Continuous Probability distributions

1. Uniform distribution

Let X be a cts r.v. Then X is said to have a cts uniform distribution in any finite interval $[a, b]$, if its prob. dist is constant over the entire interval.

Let $f(x) = \begin{cases} k, & a \le x \le b \\ 0 & \text{o.w.} \end{cases}$

Since $\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow k \int_{a}^{b} f(x)\,dx = 1$ or $k(b-a) = 1$

$\Rightarrow k = \frac{1}{b-a}$

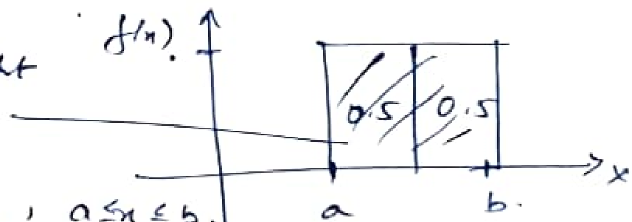
$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & \text{o.w.} \end{cases}$$

Area = base × height

Area under the entire curve = 1

$1 = (b-a) \times f(x)$

$\therefore f(x) = \frac{1}{b-a}, \quad a \le x \le b.$



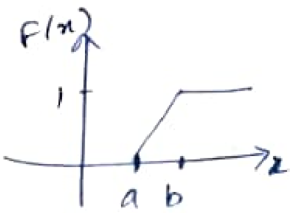
a & b are parameters of the distribution.

* This dist. is also called a rectangular dist since the curve $y = f(x)$ describes a rectangle over x-axis & b/w the ordinates $x = a$ & $x = b$.

$X \sim U[a, b]$

* The distribution function $F(x)$ is

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b \end{cases}$$



* $F(x)$ is not cts at a & b, it isn't diff at a & b.

$\frac{d}{dx} F(x) = f(x) = \frac{1}{b-a} \neq 0$

exists everywhere except at a & b.

· Median of the dist, - value x that splits the dist half ie. mid pt of a & b here

· Since it is a symmetric dist. & for a symmetric dist mean & median are same.

$\therefore$ Mean = median = $\frac{a+b}{2}$

To verify

$$E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\left.\frac{x^2}{2}\right)_a^b = \frac{(b-a)^2}{2(b-a)} = \frac{b+a}{2}$$

$$E(X^2) = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{3(b-a)}(b^3-a^3) = \frac{b^2+ab+a^2}{3}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2$$

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

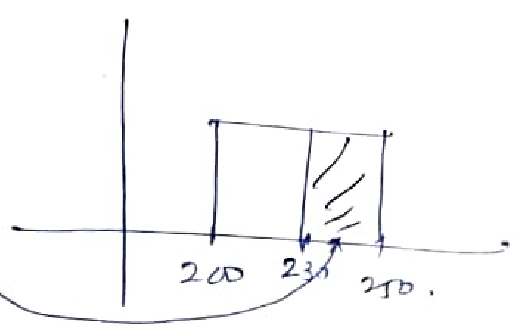
MGF of UD

$$M_X(t) = \int_a^b e^{tx} f(x)\,dx = \frac{1}{b-a}\int_a^b e^{tx}\,dx$$

$$= \frac{1}{(b-a)t}\left(e^{bt} - e^{at}\right),\quad t \neq 0$$

Ex If $a = 200$, $b = 250$. What is $f(x)$? Also find $P(X > 230)$

$$f(x) = \frac{1}{b-a} = \begin{cases} \frac{1}{50}, & 200 \le x \le 250 \\ 0 & \text{ow.} \end{cases}$$



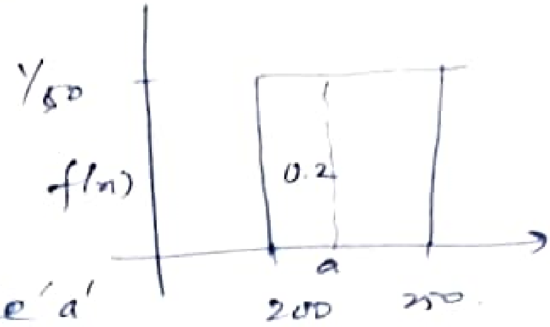
$P(X > 230)$ = Area of a rectangle

$= (250-230) \times \frac{1}{50} = \frac{20}{50} = \frac{2}{5}$

* For UD, areas under the curve is simply area of rectangle.

3. What is the 20th percentile of this distribution?

20th percentile is the value of the variable such that area to the left is 20% i.e. 0.2



Let the value of the variable be 'a'

$$(a-200) \times \frac{1}{50} = 0.2$$

$$a - 200 = 10 \Rightarrow a = 210$$

So the 20th percentile is 210.

Ex Suppose that a large conference room at a certain company can be reserved for no more than 4 hrs. Both long and short conferences occur quite often. In fact, it can be assumed that the length X of a conference has a uniform dist. on the interval [0, 4].

(i) what is the pdf?

(ii) what is the prob. that any given conference lasts at least 3 hrs.?

The density fn for the uniformly distributed r.v X is

$$f(x) = \begin{cases} \frac{1}{4-0} = \frac{1}{4}, & 0 \le x \le 4 \\ 0, & \text{o w} \end{cases}$$

$$P(X \geq 3) = \frac{1}{4}\int_3^4 dx = \frac{1}{4}.$$

predicts the wait time until the first event occurs

Exponential distribution

This distribution play an important role in both queue theory and reliability problems. Time b/w arrivals at service facilities and time to failure of component parts and electronic systems often are nicely modelled by exponential dist. Ex long distance telephone calls, the amount of time, car battery lasts. It is concerned with the amount of time until some specific event occurs.

The cts r.v X has an exponential distribution with parameter β, if its density function is given by

$$f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0 & \text{ow} \end{cases} \qquad \text{where } \beta > 0$$

or $\theta e^{-\theta x}$, $\theta > 0$

$\theta = 1/\beta$

Mgf of exponential dist

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx}\,\theta e^{-\theta x}\,dx = \theta \int_0^\infty e^{-(\theta - t)x}\,dx$$

$$= \frac{\theta}{\theta - t} = \left(1 - \frac{t}{\theta}\right)^{-1} = \sum_{r=0}^{\infty} \left(\frac{t}{\theta}\right)^r, \quad \theta > t$$

$M_X(t) = E(e^{tX}) = E(1 + tX + \frac{t^2X^2}{2!} + \dots)$
$= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \dots$
$= \sum \frac{t^r}{r!} \mu_r'$

$$\mu_r' = E(X^r) = \text{coeff of } \frac{t^r}{r!} \text{ in } M_X(t) = \frac{r!}{\theta^r}, \quad r = 1, 2, \dots$$

$$\text{Mean} = \mu_1' = \frac{1}{\theta}, \qquad \text{Variance} = \mu_2$$

$$= \mu_2' - \mu_1'^2$$

$$= \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

Hence if $X \sim \exp(\theta)$, Mean $= \frac{1}{\theta}$

Variance $= \frac{1}{\theta^2}$

$$\text{Var} = \frac{1}{\theta^2} = \frac{1}{\theta} \cdot \frac{1}{\theta} = \frac{\text{Mean}}{\theta}$$

Var > Mean if $0 < \theta < 1$
Var = Mean if $\theta = 1$
Var < Mean if $\theta > 1$

Suppose that a system contains a certain type of component whose time, in years, to failure is given by T. The random variable T is modeled nicely by exponential distribution with mean time to failure $\beta = 5$ $\left(= \frac{1}{\theta}\right)$.
If 5 of these components are installed in different systems, what is the prob. that atleast 2 are still functioning at the end of 8 years?

The prob. that a given component is still functioning after 8 yrs

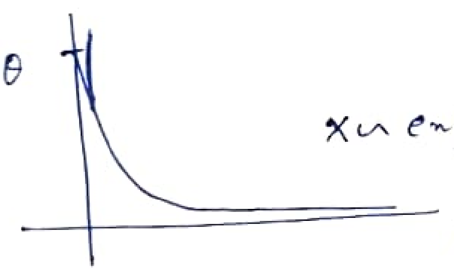
$$P(T>8) = \frac{1}{5}\int_8^\infty e^{-t/5}\,dt = e^{-8/5} \simeq 0.2$$

Let X represent the no. of components functioning after 8 yrs. Then using binomial dist

$$P(X \geq 2) = \sum_{x=2}^{5} b(x;5,0.2) = 1 - \sum_{x=0}^{1} b(x;5,0.2)$$

$$= 1 - 0.7373 = 0.2627$$

θ



$X \sim \exp(\theta)$, $x \geq 0$

$\theta > 0$, cannot be 0

As x increases, $f(x)$ becomes smaller.

To show it is a pdf.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^\infty \theta e^{-\theta x}\,dx = \theta \left.\frac{e^{-\theta x}}{-\theta}\right]_0^\infty$$

$$= -(0-1) = 1$$

OR $\longrightarrow$ $M_X(t) = \dfrac{\theta}{\theta - t}$

$$E(X) = \frac{d}{dt}\left(\frac{\theta}{\theta-t}\right)_{t=0} = \left.\frac{+\theta}{(\theta-t)^2}\right|_{t=0} = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

$$E(x^2) = \theta\left(\frac{2}{(\theta-t)^3}\right)_{t=0} = \frac{2\theta}{\theta^3} = \frac{2}{\theta^2}$$

$$Var(x) = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

Ex Median of $f(x) = 2e^{-2x}$

Median is 50%

$$\int_0^m 2e^{-2x}\,dx = 0.5$$

$$\frac{2}{-2}\left(e^{-2x}\right)_0^m = 0.5 \quad \text{or} \quad 1-e^{-2m} = 0.5$$

$$e^{-2m} = 0.5$$

$$-2m = \ln(0.5)$$

$$m = \frac{-0.693}{2} = .3465$$

Questions based on Uniform & Exponential distributions

Ex The time required to repair a machine is exponentially distributed with parameters $\frac{1}{2}$. What is the probability that a repair time exceeds 2 hr? What is the conditional prob. that a repair time takes atleast 10 hrs given that its duration exceeds 9 hrs.

X: time reqd. to repair a machine.

$\alpha = 1/2$

$X \sim exp(\alpha)$

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x}, & x>0,\ \alpha>0 \\ 0 & o.w \end{cases}$$

$$P(X>2) = \int_2^\infty \alpha e^{-\alpha x}\,dx = \frac{1}{2}\int_2^\infty e^{-x/2}\,dx$$

$$= \frac{1}{2}\left.\frac{e^{-x/2}}{(-1/2)}\right|_2^\infty = -\left[0 - e^{-1}\right) = \frac{1}{e}$$

$$P(X \ge 10 / X > 9) = \frac{P(X \ge 10)}{P(X > 9)} \quad \frac{P(A \cap B)}{P(B)} = \frac{\int_{10}^{\infty} \lambda e^{-\lambda x} dx}{\int_{9}^{\infty} \lambda e^{-\lambda x} dx}$$

$$= \frac{e^{-1/2 x}\Big|_{10}^{\infty}}{e^{-1/2 x}\Big|_{9}^{\infty}} = \frac{(e^{-5})}{e^{-9/2}} = e^{-5+9/2} = e^{-1/2}$$

OR memoryless property on page 6

$$= P(X > 1) = \frac{1}{2}\int_{1}^{\infty} e^{-x/2} dx = \frac{1}{2}\left(\frac{e^{-x/2}}{-1/2}\right)\Big|_1^{\infty} = -(0 - e^{-1/2}) = e^{-1/2}$$

$$E(x) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \lambda\left[\frac{x e^{-\lambda x}}{-\lambda}\Big|_0^{\infty} + \frac{1}{\lambda}\int e^{-\lambda x}\right]_0^{\infty}$$

$$= \lambda\left[-\frac{1}{\lambda}(0-0) - \frac{1}{\lambda^2}(0-1)\right] = \frac{1}{\lambda}$$

CDF $\quad F(a) = P(X \le a) = \int_{-\infty}^{a} f(x)dx = \lambda\int_{0}^{a} e^{-\lambda x} dx = -e^{-\lambda x}\Big)_0^a$

$$= -(e^{-\lambda a} - 1)$$

$$= 1 - e^{-\lambda a}$$

Ex exp Suppose that the length of a phone call in min is an exponential r.v. with parameter $\lambda = \frac{1}{5}$. If someone arrives immediately ahead of you at a public phone booth, find the prob. that you will have to wait

a) more than 5 min

b) b/w 5 & 10 min.

X : length of a phone call in min

$X \sim exp(1/5)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0, \lambda > 0 \\ 0 & ew. \end{cases}$$

$E(X) = \frac{1}{\lambda} = 5$, $\quad Var = \frac{1}{\lambda^2} = 25$

a) $P(X > 5) = \int_5^{\infty} \lambda e^{-\lambda x} dx = \frac{1}{5}\frac{e^{-x/5}}{-1/5}\Big|_5^{\infty} = -(0 - e^{-1}) = e^{-1}$

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

$$\lambda\int_5^{10} e^{-\lambda x} dx = -e^{-x/5}\Big)_5^{10} = e^{-1} - e^{-2} = \frac{e-1}{e^2}$$

$x > 0$

Ex Rock noise in an underground mine occurs at an average rate of 3 per hr. Find the prob that no rock noise will be recorded for atleast 30 mins. wait

X no of rock noise occur (No. of arrivals in 30 mins is zero)

Poisson $\lambda = 3/hr$

$\beta = 1/3$

$f(x) =$

exp X: the time to the first arrival

$\lambda = 1/3$ $\theta = 3$

$P(X > 30) = P(X > 1/2) = P(X > 0.5)$

$f(x) = 3e^{-3x}$

$P(X > b) = \int_b^\infty f(x)\,dx = \lambda \int_b^\infty e^{-\lambda x}\,dx = \lambda \left.\frac{e^{-\lambda x}}{-\lambda}\right)_b^\infty = e^{-\lambda b}$

Memoryless

$X \sim \exp(\beta)$, the amount of time the component has been in service has no effect on the amount of time until it fails

$$P(X > a+b \mid X > a) = P(X > b)$$

Pf: LHS $\dfrac{P((X > a+b) \cap (X > a))}{P(X > a)} = \dfrac{P(X > a+b)}{P(X > a)}$

$$= \frac{e^{-(a+b)/\beta}}{e^{-a/\beta}} = e^{-b/\beta} = P(X > b)$$

Given a device has already worked 5 hrs, the prob of it failing in the next hour is the same as it was at $t=0$

$$\frac{P(X > a+b)}{P(X>a)} = \frac{1-F_x(a+b)}{1-F_x(a)} = \frac{1-(1-e^{-\lambda(a+b)})}{1-(1-e^{-\lambda a})} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}}$$

$$= e^{-\lambda b}$$

$$= P(X > b)$$

Ex Uniform An examination paper has 150 MCQs of 1 marks each, with each question having 4 choices. Each incorrect answer fetches -0.25 marks. Suppose 1000 students choose all their answers randomly with uniform prob. The sum of all expected marks obtained by all these students is

Soln This is the case of discrete Uniform dist.

Not cts " ".

As per question, students choose their answer with uniform prob.

4 choices, $P(\text{choosing answer}) = 1/4$

Let X: Marks obtained by the student / question.

$$f(x) = \begin{cases} 1/4 & x = 1 \\ 3/4 & x = -0.25 \end{cases}$$

$$E(X) = \sum x f(x)$$
$$= 1 \times \frac{1}{4} + (-0.25) \times \frac{3}{4} = \frac{1}{4} - \frac{3}{16} = \frac{1}{16}$$

Let $X_1, X_2, \ldots X_{150}$ be 150 r.v.

$E(X_1) = E(X_2) = \ldots = E(X_{150}) = E(X)$

All are independent and identically distributed.

$\therefore$ S: marks obtained by 1 student

$$E(S) = E(X_1 + X_2 + \ldots + X_n)$$
$$= E(X_1) + E(X_2) + \ldots + E(X_n)$$
$$= 150\, E(X)$$

Now, total no. of students is 1000

TS: marks obtained by 1000 students

$$E(TS) = E(S_1) + E(S_2) + \ldots + E(S_{1000})$$
$$= 1000 \times 150\, E(X) = 1000 \times 150 \times \frac{1}{16} = \ldots$$

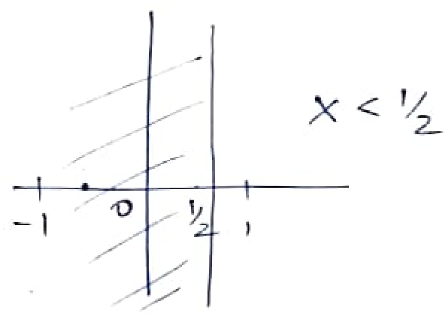
Ex Uniform

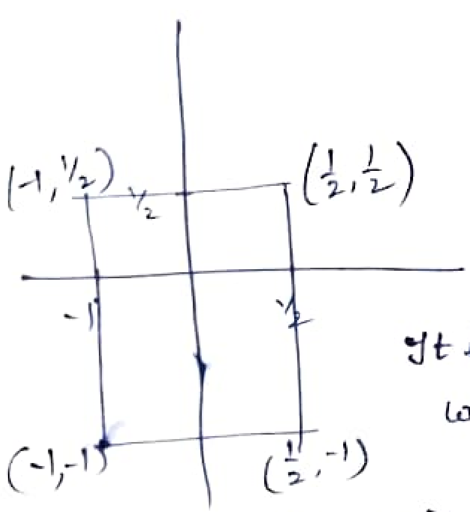
Two independent r.v X & Y are uniformly distributed in $[-1, 1]$. Prob that $\max(X, Y) < \frac{1}{2}$.

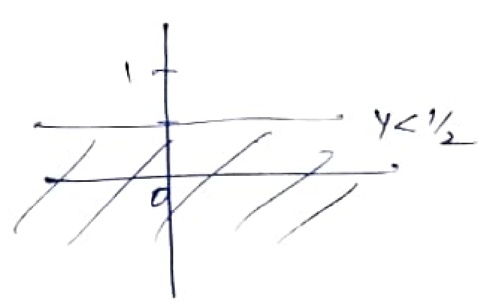
$\max(X, Y) < \frac{1}{2}$ only if

$X < \frac{1}{2}$ & $Y < \frac{1}{2}$

X & Y are uniformly distributed in $[-1, 1]$







It is a square with side $3/2$

$$P[\max(X, Y) < \tfrac{1}{2}] = \frac{\text{fav. outcome}}{\text{Total outcome}} = \frac{\text{shaded area}}{\text{Total area}}$$

$$= \frac{3/2 \times 3/2}{2 \times 2} = \frac{9}{4 \times 4} = \frac{9}{16}$$

Normal distribution / Gaussian dist Also

The most important cts prob dist in the entire field of statistics is called normal dist.

Its graph, called the normal curve, is the bell shaped curve which approximately describes many phenomena that occur in nature, industry and research.

For eg physical measurements in areas such as meteorological experiments, rainfall studies & measurements of manufactured parts are often more than adequately explained with a normal dist.

In addition, errors in scientific measurements are extremely well approximated by a normal dist.

It also forms a basis from which much of the theory of inductive stats is founded.

The normal dist is also called Gaussian dist.

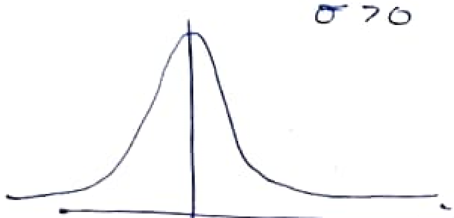
A cts r.v. X having the bell-shaped dist (of fig below) is called a normal r.v; with parameters μ (mean) & σ^2 (variance), if its pdf is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad \begin{array}{l} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0 \end{array}$$

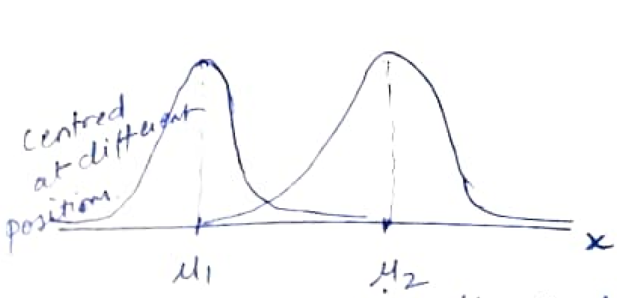
where $\pi = 3.14159\ldots$

$e = 2.71828\ldots$

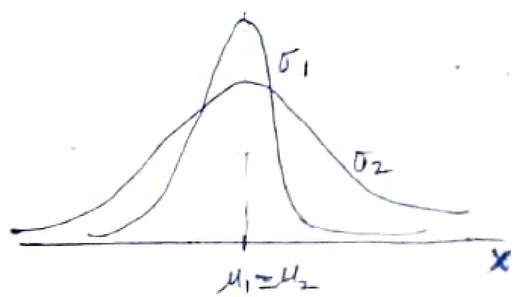
$X \sim N(\mu, \sigma^2)$



Once μ & σ are specified, the normal curve is completely determined.



Normal curves with $\mu_1 < \mu_2$ & $\sigma_1 = \sigma_2$



$\mu_1 = \mu_2$, $\sigma_1 < \sigma_2$

curve with larger s.d is lower and spreads out farther. Since the area under prob curve must be equal to 1 and therefore the more variable the set of observations, the lower and wider the corresponding curve will be.

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

let $\frac{x-\mu}{\sqrt{2}\sigma} = t$. Then $dx = \sigma\sqrt{2}\,dt$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2}\sigma\sqrt{2}\,dt = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2}\,dt$$

$$= \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1 \qquad \text{since } \int_{-\infty}^{\infty} e^{-t^2}\,dt = \sqrt{\pi}$$

<u>Properties of the normal curve</u>

1. The mode, which is the point on the horizontal axis where the curve is max, occurs at $x = \mu$.
2. The curve is symmetric about the vertical axis through mean μ.
3. The curve has its points of inflection ($y = f(x)$, $y'' = 0$) at $x = \mu \pm \sigma$, it is concave downward if $\mu - \sigma < X < \mu + \sigma$ and is
   " upward ~~if~~ otherwise.
4. The normal curve approaches the horizontal axis asymptotically as we proceed to either direction away from the mean.
5. The total area under the curve and above the horizontal axis is equal to 1.

Mean and variance of $N(\mu, \sigma^2)$

If $X \sim N(\mu, \sigma^2)$, then

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} dx$$

Let $\frac{x-\mu}{\sigma\sqrt{2}} = t$, $dx = \sqrt{2}\,\sigma\, dt$

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma\sqrt{2}\,t) e^{-t^2} \sqrt{2}\,\sigma\, dt$$

$$= \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt + \frac{2\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-t^2} dt$$

$$= \frac{\mu}{\sqrt{\pi}} \sqrt{\pi} + 0 = \mu$$

Using the same substitution

$$E(X^2) = \frac{\sqrt{2}\,\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma\sqrt{2}\,t)^2 e^{-t^2} dt$$

$$= \frac{1}{\sqrt{\pi}} \left[\int_{-\infty}^{\infty} \mu^2 e^{-t^2} dt + 2\sqrt{2}\,\sigma\mu \int_{-\infty}^{\infty} t e^{-t^2} dt + 2\sigma^2 \int_{-\infty}^{\infty} t^2 e^{-t^2} dt \right]$$

$$= \frac{1}{\sqrt{\pi}} \left[\mu^2\sqrt{\pi} + 0 + 4\sigma^2 \int_{0}^{\infty} t^2 e^{-t^2} dt \right]$$

Let $t^2 = u$, $2t\, dt = du$

$$= \frac{1}{\sqrt{\pi}} \left[\sqrt{\pi}\,\mu^2 + 2\sigma^2 \int_{0}^{\infty} u^{1/2} e^{-u} du \right]$$

$$= \frac{1}{\sqrt{\pi}} \left[\sqrt{\pi}\,\mu^2 + 2\sigma^2 \Gamma(3/2) \right] = \frac{1}{\sqrt{\pi}} \left[\sqrt{\pi}\,\mu^2 + 2\sigma^2 \frac{1}{2}\sqrt{\pi} \right]$$

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

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

normal dist has an interesting property that mean, median and mode of the dist. have the same value.

<u>Median</u> M of the dist is defined by

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

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{M}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx = \frac{1}{2}$$

$$\frac{1}{\sigma\sqrt{2\pi}} \left[\int_{M}^{\mu} e^{-(x-\mu)^2/2\sigma^2}\,dx + \int_{\mu}^{\infty} \text{''} \right] = \frac{1}{2}$$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{M}^{\mu} \text{''} \; dx + \frac{1}{2} = \frac{1}{2}$$

$$\int_{M}^{\mu} f(x)\,dx = 0$$

$$\therefore \underline{\mu = M}.$$

<u>Mode</u> of the dist. is that value of x for which the density for $f(x)$ is maximum.

$f'(x)=0$ & $f''(x)<0$

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

$$\log f(x) = \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{(x-\mu)^2}{2\sigma^2}$$

Differentiating wrt 'x',

$$\frac{f'(x)}{f(x)} = \frac{-1}{\sigma^2}(x-\mu)$$

Setting $f'(x)=0$, we get

$$f'(x) = -\frac{1}{\sigma^2}(x-\mu) f(x) = 0 \quad \text{or} \quad x=\mu$$

$$f''(x) = -\frac{1}{\sigma^2}\left[f(x) + (x-\mu) f'(x) \right]$$

$f''(x)\Big|_{x=\mu} = f''(\mu) = \frac{-1}{\sigma^2} f(\mu) < 0$

So, max $f(x)$ is obtained when $x = \mu$.

$\therefore$ mode of the dist $= \mu$.

<u>mgf of the dist</u>

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx}\, e^{-(x-\mu)^2/2\sigma^2}\, dx$$

let $z = \frac{x-\mu}{\sigma}$. Then $dx = \sigma\, dz$ and

$$M_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\mu+\sigma z)}\, e^{-z^2/2}\, dz$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z^2 - 2\sigma t z)/2}\, dz$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-[(z-\sigma t)^2 - \sigma^2 t^2]/2}\, dt$$

$$= \frac{e^{\mu t + \sigma^2 t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z-\sigma t)^2/2}\, dt \qquad \text{set } \frac{z-\sigma t}{\sqrt{2}} = u$$

$$= \frac{e^{\mu t + \sigma^2 t^2/2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2}\, du = e^{\mu t + \sigma^2 t^2/2}$$

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

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

$$= 1 + \mu t + \frac{t^2}{2!}(\mu^2 + \sigma^2) + \frac{t^3}{3!}(\mu^3 + 3\mu\sigma^2) + \cdots$$

$E(X) =$ coeff of $\frac{t'}{1!} = \mu$, $\quad E(X^3) = \mu^3 + 3\mu\sigma^2$ etc.

$E(X^2) =$ " $\frac{t^2}{2!} = \sigma^2 + \mu^2$,

...th order central moment is

$$\mu_k = E\left[(X-\mu)^k\right] = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\mu)^k e^{-(x-\mu)^2/2\sigma^2}\,dx$$

we obtain $\mu_1 = 0$, $\mu_2 = \sigma^2$, $\mu_3 = 0$, $\mu_4 = 3\sigma^4$ etc.

* An important property –
let X_i, $i = 1, 2, \ldots, n$ be n independent normal variables with mean μ_i & variance σ_i^2, resp., i.e., $X_i \sim N(\mu_i, \sigma_i^2)$.

Then $\displaystyle\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$

the additive property of the normal dist.

Pf $\quad M_{X_i}(t) = e^{\mu_i t + \sigma_i t^2/2}$

mgf of $a_1X_1 + a_2X_2 + \ldots + a_nX_n$ where a_i are constants is

$$\begin{aligned}
M_{a_1X_1+\cdots a_nX_n}(t) &= M_{a_1X_1}(t) \cdot M_{a_2X_2}(t) \cdots M_{a_nX_n}(t) \\
&= M_{X_1}(a_1t)\, M_{X_2}(a_2t) \cdots M_{X_n}(a_nt) \\
&= e^{a_1\mu_1 t + \sigma_1^2 a_1^2 t^2/2}\, e^{a_2\mu_2 t + \sigma_2^2 a_2^2 t^2/2} \cdots e^{a_n\mu_n t + \sigma_n^2 a_n^2 t^2/2} \\
&= e^{(a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n)t + (\sigma_1^2a_1^2 + \sigma_2^2a_2^2 + \cdots + \sigma_n^2a_n^2)t^2/2}
\end{aligned}$$

which is the mgf of normal variate with mean $\displaystyle\sum_{i=1}^{n} a_i\mu_i$ & variance $\displaystyle\sum_{i=1}^{n} a_i^2\sigma_i^2$

Particular cases

(i) Let $a_1 = a_2 = 1$ & $a_3 = a_4 = \cdots = a_n = 0$. Then
$X_1 + X_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

(ii) $a_1 = 1, a_2 = -1$, others $= 0$ $\quad X_1 - X_2 \sim N(\mu_1-\mu_2, \sigma_1^2+\sigma_2^2)$

(iii) $a_1 = a_2 = \cdots = a_n = \frac{1}{n}$
$\displaystyle\frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\frac{1}{n}\sum_{i=1}^{n}\mu_i, \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\right)$

LHS in $\bar{x}$ Thus

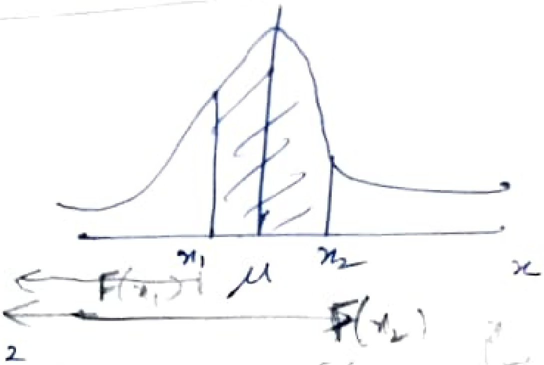
$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$\therefore$ if X_i are independent & identically distributed normal variates with mean μ & variance σ^2, then their mean $\bar{x}$ is also a normal variate & $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Area under the normal Curve

The curve of any cts prob. dist or density fn is constructed so that the area under the curve bounded by $x = x_1$ and $x = x_2$ equals the prob that r.v X assumes a value b/w $x = x_1$ & $x = x_2$.

Thus for the normal curve



$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} N(\mu, \sigma^2)\, dx \quad f(x) \text{ or}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

$F(x_2) - F(x_1) =$

is represented by the area of the shaded region.

The difficulty encountered in solving integrals of normal density functions necessitates the tabulation of normal curve areas for quick reference.

However, in order to avoid set up separate tables for every value of μ & σ, we can transform all the observations of any normal r.v X into a new set of observations of a normal random variable Z with mean 0 & variance 1. This can be done by using the transformation

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

whenever X assumes the value x, the corresponding value of Z is given by $\frac{x-\mu}{\sigma}$

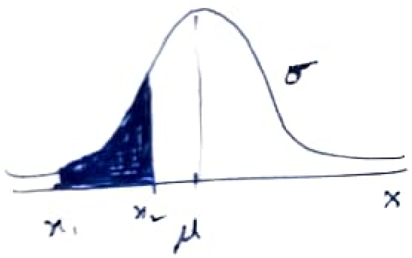
$\therefore$ if X falls b/w $x = x_1$ & $x = x_2$, the r.v Z will fall b/w corresponding values $z_1 = \frac{x_1 - \mu}{\sigma}$ & $z_2 = \frac{x_2 - \mu}{\sigma}$

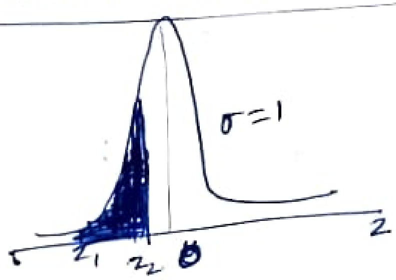
$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{x_1}^{x_2} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{z_1}^{z_2} e^{-z^2/2} dz = \int_{z_1}^{z_2} n(z;0,1) dz$$

$$= P(z_1 < Z < z_2)$$

* The distribution of a normal r.v with mean 0 & variance 1 is called standard normal distribution.





We have now reduced the reqd no. of tables of normal curve areas to one, that of the standard normal dist. Table — indicates the area under the standard normal curve corresponding to $P(Z < z)$ for values of z ranging from -3.49 to 3.49.

for eg $P(Z < 1.74)$

first we locate $z = 1.7$ in left column

then we move across row to the column under 0.04,

we see 0.9591

$P(Z < 1.74) = 0.9591$

To find a z value corresponding to given prob., the ~~process~~ is reversed.

For eg: z value leaving an area of 0.2148 under the curve to the left of z is seen to be -0.79

$Z \sim N(0,1)$ → This doesn't change the shape of the dist.

$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ $-\infty < z < \infty$ → 'Standard normal prob. curve'

The dist fn. $F(z)$

$$F(z) = P(Z \le z) = \int_{-\infty}^{z} \phi(u)\,du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\,du$$

Properties of $F(z)$

(i) $F(-z) = 1 - F(z)$

$$F(-z) = P(Z \le -z) = P(Z \ge z) \quad \text{(by symmetry)}$$
$$= 1 - P(Z \le z) = 1 - F(z)$$

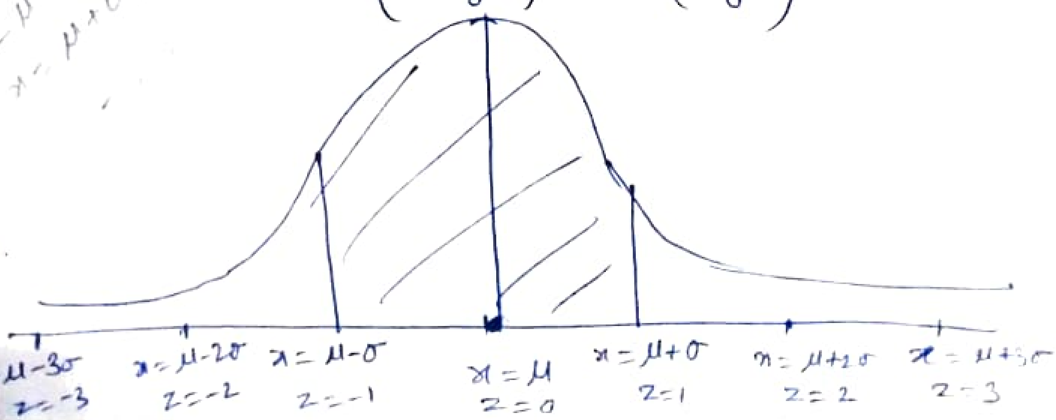
(ii) $P(a \le X \le b) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$

$X \sim N(\mu, \sigma^2)$. $Z = \frac{X-\mu}{\sigma^2}$

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$
$$= P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right)$$
$$= F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

$z = \frac{x-\mu}{\sigma}$, $x = \mu - \sigma$, $z = -1$, $x = \mu + \sigma$, $z = 1$

Normal prob. integral



$$P(\mu < X < x_1) = P(0 < Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_0^{z_1} e^{-z^2/2} \, dz$$

$$= \int_0^{z_1} \phi(z) \, dz$$

Def int $\int_0^{z_1} \phi(z) \, dz$ is called the normal prob. integral.

It gives the areas under the standard normal curve $y = \phi(z)$ b/w $z = 0$ & $z = z_1$. These areas are tabulated for different values of z_1, at intervals of 0.01.

- The prob. that the r.v X lies in $(\mu - \sigma, \mu + \sigma)$ is

$$P(\mu - \sigma < X < \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x) \, dx$$

or $$P(-1 < Z < 1) = \int_{-1}^{1} \phi(z) \, dz \qquad z = \frac{x - \mu}{\sigma}$$

$0.8413 - (+0.1587)$

$$= 2 \int_0^1 \phi(z) \, dz \quad \text{by symmetry.}$$

$$= 2(0.3413) = 0.6826 \quad \text{(from table)}$$

Thus, approx there is a prob of 68% that a normal variate lies in the interval $(\mu - \sigma, \mu + \sigma)$

- $$P(\mu - 2\sigma < X < \mu + 2\sigma) = \int_{\mu-2\sigma}^{\mu+2\sigma} f(x) \, dx$$

or $$P(-2 < Z < 2) = \int_{-2}^{2} \phi(z) \, dz = 2 \int_0^2 \phi(z) \, dz$$

$$= 2(0.4772) = 0.9544$$

Thus, approx, there is a prob. of 95% that a normal variate lies in the interval $(\mu - 2\sigma, \mu + 2\sigma)$.

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = \int_{\mu-3\sigma}^{\mu+3\sigma} f(x)\,dx$$

or $$P(-3 < Z < 3) = \int_{-3}^{3} \phi(z)\,dz = 2\int_{0}^{3} \phi(z)\,dz$$

$$= 2(0.49865) = 0.9973$$

99.7% X $(\mu - 3\sigma, \mu + 3\sigma)$.

Given a standard normal dist find the area under the curve that lies

a) to the right of $z = 1.84$ and b) b/w $z = -1.97$ and $z = 0.86$

$1 - F(1.84)$

$1 -$ Area to the left of $z = 1.84$

$1 - 0.9671 = 0.0329$

$F(0.86) - F(-1.97)$

Area to the left of $z = 0.86$

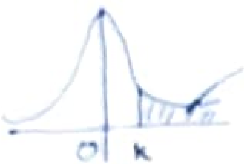
$-$ Area $z = -1.97$

$= 0.8051 - 0.0244$

$= 0.7807$

Ex Given a standard normal dist, find the value of k s.t.

a) $P(z > k) = 0.3015$



$P(z \leq k) = 1 - 0.3015$

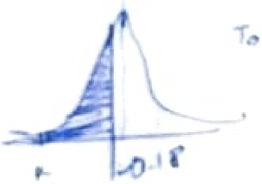
$= 0.6985$

$\Rightarrow k = 0.52$ (from table)

b) $P(k < z < -0.18) = 0.4197$

$F(-0.18) - F(k) = 0.4197$

Total area to the left of $z = -0.18$ is $P(z < -0.18) = 0.4286$



Area b/w k & -0.18 is 0.4197

$\Rightarrow$ Area to the left of $k = 0.4286 - 0.4197$

$= 0.0089$

$k = -2.37$ (from table)

Ex A r.v X is normally distributed with mean 9 and s.d 3

a) find the prob (i) $X \geq 15$ (ii) $X \leq 15$ (iii) $0 \leq X \leq 9$

$\mu = 9$, $\sigma = 3$ & $X \sim N(9, 9)$.

The standard normal variate $z = \frac{X - 9}{3}$

(i) $X = 15$, $z = \frac{15 - 9}{3} = 2$

$P(X \geq 15) = P(z \geq 2) = 0.5 - P(0 \leq z \leq 2)$

$= 0.5 - 0.4772 = 0.0228$

or $1 - P(z \leq 2) = 1 - 0.9772$

(ii) $P(X \le 15) = 1 - P(X \ge 15) = 1 - 0.0228 = 0.9772$

(iii) $P(0 \le X \le 9) = P(-3 \le Z \le 0) = P(0 \le Z \le 3) = 0.4$

$= P(Z \ge 0) - P(Z \ge 3) = 0.5 - 0.0013$

$= 0.$

(b) find x^* when $P(X > x^*) = 0.16$

$P(X > x^*) = P(Z > z^*) = 0.16$

$P(0 < Z < z^*) = 0.5 - 0.16 = 0.34$

from table $z^* = 1$

$\frac{x^* - \mu}{\sigma} = 1$ or $x^* = 9 + 3 = 12$

Ex r.v. X, $\mu = 50$, $\sigma = 10$

find the prob that X assumes a value b/w 45 and 62.

z values corresponding to $x_1 = 45$ and $x_2 = 62$ are

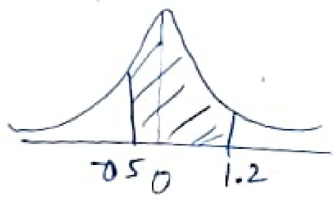
$$z_1 = \frac{45 - 50}{10} = -0.5, \quad z_2 = \frac{62 - 50}{10} = 1.2$$

$\therefore P(45 < X < 62) = P(-0.5 < Z < 1.2)$

$= P(Z < 1.2) - P(Z < -0.5)$

$= 0.8849 - 0.3085$

$= 0.5764$



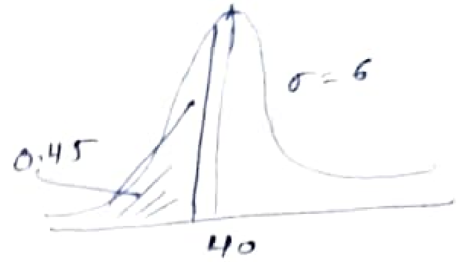
Using the normal curve in reverse

Sometimes, we are reqd to find the value of z corresponding to a specified probability that falls b/w values listed in Table. For convenience, we shall always choose z value corresponding to the tabular probability that comes closest to the specified probability.

In previous 2 ex., we solved by first going from a value of x to a z value & then computing the desired area. Now, we reverse the process & begin with known area/prob. find z value & determine x by $z = \frac{x - \mu}{\sigma}$. to give $x = \mu + z\sigma$

Q. In a normal dist with $\mu = 40$ and $\sigma = 6$, find the value of x that has

i) 45% of the area to the left



from table

$P(z < -0.13) = 0.45$

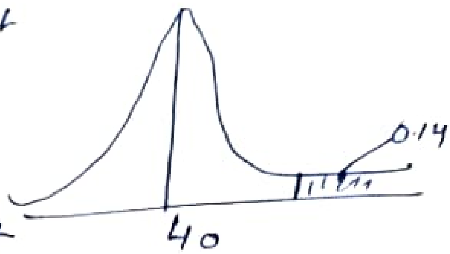
We require a z value that leaves an area of 0.45 to the left.

$\therefore z = -0.13$

$\therefore x = 6 \times -0.13 + 40$

$= 39.22$

(ii) 14% of the area to the right



we require z value that leaves 0.14 of the area to the right & hence an area of 0.86 to the left

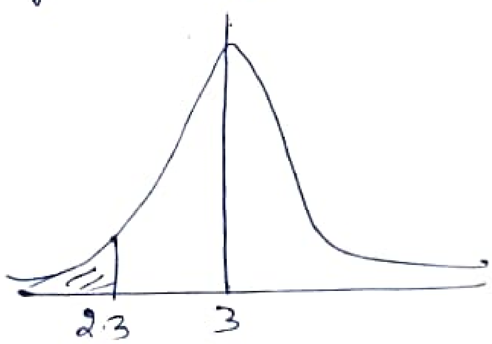
from table, $P(z < 1.08) = 0.86$

$\therefore z = 1.08$

$x = 6(1.08) + 40 = 46.48$

Ex A certain type of storage battery lasts, on average, 3 yrs with a s.d. of 0.5 yr. Assuming that battery life is normally distributed, find the prob that a given battery will last less than 2.3 yrs.

$P(x < 2.3) = ?$



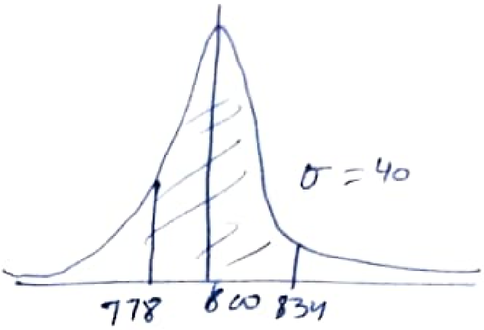
we need to evaluate the area under normal curve to the left of 2.3.

$$z = \frac{2.3 - 3}{0.5} = -1.4$$

$P(x < 2.3) = P(z < -1.4) = 0.0808$

Ex An electrical firm manufactures light bulbs that have a life, before burn out, that is normally distributed with mean = 800 hrs & s d = 40 hrs. Find the prob. that a bulb burns b/w 778 and 834 hrs.

$x_1 = 778$, $x_2 = 834$



$$z_1 = \frac{778 - 800}{40} = -0.55$$

$$z_2 = \frac{834 - 800}{40} = 0.85$$

$$P(778 < X < 834) = P(-0.55 < z < 0.85)$$
$$= P(z < 0.85) - P(z < -0.55)$$
$$= 0.8023 - 0.2912 = 0.5111$$

Ex There are 500 students taking a Mathematics course in an engg. college. The prob ~~that~~ for any student to need a particular book from a college library on any day is 0.07. How many copies of a book should be kept in the library so that the prob. may be greater than 0.95 that none of the students needing a copy from the library has to go back disappointed? Assume normal dist.

$n = 500$, $p = 0.07$, $\mu = np = 35$, $\sigma^2 = npq = 32.55$

$\sigma = 5.7$

find x^* s.t.

$$P(X < x^*) \geq 0.95$$

set $z_1 = \dfrac{x^* - 35}{5.7}$. Hence $P(0 < z < z_1) \geq 0.45$

$P(z < z_1) > 0.95$

from table, $z_1 > 1.65$ or $x^* > 35 + 5.7(1.65)$

$> 44.4 \simeq 45$

Hence the library should keep atleast 45 copies.