Unit - 1

Random Variable and Probability Distributions

Consider two experiments.

E-1 Number of students getting 98% or above in an exam.

$S = \{0, 1, 2, 3, ---\}$

E-2 Testing 3 electronic Components as defective or non-defective.

$S = \{DDD, DDN, DND, DNN, NDD, NDN, NND, NNN\}$.

Thus, outcomes of an experiment may be numerical or non-numerical in nature. As it is often useful to describe the outcome of a random experiment by a number, we will assign a number to each non-numerical outcome of an experiment.

In E-2, we are interested in number of defective components. So, the numerical values assigned with each element of the sample space are 0, 1, 2 and 3.

We represent these numerical values by the random variable X.

Definition : A random variable is a function that associates a real number with each element in the sample space.

or A random variable $X: S \to \mathbb{R}$ is a function, whose domain is S and range is subset of $\mathbb{R}$.

* Each possible value of X represents an event.

Random Variable

Discrete Random Variable — Continuous Random Variable

Discrete Random Variable

If X is a RV which can take a finite number or countably infinite number of values, X is called a discrete RV. The possible values of X may be assumed as $x_1, x_2, ---, x_n, ----$.

associated with natural numbers

Ex-: 1. Number of students in a class.
2. Number of defective items.
3. Number of accidents in June.
4. Number of heads, when a coin is tossed.
5. Number of items observed before a defective item is observed.

$S = \{D, NND, NNND, ---\}$

$X = \{1, 2, 3, ---\}$

Discrete Probability Distributions

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

Let X represents the number of heads/tails in the toss of two coins.

X takes values 0, 1, 2.

The probability distribution of X is given by

X	0	1	2
$f(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

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

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

$f(2) = P(X=2) = \cancel{\frac{2}{2}}\ \frac{1}{4}$

In E-2, The probability distribution is

X	0	1	2	3
$f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Note :- In some books, $P_X(x)$ or $P(X=x)$ is used for $f(x)$.

Probability Function | Probability Mass Function | Probability Distribution

Definition :- The set of ordered pairs $(x, f(x))$ is a probability mass function or probability function or probability distribution of the discrete random variable X, if for each possible outcome x,

1. $f(x) \geq 0$,
2. $\sum_x f(x) = 1$,
3. $P(X=x) = f(x)$.

Mathematical Expectation | Expected Value | Mean of DRV

Let X be a random variable with pmf $f(x)$. The mean or expected value of X is

$$\mu = E(X) = \sum_x x f(x).$$

Variance of Discrete RV

Let X be a discrete random variable with probability distribution function $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$

or $\sigma^2 = E(X^2) - \mu^2$.

Note : The positive square root of variance is called standard deviation (σ)

Ex : A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution, mean and variance of defectives.

Sol : Let X be the random variable representing the possible number of defective computers purchased by the school.

$\therefore$ X can take values 0, 1, 2.

The probability distribution of X is

X	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

$$f(0) = P(X=0) = \frac{{}^{17}C_2}{{}^{20}C_2} = \frac{68}{95}$$

$$f(1) = P(X=1) = \frac{{}^{17}C_1 \times {}^{3}C_1}{{}^{20}C_2} = \frac{51}{190}$$

$$f(2) = P(X=2) = \frac{{}^{3}C_2}{{}^{20}C_2} = \frac{3}{190}$$

Now, $\mu = E(X) = \sum_{x=0}^{2} x f(x)$

$$= 0 + \frac{51}{190} + \frac{6}{190} = \frac{57}{190} = 0.3$$

$$\boxed{\mu = 0.3}$$

and $\sigma_X^2 = E((X-\mu)^2) = \sum_{x=0}^{2} (x-\mu)^2 f(x)$

$$= (0.09)\frac{68}{95} + (0.49)\frac{51}{190} + (2.89)\frac{3}{190}$$

$$= 0.064 + 0.131 + 0.046$$

$$= 0.241$$

$$\boxed{\sigma_X^2 = 0.241}$$

or $\sigma_X^2 = E(X^2) - \mu^2 = \frac{51}{190} + \frac{12}{190} - (0.3)^2$

$$= 0.331 - 0.09$$

$$= 0.241$$

$$\boxed{\sigma^2 = 0.241}$$

Ex: A lot containing 7 components is sampled by a quality inspector, the ~~set~~ lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value and variance of the number of good components in this sample.

Let X represent the number of good components.

X takes values 0, 1, 2, 3.

The pmf is given by

X	0	1	2	3
$f(x)$	$\frac{1}{35}$	$\frac{12}{35}$	$\frac{18}{35}$	$\frac{4}{35}$

$$f(0) = P(X=0) = \frac{4C_0 \; 3C_3}{7C_3} = \frac{1}{35}$$

$$f(1) = P(X=1) = \frac{4C_1 \; 3C_2}{7C_3} = \frac{12}{35}$$

$$f(2) = P(X=2) = \frac{4C_2 \; 3C_1}{7C_3} = \frac{18}{35}$$

$$f(3) = P(X=3) = \frac{4C_3 \; 3C_0}{7C_3} = \frac{4}{35}$$

$$\therefore \mu = E(X) = \frac{12}{35} + \frac{36}{35} + \frac{12}{35} = \frac{60}{35} = 1.71 \Rightarrow \boxed{\mu = 1.71}$$

$$\sigma^2 = E(X^2) - \mu^2 = \frac{12}{35} + \frac{72}{35} + \frac{36}{35} - \left(\frac{60}{35}\right)^2$$

$$= 3.428 - 2.924$$

$$\boxed{\sigma^2 = 0.504}$$

Let X be a random variable with probability distribution function $f(x)$. The expected value of the random variable $g(X)$ is

$$\mu_{g(X)} = E(g(X)) = \sum_x g(x) f(x).$$

Thm Let X be a random variable with probability distribution $f(x)$. The variance of random variable $g(X)$ is

$$\sigma^2_{g(X)} = E\{(g(X)-\mu_{g(X)})^2\} = \sum_x [g(x)-\mu_{g(X)}]^2 f(x).$$

Ex Calculate the variance of $g(X) = 2X+3$, where X is a random variable with probability distribution

X	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

Sol:

$$\mu_{g(X)} = E(2X+3) = 2E(X)+3 = 2\left[\frac{1}{8}+1+\frac{3}{8}\right]+3$$

$$= 2\left[\frac{12}{8}\right]+3$$

$$= 6$$

$$\boxed{\mu_{g(X)} = 6}$$

$$= 4\left[\frac{26}{8}-\frac{9}{4}\right] = 4\left[\frac{13}{4}-\frac{9}{4}\right]$$

$$= 4\left[\frac{4}{4}\right]$$

$$\sigma^2_{2X+3} = E\{[(2X+3)-6]^2\} = E[(2X-3)^2]$$

$$= E[4X^2+9-12X] \quad \boxed{\sigma^2 = 4}$$

$$= 4E(X^2)+E(9)-12E(X)$$

$$= 4E(X^2)+9-12E(X)$$

$$4\sigma_X^2 = 4\left[E(X^2) - \left(\frac{1}{8}+1+\frac{3}{8}\right)^2\right]$$

$$= 4\left[\frac{1}{8}+2+\frac{9}{8}-\frac{9}{4}\right]$$

$$E(X) = \frac{12}{8} = \frac{3}{2}$$

$$E(X^2) = 0 + \frac{1}{8} + \frac{4}{2} + \frac{9}{8} = \frac{1}{8} + 2 + \frac{9}{8} = \frac{26}{8} = \frac{13}{4}$$

$$\therefore \sigma^2_{g(x)} = 4\left(\frac{13}{4}\right) + 9 - 12\left(\frac{3}{2}\right) = 13 + 9 - 18 = 4$$

$$\boxed{\sigma^2_{g(x)} = 4}$$

Properties

1. $E(aX+b) = aE(X)+b$, where a and b are constants.

   If $a=0$, $E(b)=b$

   If $b=0$, $E(aX) = aE(X)$.

2. $E[g(X) \pm h(X)] = E[g(X)] \pm E(h(X))$

   If $Y = (X-1)^2 \Rightarrow E(Y) = E(X^2) - 2E(X) + 1$.

3. $\sigma^2_{aX+c} = a^2\sigma_X^2 = a^2\sigma^2$

4. $\sigma^2_{aX} = a^2\sigma_X^2$

5. $\sigma^2_{X+c} = \sigma_X^2$

6. $\sigma^2_c = 0$

Cumulative Distribution Function

Cumulative → Increasing by successive addition

The cdf $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

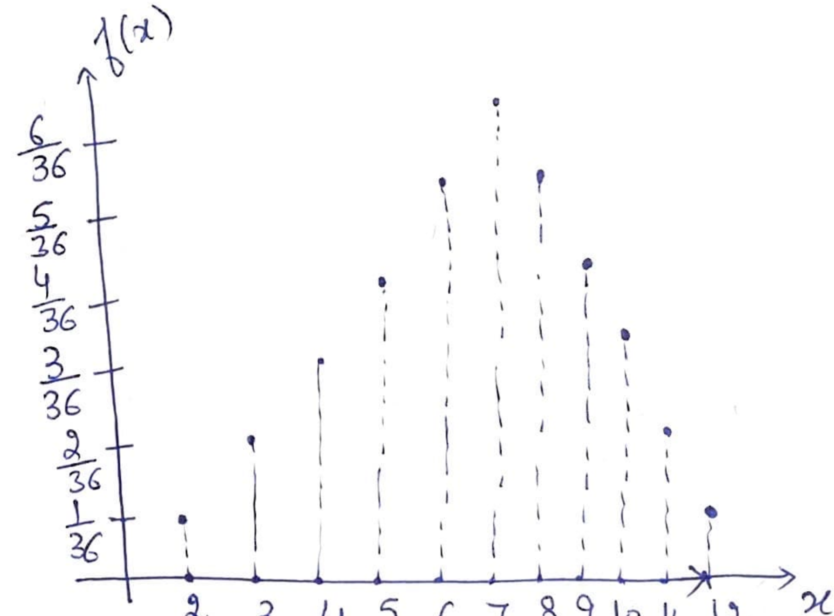
$$F(x) = P(X \le x) = \sum_{t \le x} f(t) \text{ for } -\infty < x < \infty.$$

$$\boxed{0 \le F(x) \le 1}$$

Ex Two dice are rolled. Let X denote the random variable which counts the total number of points on the upturned faces, Construct a table giving the non-zero values of the probability mass function and also draw the probability chart and find distribution function of X.

Sol:

X	$f(x)$
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$



p.m.f. plot

Properties: 1. If F is the cdf of X and $a < b$, then $P(a < X \le b) = F(b) - F(a)$.

2. $0 \le F(x) \le 1$

Distribution function

it

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

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

$F(2) = P(X \le 2) = \frac{1}{36}$

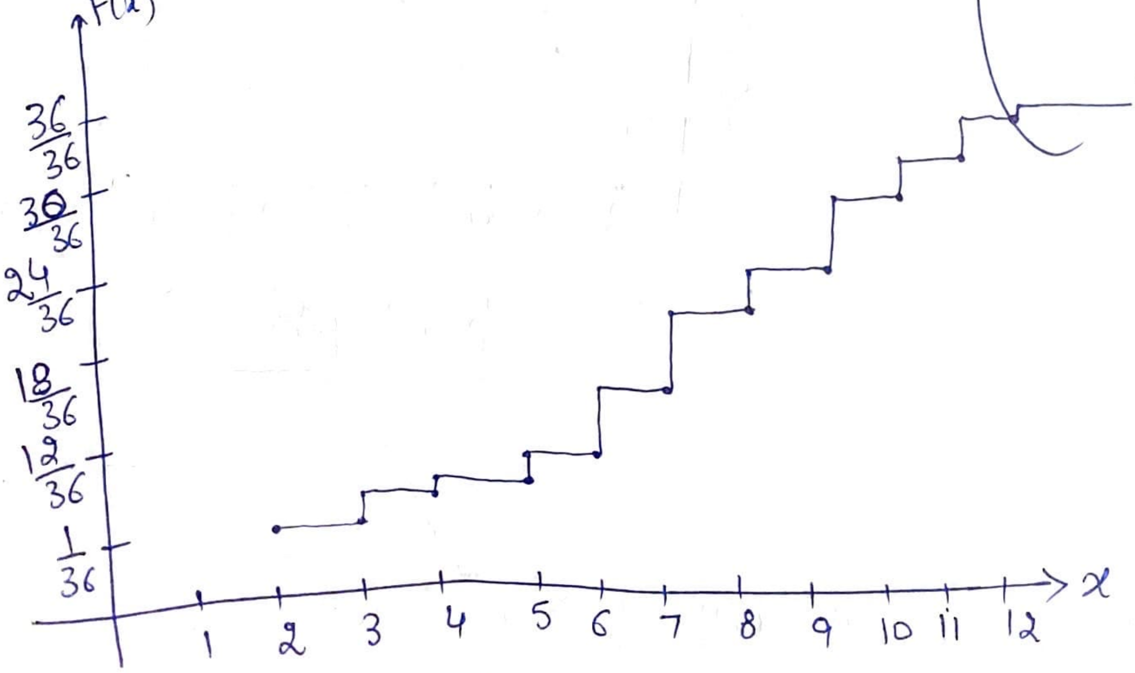
$F(3) = P(X \le 3) = \frac{3}{36}$

$F(4) = \frac{6}{36}$, $F(5) = \frac{10}{36}$, $F(6) = \frac{15}{36}$,

$F(7) = \frac{21}{36}$, $F(8) = \frac{26}{36}$, $F(9) = \frac{30}{36}$,

$F(10) = \frac{33}{36}$, $F(11) = \frac{35}{36}$, $F(12) = 1.$

$$F(x) = \begin{cases} 0, & \text{for } x < 2 \\ \frac{1}{36}, & \text{for } 2 \le x < 3 \\ \frac{3}{36}, & \text{for } 3 \le x < 4 \\ \frac{6}{36}, & \text{for } 4 \le x < 5 \\ \frac{10}{36}, & \text{for } 5 \le x < 6 \\ \frac{15}{36}, & \text{for } 6 \le x < 7 \\ \frac{21}{36}, & \text{for } 7 \le x < 8, \\ \frac{26}{36}, & \text{for } 8 \le x < 9 \\ \frac{30}{36}, & \text{for } 9 \le x < 10 \\ \frac{33}{36}, & \text{for } 10 \le x < 11 \\ \frac{35}{36}, & \text{for } 11 \le x < 12 \\ 1, & \text{for } x \ge 12. \end{cases}$$



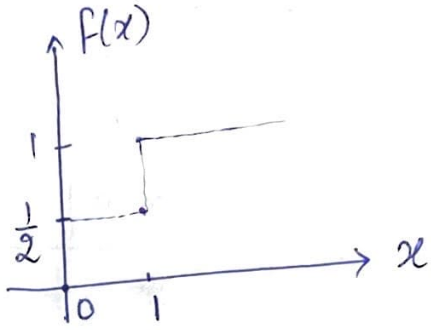
discrete c.d.f.

Let X be the RV representing the number of heads in a single toss of a fair coin. Find cdf.

Sol:

$$F(x) = \begin{cases} 0 & ; x < 0 \\ \frac{1}{2} & ; 0 \le x < 1 \\ 1 & ; x \ge 1 \end{cases}$$

X	0	1
$f(x)$	1/2	1/2
$F(x)$	1/2	1



Q: Find c.d.f. if pmf is

x	0	1	2	3	4
$f(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Also plot ~~pdf~~ pmf and draw probability histogram.

Sol:

$F(0) = P(X \le 0) = P(X=0) = f(0) = \frac{1}{16}$

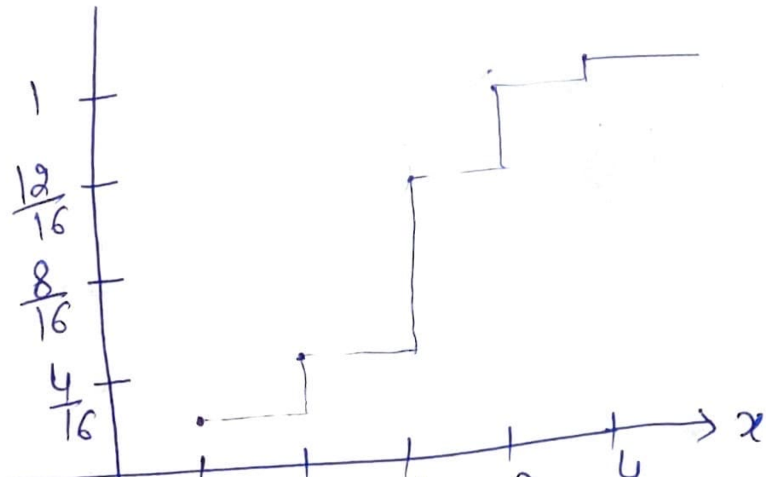
$F(1) = P(X \le 1) = P(X=0) + P(X=1) = f(0) + f(1) = \frac{5}{16}$

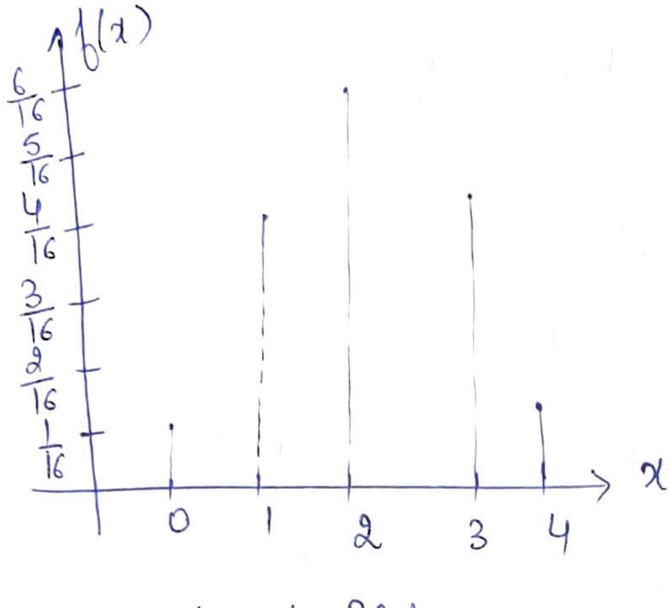
$F(2) = P(X \le 2) = P(X=0) + P(X=1) + P(X=2) = \frac{11}{16}$

$F(3) = P(X \le 3) = \frac{15}{16}$

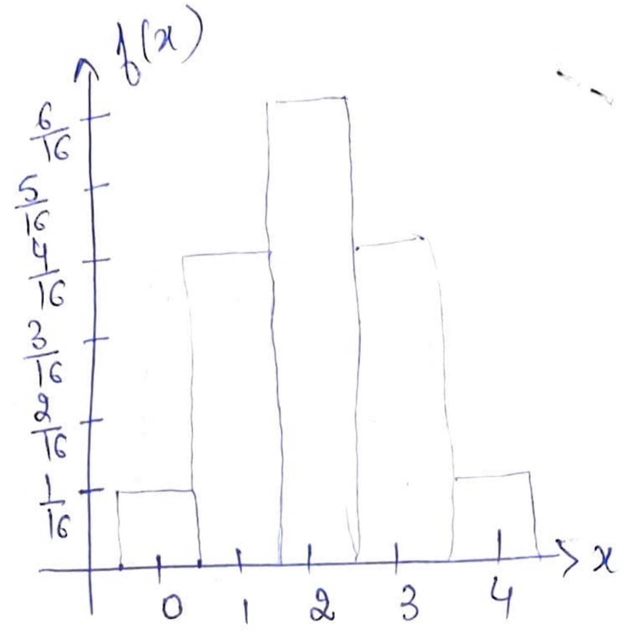
$F(4) = 1.$

$$F(x) = \begin{cases} 0 & ; x < 0 \\ \frac{1}{16} & ; 0 \le x < 1 \\ \frac{5}{16} & ; 1 \le x < 2 \\ \frac{11}{16} & ; 2 \le x < 3 \\ \frac{15}{16} & ; 3 \le x < 4 \\ & ; x \ge 4 \end{cases}$$





p.m.f. Plot



Probability histogram

Continuous Random Variable

A RV X is said to be continuous random variable if it takes all the possible values in a given interval.

Ex: (i) Weight of a group of individuals.
(ii) Height of a group of individuals.
(iii) Length of time for a chemical reaction to take place
(iv) Temparature in a given day.
(v) Age of a person.

Continuous Probability Distributions | Probability Density Function | Probability Function

When, we have a discrete random variable, we cant list all the values and their probabilities, even if the list is infinite.

But, when we have a continuous set, e.g, $[0,1]$, how to list the numbers in $[0,1]$?

- The smallest number is 0, but what is the next smallest?

$$0.01,\ 0.0001,\ 0.0000001, \ldots$$

* In fact, there are so many numbers in any continuous set that each of them must have probability 0.

When X is continuous RV,
$$P(X=x)=0 \text{ for all } x.$$

A continuous random variable takes values in a continuous interval.

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

i.e, it does not matter whether we conclude an endpoint the interval or not.

Def: The function $f(x)$ is a probability density function (pdf) for the continuous random variable X, defined over the set of real numbers if

1. $f(x) \geq 0$ for all $x \in \mathbb{R}$,
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$,
3. $P(a < X < b) = \int_{a}^{b} f(x)\,dx$.

Cumulative distribution function

The c.d.f. $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt, \quad \text{for } -\infty < x < \infty.$$

Remark: $P(a<X<b) = F(b) - F(a)$ and $f(x) = \dfrac{dF(x)}{dx}$.

Mean and Variance of CRV

Let X be a CRV with pdf $f(x)$. The mean and variance of X is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx \quad \text{and} \quad \sigma^2 = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx.$$

$$\text{or } \sigma^2 = E(X^2) - \mu^2.$$

Suppose that the error in the reaction temperature for a laboratory experiment is a continuous random variable X having pdf $f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

(c) Find $F(x)$ and use it to evaluate $P(0 < X \leq 1)$

(d) Find mean and variance of X.

(e) Find the variance of $g(X) = 4X + 3$.

Sol:- (a) Obviously, $f(x) \geq 0$

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{-1} f(x)\,dx + \int_{-1}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx$$

$$= \int_{-1}^{2} \frac{x^2}{3}\,dx = \frac{1}{3}\left[\frac{x^3}{3}\right]_{-1}^{2} = \frac{1}{9}\left[8+1\right] = 1.$$

$\therefore$ $f(x)$ is a density function.

(b) $$P(0 < X \leq 1) = \int_{0}^{1} f(x)\,dx = \int_{0}^{1} \frac{x^2}{3}\,dx = \frac{1}{3}\left[\frac{x^3}{3}\right]_{0}^{1} = \frac{1}{9}.$$

(c) $$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-1} f(t)\,dt + \int_{-1}^{x} f(t)\,dt$$

$$= \int_{-1}^{x} \frac{t^2}{3}\,dt = \frac{1}{3}\left[\frac{t^3}{3}\right]_{-1}^{x}$$

$$= \frac{1}{9}\left[x^3+1\right], \text{ for } -1 \leq x < 2.$$

$$\therefore F(x) = \begin{cases} 0, & x < -1 \\ \dfrac{x^3+1}{9}, & -1 \le x < 2 \\ 1, & x \ge 2 \end{cases}$$

For $x \ge 2$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-1} f(t)\,dt + \int_{-1}^{2} f(t)\,dt + \int_{2}^{x} f(t)\,dt$

$$= 0 + \int_{-1}^{2} \frac{t^2}{3}\,dt + 0$$

$$= \frac{1}{3}\left[\frac{t^3}{3}\right]_{-1}^{2} = \frac{1}{9}[8+1] = 1.$$

$P(0 < X \le 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}.$

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

$$= \int_{-\infty}^{-1} x f(x)\,dx + \int_{-1}^{2} x f(x)\,dx + \int_{2}^{\infty} x f(x)\,dx$$

$$= 0 + \int_{-1}^{2} x \cdot \frac{x^2}{3}\,dx + 0$$

$$= \frac{1}{3}\left[\frac{x^4}{4}\right]_{-1}^{2} = \frac{1}{12}[16-1] = \frac{15}{12}$$

$$\boxed{\mu = \frac{15}{12} = \frac{5}{4}}$$

$$\sigma^2 = E\left[\left(X-\frac{15}{12}\right)^2\right] = E(X^2) - \mu^2$$

$$= \int_{-1}^{2} \frac{x^4}{3}\,dx - \left(\frac{15}{12}\right)^2$$

$$= \frac{1}{3}\left[\frac{x^5}{5}\right]_{-1}^{2} - \left(\frac{5}{4}\right)^2$$

$$= \frac{1}{5\cdot 3}\left[32+1\right] - \frac{25}{16}$$

$$= \frac{33}{15} - \frac{25}{16} = \frac{11}{5} - \frac{25}{16} = \frac{176-125}{80} = \frac{51}{80} =$$

$$\boxed{\sigma^2 = 0.6375}$$

Ex Let X be a continuous random variable with pdf given by

$$f(x) = \begin{cases} kx, & 0 \le x < 1 \\ k, & 1 \le x < 2 \\ -kx+3k, & 2 \le x < 3 \\ 0, & \text{elsewhere} \end{cases}$$

(i) Determine the constant k.

(ii) Determine F(x).

Sol: (i) Since $f(x)$ is pdf of X,

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

$$\Rightarrow \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} kx\,dx + \int_{1}^{2} k\,dx + \int_{2}^{3} (-kx+3k)\,dx + \int_{3}^{\infty} 0\,dx = 1$$

$$\Rightarrow k\left[\frac{x^2}{2}\right]_0^1 + k\left[x\right]_1^2 + \left[-\frac{kx^2}{2} + 3kx\right]_2^3 = 1$$

$$\Rightarrow \frac{k}{2} + k - \frac{k}{2}(9-4) + 3k = 1$$

$$\Rightarrow 4k + \frac{k}{2} - \frac{5k}{2} = 1$$

$$\Rightarrow 4k - 2k = 1 \Rightarrow \boxed{k = \frac{1}{2}}$$

(ii) For $-\infty < x < 0$, $F(x) = 0$

For, $0 \le x < 1$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{0} f(t)\,dt + \int_{0}^{x} f(t)\,dt$

$$= \int_{0}^{x} \frac{1}{2}t\,dt = \frac{1}{2}\frac{x^2}{2} = \frac{x^2}{4}$$

For $1 \le x < 2$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_{0}^{1} \frac{t}{2}\,dt + \int_{1}^{x} \frac{1}{2}\,dt$

$$= \frac{1}{2}\frac{t^2}{2}\Big]_0^1 + \frac{1}{2}(x-1)$$

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

For $2 \le x < 3$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_{0}^{1} \frac{t}{2}\,dt + \int_{1}^{2} \frac{1}{2}\,dt +$

$$\int_{2}^{x}\left(-\frac{t}{2} + \frac{3}{2}\right)dt$$

$$= \frac{t^2}{4}\Big|_0^1 + \frac{t}{2}\Big|_1^2 + \left(-\frac{t^2}{4}\right)_2^x + \left(\frac{3}{2}t\right)_2^x$$

$$= \frac{1}{4} + \frac{1}{2} - \frac{1}{4}(x^2-4) + \frac{3}{2}(x-2)$$

$$= \frac{1}{4} + \frac{1}{2} - \frac{x^2}{4} + 1 + \frac{3}{2}x - 3$$

$$= \frac{-x^2+6x-5}{4}$$

For $x \geq 3$, $F(x) = \int_{-\infty}^{0} 0\,dt + \int_0^1 \frac{1}{2}t\,dt + \int_1^2 \frac{1}{2}\,dt + \int_2^3 \left(-\frac{1}{2}t + \frac{3}{2}\right)dt$

$$+ \int_3^{\infty} 0\,dt$$

$$\cancel{= \frac{1}{2}\frac{t^2}{2}\Big|_0^1 + \frac{1}{2}t\Big|_1^2 - \frac{3}{2}\frac{t^2}{2}\Big|_2^3 + \frac{3}{2}t\Big|_2^3}$$

$$\cancel{= \frac{1}{4} + \frac{1}{2} - \frac{3}{4}(9-4) + \frac{3}{2}}$$

$$\cancel{= \frac{1}{4} + \frac{1}{2} - \frac{15}{4} + \frac{3}{2}}$$

$$\cancel{= -3 +}$$

$\cancel{F(x) =}$

$$\boxed{\begin{aligned} F(x) &= \frac{1}{2}\frac{t^2}{2}\Big|_0^1 + \frac{1}{2}t\Big|_1^2 - \frac{1}{2}\frac{t^2}{2}\Big|_2^3 + \frac{3}{2}t\Big|_2^3 \\ &= \frac{1}{4} + \frac{1}{2} - \frac{1}{4}(5) + \frac{3}{2} \\ &= 1 \end{aligned}}$$

$$F(x) = \begin{cases} 0 & ; x < 0 \\ \frac{x^2}{4} & ; 0 \le x < 1 \\ \frac{2x-1}{4} & ; 1 \le x < 2 \\ \frac{-x^2+6x-5}{4} & ; 2 \le x < 3 \\ 1 & ; x \ge 3 \end{cases}$$

(e) $g(X) = 4X+3.$

$$\mu_{g(X)} = E(4X+3) = \int_{-\infty}^{\infty} (4x+3) \cancel{\frac{x^2}{3}} f(x)\,dx$$

$$= \int_{-1}^{2} (4x+3) \frac{x^2}{3}\,dx = \frac{4}{3}\int_{-1}^{2} x^3\,dx + \int_{-1}^{2} x^2\,dx$$

$$= \frac{4}{3}\left[\frac{x^4}{4}\right]_{-1}^{2} + \left[\frac{x^3}{3}\right]_{-1}^{2}$$

$$= \frac{1}{3}[16-1] + \frac{1}{3}[8+1]$$

$$= \frac{15+9}{3} = 8$$

$$\boxed{\mu_{g(X)} = E[4X+3] = 8}$$

$$\sigma^2_{g(X)} = E[(4X+3-8)^2]$$

$$= E[(4X-5)^2]$$

$$= \int_{-1}^{2} (4x-5)^2 \frac{x^2}{3} dx$$

$$= \int_{-1}^{2} (16x^2+25-40x) \frac{x^2}{3} dx$$

$$= \frac{1}{3}\left[\frac{16x^5}{5} + \frac{25x^3}{3} - \frac{40x^4}{4}\right]_{-1}^{2}$$

$$= \frac{1}{3}\left[\frac{16}{5}(33) + \frac{25}{3}(9) - 10(15)\right]$$

$$= \frac{1}{3}\left[\frac{528}{5} + \frac{225}{3} - 150\right]$$

$$= \frac{1}{3}\left[\frac{528}{5} + 75 - 150\right]$$

$$= \frac{1}{3}\left[\frac{528-375}{5}\right] = \frac{\cancel{150}\,153}{3 \times \cancel{3}5} = \frac{51}{5}$$

$$\boxed{\sigma^2_{g(X)} = \frac{51}{5}}$$

Mean and Variance of $g(X)$

Let X be CRV with pdf $f(x)$. The mean and variance of $g(X)$ is

$$\mu_{g(X)} = E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx \quad \text{and}$$

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(X) - \mu_{g(X)}]^2 f(x) dx.$$

Joint Probability Distributions

Two dimensional Random Variables

Let X and Y be two RV defined on the same sample space S, then the function (X, Y) that assigns a point in $\mathbb{R}^2$, is called a two dimensional RV.

Ex: (i) (X, Y), where X is associated with height of persons and Y is associated with weight of persons in an educational institute.

(ii) (X, Y), where X denotes the number of heads obtained in the first toss and Y denote the number of heads in second toss.

(X, Y) takes values $\quad S = \{HH, HT, TH, TT\}$

$\{(1,1), (1,0), (0,1), (0,0)\}$.

(iii) (X, Y); X → Gender, Y → Sports.

(iv) (X, Y); X → Boy child in a family ~~of~~ having 2 children
Y → Girl child

(v) (X, Y); X → Room temparature, Y → Outside temp.

(vi) (X, Y); X → Rainfall in Jan, Y → Rainfall in July.

Joint Probability Distributions

Joint p.m.f.

Let (X,Y) be a DRV, then $f(x,y)$ is joint p.m.f. of (X,Y) if

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

(ii) $\sum_x \sum_y f(x,y) = 1$,

(iii) $P(X=x, Y=y) = f(x,y)$.

Joint p.d.f.

Let (X,Y) be a CRV, then $f(x,y)$ is joint density function if

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

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

(iii) $P[(X,Y) \in A] = \iint_A f(x,y)\,dx\,dy$,

for any region in the xy plane

Mean and Variance of $g(X,Y)$

Let X and Y be RV with joint probability distribution $f(x,y)$. The mean or expected value of the RV $g(X,Y)$ is

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y) f(x,y) \text{ if } (X,Y) \text{ is discrete RV.}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\,dx\,dy \text{ if } (X,Y) \text{ is CRV.}$$

Covariance

Covariance refers to the measure of how two RV will change when they are compared to each other.

The Covariance between two RV is a measure of the nature of the association between the two.

The sign of the Covariance indicates whether the relationship between two dependent RV is +ve or -ve.

If X and Y are independent, then the covariance is zero.

Covariance

Let (X,Y) be a RV with joint prob function $f(x,y)$. The Covariance of (X,Y) is

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)]$$

or $\boxed{\sigma_{XY} = E(XY) - \mu_X\mu_Y}$

$$= \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x,y) & \text{if } (X,Y) \text{ is DRV} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f(x,y) & \text{if } (X,Y) \text{ is CRV} \end{cases}$$

Q:- Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find (a) Joint probability distribution $f(x,y)$.

(b) $P[(X,Y) \in A]$, where A is the region $\{(x,y) \mid x+y \le 1\}$.

(c) Find the expected value of $g(X,Y) = XY$ and Covariance of (X,Y).

Sol: The possible values of (x,y) are $(0,0), (0,1), (1,0), (0,2), (2,0)$.

$f(x,y)$		$x=0$	$x=1$	$x=2$	Row total
y	0	3/28	9/28	3/28	15/28
	1	3/14	3/14	0	3/7
	2	1/28	0	0	1/28
Col Total		10/28	$\frac{15}{28}$	$\frac{3}{28}$	1

$$f(0,0) = \frac{3C_2}{8C_2} = \frac{3}{28}$$

$$f(0,1) = \frac{3C_1 \times 2C_1}{8C_2} = \frac{3}{14}$$

$$f(0,2) = \frac{2C_2}{8C_2} = \frac{1}{28}$$

$$f(1,0) = \frac{3C_1 \times 3C_1}{8C_2} = \frac{9}{28}$$

$$f(1,1) = \frac{3C_1 \times 2C_1}{8C_2} = \frac{3}{14}$$

$$f(2,0) = \frac{3C_2}{8C_2} = \frac{3}{28}$$

$P((X,Y) \in A) = P(X+Y \le 1)$

$= P(X+Y \le 1)$

$= f(0,0) + f(0,1) + f(1,0)$

$= \frac{3}{28} + \frac{3}{14} + \frac{9}{28}$

$= \frac{18}{28} = \frac{9}{14}.$

(C) ~~E(X,Y)~~ $E(g(X,Y)) = E(XY)$

$$= \sum_{y=0}^{2} \sum_{x=0}^{2} xy\, f(x,y).$$

$$= \sum_{y=0}^{2} 0 + y f(1,y) + 2y f(2,y)$$

$= 0 + f(1,1) + 2f(2,1) + 2f(1,2)$

$+ 4f(2,2)$

$= \frac{3}{14} + 2(0) + 0 + 4(0)$

$= \frac{3}{14}.$

$\sigma_{XY} = E(XY) - \mu_X \mu_Y$

$$\mu_X = \sum_{x=0}^{2} x g(x)$$

x	0	1	2
$g(x)$	$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$

$= \frac{15}{28} + \frac{2(3)}{28} + 0 = \frac{21}{28} = \frac{3}{4}$

y	0	1	2
$h(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

$\boxed{\mu_X = \frac{3}{4}}$

$$\mu_Y = \sum_{y=0}^{2} y\, h(y) = \frac{3}{7} + \frac{2}{28} = \frac{3}{7} + \frac{1}{14} = \frac{7}{14} = \frac{1}{2}$$

$$\boxed{\mu_Y = \frac{1}{2}}$$

$$\sigma_{XY} = \frac{3}{14} - \frac{3}{4}\cdot\frac{1}{2} = \frac{12-21}{56} = \frac{-9}{56} \qquad \boxed{\sigma_{XY} = \frac{-9}{56}}$$

Or

$$\sigma_{XY} = \sum_{x=0}^{2}\sum_{y=0}^{2}\left(x-\frac{3}{4}\right)\left(y-\frac{1}{2}\right) f(x,y)$$

$$= \sum_{x=0}^{2}\Big[\left(x-\frac{3}{4}\right)\left(\frac{-1}{2}\right) f(x,0) + \frac{1}{2}\left(x-\frac{3}{4}\right) f(x,y)$$

$$+ \left(\frac{3}{2}\right)\left(x-\frac{3}{4}\right) f(x,2)\Big]$$

$$= \frac{3}{8} f(0,0) + \left(\frac{-3}{8}\right) f(0,1) + \frac{3}{2}\left(\frac{-3}{4}\right) f(0,2)$$

$$+ \left(\frac{1}{4}\right)\left(\frac{-1}{2}\right) f(1,0) + \frac{1}{2}\left(\frac{1}{4}\right) f(1,1) + \frac{3}{2}\left(\frac{1}{4}\right) f(1,2)$$

$$+ \frac{5}{4}\left(\frac{-1}{2}\right) f(2,0) + \frac{1}{2}\left(\frac{5}{4}\right) f(2,1) + \frac{3}{2}\left(\frac{5}{4}\right) f(2,2)$$

$$= \frac{3}{8}\cdot\frac{3}{28} - \frac{3}{8}\cdot\frac{3}{14} - \frac{9}{8}\frac{1}{28} - \frac{1}{8}\frac{9}{28} + \frac{1}{8}\cancel{\frac{9}{28}}\frac{3}{14}$$

$$+ \frac{3}{8}(0) \cancel{+ \frac{5}{16}} - \frac{5}{8}\frac{3}{28}$$

$$= -\frac{9}{56}$$

A privately owned business operates both a drive in facility and walk in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive in and the walk in facilities are in use and suppose that the joint density function of (X,Y) is

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & 0<x<2,\ 0<y<1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify $f(x,y)$ is pdf.

(b) Find $E\left(\frac{Y}{X}\right)$ and Covariance of (X,Y).

(c) Find $P[(X,Y) \in A]$, where $A = \{(x,y) \mid 0<x<\frac{1}{2}, \frac{1}{4}<y<\frac{1}{2}\}$.

Sol: (a) $f(x,y) \geq 0$.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_{x=0}^{2}\int_{y=0}^{1} x\left(\frac{1+3y^2}{4}\right) dx\,dy$$

$$= \frac{1}{4}\int_{x=0}^{2} x\left(y + \frac{3y^3}{3}\right)_0^1 dx$$

$$= \frac{1}{4}\int_0^2 x(1+1)-0\,dx$$

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

(b) $g(X,Y) = \frac{Y}{X}$

$$E\left[\frac{Y}{X}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{y}{x}\right) f(x,y)\,dx\,dy$$

$$= \int_{x=0}^{2}\int_{y=0}^{1} \frac{y}{x}\, x\left(\frac{1+3y^2}{4}\right) dx\,dy$$

$$= \frac{1}{4}\int_{x=0}^{2}\int_{0}^{1} (y+3y^3)\,dy\,dx$$

$$= \frac{1}{4}\int_{0}^{2} \left[\frac{y^2}{2} + \frac{3y^4}{4}\right]_0^1 dx$$

$$= \frac{1}{4}\int_{0}^{2} \left(\frac{1}{2}+\frac{3}{4}\right) dx = \frac{1}{4}\int_0^2 \frac{5}{4}\,dx = \frac{5}{16}\left[2-0\right] = \frac{5}{8}.$$

$$E\left[Y/X\right] = \frac{5}{8}.$$

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y.$$

$$\mu_X = \int_{-\infty}^{\infty} x\, g(x)\,dx$$

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

$$= \frac{1}{6}(8) = \frac{4}{3}.$$

for $0<x<2$,

$$g(x) = \int_0^1 x\left(\frac{1+3y^2}{4}\right) dy$$

$$= \frac{x}{4}\left[y + \frac{3y^3}{3}\right]_0^1$$

$$= \frac{x}{4}\left[1+1\right] = \frac{x}{2}$$

$$g(x) = \begin{cases} \frac{x}{2}, & 0<x<2 \\ 0, & \text{elsewhere} \end{cases}$$

$$(y = \int_{-\infty}^{\infty} y\, h(y)\, dy$$

$$= \int_{0}^{1} y\left(\frac{1+3y^2}{2}\right) dy$$

$$= \frac{1}{2}\int_{0}^{1} (y+3y^3)\, dy$$

$$= \frac{1}{2}\left[\frac{1}{2}+\frac{3}{4}\right] = \frac{1}{2}\left(\frac{5}{4}\right)$$

$$= \frac{5}{8}.$$

for $0<y<1$,

$$h(y) = \int_{0}^{2} \frac{x(1+3y^2)}{4}\, dx$$

$$= \frac{1+3y^2}{4}\left[\frac{x^2}{2}\right]_0^2$$

$$= \frac{1+3y^2}{2}, \quad 0<y<1$$

$$h(y) = \begin{cases} \dfrac{1+3y^2}{2}, & 0<y<1 \\ 0, & \text{elsewhere} \end{cases}$$

$$\sigma_{XY} = \int_{x=0}^{2}\int_{y=0}^{1} \frac{x(1+3y^2)}{4}\, xy\, dx\, dy - \frac{4}{3}\cdot\frac{5}{8}$$

$$= \int_{0}^{2} \frac{x^2}{4}\, dx \int_{0}^{1} (y+3y^3)\, dy - \frac{\cancel{20}}{\cancel{24}}\,\frac{5}{6}$$

$$= \frac{1}{4}[8]\left[\frac{1}{2}+\frac{3}{4}\right] - \frac{5}{6}$$

$$= 2\left(\frac{5}{4}\right) - \frac{5}{6} = \frac{5}{2} - \frac{5}{6} = \frac{15-5}{6} = \frac{10}{6} = \frac{5}{3}.$$

$$\boxed{\sigma_{XY} = \frac{5}{3}}$$

(c) $P((X,Y) \in A) = \int_{x=0}^{1/2} \int_{y=1/4}^{1/2} \frac{x}{4}(1+3y^2)\, dx\, dy$

$$= \frac{1}{4}\left[\frac{x^2}{2}\right]_0^{1/2} \left[y + \frac{3y^3}{3}\right]_{1/4}^{1/2}$$

$$= \frac{1}{8}\left(\frac{1}{4}\right)\left(\frac{1}{2}+\frac{1}{8}-\frac{1}{4}-\frac{1}{64}\right)$$

$$= \frac{1}{32}\left(\frac{32+8-16-1}{64}\right) = \frac{23}{32 \cdot 64} = \frac{23}{2048}$$

<u>Properties</u>

1. $E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)]$
2. If X and Y are two ind random variables,
   $E(XY) = E(X)\,E(Y)$.

(b) $P[(X,Y) \in A] = P(X+Y \leq 1)$

$= f(0,0) + f(0,1) + f(1,0)$

$= \frac{3}{28} + \frac{3}{14} + \frac{9}{28}$

$= \frac{18}{28}$

$= \frac{9}{14}.$

Ex If the joint probability distribution of X and Y is given by

$f(x,y) = \frac{x+y}{30}$, for $x = 0, 1, 2, 3;\ y = 0, 1, 2,$

find

(a) $P(X \leq 2, Y = 1)$

(b) $P(X > 2, Y \leq 1)$

(c) $P(X > Y)$

(d) $P(X + Y = 4)$.

Sol :

	$f(x,y)$	x: 0	1	2	3	Row Totals
y	0	0	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{6}{30}$
	1	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{4}{30}$	$\frac{10}{30}$
	2	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{4}{30}$	$\frac{5}{30}$	$\frac{14}{30}$
Column Totals		$\frac{3}{30}$	$\frac{6}{30}$	$\frac{9}{30}$	$\frac{12}{30}$	1

(a) $P(X \leq 2, Y = 1) = f(2,1) + f(1,1) + f(0,1)$

$= \frac{3}{30} + \frac{2}{30} + \frac{1}{30}$

$= \frac{6}{30} = \frac{1}{5}.$

(b) $P(X>2, Y\le 1) = f(3,0)+f(3,1)$

$$= \frac{7}{30}$$

(c) $P(X>Y) = f(1,0)+f(2,0)+f(3,0)+f(2,1)+f(3,1)+f(3,2)$

$$= \frac{18}{30} = \frac{3}{5}$$

(d) $P(X+Y=4) = f(2,2)+f(3,1) = \frac{8}{30} = \frac{4}{15}$

Joint density function

The function $f(x,y)$ is a joint density function of the continuous random variables X and Y if

1. $f(x,y) \ge 0$ for all (x,y),
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$,
3. $P[(X,Y)\in A] = \iint_A f(x,y)\,dx\,dy$, for any region A in the xy plane.

Ex: A privately owned business operates both a drive-in facility and walk in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0\le x\le 1,\ 0\le y\le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that $f(x,y)$ is a joint density function.

(b) Find $P[(X,Y)\in A]$, where $A = \{(x,y) \mid 0<x<\frac{1}{2}, \frac{1}{4}<y<\frac{1}{2}\}$.

2: (a) $f(x,y) \geq 0$ for all (x,y).

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_0^1\int_0^1 \frac{2}{5}(2x+3y)\,dx\,dy$$

$$= \frac{2}{5}\int_0^1 \left[\frac{2x^2}{2} + 3xy\right]_0^1 dy$$

$$= \frac{2}{5}\int_0^1 (1+3y)\,dy$$

$$= \frac{2}{5}\left[y + \frac{3y^2}{2}\right]_0^1$$

$$= \frac{2}{5}\left[1+\frac{3}{2}\right] = \frac{2}{5} \times \frac{5}{2} = 1.$$

$\therefore$ $f(x,y)$ is a joint density function.

(b) $$P[(X,Y) \in A] = \int_{1/4}^{1/2}\int_0^{1/2} \frac{2}{5}(2x+3y)\,dx\,dy$$

$$= P\left(0<X<\frac{1}{2}, \frac{1}{4}<Y<\frac{1}{2}\right) = \frac{2}{5}\int_{1/4}^{1/2}\left[\frac{2x^2}{2} + 3xy\right]_0^{1/2} dy$$

$$= \frac{2}{5}\int_{1/4}^{1/2}\left(\frac{1}{4} + \frac{3}{2}y\right) dy$$

$$= \frac{2}{5}\left[\frac{y}{4} + \frac{3y^2}{4}\right]_{1/4}^{1/2}$$

$$= \frac{2}{5}\left[\frac{1}{4}\left(\frac{1}{2}\right) + \frac{3}{4}\left(\frac{1}{4}\right) - \frac{1}{4}\left(\frac{1}{4}\right) - \frac{3}{4}\left(\frac{1}{16}\right)\right]$$

$$= \frac{2}{5}\left[\frac{1}{8} + \frac{3}{16} - \frac{1}{16} - \frac{3}{64}\right]$$

$$= \frac{2}{5}\left[\frac{8+12-4-3}{64}\right] = \frac{2}{5}\,\frac{13}{64} = \frac{13}{160}.$$

Although we often collect data for two variables, som we have specific questions about just one variable. In such situations, we use marginal distributions.

Marginal Distributions

The marginal distributions of X alone and Y alone are

$$g(x) = \sum_{y} f(x,y) \quad \text{and} \quad h(y) = \sum_{x} f(x,y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

for the continuous case.

Conditional Distribution

Let X and Y be two random variables, discrete or continuous. The conditional distribution of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \quad \text{provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y=y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \quad \text{provided } h(y) > 0.$$

The Joint Probability Distribution of two random variables X and Y is given by

$P(X=0, Y=1) = \frac{1}{3}$, $P(X=1, Y=-1) = \frac{1}{3}$, $P(X=1, Y=1) = \frac{1}{3}$.

Find (i) Marginal distributions of X and Y

(ii) Conditional probability distribution of X given $Y=1$.

<u>Sol</u>:-

		x: -1	0	1	Marginal (Y)
y	-1	0	0	$\frac{1}{3}$	$\frac{1}{3}$
	0	0	0	0	0
	1	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
Marginal (X)		0	$\frac{1}{3}$	$\frac{2}{3}$	1

(i)

$$g(-1) = \sum_y f(x,y)$$

$$= f(-1,-1) + f(-1,0) + f(-1,1)$$

$$= 0 + 0 + 0$$

$$= 0$$

x	-1	0	1
g(x)	0	$\frac{1}{3}$	$\frac{2}{3}$

$$g(0) = f(0,-1) + f(0,0) + f(0,1)$$

$$= 0 + 0 + \frac{1}{3}$$

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

$$g(1) = f(1,-1) + f(1,0) + f(1,1)$$

$$= \frac{1}{3} + 0 + \frac{1}{3}$$

$$= \frac{2}{3}$$

y	-1	0	1
$g(y)$	$\frac{1}{3}$	0	$\frac{2}{3}$

$$h(-1) = \sum_{x} f(x,y)$$
$$= f(-1,-1) + f(0,-1) + f(1,-1)$$
$$= \frac{1}{3}$$

$$h(0) = f(-1,0) + f(0,0) + f(1,0)$$
$$= 0$$

$$h(1) = f(-1,0) + f(0,1) + f(1,1)$$
$$= \frac{2}{3}.$$

(ii) $$f(x|y) = \frac{f(x,y)}{g(y)}$$

$\Rightarrow$ ~~$f(x|y) =$~~

$$f(x|1) = \frac{f(x,1)}{g(1)}$$

x	-1	0	1
$f(x\|1)$	0	$\frac{1}{2}$	$\frac{1}{2}$

$$f(-1/1) = \frac{f(-1,1)}{g(1)} = 0$$

$$f(0|1) = \frac{f(0,1)}{g(1)} = \frac{1/3}{2/3} = \frac{1}{2}$$

$$f(1/1) = \frac{f(1,1)}{g(1)} = \frac{1/3}{2/3} = \frac{1}{2}$$

Ex The Joint density for the random variables (X,Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces, is

$$f(x,y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{elsewhere}. \end{cases}$$

a) Find the marginal densities $g(x)$, $h(y)$ and conditional density $f(y|x)$.

(b) Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25.

Sol: (a) $g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{x}^{1} 10xy^2\,dy$

$$= \frac{10x}{3}(1-x^3), \quad 0<x<1.$$

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{0}^{y} 10xy^2\,dy = \frac{10y^2}{2}(y^2-0)$$

$$= 5y^4, \quad 0<y<1.$$

$$f(y|x) = \frac{f(x,y)}{g(x)} = \frac{10xy^2}{\frac{10x}{3}(1-x^3)} = \frac{3y^2}{1-x^3}, \quad 0<x<y<1.$$

(b) $$P\left(Y>\frac{1}{2} \,\middle|\, X=0.25\right) = \int_{1/2}^{1} f(y|x=0.25)\,dy$$

$$= \int_{1/2}^{1} \frac{3y^2}{1-(0.25)^3}\,dy$$

$$= \frac{3}{0.98}\left.\frac{y^3}{3}\right|_{1/2}^{1}$$

$$= \frac{1}{0.98}\left(1-\frac{1}{8}\right) = \frac{1}{0.98}\left(\frac{7}{8}\right) = 0.89$$

Statistical Independence

Let X and Y be two random variables, discrete or cont with joint probability distribution $f(x,y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be statistically independent iff

$$f(x,y) = g(x)\, h(y)$$

for all (x,y) within their range.

Ex: The Joint probability density function of a random variable (X,Y) is

$$f(x,y) = \begin{cases} 2; & 0<x<1,\ 0<y<x; \\ 0, & \text{elsewhere} \end{cases}$$

(i) Find the marginal density functions of X and Y.

(ii) Find the conditional density function of Y given $X = x$ and conditional density function of X given $Y = y$.

(iii) Check for independence of X and Y.

Sol: (i) $g(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$, $h(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$

$$g(x) = \begin{cases} \int_0^x 2\, dy = 2x, & 0<x<1 \\ 0, & \text{elsewhere} \end{cases}$$

$0<y<x<1$

$$h(y) = \begin{cases} \int_y^1 2\, dx = 2(1-y), & 0<y<1 \\ 0, & \text{elsewhere} \end{cases}$$

(ii) $f(y|x) = \frac{f(x,y)}{g(x)} = \frac{2}{2x} = \frac{1}{x}, \quad 0<y<x<1$

$f(x|y) = \frac{f(x,y)}{h(y)} = \frac{2}{2(1-y)} = \frac{1}{1-y}, \quad 0<y<x<1.$

(iii) $g(x)\,h(y) = 4x(1-y) \neq f(x,y)$

$\Rightarrow$ X and Y are not independent.

2. Ex The fraction X and Y of male and female runners who compete in marathon races are described by the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y

Sol:

$$g(x) = \int_0^x 8xy \, dy = 8x\left[\frac{y^2}{2}\right]_0^x = 4x^3, \quad 0 \le x \le 1$$

$$g(x) = \begin{cases} 4x^3, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

$$h(y) = \int_y^1 8xy \, dx = 8y\left[\frac{x^2}{2}\right]_y^1 = 4y(1-y^2), \quad 0 \le y \le 1.$$

$$h(y) = \begin{cases} 4y(1-y^2), & 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

$$\mu_X = E(X) = \int_0^1 (4x^3)\, x\, dx = 4\left[\frac{x^5}{5}\right]_0^1 = \frac{4}{5}$$

$$\mu_Y = E(Y) = \int_0^1 4y^2(1-y^2)\, dy = 4\int_0^1 \left[y^2 - y^4\right] dy$$

$$= 4\left[\frac{y^3}{3} - \frac{y^5}{5}\right]_0^1$$

$$= 4\left[\frac{1}{3} - \frac{1}{5}\right] = 4\left(\frac{2}{15}\right) = \frac{8}{15}$$

$$E(XY) = \int_0^1 \int_y^1 8x^2y^2\, dx\, dy = 8\int_0^1 y^2 \left[\frac{x^3}{3}\right]_y^1 dy$$

$$= \frac{8}{3}\int_0^1 y^2(1-y^3)\, dy$$

$$= \frac{8}{3}\int_0^1 (y^2 - y^5)\, dy$$

$$= \frac{8}{3}\left[\frac{y^3}{3} - \frac{y^6}{6}\right]_0^1$$

$$= \frac{8}{3}\left[\frac{1}{3} - \frac{1}{6}\right] = \frac{8}{3}\left[\frac{1}{6}\right]$$

$$= \frac{4}{9}$$

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{4}{9} - \frac{4}{5}\cdot\frac{8}{15} = \frac{100 - \overset{96}{\cancel{288}}}{225} = \frac{4}{225}$$

Correlation Coefficient

Covariance provides the information regarding the nature of the relationship, it does not indicate anything regarding the strength of relationship, since σ_{XY} is not scale free.

There is a scale free version of Covariance called the correlation coefficient.

Correlation coefficient indicates about the directional relationship and strength of the relationship.

Note: $-1 \leq \rho_{XY} \leq 1$.

Def: Let X and Y be random variables with Covariance σ_{XY} and standard deviations σ_X, σ_Y, resp.

The correlation coefficient of X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Ex Find the correlation coefficient between X and Y in ex. 1.

Sol: $\sigma_{XY} = -\frac{9}{56}, \quad \sigma_X^2 = E(X^2) - \mu_X^2$

$\sigma_Y^2 = E(Y^2) - \mu_Y^2$

$$E(X^2) = \sum_{x=0}^{2} x g(x) = 0\left(\frac{5}{14}\right) + (1)\left(\frac{15}{28}\right) + 2^2\left(\frac{3}{28}\right) = \frac{27}{28}$$

$$E(Y^2) = \sum_{y=0}^{2} y h(y) = (0)\frac{15}{28} + (1)\left(\frac{3}{7}\right) + (2^2)\left(\frac{1}{28}\right) = \frac{16}{28} = \frac{4}{7}$$

$$\sigma_X^2 = \frac{27}{28} - \left(\frac{3}{4}\right)^2 = \frac{108-63}{112} = \frac{45}{112}$$

$$\sigma_Y^2 = \frac{4}{7} - \frac{1}{4} = \frac{16-7}{28} = \frac{9}{28}$$

$$\therefore \rho_{XY} = \frac{-\frac{9}{56}}{\sqrt{\frac{45}{112} \times \frac{9}{28}}} = \frac{-\frac{9}{56}}{\sqrt{\frac{9 \times 5 \times 9}{28 \times 4 \times 28}}} = \frac{-\frac{9}{56}}{\frac{9}{28 \times 2}\sqrt{5}} = \frac{-1}{\sqrt{5}}$$

~~Ex Find ρ_{XY} in ex. 2.~~

~~Sol: $E(X^2) = \int x^2 8xy\,dy$~~

Ex Find ρ_{XY} in Ex. 2.

Sol: $E(X^2) = \int_0^1 x^2(4x^3)\,dx = 4\left[\frac{x^6}{6}\right]_0^1 = \frac{4}{6} = \frac{2}{3}.$

$E(Y^2) = \int_0^1 y^2\, 4y(1-y^2)\,dy = 4\int_0^1 (y^3 - y^5)\,dy = 4\left[\frac{y^4}{4} - \frac{y^6}{6}\right]_0^1$

$= 4\left[\frac{1}{4} - \frac{1}{6}\right]$

$= 4\left[\frac{3-2}{12}\right]$

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

$\sigma_X^2 = E(X^2) - \mu_X^2 = \frac{2}{3} - \frac{16}{25} = \frac{50-48}{75} = \frac{2}{75}.$

$\sigma_Y^2 = E(Y^2) - \mu_Y^2 = \frac{1}{3} - \frac{64}{225} = \frac{75-64}{225} = \frac{11}{225}.$

$\therefore \rho_{XY} = \dfrac{\frac{4}{225}}{\sqrt{\frac{2}{75} \times \frac{11}{225}}} = \frac{4}{\sqrt{66}}.$

Properties

1. If X and Y are random variables with joint probability function $f(x,y)$ and a, b, c are constants, then

$$\sigma^2_{aX+bY+c} = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}.$$

Letting $b=0$ in (1)

$$\sigma^2_{aX+c} = a^2\sigma_X^2 = a^2\sigma^2$$

4. Letting $b=0$, $c=0$

$$\sigma^2_{aX} = a^2\sigma_X^2 = a^2\sigma^2.$$

3. Letting $a=1, b=0$ in (1),

$$\sigma^2_{X+c} = \sigma_X^2 = \sigma^2$$

$\Rightarrow$ Variance is unchanged if a constant is added to or subtracted from a random variable. (1), (3)

However, if a random variable is multiplied or divided by a constant, then the variance is multiplied or divided by the square of the constant. (2), (4)

5. If X and Y are independent,

$$\sigma^2_{aX+bY} = a^2\sigma_X^2 + b^2\sigma_Y^2$$

and $$\sigma^2_{aX-bY} = a^2\sigma_X^2 + b^2\sigma_Y^2.$$

6. If $X_1, X_2, \ldots, X_n$ are independent random variables,

$$\sigma^2_{a_1X_1+a_2X_2+\cdots+a_nX_n} = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + \cdots + a_n^2\sigma_{X_n}^2.$$

Chebyshev's Theorem

If X is a RV with mean μ and variance σ^2, then for any positive real number k, we have,

$$P\{|X-\mu| < k\sigma\} \geq 1-\frac{1}{k^2}$$

or $$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}.$$

Ex A random variable X has mean $\mu=10$ and variance $\sigma^2=4$. Find

(a) $P(|X-10| \geq 3)$ (b) $P(|X-10|<3)$ (c) $P(5<X<15)$

(d) The value of constant c s.t. $P(|X-10| \geq c) \leq 0.04$.

(e) Find lower bound for $P(5<X<15)$.

Sol: $\mu=10,\ \sigma^2=4 \Rightarrow \sigma=2$

(a) $k\sigma=3 \Rightarrow k=\frac{3}{2}$

$\therefore P(|X-10| \geq 3) \leq \frac{1}{\left(\frac{3}{2}\right)^2} = \frac{4}{9} \Rightarrow P(|X-10| \geq 3) \leq \frac{4}{9}.$

(b) $P(|X-10|<3) \geq 1-\frac{1}{\left(\frac{3}{2}\right)^2} = \frac{5}{9}$

$\Rightarrow P(|X-10|<3) \geq \frac{5}{9}.$

(c) $P(5<X<15) = P(5-10< X-10< 15-10)$

$= P(-5< X-10 < 5)$

$= P(|X-10|<5)$

$k\sigma=5$

$k=\frac{5}{2}$

$\geq 1-\frac{1}{\left(\frac{5}{2}\right)^2} = \frac{21}{25}$

$\Rightarrow P(5<X<15) \geq \frac{21}{25}.$

Q: If X is a RV with $E(X) = 3$ and $E(X^2) = 13$. Use the Chebyshev's Inequality to determined lower bound for $P(-2 < X < 8)$.

Given $E(X) = 3$, $E(X^2) = 13$.

Since, $\sigma_X^2 = E(X^2) - [E(X)]^2$

$= 13 - 9 = 4$

$\boxed{\sigma_X = 2}$

To find lower bound

$P(-2 < X < 8) = P(-2 - 3 < X - 3 < 8 - 3)$

$= P(-5 < X - 3 < 5)$

$= P(|X - 3| < 5)$

Using, Chebyshev's inequality,

$P\{|X - \mu| < K\sigma\} \geq 1 - \frac{1}{K^2}$

Comparing the above equations, $K\sigma = 5 \Rightarrow K = \frac{5}{2}$.

$\therefore$ The lower bound is

$1 - \frac{1}{K^2} = 1 - \frac{1}{\left(\frac{5}{2}\right)^2}$

$= 1 - \frac{4}{25} = \frac{21}{25}$.

Ex Compute $P(\mu - 2\sigma < X < \mu + 2\sigma)$, where X has the density function

$$f(x) = \begin{cases} 6x(1-x), & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

and compare with the result given in Chebyshev's theorem.

Sol:

$$\mu = E(X) = \int_0^1 6x^2(1-x)\,dx = 6\int_0^1 (x^2 - x^3)\,dx = 6\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1$$

$$\cancel{= 6\left(\frac{1}{2} - \frac{1}{3}\right)}$$

$$\cancel{= 6 \cdot \frac{1}{6}}$$

$$= 6\left[\frac{1}{3} - \frac{1}{4}\right]$$

$$= 6\left[\frac{1}{12}\right] = \frac{1}{2} = 0.5$$

$$E(X^2) = 6\int_0^1 x^2 x(1-x)\,dx = 6\int_0^1 (x^3 - x^4)\,dx$$

$$= 6\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1$$

$$= 6\left[\frac{1}{4} - \frac{1}{5}\right] = 6 \cdot \frac{1}{20} = 0.3$$

$\sigma^2 = 0.3 - 0.25 = 0.05$

$\sigma = 0.2236$

$\therefore P(0.5 - 2(0.2236) < X < 0.5 + 2(0.2236)) = P(0.0528 < X < 0.9472)$

$$= \int_{0.0528}^{0.9472} 6x(1-x)\,dx = 6\int_{0.0528}^{0.9472} (x - x^2)\,dx$$

$$= 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{0.0528}^{0.9472}$$

$$= 6\left[0.4486 - 0.2833 - 0.0014 + 0.00005\right]$$

$$= 0.9837$$

By Chebyshev's th^m,

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{4} = \frac{3}{4} = 0.75.$$