SCHOOL OF ENGINEERING

Department of Mathematics

Course Title: **PROBABILITY AND STATISTICS**

Module 2: **Random Variables and their Properties and Probability Distributions**

CONTINUOUS PROBABILITY DISTRIBUTIONS

Definitions:

- **Recall:** If a random variable takes uncountable number of possible values, then it is continuous. Here, we shall concern ourselves with computing probabilities for various intervals of continuous random variables.
- 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) $f(x) \geq 0$
  2) $\int_{-\infty}^{\infty} f(x)\, \mathrm{dx} = 1$
  3) $\mathrm{P}(a < X < b) = \int_a^b f(x)\ dx$

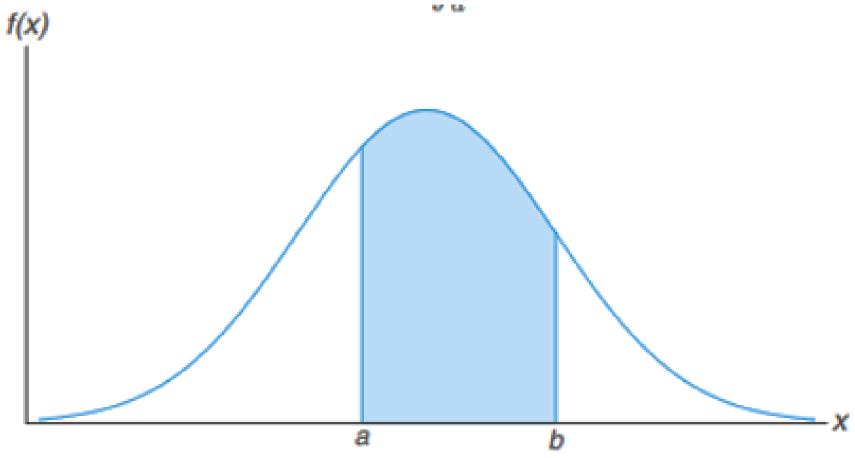
- Note : (i) When X is continuous, $P(X = a) = \int_a^a f(x)\ dx = 0$ (i.e.,) the probability that a continuous random variable will assume any fixed value is zero, and hence
  - $\mathrm{P}(\mathrm{a} < \mathrm{X} \leq b) = \mathrm{P}(\mathrm{a} < \mathrm{X} < \mathrm{b}) + \mathrm{P}(\mathrm{b}) = \mathrm{P}(\mathrm{a} < \mathrm{X} < \mathrm{b})$.

  (i.e.,) it does not matter whether we include an end point of the interval or not. This is not true, though, when X is discrete.

(ii) Areas will be used to represent probabilities and probabilities are positive numerical values, the density function must lie entirely above the x-axis.

(iii) The probability density function is constructed so that the area under its curve bounded by the x-axis is equal to 1 when computed over the range of X for which $f(x)$ is defined.

(iv) The probability that X assumes a value between a and b is equal to the shaded area under the density function between the ordinates at $x = a$ and $x = b$. (P$(a < X < b) = \int_a^b f(x)\ dx$)



$P(a < X < b)$.

(v) For a continuous random variable X and real number 'a',

- $P(x \geq a) = \int_a^\infty f(x)\, dx$
- $P(x < a) = 1 - P(x \geq a) = 1 - \int_a^\infty f(x)\ dx$

- **Definition:** For a continuous random variable X,

  $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\ dt$, for $-\infty < x < \infty$ is called the cumulative distribution function (c.d.f) of X.

  - <u>**Note:**</u> (i) $\frac{d}{dx}F(x) = f(x)$, if derivative exists (ii) P(a<X<b)=F(b)-F(a)

- If X is a continuous random variable having a probability density function f(x), the expectation, or the expected value of X, denoted by E[X], is defined by

  - E[X] = $\int_{-\infty}^{\infty} xf(x)\, dx \quad (Mean\ \mu)$

- If X is a random variable with mean μ, then the variance of X, denoted by Var(X), is defined by
  - Var(X) = $\sigma^2 = \mathrm{E}[X^2] - (E[X])^2 = \left(\int_{-\infty}^{\infty} x^2\ f(x)\, dx\right) - \mu^2$
  - Var(X) = $\int_{-\infty}^{\infty} (x-\mu)^2\ f(x) dx$
- Standard deviation of X, SD(X) = σ = $\sqrt{Var(X)}$

Question 1 Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} c(4x - 2x^2) & \text{if } 0 < x < 2 \text{ ,} \\ 0 & \text{otherwise} \end{cases}$$

a) What is the value of c? b) Find $P\{X > 1\}$.

Solution:

- (a) Since f is a probability density function, $\int_{-\infty}^{\infty} f(x)dx = 1$
- $\int_{-\infty}^{0} f(x)\ dx + \int_{0}^{2} f(x)\ dx + \int_{2}^{\infty} f(x)\ dx = \int_{-\infty}^{0} 0\ dx + \int_{0}^{2} c(4x - 2x^2)\ dx + \int_{2}^{\infty} 0\ dx = 1$
- $\int_{0}^{2} c(4x - 2x^2)\ dx$=1 ⇒ c$\left\{\left[4.\frac{x^2}{2}\right]_{x=0}^{x=2} - \left[2.\frac{x^3}{3}\right]_{x=0}^{x=2}\right\}$ = 1⇒ c$\left(\frac{8}{3}\right)$ = 1⇒ c = $\frac{3}{8}$
- $f(x) = \begin{cases} \frac{3}{8}(4x - 2x^2), & 0 < x < 2 \\ 0, & otherwise \end{cases}$
- (b) P(X>1) = $\int_{1}^{\infty} f(x)dx = \int_{1}^{2} \frac{3}{8}(4x - 2x^2)\ dx + \int_{2}^{\infty} 0\ dx = \frac{1}{2}$

Question 2: The total number of hours, **measured in units of 100 hours**, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

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

Find the probability that over a period of one year, a family runs their vacuum cleaner (a) less than 120 hours (b) between 50 and 100 hours.

Solution: Given that the total number of hours is measured in units of 100 hours (i.e.,) 1 unit = 100 hours

- (a) To find : P(vacuum cleaner runs less than 120 hours)=P($x < 1.2$) (120 hours=1.2 units)
  - $Formula: \ P(X < a) = \int_{-\infty}^{a} f(x)\ dx$
  - $P(X < 1.2) = \int_{0}^{1.2} f(x)\ dx = \int_{0}^{1} x\ dx + \int_{1}^{1.2} (2 - x)\ dx$
  - $= \left\{ \left[\frac{x^2}{2}\right]_{x=0}^{x=1} + \left[2x - \frac{x^2}{2}\right]_{x=1}^{x=1.2} \right\} = \left\{\frac{1}{2} + [1.68 - 1.5]\right\} = 0.68$
- (b) To find : P(vacuum cleaner runs between 50 and 100 hours)=P($0.5 < x < 1$) (50 hours=0.5 units and 100 hours=1unit)
  - $Formula:$ P$(a < X < b) = \int_{a}^{b} f(x)\ dx$
  - P$(a < X < b) = \int_{0.5}^{1} f(x)\ dx = \int_{0.5}^{1}\ x\ dx$
  - $= \left[\frac{x^2}{2}\right]_{x=0.5}^{x=1} = \left\{\frac{1}{2} - \frac{0.5^2}{2}\right\} = 0.375$

Question 3 A continuous random variable has the **Cumulative distribution function**

$$F(x) = \begin{cases} 0 & \text{if } x \leq 1\,, \\ c(x-1)^4 & 1 \leq x \leq 3 \\ 1 & x > 3 \end{cases}$$

Find c and also the probability density function.

Solution:

- By definition $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\ dt\,, -\infty < x < \infty, where\ f(t)\ is\ the\ p.d.f.$ Of X.
- Use: $\frac{d}{dx}F(x) = f(x), where\ f(x)\ is\ the\ p.d.f.$
- Differentiating F(x) with respect to x, $f(x) = \begin{cases} 0, & x \leq 1 \\ 4c(x-1)^3, & 1 \leq x \leq 3 \\ 0, & x > 3 \end{cases}$
- Since f is a probability density function, $\int_{-\infty}^{\infty} f(x)dx = 1$, $\int_{1}^{3} 4c(x-1)^3 dx = 1$
- $4c\left\{\left[\frac{(x-1)^4}{4}\right]_{x=1}^{x=3}\right\} = 1 \Rightarrow (16c) = 1 \Rightarrow$ c = $\frac{1}{16}$
- Then, the p.d.f is $f(x) = \begin{cases} 0, & x \leq 1 \\ \frac{1}{4}(x-1)^3, & 1 \leq x \leq 3 \\ 0, & x > 3 \end{cases}$

Question 4 Find k such that

$$f(x) = \begin{cases} kx^2 & \text{if } 0 < x < 3\,, \\ 0 & \text{otherwise} \end{cases}$$

is a probability density function. Also compute (i) $P\{1 < X < 2\}$ (ii) $P\{X \le 1\}$, (iii) $P\{X > 1\}$, (iv) mean and variance.

- Since f is a probability density function, $\int_{-\infty}^{\infty} f(x)dx = 1$
  - $\int_{-\infty}^{0} f(x)\ dx + \int_{0}^{3} f(x)\ dx + \int_{3}^{\infty} f(x)\ dx = \int_{-\infty}^{0} 0\ dx + \int_{0}^{3} kx^2\ dx + \int_{3}^{\infty} 0\ dx = 1$
  - $\int_{0}^{3} kx^2\ dx = 1 \Rightarrow \mathrm{k}\left\{\left[\frac{x^3}{3}\right]_{x=0}^{x=3}\right\} = 1 \Rightarrow 9\mathrm{c} = 1 \Rightarrow \mathrm{c} = \frac{1}{9}$
  - $f(x) = \begin{cases} \frac{1}{9}x^2, & 0 < x < 3 \\ 0, & otherwise \end{cases}$

- (i) P(1<X<2) = $\int_1^2 f(x)dx = \int_1^2 \frac{1}{9}x^2dx = \frac{1}{9}\left\{\left[\frac{x^3}{3}\right]_{x=1}^{x=2}\right\} = \frac{7}{27}$

- (ii) $\mathrm{P}(\mathrm{X} \leq 1) = \int_{-\infty}^{1} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_0^1 f(x)dx = \int_0^1 \frac{1}{9}x^2dx = = \frac{1}{9}\left\{\left[\frac{x^3}{3}\right]_{x=0}^{x=1}\right\} = \frac{1}{27}$

- P(X>1) = $\int_1^\infty f(x)dx = \int_1^3 \frac{1}{9}x^2dx + \int_3^\infty 0\ dx = \frac{1}{9}\left\{\left[\frac{x^3}{3}\right]_{x=1}^{x=3}\right\} = \frac{1}{27}\{27-1\} = \frac{26}{27}$

- Mean μ = $\int_{-\infty}^{\infty} xf(x)\ dx$ = $\int_{-\infty}^{0} 0\ dx + \int_0^3 x\left(\frac{1}{9}x^2\right)\ dx + \int_3^\infty 0\ dx$

  - $= \frac{1}{9}\int_0^3 (x^3)\ dx = \frac{1}{9}\left[\frac{x^4}{4}\right]_{x=0}^{x=3} \ = \ \frac{9}{4}$

- Variance σ² $= \left(\int_{-\infty}^{\infty} x^2\ f(x)\ dx\right) - \mu^2 = \left(\int_0^3 x^2\left(\frac{1}{9}x^2\right)\ dx\right) - \left(\frac{9}{4}\right)^2$

  - $= \frac{1}{9}\int_0^3 (x^4)\ dx - \left(\frac{81}{16}\right) = \frac{1}{9}\left\{\left[\frac{x^5}{5}\right]_{x=0}^{x=3}\right\} - \ \left(\frac{81}{16}\right) = \left(\frac{27}{80}\right)$

Question 5: A random variable X has the density function $f(x) = \frac{k}{1+x^2}$, $-\infty < x < \infty$, $where\ k\ is\ a\ constant$. Determine k and hence evaluate (i) $P(X \geq 0)$ (ii) $P(0 < X < 1)$.

Solution:

- Since f is a probability density function, $\int_{-\infty}^{\infty} f(x)dx = 1$
  - Since f(x) is an even function, $\int_{-\infty}^{\infty} f(x)\, dx = 2\int_{0}^{\infty} f(x)\ dx$
  - $2\int_{0}^{\infty} \frac{k}{1+x^2}\ dx = 1\ \Rightarrow 2k[\tan^{-1} x]_{x=0}^{x=\infty} = 1 \Rightarrow 2k\left(\frac{\pi}{2}\right) = 1 \Rightarrow k = \frac{1}{\pi}$
  - $f(x) = \begin{cases}\frac{1}{\pi(1+x^2)}, -\infty < x < \infty\end{cases}$
- P(X ≥ 0) = $\int_{0}^{\infty} f(x)dx = \int_{0}^{\infty} \frac{1}{\pi(1+x^2)} dx = \frac{1}{\pi}[\tan^{-1} x]_{x=0}^{x=\infty}$ = $\frac{1}{\pi}\left(\frac{\pi}{2}\right)$= $\frac{1}{2}$
- P(0<X<1) = $\int_{0}^{1} f(x)dx = \int_{0}^{1} \frac{1}{\pi(1+x^2)} dx = \frac{1}{\pi}[\tan^{-1} x]_{x=0}^{x=1}$ = $\frac{1}{\pi}\left(\frac{\pi}{4}\right)$= $\frac{1}{4}$

Question 6 : The lifetime (in hours) of a certain kind of radio tube is a random variable having a probability density function given by

$$f(x) = \begin{cases} 0, & x \leq 100 \\ \dfrac{100}{x^2}, & x > 100 \end{cases}$$

What is the probability that exactly 2 of 5 such tubes in a radio set will have to be replaced within the first 150 hours of operation? Assume that the events E_i , i = 1, 2, 3, 4, 5 that the i-th such tube will have to be replaced within this time are independent.

Solution:

- The probability that the radio tube will function for 150 hours : P(X<150) =$\int_0^{150} f(x)\, dx$
  - $\int_0^{100} 0\ dx + \int_{100}^{150} \frac{100}{x^2}$ = 100 $\int_0^{150} x^{-2} dx = 1/3$
- Let Y denotes the number of tubes to be replaced within 150 hours of operation. Then Y can be considered as a binomial random variable. Then, n=5, x=2 (two tubes to be replaced) , p= 1/3 , q=1-1/3=2/3
  - P(Y=2)=$^5C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = \frac{80}{243}$

Question 7 The amount of time, in hours, that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the probability that

(a) a computer will function between 50 and 150 hours before breaking down;

(b) it will function less than 100 hours?

Solution (a) Since

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \lambda \int_{0}^{\infty} e^{-x/100}\,dx$$

we obtain

$$1 = -\lambda(100)e^{-x/100}\Big|_0^{\infty} = 100\lambda \qquad \text{or} \qquad \lambda = \frac{1}{100}$$

Hence the probability that a computer will function between 50 and 150 hours before breaking down is given by

$$P\{50 < X < 150\} = \int_{50}^{150} \frac{1}{100} e^{-x/100}\, dx = -e^{-x/100}\Big|_{50}^{150}$$

$$= e^{-1/2} - e^{-3/2} \approx .384$$

(b) Similarly,

$$P\{X < 100\} = \int_{0}^{100} \frac{1}{100} e^{-x/100}\, dx = -e^{-x/100}\Big|_{0}^{100} = 1 - e^{-1} \approx .633$$

In other words, approximately 63.3 percent of the time a computer will fail before registering 100 hours of use. ∎

Question 8 The density function of X is given by

$$f(x) = \begin{cases} a + bx^2 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

If $E[X] = 3/5$, find a and b.

- Since f is a probability density function, $\int_{-\infty}^{\infty} f(x)dx = 1$
- $\int_0^1 (a + bx^2)\,dx = 1 \Rightarrow \left\{ a[x]_{x=0}^{x=1} + b\left[\frac{x^3}{3}\right]_{x=0}^{x=1} \right\} = 1 \Rightarrow a + \frac{1}{3}b = 1 \Rightarrow$ 3a+b=3 $-----$ (1)
- Given E[x]= $\mu = \int_{-\infty}^{\infty} x\, f(x)\,dx = \int_0^1 x\, f(x)\,dx = \int_0^1 (ax + bx^3)\,dx = \frac{3}{5}$
  - $\Rightarrow 4a + 2b = \frac{24}{5}$ ------- (2)
- Solving (1) and (2) , a=3/5 and b=6/5

Normal Distribution

Definitions:

- X is a normal random variable or X is normally distributed with parameters μ and σ^2 if the density of X is given by
  - $f(x) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$
- This density function $f(x)$ is a bell-shaped curve that is symmetric about μ (the graph is called the normal curve).
- Normal distribution is also known as Gaussian Distribution.

Note:

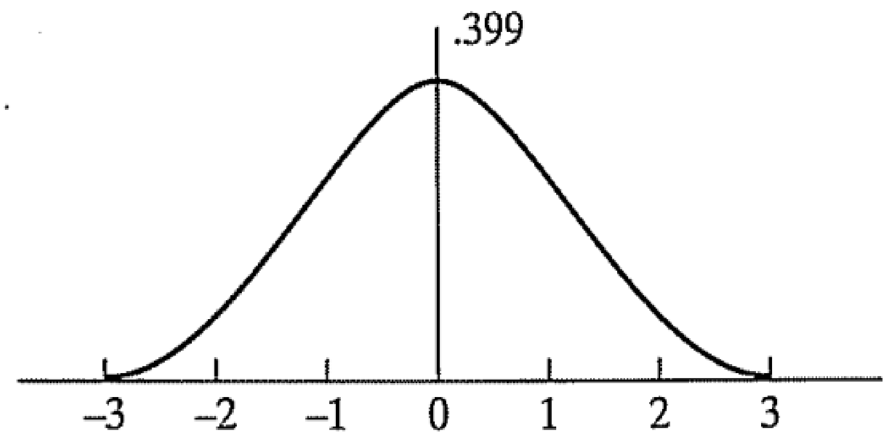
$\int_{-\infty}^{\infty} f(x)\, dx = \frac{1}{\sqrt{2\pi}\ \sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$ (For normal random variable)

- For the normal random variable X:
  - Mean= E(X) = μ, Variance= Var(X) = σ^2 , Standard Deviation= σ .

Normal density functions

(a) μ=0 and σ=1 : $f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$

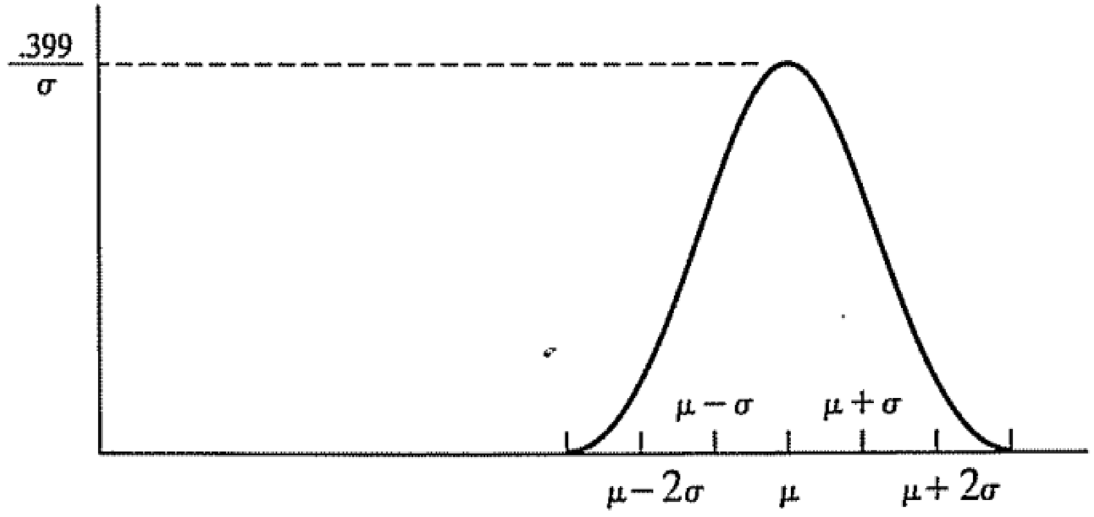
(b) arbitrary μ, σ^2 : $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



(a)

Note:

- The line $x = \mu$ divides the total area under the curve which is equal to 1 into two equal parts
- The area to the right as well as to the left of the line $x = \mu$ is 0.5



(b)

Standard normal distribution

- If X is normally distributed with parameters μ and σ^2, then $Z = \frac{X - \mu}{\sigma}$ is normally distributed with the parameter 0 and 1. Such a random variable is said to be a standard or a unit normal random variable.

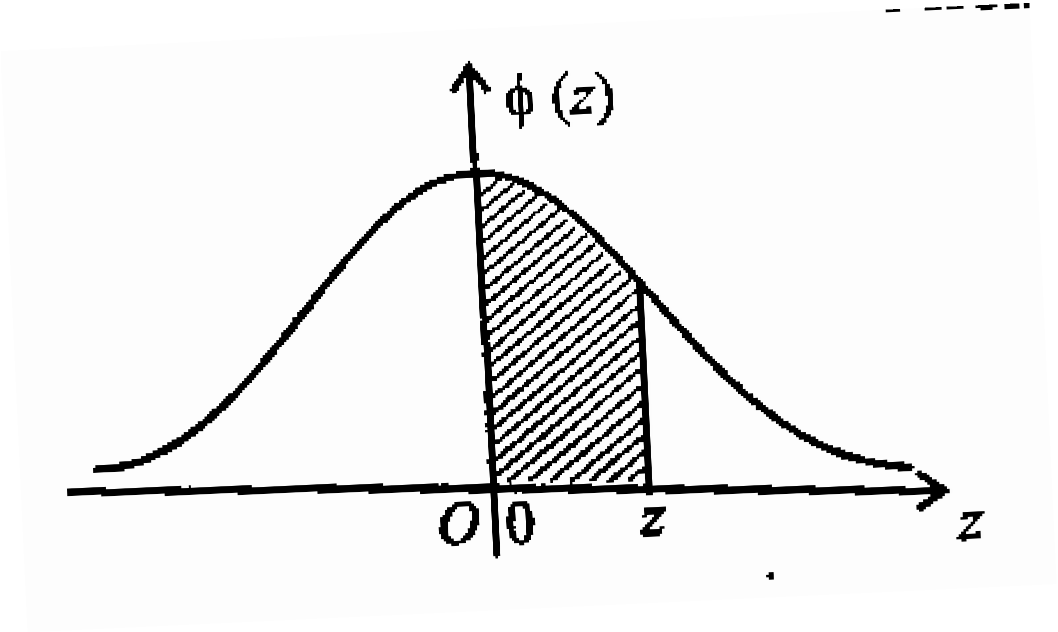
- If X is a normal variate, then $P(a \le X \le b) = \int_a^b f(x)\ dx = \frac{1}{\sqrt{2\pi}\ \sigma}\int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx\ ---- (1)$
  - Sub. $Z = \frac{X - \mu}{\sigma}$, and changing the limits to $z_1 = \frac{a - \mu}{\sigma}$ and $z_2 = \frac{b - \mu}{\sigma}$ in Equation (1)
  - $P(a \le X \le b) = P(z_1 \le Z \le z_2) = \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-\frac{z^2}{2}}\ dz$
  - Standard normal probability density function $F(z) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{z^2}{2}}$ is also called the standard normal curve which is symmetrical about the line z=0.

Note:

- $\int_{-\infty}^{\infty} f(z)\, dz = \frac{1}{\sqrt{2\pi}}\ \int_{-\infty}^{\infty} e^{-\frac{(z)^2}{2}} dz = 1$ *(For standard normal random variable)*
- $\int_{-\infty}^{0} f(z)\, dz = \frac{1}{\sqrt{2\pi}}\ \int_{-\infty}^{0} e^{-\frac{(z)^2}{2}} dz = \frac{1}{2}$ *and* $\int_{0}^{\infty} f(z)\, dz = \frac{1}{\sqrt{2\pi}}\ \int_{0}^{\infty} e^{-\frac{(z)^2}{2}} dz = 1/2$

- For the standard normal variable Z:
  - Mean= E[Z] = 0, Variance= Var(Z) = 1, Standard Deviation= 1.

- Define $\boldsymbol{\phi(z=a)=\frac{1}{\sqrt{2\pi}}\int_0^a e^{-\frac{z^2}{2}}\,dz}$ (This represents the area under the standard normal curve from Z=0 to a)



- The table which gives the area for different values of z is called normal probability table.

Normal probability table

the area under the standard normal curve from Z=0 to a

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Note:

- **For z_1 and $z_2 > 0$,**

(1) $P(-\infty \leq z \leq \infty) = 1$ (2) $P(-\infty \leq z \leq 0) = 1/2$

(3) $P(0 \leq z \leq \infty)$ or $P(z \geq 0) = 1/2$

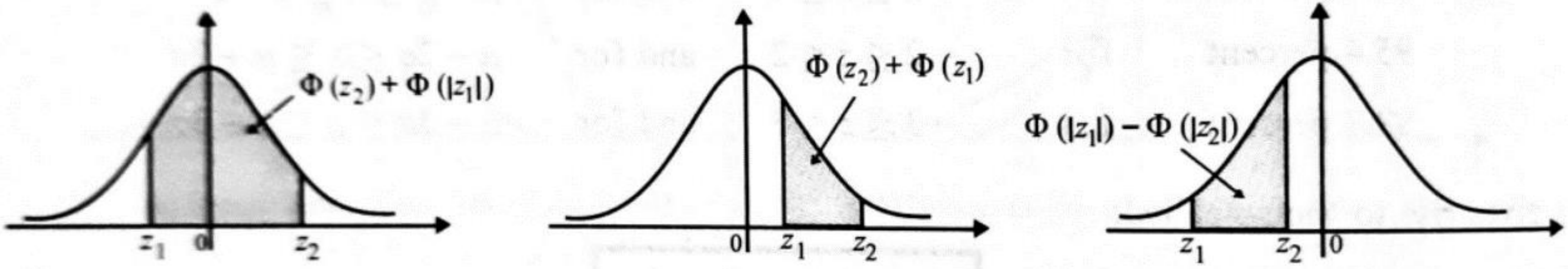
Also $P(-\infty < z < z_1) = P(-\infty < z \leq 0) + P(0 \leq z < z_1)$

i.e., $P(z < z_1) = 0.5 + \phi(z_1)$

Also $P(z > z_2) = P(z \geq 0) - P(0 \leq z < z_2)$

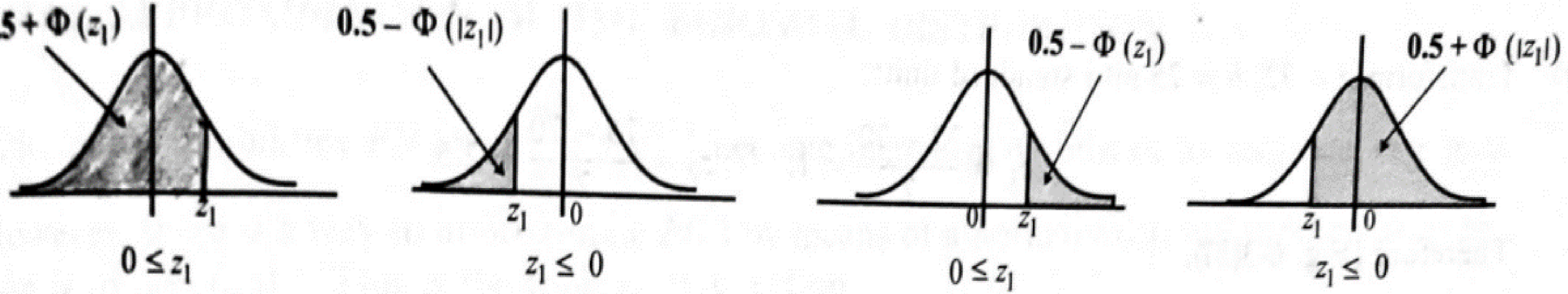
i.e., $P(z > z_2) = 0.5 - \phi(z_2)$

$$
\mathrm{P}(\mathrm{z}_1 \leq z \leq z_2) \begin{cases} \phi(z_2) + \phi(|z_1|) & \text{if } z_1 \leq 0 \leq z_2 \\ \phi(z_2) - \phi(z_1) & \text{if } 0 \leq z_1 \leq z_2 \\ \phi(|z_1|) - \phi(|z_2|) & \text{if } z_1 \leq z_2 \leq 0 \end{cases}
$$



(*a*) $z_1 \leq 0 \leq z_2$ (*b*) $0 \leq z_1 \leq z_2$ (*c*) $z_1 \leq z_2 \leq 0$

$$P(z \geq z_1) \begin{cases} 0.5 - \phi(z_1) & \text{if } z_1 \geq 0 \\ 0.5 + \phi(|z_1|) & \text{if } z_1 \leq 0 \end{cases}$$



(a) $P(Z \leq z_1)$ (b) $P(Z \geq z_1)$

Question 1: If X is a normal random variable with parameters μ = 3 and σ^2 = 9, find (a) P{2 < X < 5} (b) P{X > 0} (c) P{|X − 3| > 6}, (d) P{|X − 3| ≤ 6}.

Given: $\phi(1) = 0.3413, \phi(2) = 0.4772, \phi(0.67) = 0.2486, \phi(0.33) = 0.1293.$

Solution:

- Given: $\mu = 3, \sigma^2 = 9 \Rightarrow \sigma=3$.
- Standard normal variate : $z = \frac{x-\mu}{\sigma} = \frac{x-3}{3}$
- (a) P$(2 < x < 5)$:
  - When x =2, $z = \frac{2-3}{3} = -0.33$ and when x =5, $z = \frac{5-3}{3} = 0.66$
  - P$(2 < x < 5)$ = P(-0.33< z <0.67)=$P(-0.33 < z < 0) + P(0 < z < 0.67)$
  - $= P(0 < z < 0.33) + P(0 < z < 0.67)$ (By symmetry of standard normal curve)
  - =$\phi(0.33) + \phi(0.67)$ $\left(Recall: by\ defn.\ \phi(z = a) = \frac{1}{\sqrt{2\pi}} \int_0^a e^{-\frac{z^2}{2}} dz \right)$
  - $= 0.1293 + 0.2486 = 0.3779$
- (b) P$(x > 0)$:
  - When x =0, $z = -1$
  - P$(x > 0)$ = P$(z > -1)$=$P(-1 < z < 0) + P(z \geq 0)$
  - $= P(0 < z < 1) + 0.5$ (By symmetry of standard normal curve)
  - $\phi(1) + 0.5 = 0.3413 + 0.5 = 0.8413$

- (c) P{|X − 3| > 6}:
  - $|x-3| = \begin{cases} x-3, & x-3>0, \\ -(x-3), & x-3<0 \end{cases}$
  - $|x-3| > 6 \Rightarrow \begin{cases} x-3>6 \\ -(x-3)>6 \end{cases}$
  - $\Rightarrow \begin{cases} x>9 \\ -x>3 \end{cases} \Rightarrow \begin{cases} x>9 \\ x<-3 \end{cases}$
  - $P(|x-3|>6) = P(x<-3) + P(x>9)$

- When $x = -3, \mathrm{z} = \frac{-3\ -3}{3} = -2$
- P($x < -3$) = P($\mathrm{z} < -2$) = P($\mathrm{z} > 2$)(By symmetry)
- $= \mathrm{P}(\mathrm{z} \geq 0) - \mathrm{P}(0 < \mathrm{z} < 2) = 0.5 - \phi(2)$
- $= 0.5 - 0.4772 = 0.0228$

- When $x = 9, \mathrm{z} = \frac{9\ -3}{3} = 2$
- P($x > 9$) = P($\mathrm{z} > 2$) =P($\mathrm{z} \geq 0$) − P($0 < \mathrm{z} < 2$)
- $= 0.5 - \phi(2)$
- $= 0.5 - 0.4772 = 0.0228$

- $P(|x-3|>6) = P(x<-3) + P(x>9)$ = 0.0228+0.0228=0.0456

(d) P{|X − 3| ≤ 6}= 1-P{|X − 3| > 6} = 1-0.0456=0.9544

Question 2: The marks of 1000 students in an examination follows a normal distribution with mean 70 and standard deviation 5. Find the number of students whose mark will be (i) less than 65, (ii) more than 75, (iii) between 65 and 75. Given: $\phi(1) = 0.3413$

Solution:

- Let x represents the marks of students.
- Given: $\mu = 70$, σ=5.
- Standard normal variate : $Z = \frac{x - \mu}{\sigma} = \frac{x - 70}{5}$
- (i) To find P(x <65) : When x =65, z = $\frac{65 - 70}{5} = -1$
- $P(x < 65) = P(Z<-1) = P(Z\leq 0)-P(-1<Z<0)$
  $= 0.5-P(0<Z<1) = 0.5-\phi(1)$
  $= 0.5 - 0.3413 = 0.1587$

$$\left(by\ defn.\ \phi(z = a) = \frac{1}{\sqrt{2\pi}} \int_0^a e^{-\frac{z^2}{2}}\, dz \right)$$

(OR)

$$P(z < -1) = P(z > 1)$$
$$= P(z \geq 0) - P(0 < z < 1)$$
$$= 0.5 - \phi(1) = 0.5 - 0.3413 = 0.1587$$

- **Number of students whose marks will be less than 65 = 1000 * 0.1587 = 158.7 ≈ 159**

- (ii) To find P(x >75)
  - When x =75, z = $\frac{75 - 70}{5} = 1$
  - P(x >75)=P(z >1)=P(1< z <∞)
  - =P(z ≥0)-P(0< z <1)=0.5-P(0< z <1)=0.5-$\phi(1) = 0.5 - 0.3413 = 0.1587$
  - **Number of students whose marks will be more than 75 = 1000 * 0.1587 = 158.7 ≈ 159**

- (iii) To find P($65 < x < 75$)
  - When x =65, $z = -1$, x =75, $z = 1$
  - P($65 < x < 75$) = P(-1< z <1)=2*P(0< z <1) = 2*$\phi(1) = 2(0.3413) = 0.6826$
  - **Number of students scoring marks between 65 and 75 = 1000 * 0.6826 = 682.6 ≈ 683**

Question 3: A lawyer commutes daily from his suburban home to his midtown office. The average time for a one-way trip is 24 minutes, with a standard deviation of 3.8 minutes. Assume the distribution of trip times to be normally distributed.

(i) What is the probability that a trip will take at least ½ hour?

(ii) Find the probability that 2 of next 3 trips will take at least ½ hour?

(iii) If the office opens at 9:00AM and the lawyer leaves his house at 8:45 AM daily, what percentage of the time is he late for work?

(iv) If he leaves the house at 8:35 AM and coffee is served at the office from 8:50AM until 9AM, what is the probability that he misses coffee?

Given: $\mathrm{P}(0 < \mathrm{Z} < 1.58) = \phi(1.58) = 0.4429, \mathrm{P}(0 < \mathrm{Z} < 2.37) = \phi(2.37) = 0.4911, \mathrm{P}(0 < \mathrm{Z} < 2.37) = \phi(0.26) = 0.1026.$

Solution: Let X represents the trip time.

- Given: $\mu = 24$, σ=3.8, Standard normal variate : $\mathrm{Z} = \frac{x-\mu}{\sigma} = \frac{x-24}{3.8}$
- (i) P(trip takes at least ½ hour)=P(X ≥30) :
  - When X =30, Z = $\frac{30-24}{3.8} = 1.58$
  - P(X >30)=P(Z >1.58)= P(Z ≥0)−P(0< Z <1.58)

$$=0.5 - \phi(1.58) \quad \left(by\ defn.\ \phi(z=a) = \frac{1}{\sqrt{2\pi}} \int_0^a e^{-\frac{z^2}{2}}\, dz \right)$$

$$= 0.5 - 0.4429 = 0.0571$$

(ii) P(2 of next 3 trips will take at least ½ hour)

- Let Y denotes the number of trips that will take at least ½ hour. Then Y can be considered as a binomial random variable.
- Then, n=3, X=2 (two trips will take at least ½ hour) , p= 0.0571 , q=1-0.0571 = 0.9429
  - P(Y=2)=${}^3C_2$ $(0.0571)^2\ (0.9429)^1 = 0.0092.$

(iii) what percentage of the time is he late for work ?

- P(He is late for work)=$P(travel\ time\ exceeds\ 15\ minutes) = P(X > 15)$
- When X =15, Z = $\frac{15-24}{3.8} = -2.37\ (rounded\ to\ two\ digits)$
- P(X >15)=P($\mathrm{Z} > -2.37$) = P($-2.37 < \mathrm{Z}$ <0)+P($\mathrm{Z} > 0$)

  =$\mathrm{P}(0 < \mathrm{Z} < 2.37) + 0.5$

  =$0.5 + \phi(2.37) = 0.5 + 0.4911 = 0.9911$

  - **Conclusion: 99.11 % of the time he is late for work.**

(iv) Probability that he misses coffee:

- P(he misses coffee)=$P(travel\ time\ exceeds\ 25\ minutes) = P(X > 25)$
- When X =25, Z = $\frac{25-24}{3.8} = 0.26\ (rounded\ to\ two\ digits)$
- P(X >25)=P($\mathrm{Z} > 0.26$) = P(Z ≥0)−P(0< Z <0.26)=$0.5 - \phi(0.26) = 0.5 - 0.1026 = 0.3974$

Question 4: An electrical firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 2040 hours and standard deviation of 60 hours. In a test on 2000 bulbs, Estimate the number of bulbs likely to last for (i) more than 2150 hours (ii) less than 1950 hours (iii) more than 1920 hours but less than 2160 hours.

Given: $\phi(1.83) = 0.4664, \phi(1.5) = 0.4332, \phi(2) = 0.4772.$

Solution:

- Let x represents the lifetime of the bulb.
- Given: $\mu = 2040$, σ=60.
- Standard normal variate : $z = \frac{x - \mu}{\sigma} = \frac{x - 2040}{60}$
- (i) To find P($x > 2150$)
  - When x =2150, $z = \frac{2150 - 2040}{60} = \frac{11}{6} = 1.83$
  - P($x > 2150$)=P($z > 1.83$)=P($z \geq 0$)-P(0< z <1.83)
  - = $0.5 - \phi(1.83)$ $\left(By\ defn.\ \phi(z = a) = \frac{1}{\sqrt{2\pi}} \int_0^a e^{-\frac{z^2}{2}}\, dz \right)$
  - $= 0.5 - 0.4664 = 0.0336$
  - **Number of bulbs to last more than 2150 hours = 2000 * 0.0336 = 67.2 ≈ 67**

- (ii) To find P(x <1950) : When x =1950, z = $\frac{1950-2040}{60} = -1.5$
  - $\mathrm{P}(x < 1950) = \mathrm{P}(\mathrm{z} < -1.5) = \mathrm{P}(\mathrm{z} > 1.5) = \mathrm{P}(\mathrm{z} \geq 0) - \mathrm{P}(0 < \mathrm{z} < 1.5) = 0.5 - \phi(1.5)$
    $= 0.5 - 0.4332 = 0.0668$
  - **Number of bulbs to last less than 1950 hours = 2000 * 0.0668 = 133.6 ≈ 134**

- (iii) To find P($1920 < x < 2160$)
  - When x =1920, $\mathrm{z} = -2$, x =2160, $\mathrm{z} = 2$
  - P($1920 < x < 2160$) = P(-2< z <2)=2*P(0< z <2) = 2*$\phi(2) = 2(0.4772) = 0.9544$
  - **Number of bulbs to last more than 1920 hours but less than 2160 hours = 2000 * 0.9544 = 1908.8 ≈ 1909**

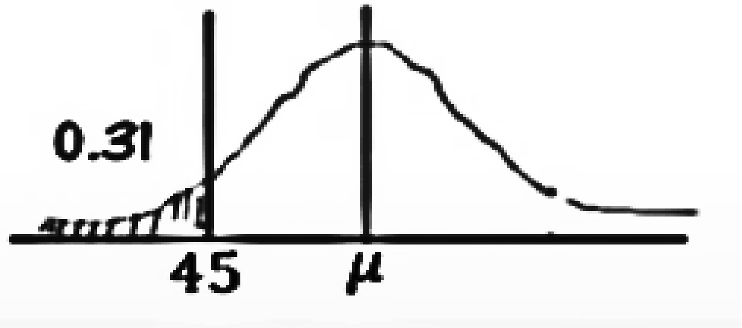
Question 5: In a normal distribution, 31% of the item are under the value 45 and 8% of the item are over the value 64. Find the mean and variance of the distribution. Given that $P(z < -1.4) = 0.08$ and $P(z > 0.5) = 0.31$

Solution:

- Let X be the normal random variable with mean μ and standard deviation σ.
- Given: $P(x < 45) = 0.31$ and $P(x > 64) = 0.08$

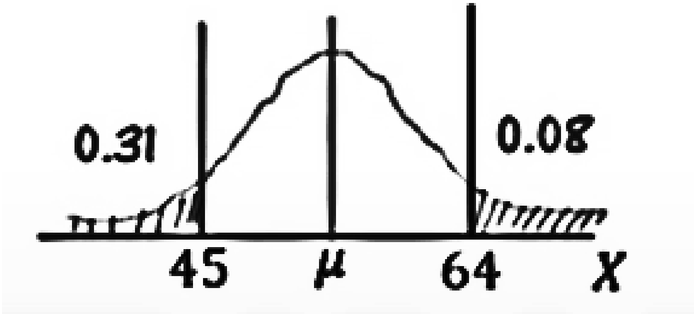
- There are two possible cases : $45 > \mu$ (or) $45 < \mu$
- If $45 > \mu$,
- $P(x < 45) = P(x \le \mu) + P(\mu < x < 45) = 0.31$

  $(\textit{which is not possible because } P(x \le \mu) = 0.5)$
- Therefore, $45 < \mu$ (or) $\mu > 45$

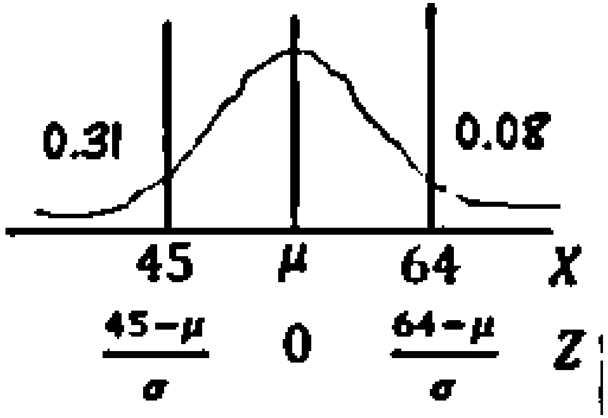


- There are two possible cases : $45 < 64 < \mu$ (or) $45 < \mu < 64$.
- If $45 < 64 < \mu$, $P(x > 64) = P(64 < x < \mu) + P(x \ge \mu) = 0.08$

  (which is not possible because $P(x \ge \mu) = 0.5$)
- Therefore, $45 < \mu < 64$



- Standard normal variate : $z = \frac{x-\mu}{\sigma}$
- When $x = 45$, $z = \frac{x-\mu}{\sigma} = \frac{45-\mu}{\sigma}$ and when $x = 64, z = \frac{64-\mu}{\sigma}$
- $P(x < 45) = P\left(z < \frac{45-\mu}{\sigma}\right) = 0.31$ ----- (1) and
- $P(x > 64) = P\left(z > \frac{64-\mu}{\sigma}\right) = 0.08$ ------ (2)



- **Given :** $P(z < -1.4) = 0.08 \Rightarrow P(z > 1.4) = 0.08$ ------ (3) and
  - $P(z > 0.5) = 0.31 \Rightarrow P(z < -0.5) = 0.31$-------- (4)

- Comparing (2) and (3) , $\frac{64-\mu}{\sigma} = 1.4 \Rightarrow \mu + 1.4\sigma = 64$ ---- (5)

- Comparing (1) and (4) , $\frac{45-\mu}{\sigma} = -0.5 \Rightarrow \mu - 0.5\sigma = 45$---- (6)

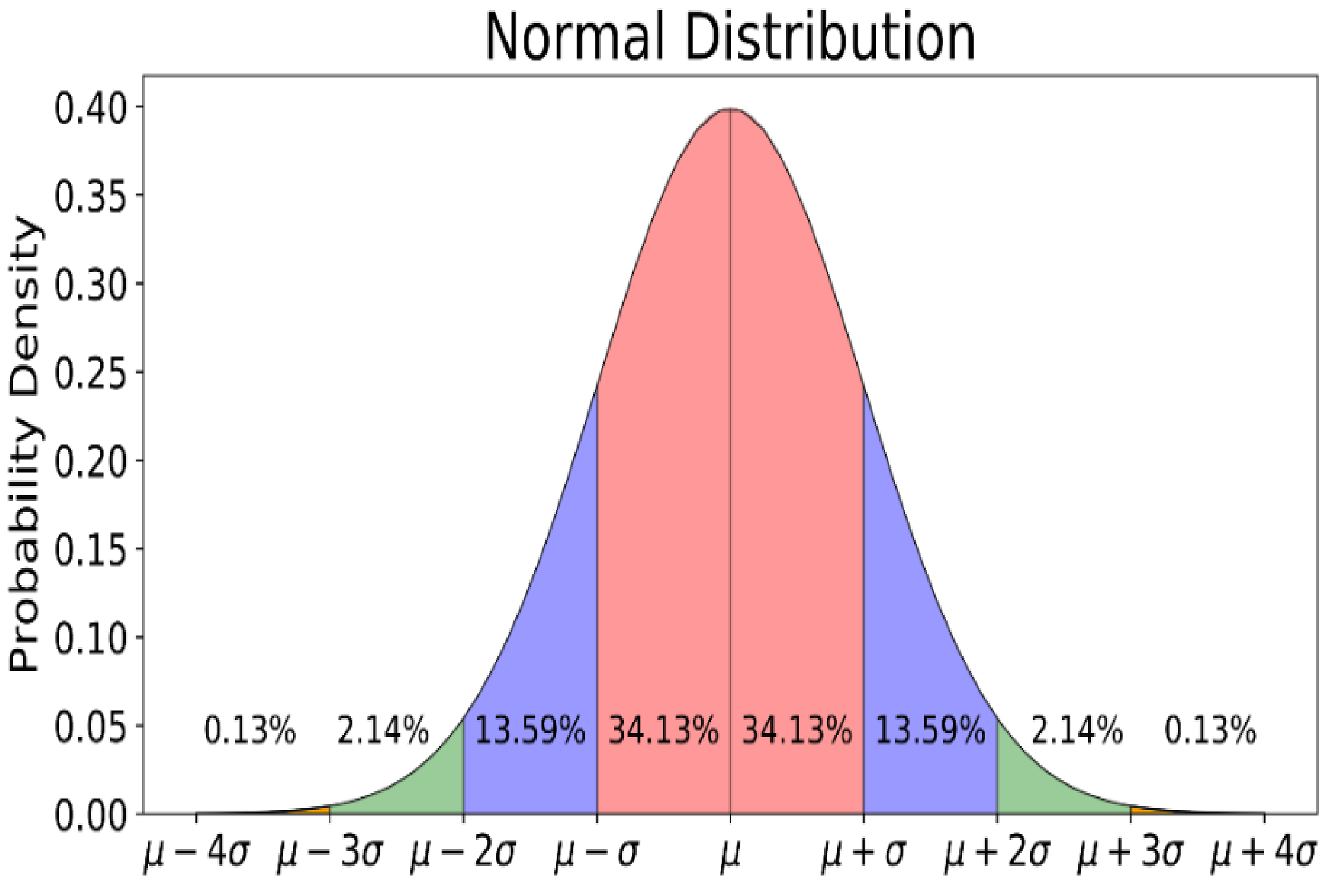
- Solving (5) and (6) , we get Mean $\mu=50$, Standard deviation $\sigma = 10$.

Question 6: If $f(x) = c\, e^{-\frac{x^2-6x+4}{24}}$ is the probability density function of a normal variate, then find c, mean and variance.

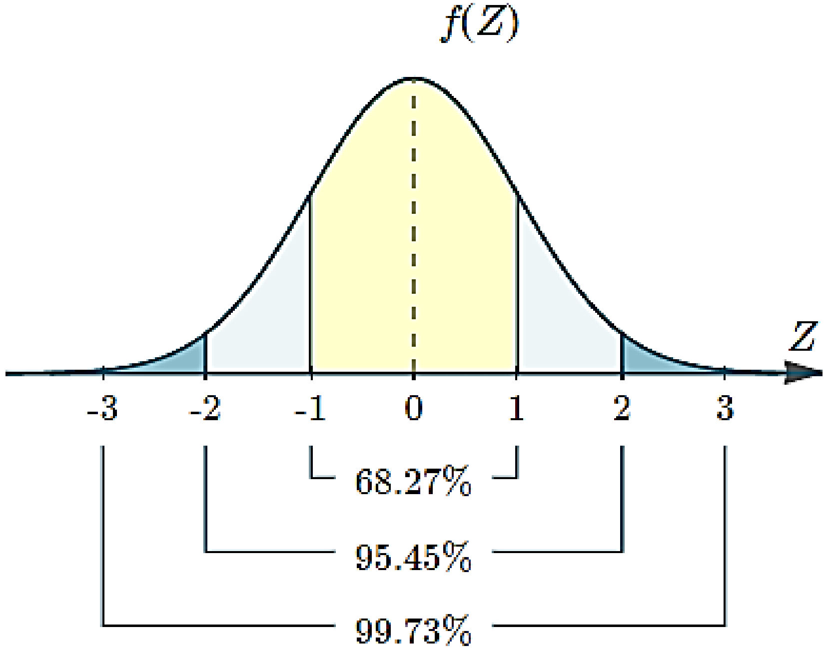
Solution:

- The probability density function of a normal variable X is given by $f(x) = \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ ------ (1)
- Given $f(x) = c\, e^{-\frac{x^2-6x+4}{24}}$
- Rewriting given $f(x)$:
  - $c\, e^{-\frac{1}{24}(x^2-(2)(3)(x)+9-9+4)} = c\, e^{-\frac{1}{24}((x-3)^2-5)} = c\, e^{5/24}\, e^{-\frac{(x-3)^2}{24}}$
  - $= c\, e^{5/24} e^{-\frac{(x-3)^2}{2\,(\sqrt{12})^2}}$ $-----(2)$
- Comparing the exponents in (1) and (2),
  - Mean μ = 3 , Standard deviation σ = $\sqrt{12}$ provided the constant c = $\frac{e^{-\frac{5}{24}}}{\sqrt{24\pi}}$

Percentages of the Area Under the Normal Curve



Percentages of the Area Under the Standard Normal Curve



Standard Normal Curve showing percentages $\mu = 0$, $\sigma = 1$

This means that 68.27% of the scores lie within 1 standard deviation of the mean.

This comes from: $\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = 0.68269$

Also, 95.45% of the scores lie within 2 standard deviations of the mean.

This comes from: $\int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = 0.95450$

Finally, 99.73% of the scores lie within 3 standard deviations of the mean.

This comes from: $\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = 0.9973$

The total area from $-\infty < z < \infty$ is 1.

Exponential Distribution

- The exponential distribution often arises, in practice, as being the distribution of the amount of time until some specific event occurs.

  Eg: Time between arrivals at a congested intersection during rush hour in a large city.

  Amount of time until a phone call you receive turns out to be a wrong number.

  Amount of time before a certain type of component in a system fails.(time to failure).

- **Definition:** A continuous random variable whose **probability density function** is given by
  - $f(x) = \begin{cases} \alpha e^{-\alpha x}, & x \geq 0 \\ 0, & otherwise \end{cases}$ $\quad for\ some\ \alpha > 0,$ is said to be exponential random

  variable with parameter α.

- The cumulative distribution function F(a) of an exponential random variable is given by
  - $F(a) = P(X \leq a) = 1 - e^{-\alpha a}, a \geq 0.$
  - $\mathrm{P(X > a)} = 1 - P(X \leq a) = 1 - (1 - e^{-\alpha a}) = e^{-\alpha a}, a \geq 0.$
  - $P(a < X < b) = F(b) - F(a) = \left(1 - e^{-\alpha b}\right) - (1 - e^{-\alpha a}) = e^{-\alpha a} - e^{-\alpha b}$

Note:

- $\int_{-\infty}^{\infty} f(x)\, dx = \int_{\mathbf{0}}^{\infty} \alpha e^{-\alpha x}\, dx = 1$
- Mean (μ) : E(X) = $\frac{1}{\alpha}$ (Mean of the exponential is the reciprocal of its parameter α),
- Variance (σ^2) : Var (X) = $\frac{1}{\alpha^2}$, Standard Deviation (σ) = $\frac{1}{\alpha}$.

Question 1: If a random variable X follows exponential distribution with **mean 5**, find (i) P{0 < X < 1} (ii) P{X < 10} (iii) P{X ≤ 0 or X ≥ 1}

Solution:

- Given Mean = $\frac{1}{\alpha} = 5 \;\Rightarrow \alpha = \frac{1}{5}$.
- $f(x) = \frac{1}{5}e^{-\frac{x}{5}},\;\; x \geq 0,$ is the probability density function of exponential random variable X.
- (i) P(0<X<1) $=\int_0^1 f(x)\,dx = \frac{1}{5}\int_0^1 e^{-\frac{x}{5}}\,dx = 1-\left(e^{-\frac{1}{5}}\right) = 0.1813$ (OR)
- P(0<X<1) = F(1)-F(0) = $e^{-\alpha(0)} - e^{-\alpha(1)}$ $\left(\because P(a < X < b) = F(b) - F(a)\right)$

$$= 1 - e^{-\frac{1}{5}} = 0.1813$$

- (ii) P(X<10) $=\int_{-\infty}^{10} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_0^{10} f(x)\,dx = \frac{1}{5}\int_0^{10} e^{-\frac{x}{5}}\,dx = 1-\left(\frac{1}{e^2}\right) = 0.8647$ (OR)
- P(X<10) = P(X ≤10)=F(10) = $1 - e^{-\alpha(10)}$ $(\because F(a) = P(X \leq a) = 1 - e^{-\alpha a}, a \geq 0.)$

$$= 1 - e^{-\frac{10}{5}} = 1 - e^{-2} = 0.8647$$

- (iii) P{X ≤ 0 or X ≥ 1} = P(X ≤ 0) + P(X ≥ 1} = 0 + $\int_1^{\infty} f(x)\,dx = \frac{1}{5}\int_1^{\infty} e^{-\frac{x}{5}}\,dx = e^{-\frac{1}{5}} = 0.8187$ (OR)
- P(X ≤ 0) + P(X ≥ 1} = 0 + (1-P(X<1))=1-F(1)=1-($1 - e^{-\alpha(1)}$)= $e^{-\frac{1}{5}}$=0.8187

Question 2: Suppose that a system contains a certain type of component whose time to failure(in years) is given by T. The random variable T follows an exponential distribution with **mean time to failure 5**. If 5 of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years?

Solution:

- Let T denote the time to failure. As it follows exponential distribution,
  - $f(t) = \alpha e^{-\alpha t}, \quad t \geq 0$
- Given Mean = $\frac{1}{\alpha} = 5 \Rightarrow \alpha = \frac{1}{5}$, $\therefore f(t) = \frac{1}{5} e^{-\frac{t}{5}}, \quad t \geq 0$ is the probability density function of T.

- P(a given component is still functioning after 8 years) = P(time to failure > 8)
  - P(T>8) = $\int_8^{\infty} f(t)\, dt = \frac{1}{5}\int_8^{\infty} e^{-\frac{t}{5}}\, dt = e^{-\frac{8}{5}} = 0.2019$

- Let Y denotes the number of components functioning after 8 years. Then Y can be considered as a binomial random variable. Then, n=5, x≥2 (at least two) , p= 0.2019, q=1- 0.2019 =0.7981
  - P(Y ≥ 2)=$1 - \{P(Y = 0) + P(Y = 1)\}$

$$= 1 - \{{}^5C_0\, (0.2019)^0\, (0.7981)^5 + {}^5C_1\, (0.2019)^1\, (0.7981)^4\} = 1 - 0.7333 = 0.2667$$

Question 3: Based on extensive testing, it is determined that the time X, in years, before a major repair is required for a certain washing machine is characterized by the density function

$$f(x) = \begin{cases} \frac{1}{4} e^{-\frac{x}{4}}, & x \geq 0 \\ 0, & otherwise. \end{cases}.$$

The machine is considered a good purchase if it is unlikely to require a major repair before the sixth year. Conclude whether or not the washing machine is a good purchase.

Solution: Let X be the time before a major repair is required.

- Given X is exponentially distributed with mean α = 1/4

- The machine is a good purchase if the probability that it will require major repair after sixth year is more than the probability that it will require a repair before six years.
- P(it requires major repair after sixth year) = P(X>6) $= \frac{1}{4} \int_6^{\infty} e^{-\frac{x}{4}} dx = e^{-\frac{3}{2}} = 0.2231.$

- P(it requires major repair before sixth year)= P(X<6)=1-0.2231=0.7777

- Conclusion: The machine is not really a good purchase.

Question 4: The length of a phone call (in minutes) arriving at a particular center follows an exponential distribution with an average of 5 minutes. Find the probability that a random call made to this center (i) ends less than 5 minutes (ii) ends between 5 and 10 minutes.

Solution:

- Let X denote the length of the call. As it follows exponential distribution,
  - $f(x) = \alpha e^{-\alpha x}, \quad x \geq 0$

- Given Mean = $\frac{1}{\alpha} = 5 \Rightarrow \alpha = \frac{1}{5}$, $\therefore f(x) = \frac{1}{5} e^{-\frac{x}{5}}, \quad x \geq 0$ is the probability density function of X.

- (i) P(x<5) = $\int_0^5 f(x)\, dx = \frac{1}{5} \int_0^5 e^{-\frac{x}{5}}\, dx = 1 - \left(\frac{1}{e}\right) = 0.6321$

- (ii) P(5<x<10) = $\int_5^{10} f(x)\, dx = \frac{1}{5} \int_5^{10} e^{-\frac{x}{5}}\, dx = \frac{1}{e} - \left(\frac{1}{e^2}\right) = 0.2325$

JOINT PROABABILITY DISTRIBUTIONS

Definitions: **(Discrete Case)**

- If X and Y are two discrete random variables, we define the joint probability function (or joint probability mass function) of X and Y by P(X=x,Y=y)=$p(x,y)$ where $p(x,y)$ satisfy the conditions

  $p(x,y) \geq 0$ and $\sum_x \sum_y p(x,y) = 1$, *the summation is taken over all the values of x and y.*

  (i.e.,) that values $p(x,y)$ give the probability that outcomes x *and* y occur at the same time.
- Suppose X={$x_1, x_2, \ldots, x_m$} , Y={$y_1, y_2, \ldots, y_n$} , then $P(X = x_i, Y = y_j) = p(x_i, y_j)$ *denoted by* J_{ij}.

- The set of values of the function $p(x_i, y_j) = J_{ij}$, $i = 1,2, \ldots m,\ j = 1,2, \ldots n$ is called the joint probability distribution of X and Y. These values are presented in the form of a two way table called the joint probability table.

X \ Y	y_1	y_2	…	y_n	Sum
x_1	J_{11}	J_{12}	…	J_{1n}	$f(x_1)$
x_2	J_{21}	J_{22}	…	J_{2n}	$f(x_2)$
…	..	..	..	..	..
x_m	J_{m1}	J_{m2}	…	J_{mn}	$f(x_m)$
Sum	$g(y_1)$	$g(y_2)$	…	$g(y_n)$	1

- **Note:** The function p is defined on the set $X \times Y = \{(x_1, y_1), (x_1, y_2), \ldots (x_m, y_n)\}$ (cartesian product of the sets X and Y).

- In the joint probability table, $f(x_1), f(x_2), \dots f(x_m)$ respectively represents the sum of all the entries in the first row, second row,…, m^{th} row and $g(y_1), g(y_2), \dots g(y_n)$ respectively represents the sum of all the entries in the first column, second column,…, n^{th} column (i.e.,)

$$f(x_1) = J_{11} + J_{12} + \cdots + J_{1n} \; ; \; g(y_1) = J_{11} + J_{21} + \cdots + J_{m1}$$

$$f(x_2) = J_{21} + J_{22} + \cdots + J_{2n} \; ; \; g(y_2) = J_{12} + J_{22} + \cdots + J_{m2}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots \; ; \; \cdots\cdots\cdots\cdots\cdots\cdots$$

$$f(x_m) = J_{m1} + J_{m2} + \cdots + J_{mn} \; ; \; g(y_n) = J_{1n} + J_{2n} + \cdots + J_{mn}$$

- $\{f(x_1), f(x_2), \dots f(x_m)\}$ and $\{g(y_1), g(y_2), \dots g(y_n)\}$ are called marginal probability distributions of X alone and Y alone respectively.
- **Note:**
  - $f(x_1) + f(x_2) + \cdots + f(x_m) = 1 \quad and g(y_1) + g(y_2) + \cdots + g(y_n) = 1$
  - In other words, $\sum_{i=1}^{m} \sum_{j=1}^{n} p(x_i, y_j) = \sum_{i=1}^{m} \sum_{j=1}^{n} J_{ij} = 1$ (i.e.,) total of all entries in the joint probability table is equal to 1.

- The discrete random variables X and Y are said to be independent random variables if

$$P(X = x, Y = y) = P(X = x).P(Y = y) \;\; and \;\; conversely.$$

- $P\left(X = x_i, Y = y_j\right) = P(X = x_i).P(Y = y_j) \Rightarrow f\left(x_i\right).g\left(y_j\right) = J_{ij}$ in the joint probability table (i.e.,) X and Y are independent if each entry J_{ij} in the table is equal to the product of its marginal entries. Otherwise, X and Y are said to be dependent.

- If X and Y are two discrete random variables having the joint probability function $p(x, y)$ then the Expectations of X and Y are defined as
  - $\mu_X = E(X) = \sum_x \sum_y x\, p(x, y) = \sum_i x_i f(x_i)$
  - $\mu_Y = E(Y) = \sum_x \sum_y y\, p(x, y) = \sum_i y_i g(y_i)$
  - $E(XY) = \sum_{i,j} x_i\, y_j\, J_{ij}$
  - If $Z = \phi(X, Y)$ and $p(x, y)$ is the joint distribution of X and Y, the Expectation of Z in the joint distribution of X,Y is defined as $\mathrm{E(Z)} = \sum_{i,j} \phi(x_i\,, y_j)\, J_{ij}$

- If X and Y are two discrete random variables having mean μ_X and μ_Y respectively, then the covariance of X and Y denoted by $cov(X,Y)$ is defined as
  - $cov(X,Y) = \sum_i \sum_j (x_i - \mu_X)(y_j - \mu_Y)\, J_{ij} = E[(X - \mu_X)(Y - \mu_Y)]$
  - $cov(X,Y) = \sum_i \sum_j x_i y_j\, J_{ij} - \mu_X \mu_Y = E(XY) - \mu_X \mu_Y$

- Correlation of X and Y : $\rho(x,y) = \frac{cov(X,Y)}{\sigma_X\, \sigma_Y}$, $where$ σ_X and σ_Y denotes standard deviation of X and

Y respectively (σ_X^2=$\mathrm{E}(X^2) - \mu_X^2$ and σ_Y^2=$\mathrm{E}(Y^2) - \mu_Y^2$).

- <u>Note:</u>

1) If X and Y are independent random variables, then
   - $E(XY) = E(X)E(Y)$
   - $cov(X,Y) = 0$ and $\rho(x,y)$ =0
   - $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$

2) $cov(X,X) = E[(X - \mu_X)^2] = V(X) = \sigma_X^2$

Question 1: The joint distribution of two random variables X and Y is as follows.

X \ Y	-4	2	7
1	1/8	1/4	1/8
5	1/4	1/8	1/8

Compute the following:

(a) E(X) and E(Y) (b) E(XY) (c) σ_X and σ_Y (d) $cov(X,Y)$ (e) $\rho(X,Y)$

Solution:

- Given: $x_1 = 1, x_2 = 5, y_1 = -4, y_2 = 2, y_3 = 7$, $J_{11}=\frac{1}{8}, J_{12}=\frac{1}{4}, J_{13}=\frac{1}{8}, J_{21}=\frac{1}{4}, J_{22}=\frac{1}{8}, J_{23}=\frac{1}{8}$
- **The marginal distributions of X and Y :**
  - Sum of entries in each row : $f(x_1) = J_{11}+ J_{12}+J_{13}$=1/2 and $f(x_2) = J_{21}+ J_{22}+J_{23}$=1/2
  - Sum of entries in each column : $g(x_1) = J_{11}+ J_{21}$=3/8, $g(x_2) = J_{12}+ J_{22}$=3/8, $g(x_3) = J_{13}+ J_{23}$= 1/4

Distribution of X:

x_i	1	5
$f(x_i)$	1/2	1/2

Distribution of Y:

y_j	-4	2	7
$g(y_j)$	3/8	3/8	1/4

- (a) $E(X) = \mu_X = \sum_{i=1}^{2} x_i f(x_i) = 1\left(\frac{1}{2}\right) + 5\left(\frac{1}{2}\right) = 3$ and $E(Y) = \mu_Y = \sum_{j=1}^{3} y_j g(y_j) = (-4)\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 7\left(\frac{1}{4}\right) = 1$
- (b) $E(XY) = \sum_{i,j} x_i\, y_j\, J_{ij} = \sum_{i=1}^{2}\sum_{j=1}^{3} x_i\, y_j\, J_{ij}$

  $= (1)(-4)\left(\frac{1}{8}\right) + (1)(2)\left(\frac{1}{4}\right) + (1)(7)\left(\frac{1}{8}\right) + (5)(-4)\left(\frac{1}{4}\right) + (5)(2)\left(\frac{1}{8}\right) + (5)(7)\left(\frac{1}{8}\right) = 3/2$
- (c) $\sigma_X^2 = \mathrm{E}(X^2) - \mu_X^2$

  $\mathrm{E}(X^2) = \sum_{i=1}^{2} x_i^2 f(x_i) = 1\left(\frac{1}{2}\right) + 25\left(\frac{1}{2}\right) = 13$ and $\mu_X^2 = 3^2 = 9$

  $\sigma_X^2 = 13 - 9 = 4$
- $\sigma_Y^2 = \mathrm{E}(Y^2) - \mu_Y^2$

  $\mathrm{E}(Y^2) = \sum_{j=1}^{3} y_j^2 g(y_j) = 16\left(\frac{3}{8}\right) + 4\left(\frac{3}{8}\right) + 49\left(\frac{1}{4}\right) = \frac{79}{4}$ and $\mu_Y^2 = 1$

  $\sigma_Y^2 = \frac{79}{4} - 1 = \frac{75}{4}$

- (d) $cov(X,Y) = E(XY) - \mu_X\, \mu_Y = \left(\frac{3}{2}\right) - (3)(1) = -\frac{3}{2}$
- (e) $\rho(x,y) = \frac{cov(X,Y)}{\sigma_X\, \sigma_Y} = \frac{-3/2}{(\sqrt{4})(\sqrt{75/4})} = -0.1732$

Question 2: A fair coin is tossed thrice. The random variables X and Y are defined as follows:

X= 0 (or) 1 according as head or tail occurs on the first toss.

Y= Number of heads.

a) Determine the marginal distribution of X and Y
b) Determine the joint distribution of X and Y
c) Obtain the expectations of X,Y and XY. Also find standard deviations of X and Y
d) Compute the covariance and correlation of X and Y

Solution: The sample space S and the values of random variables X and Y are as follows:

S	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
X	0	0	0	0	1	1	1	1
Y	3	2	2	1	2	1	1	0

a) The probability distributions of X and Y:

- X={0,1} , Y={0,1,2,3}
- $P(X=0)=\frac{4}{8}=\frac{1}{2}, P(X=1)=\frac{4}{8}=\frac{1}{2},$
- $P(Y=0)=\frac{1}{8}, P(Y=1)=\frac{3}{8}, P(Y=2)=\frac{3}{8}, P(Y=3)=\frac{1}{8}$

Distribution of X:

x_i	0	1
$f(x_i)$	1/2	1/2

Distribution of Y:

y_j	0	1	2	3
$g(y_j)$	1/8	3/8	3/8	1/8

(b) The joint distribution of X and Y is found by computing

$J_{ij} = P(X = x_i, Y = y_j)$ where we have

$x_1 = 0, x_2 = 1$ and $y_1 = 0, y_2 = 1, y_3 = 2, y_4 = 3$

$J_{11} = P(X = 0, Y = 0) = 0$

($X = 0$ *implies that there is a head turn out and* Y *the total number heads* 0 *is impossible)*

$J_{12} = P(X = 0, Y = 1) = 1/8$ corresponding to the outcome HTT

$J_{13} = P(X = 0, Y = 2) = 2/8 = 1/4$; out comes are HHT and HTH

$J_{14} = P(X = 0, Y = 3) = 1/8$; outcome is HHH

$J_{21} = P(X = 1, Y = 0) = 1/8$; outcome is TTT

$J_{22} = P(X = 1, Y = 1) = 2/8 = 1/4$; out comes are THT, TTH

$J_{23} = P(X = 1, Y = 2) = 1/8$; outcome is THH

$J_{24} = P(X = 1, Y = 3) = 0$; since the outcome is impossible.

X \ Y	$y_1 = 0$	$y_2 = 1$	$y_3 = 2$	$y_4 = 3$	$f(x_i)$
$x_1 = 0$	0	1/8	1/4	1/8	1/2
$x_2 = 1$	1/8	1/4	1/8	0	1/2
$g(y_j)$	1/8	3/8	3/8	1/8	1

c) Expectation of X, Y, XY and standard deviations of X and Y:

- $E(X) = \mu_X = \sum_{i=1}^{2} x_i f(x_i) = 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = \frac{1}{2}$
- $E(Y) = \mu_Y = \sum_{j=1}^{4} y_j g(y_j) = (0)\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = \frac{12}{8} = \frac{3}{2}$
- $E(XY) = \sum_{i,j} x_i\, y_j\, J_{ij} = \sum_{i=1}^{2} \sum_{j=1}^{4} x_i\, y_j\, J_{ij} = 0 + \frac{1}{4} + 2\left(\frac{1}{8}\right) = \frac{1}{2}$
- $\sigma_X^2 = \mathrm{E}(X^2) - \mu_X^2 = \sum_{i=1}^{2} x_i^2 f(x_i) - \mu_X^2 = 0 + 1\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$
- $\sigma_Y^2 = \mathrm{E}(Y^2) - \mu_Y^2 = \sum_{j=1}^{4} y_j^2 g(y_j) - \mu_Y^2 = 0 + 1\left(\frac{3}{8}\right) + 4\left(\frac{3}{8}\right) + 9\left(\frac{1}{8}\right) - \left(\frac{3}{2}\right)^2 = \frac{3}{4}$

(d) $cov(X,Y) = E(XY) - \mu_X\, \mu_Y = \frac{1}{2} - \frac{3}{4} = -\frac{1}{4}$

$$\rho(x,y) = \frac{cov(X,Y)}{\sigma_X\, \sigma_Y} = \frac{-1/4}{(\sqrt{3})/4} = -\frac{1}{\sqrt{3}}$$

Question 3: Suppose X and Y are independent random variables with the following respective distribution, find the joint distribution of X and Y. Also verify that $cov(X,Y) = 0$.

x_i	1	2
$f(x_i)$	0.7	0.3

y_j	-2	5	8
$g(y_j)$	0.3	0.5	0.2

Solution:

- Since X and Y are independent, $f(x_i).\,g(y_j) = J_{ij}\ (i = 1,2\ and\ j = 1,2,3)$ (i.e.,) J_{ij} is obtained by multiplying the marginal entries.

- $J_{11} = f(x_1)g(y_1) = (0.7)(0.3) = 0.21,\quad J_{12} = f(x_1)g(y_2) = (0.7)(0.5) = 0.35,$

$J_{13} = f(x_1)g(y_3) = (0.7)(0.2) = 0.14,\quad J_{21} = f(x_2)g(y_1) = (0.3)(0.3) = 0.09,$

$J_{22} = f(x_2)g(y_2) = (0.3)(0.5) = 0.15,\quad J_{23} = f(x_2)g(y_3) = (0.3)(0.2) = 0.06.$

Y \ X	$y_1 = -2$	$y_2 = 5$	$y_3 = 8$	$f(x_i)$
$x_1 = 1$	0.21	0.35	0.14	0.7
$x_2 = 2$	0.09	0.15	0.06	0.3
$g(y_j)$	0.3	0.5	0.2	1

- $cov(X,Y) = E(XY) - \mu_X \, \mu_Y$
  - $E(X) = \mu_X = \sum_{i=1}^{2} x_i f(x_i) = 1(0.7) + 2(0.3) = 1.3$
  - $E(Y) = \mu_Y = \sum_{j=1}^{3} y_j g(y_j) = (-2)(0.3) + 5(0.5) + 8(0.2) = 3.5$
  - $E(XY) = \sum_{i,j} x_i \, y_j \, J_{ij} = \sum_{i=1}^{2} \sum_{j=1}^{3} x_i \, y_j \, J_{ij}$

$= (1)(-2)(0.21) + (1)(5)(0.35) + (1)(8)(0.14) + (2)(-2)(0.09) + (2)(5)(0.15) + (2)(8)(0.06) = 4.55$

  - $cov(X,Y) = E(XY) - \mu_X \, \mu_Y = 4.55 - (1.3)(3.5) = 0$

Question 4: Let X and Y are independent random variables. X take values 2,5,7 with probability 1/2, 1/4, 1/4 respectively. Y take the values 3,4,5 with the probability 1/3, 1/3, 1/3.

(i) Find the joint probability distribution of X and Y.

(ii) Show that $cov(X, Y) = 0$.

Solution: (i)

- **Given data:**

x_i	**2**	**5**	**7**
$f(x_i)$	**1/2**	**1/4**	**1/4**

y_j	**3**	**4**	**5**
$g(y_j)$	**1/3**	**1/3**	**1/3**

- Since X and Y are independent, $f(x_i).g(y_j) = J_{ij}\ (i = 1,2,3\ and\ j = 1,2,3)$ (i.e.,) J_{ij} is obtained by multiplying the marginal entries.

X \ Y	$y_1 = 3$	$y_2 = 4$	$y_3 = 5$	$f(x_i)$
$x_1 = 2$	1/6	1/6	1/6	1/2
$x_2 = 5$	1/12	1/12	1/12	1/4
$x_3 = 7$	1/12	1/12	1/12	1/4
$g(y_j)$	1/3	1/3	1/3	1

- $(ii)\ cov(X,Y) = E(XY) - \mu_X\ \mu_Y$
  - $E(X) = \mu_X = \sum_{i=1}^{3} x_i f(x_i) = 2\left(\frac{1}{2}\right) + 5\left(\frac{1}{4}\right) + 7\left(\frac{1}{4}\right) = 4$
  - $E(Y) = \mu_Y = \sum_{j=1}^{3} y_j g(y_j) = (3)\left(\frac{1}{3}\right) + 4\left(\frac{1}{3}\right) + 5\left(\frac{1}{3}\right) = 4$
  - $E(XY) = \sum_{i,j} x_i\, y_j\, J_{ij} = \sum_{i=1}^{3}\sum_{j=1}^{3} x_i\, y_j\, J_{ij}$

$$= (2)(3)\left(\frac{1}{6}\right) + (2)(4)\left(\frac{1}{6}\right) + (2)(5)\left(\frac{1}{6}\right) + (5)(3)\left(\frac{1}{12}\right) + (5)(4)\left(\frac{1}{12}\right) + (5)(5)\left(\frac{1}{12}\right) + (7)(3)\left(\frac{1}{12}\right) + (7)(4)\left(\frac{1}{12}\right) + (7)(5)\left(\frac{1}{12}\right) = 16$$

  - $cov(X,Y) = E(XY) - \mu_X\ \mu_Y 16 - (4)(4) = 0$