12/02/2021

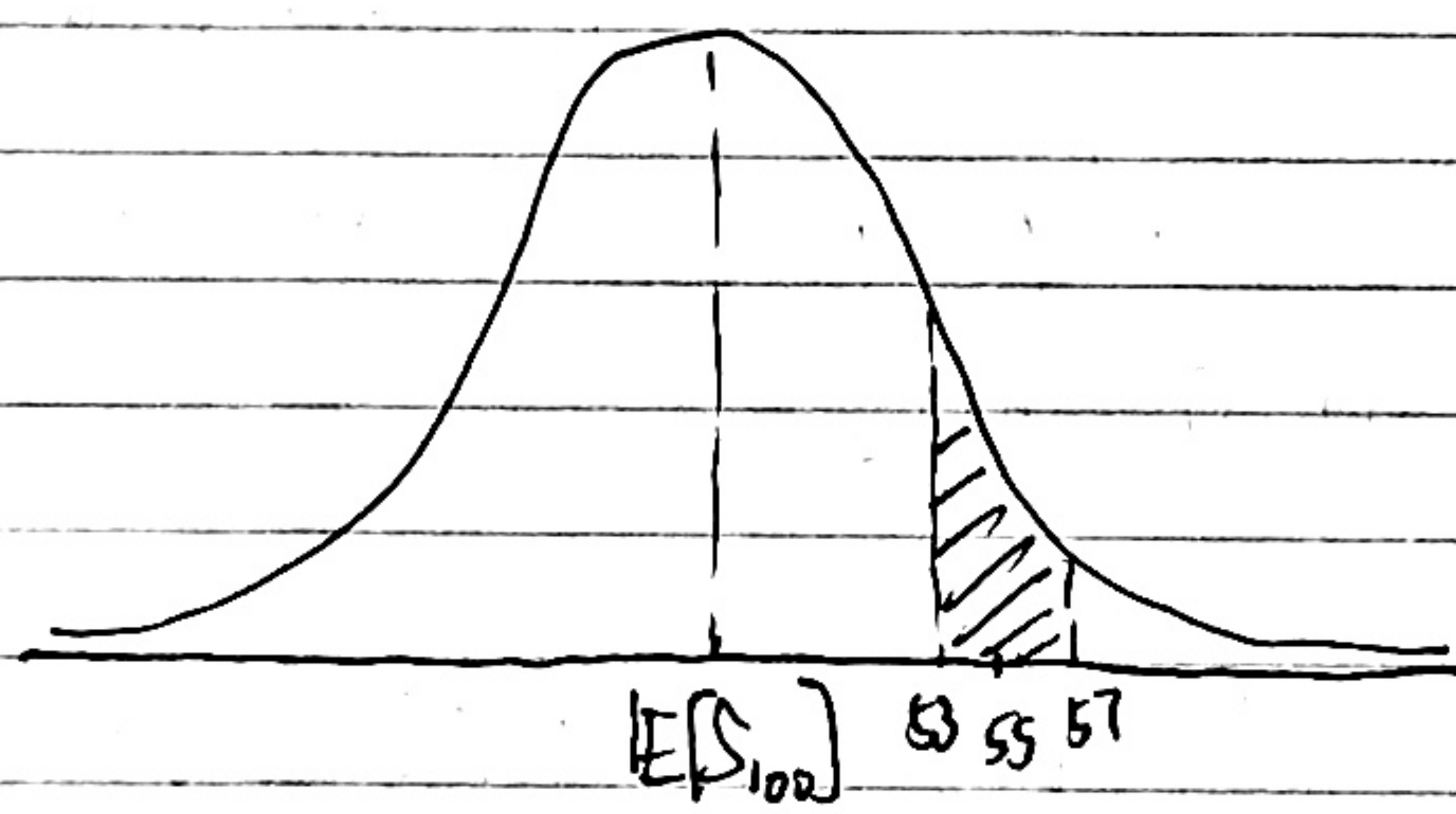
CS70 LECTURE NOTES 26: ESTIMATION

$X_1, X_2, \cdots, X_n$ be identically independently distributed variables.
$\mathbb{E}[X_i] = \mu \qquad \sigma_{X_i}^2 = \sigma^2$

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$Z_n = \frac{M_n - \mu}{\sigma_{M_n}} = \frac{\frac{X_1+X_2+\cdots+X_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

The central limit theorem says $\lim_{n\to\infty} \mathbb{P}[Z_n \le z] = \mathbb{P}[Z \le z]$
where Z is the standard Gaussian



$$\mathbb{P}[53 \le S_{100} \le 57]$$
$$= \mathbb{P}\left[\frac{53-50}{\frac{10}{2\sqrt{3}}} \le \frac{S_{100}-50}{\frac{10}{2\sqrt{3}}} \le \frac{57-50}{\frac{10}{2\sqrt{3}}}\right]$$
$$= \mathbb{P}\left[\frac{3}{\frac{10}{2\sqrt{3}}} \le Z_{100} \le \frac{7}{\frac{10}{2\sqrt{3}}}\right]$$

The approximation of Z_{100} to Z (the standard gaussian uses CLT)

$$\approx \mathbb{P}\left[\frac{3}{\frac{10}{2\sqrt{3}}} \le Z \le \frac{7}{\frac{10}{2\sqrt{3}}}\right]$$
$$= \mathbb{P}\left[\frac{3\sqrt{3}}{5} \le Z \le \frac{7\sqrt{3}}{5}\right]$$

$= \mathbb{P}[1.039 \le Z \le 2.425]$ ~~= 0.9924~~ $= \mathbb{P}[Z \le 2.425] - \mathbb{P}[Z \le 1.039]$
$= 0.9924 - 0.8508 = 0.1416.$

$\Phi(2.425) = 0.9924$
$\Phi(1.039) \approx 0.8508$

Estimation

Suppose there is a random variable Y that I want to estimate.
No observation. Want to estimate Y using a fixed number $\hat{y}$. ~~s.t.~~
for the random variable.

Define error $\varepsilon = Y - \hat{y}$
Criterion: Minimize the mean of the squared error (MMSE)

Squares are great because differentiation gives linear (Mathematical convenience)
Might not be the best for all scenarios (audio, visual)

i.e. Minimize

$$\mathbb{E}[\varepsilon^2] = \mathbb{E}[(Y-\hat{y})^2]$$

let $Z = Y - \hat{y}$.

$\mathbb{E}[Z] = \mathbb{E}[Y] - \hat{y}$

$\mathrm{Var}[Z] = \mathrm{Var}[Y]$ (Same variance, since $\hat{y}$ is a constant)

$\sigma_Z^2 = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 \Rightarrow \mathbb{E}[Z^2] = \sigma_Z^2 + \mathbb{E}^2[Z]$

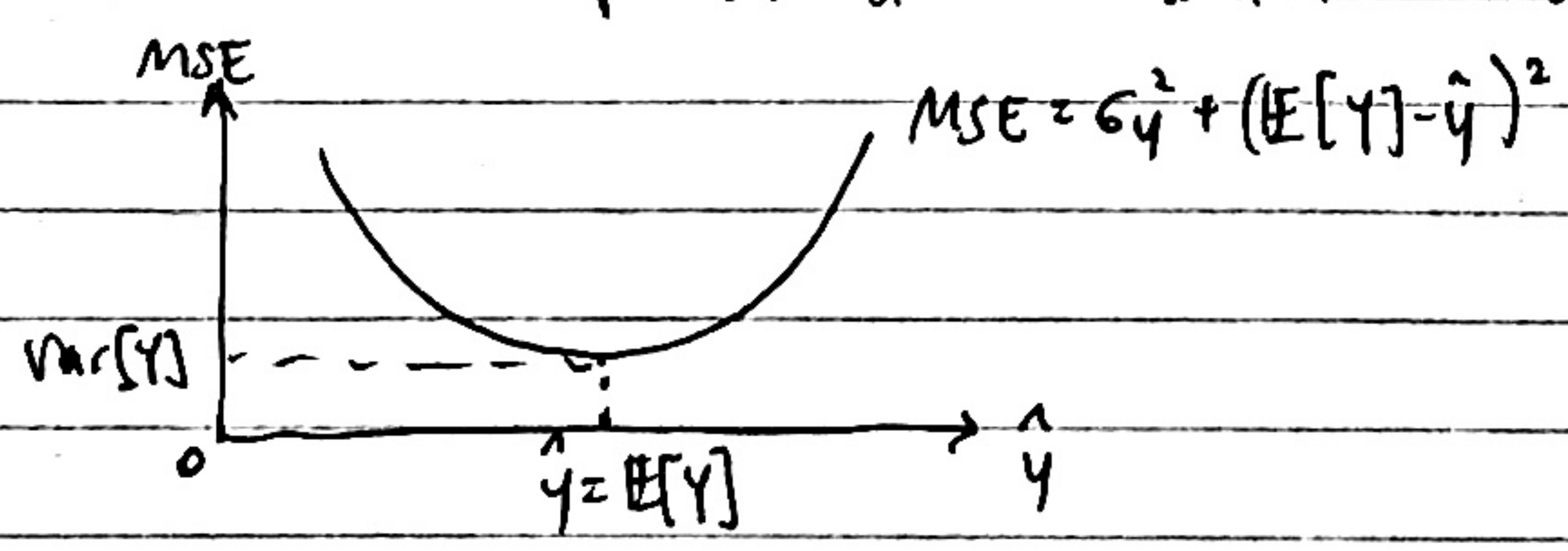
$\Rightarrow \mathbb{E}[(Y-\hat{y})^2] = \mathrm{Var}[Y] + (\mathbb{E}[Y] - \hat{y})^2$ → BUT can influence this!

Cannot change $\mathrm{Var}(Y)$ since Y comes to you as a random variable

~~To minimize $\mathbb{E}[(Y-\hat{y})^2]$, set $\hat{y} = \mathbb{E}[Y]$~~

Optimized Mean Squared Error: $\mathrm{MSE} = \mathbb{E}[\varepsilon^2] = \mathbb{E}[(Y-\mathbb{E}[Y])^2] = \mathrm{Var}[Y]$ (!)

∴ The mean is the optimal estimator and it results in an MSE σ_Y^2.



$\mathbb{E}[(Y-\mathbb{E}[Y])^2] \leq \mathbb{E}[(Y-\hat{y})^2] \quad \forall \hat{y}.$

What if we have an observation X? Entering conditional universe.

$\hat{y} = \underbrace{\mathbb{E}[Y \mid X = x]}_{\text{point estimate of } Y \text{ in the MMSE sense}} = g(x)$ → some function of x.

$\mathbb{E}[(Y - \mathbb{E}[Y|X])^2] \leq \mathbb{E}[(Y-g(X))^2] \quad \forall$ functions $g(x)$

Linear MSE Estimator

$\hat{y}(X) = aX + b$ Restrict to searching over all linear functions instead of over all functions

$\varepsilon = Y - \hat{y}(X) = Y - (aX + b)$

$MSE = \mathbb{E}[(Y - \hat{y}(X))^2] = \mathbb{E}[(Y - aX - b)^2]$

let $Z = Y - aX$

$\Rightarrow MSE = \mathbb{E}[(Z - b)^2]$

The optimal b is $b = \mathbb{E}[Z] = \mathbb{E}[Y - aX] = \mathbb{E}[Y] - a\mathbb{E}[X]$.

$$\begin{aligned} MSE = \mathbb{E}[(Z-b)^2] &= \mathbb{E}[(Y - aX - b)^2] \\ &= \mathbb{E}[(Y - aX - \mathbb{E}[Y] + a\mathbb{E}[X])^2] \\ &= \mathbb{E}[((Y - \mathbb{E}[Y]) - a(X - \mathbb{E}[X]))^2] \end{aligned}$$

$$\begin{aligned} &= \mathbb{E}[(Y - \mathbb{E}[Y])^2] + \mathbb{E}[(X - \mathbb{E}[X])^2] - 2a\,\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= Var[Y] + a^2 Var[X] - 2a\,Cov[X, Y] \end{aligned}$$

$$\Rightarrow \frac{dMSE}{da} = 2a\,Var[X] - 2\,Cov[X, Y] = 0 \Rightarrow \boxed{a = \frac{Cov[X, Y]}{Var[X]}} = \frac{\sigma_{xy}}{\sigma_x^2}$$

$b = \mathbb{E}[Y] - a\mathbb{E}[X]$

$$\therefore \hat{y}_L(X) = aX + b = \frac{\sigma_{xy}}{\sigma_x^2}X + \mathbb{E}[Y] - \frac{\sigma_{xy}}{\sigma_x^2}\mathbb{E}[X]$$

$$= \mathbb{E}[Y] + \frac{\sigma_{xy}}{\sigma_x^2}(X - \mathbb{E}[X]).$$

$$\therefore \boxed{\hat{y}_L(X) = \mathbb{E}[Y] + \underbrace{\frac{\sigma_{xy}}{\sigma_x^2}(X - \mathbb{E}[X])}_{\text{correction term.}}}$$

If X, Y are not related, $Cov[X, Y] = 0 \Rightarrow$ correction term is 0.
Linear estimation only uses first and second order properties.

Homework: Write $\hat{Y}_L(x)$ in terms of $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$.

Think of $\rho = 1, 0, -1$.