Probability (MA2110)

Random experiment properties ($\mathcal{E}$):-

→ Repeatable
→ Entirety
→ Regularity

$\mathcal{L} \subseteq \mathcal{S}$

events   sample space

+ $P : P(\mathcal{S}) \longrightarrow [0, 1]$

i.e. $P(\phi) = 0$ and $P(\mathcal{S}) = 1$

Also, if $A \subseteq B$; then $P(A) \leq P(B)$

* If $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \ldots, \mathcal{L}_n \ldots$ $(\mathcal{L}_i \cap \mathcal{L}_j = \phi)$ is a disjoint set,

Then $$\boxed{P\left(\bigcup_{n=1}^{\infty} \mathcal{L}_n\right) = \sum_{n=1}^{\infty} P(\mathcal{L}_1)}$$

* Elementary event is a singleton set (only 1 outcome).

For elementary events of a finite sample space, $P(\mathcal{L}) = \dfrac{|\mathcal{L}|}{|\mathcal{S}|}$

Relative Frequency:-

$$\boxed{f_{\mathcal{L}} = \lim_{n \to \infty} \frac{n_{\mathcal{L}}}{n} \simeq P(\mathcal{L})}$$

1) Throwing Two dice :-

$\mathcal{S} = \{(1,1), (1,2), \ldots \ldots (6,5), (6,6)\}$

A: $x_1 + x_2 = 10$; then $A = \{(6,4), (5,5), (4,6)\}$

$\therefore P(A) = \dfrac{3}{36} = \dfrac{1}{12}$

Now, $B: x_1 > x_2$; Then $P(B) = \frac{15}{36} = \underline{\underline{\frac{5}{12}}}$

- $S_1 = \{2, 3, 4, ..., 12\}$ is a sample space for 'A' but not 'B'
- $S_2 = \{0, 1\}$ is a sample space for 'B' but not 'A'

* Urn : 50 B + 40 W + 30 R Balls are there

Total = 120 . We need to pick 25 balls.

$$P(10B + 5W + 10R) = \frac{{}^{50}C_{10} * {}^{40}C_{5} \times {}^{30}C_{10}}{{}^{150}C_{25}}$$

* $P(B/A) = \frac{P(A \cap B)}{P(A)}$

Joint v/s Conditional :-

Joint $\rightarrow A \cap B$

Conditional $\rightarrow B/A$ or A/B

- Time component
- Space component

Q: 3 factories. 1^{st} one produces twice as much as 2^{nd} & 3^{rd}. 4% from 1^{st} one are defective. 2% from 2^{nd} and 3^{rd} each are defective. All products are combined in a godown. Probability that a product chosen is defective.

Ans. $P(F_1) = \frac{1}{2}$ $P(F_2) = \frac{1}{4}$ $P(F_3) = \frac{1}{4}$

$P(D/F_1) = \frac{4}{100} = 0.04$

$P(D/F_2) = 0.02$, $P(D/F_3) = 0.02$

Now, $P(F_1/D) = \frac{P(F_1 \cap D)}{P(D)} = \frac{P(F_1) \cdot P(D/F_1)}{\sum_{i=1} P(F_i) P(D/F_i)}$

$P(D) = P(F1)\, P(D/F1) + P(F2)\, P(D/F2) + P(F3)\, P(D/F2)$

* Indipendent events :-

A & B are indipendent if $P(A \cap B) = P(A) \cdot P(B)$

i.e. $P(B/A) = P(B)$ [B is not affected by A]

Q: Two coins are tossed

A : Both Heads

B : First one is heads

$P(A/B) = ?$

Ans: $P(A/B) = \dfrac{P(A \cap B)}{P(B)}$

$P(A) = \dfrac{1}{4}$, $P(B) = \dfrac{1}{2}$

$P(A \cap B) = \dfrac{1}{4}$

$\therefore P(A/B) = \dfrac{\frac{1}{4}}{\frac{1}{2}} = \dfrac{1}{2}$

$\boxed{P(A/B) = \dfrac{1}{2}}$

But $P(A \cap B) \neq P(A) \cdot P(B)$

'A' & 'B' are dependent events

Q: An insurance company finds that 20% of applires are accident prone. Out of these 20%, only 80% people actually meet with an accident.

Out of remaining 80% people, 20% meet with an accident.

(i) What is probability that someone will claim?

Ans: $P(A) = 0.2$, $P(A^c) = 0.8$ [Prior]

A → accident prone

C → claim

$P(C/A) = 0.8$, $P(C/A^c) = 0.2$ [Conditional]

We have to find $P(C)$ [Total]

$\Rightarrow P(C) = P(C/A)\,P(A) + P(C/A^c)\,P(A^c)$

$\Rightarrow P(C) = 0.8 \times 0.2 + 0.2 \times 0.8$

$\Rightarrow \boxed{P(C) = 0.32}$

iii) Given that a claim is made, what is probability that he is accident prone?

Ans: $P(A/C)$ has to be found

$$\Rightarrow P(A/C) = \frac{P(A \cap C)}{P(C)} = \frac{P(C/A) \cdot P(A)}{P(C)} \quad \text{[Bayesian]}$$

$$\Rightarrow P(A/C) = \frac{0.8 \times 0.2}{0.32} = 0.5$$

$$\boxed{P(A/C) = 0.5}$$

** Bayesian probability is of the form $\frac{\alpha}{\alpha+\beta}$ i.e. $P(C/A) \cdot P(A) = \alpha$, $P(C/A^c)\,P(A^c) = \beta$

$P(C) = P(C/A)\,P(A) + P(C/A^c) \cdot P(A^c)$

$P(C) = \alpha + \beta$

** Bayesian probability gives us a chance to change/adjust the prior probability in order to refine the problem.

Random Variables :-

$X : S \longrightarrow T \subseteq \mathbb{R}$

(i) It helps in standardising notations.

Ex:-

$$\begin{tikzcd} S \arrow[r, "P"] \arrow[d, "X"'] & {[0,1]} \\ T = \{t_1, t_2, \ldots, t_k\} \arrow[ur, dashed, "M"] & \end{tikzcd}$$

$M(t_1) = P(X^{-1}(t_1)) \in [0,1] \quad = P\{s \in S \mid X(s) = t_k\}$
$\forall s \in S$

Hence, it can be modified Directly from $T \longrightarrow [0,1]$

(ii) Captures Patterns

$$\boxed{P(X=k) = (1-p)^{k-1}\, p}$$

$p \to$ probability of success

This type of random variables are called geometric random variables.

Ex:- Urn : $3W + 3B + 4R$
$\downarrow \quad \downarrow \quad \downarrow$
$+\$1 \quad -\$1 \quad \$0$

2 balls are picked
~~[illegible]~~ Probability that he will lose \$2

$X=$	-2	-1	0	1	2
	BB	BR, RB	WB, BW, RR	WR, RW	WW

$X \rightarrow$	-2	-1	0	1	2
Cases $\rightarrow$	BB	BR RB	WB BW RR	WR RW	WW
Probability $\rightarrow$	$\frac{3}{10} \cdot \frac{2}{9}$	$2 \cdot \frac{3}{10} \cdot \frac{4}{9}$	$\left(2 \cdot \frac{3}{10} \cdot \frac{3}{9}\right) + \frac{4}{10} \cdot \frac{3}{9}$	$2 \cdot \frac{3}{10} \cdot \frac{4}{9}$	$\frac{3}{10} \cdot \frac{2}{9}$

In this case, there is no pattern observable.

$\therefore$ The random variable doesn't always capture patterns.

Note: Sum of all individual probabilities = 1

$$\sum_{i=1}^{n} P(X_i) = 1$$

* $X : \xi \longrightarrow \mathbb{T} \subseteq \mathbb{R} \quad (|T|^n < \infty)$

$$\mathbb{T} = \{t_1, t_2, \ldots, t_n\}$$

Any random variable (R.V) whose set $\mathbb{T}$ is a finite set or countably infinite is called **Discrete R.V**.

* $P(X = k) = P\{X^{-1}(k)\} = p(k) \geq 0$

$$p : \{1, 2, \ldots, n\} \longrightarrow [0, 1]$$

'p' is probability mass function (p.m.f)

* Attached to every discrete R.V, there is a p.m.f

$$P(\xi) = P\left(\bigcup_{k=1}^{n} X^{-1}(k)\right) = \sum_{k=1}^{n} p(k) = 1$$

$$\sum_{k=1}^{n} p(k) = \sum_{k=1}^{n} P[X^{-1}(k)]$$

* Attached to every discrete R.V, there is a p.m.f (probability mass function)

& a c.d.f (cumulative distribution function)

* A c.d.f is $F: \pi \longrightarrow [0, 1]$

$$F(k) = \sum_{i=1}^{k} p(i) = P[X \leq k]$$

* Types of R.Vs :-

1) X - Bernoulli :-

Success (or) Failure

Success $\rightarrow p$

Failure $\rightarrow 1-p$

Ex:- i) E_1 - 5 tosses of a coin

X_1 - no. of heads

$X_1 \rightarrow$	0	1	2	3	4	5
	$\left(\frac{1}{2}\right)^5$	$\frac{5c_1}{2^5}$	$\frac{5c_2}{2^5}$	$\frac{5c_3}{2^5}$	$\frac{5c_4}{2^5}$	$\frac{5c_5}{2^5}$

(ii) E_2 - 10 shots at a target

Probability that he misses is '0.2'

X_2 - no. of misses

~~X_2 = 0 | 1 | 2 | 3 | 4 . . . . . . 10~~

~~$(0.2)^{10}$ | $10c_1 (0.8)(0.2)$~~

$X_2 =$	0	1	2	9	10

$^nC_0\left(\frac{5}{6}\right)^n$	$^nC_1\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{n-1}$	$^nC_2\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{n-2}$	.....	$^nC_k\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k}$	....	$^nC_n\left(\frac{1}{6}\right)^n$

For k, it is: $^nC_k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$

Q. Urn: 3W + 3B + 4R

3 picks

X_u – no. of white balls

X_u –	0	1	2	3
with replacement	$^3C_0\left(\frac{3}{10}\right)^0\left(\frac{7}{10}\right)^3$	$^3C_1\left(\frac{3}{10}\right)^1\left(\frac{7}{10}\right)^2$	$^3C_2\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)$	$^3C_3\left(\frac{3}{10}\right)^3\left(\frac{7}{10}\right)^0$
without replacement	$\frac{7}{10}\cdot\frac{6}{9}\cdot\frac{5}{8}$	$\frac{7}{10}\cdot\frac{6}{9}\cdot\frac{3}{8} + \frac{7}{10}\cdot\frac{3}{9}\cdot\frac{6}{8} + \frac{3}{10}\cdot\frac{7}{9}\cdot\frac{6}{8}$	$3\cdot\frac{7}{10}\cdot\frac{3}{9}\cdot\frac{2}{8}$	$\frac{3}{10}\cdot\frac{2}{9}\cdot\frac{1}{8}$

We can see that the R.V is Bernoulli only if there is replacement.

This is because in without replacement, the 'n' & 'p' keep on changing. That means the 1st pick & 2nd pick & 3rd pick are **not** indipendent.

0 1 2 . . . . k . . . . n ($X \to$)

Q: There are 2 planes — (i) 4 engines (atleast 2 must work for flying)

(ii) 2 engines (atleast 1 must work for flying)

Probability that engine will fail is 'p' for all engines.

When should 'I' prefer 2 engines over 4 engines?

Ans:

<u>2 engines</u>

X_1 - no. of engines working

X_1 - 0 1 2

$P[X_1 \geq 1]$

$= 1 - P[X=0]$

$= 1 - p^2$

<u>4 engines</u>

X_2 - no. of engines working

X_2 - 0 1 2 3 4

$P[X_2 \geq 2]$

$= 1 - P[X=0] - P[X=1]$

$= 1 - {}^4C_0\, p^4 - {}^4C_1\, p^3(1-p)$

Now, $\cancel{1} - 2p^2 > \cancel{1} - p^4 - 4p^3 + 4p^4$,

$\Rightarrow -2p^2 > 3p^4 - 4p^3$

$\Rightarrow 3p^4 - 4p^3 - 2p^2 < 0$

$p^2(3p^2 - 4p - 2) < 0$

$\Rightarrow 3p^2 - 4p - 2 < 0$

$\dfrac{4 \pm 2\sqrt{10}}{6}$

$\dfrac{2 \pm \sqrt{10}}{3}$

$\Rightarrow \left[p - \left(\dfrac{2-\sqrt{10}}{3}\right)\right]\left[p - \dfrac{2+\sqrt{10}}{3}\right] < 0$

$\boxed{0 < p < \dfrac{2+\sqrt{10}}{3}}$

Scanned by CamScanner

Q: 'P' is the probability that a product is defective.

There are 10 producs in a box. They are sold. What is probability of a box being returned, if more than 1 product is defective in a box of 10?

Ans: X - no. of defective items in a box of 10

$P[X>1] = 1 - P[X=0] - P[X=1]$

$P = 1 - {}^{10}C_0 (1-p)^{10} - {}^{10}C_1 p(1-p)^9$

---

* X_1 - no. of tosses till 1st head

X -	1	2	3	4	- - - - -
	p	$(1-p)(p)$	$(1-p)^2 p$	$(1-p)^3 p$	- - - - -

These kinds of R.V are called Geometric R.V.

(ii) Geometric R.V

(iii) Binomial R.V - $^nC_k (p)^k (1-p)^{n-k}$

* All the trials of a geometric R.V are independent.

* Bernoulli → $X \sim b(p)$
  ($\sim$ distributed)

* Binomial → $X \sim B(n,p)$

* Geometric → $X \sim G(p)$

* Continuous R.V :-

$X : \xi \longrightarrow \Pi \subseteq \mathbb{R}$

$|\Pi| = \infty$

$f_x : \Pi \longrightarrow \mathbb{R}^{\geq 0}$

Probability density function (pdf)

cdf $\longrightarrow$ $F_X(t_0) = \int_{-\infty}^{t_0} f_x(t)\, dt$

Ex :- $X \sim (0, 1)$

every point is equiprobable.

$$f_x(t_0) = \begin{cases} 0 , & t<0 \\ c , & t \in [0,1] \\ 0 , & t>1 \end{cases}$$ [c is some const.]

pdf of the given distribution.

$\Rightarrow \int_{-\infty}^{\infty} f_x(t)\, dt = 1$

$\Rightarrow \cancel{\int_{-\infty}^{0} f_x(t)\, dt}^{\,0} + \int_{0}^{1} f_x(t)\, dt + \cancel{\int_{1}^{\infty} f_x(t)\, dt}^{\,0} = 1$

$\Rightarrow \int_{0}^{1} f_x(t)\, dt = 1$

$\Rightarrow c = 1$

Here the pdf is $c = 1$ for $t \in [0, 1)$

- Note :- pmf gives the actual probability of the event occuring
  But pdf doesn't give the actual probability.

- pmf $\in [0,1]$

pdf can be greater than '1' also

- Point probabilities are '0'

* $(a,b) \subseteq \pi$

$$P[a<x<b] = \int_a^b f_x(t)\,dt = F_x(b) - F_x(a) = 1$$

$$\Rightarrow f_x(t) = \begin{cases} 0, & t<a \\ c, & t\in(a,b) \\ 0, & t>b \end{cases}$$

$$\Rightarrow \int_a^b c\,dt = 1$$

$$\Rightarrow c(b-a) = 1$$

$$\Rightarrow \boxed{c = \frac{1}{b-a}}$$

Q: $f(x) = \frac{x^2}{9}$ $[0,b]$

For this function to be a valid pdf, find 'b'.

Ans: $\int_0^b f(x)\,dx = 1$

$$\Rightarrow \frac{b^3}{27} = 1$$

$$\Rightarrow \boxed{b=3}$$

* $f(x) = \frac{x^2}{9}$ $[0,3]$

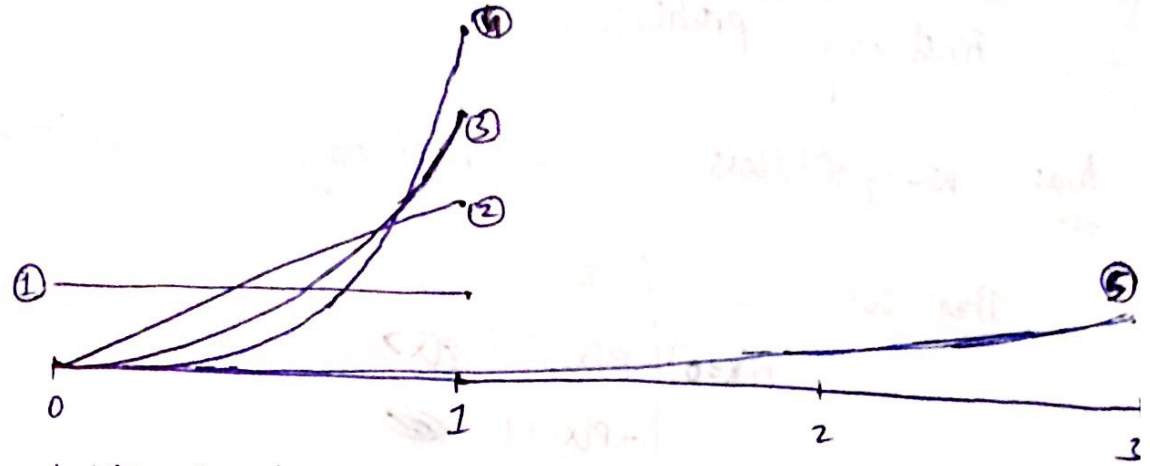
$f(x) = 2x$ $[0,1]$

$f(x) = 3x^2 [0,1]$

$f(x) = 1$ $[0,1]$

all are valid pdfs.

Graphs :



• Over [0,1], the probability of picking any number according to the various pdfs varies.

Ex:- (i) For $f(x)=1$, the probability is same from $x \in [0,1]$

(ii) For $f(x)=2x$, $f(x)=3x^2$ & $f(x)=4x^3$; the probability of picking increases for numbers from [0.5,1] than [0, 0.5).

* $p(x=h) = p(h-\varepsilon < x < h+\varepsilon)$ [ε is very small]

Q: Any wire is generally cut between 10–12 mm.

If wire is between 10.5–11.5 mm, it can be sold at a profit of 200 rupees.

If wire >11.5 but <12, it can be sold at a point of 100 ruppes

If wire < 10.5, then it results in a loss of Rs. 50.

On an average, what is the profit ?

Ans: $-\int_{10}^{10.5} 50\,dx + \int_{10.5}^{11.5} 200\,dx + \int_{11.5}^{12} 100\,dx$

$= -25 + 200 + 50$

$= 225$

(OR) W (profit): −50 100 200

$p(x<10.5)$ $\frac{1}{2}$ $\frac{1}{4}$

$= \frac{0.5}{2}$

$= \frac{1}{4}$

~~Avg = $-\frac{50}{4} + \frac{100}{4} + \frac{200}{4}$~~

Q: P → probability of breaking down (5 day week)

Runs for 5 days → profit = S

Breaks in ≤ 2 days → Profit = R

Breaks in ≥ 3 days → Loss = L

Find avg. profit/loss ?

Ans: $W \to$ profit/loss $\qquad X \to$ no. of days in which it was broken down

Then W:

S	R	$-L$
$P[X=0]$	$P[X=1] + P[X=2]$	$P[X \geq 3]$

$$\boxed{\begin{aligned} Avg. = {} & S(1-p)^5 + 5C_1 R p(1-p)^4 + R\,5C_2\, p^2(1-p)^3 + \cancel{P[X \geq 1] - P[X=0]} \\ & L - L\,5C_1\, p(1-p)^4 \\ & - L\,5C_2\, p^2(1-p)^3 \\ & - L(1-p)^5 \end{aligned}}$$

~~Avg~~

Expectation Values :-

$$\boxed{E[X] = \sum_{i=1}^{n} t_i\, p(t_i) = \int_{-\infty}^{\infty} t\, f_X(t)\, dt}$$

Ex:- (i) $f(x) = 1,\ [0,1] \longrightarrow E[X] = \int_0^1 x\, dx = \frac{1}{2}$

(ii) $f(x) = 2x,\ [0,1] \longrightarrow E[X] = \int_0^1 2x^2\, dx = \frac{2}{3}$

(iii) $f(x) = 3x^2,\ [0,1] \longrightarrow E[X] = \int_0^1 3x^3\, dx = \frac{3}{4}$

(iv) $f(x) = 4x^3,\ [0,1] \longrightarrow E[X] = \int_0^1 4x^4\, dx = \frac{4}{5}$

II] $X \sim B(n,p)$ [Binomial] :-

$$\mathcal{E}[X] = \sum_{k=0}^{n} k \cdot p(k)$$

$$= \sum_{k=0}^{n} k \cdot {}^{n}C_k \, p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} k \cdot \frac{n!}{(n-k)!\,(k+1)!} p^k (1-p)^{n-k}$$

Let $s = k-1$

$$= \sum_{s=0}^{n-1} \frac{n \cdot (n-1)!}{(n-1-s)!\, s!} p^s \cdot p \, (1-p)^{(n-1)-s}$$

$$= np \left(\sum_{s=0}^{n-1} \cancel{\frac{(n-1)!}{(n-1-s)!\, s!} p^s (1-p)^{n-1-s}} \right)^{1}$$

$$= np$$

$$\therefore \boxed{\mathcal{E}[X] = np \text{ for Binomial}}$$

* $\boxed{\mathcal{E}[aX] = a \sum_{i=0}^{n} t_i \, p(t_i)}$

$\boxed{\mathcal{E}[b] = b}$

---

* Ex :- $f(x) = 1/2$, $[-1, 1]$

$$\mathcal{E}[X] = \int_{-1}^{1} x \left(\tfrac{1}{2}\right) dx = \frac{1}{2} \frac{x^2}{2} \Big|_{-1}^{1} = 0$$

Ans: $\int_{-1}^{1} f(x)\,dx = \int_{-1}^{0}(1+x)\,dx + \int_{0}^{1}(1-x)\,dx$

$$= x+\frac{x^2}{2}\Big]_{-1}^{0} + x-\frac{x^2}{2}\Big]_{0}^{1}$$

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

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

Now, $E[X] = \int_{-1}^{1} x f(x)\,dx = \int_{-1}^{0}(x+x^2)\,dx + \int_{0}^{1}(x-x^2)\,dx$

$$= \frac{x^2}{2}+\frac{x^3}{3}\Big]_{-1}^{0} + \frac{x^2}{2}-\frac{x^3}{3}\Big]_{0}^{1}$$

$$= \left[0-\left(\frac{1}{2}-\frac{1}{3}\right)\right] + \left[\left(\frac{1}{2}-\frac{1}{3}\right)-0\right]$$

$$= \frac{-1}{6}+\frac{1}{6}$$

$$= 0$$

As $\int_{-1}^{1} f(x)\,dx = 1$, it is valid pdf.

$X \sim Exp(\lambda) \rightarrow$ Time to failure

$T = R^{\geq 0}$

$\rightarrow$ Continuous Random variable

$$\boxed{f_x(t) = \lambda e^{-\lambda t}, t \geq 0}$$

$$F_x(t_0) = \int_{-\infty}^{t_0} \lambda e^{-\lambda t} dt$$

$$E(x) = \int_{-\infty}^{\infty} (t) \lambda e^{-\lambda t} dt$$

$$= \int_{0}^{\infty} \lambda(t) e^{-\lambda t} dt$$

$$= -\lambda \frac{e^{-\lambda t}}{\lambda^2} (t\lambda - 1) \Big|_0^{\infty} = \frac{1}{\lambda} \quad \boxed{E(x) = \frac{1}{\lambda}}$$

$$\boxed{F_x(t_0) = 1 - e^{-\lambda t_0}}$$

* $X \sim G(P)$ [Geometric] :-

$$P(X = k) = p + p(1-p) + p(1-p)^2 \ldots + p(\ldots$$

$$\boxed{E(X) = \frac{1}{P}}$$

* $X \sim B(n, P)$:-

$$\left.\begin{matrix} n \to \infty \\ p \to 0 \end{matrix}\right\} \boxed{np = \lambda}$$

$$p(x = k) = {}^nC_k \, p^k (1-p)^{n-k}$$

$$= \frac{n}{(n-k)!\,k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n}{(n-k)!\,k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

$$= \left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right) \cdots \left(\frac{n-k+1}{n}\right) \frac{(\lambda)^k}{k!} \left(\frac{-\lambda}{n}\right)^n \underbrace{\left(1 - \cancel{\frac{\lambda}{n}}\right)^{-k}}_{1}$$

$$= 1\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\left(1 - \frac{3}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n$$

$$= e^{-\lambda} \frac{\lambda^k}{k!}$$

Exam Question from Oliver. C. Ibe

Anybody comes is Poisson Distribution

Interval in which person comes is called Exponential Distribution

* $X \sim U(a,b)$ uniform function $f_X = \frac{1}{b-a}$, (a,b) $\mu_X = \frac{a+b}{2} = E(X)$

$X \sim Exp(\lambda)$ $f_X = \lambda e^{-\lambda x}$, $x \geq 0$ $\mu_X = \frac{1}{\lambda} = E(X)$

* The exponential distribution has memory-less property i.e. $P(X > s+t \mid X > t) = P(X > s)$

i.e. The fact that it hasn't failed until now has no effect on how it performs later.

Functions of Random Variables :-

$Y = H(X)$

$X \rightarrow$ discrete random variables (pmf exists)

Let $Y = 3X+1$

Now, $P[Y>10] = P[3X+1>10] = P[X>3]$

$Y>10$ iff $X>3$

$\Rightarrow$ $P[Y>10]$ & $P[X>3]$ are equivalent.

As 'X' is discrete R.V, 'Y' is also a discrete R.V.

Ex :- X:

-1	0	1
1/3	1/6	1/2

$Y = X^2$:

0	1
1/6	$\frac{1}{3} + \frac{1}{2} = \frac{5}{6}$

let $X = \{x_1, x_2, \dots x_m\}$ & $Y = \{y_1, y_2, \dots, y_n\}$

Now $P[Y=y_i] = P[H(X)=y_i]$

$$\Rightarrow \boxed{P[Y=y_i] = P[\{x_j \mid H(x_j)=y_i\}]}$$

$$\Rightarrow \boxed{P[Y=y_i] = \sum_{H(x_j)=y_i} p(x_j)}$$

Now, let $p(X=k) = \frac{1}{2^k}$; $k=1,2,\dots$ [example]

$$\Rightarrow Y = \begin{cases} 1, & x \text{ is even} \\ -1, & x \text{ is odd} \end{cases}$$

$P[Y=1] = P[X=2] + P[X=4] + \dots$

$\Rightarrow P[Y=1] = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \dots = \frac{1}{2^2}\left[\frac{1}{1} + \frac{1}{2^2} + \frac{1}{2^4} + \dots\right]$

$\Rightarrow P[Y=1] = \frac{1}{2^2} \frac{1}{1-\frac{1}{4}} = \frac{1}{4} \times \frac{4}{3} = \frac{1}{3}$

$\therefore P[Y=1] = \frac{1}{3}$

• Finding pdf of Y, $f(X) = 2X$

(i) Step ①:- Find cdf

1) $G(y) = P[Y \le y] = P[3X+1 \le y] = P\left[X \le \frac{y-1}{3}\right]$

$= F\left(\frac{y-1}{3}\right) = \int_0^{\frac{y-1}{3}} 2x\,dx = \frac{(y-1)^2}{9}$

Now $g(y) = G'(y) = \frac{2(y-1)}{9}$ ~~[illegible]~~

(OR) 2) $Y = H(x) = e^{-x}$

$$G(y) = P[Y \le y] = P[e^{-x} \le y] = P[-x \le \ln y] = P[x \ge -\ln y]$$

$$= 1 - F(-\ln y) = 1 - \int_0^{-\ln y} 2x\,dx = 1 - (\ln y)^2$$

$$\therefore g(y) = G'(y) = \frac{-2\ln y}{y}, \quad y \in (1/e, 1)$$

Ex:- $Y = H(X)$

$$G(y) = P[Y \le y] = P[H(X) \le y] = P[X \le H^{-1}(y)]$$

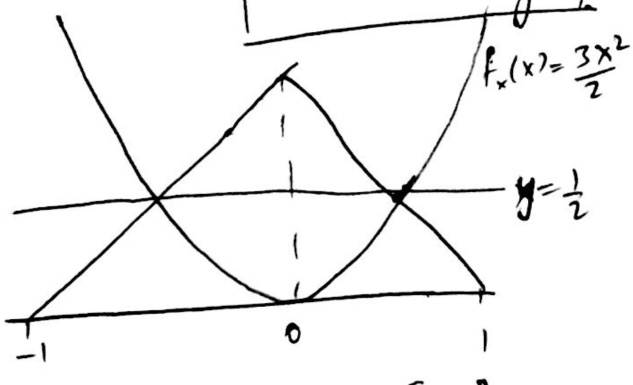
$$= F(H^{-1}(y)) = F(u)$$

where $\boxed{u = H^{-1}(y)}$

$$\therefore g(y) = \frac{d\,G(y)}{dy} = \frac{d\,F(u)}{dy}\cdot\frac{du}{dy} = f(u)\cdot\frac{du}{dy}$$

$$\therefore \boxed{g(y) = f(u)\cdot\frac{du}{dy}}$$

Q:-



$$f(x) = \begin{cases} 1+x & ; x \in [-1, 0) \\ 1-x & ; x \in [0, 1] \end{cases}$$

Find $P[-\varepsilon < x < \varepsilon]$.

$$f_x(x) = \frac{3x^2}{2}; (-1, 1)$$

Ans: $P[-\varepsilon < x < \varepsilon] = \int_{-\varepsilon}^{0} (1+x)\,dx + \int_{0}^{\varepsilon} (1-x)\,dx$

$$= x + \frac{x^2}{2}\Big|_{-\varepsilon}^{0} + x - \frac{x^2}{2}\Big|_{0}^{\varepsilon}$$

$$= \varepsilon - \cancel{\frac{\varepsilon^2}{2}}^{0} + \varepsilon - \cancel{\frac{\varepsilon^2}{2}}^{0}$$

$$= \underline{\underline{2\varepsilon}}$$

Now using $f(x) = \frac{3x^2}{2}$

$$\therefore P[-\varepsilon < x < \varepsilon] = \int_{-\varepsilon}^{\varepsilon} \frac{3x^2}{2} dx$$

$$= \frac{x^3}{2}\Big|_{-\varepsilon}^{\varepsilon}$$

$$= \frac{\varepsilon^3}{2} + \frac{\varepsilon^3}{2} = \cancel{\varepsilon^3}^{0} = \underline{\underline{0}}$$

Using $f(x) = \frac{1}{2}$

$$P[-\varepsilon < x < \varepsilon] = \int_{-\varepsilon}^{\varepsilon} \frac{1}{2} dx$$

$$= \frac{x}{2}\Big|_{-\varepsilon}^{\varepsilon}$$

$$= \frac{\cancel{2}\varepsilon}{\cancel{2}} = \underline{\underline{\varepsilon}}$$

$\underline{\underline{Q}}$: Can continuous random variable have an improper expectation i.e. $E(x) = \infty$?

$\boxed{\text{exam question}}$

* Markov's Inequality -

If $x \geq 0$ & $E[x]$ exists & $a > 0$

then $P[x > a] \leq \frac{E[x]}{a}$

Pf - $E[x] = \int_0^\infty x f(x) dx \geq \int_a^\infty a f(x) dx \geq \int_a^\infty a f(x) dx$

$= a \int_a^\infty f(x) dx$

$\Rightarrow E[x] \geq a \int_a^\infty f(x) dx$

$\Rightarrow \boxed{\frac{E[x]}{a} \geq P[x > a]}$

$\therefore$ HTP

* μ_x is mean

$\Rightarrow P[x > \mu_x] \leq \frac{\cancel{E[x]}}{\cancel{\mu_x}} = 1$

$\Rightarrow P[x > \mu_x] \leq 1$

* Variance -

How far the values are spread about 'μ'.

i.e. longterm average value of squared & centred random variable

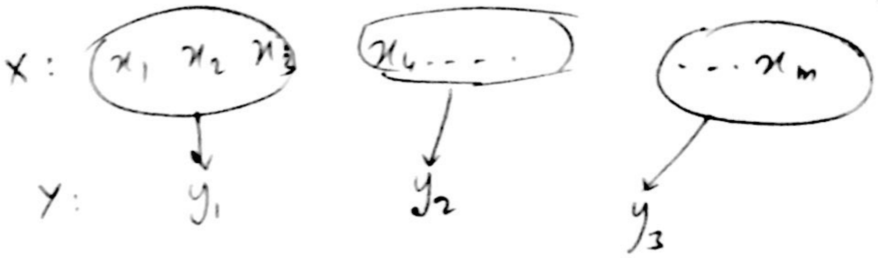
$$\boxed{Var(x) = E[(x-\mu)^2]}$$

$\Rightarrow Var(x) = E[x^2] - (E[x])^2 = E[x^2 - 2x\mu - \mu^2]$

- $E[ax+b] = aE[X] + b = E[ax] + E[b]$

- $Y = H(X)$

$$E[Y] = \sum_{i=1}^{n} y_i \, p(y_i)$$



$$y_1[p(x_1) + p(x_2) + p(x_3)] + y_2 p(y_2) + y_3 p(y_3)$$

$$= H(x_1)p(x_1) + H(x_2)p(x_2) + H(x_3)p(x_3) + \dots$$

$$\Rightarrow \boxed{E[Y] = \sum_{i=1}^{n} y_i \, p(y_i) = \sum_{j=1}^{m} H(x_j)\, p(x_j) = E[H(x)]}$$

Now, $E[X] = \sum_{j=1}^{m} x_j \, p(x_j)$

* i) $X \sim b(p)$ (Bernoulli):

$$\Rightarrow \text{var}(X) = E[X^2] - p^2 = p(1-p) = pq \qquad [q = 1-p]$$

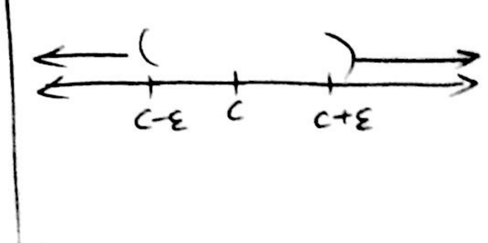
ii) $X \sim B(n,p)$ (Binomial):

Calculate yourself (Ans: npq)

---

* Chebisher's Inequality:

$$P\{|X-c| > \varepsilon\} \le \frac{E[(X-c)^2]}{\varepsilon^2}$$



We know that $\{|X-c| > \varepsilon\} \Leftrightarrow \{(X-c)^2 > \varepsilon^2\}$

$$\Rightarrow P\{|X-c| > \varepsilon\} = P\{(X-c)^2 > \varepsilon^2\}$$

From Markov's Inequality, $P\{(X-c)^2 > \varepsilon^2\} \le \frac{E[(X-c)^2]}{\varepsilon^2}$

$$\Rightarrow \boxed{P\{|X-c| > \varepsilon\} \le \frac{E[(X-c)^2]}{\varepsilon^2}}$$

$$\Rightarrow P\{|X-c| < \varepsilon\} \ge 1 - \frac{E[(X-c)^2]}{\varepsilon^2}$$

$$\therefore \boxed{P\{|X-c| < \varepsilon\} \ge 1 - \frac{E[(X-c)^2]}{\varepsilon^2}}.$$

* If $c = \mu$

$$\Rightarrow P\{|X-\mu| > \varepsilon\} \le \frac{E[(X-\mu)^2]}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}$$

$$\therefore \boxed{P\{|X-\mu| > \varepsilon\} \le \frac{Var(X)}{\varepsilon^2}}$$

$$\& \boxed{P\{|X-\mu| < \varepsilon\} \ge 1 - \frac{Var(X)}{\varepsilon^2}}$$

} only if μ exists for the R.V.

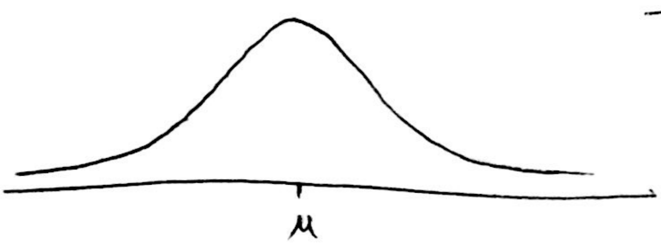
---

* Normal Distribution:-

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

• 'x' is said to be normal/gaussian distribution if $\boxed{\underbrace{f(x)}_{\text{pdf}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; x \in}$

• Graph →



We need to prove the following -

i) Is it a valid pdf ? -

$\Rightarrow$ We have to show $\int_{-\infty}^{\infty} f(x)\, dx = 1$

Let $t = \frac{x-\mu}{\sigma}$ $\Rightarrow$ $dt = \frac{1}{\sigma} dx$

$$\therefore \int_{-\infty}^{\infty} (f(x)\, dx) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = I$$

$$\Rightarrow I^2 = \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt\right) \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}s^2} ds\right)$$

$$\Rightarrow I^2 = \frac{1}{2\pi} \int_{t=-\infty}^{\infty} \int_{s=-\infty}^{\infty} e^{-\frac{1}{2}(t^2+s^2)} ds\, dt$$

Let $t = r\cos\theta$, $s = r\sin\theta$

$\Rightarrow$ $ds\, dt = r\, dr\, d\theta$

$$\Rightarrow I^2 = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} e^{-\frac{1}{2}r^2} \cdot r\, dr\, d\theta$$

Let $u = \frac{+1}{2} r^2$

$\Rightarrow$ $du = r\, dr$

$$\Rightarrow I^2 = \frac{1}{2\pi} \int_{0}^{2\pi} \left(e^{-u}\Big|_{0}^{\infty}\right) d\theta = \frac{1}{2\pi} \int_{0}^{2\pi} 1\, d\theta$$

$$\Rightarrow I^2 = \frac{2\pi}{2\pi} = 1$$

$\Rightarrow I^2 = 1$

$\Rightarrow I = \pm 1$

but area can't be $-ve$.

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

$\therefore$ It is a valid pdf.

Note:- Area under pdf from $-\infty$ to μ is $\frac{1}{2}$

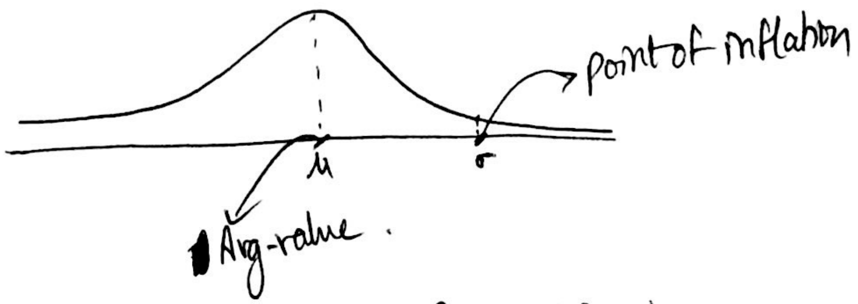
& from μ to ∞ is $\frac{1}{2}$ [$\because$ graph is symmetric about μ]

(ii) What do 'μ' & 'σ^2' stand for?:-

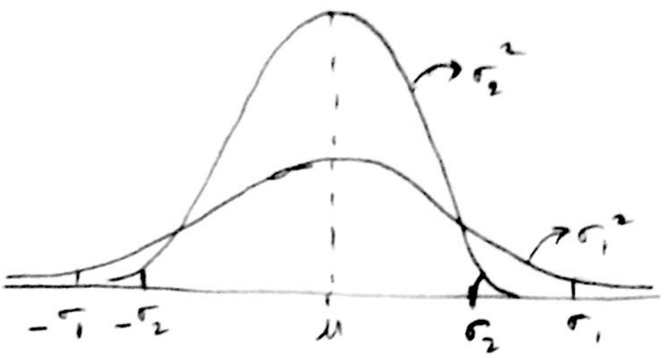
$$\boxed{\begin{aligned} \mu &= E(X) \\ \& \quad \sigma^2 &= Var(X) \end{aligned}}$$

- 'μ' may be $+ve$ or $-ve$
- 'σ' is $+ve$

- Also, the point of inflation (change of graph from concave to convex) occurs at 'σ'.



- Let $X \sim N(\mu, \sigma_1^2)$ & $X \sim N(\mu, \sigma_2^2)$ be two distributions &



Out of the above two distributions, σ_2^2 is generally preferred over σ_1^2 as the values are concentrated closer to 'μ'.

- Standard Normal Distribution :- $Z \sim N(0,1)$ [i.e. $\mu = 0$ & $\sigma^2 = 1$]

$$\boxed{f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}}$$

$$F_Z(s) = \int_{-\infty}^{s} f_Z(t)\,dt = \Phi(s) \; ; \; s \in (-\infty, \infty)$$

Now, $\boxed{1 - \Phi(s) = \Phi(-s)}$

$\therefore$ $F_Z(s)$ can be found only for $s \in (0, \infty)$ because if we know $\Phi(s)$, we can find out $\Phi(-s)$.

Ex:- $X \sim N(0,1)$

$$P\{|X - 0| > 3\} = P\{|X - \mu| > 3\sigma\} \le \frac{\sigma^2}{9\sigma^2} \le \frac{1}{9} \approx 0.11 \text{ (Chebishev's inequality)}$$

$\Rightarrow$ Only 11% of the area is below '-3' & above '3'

Other 89% of the area lies between -3 to 3

Now, $P\{|X - 0| > 4\} \le \frac{1}{16} \approx 0.06$

Almost, 94% of area lies between -4 to 4

$\therefore$ $\Phi(1) = 0.8413$, $\Phi(2) = 0.9774$, $\Phi(3) = 0.99$[illegible], Φ[illegible]

* $X \sim N(\mu, \sigma^2)$ $\qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

let $Y = aX + b = H(X)$

$H(X)$ is invertible here

So, $u = \frac{y-b}{a}$

$\Rightarrow g(y) = f(u)\left|\frac{du}{dy}\right|$

$$\Rightarrow g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)^2} \left|\frac{1}{a}\right|$$

$$\Rightarrow g(y) = \frac{1}{\sigma|a|\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{y-(a\mu+b)}{a\sigma}\right]^2}$$

let $a\mu + b = \mu^*$ & $a\sigma = \sigma^*$

$$\Rightarrow \boxed{g(y) = \frac{1}{\sigma^*\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu^*}{\sigma^*}\right)^2}}$$

$\therefore$ $X \sim N(\mu, \sigma^2) \equiv Y \sim N(\mu^*, \sigma^{*2}) \equiv Y \sim N(a\mu+b, a^2$

$\& b = -a\mu$

$$\boxed{b = \frac{-\mu}{\sigma}}$$

$$\therefore \boxed{Y = \frac{X-\mu}{\sigma}}$$

* Now $P[a<X] = P[a-\mu < X-\mu]$

$\& \; P[a-\mu < X-\mu] = P\left[\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma}\right]$

$$= P\left[\frac{a-\mu}{\sigma} < Y\right]$$

where $Y = \frac{X-\mu}{\sigma}$

Let $\frac{a-\mu}{\sigma} = s$

$$\boxed{P[a<X] = P\left[\frac{a-\mu}{\sigma} < Y\right] = P[s<Y] = 1-\Phi(s)}$$

Q: 'X' denotes tensility of a material such that $X \sim N(165, 9)$. If tensility is less than '162', then it is defective. What is the probability that an item is defective?

Ans: We need to find $P[X<162]$

$$P[X<162] = P\left[\frac{X-165}{3} < \frac{162-165}{3}\right]$$

$$= P[s<-1] = \Phi(-1)$$

$$P[X<162] = \Phi(-1) = 1-\Phi(1) = 1-0.8413$$

$$\boxed{P[X<162] = 0.1587}$$

* $\Phi(z)$ is same for any distribution as it is derived from standard normal distribution.

Q: In the previous question, find $P[X<0]$.

Ans: $P[X<0] = P\left[\frac{X-165}{3} < \frac{0-165}{3}\right] = P[z<-55] = \Phi(-55)$

$, P(X<0) = 1-\Phi(55) \simeq 0$

$(\because \Phi(55) \simeq 1)$

$\therefore$ It proves that tensility can't be -ve.

Actually, $X \in (-\infty, \infty)$
But here $X \in (0,\infty)$ [approximately]
So, the distribution automatically adjusts so as to not violate the property

---

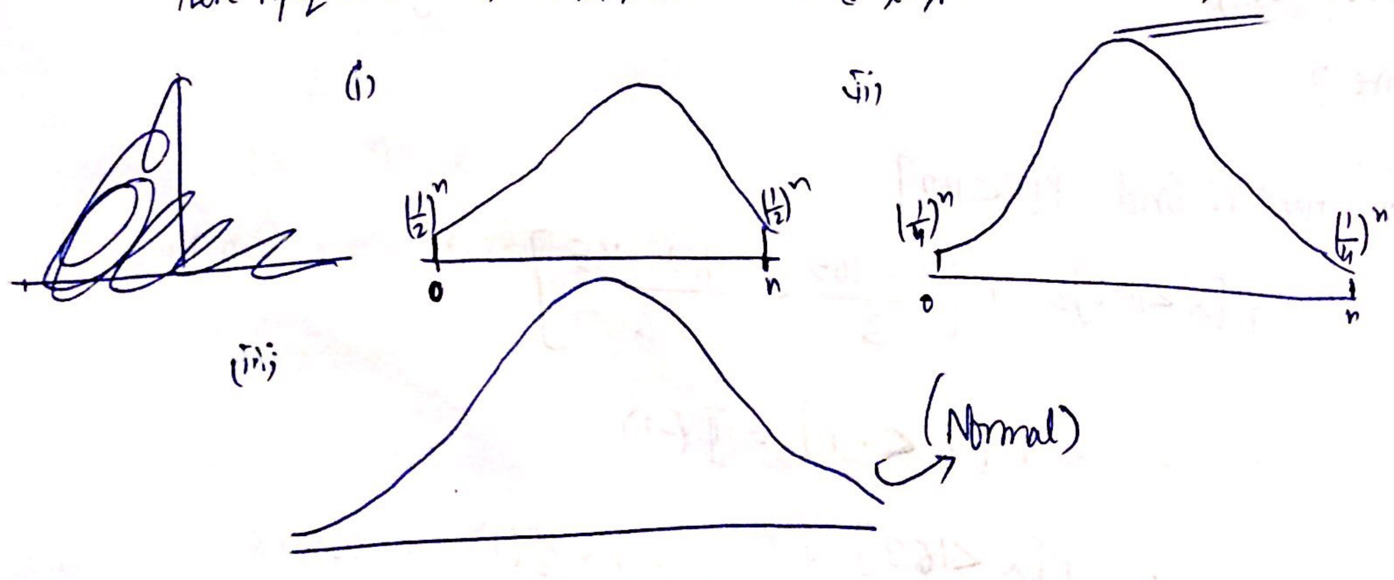
Note:- We need not worry about the values taken up by 'X' while defining a normal distribution as it automatically adjusts so that it doesn't happen

---

De Moivre's Theorem :-

In Binomial Distribution, if $npq >>> 1$, then the binomial distribution closely replicates a normal distribution.

Here $npq = \sigma^2$ " $\sigma^2 >>> 1 \Rightarrow X \sim B(n,p) \simeq X \sim N(\mu, \sigma^2)$



Note:- Generally normal approximates Binomial if $\sigma^2 \geq 10$

Q: $P[X=20] = {}^{40}C_{20}\left(\frac{1}{2}\right)^{40} = 0.125$

Now $npq = 40\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 10 = \sigma^2$

$X \sim N(20,10)$ approximates $X \sim B(40, \frac{1}{2})$

Find $P[19.5 < X < 20.5]$

Ans: $P[19.5 < X < 20.5) = P\left[\frac{19.5-20}{\sqrt{10}} < z < \frac{20.5-20}{\sqrt{10}}\right]$

$= \Phi(0.16) - \Phi(-0.16]$

$= 2\Phi(0.16) - 1$

$= 2(0.563) - 1$

$= 1.126 - 1$

$= 0.126 \simeq 0.125 = P[x=20]$