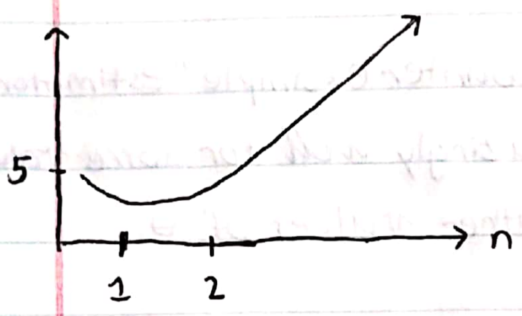
Lecture 9 10/5/2020

Define "relative efficiency" (RE) as the ratio of variances:

$$RE = \frac{Var[\hat{\theta}^{MM}]}{Var[\hat{\theta}^{MLE}]} = \frac{\cancel{\theta^2}\frac{1}{3n}}{\cancel{\theta^2}\frac{n}{(n+1)^2(n+2)}} = \frac{(n+1)^2(n+2)}{3n^2} > 1 \Rightarrow$$

MLE is "better" as measured by variance.



this means the higher the sample size the bigger the MLE's advantage is over the MM estimator.

Maybe we should be comparing the ratio of MSE's? True... but in this case the tiny amount of bias in the MLE (see simulation) won't matter if n is large.

Two really important questions:

1) Is there a theoretical minimum MSE (best) when estimating θ for a given DGP?

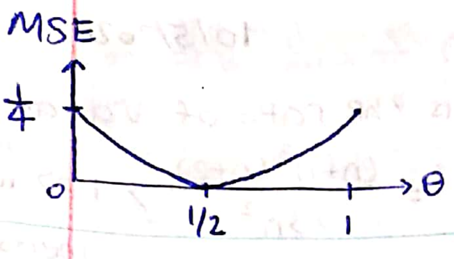
2) If (1) is true, then for any DGP/θ, is there a procedure for locating that estimator with the best MSE?

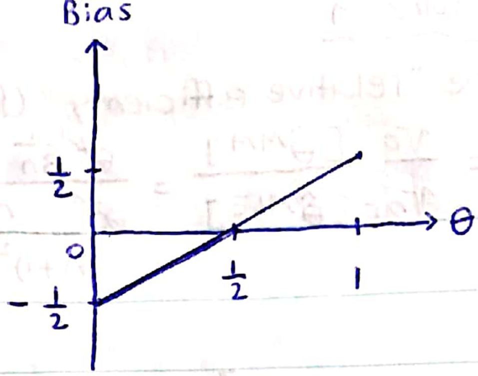
The answer to both is ... NO! (p. 334 C&B textbook)
Why? Because the class of "all" estimators is too big. For example...

$$X_1, \dots, X_n \overset{i.i.d.}{\sim} \text{Bern}(\theta).$$

$\hat{\theta}_{bad} = \frac{1}{2}$ $\quad MSE[\hat{\theta}_{bad}](\theta = \frac{1}{2}) = E[(\hat{\theta}_{bad} - \theta)^2] = E[(\frac{1}{2} - \frac{1}{2})^2] = 0$

This means that $\hat{\theta}_{bad}$ does amazingly well at $\theta = \frac{1}{2}$.





I can always create a "counterexample" estimator like this one that does amazingly well for some values of θ and very badly for other values of θ.

For all *unbiased* estimators (this limits the scope of possible estimators and closes the loophole of the above counterexample)...

1) Is there a theoretical minimum MSE (best) when estimating θ for a given DGP?

2) If (1) is true, then for any DGP / θ, is there a procedure for locating that estimator with the best MSE?

Define: a uniformly minimum variance unbiased estimator (UMVUE) is the estimator $\hat{\theta}^*$ s.t. for all θ and all other unbiased estimators $\hat{\theta}$,

$$Var[\hat{\theta}^*] \leq Var[\hat{\theta}].$$

Let's rephrase the two questions... For all *unbiased* estimators,

1) Is there a theoretical lower bound on the

variance of the UMVUE?

Yes. It is called the Cramer-Rao Lower Bound (CRLB)

Proven in 1945-1946. *

2) Is there a procedure for locating the UMVUE?

Sometimes... unsure if we will get to it in this class.

DGP i.i.d. normal

$\text{var}[\bar{x}] = \frac{\sigma^2}{n}$

CRLB : $X_1, \dots, X_n \overset{i.i.d.}{\sim} DGP(\theta)$, continuous

for any unbiased estimator $\hat{\theta}$,

$$\text{var}[\hat{\theta}] \geq \frac{I(\theta)^{-1}}{n}$$

the numerator is an irreducible core quantity based on the DGP and based on θ.

$I(\theta) := E^X[\ell'(\theta; X)^2]$ and it's called the "Fisher Information" defined by Fisher in 1922

↑

expectation of the squared log-likelihood

Proof This pure probability fact is proved in 368. The Cauchy-Schwartz Inequality for any two R.V.'s Q and S are:

$$\text{cov}[Q,S]^2 \leq \text{Var}[Q]\,\text{Var}[S]$$

$$\Rightarrow \text{var}[Q] \geq \frac{\text{Cov}[Q,S]^2}{\text{var}[S]} = \frac{(E[QS] - E[Q]E[S])^2}{E[S^2] - E[S]^2}$$

Let $Q = \hat{\theta} \Rightarrow E[\hat{\theta}] = \theta$ due to unbiasedness

Define the "Score function" S as:

$S : \frac{\partial}{\partial\theta}[\ln f(X_1, \dots, X_n; \theta)]$ (def 1) chain rule $\rightarrow$ $$= \frac{\frac{\partial}{\partial\theta}[f(X_1,\dots,X_n;\theta)]}{f(X_1,\dots,X_n;\theta)} \quad \text{(def 2)}$$

by i.i.d., mult. rule $\rightarrow$ $$= \underbrace{\frac{\partial}{\partial\theta}\left[\ln \prod_{i=1}^{n} f(X_i;\theta)\right]}_{\text{(def 3)}} \overset{\text{pre-calc}}{=} \underbrace{\frac{\partial}{\partial\theta}\left[\sum_{i=1}^{n} \ln f(X_i;\theta)\right]}_{\text{(def 4)}}$$

linearity of derivative $\rightarrow$ $$= \underbrace{\sum_{i=1}^{n} \frac{\partial}{\partial\theta}[\ln f(X_i;\theta)]}_{\text{(def 5)}}$$

$\mathcal{L} = f$, $\ell = \ln(\mathcal{L}) = \ln(f)$, def 1

$\downarrow$

$$= \underbrace{\frac{\partial}{\partial\theta}[\ell(\theta;X_1,\dots,X_n)]}_{\text{(def 6)}} = \underbrace{\ell'(\theta;X_1,\dots,X_n)}_{\text{(def 7)}} = \sum_{i=1}^{n} \ell'(\theta;X_i)$$

NOTE: S is a R.V., hence all X_i's are also R.V.'s hence capital letters.

We need $E[\hat{\theta}^S]$, $E[S^2]$, $E[S]$, then we're done!

$$E[S] \overset{\text{def 2}}{=} E\left[\frac{\frac{\partial}{\partial\theta}[f(X_1,\dots X_n;\theta)]}{f(X_1,\dots,X_n;\theta)}\right] = \underbrace{\int\cdots\int}_{n} \frac{\frac{\partial}{\partial\theta}[f(X_1,\dots,X_n;\theta)]}{\cancel{f(X_1,\dots,X_n;\theta)}} \cancel{f(X_1,\dots,X_n;\theta)}\, dx_1,\dots,dx_n$$

support of the n-dim r.v. $X_1,\dots,X_n$

if you can interchange the derivative with the integral...

$$= \frac{\partial}{\partial\theta}\Big[\overbrace{\int\cdots\int_{\text{Supp}[X]} f(X_1,\dots,X_n;\theta)\,dx_1,\dots,dx_n}^{1}\Big] = \frac{\partial}{\partial\theta}[1] = 0 \quad \text{(Fact 1a)}$$

$$E[S] \overset{\text{def 7}}{=} E[\ell'(\theta;X_1,\dots,X_n)] = 0$$

$$E[S] \overset{\text{def 8}}{=} E[\textstyle\sum \ell'(\theta;X_i)] \overset{\text{i.i.d.}}{=} nE[\ell'(\theta;X_i)] = 0 \Rightarrow E[\ell'(\theta;X_i)] = 0 \quad \text{(Fact 1b)}$$

$$\mathrm{Var}[S] = E[S^2] - \underbrace{\cancel{E[S]^2}}_{0} \underset{\text{def } 8}{=} E\Bigg[\underbrace{\Big(\sum_{i=1}^{n} \ell'(\theta; X_i)\Big)^2}_{(a_1 + a_2 + \dots + a_n)^2}\Bigg]$$

$$(a_1 + a_2 + \dots + a_n)^2 \Downarrow \sum_{i=1}^{n} a_i^2 + \sum_{i \neq j} a_i a_j$$

end linearity of expectation

$$= \sum_{i=1}^{n} E[\ell'(\theta; X_i)]^2 + \sum_{i \neq j} E[\ell'(\theta; X_i)\, \ell'(\theta; X_j)]$$