2 RVs ✓

n RVs → n iids

iids → independent & identically distributed

$X_1, X_2 \ldots X_n$ are iids

$$W = \sum_{i=1}^{n} X_i \longrightarrow$$ approaches a Gaussian as $n \to \infty$

↓

Central Limit thm (CLT)

Using CLT

W is approximately $N(\mu_W, \sigma_W)$

$$\mu_W = E[W] = E[\cancel{\ }X_1 + X_2 \cdots + X_n]$$

$$= n\mu_x$$

$\mu_x \rightarrow$ mean of any X_i

$\downarrow$

all iids

$$\sigma_W = ?$$

$$Var[W] = Var[X_1 + X_2 \cdots + X_n]$$

$$\left[\text{Recall: } Var[X+Y] = Var[X] + Var[Y] + \underbrace{2Cov.[X,Y]}_{\substack{=0 \text{ if } X \& Y \\ \text{are uncorrelated}}} \right]$$

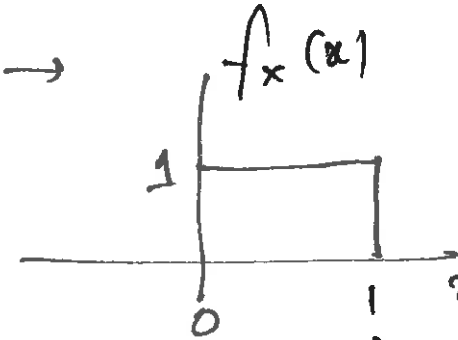
if independent $\rightarrow$ X & Y are uncorrelated.

when X_is are iids

$$Var[W] = n\, Var[X]$$

eg:- X_is $i=1,\cdots 100$ are iids

Each X_i has the pdf $\rightarrow$



Find $P\Big[\underbrace{\sum_{i=1}^{100} X_i}_{W} > 80\Big]$

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

$Var[X] = \frac{1^2}{12}$

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

Using CLT

W is $N(\mu_W, \sigma_W)$

$$\mu_W = 100\,\mu_x = 50$$

$$\sigma_W^2 = (100)\,Var[x] = \frac{100}{12}$$

$P[W>80] = ?$

$$\Phi\left(\frac{80-50}{\sigma_W}\right)$$

Find $P\left[\left|\frac{1}{100}\sum_{i=1}^{100} x_i - 0.5\right| > 0.1\right]$ ④

$$M = \underbrace{\frac{X_1 + X_2 \cdots + X_n}{n}}_{\rightarrow \text{ Sample Mean}} = \frac{W}{n}$$

$$\mu_M = \frac{1}{n} \cdot \mu_W = \mu_X$$

$$\left[\text{Recall: } \begin{array}{l} Y = aX + b \\ \mu_Y = a\mu_X + b, \quad Var[Y] = a^2 Var[X] \end{array}\right]$$

$$Var[M] = Var\left[\frac{W}{n}\right] = \frac{1}{n^2} Var[W]$$

$$= \frac{1}{n^2}\left(n \, Var[X]\right)$$

$$= \frac{Var[X]}{n}$$

$$M \text{ is } N\left(\mu_X, \frac{Var[X]}{n}\right)$$

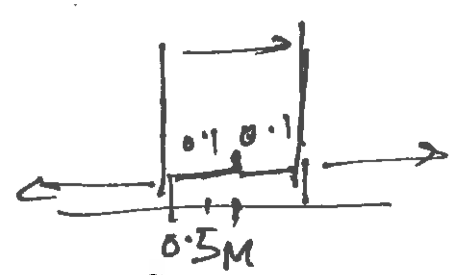
Note $Var[M] \xrightarrow[n \to \infty]{} 0$

$\therefore$ M approaches μ_x as $n \to \infty$

(As stated before under the freq. interpretation)

$$P\left[\left| \underbrace{\frac{1}{100} \sum_{i=1}^{100} x_i}_{M} - 0.5 \right| > 0.1 \right] = ?$$

$$\mu_M = \mu_x = 0.5$$



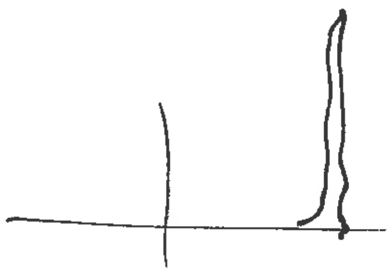
$$P\left[|M - 0.5| > 0.1 \right]$$

$$= P\left[M < 0.4 \text{ or } M > 0.6 \right]$$

M is $N(0.5, ?\sigma_M)$

~~Va[M]~~

$$\sigma_M^2 = Va[M] = \frac{Va[X]}{n} = \frac{\frac{1}{12}}{100} = \frac{1}{1200}$$



$$P[M < 0.4 \text{ or } M > 0.6]$$

$$= P[M < 0.4] + P[M > 0.6]$$

$$= 2\Phi\left(\frac{0.1}{\sigma_M}\right)$$

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

confidence interval

$c = 0.1$

p

Eg:- 0.001

with prob. $(1-p)$

0.999

find from the expt

$|M - \mu_0| \leq c$

Note in CLT

the individual pdf of X_is can be anything.

$f_x(x)$ can be $\longrightarrow$ exponential
uniform.

X_is can even be discrete.

eg:- A company manufactures 1000 products

$P[\text{Defective}] = 0.1 \longrightarrow$ for any product

products are indepd.

P[Manufactures more than 200 Defective products]

P[Manufacturing k defective products out of 1000 products]

$$= \binom{1000}{k} (0.1)^k (0.9)^{1000-k}$$

W = No. of Defective from (8)

P[Manufactuy more than 200 Defective products]

$$= 1 - P[W \leq 200]$$

$$= 1 - \sum_{k=0}^{200} \binom{1000}{k} (0.1)^k (0.9)^{1000-k}$$

$$X_i = \begin{cases} 1 \longrightarrow & \text{if } i^{th} \text{ product is Defective} \\ 0, \longrightarrow & \text{if } i^{th} \text{ product is Good} \end{cases}$$

$i = 1, \cdots 1000$

X_i's are Bernoulli $\hookrightarrow$ iids

Note: $W = \sum_{i=1}^{1000} X_i \longrightarrow$ Total number of Defective products

Using CLT,

W is approximately $N(\mu_W, \sigma_W)$

$$\mu_W = n\,\mu_X$$

Recall: if X is Bernoulli (p)

$$\mu_X = p, \quad Var[X] = p(1-p)$$

$$\mu_W = (1000)(0.1) = 100$$

$$Var[W] = \sigma_W^2$$

$$= n\,Var[X]$$

$$\sigma_W^2 = 1000\,(0.1)(0.9) \rightarrow \sigma_W$$

$$P[W > 200] = \Phi\left(\frac{100}{\sigma_W}\right)$$

Find the prob. of manufacty exactly 100 defective products

$$= \binom{1000}{100} (0.1)^{100} (0.9)^{900}$$

Note $P[W = 100] = 0$

↳ continuous

results from the fact that X_i s are discrete

approximate W → is continuous

To fix

Approximate $P[W=100] \simeq P[99.5 < W < 100.5]$

Non-zero.

Linear Regression (Estimation)

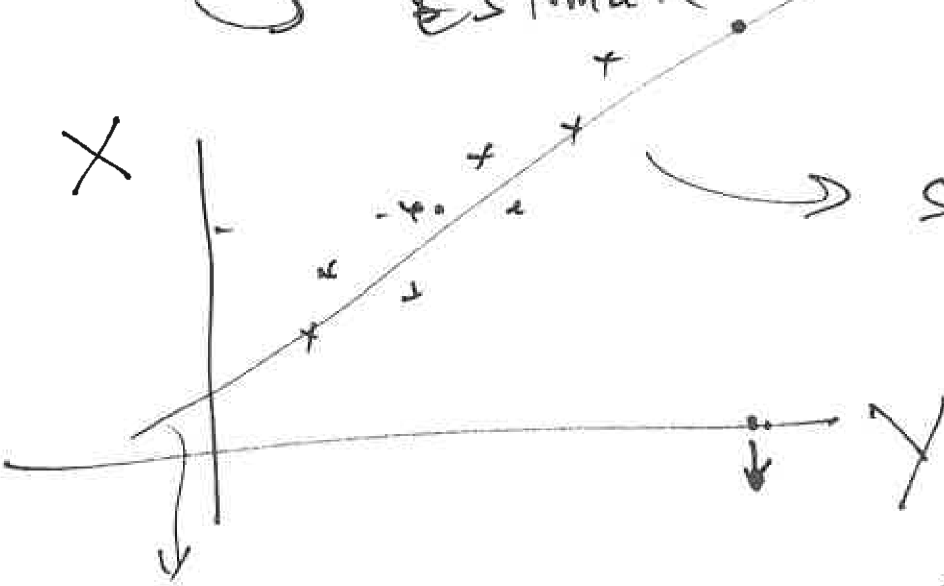
Recall: Joint Statistics

~~§~~ 12.2 (9.2) → 2nd Edi

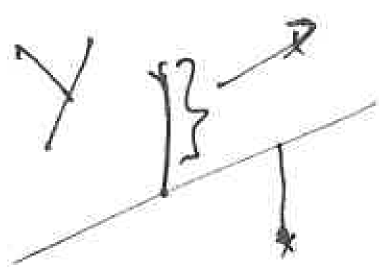
X & Y → 2 RVs

Observe Y value (y)

↪ Estimate the corresponding X



→ scatter plot.



$$\hat{x}_L(y) = ay + b$$

Goal: Minimise the Squared Error

$$E\{[x - \hat{x}_L(y)]^2\}$$ → Minimum Mean Squared Error

$$e_L = E\left[\{x - \hat{x}_L(y)\}^2\right]$$

$$= E\left[\{x - (ay+b)\}^2\right]$$

$$\frac{\partial e_L}{\partial a} = 0 \quad \& \quad \frac{\partial e_L}{\partial b} = 0.$$

optimal $a \rightsquigarrow a^*$
$b \rightsquigarrow b^*$

$$\hat{x}_L(y) = a^* y + b^*$$

It can be shown

$$a^* = \frac{\rho_{x,y}\,\sigma_x}{\sigma_y}, \quad b^* = \mu_x - a^* \mu_y$$

$$\hat{x}_L(y) - \mu_x = \frac{\rho_{x,y}\,\sigma_x}{\sigma_y}(y - \mu_y)$$

Minimum Squared Error $e_L^* = (1-\rho_{x,y}^2)\,\sigma_x^2$

~~eg~~ $a^* = \frac{\rho_{x,y}\,\sigma_x}{\sigma_y} = \frac{Cov[x,y]}{\sigma_x \sigma_y}\cdot\frac{\sigma_x}{\sigma_y}$ 13

$$a^* = \frac{Cov[x,y]}{Var[y]}$$

$$\rho_{x,y} = \frac{Cov[x,y]}{\sigma_x \sigma_y}$$

eg:- (X,Y): $(1,1)$, $(2,3)$, $(3,4)$

$X : \{1, 2, 3\} \longrightarrow \mu_x, \sigma_x, \mu_x = \frac{1+2+3}{3}$

$Y : \{1, 3, 4\} \longrightarrow \mu_y, \sigma_y, \mu_y = \frac{1+3+4}{3}$

Find $E[xy] = \frac{(1)(1) + (2)(3) + (3)(4)}{3}$

$Cov[x,y] = \checkmark$

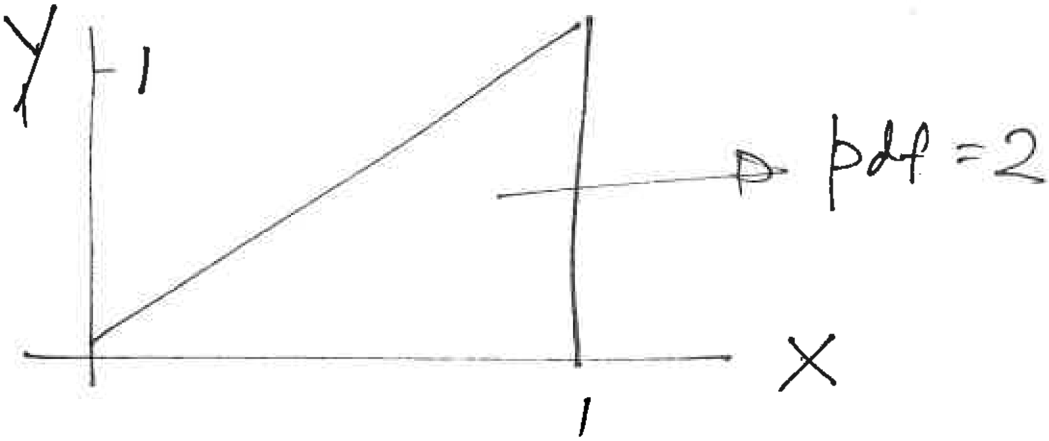
$\rho_{x,y} = \checkmark$

or ~~Cov~~

$$a^* = \frac{Cov[x,y]}{Var[y]}$$

Complete at home.

eg:-



Find $\hat{x}_L(y)$?

$$f_{X,Y}(x,y) = \begin{cases} 2, & 0 \le y \le x \le 1 \\ 0, & \text{oth.} \end{cases}$$

<u>Complete at home</u>

See Ex 12.6 (9.6)

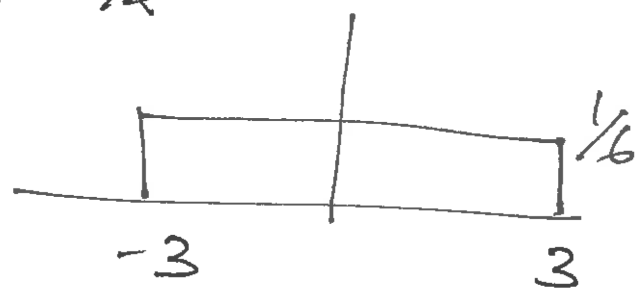
Q 12.2 (9.2)

T is $N(0,3)$, $E[T]=0$

$Var[T]=9$

$R = T + X$

$\hookrightarrow \sim f_X(x)$



T & X are independent

(1) $E[R] = \mu_R = E[T+X]$

$= \mu_T + \mu_X = 0 + 0 = 0$

(2) $Var[R] = Var[T+X]$

$= Var[T] + Var[X] + 2\underbrace{Cov[T,X]}_{=0}$

(T & X are indep)

$= 9 + \frac{6^2}{12}$

$= 12$

(3) $\mathrm{Cov}[T, R] = E[TR] - \mu_T \mu_R$

$= E[T(T+X)] - \cancel{\mu}\, 0$

$= \underbrace{E[T^2]} + \underbrace{E[TX]}$ → T & X are indep.

$= E[T]E[X]$

$= 0$.

$\underbrace{\mathrm{Var}[T]} = E[T^2] - \mu_T^2$

$9 = E[T^2] - 0 \rightarrow E[T^2] = 9$

(4) $\rho_{T,R} = \dfrac{\mathrm{Cov}[T, R]}{\sigma_T \sigma_R} = \dfrac{9}{3\sqrt{12}}$

(5) $a^* = \dfrac{\rho_{T,R} * \sigma_T}{\sigma_R} =$ ✓

$b^* =$ ✓ $(\mu_T - a^* \mu_R)$

(6) Min $e_L = (1 - \rho_{\cancel{T},R}^2)\, \sigma_{\cancel{T}}^2$

Confidence Intervals

X_is ~~are iids~~

~~can also be realised~~

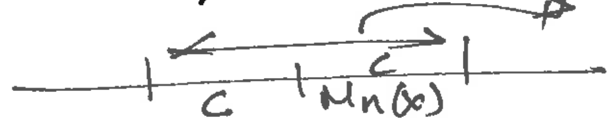
X_is are realisations of a RV X

$$M_n(X) = \underbrace{\frac{X_1 + X_2 \cdots + X_n}{n}}_{\text{Sample Mean}}$$

Try to estimate μ_X and provide the confidence interval for μ_X

$$P\left[\, |M_n(X) - \mu_X| \geq c \,\right] \leq p$$

$M_n(X)$ & μ_X lie in the region → confidence interval



Chebyshev Inequality.

X is a RV, X is any RV

$$P\left[\,|X-\mu_X| \geq c\,\right] \leq \frac{\mathrm{Var}[X]}{c^2}$$

$$M_n(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

use $M_n(X)$ in place of X

$$\mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}$$

$$E[M_n(X)] = \mu_X$$

$$P\left[\,|M_n(X)-\mu_X| \geq c\,\right] \leq \frac{\mathrm{Var}[X]}{nc^2}$$

Note: for any c, $\frac{\mathrm{Var}[X]}{nc^2}$ can be made as small as possible by increasing n