* Mean-Squared estimation of Random variable :

[A]. Some random experiment is conducted.

Now without using any model, we are interested in knowing what is the next outcome?

→ let us first consider the estimate of X by using estimate as some constant, b.

∴ We need to use some kind of "measure" so that error in estimate is minimum/zero

e.g.,
- $E(X-b)$
- $E|(X-b)|$
- $E[(X-b)^2]$
- ⋮ etc.

} all this measures lead to different answers.

- Consider, $E[(X-b)^2]$ as a measure. and try to minimize the error.

i.e., $\frac{d}{db} E[(X-b)^2]$ has to be minimum.

$$\therefore \frac{d}{db} E[(X-b)^2] = 0.$$

This is called "<u>minimum mean-squared estimation</u>" (<u>MMSE</u>).

$$\frac{d}{db} E[(X-b)^2] = \frac{d}{db}\left[\cancel{E(X^2)}^{\to 0} + E(b^2) - 2b\,E(X)\right] = 0.$$

$$\Leftrightarrow \quad 2b - 2E(X) = 0$$

$$\Leftrightarrow \quad \boxed{b = E(X)}$$ → next estimated value will be $E(X)$.

Therefore, $\min E\left[(X-b)^2\right] \Rightarrow \boxed{\hat{X} = E(X)}$

normally, written as $\hat{X}$ (estimate of X).

Lab experiment:

- Take $X \sim$ known pdf (to verify)
- find $E(X)$
- mean-sub $(X - E(X))$
- square it $(X - E(X))^2$
- do $E\left[(X - E(X))^2\right]$.

Here, we are not knowing anything about $f_X(x)$

only X is available

Then, mean is the best estimate (MSE sense)

Check: take some other value of b, i.e., $b \neq E(X)$, then error will be more w.r.t. $b = E(X)$

[B]. Now, consider (measured) received data

X as $\boxed{X = Y + N}$ model

Measured (available) — X

Actual measurements — Y

noise (perturbations) — N

problem : given, X

estimate, Y

- Let's assume linear relationship between X and Y.

i.e., $\hat{Y} = aX + \underset{\substack{\uparrow \\ \text{affine}}}{b}$ { a, b are unknowns to be estimated }

$\therefore$ Minimize $E[(Y-\hat{Y})^2]$

$$\Leftrightarrow \arg\min_{\forall a,b} E[(Y-(aX+b))^2]$$

$$\Leftrightarrow \arg\min_{\forall a,b} E[((Y-aX)-b)^2]$$

$$\Leftrightarrow \arg\min_{\forall a,b} E[Y^2-2aXY+a^2X^2-2bY+2abX+b^2]$$

$$\Leftrightarrow \arg\min_{\forall a,b} E[Y^2]-2aE[XY]+a^2E(X^2)-2bE(Y)+2abE(X)+b^2$$

now, $\frac{\partial}{\partial a}=0$ and $\frac{\partial}{\partial b}=0$.

$$\frac{\partial \mathcal{E}}{\partial a} = 0 \Rightarrow -2E(XY) + 2a\,E(X^2) + 2b\,E(X)$$

$$\frac{\partial \mathcal{E}}{\partial b} = 0 \Rightarrow -2E(Y) + 2a\,E(X) + 2b$$

$$\therefore \boxed{a_{opt} = \frac{Cov(X,Y)}{Var(X)}} \qquad \boxed{b_{opt} = E(Y) - E(X)\frac{Cov(X,Y)}{Var(X)}}$$

Recall, $\hat{Y} = a_{opt} X + b_{opt}$

$$= \frac{Cov(X,Y)}{Var(X)} \cdot X + E(Y) - E(X)\frac{Cov(X,Y)}{Var(X)}$$

$$\boxed{\therefore \hat{Y} = E(Y) + \frac{Cov(X,Y)}{Var(X)}\left[x - E(X)\right]}$$

Remarks:

- $E(Y)$, $\mathrm{cov}(X,Y)$, $\mathrm{Var}(X)$, $E(X)$

  should be known.

  $\therefore$ Need kind of "training" or "machine learning"

  prior to actual measurements

  or actual data X^n.

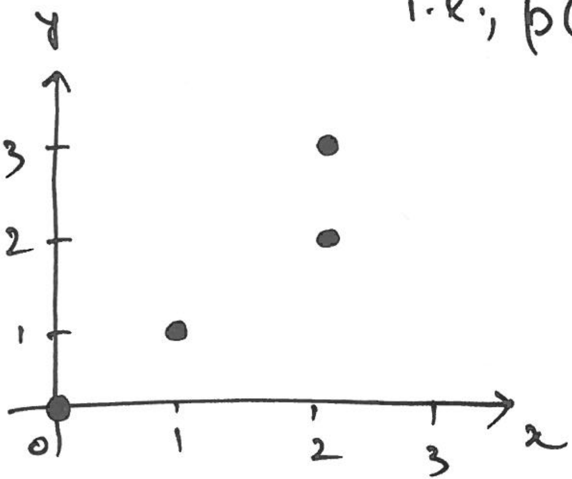
- The eqn of $\hat{Y}$ represents the best linear estimate of Y, given X.

Example: Sample space $S_{XY} = \{(0,0), (1,1), (2,2), (2,3)\}$

with equiprobable outcomes,

i.e., $p(0,0) = p(1,1) = p(2,2) = p(2,3)$
$= 1/4$.

Joint density funⁿ:



X	Y
0	0
1	1
2	2
2	3

Aim is to estimate Y, i.e., $\hat{Y}$, given x.

$$\hat{Y} = E(Y) + \frac{cov(X,Y)}{Var(X)}\left[x - E(X)\right].$$

$$E(X) = \sum_i x_i \, p(x_i) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} = \frac{5}{4}$$

$$E(Y) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = 3/2.$$

$$E(X^2) = 9/4.$$

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

$$= 9/4 - 25/16 = 11/16.$$

$$E(XY) = 11/4.$$

$$\mathrm{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y)$$

$$= 11/4 - 5/4 \cdot 3/2$$

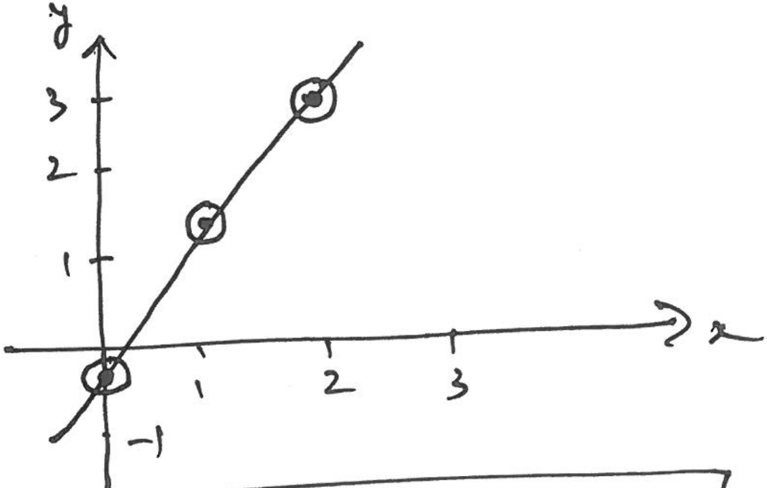
$$= 7/8.$$

Hence, $\hat{Y} = \frac{3}{2} + \frac{7/4}{11/16}\left[x - 5/4\right]$

$$\therefore \boxed{\hat{Y} = \frac{14}{11}x - \frac{1}{11}}$$

$$\left.\begin{aligned} a_{opt} &= 14/11 \\ b_{opt} &= -1/11 \end{aligned}\right\} \text{estimated unknowns.}$$

X	Y	$\hat{Y}$
0	0	$-1/11$
1	1	$13/11$
2	2, 3	$27/11$



$$\{Y - \hat{Y}\} = \left\{\frac{1}{11}, -\frac{2}{11}, -\frac{5}{11}, +\frac{6}{11}\right\}$$

all are equiprobable

$$\therefore \boxed{E\{(Y - \hat{Y})^2\} = \frac{33}{242}}$$

MSE

<u>Remark:</u> we have done $\underset{\text{minimum}}{E[(Y-(aX+b))^2]}$

$$\therefore \frac{\partial}{\partial a}[\cdot] = -2E[\underbrace{(Y-(aX+b))}_{\text{error}}\underbrace{X}_{\text{data}}] = 0$$

For MMSE, error is "<u>orthogonal</u>" to data (r.v.)

$\therefore$ $E[\text{error} \cdot \text{data}] = 0.$

Hence, Geometrically we can solve the problem by making error orthogonal to the data.

(<u>Least-Squares projection</u>)

Check:

error	data (X)
1/11	0
-2/11	1
-5/11	2
6/11	2

dot product

$$(Y-\hat{Y}) \cdot X = \left(\cancel{\tfrac{1}{11} \cdot 0}\right)^{\to 0} + \left(-\frac{2}{11} \cdot 1\right) + \left(-\frac{5}{11} \cdot 2\right) + \left(\frac{6}{11} \cdot 2\right)$$

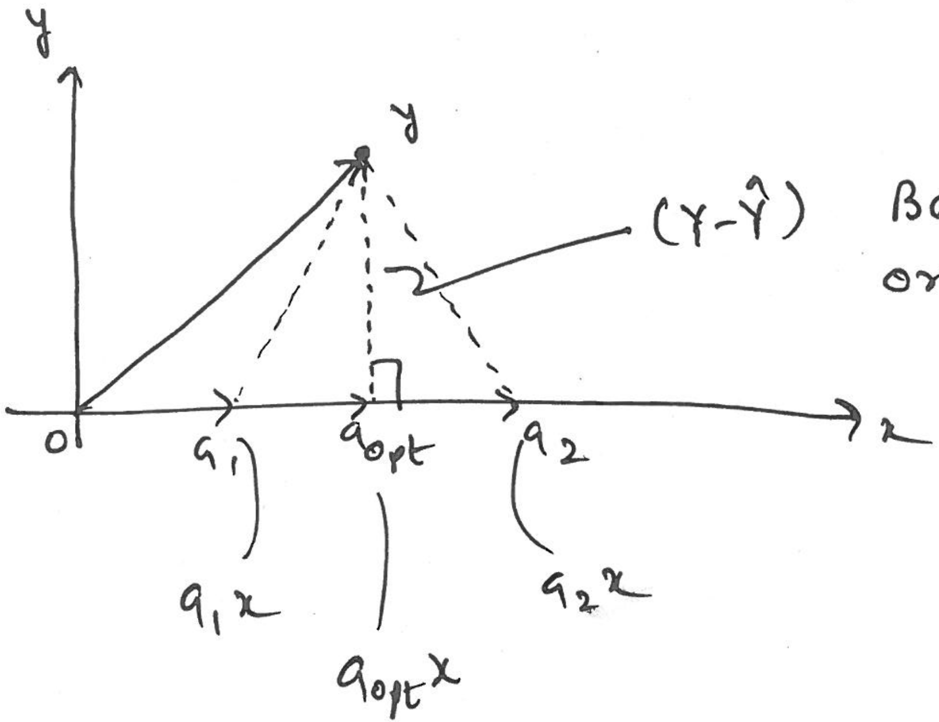
$$= -\frac{2}{11} - \frac{10}{11} + \frac{12}{11} = 0.$$

dot product is zero $\therefore$ $\boxed{\text{error} \perp \text{data}}$

MMSE / LSE

Say, we have $\hat{Y} = aX$

$\therefore \hat{Y} = a_{opt} X$



Best estimate
orthogonal projection
onto data vector

* Least-squares estimate (LSE)

- $A\underline{x} = \underline{b}$ given, $A, \underline{b}$, estimate $\underline{x}$

  when $\|A\underline{x} - \underline{b}\| \neq 0$

pre-multiply both the sides by A^T,

$$\Leftrightarrow \quad A^T A \underline{x} = A^T \underline{b}$$

$$\Leftrightarrow \quad \hat{\underline{x}} = (A^T A)^{-1} A^T \underline{b}$$

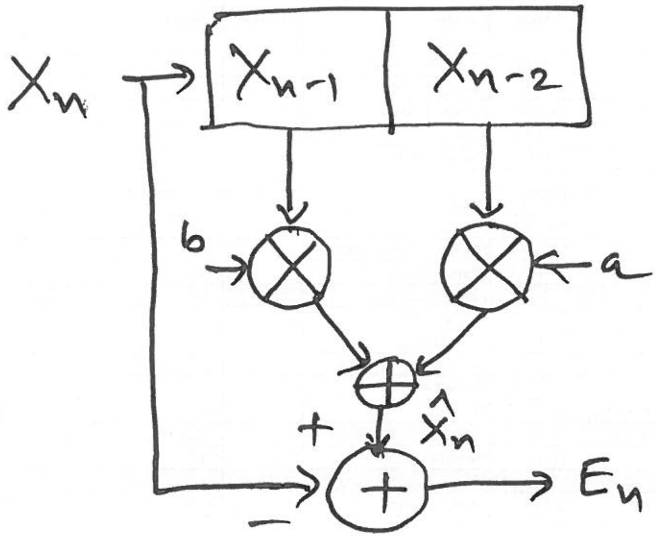
$\Rightarrow$ this will make

$\|A\underline{x} - \underline{b}\|_2$ minimum

This is algebraic version of MMSE / LSE.

Lab expt:

Application: Second-order prediction of Speech.



Two-tap linear predictor for speech processing.

$X_1, X_2, \ldots X_n$ sequence of speech samples

estimate a, b

→ predictor coefficients

s.t. $E(X_n - \hat{X}_n)^2$ is min.

where, $\hat{X}_n = aX_{n-2} + bX_{n-1}$

e.g., $X_3 = aX_1 + bX_2$

So & so forth.

- It is common practice to model speech samples as
  - zero mean
  - σ^2 variance
  - $\text{cov}(\cdot)$ that does not depend on the specific index of the samples, but rather on the separation bet^{n} them
    i.e., $\text{cov}(X_j, X_k) = \rho_{|j-k|}\,\sigma^2$.

- The eq^{n} for optimum lin. predictor coeff. becomes

$$\left.\begin{array}{l}\rho_1 = .825 \\ \rho_2 = .562\end{array}\right\} \qquad \sigma^2 \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \sigma^2 \begin{bmatrix} \rho_2 \\ \rho_1 \end{bmatrix}$$

where, $a = \dfrac{\rho_2 - \rho_1^2}{1-\rho_1^2}$; $b = \dfrac{\rho_1(1-\rho_1^2)}{1-\rho_1^2}$.