Theorem: MSE = bias^2 + Var

Pf:

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

$$= E[(\underbrace{\hat{\theta} - E\hat{\theta}}_{a} + \underbrace{E\hat{\theta} - \theta}_{b})^2]$$

$$= E[a^2] + E[b^2] + 2E[ab]$$

$$= \underbrace{E[(\hat{\theta} - E\hat{\theta})^2]}_{1} + \underbrace{E[(E\hat{\theta} - \theta)^2]}_{2} + \underbrace{2E[(\hat{\theta} - E\hat{\theta})(E\hat{\theta} - \theta)]}_{3}$$

$$1 = Var(\hat{\theta})$$

$$2 = \cancel{E[}(E\hat{\theta} - \theta)^2\cancel{]} = (E\hat{\theta} - \theta)^2$$

$$= Bias(\hat{\theta})^2$$

$$③ = 2\left[E(\hat{\theta}) - \theta\right] E\left[\hat{\theta} - E\hat{\theta}\right]$$

$$= 2(E\hat{\theta} - \theta)\underbrace{(E\hat{\theta} - E\hat{\theta})}_{0} = 0.$$

Ex. $X_n \overset{iid}{\sim} f$ where $\mu = EX_n$

$\sigma^2 = Var(X_n)$.

Prev. showed if $\hat{\mu} = \bar{X}$

$$E\hat{\mu} = \mu$$

$$Var(\hat{\mu}) = \sigma^2/N.$$

So by prev. thm:

$$MSE(\hat{\mu}) = Bias^2 + Var$$

$$= (E[\hat{\mu}] - \mu)^2 + \sigma^2/N$$

$$= (\mu - \mu)^2 + \sigma^2/N$$

$$= \sigma^2/N.$$

$= \sigma^2/N.$

Ex. Consider $S^2 = \frac{1}{N-1} \sum_n (X_n - \bar{X})^2$

Showed: $E[S^2] = \sigma^2$

So, S^2 is unbiased for σ^2.

Consider $X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ then we showed

$$\frac{N-1}{\sigma^2} S^2 \sim \chi^2(N-1)$$

So

$$Var\left(\frac{N-1}{\sigma^2} S^2\right) = Var\left(\chi^2(N-1)\right)$$

$$= 2(N-1)$$

$$\Rightarrow \frac{(N-1)^2}{\sigma^4} Var(S^2) = 2(N-1)$$

$$\Rightarrow \boxed{Var(S^2) = \frac{2\sigma^4}{\ldots}}$$

$$\Rightarrow \boxed{\mathrm{Var}(S^2) = \frac{2\sigma}{N-1}}$$

Since S^2 unbiased, MSE = Var

So $$MSE(S^2) = \frac{2\sigma^4}{N-1}.$$

The MLE for σ^2 was

$$\hat{\sigma}^2 = \frac{1}{N}\sum_n (X_n - \bar{X})^2. = \frac{N-1}{N} S^2.$$

Which has lower MSE?

<u>Bias</u>

$$E[\hat{\sigma}^2] = E\left[\frac{N-1}{N} S^2\right] = \frac{N-1}{N} E[S^2] = \frac{N-1}{N}\sigma^2$$

$$\mathrm{Bias}(\hat{\sigma}^2) = E[\hat{\sigma}^2] - \sigma^2$$

$$= \frac{N-1}{N}\sigma^2 - \sigma^2 = \boxed{-\frac{1}{N}\sigma^2}$$

$$= \frac{N-1}{N}\sigma^2 - \sigma^2 = \boxed{-\frac{1}{N}\sigma^2}.$$

<u>Var</u>

$$Var(\hat{\sigma}^2) = Var\left(\frac{N-1}{N}S^2\right)$$

$$= \frac{(N-1)^2}{N^2} Var(S^2)$$

$$= \frac{(N-1)^2}{N^2} \frac{2\sigma^4}{N-1}$$

$$\boxed{= \frac{2(N-1)\sigma^4}{N^2}}.$$

$$MSE(\hat{\sigma}^2) = Bias^2 + Var$$

$$= \left(-\frac{1}{N}\sigma^2\right)^2 + \frac{2(N-1)\sigma^4}{N^2}$$

$$= \dots$$

$= \ldots$

$$\boxed{= \frac{(2N-1)\sigma^4}{N^2}}$$

Which is smaller?

$$MSE(\hat{\sigma}^2) = \underbrace{\frac{(2N-1)}{N^2}\left(\frac{N-1}{2}\right)}_{>1?\ <1?}\underbrace{\left(\frac{2}{N-1}\right)\sigma^4}_{MSE(S^2)}$$

$$\frac{2N^2-3N+1}{2N^2} < 1$$

So $MSE(\hat{\sigma}^2) < MSE(S^2)$.

More generally, what do I multiply S^2 by to minimize MSE?

$MSE(cS^2)$

$= Bias(cS^2)^2 + Var(cS^2)$

$= \left(E[cS^2] - \sigma^2\right)^2 + c^2 Var(S^2)$

$= \left(c E[S^2] - \sigma^2\right)^2 + c^2 Var(S^2)$

$= \left(c\sigma^2 - \sigma^2\right)^2 + c^2 \frac{2\sigma^4}{N-1}$

$= (c-1)^2 \sigma^4 + \frac{2c^2\sigma^4}{N-1}$

take deriv wrt c, set to zero.

$\frac{\partial}{\partial c} MSE = 2(c-1)\cancel{\sigma^4} + \frac{4c\cancel{\sigma^4}}{N-1} = 0$

$\Rightarrow \cancel{2}c - \cancel{2}^{1} + \frac{\cancel{4}^{2}c}{N-1} = 0$

$\Rightarrow \ldots\ldots$

$$\Rightarrow \boxed{c^* = \frac{N-1}{N+1}}$$

So min MSE is actually at

$$c^* s^2 = \frac{N-1}{N+1} s^2 = \boxed{\frac{1}{N+1} \sum_n (x_n - \bar{x})^2}.$$

MID1
SP4
QP4

I want to find a "best" estimator.

Problem: If I am too permissive in what I call an estimator, there is no "best".

Ex. $x_n \overset{iid}{\sim} N(\mu, 1)$

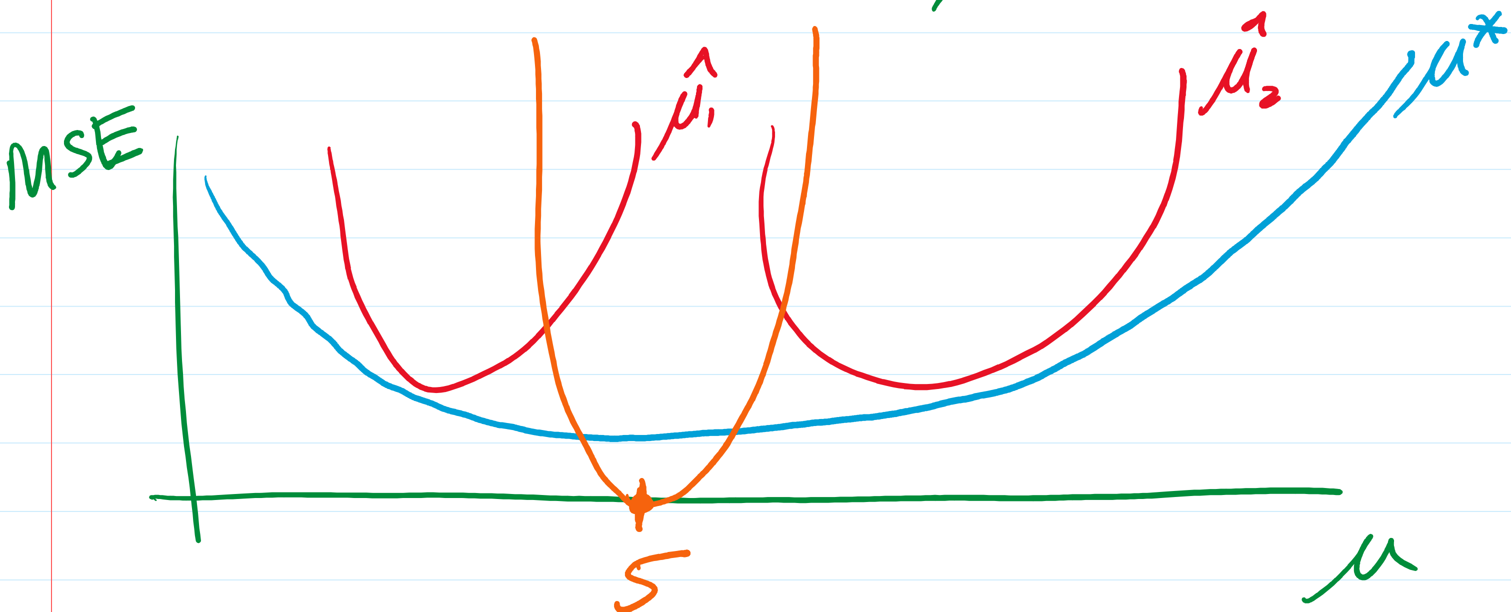
Want some optimal μ^* so that

Want some optimal μ so that

$$MSE_\mu(\mu^*) \le MSE_\mu(\hat{\mu})$$

$\forall$ possible $\hat{\mu}$

$\forall \mu$



$\hat{\mu}_3 = 5$

Need to restrict class of allowable ests.

Defn: UMVUE

Uniformly Minimum Variance Unbiased Est.

Note: If Bias $= 0$ then MSE $=$ Var.

Note: If Bias $= 0$ then MSE $=$ Var.

We call θ^* the UMVUE of $\tau(\theta)$ if

① it's unbiased for $\tau(\theta)$

$$E[\theta^*] = \tau(\theta)$$

② minimum variance – uniformly

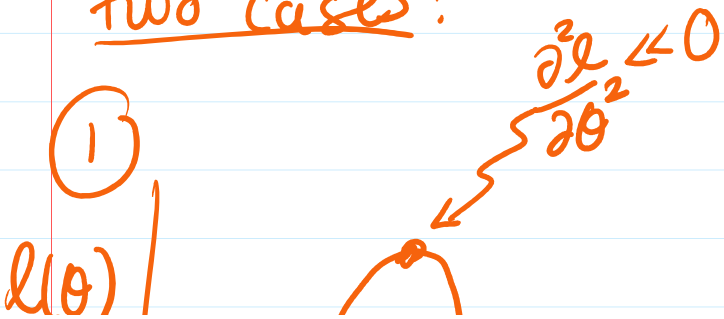
$$Var_\theta(\theta^*) \le Var_\theta(\hat{\theta})$$

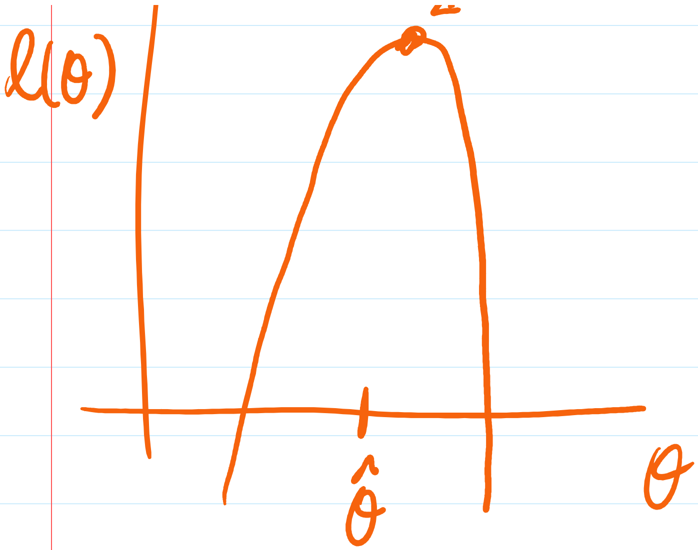
$$\forall \text{ unbiased } \hat{\theta}$$

$$\forall \theta \in \Theta.$$

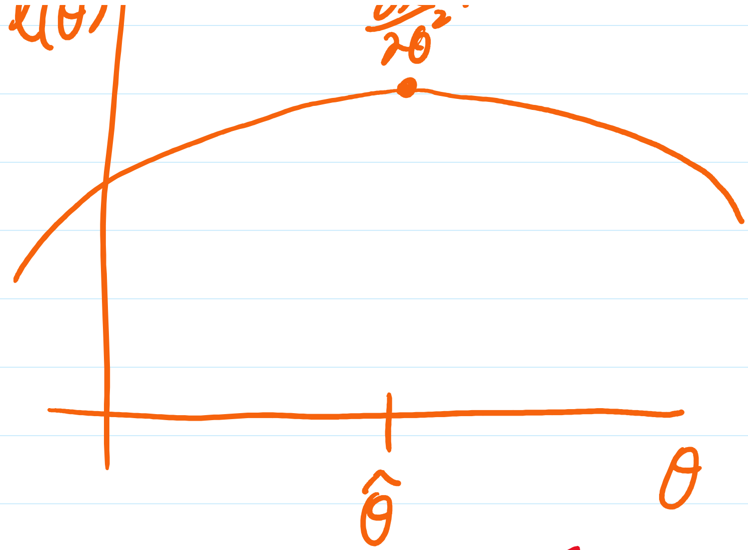
Consider estimating θ w/ MLE

two cases:





Strongly prefer $\hat{\theta}$ over other vals



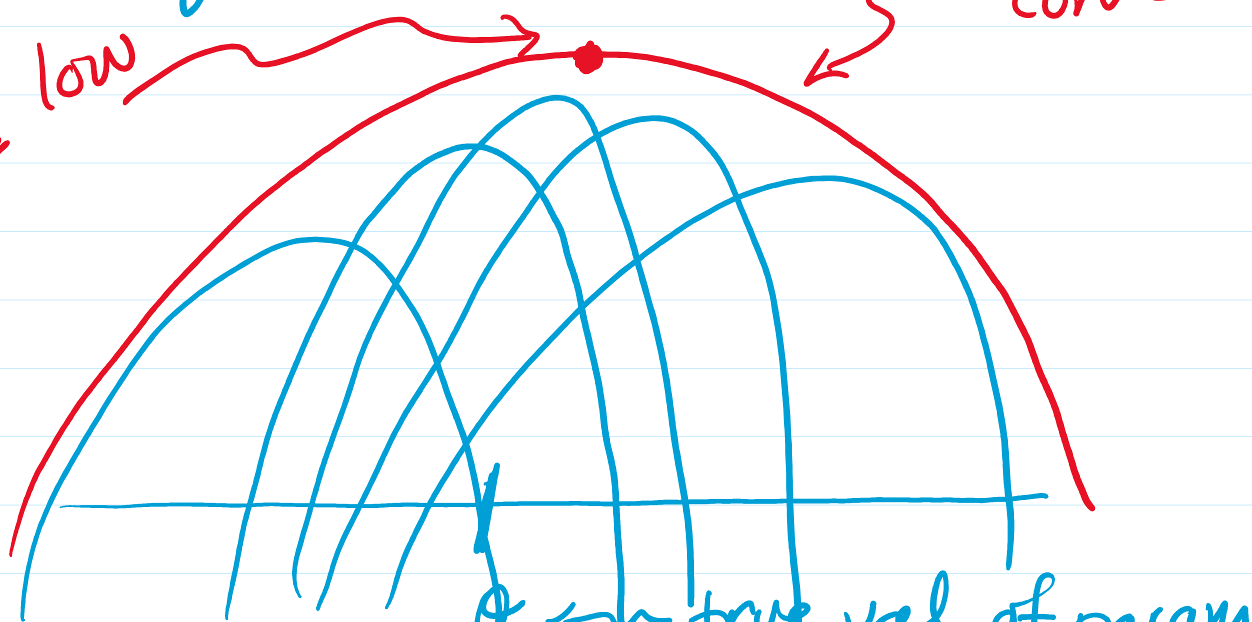
weakly prefer $\hat{\theta}$.

What about est. accuracy

(1) $\hat{\theta}$ highly variable

$\frac{\partial^2}{\partial \theta^2}$ low

expected curve

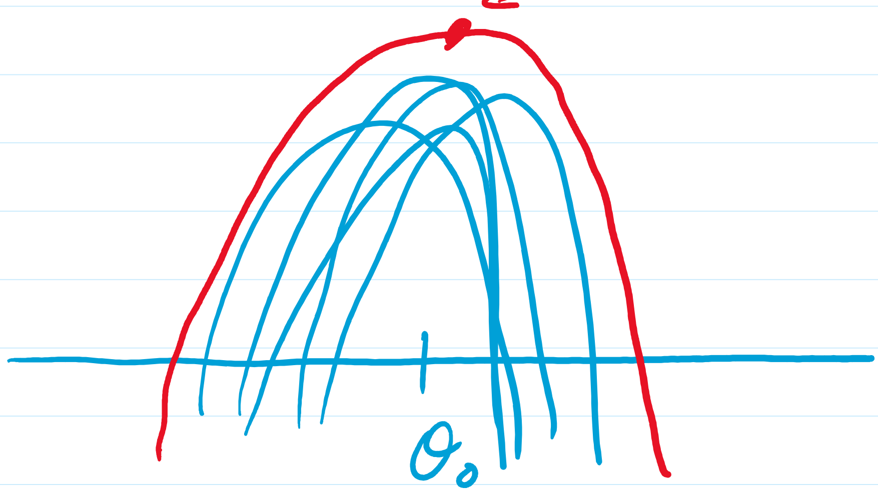


θ_0 ← true val of param

(2) $\hat{\theta}$ ... variable

② $\hat{\theta}$ not variable

$\frac{\partial^2}{\partial \theta^2} \ll 0$



Claim: $\frac{\partial^2}{\partial \theta^2}$ at θ_0 tells us about Var/accuracy/MSE of $\hat{\theta}$.

Really only useful for "nice" dists.

Defn: Fisher Information (*) Always makes sense for exp. fams.

For a single obs ($N = 1$)

so that $X \sim f_\theta$

we define the Fisher Info. about θ

we define the Fisher info

contained in $\mathbb{X}$ as

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(\mathbb{X})\right] = -E\left[\frac{\partial^2 \ell}{\partial\theta^2}\right]$$

For $N > 1$ observations we define

$$I_N(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(\mathbb{X})\right] = -E\left[\frac{\partial^2 \ell}{\partial\theta^2}\right]$$