GATE

DATA SCIENCE + CS & IT

Engineering Mathematics

SUPER 1500

Lec : 02 Statistics-1

By - Dr. Puneet Sharma Sir

Topics *to be covered*

STATISTICS - I

1. p.d.f & c.D.f
2. Exponential Dist
3. Uniform Dist
4. Normal Dist
5. Correlation - Regression

CONTINUOUS RANDOM VARIABLE →

(1) $f(x) \geq 0$ (2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

(3) $F(x) = \int_{-\infty}^{x} f(x)\,dx$ & $f(x) = \frac{d}{dx} F(x)$, $\boxed{0 \leq F(x) \leq 1}$

(4) $P(a < x < b) = \int_{a}^{b} f(x)\,dx = F(b) - F(a)$

(5) $\text{Mean}(x) = E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$ (7) $Var(x) = E(x^2) - E^2(x)$

(6) $E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$ (8) $\sigma = +\sqrt{var(x)}$

$P_x(X) = Me^{(-2|x|)} + Ne^{(-3|x|)}$ is the probability density function for the real random variable X, over the entire x-axis, M and N are both positive real numbers. The equation relating M and N is

(a) $M + \frac{2}{3}N = 1$ (b) $2M + \frac{1}{3}N = 1$

(c) $M + N = 1$ (d) $M + N = 3$

$$f(x) = Me^{-2|x|} + Ne^{-3|x|}$$

$$= \begin{cases} Me^{2x} + Ne^{3x}, & x < 0 \\ Me^{-2x} + Ne^{-3x}, & x > 0 \end{cases}$$

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

$$\left(M\frac{e^{2x}}{2} + N\frac{e^{3x}}{3}\right)_{-\infty}^{0} + \left(\frac{Me^{-2x}}{-2} + \frac{Ne^{-3x}}{-3}\right)_{0}^{\infty} = 1$$

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

$$M + \frac{2N}{3} = 1$$

(a)

Let X be a random variable with probability density function

$$f(x)=\begin{cases}0.2, & \text{for } |x|\le 1 \\ 0.1, & \text{for } 1<|x|\le 4 \\ 0, & \text{otherwise}\end{cases}$$

$\Rightarrow -1\le x\le 1$

$\Rightarrow (-4\le x<-1)\cup(1<x\le 4)$

The probability $P(0.5 < X < 5)$ is ______.

W.K.that;

$|x|\le a \Rightarrow -a\le x\le a$

& $|x|>a \Rightarrow x<-a$ or $x>a$

$|x|>1 \Rightarrow x<-1$ or $x>1$

$|x|\le 4 \Rightarrow -4\le x\le 4$

$$P(0.5<x<5)=\int_{0.5}^{5} f(x)\,dx=\int_{0.5}^{1}(0.2)\,dx+\int_{1}^{4}(0.1)\,dx$$

$$=0.2\left(1-\frac{1}{2}\right)+0.1(4-1)$$

$$=0.1+0.3=0.40$$

The variance of the random variable X with probability density function $f(x) = \frac{1}{2}|x|e^{-|x|}$ is

$= \text{Even func}^n$

______.

$$E(x) = \int_{-\infty}^{\infty} \underbrace{x f(x)}_{\text{odd func}^n} dx = 0$$

$$E(x^2) = \int_{-\infty}^{\infty} \underbrace{x^2 f(x)}_{\text{Even f.}} dx = 2\int_0^{\infty} x^2 \left(\frac{1}{2} x \cdot e^{-x}\right) dx$$

$$= \int_0^{\infty} (x^3 e^{-x}) dx = 6 \quad (\text{P.T.O})$$

(M-I) $\int uv\,dx = uv_1 - u'v_2 + u''v_3 - u'''v_4 + \cdots\cdots$

$$\int_0^\infty x^3 e^{-x}\,dx = \left[x^3(-e^{-x}) - 3x^2(+e^{-x}) + 6x(-e^{-x}) - 6(+e^{-x}) + 0\right]_0^\infty$$

$$= \left[(0-0+0-0) - (0-0+0-6)\right] = 6$$

(M-II) $\int_0^\infty e^{-x} x^{n-1}\,dx = \Gamma n$

$$\int_0^\infty e^{-x} x^{4-1}\,dx = \Gamma 4 = 3! = 6$$

A random variable X has a probability density function

$$f(x)=\begin{cases} kx^n e^{-x}; & x \geq 0 \\ 0; & \text{otherwise} \end{cases}$$ (n is an integer)

with mean 3. The values of {k, n} are

(a) $\left\{\frac{1}{2}, 1\right\}$ (b) $\left\{\frac{1}{4}, 2\right\}$

(c) $\left\{\frac{1}{2}, 2\right\}$ (d) {1, 2}

$\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_0^{\infty} kx^n e^{-x}\,dx = 1$

$k\int_0^{\infty} e^{-x}\cdot x^{(n+1)-1}\,dx = 1$

$\boxed{k\,\Gamma(n+1) = 1}$ → (1)

Let $n=2$, $k\,\Gamma(2+1)=1$

$k(2!)=1$

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

M-II Mean = 3 ⇒ $\int_{-\infty}^{\infty} x f(x)\,dx = 3$

$\int_0^{\infty} x\cdot kx^n e^{-x}\,dx = 3 \Rightarrow k\int_0^{\infty} e^{-x} x^{(n+2)-1}\,dx = 3 \Rightarrow \boxed{k\,\Gamma(n+2) = 3}$ → (2)

Suppose the random variable X has distribution function $F(x) = \begin{cases} 0 & x \le 0 \\ 1 - \exp(-x^2) & x > 0 \end{cases}$. What is the probability that X exceed 1?

(a) e^{-2} (b) e^{-1}

(c) e^{-3} (d) e^{-4}

$\Gamma(n+1) = \begin{cases} n!, & n \in I^+ \\ n\Gamma n, & n \in Q^+ \end{cases}$

eg $\Gamma 4 = 3!$, $\Gamma 5 = 4!$, $\Gamma 2 = 1!$

$\Gamma 1 = 0!$, $\Gamma 0 = ND$

$$P(x>1) = ? = 1 - P(x \le 1)$$
$$= 1 - P(-\infty < x \le 1)$$
$$= 1 - F(1)$$
$$= 1 - (1 - e^{-1})$$
$$= e^{-1}$$

(b)

Q. if $f(x)=\begin{cases} \frac{1}{8}, & 0<x\le 2 \\ \frac{3}{8}, & 2<x\le 4 \\ 0, & \text{otherwise} \end{cases}$ is density function for Random Variable x then it's distribution function will be?

(a) $F(x)=\begin{cases} 0, & x\le 0 \\ \frac{x}{8}, & 0<x\le 2 \\ \frac{3x}{8}+\frac{1}{4}, & 2<x\le 4 \\ 1, & x>4 \end{cases}$

(b) $F(x)=\begin{cases} 0, & x\le 0 \\ \frac{x}{8}, & 0<x\le 2 \\ \frac{3x}{8}+\frac{1}{2}, & 2<x\le 4 \\ 1, & x>4 \end{cases}$

(c) $F(x)=1 \ \forall x$

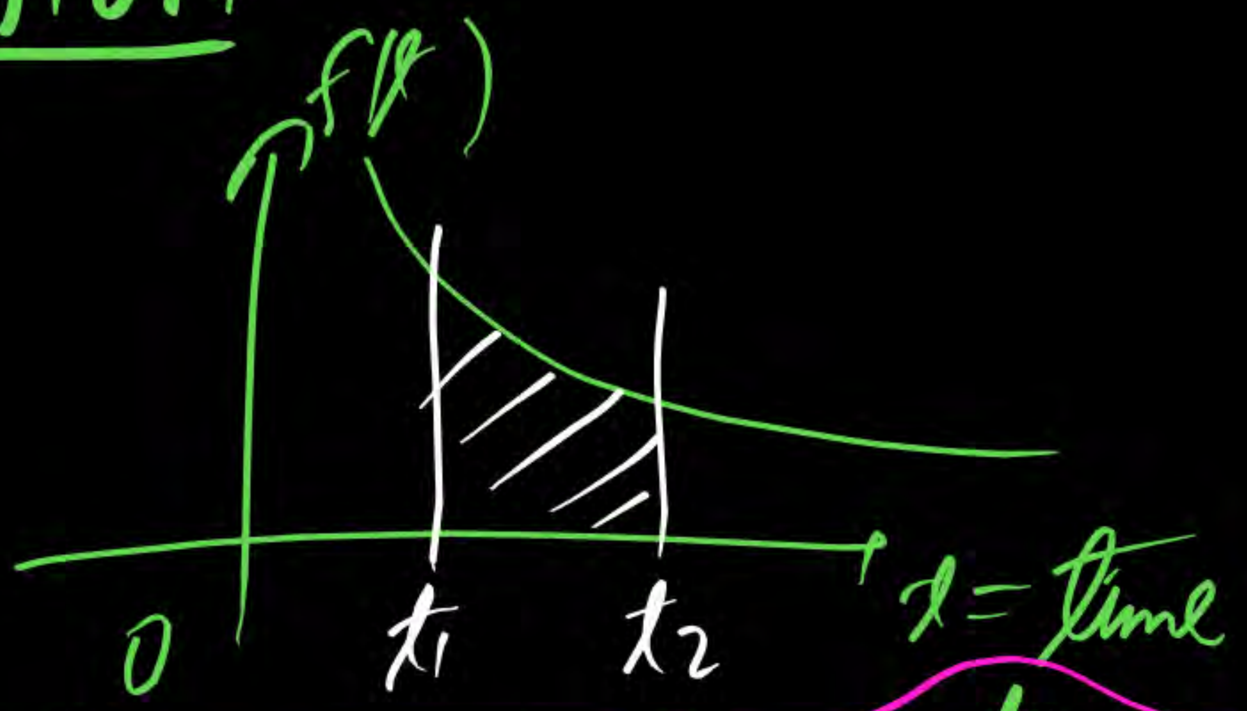
(d) $F(x)=\begin{cases} 0, & x\le 0 \\ x/8, & 0<x\le 2 \\ \frac{3x}{8}-\frac{1}{2}, & 2<x\le 4 \\ 1, & x>4 \end{cases}$

$2<x\le 4$: $F(x)=\int_{-\infty}^{x} f(x)\,dx=\int_{-\infty}^{0}(0)+\int_{0}^{2}\left(\frac{1}{8}\right)dx+\int_{2}^{x}\left(\frac{3}{8}\right)dx$

$=\left(\frac{2}{8}-0\right)+\frac{3}{8}(x-2)=\frac{1}{4}+\frac{3x}{8}-\frac{3}{4}=\boxed{\frac{3x}{8}-\frac{1}{2}}$

(d)

EXPONENTIAL DISTRIBUTION



(1) $f(t) = \begin{cases} \mu e^{-\mu t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$

(2) Mean $(t) = E(t) = \dfrac{1}{\mu} = S.D$

(3) $\mu \rightarrow$ Service Rate

& $\dfrac{1}{\mu} \rightarrow$ Av. waiting time.

(4) $P(t_1 < t < t_2) = \displaystyle\int_{t_1}^{t_2} f(t)\,dt$

(5) In P-Dist Mean = Var = λ & SD = $+\sqrt{\lambda}$

In Exp Dist Mean = SD = $\dfrac{1}{\mu}$

& Var = $\dfrac{1}{\mu^2}$

Assume that the duration in minutes of a telephone conversation follows the exponential distribution $f(x)=\frac{1}{5}e^{-x/5}, x \geq 0$. The probability that the conversation will exceed five minutes is

(a) $\frac{1}{e}$ (b) $1-\frac{1}{e}$

(c) $\frac{1}{e^2}$ (d) $1-\frac{1}{e^2}$

$f(x)=\frac{1}{5}e^{-\frac{x}{5}} \simeq \mu e^{-\mu x}$

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

$x = \{\text{length of Conversation}\}$

$P(x>5) = 1-P(0<x\leq 5)$

$= 1-\int_0^5 f(x)dx = 1-\int_0^5 \frac{1}{5}e^{-x/5}\cdot dx$

$= 1-\frac{1}{5}\left(\frac{e^{-x/5}}{-1/5}\right)_0^5 = 1+\left(e^{-1}-e^{0}\right) = e^{-1} = $ (a)

M-II

$$p(x>5) = \int_5^\infty f(x)\,dx = \int_5^\infty \frac{1}{5} e^{-x/5}\,dx$$

$$= \frac{1}{5}\left[\frac{e^{-x/5}}{-1/5}\right]_5^\infty = -\left[e^{-\infty} - e^{-1}\right] = -(-e^{-1}) = e^{-1}$$

$$= \frac{1}{e}$$

Vehicles arriving at an intersection from one of the approach roads follow the Poisson distribution. The mean rate of arrival is 900 vehicles per hour. If a gap is defined as the time difference between two successive vehicle arrivals (with vehicles assumed to be points), the probability (up to four decimal places) that the gap is greater than 8 seconds is ______.

$\lambda = 900$ veh/hr

$= \frac{900}{60}$ veh/min $= 15$ veh/min

$= \frac{15}{60}$ veh/sec $= \frac{1}{4}$ veh/sec

Av time Gap b/n two successive vehicles (in sec) $= \boxed{4 \text{ sec} = \frac{1}{\mu}}$

$x = \{$ time Gap b/n two successive vehicles in sec $\}$

$$P(x > 8) = P(8 < x < \infty) = \int_8^\infty f(x)\,dx = \int_8^\infty \mu e^{-\mu x}\,dx = ? =$$

$= 0.135$

M-II

$\lambda = \frac{1}{4}$ veh in 1 sec

$= \frac{2}{}$ veh in 8 sec

$\lambda = 2$ veh / (8 sec)

$X = \{$No. of veh in 8 sec$\}$ → success

$$P(X=0) = \frac{e^{-\lambda}\lambda^{r}}{r!} = \frac{e^{-2} \cdot 2^{0}}{0!} = e^{-2} = 0.135$$

For a single server with Poisson arrival and exponential service time, the arrival rate is 12 per hour. Which one of the following service rates will provide a steady state finite queue length?

(a) 6 per hour (b) 10 per hour

(c) 12 per hour (d) 24 per hour

Queue Theory

Traffic Intensity in Queue;

$$\rho = \frac{\text{Arrival Rate}}{\text{Service Rate}} = \frac{\lambda}{\mu}$$

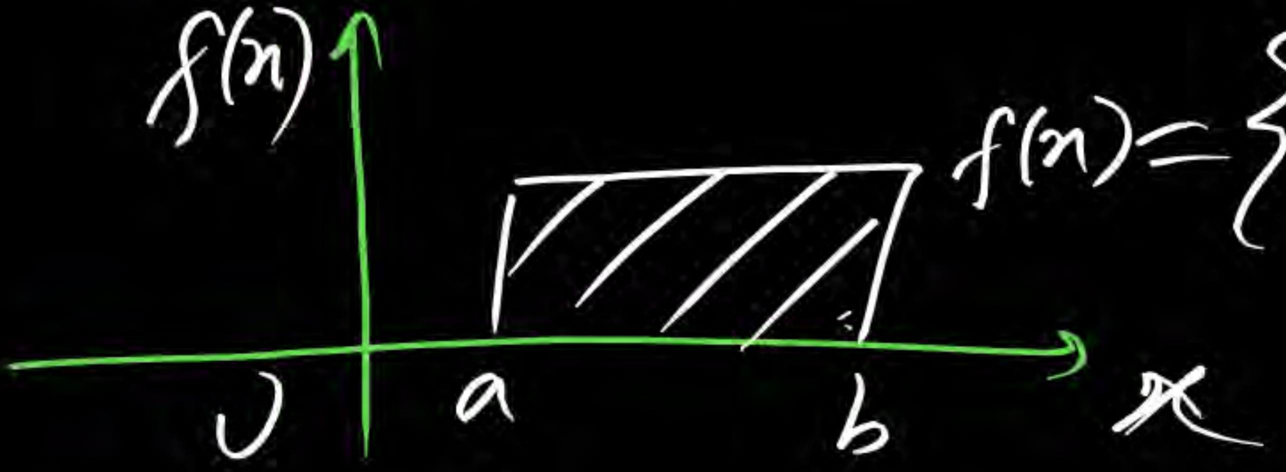
for finite Queue; $\rho < 1$

$\Rightarrow \lambda < \mu$

$\because \lambda = 12$ so $\mu > 12 \Rightarrow$ (d)

UNIFORM DISTRIBUTION

if $x \in (a,b)$ then



$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a<x<b \\ 0, & o/w. \end{cases}$$

(2) Mean $= \dfrac{a+b}{2}$, (3) $Var = \dfrac{(b-a)^2}{12}$

(4) $SD = \dfrac{b-a}{\sqrt{12}}$, (5) $P(\alpha < x < \beta) = \int_{\alpha}^{\beta} f(x)\,dx$

X is a uniformly distributed random variable that takes values between 0 and 1. The value of $E(X^3)$ will be

(a) 0 (b) 1/8

(c) 1/4 (d) 1/2

$\because x \in (0,1)$ & x is U.R.V

so $f(x) = \begin{cases} \frac{1}{1-0}, & 0<x<1 \\ 0, & \text{otw} \end{cases}$

NENO

$$E(x^3) = \int_{-\infty}^{\infty} x^3 f(x)\,dx = \int_0^1 x^3(1)\,dx$$

$$= \frac{1}{4}$$

Consider the random process

$$X(t) = U + Vt.$$

where U is a zero-mean Gaussian random variable and V is a random variable uniformly distributed between 0 and 2. Assume that U and V are statistically independent. The mean value of the random process at t = 2 is ______

(a) 2 (b) 1

(c) 0 (d) None

$E(U) = 0$ (given) & V is U.R.V in (0, 2)

$\Rightarrow E(V) = \frac{a+b}{2} = \frac{0+2}{2} = 1$

$E(X) = E(U + Vt) = E(U) + tE(V)$

$= 0 + t(1)$

i.e. $E(X) = t$

So $(E(X))_{t=2} = 2$

Suppose Y is distributed uniformly in the open interval (1, 6). The probability that the polynomial $3x^2 + 6xy + 3y + 6$ has only real roots is (rounded off to 1 decimal place) _____.

$\because y \in (1,6)$ & y is U.R.V so $f(y) = \begin{cases} \frac{1}{6-1}, & 0<y<6 \\ 0, & \text{otw} \end{cases}$

$3x^2 + (6y)x + (3y+6) = 0$

for Real Roots, Disc ≥ 0

$B^2 - 4AC \geq 0$

$(6y)^2 - 4(3)(3y+6) \geq 0$

$36y^2 - 36y - 72 \geq 0$

$y^2 - y - 2 \geq 0$

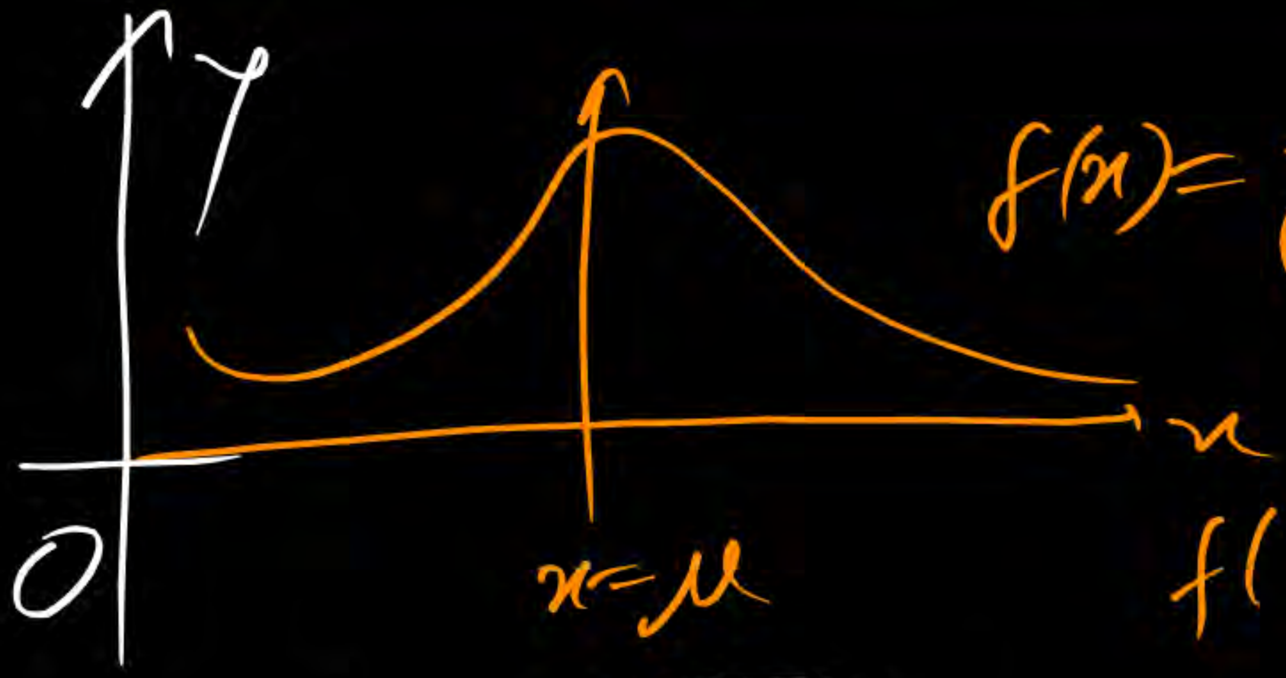
$(y-2)(y+1) \geq 0$

$y \leq -1$ or $y \geq 2$

$$P\{Y \le -1 \cup Y \geq 2\} = \int_{-\infty}^{-1} f(y)\,dy + \int_{2}^{6} f(y)\,dy$$

$$= 0 + \frac{1}{5}\int_{2}^{6} (1)\,dy = \frac{4}{5} = 0.8$$

NORMAL DISTRIBUTION



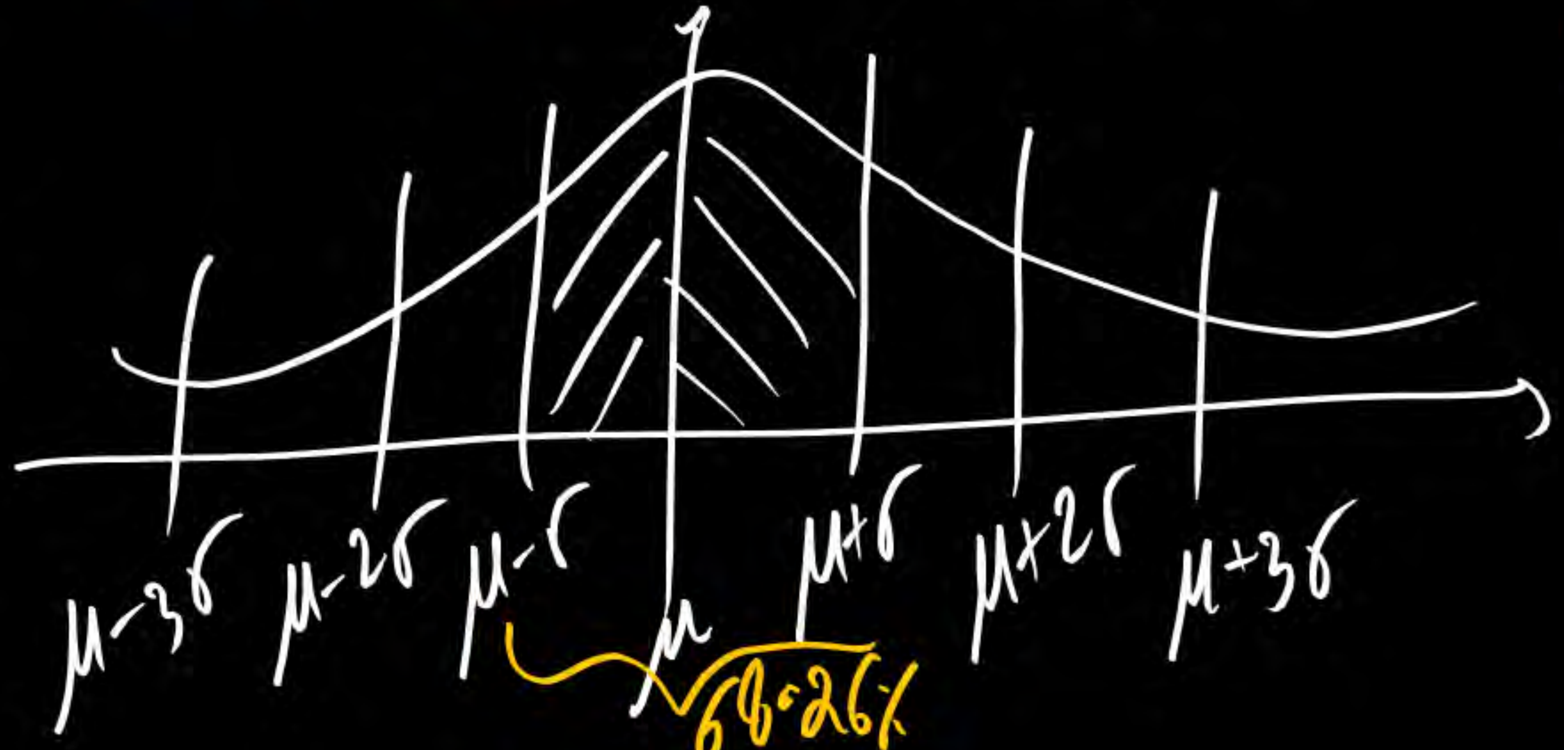
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \text{NENO}$$

i.e. $x \sim N\{\mu, \sigma^2\}$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{z^2}{2}\right)} = \text{even func}^n$$

i.e. $z \sim N\{0, 1\}$

(*) $\frac{x-\mu}{\sigma} = z$



68.26%

95.5%

99.7%

Which one of the following statements is not true?

(a) The measure of skewness is dependent upon the amount of dispersion (T)

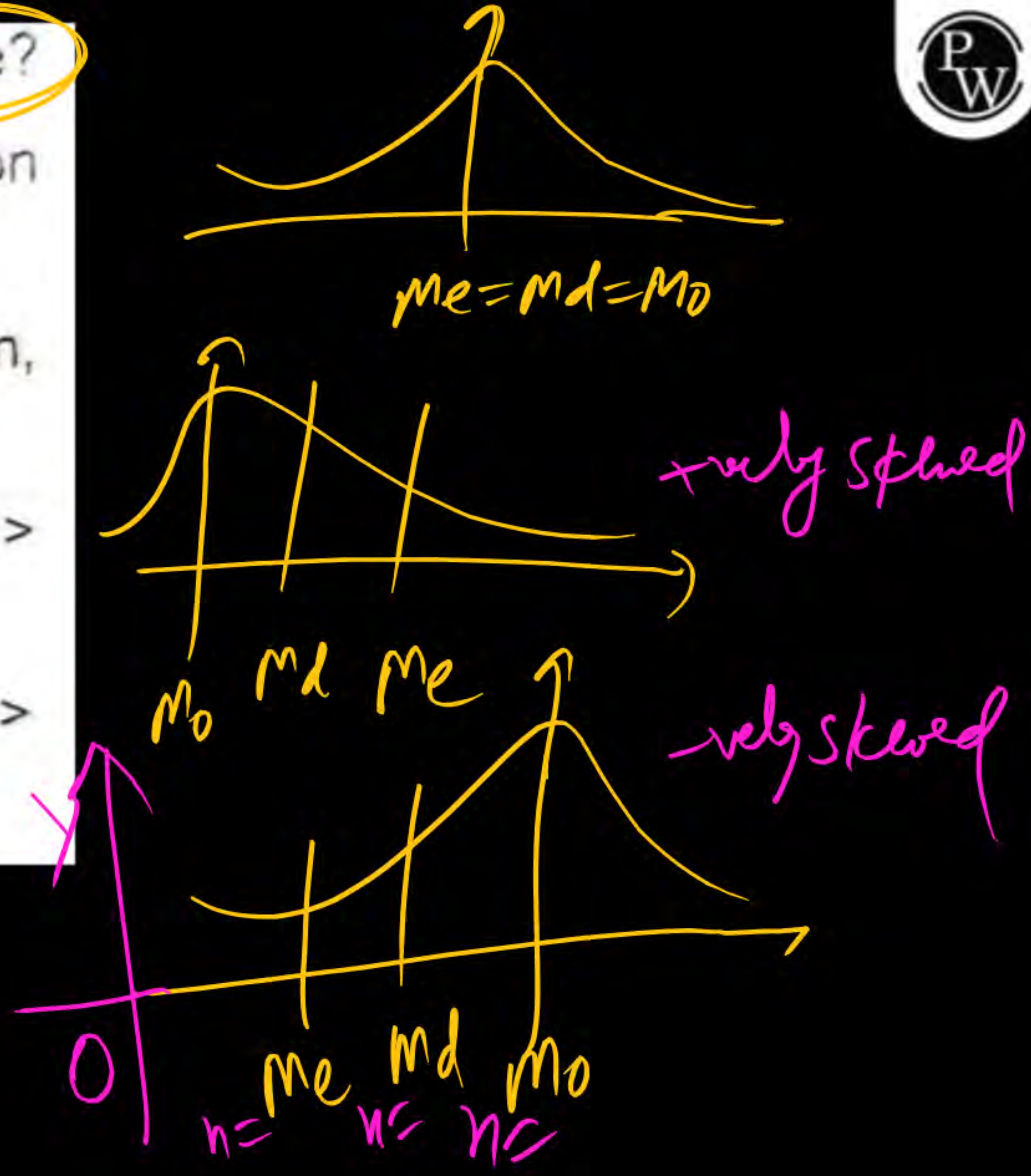
(b) In a symmetric distribution, the values of mean, mode and median are the same (T)

(c) In a positively skewed distribution, mean > median > mode (T)

(d) In a negatively skewed distribution, mode > mean > median (F)

?.?

$\because M_o > M_d > M_e$



Let X be a normal random variable with mean 1 and variance 4. The probability P(X < 0) is

(a) 0.5

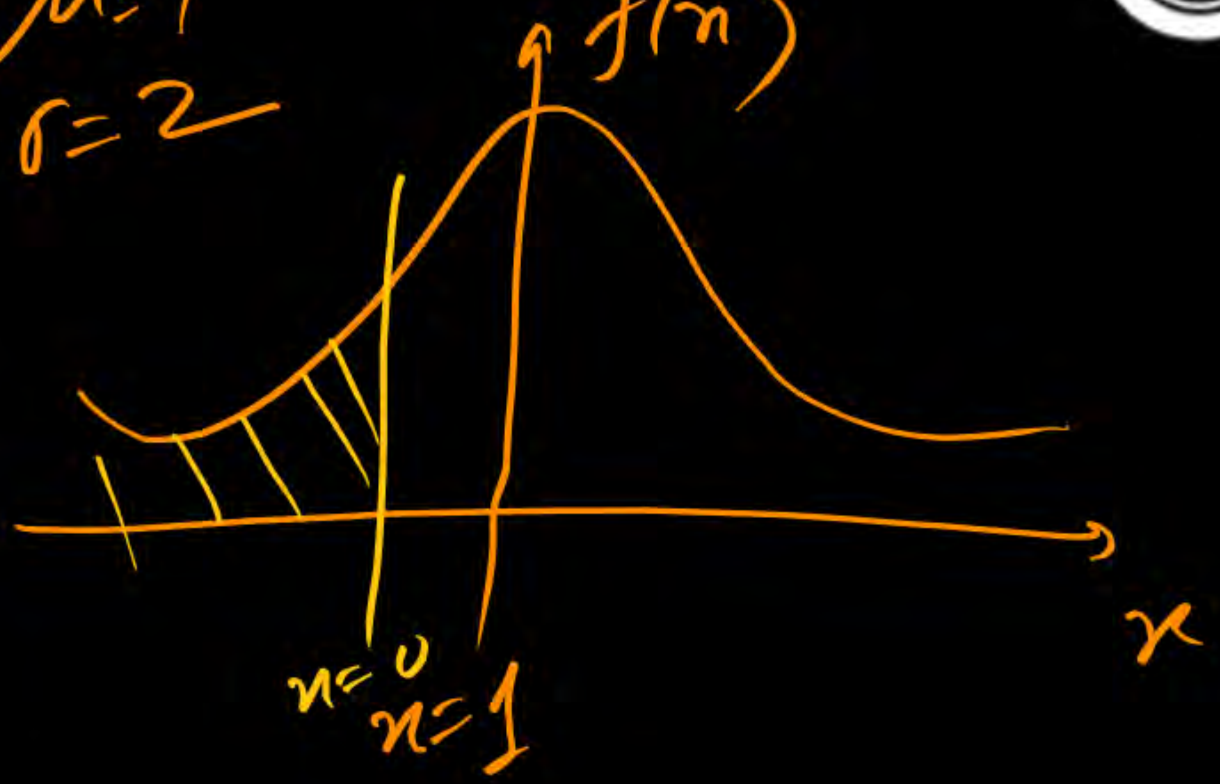
(b) greater than zero and less than 0.5

(c) greater than 0.5 and less than 1.0

(d) 1.0

$\mu = 1$

$\sigma = 2$



$P(x < 0) < 0.5$

i.e $0 < P < 0.5$ i.e (b)

A normal random variable X has the following probability density function

$$f_X(x) = \frac{1}{\sqrt{8\pi}} e^{-\left\{\frac{(x-1)^2}{8}\right\}}, -\infty < x < \infty$$

Then $\int_1^{\infty} f_X(x)\,dx$?

(a) 0

(b) $\frac{1}{2}$

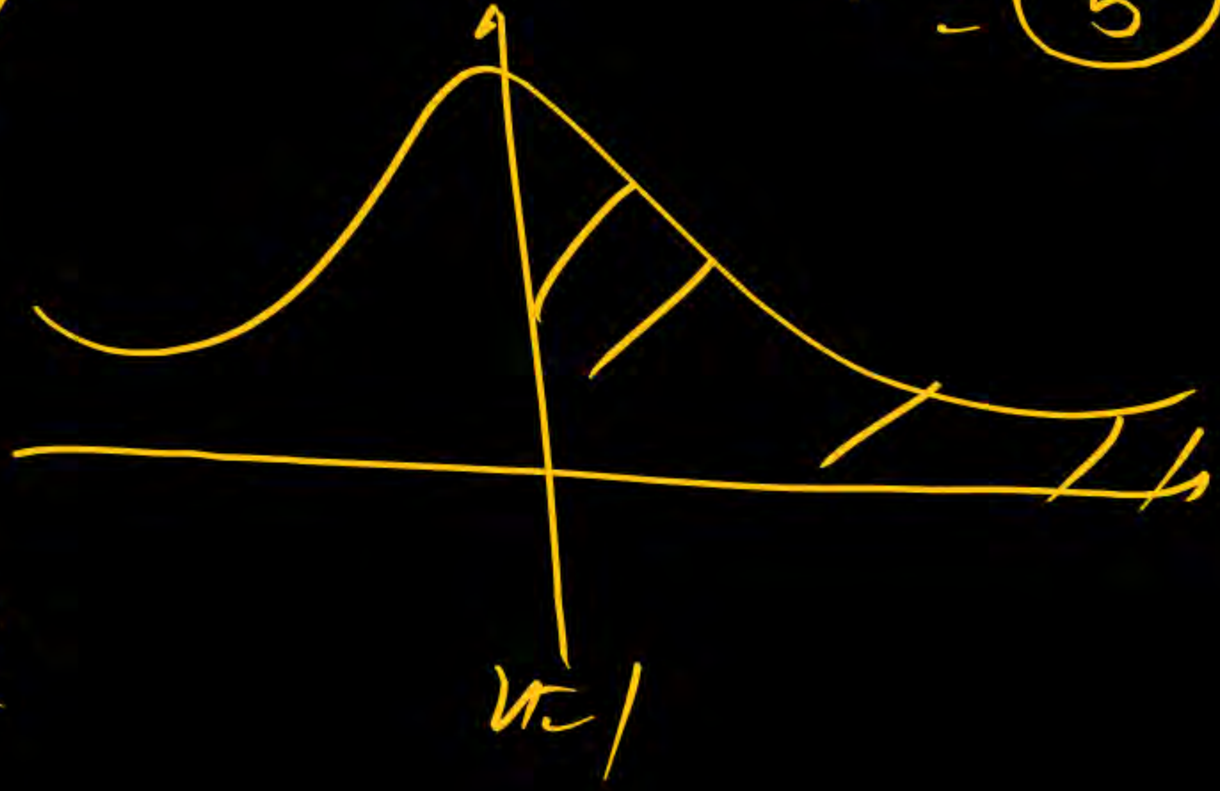
(c) $1 - \frac{1}{e}$

(d) 1

$$f(x) = \frac{1}{\sqrt{8\pi}} e^{-\frac{1}{8}(x-1)^2}$$

$$\int_1^{\infty} f(x)\,dx = ? = 0.5$$

i.e. (b)



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

$\Rightarrow \mu = 1, \sigma = 2$

Let X_1, X_2 and X_3 be independent and identically distributed random variables with the uniform distribution on $[0, 1]$. The probability $P\{X_1 + X_2 \leq X_3\}$ $= ?$ is ______.

$P(x_1+x_2 \leq x_3) = P(x_1+x_2-x_3 \leq 0) = P(X \leq 0) = P(Z_X \leq ?)$

let $\boxed{X = x_1+x_2-x_3}$, then X is N.R.V by using Normal Sum theorem.

$\because x_1, x_2, x_3$ are U.R.V in $(0,1)$ then $E(x_1) = E(x_2) = E(x_3) = \frac{a+b}{2} = \frac{1}{2}$

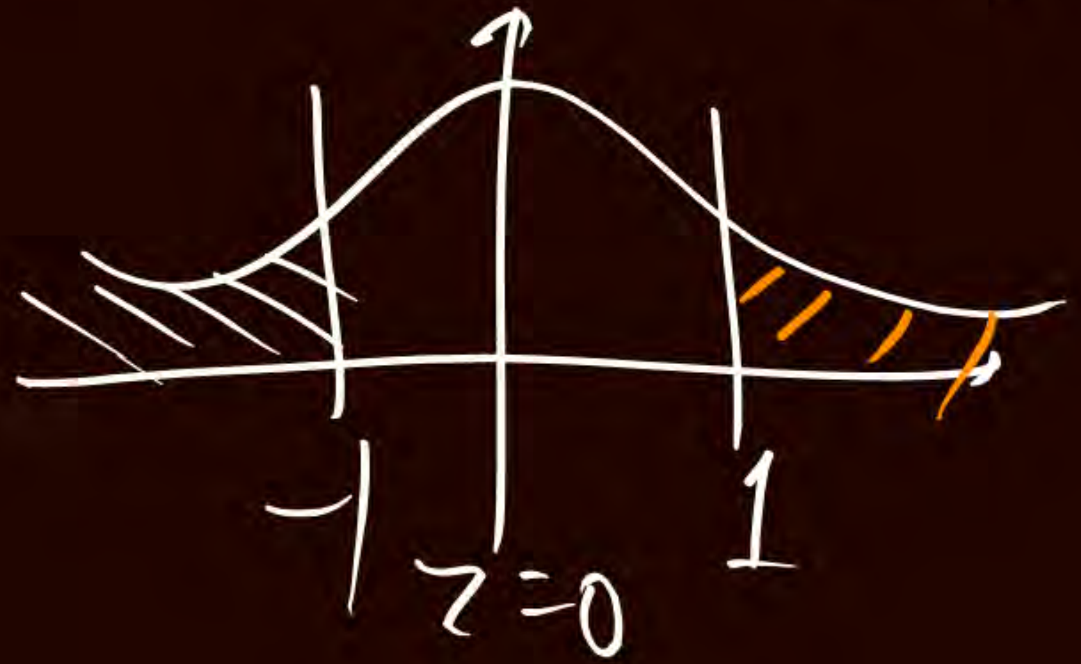
$\mu = E(X) = \frac{1}{2} + \frac{1}{2} - \frac{1}{2} = \frac{1}{2}$ & $Var(x_1) = Var(x_2) = Var(x_3) = \frac{(b-a)^2}{12} = \frac{1}{12}$

& $Var(X) = \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + 0 = \frac{1}{4}$

$X \sim N\{\mu, \sigma^2\}$

ie $X \sim N\{\frac{1}{2}, \frac{1}{4}\}$

$$Z = \frac{x - \mu_x}{\sigma_x} = \frac{0 - \frac{1}{2}}{(1/2)} = -1$$



$P(X_1 + X_2 - X_3 \le 0)$

$= P(X \le 0)$

$= P(Z \le -1)$

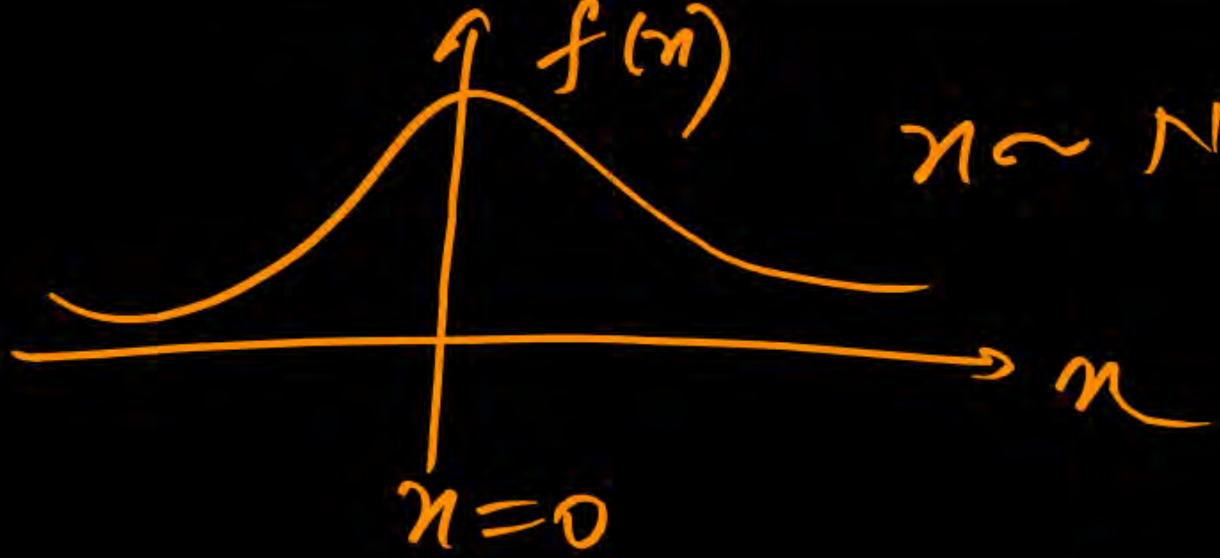
$= P(Z \geq 1)$

$= \frac{1}{2} - P(0 \le Z \le 1)$

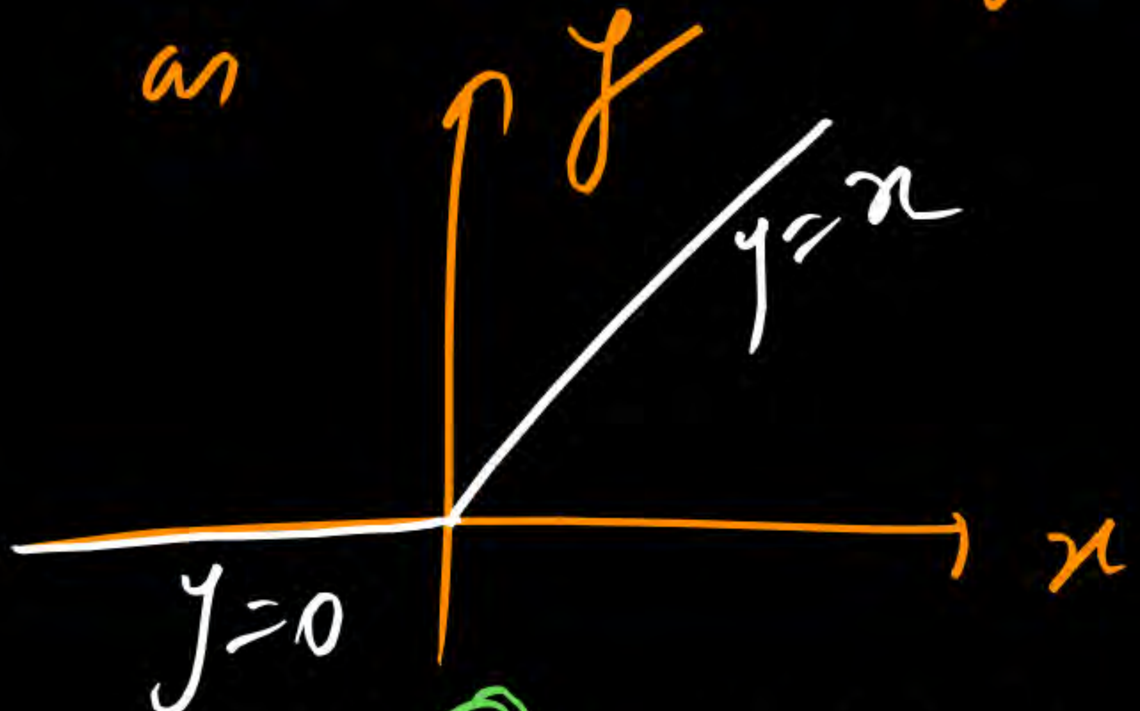
$= 0.5 - 0.3413 = 0.1587$

Let X be a Gaussian random variable mean 0 and variance σ^2. Let $\boxed{Y = \max(X, 0)}$ where max (a, b) is the maximum of a and b. The median of Y is _______________.

i.e. y is also R-V. defined as



$x \sim N\{0, \sigma^2\}$



when $x<0$, $y = \max\{x,0\} = 0$

when, $x>0$, $y = \max(x,0) = x$

$x:$ ....-2 -1 0 1 2 3 4....

$y:$ ....0, 0 0 1 2 3 4....

$\Longrightarrow$ Md of $y = 0$

Consider a binomial random variable X. If X_1, X_2, ... X_n are independent and identically distributed samples from the distribution of X with sum $Y = \sum_{i=1}^{n} X_i$, then the distribution of Y as $n \to \infty$ can be approximated as

(a) Exponential (b) Bernoulli

(c) Binomial (d) Normal

By N. Sumanth

CORRELATION & REGRESSION

1. $Cov(x,y) = E(xy) - E(x)\cdot E(y)$

2. $r = \dfrac{Cov(x,y)}{\sigma_x , \sigma_y}$, $-1 \le r \le 1$

3. $\begin{cases} y - \bar{y} = b_{yx}(x - \bar{x}) \\ x - \bar{x} = b_{xy}(y - \bar{y}) \end{cases}$, $b_{yx} = r\dfrac{\sigma_y}{\sigma_x}$, $b_{xy} = r\dfrac{\sigma_x}{\sigma_y}$

4. $r = \pm\sqrt{b_{yx}\, b_{xy}}$

For the regression equations

$$y = 0.516x + 33.73 \quad \text{— (1)}$$

and $$x = 0.512y + 32.52 \quad \text{— (2)}$$

the means of x and y are nearly

(a) 67.6 and 68.6 (b) 68.6 and 68.6

(c) 67.6 and 58.6 (d) 68.6 and 58.6

Mean = Intersecting Point of (1) & (2)

$(\bar{X}, \bar{Y}) = $ " " "

i.e. solving (1) & (2)

$\bar{x} = 67.6$

$\bar{y} = 68.6$

5 Consider the following regression equations obtained from a correlation table :

$$y = 0.516\, x + 33.73$$

$$x = 0.512\, y + 32.52$$

The value of the correlation coefficient will be

(a) 0.514 (b) 0.586

(c) 0.616 (d) 0.684

$b_{yx} = 0.516$ given

$b_{xy} = 0.512$ "

$r = +\sqrt{0.516 \times 0.512}$

$= 0.514$

Three points in the x-y plane are $(-1, 0.8)$, $(0, 2.2)$ and $(1, 2.8)$. The value of the slope of the best fit straight line in the least square sense is ______ (Round off to 2 decimal places).

$x =$	-1	0	1
$y =$	0.8	2.2	2.8

$n = 3$

$$\boxed{y = a + bx}$$

So Slope of line of Best fit $= b = ?$

$$\Sigma y = 3a + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

	x	y	Σxy	Σx^2
	-1	0.8	-0.8	1
	0	2.2	0	0
	1	2.8	2.8	1
Σ	0	5.8	2	2

$\Rightarrow$ $5.8 = 3a + 0$

$2 = \frac{5.8}{3}(0) + 2b$

$\Rightarrow b = 1$

Ans

A lot contains 20 articles. The probability that the lot contains exactly 2 defective articles is 0.4 and the probability that the lot contains exactly 3 defective articles is 0.6. Articles are drawn at random one by one without replacement and tested till all the defective articles are found. What is the probability that the testing procedure ends at the twelfth testing? ________ Sol: 0.05

Article (20)

0.4 → Articles are drawn from 2 Def lot (E_1)

0.6 → Articles are drawn from 3 def (E_2)

$$P(A) = P(E_1)P(A/E_1) + P(E_2)P(A/E_2)$$

$$=$$

A : (test ends in 12th test)

A/E_1: $(\quad) \times (?)$

A/E_2: $(\quad) \times (?)$

$$P(A/E_1) = \left(\frac{{}^{2}C_1 \times {}^{18}C_{10}}{{}^{20}C_{11}}\right) \times \frac{1}{9} \quad \text{--- (1)}$$

$$P(A/E_2) = \left(\frac{{}^{3}C_2 \times {}^{17}C_9}{{}^{20}C_{11}}\right) \times \frac{1}{9} \quad \text{--- (2)}$$

$$\text{Re Prob} = (0.4)(\text{Case I}) + (0.6) \times (\text{Case II})$$

$$= 0.054 \quad \underline{\text{Ans}}$$

Thank

you

Keep Hustling!