$E(X) = \sum_{i=1}^{n} x_iP_i$

Variance of Discrete r.v

$Var(X) = E(X^2) - \underbrace{[E(X)]^2}_{\text{Mean}}$

$E(X^2) = x_1^2P_1 + x_2^2P_2 + \cdots + x_n^2P_n$

$E(X^2) = \sum_{i=1}^{n} x_i^2P_i$

Q: Calculate Mean and Variance of Previous example of two coin toss????

Example-I

Find the probability distribution and distribution function for the number of heads when 3 coins are tossed. Construct a probability histogram and a graph of cdf.

* Solve it Yourself

Example-II:

a) Find the Probability distribution of the sum of the dots when two fair dice are thrown.

b) Use the probability distribution to find the probabilities of obtaining

i) a sum of 8 or 11

ii) a sum that is greater than 8

iii) a sum that is greater than 5 but less than or equal to 10.

★ **Solve it**

Probability Distribution of Continuous random variable (Pdf, cdf, mean, Variance, S.D)

★ Continuous Random Variable

A random variable "X" is defined to be continuous if it can assume every possible value in an interval $[a, b]$, $a < b$, where a and b may be $-\infty$ to ∞ respectively.

e.g. The height of a person.

The amount of rainfall.

The temperature at a place.

Relationship b/w Cdf and pdf

$$\frac{dF(x)}{dx} = f(x)$$

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

Why the probability at single point for a continuous random variable is zero?

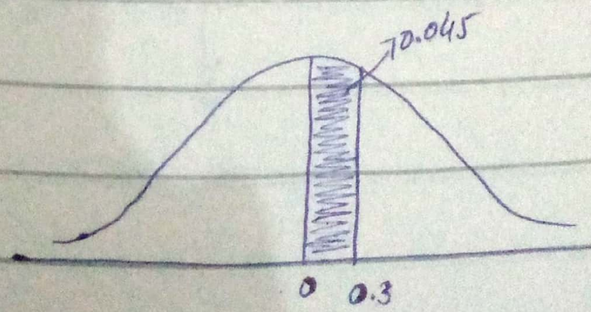
If we have a continuous random variable and we calculate its probabilities, it will provide area under the curve as we using integration to calculate probabilities of the continuous random variable.

For example :

Let $P(0 \le X \le 0.3) = \int_0^{0.3} x\,dx = \left.\frac{x^2}{2}\right|_0^{0.3}$

$$= \frac{(0.3)^2}{2} - \frac{(0)^2}{2}$$

$$= 0.045$$



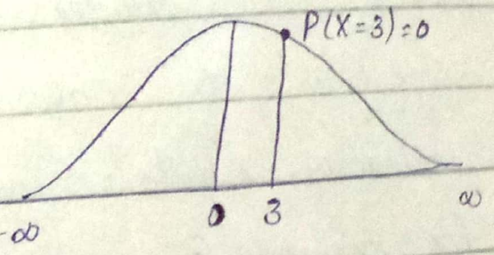
Here, the answer shows that the area between 0 and 0.3 is equal to 0.045.

Now,
if we calculate probability at single point, it must be zero.

i) One of its reason is that we can't calculate area at a single point, it just draw a single line in normal curve.

e.g

$P(X=3)=0$



$$P(X=3) = \frac{n(A)}{n(S)} = \frac{n(A)}{\infty} = 0$$

ii) One of the other reason is that mathematically we can't calculate prob. at single value by using integration for continuous r.v.

e.g

$$P(X=3) = \int_3^3 x\,dx = \left.\frac{x^2}{2}\right|_3^3 = \frac{9}{2} - \frac{9}{2} = 0$$

Scanned with CamScanner

* When we put a single point in Probability density function of any distribution and solve it, then its answer is not a probability at that point, its called value of function, height, ordinate or density.

Example:-

Gamma Distribution

$$f(y;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma\alpha} x^{\alpha-1} e^{-\beta x}$$

Let $\alpha = 1$

$\beta = 2$

$x = 3$

$$f(3) = \frac{(2)^1}{\Gamma 1} (3)^{1-1} e^{-2(3)}$$

$$= \frac{2}{1} (1)(0.002)$$

$$f(3) = 0.005$$

Here, 0.005 is not the probability at $x=3$, we can call it density, value of function, ordinate or height.

Mean and Variance of Continuous r.v

$$\star\ E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$\star\ Var(X) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - [E(X)]^2$$

Properties of Probability Density Function

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

ii) $f(x) \geq 0$, non-negative

iii) $P(X = c) = \int_{c}^{c} f(x)\, dx = 0$, where "c" is constant.

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

Example:-

(a) A continuous random variable X has a density function $f(x) = 2x$ where $0 \le x \le 1$ and zero otherwise. Find (i) $P(X < \frac{1}{2})$ (ii) $P(1/4 < X < 1/2)$

b) If $f(x)$ has probability density kx^2, $0 < x < 1$, determine k and find the probability that $1/3 < X < 1/2$.

Solution:-

a) $f(x) = 2x, \quad 0 \le x \le 1$

$\quad = 0, \quad$ otherwise

i) $P(X < \frac{1}{2}) = \int_0^{1/2} f(x)\,dx = \int_0^{1/2} 2x\,dx$

$$= 2\left[\frac{x^2}{2}\right]_0^{1/2}$$

$$= [x^2]_0^{1/2} = \left[(1/2)^2 - (0)^2\right] = \frac{1}{4}$$

ii) $P(\frac{1}{4} < X < \frac{1}{2}) = \int_{1/4}^{1/2} 2x\,dx = 2\left[\frac{x^2}{2}\right]_{1/4}^{1/2}$

$$= [x^2]_{1/4}^{1/2} = \left(\frac{1}{2}\right)^2 - \left(\frac{1}{4}\right)^2$$

$$= \left(\frac{1}{4} - \frac{1}{16}\right) = \frac{3}{16}$$

b) We know that

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

So,

$$1 = \int_0^1 f(x)\,dx = \int_0^1 kx^2\,dx = k\left[\frac{x^3}{3}\right]_0^1 = k\left(\frac{1}{3}-0\right) = \frac{k}{3}$$

$$1 = \frac{k}{3}$$

$$\boxed{k = 3}$$

$$P\left(\frac{1}{3} < X < \frac{1}{2}\right) = \int_{1/3}^{1/2} f(x)\,dx = \int_{1/3}^{1/2} 3x^2\,dx$$

$$= \cancel{3}\left[\frac{x^3}{\cancel{3}}\right]_{1/3}^{1/2}$$

$$= \left[\left(\frac{1}{2}\right)^3 - \left(\frac{1}{3}\right)^3\right]$$

$$= \left(\frac{1}{8} - \frac{1}{27}\right)$$

$$P\left(\frac{1}{3} < X < \frac{1}{2}\right) = \frac{19}{216}$$

Mean and Variance

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

$$= \int_0^1 x \cdot 2x\, dx = \int_0^1 2x^2 dx = 2\left[\frac{x^3}{3}\right]_0^1$$

$$= 2\left(\frac{1}{3} - \frac{0}{3}\right) = 2\left(\frac{1}{3}\right) = \frac{2}{3}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$= \int_0^1 x^2 \cdot 2x\, dx = \int_0^1 2x^3 dx = 2\left[\frac{x^4}{4}\right]_0^1$$

$$= \left(\frac{x^4}{2}\right)_0^1 = \left(\frac{1}{2} - \frac{0}{2}\right) = \frac{1}{2}$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= \frac{1}{2} - \left(\frac{2}{3}\right)^2$$

$$= \frac{1}{2} - \frac{4}{9}$$