Exercise

$$\mathbb{E}[\hat{\theta}_e | \theta_0] = \mathbb{E}\,\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, y_i = \frac{1}{n}\sum \theta = \theta$$

$\hat{\theta}_e$ is unbiased.

$$\mathbb{E}[\hat{\theta}_b | \theta_0] = \mathbb{E}[w\bar{y} + (1-w)\mu_0]$$

$$\boxed{w\theta + (1-w)\mu_0}$$

If $\mu_0 \neq \theta_0$ then $\hat{\theta}_b$ is biased.

Which has lower variance?

$$\mathbb{V}(\hat{\theta}_e | \theta_0) = \mathbb{V}(\bar{y}) = \mathbb{V}\left(\frac{1}{n}\sum y_i\right) = \frac{1}{n^2}\sum \mathbb{V}(y_i)$$

$$= \frac{1}{n^2}\sum \sigma^2 = \boxed{\frac{1}{n}\cdot\sigma^2}$$

$$\mathbb{V}(\hat{\theta}_b | \theta_0) = \mathbb{V}(w\bar{y} + (1-w)\mu_0)$$

$$= \mathbb{V}(w\bar{y})$$

$$= w^2\,\mathbb{V}(\bar{y})$$

$$= \boxed{w^2 \cdot \frac{\sigma^2}{n}}$$

$\hat{\theta}_b$ has smaller variance.

$$MSE = \mathbb{E}\left[(\tilde{\theta} - \theta_0)^2 \mid \theta_0\right] \quad ☆$$

$$= var(\hat{\theta} \mid \theta_0) + Bias^2(\hat{\theta} \mid \theta_0)$$

$$\boxed{MSE(\hat{\theta}_e \mid \theta_0) = \overset{var}{\frac{\sigma^2}{n}}} \overset{+ \ bias}{+ \ 0^2}$$

$$MSE(\hat{\theta}_b \mid \theta_0)$$

$\boxed{\text{TRICK}}$ $\theta_0 = w\theta_0 + (1-w)\theta_0$

w/ ☆ :

$$\mathbb{E}\left[\left(\underline{w(\bar{y} - \theta_0)} + \underline{(1-w)(\mu_0 - \theta_0)}\right)^2\right]$$

$$= \mathbb{E}\left[w^2 \underbrace{(\bar{y} - \theta_0)^2}_{var\ \bar{y}} + 2w(\bar{y} - \theta_0)(1-w)(\mu_0 - \theta_0) + (1-w)^2(\mu_0 - \theta_0)^2\right]$$

$$\boxed{w^2 \cdot \frac{\sigma^2}{n} + 0 + (1-w)^2(\mu_0 - \theta_0)^2}$$

$\text{MSE}\left[\hat{\theta}_b \mid \theta_0\right] < \text{MSE}\left[\hat{\theta}_e \mid \theta_0\right]$ when

$$(\mu_0 - \theta_0)^2 < \frac{\sigma^2}{n}\frac{1+w}{1-w}$$

$$= \sigma^2\left(\frac{1}{n} + \frac{2}{K_0}\right)$$

Rule of thumb: if prior guess of μ_0 is within 2 sd of θ_0 & we pick $K_0 = 1$ then Bayes estimator probably has lower MSE.

$$MSE(\hat{\theta} \mid \theta_0) = \underset{\uparrow \text{ mean}}{\mathbb{E}}\Big[\underbrace{(\hat{\theta} - \theta_0)^2}_{\text{squared error}} \mid \theta_0\Big]$$

let $m = \mathbb{E}[\hat{\theta} \mid \theta_0]$

TRICK:

$$MSE(\hat{\theta} \mid \theta_0) = \mathbb{E}\Big[(\underline{\hat{\theta} - m} + \underline{m - \theta_0})^2 \mid \theta_0\Big]$$

$$= \mathbb{E}\Big[(\hat{\theta} - m)^2 + 2(\hat{\theta} - m)(m - \theta_0) + (m - \theta_0)^2 \mid \theta_0\Big]$$

$$\underbrace{\mathbb{E}\big[(\hat{\theta} - m)^2\big]}_{\underline{Var(\hat{\theta} \mid \theta_0)}} + \underbrace{2\,\mathbb{E}\big[(\hat{\theta} - m)(m - \theta_0)\big]}_{0} + \underbrace{\mathbb{E}\big[(m - \theta_0)^2 \mid \theta_0\big]}_{(m - \theta_0)^2}$$

$+$

$\underline{Bias^2}(\hat{\theta} \mid \theta_0)$

Frequentist risk

How well does $\hat{\theta}$ perform on avg across different datasets?

function of y

$$R(\theta, \hat{\theta}) = \mathbb{E}\left[L(\theta, \hat{\theta}) \mid \theta \right]$$

If $L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$ then $R(\theta, \hat{\theta}) = MSE(\theta, \hat{\theta})$.