Random Variables

Consider two experiments.

E.1 : Number of Iphones sold by an Apple store in Delhi in 2022.

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

E.2 : 2 coins are tossed.

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

E.3 : A ~~Two~~ coin is ~~are~~ tossed until we get a tails.

$S = \{T, HT, HHT, HHHT, HHHHT, \ldots\}$

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 the experiment.

For ex, in E.2, we are interested in number of tails occur.

So, the numerical values assigned with each element of the sample space are 0, 1, 2.

We represent these numerical values by the random variable X.

Def : 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 a subset of $\mathbb{R}$.

* Each possible value of X respresents an event.

Random Variable

It assumes only a finite or at most countable number of values → Discrete Random Variable (certain value 5)

1. No of students in a class
2. No of accidents
3. No of defective items
4. No of heads, when we toss a coin.

It assumes infinite number of possible values → Continuous Random Variable (3.265)

Eg : life period, height, weight, temparature

Discrete Random variable

A r.v. is called a discrete random variable if its set possible outcomes is countable.

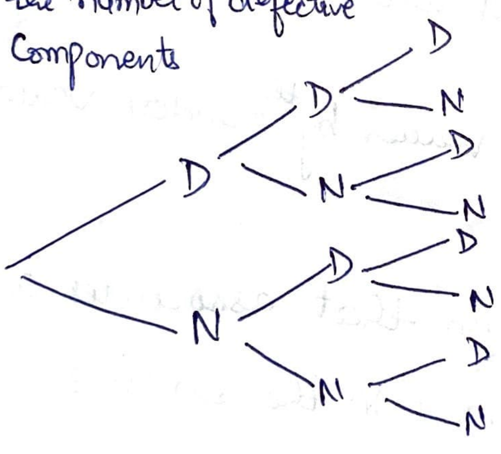
Discrete Sample Space

If a sample space contains a finite number of possibilities or an unending sequence will as many elements as there are whole numbers, it is called discrete sample space.

Ex. 1 E: Testing three electronic components.

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

X : a r.V. which describes the number of defective components



Outcomes	X
DDD	3
DDN	2
DND	2
DNN	1
NDD	2
NDN	1
NND	2
NNN	0

X : 0, 1, 2, 3.

Ex. 2 A stockroom clerk returns three safety helmets at random to three steel mill employees who had previously checked them. If Smith, Jones and Brown, in that order, receive one of the three hats, list the sample points for the possible orders of returning the helmets, and find the value m of the random variable M that represents the number of correct matches.

Sample space	m
SJB	3
SBJ	1
JSB	1
JBS	0
BSJ	0
BJS	1

Ex.3 Consider the simple condition in which components are arriving from the production line and they are stipulated to be defective or not defective.

Define the random variable

$$X = \begin{cases} 0, & \text{if the component is not defective} \\ 1, & \text{if the component is defective} \end{cases}$$

Ex.4 E: Sampling items until a defective item is observed.

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

$X = \{1, 2, 3, 4, ---\}$, where X be a random variable defined by the number of items observed before a defective is found.

Continuous Random Variable

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

Ex: (1) Weight of a group of individuals. amount of milk produced by a cow

(2) Height of a group of individuals.

(3) Length of time for a chemical reaction to take place.

Continuous Sample Space

If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a continuous sample space.

Continuous R.V.	Discrete R.V.
Represent measured data heights, weights, temperatures, distance, or life periods.	Represent count data Number of defective items or number of highway casualties.

Discrete probability distributions

X -: random variable associated with number of heads in tossing two coins.

X	0	1	2
p(X)	1/4	2/4	1/4

$\downarrow$ $f(0) = P(X=0)$ $\quad$ $f(1) = P(X=1)$ $\quad$ $f(2)$ $P(X=2)$

Probability Function, Probability Mass function or Probability distribution

Def: The set of ordered pairs $(x, f(x))$ is a probability mass function, 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. $P(X=x) = f(x)$.

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 for the number of defectives.

Sol: Let X be the random variable whose values x are the possible numbers of defective computer purchased by the school.

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

$$f(0) = P(X=0) = \frac{{}^{17}C_2}{{}^{20}C_2} = \frac{\frac{17!}{2!\,15!}}{\frac{20!}{2!\,18!}} = \frac{17.16}{20.19} = \frac{68}{95}$$

$$f(1) = P(X=1) = \frac{{}^{17}C_1\ {}^{3}C_1}{{}^{20}C_2} = \frac{17.3}{10.19} = \frac{51}{190}$$

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

The probability distribution of X is

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

Ex: If a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags, find the formula for the probability distribution of the number of cars with side airbags among the next 4 cars sold by the agency?

$x=0$ 0 car is sold with airbags.

NNNN $\frac{4!}{4!\,0!} = {}^4C_0$

$= 1$

$x=1$ 1 car is sold with airbags.

ANNN, NANN, NNAN, NNNA

$\frac{4!}{1!\,3!} = {}^4C_1$

$x=2$ 2 cars are sold with airbags.

$\frac{4!}{2!\,2!} = {}^4C_2$

$\frac{{}^4C_x}{16}$

$x=3$ 3 cars with sidebags.

$\frac{4!}{3!\,1!} = {}^4C_3$

$x=4$ 4 cars with side airbags.

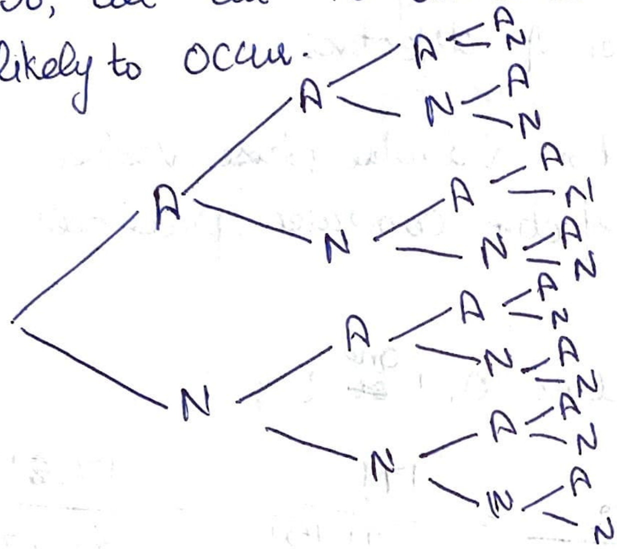
$\frac{4!}{4!\,0!} = {}^4C_4$

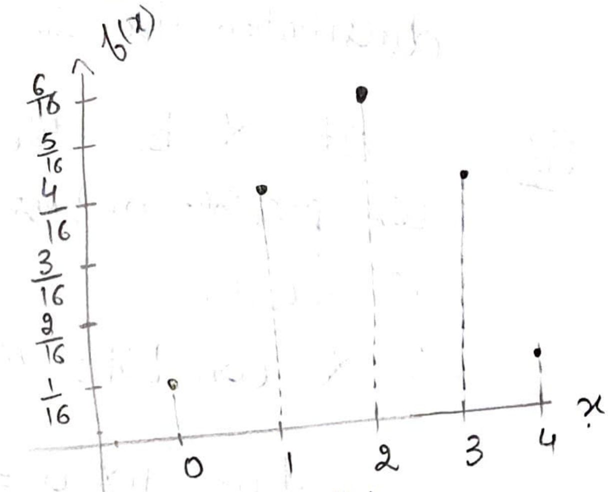
${}^4C_x$, $x=0,1,$

Sol: P(car is equipped with side airbags) = $\frac{1}{2}$.

P(car is without ~~air~~ airbags) = $\frac{1}{2}$.

So, all the 16 outcomes of the sample space are equally likely to occur.





p.m.f. plot

Let X be the random variable which represents the number of cars sold with side airbags.

$\therefore$ X takes the values 0, 1, 2, 3, 4.

$P(X=0) = \frac{^4C_0}{16} = \frac{1}{16}$, $P(X=3) = \frac{^4C_3}{16} = \frac{1}{4} = \frac{4}{16}$

$P(X=1) = \frac{^4C_1}{16} = \frac{1}{4} = \frac{4}{16}$, $P(X=4) = \frac{^4C_4}{16} = \frac{1}{16}$

$P(X=2) = \frac{^4C_2}{16} = \frac{3}{8} = \frac{6}{16}$

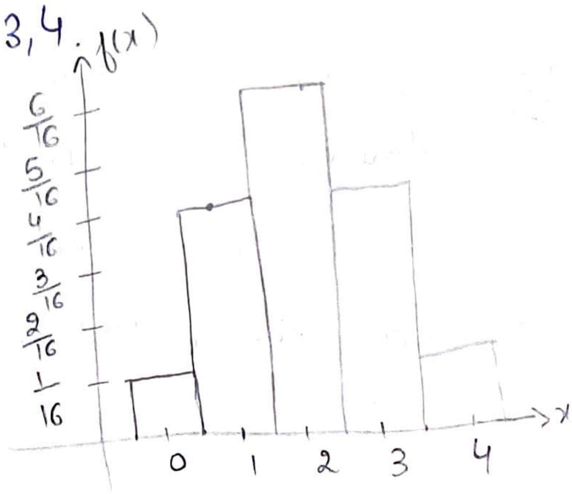
x	0	1	2	3	4
$f(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Thus, the probability distribution $f(x) = P(X=x)$ is

$$f(x) = \frac{^4C_x}{16} \quad \text{for } x = 0, 1, 2, 3, 4$$

* bases are centered at each value of x, heights are equal to corresponding probabilities

Probability histogram →



Cumulative Distribution Function

In some problems, we want to find the probability that the value of random variable X will be less than or equal to some real number x.

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

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

Increasing by successive additions

Ex Find c.d.f. in previous example. Using $F(x)$, verify that $f(2) = \frac{3}{8}$.

Sol: $f(0) = \frac{1}{16}$, $f(1) = \frac{1}{4}$, $f(2) = \frac{3}{8}$, $f(3) = \frac{1}{4}$, $f(4) = \frac{1}{16}$.

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

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

$= f(0) + f(1) = \frac{5}{16}$

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

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

$F(4) = 1.$

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

$-\infty < x < \infty$

Hence,

$-\infty < x < 0$

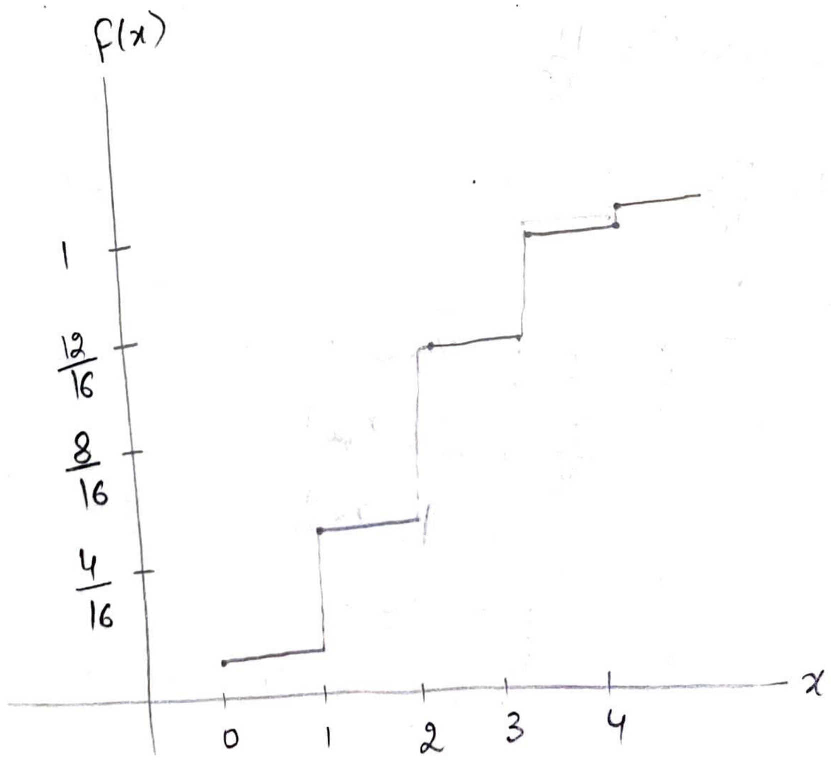
$$F(x) = \begin{cases} 0, & \text{for } x < 0 \\ \frac{1}{16}, & \text{for } 0 \le x < 1, \\ \frac{5}{16}, & \text{for } 1 \le x < 2, \\ \frac{11}{16}, & \text{for } 2 \le x < 3 \\ \frac{15}{16}, & \text{for } 3 \le x < 4 \\ 1, & \text{for } x \ge 4. \end{cases}$$

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

$= f(2) + F(1)$

$\Rightarrow f(2) = F(2) - F(1)$

$= \frac{11}{16} - \frac{5}{16} = \frac{6}{16} = \frac{3}{8}.$

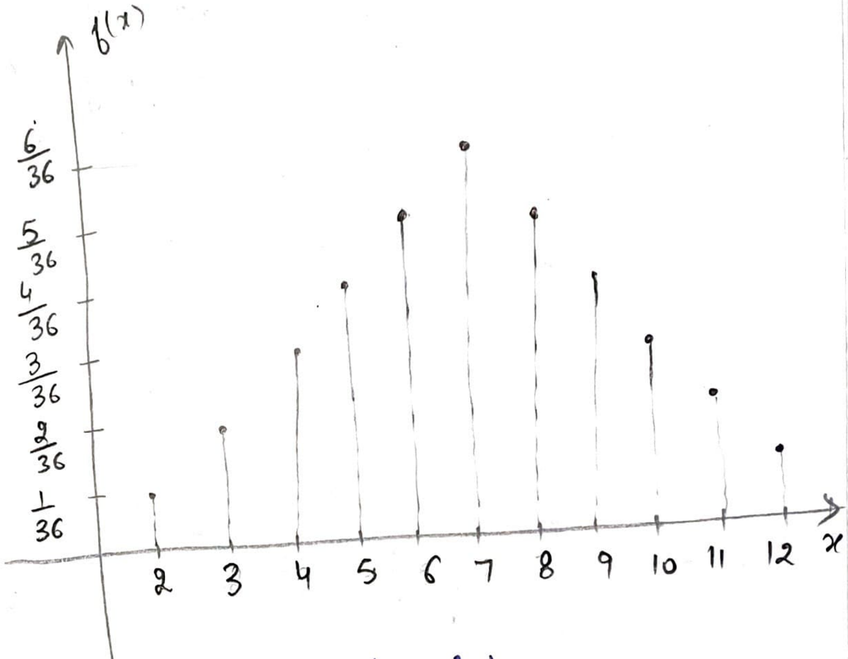


Discrete cumulative distribution function

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. Also, find the 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

Distribution function

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

$\cancel{F(0) = P(X \le 0)}$ $F(1) = P(X \le 1) = 0.$

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

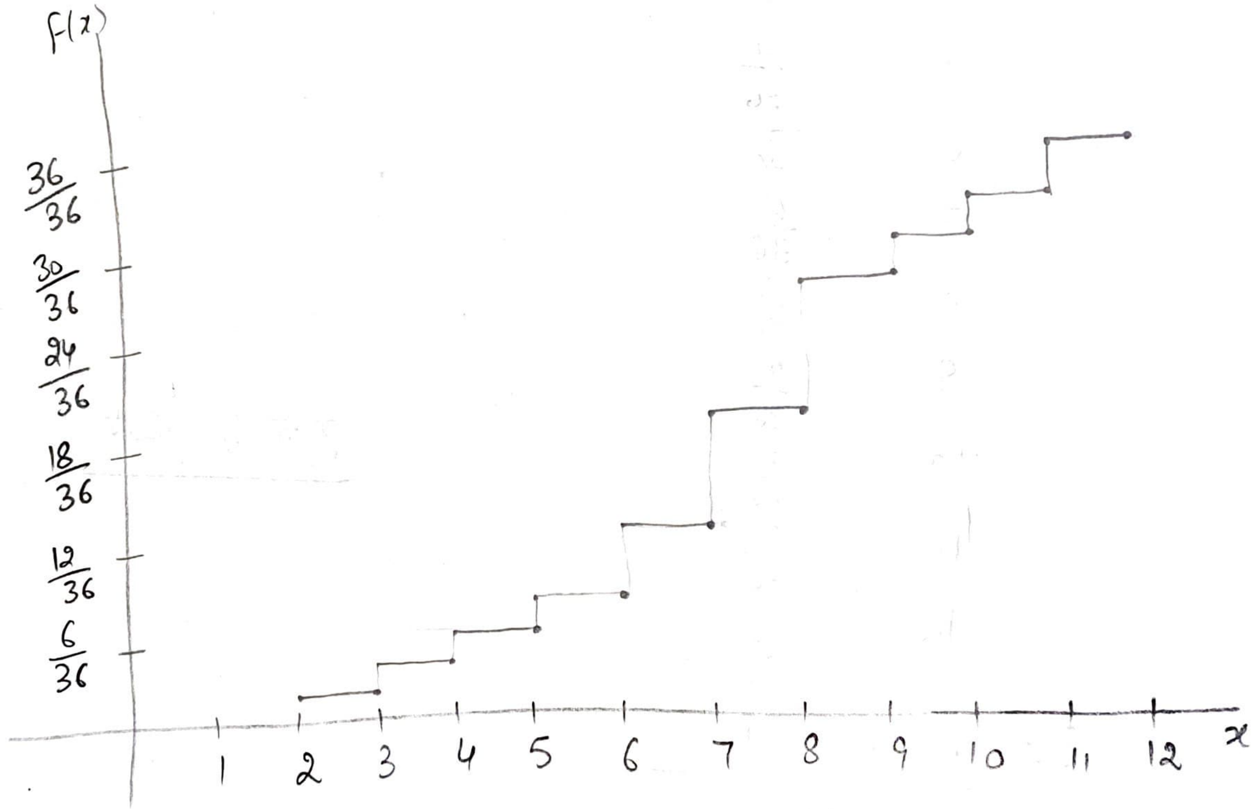
$F(3) = P(X \le 3) = \frac{3}{36}$

$F(4) = P(X \le 4) = \frac{6}{36}$

$F(5) = \frac{10}{36}$, $F(6) = \frac{15}{36}$, $F(7) = \frac{21}{36}$, $F(8) = \frac{26}{36}$, $F(9) = \frac{30}{36}$,

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

$$F(x) = \begin{cases} 0 & , \text{for } x<2, \\ \frac{1}{36} & , \text{for } 2 \le x < 3 \\ \frac{3}{36} & , \text{for } 3 \le x < 4 \\ \frac{6}{36} & , \text{for } 4 \le x < 5 \\ \vdots & \\ \frac{35}{36} & , \text{for } 11 \le x < 12 \\ 1 & , \text{for } x \ge 12 \end{cases}$$



<u>discrete c.d.f.</u>

Continuous Probability Distributions

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

Ex: If X is a random variable defined by the number of items observed before a defective is found.

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

X can take value 1, 2, 3, -----

But, when we have a continuous set, e.g., the interval $[0,1]$, 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, ? -

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

For ex, Let X be the random variable whose values are heights of all people over 21 years of age.

Between 163.99 and 164.01 cm, there are infinite number of heights, one of which is ~~exactly~~ 164 cm.

The probability of selecting a person at random who is exactly 164 cm tall and not one of the infinitely large set of heights so close to 164 cm that you cannot humanly measure the difference.

Thus, we assign a probability 0 to the event.

We can find the probability of selecting a person who is at least 163 cm but not more than 165 cm tall.

i.e. When, we are dealing with an interval rather than a point

When X is Continuous random variable,

$$P(X=x) = 0 \text{ for all } x.$$

A continuous random variable takes values in a continuous interval (a,b).

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

i.e., It does not matter whether we include an endpoint of the interval or not.

Probability density function

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

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

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

Ex Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a Continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere}. \end{cases}$$

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

$\int_{-\infty}^{-1} 0 \quad \int_{-1}^{2} \quad + \int_{2}^{\infty} 0$

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

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

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

$x < 0$

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

(b) $$P(0 \leq X \leq 1) = \int_0^1 f(x)\,dx = \int_0^1 \frac{x^2}{3}\,dx = \frac{1}{3}\left[\frac{x^3}{3}\right]_0^1 = \frac{1}{9}$$

$P(X \leq x)$ (c) $$F(x) = \int_{-1}^{x} \frac{t^2}{3}\,dt = \frac{1}{3}\left[\frac{t^3}{3}\right]_{-1}^{x} = \frac{1}{9}\left[x^3+1\right] = \frac{x^3+1}{9}, \text{ for } -1 \leq x < 2$$

$2 \leq x < \infty$

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

$$\therefore \quad F(x) = \begin{cases} 0, & x < -1 \\ \frac{x^3+1}{9}, & -1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

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

$= \left.\frac{x^3}{9}\right|_{-1}^{2}$

$= \frac{1}{9}(8+1) = 1$

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

Ex: The department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the density function of the winning (low) bid is

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \le y \le 2b, \\ 0, & \text{elsewhere}. \end{cases}$$

Find $F(y)$ and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate b.

Sol: For $\frac{2b}{5} \le y \le 2b$,

$$F(y) = \int_{2b/5}^{y} \frac{5}{8b}\, dt = \frac{5}{8b}\left[t\right]_{2b/5}^{y} = \frac{5}{8b}\left[y - \frac{2b}{5}\right]$$

$$= \frac{5y}{8b} - \frac{1}{4}.$$

$$\therefore F(y) = \begin{cases} 0, & y < \frac{2}{5}b \\ \frac{5y}{8b} - \frac{1}{4}, & \frac{2b}{5} \le y < 2b \\ 1, & y \ge 2b \end{cases}$$

Now, $P(Y \le b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}$.

Ex: If X is a random variable taking the value $\frac{1}{4}$ in the interval $[1,5]$, prove that $f(x) = \frac{1}{4}$ is the pdf of X. ~~Find~~ ~~also~~.

Sol: Clearly $f(x) \ge 0$ for all $x \in [1,5]$.

and $\int_{-\infty}^{\infty} f(x)\, dx = \int_{1}^{5} \frac{1}{4}\, dx = \frac{1}{4}(5-1) = 1.$

$\therefore f(x) = \frac{1}{4}$ is the pdf of X.

Ex If $f(x)=\begin{cases} \frac{x}{15}; & x=1,2,3,4,5 \\ 0, & \text{elsewhere} \end{cases}$

Find (i) $P\{X=1 \text{ or } 2\}$

(ii) $P\left\{\frac{1}{2}<X<\frac{5}{2} \middle| X>1\right\}$

Sol: (i) $P(X=1 \text{ or } 2) = P(X=1)+P(X=2) = \frac{1}{15}+\frac{2}{15}=\frac{1}{5}.$

(ii) $$P\left\{\frac{1}{2}<X<\frac{5}{2} \middle| X>1\right\} = \frac{P\left\{\left(\frac{1}{2}<X<\frac{5}{2}\right)\cap X>1\right\}}{P(X>1)}$$

$P(X=2)+P(X=3)+P(X=4)+P(X=5)$

$\frac{2}{15}+\frac{3}{15}+\frac{4}{15}+\frac{5}{15}$

$\frac{14}{15}$

$$= \frac{P\{(X=1 \text{ or } 2)\cap(X>1)\}}{1-P(X=1)}$$

$$= \frac{P(X=2)}{1-P(X=1)} = \frac{\frac{2}{15}}{1-\frac{1}{15}} = \frac{2}{14} = \frac{1}{7}.$$

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 $F(x)$, the c.d.f. and

(iii) If x_1, x_2 and x_3 are three independent observations from X, what is the probability that exactly one of these numbers is larger than 1.5?

Sol: (i) Since $f(x)$ is p.d.f. of X,

$$\therefore \int_0^3 f(x)\,dx = 1$$

$$\Rightarrow \int_0^1 kx\,dx + \int_1^2 k\,dx + \int_2^3 (-kx+3k)\,dx = 1$$

$$\Rightarrow k\frac{x^2}{2}\Big|_0^1 + kx\Big|_1^2 - k\frac{x^2}{2}\Big|_2^3 + 3kx\Big|_2^3 = 1$$

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

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

(ii) For $-\infty < x < 0$, $F(x) = \int_{-\infty}^{x} 0\,dt = 0$.

$0 \le x < 1$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{0} f(t)\,dt + \int_0^x f(t)\,dt$

$$= \int_0^x \frac{1}{2}t\,dt = \frac{1}{2}\frac{t^2}{2}\Big|_0^x = \frac{1}{4}x^2$$

$1 \le x < 2$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^1 \frac{1}{2}t\,dt + \int_1^x \frac{1}{2}\,dt$

$$= \frac{t^2}{4}\Big|_0^1 + \frac{1}{2}t\Big|_1^x$$

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

$2 \le x < 3$, $F(x) = \int_0^1 \frac{t}{2}\,dt + \int_1^2 \frac{1}{2}\,dt + \int_2^x \left(-\frac{t}{2} + \frac{3}{2}\right)dt$

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

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

$$= -\frac{x^2}{4} + \frac{3x}{2} - \frac{5}{4}$$

$3 \le x < \infty$, $F(x) = \int_0^1 \frac{t}{2}\,dt + \int_1^2 \frac{1}{2}\,dt + \int_2^3 \left(-\frac{t}{2} + \frac{3}{2}\right) dt + \int_3^x 0\,dt$

$$= \frac{1}{4} + \frac{1}{2} - \frac{1}{2}\frac{t^2}{2}\Big|_2^3 + \frac{3}{2}t\Big|_2^3$$

$$= \frac{1}{4} + \frac{1}{2} - \frac{9}{4} + \frac{4}{4} + \frac{9}{2} - \frac{3}{2}\cdot 2$$

$$= 1.$$

$$F(x) = \begin{cases} 0, & \text{for } -\infty < x < 0 \\ \frac{x^2}{4}, & \text{for } 0 \le x < 1 \\ \frac{2x-1}{4}, & \text{for } 1 \le x < 2 \\ -\frac{x^2}{4} + \frac{3x}{2} - \frac{5}{4}, & \text{for } 2 \le x < 3 \\ 1, & \text{for } 3 \le x < \infty \end{cases}$$

(iii) $P(X > 1.5) = 1 - P(X < 1.5) = 1 - F(1.5) = 1 - \left(\frac{2\left(\frac{3}{2}\right) - 1}{4}\right)$

$\int_{1.5}^{\infty}$

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

Hence, the probability that exactly one is larger than 1.5 out of x_1, x_2, x_3 is

Joint Probability Distributions

Two-dimensional Random Variable

Let X and Y be two random variables defined on the same sample space S, then the function (X,Y) that assigns a point in R^2, is called a two dimensional random variable.

Ex ① Let X is associated with height of persons and Y is associated with weight of persons in an educational institute.
Then, (X,Y) represents the height and weight of every person of the institute.

② Let X represents the boy child and Y represents the girl child of a family having 2 children.
(X,Y) can take values (0,1), (1,0), (0,0), (1,1)

③ (X,Y), X: Gender, Y: Sports.

Joint Probability Distribution | Probability Mass Function

The function $f(x,y)$ is a joint probability distribution or p.m.f. of the discrete random variables X and Y if

1. $f(x,y) \geq 0$ for all (x,y),
2. $\sum_x \sum_y f(x,y) = 1$,
3. $P(X=x, Y=y) = f(x,y)$.

Ex 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) the joint probability function $f(x,y)$.

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

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

$f(x,y)$		$x=0$	$x=1$	$x=2$	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

$$f(0,0) = \frac{3c_0 \times 2c_0 \times 3c_2}{8c_2} = \frac{3}{28}$$

$$f(0,1) = \frac{3c_0 \times 2c_1 \times 3c_1}{8c_2} = \frac{3}{14}$$

$$f(0,2) = \frac{3c_0 \times 2c_2 \times 3c_0}{8c_2} = \frac{1}{28}$$

$$f(1,0) = \frac{3c_1 \times 2c_0 \times 3c_1}{8c_2} = \frac{9}{28}$$

$$f(1,1) = \frac{3c_1 \times 2c_1 \times 3c_0}{8c_2} = \frac{3}{14}$$

$f(1,2) = 0$

$$f(2,0) = \frac{3c_2 \times 2c_0 \times 3c_0}{8c_2} = \frac{3}{28}$$

$f(2,1) = 0$

$f(2,2) = 0$

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

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

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

$$= \frac{18}{28}$$

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

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

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

find

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

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

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

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

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

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

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

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

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

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

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

Sol: (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, sometimes 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

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

Ex 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 $Y=1$.

Sol:-

		x			Marginal (Y)
		-1	0	1	
	-1	0	0	$\frac{1}{3}$	$\frac{1}{3}$
y	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) $g(-1) = \sum_{y} f(x,y)$

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

$= 0 + 0 + 0$

$= 0$

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

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

$= \frac{1}{3}$

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

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

$= \frac{2}{3}$

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

$$h(-1) = \sum_x f(x,y)$$

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

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

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

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

$$= 0$$

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

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

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

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

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

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

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

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

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

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

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

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

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

Statistical Independence

Let X and Y be two random variables, discrete or continuous, 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) $f(y|x) = \dfrac{f(x,y)}{g(x)} = \dfrac{2}{2x} = \dfrac{1}{x}, \quad 0<y<x<1$

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

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

$\Rightarrow$ X and Y are not independent.

Mathematical Expectation | Expected Value | Mean of a Random Variable

Suppose a gambler is interested in his average winnings at a game, a businessman is interested in his average profits on a product or a salesperson is interested in his expected commission on the sale of products, and so on. In such situations, the concept of mathematical expectation can be used.

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

$$\mu = E(X) = \sum_x x f(x), \text{ if } X \text{ is discrete random variable.}$$

and $$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx, \text{ if } X \text{ is continuous random variable.}$$

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

Sol: Let X represent the number of good components.

$$f(0) = \frac{{}^4C_0\, {}^3C_3}{{}^7C_3} = \frac{1}{35}$$

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

$$f(1) = \frac{{}^4C_1\, {}^3C_2}{{}^7C_3} = \frac{12}{35}$$

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

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

$$\therefore \; \mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + 3\left(\frac{4}{35}\right)$$

$$= \frac{60}{35} = 1.7.$$

Thus, if a sample of size 3 is selected at random over and over again from a lot of 4 good components and 3 defective components, it will contain, on average, 1.7 good components.

<u>Ex</u> (a) Find the expectation of the number on a die when thrown.

(b) Two unbiased dice are thrown. Find the expected values of the sum of numbers of points on them.

<u>Sol</u>: (a) Let X be a random variable representing the number on a die when thrown.

X	1	2	3	4	5	6
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\mu = E(X) = (1)\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6}$$

$$= \frac{21}{6} = \frac{7}{2}.$$

(b) Let X be a random variable representing the sum of numbers on two dice.

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

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

$$= 2\times\frac{1}{36} + 3\times\frac{2}{36} + 4\times\frac{3}{36} + 5\times\frac{4}{36} + 6\times\frac{5}{36} + 7\times\frac{6}{36} + 8\times\frac{5}{36}$$

$$+ 9\times\frac{4}{36} + 10\times\frac{3}{36} + 11\times\frac{2}{36} + 12\times\frac{1}{36}$$

$$= \frac{252}{36} = 7.$$

Ex Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is

$$f(x) = \begin{cases} \frac{20,000}{x^3}, & x > 100 \\ 0, & \text{elsewhere} \end{cases}$$

Find the expected life of this type of device.

Sol: $$\mu = E(X) = \int_{100}^{\infty} x\left(\frac{20000}{x^3}\right) dx = \int_{100}^{\infty} \frac{20,000}{x^2} dx = 20,000 \int_{100}^{\infty} \frac{dx}{x^2}$$

$$= 20,000 \left[\frac{x^{-1}}{-1}\right]_{100}^{\infty} = -20000\left[0 - \frac{1}{100}\right]$$

$$= \frac{20,000}{100}$$

$$= 200.$$

∴ We can expect that the device has an average life of 200 hours.

Consider a new random variable $g(X)$, which depends on X.

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

$$\mu_{g(X)} = E(g(X)) = \sum_{x} g(x) f(x) \text{ if } X \text{ is discrete,}$$

and

$$\mu_{g(X)} = E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx, \text{ if } X \text{ is Continuous.}$$

Ex Suppose that the number of cars X that pass through a car wash between 4:00 pm to 5:00 pm on any sunny Friday has the following probability distribution:

x	4	5	6	7	8	9
$P(X=x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(X) = 2X-1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Sol.

$$E[g(X)] = E[2X-1] = \sum_{x=4}^{9} (2x-1) f(x)$$

$$= 7\left(\tfrac{1}{12}\right) + 9\left(\tfrac{1}{12}\right) + 11\left(\tfrac{1}{4}\right) + 13\left(\tfrac{1}{4}\right) + 15\left(\tfrac{1}{6}\right) + 17\left(\tfrac{1}{6}\right)$$

$$= \frac{7+9+33+39+30+\cancel{28}\,34}{12} = \frac{152}{12}$$

$$= 12.67.$$

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

$$E(X) = \sum_{x=4}^{9} x f(x) = (4)\left(\frac{1}{12}\right) + 5\left(\frac{1}{12}\right) + 6\left(\frac{1}{4}\right) + 7\left(\frac{1}{4}\right) + 8\left(\frac{1}{6}\right) + 9\left(\frac{1}{6}\right)$$

$$= \frac{4+5+18+21+16+18}{12} = \frac{82}{12}$$

$$E(2X-1) = 2E(X)-1 = 2\left(\frac{82}{12}\right) - 1 = \frac{152}{12} = 12.67.$$

Ex Let X be a random variable with density function,

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere} \end{cases}$$

Find the expected value of $g(X) = 4X+3.$

Sol:

$$E(4X+3) = \int_{-1}^{2} (4x+3)\frac{x^2}{3}\,dx = \frac{1}{3}\left[\frac{4x^4}{4} + \frac{3x^3}{3}\right]_{-1}^{2}$$

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

$$= \frac{24}{3} = 8.$$

Or

$$E(4X+3) = 4E(X)+3 = 4\int_{-1}^{2} x\left(\frac{x^2}{3}\right)dx + 3$$

$$= \frac{4}{3}\left[\frac{x^4}{4}\right]_{-1}^{2} + 3 = \frac{1}{3}\left[16-1\right] + 3 = \frac{15+9}{3} = 8.$$

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

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y)\, f(x,y) \text{ if } X \text{ and } Y \text{ are}$$

discrete random variables.

and

$$\mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dx\, dy \text{ if } X \text{ and } Y \text{ are}$$

continuous random variables.

Ex Let X and Y be the random variables with joint probability distribution

$f(x,y)$		$x=0$	$x=1$	$x=2$	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

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

Sol:

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

$$= \sum_{x=0}^{2} \left[x(0)\, f(x,0) + x(1)\, f(x,1) + x(2)\, f(x,2)\right]$$

$$= \cancel{(0)(0)\, f(0,0) + (1)(1)\, f(1,1) + (2)(2)\, f(2,2)}$$

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

$= f(1,1) + 0$

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

Ex:- Find $E(Y/X)$ for the density function

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

Sol:

$g(X,Y) = Y/X$

$$E\left(\frac{Y}{X}\right) = \int_0^1 \int_0^2 \frac{y}{\cancel{x}} \frac{\cancel{x}(1+3y^2)}{4}\, dx\, dy$$

$$= \int_0^1 \frac{y(1+3y^2)}{4}(2)\, dy = \frac{1}{2}\int_0^1 (y+3y^3)\, dy$$

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

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

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

Properties of Means

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

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

Ex: $Y = (X-1)^2$

$$E(Y) = E[(X-1)^2] = E(X^2+1-2X) = E(X^2) - 2E(X) + E(1).$$

Cor of 1. If $a=0$, $E(b)=b$.

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

3. $E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)]$

4. If X and Y are two independent random variables,

$$E(XY) = E(X)\,E(Y).$$

Variance and Covariance of Random Variables

Variance refers to the spread of a data set. It's a measurement used to identify how far each number in the data set is from the mean.

A variance value of zero represents that all of the values within a data set are identical.

A large variance means that the numbers in the set are far from mean and each other.

A small variance means that the numbers are closer together in value.

Variance

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

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \text{ if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx, \text{ if } X \text{ is continuous.}$$

The positive square root of the variance σ is called the standard deviation of X.

Alternative formula:

$$\boxed{\sigma^2 = E(X^2) - \mu^2}$$

Ex: Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Find σ_X^2.

Sol: $\mu_X = (0)(0.51) + (1)(0.38) + 2(0.10) + 3(0.01)$ $\boxed{\mu = \sum_x x f(x)}$

$= 0.38 + 0.2 + 0.03$

$= 0.61$

$$\sigma_X^2 = \sum_{x=0}^{3} (x-\mu_X)^2 f(x) = (0-0.61)^2 f(0) + (1-0.61)^2 f(1)$$

$$+ (2-0.61) f(2) + (3-0.61) f(3)$$

$$= (0.3721)(0.51) + (0.1521)(0.38) + (\overset{(1.9321)}{\cancel{1.39}})(0.10) +$$

$$(5.7121)\,\cancel{(2.39)}\,(0.01)$$

$$= (0.1898) + (0.0578) + \overset{0.19321}{\cancel{0.139}} + \overset{0.0571}{\cancel{0.0239}}$$

$$= 0.4979$$

or

$$E(X^2) = \sum_{x=0}^{3} x^2 f(x) = (0)(0.51) + 1(0.38) + 4(0.10) + 9(0.01)$$

$$= 0.38 + 0.4 + 0.09$$

$$= 0.87$$

$$\sigma_X^2 = E(X^2) - \mu_X^2 = 0.87 - 0.3721 = 0.4979.$$

Ex → The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a cont. random variable X having the probability density function

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

Find the mean and variance of X.

Sol:

$$\mu = E(X) = \int_1^2 x\, 2(x-1)\, dx = 2\int_1^2 (x^2 - x)\, dx$$

$$= 2\left[\frac{x^3}{3} - \frac{x^2}{2}\right]_1^2$$

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

$$= 2\left[\frac{16 - 12 - 2 + 3}{6}\right]$$

$$= 2\left(\frac{5}{6}\right) = \frac{5}{3}.$$

$$\sigma^2 = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = 2\int_1^2 \left(x - \frac{5}{3}\right)^2 (x-1)\, dx$$

$$= 2\int_1^2 \left(x^2 + \frac{25}{9} - \frac{10x}{3}\right)(x-1)\, dx$$

$$= \cancel{2\left[\frac{x^3}{3} + \frac{25}{9}x\right]} = 2\int_1^2 \left(x^3 - x^2 + \frac{25}{9}x - \frac{25}{9} - \frac{10x^2}{3} + \frac{10x}{3}\right) dx$$

$$= 2\left[\frac{x^4}{4} - \frac{13x^3}{9} + \frac{55}{9}\frac{x^2}{2} - \frac{25}{9}x\right]_1^2$$

$$= 2\left[\frac{16}{4} - \frac{104}{9} + \frac{110}{9} - \frac{50}{9} - \frac{1}{4} + \frac{13}{9} - \frac{55}{18} + \frac{25}{9}\right]$$

$$= 2\left[\frac{144 - 416 + 440 - 200 - 9 + 52 - 110 + 100}{36}\right]$$

$$= \frac{2}{36} = \frac{1}{18}.$$

or $$E(X^2) = \int_1^2 2x^2(x-1)\,dx = 2\int_1^2 (x^3 - x^2)\,dx$$

$$= 2\left[\frac{x^4}{4} - \frac{x^3}{3}\right]_1^2$$

$$= 2\left[\frac{16}{4} - \frac{8}{3} - \frac{1}{4} + \frac{1}{3}\right]$$

$$= 2\left[\frac{48-32-3+4}{12}\right]$$

$$= 2\left[\frac{17}{12}\right] = \frac{17}{6}.$$

$$\mu^2 = \frac{25}{9}$$

$$\sigma^2 = \frac{17}{6} - \frac{25}{9} = \frac{51-50}{18} = \frac{1}{18}.$$

$\underline{\underline{\text{Th}^m}}$ Let X be a random variable with probability distribution $f(x)$. The variance of the random variable $g(X)$ is

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

if X is discrete and

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

if X is continuous.

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:
$$\mu_{2X+3} = \sum_{x=0}^{3} (2x+3) f(x) = 3 \cdot \frac{1}{4} + 5 \cdot \frac{1}{8} + 7 \cdot \frac{1}{2} + 9 \cdot \frac{1}{8}$$
$$= \frac{6+5+28+9}{8} = \frac{48}{8} = 6.$$

$$\sigma^2_{2X+3} = E\{[(2X+3) - \mu_{2X+3}]^2\} = E[(2X+3-6)]^2$$
$$= E[(2X-3)^2] = E[4X^2 + 9 - 12X]$$
$$= 4E(X^2) + E(9) - 12E(X)$$

$$E(X) = \sum_{x=0}^{3} x f(x) = (0)\frac{1}{4} + \frac{1}{8} + 1 + \frac{3}{8} = \frac{0+1+8+3}{8} = \frac{12}{8} = \frac{3}{2}$$

$$E(X^2) = \sum_{x=0}^{3} x^2 f(x) = 0 + \frac{1}{8} + 2 + \frac{9}{8} = \frac{1+16+9}{8} = \frac{26}{8} = \frac{13}{4}.$$

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

Ex: Let X be a random variable having the density function given by
$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere.} \end{cases}$$

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

Sol: $\mu_{4X+3} = E(4X+3) = 4E(X) + E(3) = 4\left(\frac{15}{12}\right) + 3 = 8.$

$$E(X) = \int_{-1}^{2} x \cdot \frac{x^2}{3} dx = \frac{1}{3}\left[\frac{x^4}{4}\right]_{-1}^{2} = \frac{1}{12}[16-1] = \frac{15}{12}$$

$$\sigma^2_{4X+3} = E\left[[4X+3-8]^2\right] = E\{[4X-5]^2\}$$

$$= E\left[16X^2 + 25 - 40X\right]$$

$$= 16\,E(X^2) + E(25) - 40E(X)$$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-1}^{2} x \cdot \frac{x^2}{3}\,dx = \frac{1}{3}\left[\frac{x^4}{4}\right]_{-1}^{2} = \frac{1}{12}\left[16-1\right] = \frac{15}{12}.$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_{-1}^{2} x^2 \cdot \frac{x^2}{3}\,dx = \frac{1}{3}\left[\frac{x^5}{5}\right]_{-1}^{2} = \frac{1}{15}\left[32+1\right] = \frac{33}{15}.$$

$$\sigma^2_{4X+3} = 16\cdot\left(\frac{33}{15}\right) + 25 - 40\left(\frac{15}{12}\right) = \cancel{\frac{496+375-750}{15}} = \cancel{\frac{121}{15}}$$

$$= 16\cdot\frac{11}{5} + 25 - 50 = \frac{176+125-250}{5}$$

$$= \frac{51}{5}.$$

Covariance

A covariance refers to the measure of how two random variables will change when they are compared to each other.

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

The sign of the covariance indicates whether the relationship between two dependent random variables is positive or negative.

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

Covariance

The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

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

or

Let X and Y be random variables with joint probability distribution $f(x,y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x,y)$$

if X and Y are discrete, and

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f(x,y)\,dx\,dy$$

if X and Y are continuous.

1. Ex Random variables and probability distributions describes a situation involving the number of blue refills X and number of red refills Y. Two refills for a ball point pen are selected at random from a certain box, and the following is the joint probability distribution:

$f(x,y)$		$x=0$	$x=1$	$x=2$	$h(y)$
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
$g(x)$		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Find the Covariance of X and Y.

Sol:

$$\sigma_{XY} = \sum_{x=0}^{2}\sum_{y=0}^{2} (x-\mu_X)(y-\mu_Y) f(x,y)$$

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

$$= \frac{21}{28} = \frac{3}{4}.$$

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

$$= \frac{6+1}{14} = \frac{7}{14} = \frac{1}{2}.$$

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

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

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

$$= f(1,1) + 2f(1,2) + 2f(2,1) + 4f(2,2)$$

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

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

or

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

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

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

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

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

$$= \frac{3}{8}\cdot\frac{3}{28} - \frac{3}{8}\cdot\frac{3}{14} - \frac{9}{8}\,\frac{1}{28} - \frac{1}{8}\cdot\frac{9}{28} + \frac{1}{8}\cdot\frac{3}{14} + \frac{3}{8}(0)$$

$$-\frac{5}{8}\left(\frac{3}{28}\right) + 0 + 0$$

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

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Properties

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

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

2. 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

The probability that any random variable X will assume a value within k standard deviations of the mean is at least $1-\frac{1}{k^2}$.

That is, ~~$P(\mu-k\sigma)$~~ $P(\mu-k\sigma < X < \mu+k\sigma) \geq 1-\frac{1}{k^2}$.

~~$P(\mu-k\sigma)$~~

For $k=2$, $1-\frac{1}{k^2} = 1-\frac{1}{2^2} = \frac{3}{4}$.

The theorem states that three-fourths or more of the observations of any distribution lie in the interval $\mu \pm 2\sigma$.

For $k=3$, the theorem states that atleast eight-ninths of the observations of any distribution fall in the interval $\mu \pm 3\sigma$.

Ex A random variable X has a mean $\mu=8$, a variance $\sigma^2=9$, and an unknown probability distribution. Find

(a) $P(-4 < X < 20)$,

(b) $P(|X-8| \geq 6)$.

Sol: $P(-4 < X < 20) = P(8-4(3) < X < 8+4(3)) \geq 1-\frac{1}{16} = \frac{15}{16}$

$\Rightarrow P(-4 < X < 20) \geq \frac{15}{16}$.

(b) $P(|X-8| \geq 6) = 1 - P(|X-8| < 6)$

$= 1 - P(-6 < X-8 < 6)$

$= 1 - P(2 < X < 14)$

$= 1 - P(8-(2)(3) < X < 8+(2)(3))$

Now, $P(8-2(3) < X < 8+2(3)) \geq 1 - \frac{1}{4} = \frac{3}{4}$

$-P(8-2(3) < X < 8+2(3)) \leq -\frac{3}{4}$

$1 - P(2 < X < 14) \leq 1 - \frac{3}{4} = \frac{1}{4}$

$\therefore \; P(|X-8| \geq 6) \leq \frac{1}{4}.$

Ex A random variable X has a mean $\mu = 10$ and a 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 the constant C such that

$$P(|X-10| \geq C) \leq 0.04.$$

Sol: (a) $P(|X-10| \geq 3) = 1 - P(|X-10| < 3) = 1 - P(7 < X < 10)$

$= 1 - P\left[10 - \frac{3}{2}(2) < X < 10 + \frac{3}{2}(2)\right]$

$\leq 1 - \left(1 - \frac{1}{\frac{9}{4}}\right) = \frac{4}{9}.$

$\Rightarrow P(|X-10| \geq 3) \leq \frac{4}{9}.$

(b) $P(|X-10| < 3) = P(7 < X < 10) \geq 1 - \frac{4}{9} = \frac{5}{9}.$

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

$$P(5<X<15) = P\left[10-\frac{5}{2}(2) < X < 10+\frac{5}{2}(2)\right]$$

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

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

(d) $P(|X-10| \geq c) \leq 0.04$

$\Rightarrow 1-P(|X-10| < c) \leq 0.04$

$\Rightarrow P(|X-10| < c) \geq 0.96$

Solving. $0.96 = 1-\frac{1}{K^2} \Rightarrow \frac{1}{K^2} = 1-0.96 = 0.04 \Rightarrow K^2 = \frac{1}{0.04} = 25$

$\Rightarrow K = 5.$

$\therefore C = K\sigma = 10.$

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

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

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

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

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

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

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

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

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

$\sigma = 0.2236$

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

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

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

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

$$= 0.9837$$

By Chebyshev's th^m,

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