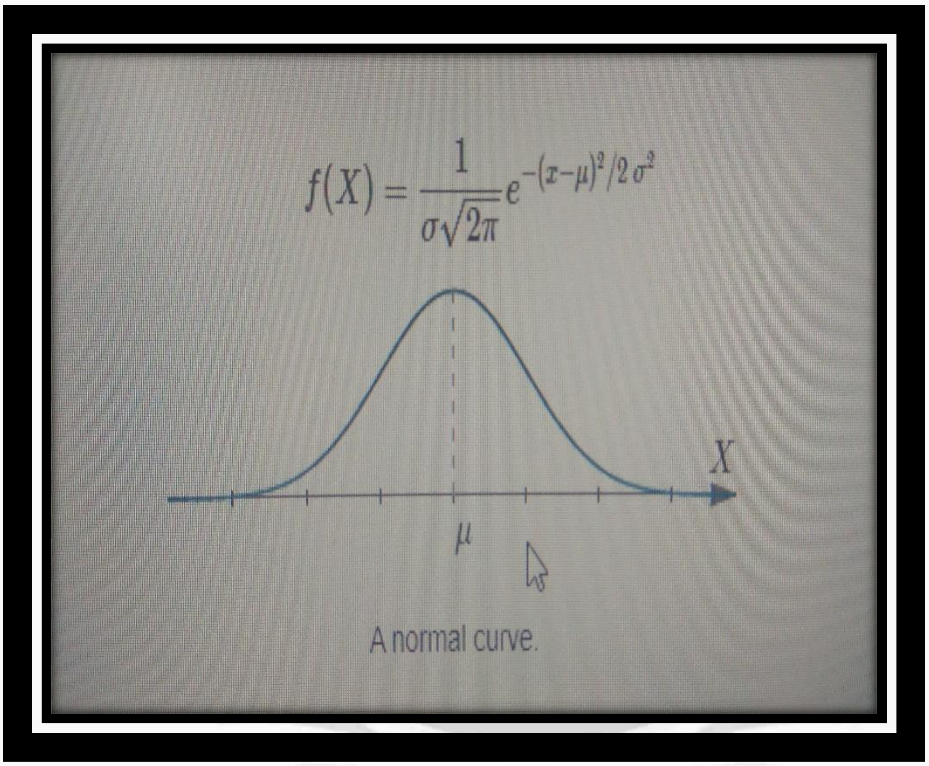
NORMAL DISTRIBUTION

The Normal Probability Distribution is very common in the field of statistics. Whenever you measure things like people's height, weight, salary, opinions or votes, the graph of the results is very often a normal curve.



Properties of a Normal Distribution:

(i) The normal curve is symmetrical about the mean

(ii) The mean is at the middle and divides the area into halves.

(iii) The total area under the curve is equal to 1.

(iv) It is completely determined by its mean and standard deviation σ (or variance σ^2).

Note:

In Normal distribution only two parameters are needed, namely μ and σ^2

Area under the Normal Curve using Integration:

The Probability of a continuous normal variable X found in a particular interval $[a, b]$ is the area under the curve bounded by $x = a$ and $x = b$ is given by $P(a < X < b) = \int_a^b f(X)dx$

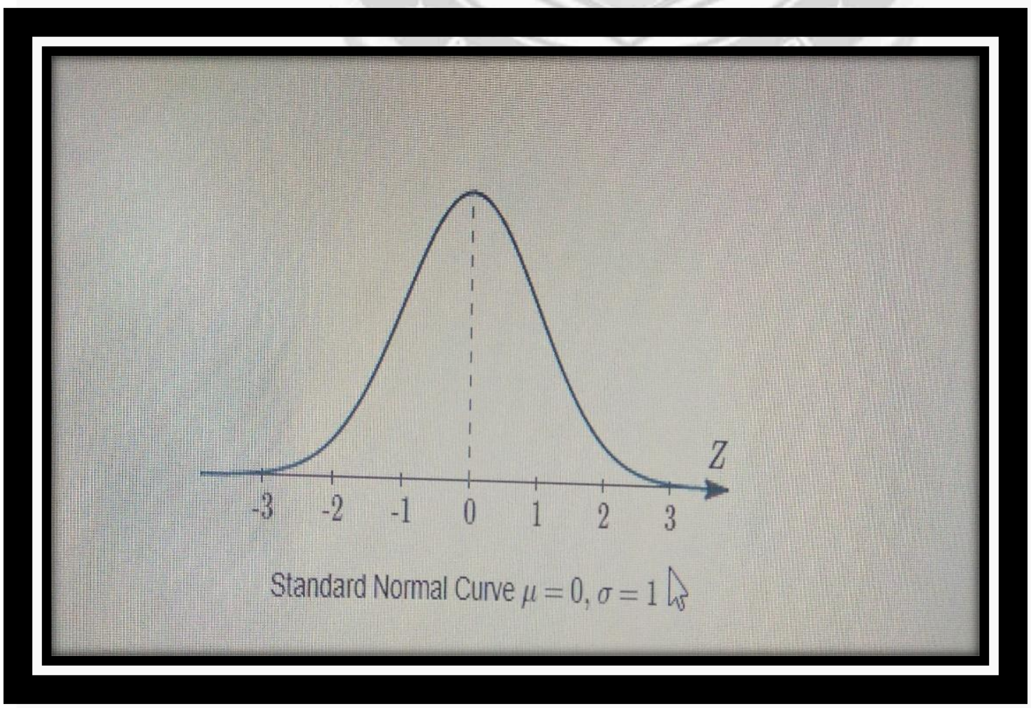
and the area depends upon the values μ and σ.

The standard Normal Distribution:

We standardize our normal curve, with a mean of zero and a standard deviation of 1 unit.

If we have the standardized situation of $\mu = 0$ and $\sigma = 1$ then we have

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



We can transform all the observations of any normal random variable X with mean μ and variance σ to a new set of observations of another normal random variable Z with mean 0 and variance 1 using the following transformation:

$$Z = \frac{X - \mu}{\sigma}$$

The two graphs have different μ and σ, but have the same area.

The new distribution of the normal random variable z with mean 0 and variance 1 (or standard deviation 1) is called a Standard normal distribution.

Formula for the Standardized normal Distribution

If we have mean μ and standard deviation σ , then

$$Z = \frac{X - \mu}{\sigma}$$

Find the moment generating function of Normal distribution

Sol: We first find the M.G.F of the standard normal distribution and hence find mean and variance.

$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} ; -\infty < z < \infty$

where $z = \frac{x-\mu}{\sigma}$

$M_z(t) = E[e^{tz}]$

$= \int_{-\infty}^{\infty} e^{tz} \phi(z) dz$

$$= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left(\frac{2tz - z^2}{2}\right)} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{z^2 - 2tz}{2}\right)} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{(z-t)^2 - t^2}{2}\right)} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{(z-t)^2}{2}\right) + \frac{t^2}{2}} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{(z-t)^2}{2}\right)} e^{\frac{t^2}{2}} dz$$

$$= \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{z-t}{\sqrt{2}}\right)^2} dz = \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-v^2} \sqrt{2} dv$$

$$= \frac{e^{\frac{t^2}{2}}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-v^2} dv = \frac{e^{\frac{t^2}{2}}}{\sqrt{\pi}} \sqrt{\pi}$$

$$M_z(t) = e^{\frac{t^2}{2}} \ldots \ldots \ldots (1)$$

$$M_X(t) = M_{\mu+\sigma z}(t); \sum z = \frac{X-\mu}{\sigma}, X = \mu + \sigma z$$

$$= e^{\mu t} M_z(\sigma t)$$

$$= e^{\mu t} \cdot e^{\frac{\sigma^2 t^2}{2}} \text{ From } (1)$$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

To find mean and variance:

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

$$= 1 + \frac{\mu t + \frac{\sigma^2 t^2}{2}}{1!} + \frac{\left(\mu t + \frac{\sigma^2 t^2}{2}\right)^2}{2!} + \cdots$$

$$= 1 + \frac{\mu t + \frac{\sigma^2 t^2}{2}}{1!} + \frac{\left(\mu^2 t^2 + \frac{\sigma^4 t^4}{4} + 2\mu t \frac{\sigma^2 t^2}{2}\right)}{2!} + \cdots$$

Coefficient of $t = \mu$ Coefficient of $t^2 = \frac{\sigma^2}{2} + \frac{\mu^2}{2}$

$$E(X) = 1! \times \text{ coefficient of } t$$

$$= \mu$$

$$E(X^2) = 2! \times \text{ coefficient of } t^2$$

$$= 2\left(\frac{\sigma^2}{2} + \frac{\mu^2}{2}\right) \Rightarrow 2\left(\frac{e^2 + u^2}{2}\right)$$

$$= \mu^2 + \sigma^2$$

$$\text{variance} = E(X^2) - [E(X)]^2$$

$$= \mu^2 + \sigma^2 - \mu^2$$

$$= \sigma^2$$

Problems based on Normal distribution

1. **X is normally distributed with mean 12 and SD is 4. Find the probability that (i) $\boldsymbol{X \geq 20}$ $\boldsymbol{(ii) X \leq 20}$ $\boldsymbol{(iii) 0 \leq X \leq 12}$.**

Solution:

Given X follows normally distribution with $\boldsymbol{\mu = 12, \sigma = 4}$

$$P(X \geq 20) = P\left(\frac{X-\mu}{\sigma} \geq \frac{20-\mu}{\sigma}\right) = P\left(Z \geq \frac{20-12}{4}\right)$$

$= P(Z \geq 2)$

$= 0.5 - P[0 < Z < 2]$

$= 0.5 - 0.4772$ (from the table)

$P(X \geq 20) = 0.0228$

(i) $P(X \leq 20) = 1 - P[\,X > 20] = 1 - 0.0228$

$P(X \leq 20) = 0.9772$

(ii) $P[0 \leq X \leq 12] = P\left(\frac{0-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \geq \frac{12-\mu}{\sigma}\right)$

$$= P\left(\frac{0-12}{4} \leq Z \geq \frac{12-12}{4}\right)$$

$= P(-3 \leq Z \leq 0)$

$= P(0 \leq Z \leq 3)$ (since the curve is symmetrical)

$P[0 \leq X \leq 12] = 0.4987$ (from the table)

2. In a normal distribution 31% of the items are under 45 and 8 % are over 64. Find the mean and the standard deviation.

Solution:

Let the mean and standard deviation of the given normal distribution be μ and .

The area lying to the left of the ordinate at x = 45 is 0.31. The corresponding value of z is negative.

The area lying to the right of the ordinates up to the mean is 0.5 – 0.31 = 0.19

The value of z corresponding to the area 0.19 is 0.5 nearly.

$$\therefore \frac{45-\mu}{\sigma} = -0.5$$

(or) $-0.5\sigma + \mu = 45$ ………………………(1)

Area to the left of the ordinate at x = 64 is 0.5 – 0.08 = 0.42 and hence the value of z corresponding to this area is 1.4 nearly.

$$\therefore \frac{64-\mu}{\sigma} = 1.4$$

(or) $1.4\sigma + \mu = 64$ ……………………..(2)

Solving (1) and (2) we get

$$-0.5\sigma + \mu = 45$$

$$1.4\sigma + \mu = 64$$

(1) – (2) $\quad - 1.9\sigma = -19$

$$\Rightarrow \sigma = 10$$

Substituting $\sigma = 10$ in (1) we get

$$-0.5\,(10) + \mu = 45$$

$$-5 + \mu = 45$$

$$\Rightarrow \mu = 50$$

3. **The weekly wages of 1000 workmen are normally distributed around a mean of Rs. 70 with a S.D of Rs. 5. Estimate the number of workers whose weekly wages will be (i) between Rs.69 and Rs.72 (ii) less than Rs.69 (iii) more than Rs.72**

Solution :

Let X be the RV denoting the weekly wages of a worker

Given $\mu = 70, \sigma = 5$

The normal variate z = $\frac{X-\mu}{\sigma} = \frac{X-70}{5}$

(i) P(69 < X < 72)

When X = 69, z = $\frac{69-70}{5} = -0.2$

When X = 72, z = $\frac{72-70}{5} = 0.4$

$$\therefore P(69 < X < 72\ ) = P(\ -0.2 < z < 0.4)$$

$$= P(-0.2 < z < 0) + P(0 < z < 0.4)$$

$$= P(0 < z < 0.2) + P(0 < z < 0.4)$$

= 0.0793 + 0.1554 (from table)

= 0.2347

Out of 1000 work men , the number of workers whose wages lies between Rs. 69 and Rs.72

= 1000 x P(69 < X < 72)

= 1000 x 0.2347 = 235

(ii) P(less than 69) = P (X < 69)

When x = 69, $z = \frac{X-\mu}{\sigma} = \frac{69-70}{5}$ = -0.2

$\therefore P(X < 69) = P(z < -0.2)$

$$= 0.5 - P(0 < z < 0.2)$$

= 0.4207

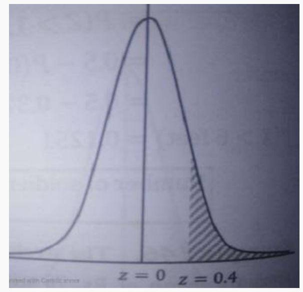
Out of 1000 workmen , the number of workers whose wages are less than Rs. 69

= 1000 x P(z < -0.2)

= 1000 x 0.4207

= 420.7

(iii) P(more than Rs. 72) = P(X > 72)



When x = 72, $z = \frac{X-\mu}{\sigma} = \frac{72-70}{5}$ = -0.2

$$\therefore P(X < 69) = P(z < -0.2)$$

$$= 0.5 - P(0 < z < 0.2)$$

= 0.4207

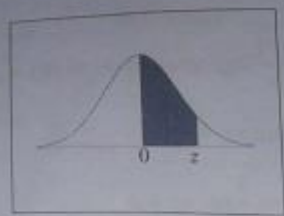
Out of 1000 workmen , the number of workers whose wages are less than Rs. 69

= 1000 x P(z < -0.2)

= 1000 x 0.4207

= 420.7

Standard Normal Distribution Table



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998

Uniform Distribution (Rectangular Distribution)

A random variable X is said to have a continuous uniform distribution if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, a < x < b \\ 0, otherwise \end{cases}$$

Find the MGF of uniform distribution and hence find its mean and variance.

Sol: Let X follows uniform distribution in (a, b).

By Definition, $f(x) = \frac{1}{b-a}; a < x < b$.

$$M_X(t) = E[e^{tx}]$$

$$= \int_a^b f(x)e^{tx}dx$$

$$= \int_a^b \frac{1}{b-a} e^{tx}dx$$

$$= \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b$$

$$M_X(t) = \frac{1}{t(b-a)}[e^{bt} - e^{at}]$$

To find mean and variance:

$$M_X(t) = \frac{1}{t(b-a)}\left\{\left[1 + \frac{bt}{1!} + \frac{(bt)^2}{2!} + \frac{(bt)^3}{3!} + \cdots\right] - \left[1 + \frac{at}{1!} + \frac{(at)^2}{2!} + \frac{(at)^3}{3!} + \cdots\right]\right\}$$

$$= \frac{1}{b-a}\left\{\left[\frac{1}{t} + \frac{b}{1!} + \frac{b^2 t}{2!} + \frac{b^3 t^2}{3!} + \cdots\right] - \left[\frac{1}{t} + \frac{a}{1!} + \frac{a^2 t}{2!} + \frac{a^3 t^2}{3!} + \cdots\right]\right\}$$

Coefficient of $t = \frac{1}{b-a}\left[\frac{b^2}{2!} - \frac{a^2}{2!}\right] = \frac{1}{b-a}\frac{b^2 - a^2}{2!}$

$$= \frac{1}{2(b-a)}[(b+a)(b-a)]$$

$$= \frac{a+b}{2}$$

Coefficient of $t^2 = \frac{1}{b-a}\left[\frac{b^3}{3!} - \frac{a^3}{3!}\right]$

$$= \frac{1}{6(b-a)}[b^3 - a^3]$$

$$= \frac{1}{6}\frac{1}{(b-a)}(b-a)(b^2 + ba + a^2)$$

$$= \frac{1}{6}(b^2 + ab + a^2)$$

$$= \frac{b^2+ab+a^2}{6}$$

Mean $= E(X) = 1! \times$ coefficient of t

$$= 1! \times \left(\frac{a+b}{2}\right) = \frac{a+b}{2}$$

$$E(X^2) = 2! \times \text{ coefficient of } t^2$$

$$= 2 \times \frac{b^2+ab+a^2}{6}$$

$$= \frac{b^2+ab+a^2}{3}$$

$$\text{Variance} = E(X^2) - [E(X)]^2$$

$$= \frac{b^2+a^2+ab}{3} - \frac{(a+b)^2}{4}$$

$$= \frac{b^2+a^2+ab}{3} - \frac{(a^2+2ab+b^2)}{4}$$

$$= \frac{4b^2+4a^2+4ab-3a^2-6ab-3b^2}{12}$$

$$= \frac{a^2+b^2-2ab}{12}$$

$$= \frac{(a-b)^2}{12}$$

$$= \frac{(b-a)^2}{12} \quad \because a < b$$

Problem based on Uniform Distribution

1. A random variable X has a uniform distribution over (0, 10) compute

(i) $P(X < 2)$ (ii) $P(X > 8)$ (iii) $P(3 < X < 9)$

Solution:

The pdf $f(x) = \begin{cases} \frac{1}{b-a}, a < x < b \\ 0, otherwise \end{cases}$

$$f(x) = \begin{cases} \frac{1}{10}, 0 < x < 10 \\ 0, otherwise \end{cases}$$

(i) $P(X < 2) = \int_0^{10} f(x)dx$

$$= \frac{1}{10}\int_0^{10} dx$$

$$= \frac{1}{10}\ [x]_0^{10}$$

$$= \frac{1}{10}[10 - 0]$$

$$= \frac{1}{5}$$

(ii) $P(X > 8) = \int_8^{10} f(x)dx$

$$= \frac{1}{10}\int_8^{10} dx$$

$$= \frac{1}{10}\ [x]_8^{10}$$

$$= \frac{1}{10}[10 - 8]$$

$$= \frac{1}{5}$$

(iii) $P(3 < X < 9) = \int_3^9 f(x)dx$

$$= \frac{1}{10}\int_3^9 dx$$

$$= \frac{1}{10}\ [x]_3^9$$

$$= \frac{1}{10}[9 - 3]$$

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

2. A random variable X has a uniform distribution over (-3, 3) compute

(i) $P(X < 2)$ (ii) $P(|X| < 2)$ (iii) $P(|X - 2| < 2)$

Solution:

The pdf $f(x) = \begin{cases} \frac{1}{b-a}, a < x < b \\ 0, otherwise \end{cases}$

$$f(x) = \begin{cases} \frac{1}{6}, -3 < x < 3 \\ 0, otherwise \end{cases}$$

(i) $P(X < 2) = \int_{-3}^{2} f(x)dx$

$$= \frac{\mathbf{1}}{\mathbf{6}} \int_{-3}^{2} dx$$

$$= \frac{1}{6} \, [x]_{-3}^{2}$$

$$= \frac{1}{6} [2 + 3]$$

$$= \frac{5}{6}$$

(ii) $P(|X| < 2) = P(-2 < X < 2)$

$$= \frac{\mathbf{1}}{\mathbf{6}} \int_{-2}^{2} dx$$

$$= \frac{1}{6} \, [x]_{-2}^{2}$$

$$= \frac{1}{6} [2 + 2]$$

$$= \frac{4}{6}$$

(iii) $P(|X - 2| < 2) = P(-2 < X - 2 < 2)$

$$= P(-2 + 2 < X < 2 + 2)$$

$$= P(0 < X < 4)$$

$$= \frac{\mathbf{1}}{\mathbf{6}} \int_{0}^{3} dx$$

$$= \frac{1}{6} \, [x]_{0}^{3}$$

$= \frac{1}{6}[3-0]$

$= \frac{3}{6}$

3. 4 buses arrive at a specified stop at 15 minute intervals starting at 7 am. That is, they arrive at 7, 7.15, 7.30, 7.45 am and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7.30 am. Find the probability that he waits (i) less than 5 minutes for a bus (ii) more than 10 minutes for a bus.

Solution:

The pdf is $f(x) = \begin{cases} \frac{1}{30}, 0 < x < 30 \\ 0, otherwise \end{cases}$

(i) P(a person arrives between 7.10 and 7.15 or 7.25 and 7.30)

$= P(10 < X < 15) + P(25 < X < 30)$

$= \int_{10}^{15} f(x)dx + \int_{25}^{30} f(x)dx$

$= \mathbf{\frac{1}{30}}\int_{10}^{15} dx + \mathbf{\frac{1}{30}}\int_{25}^{30} dx$

$= \frac{1}{30}\ [x]_{10}^{15} + \frac{1}{30}\ [x]_{25}^{30}$

$= \frac{1}{30}[15-10] + \frac{1}{30}[30-25]$

$= \frac{1}{3}$

(ii) P(a person arrives between 7.00 and 7.05 or 7.15 and 7.20)

$= P(0 < X < 5) + P(15 < X < 20)$

$$= \int_0^5 f(x)dx + \int_{15}^0 f(x)dx$$

$$= \frac{\mathbf{1}}{\mathbf{30}}\int_0^5 dx + \frac{\mathbf{1}}{\mathbf{30}}\int_{15}^{20} dx$$

$$= \frac{1}{30}\ [x]_0^5 + \frac{1}{30}\ [x]_{15}^{20}$$

$$= \frac{1}{30}[5-0] + \frac{1}{30}[20-15]$$

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

4. Subway trains on a certain line run every half an hour between mid- night and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait atleast twenty minutes.

Solution:

The pdf is $f(x) = \begin{cases} \frac{1}{30-0}, 0 < x < 30 \\ 0, otherwise \end{cases}$

(i) P(a man waiting for atleast 20 minutes) $= P(X \geq 20)$

$$= \int_{20}^{30} f(x)dx$$

$$= \frac{\mathbf{1}}{\mathbf{30}}\int_{20}^{30} dx$$

$$= \frac{1}{30}\ [x]_{20}^{30}$$

$$= \frac{1}{30}[30 - 20]$$

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