lecture 9

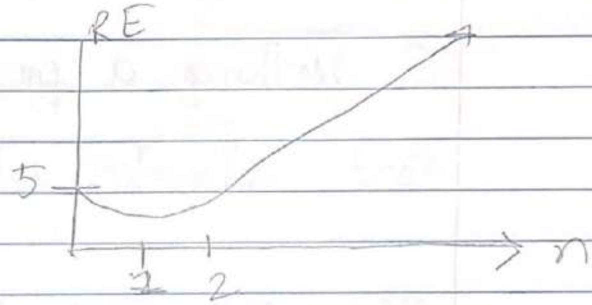
Define "relative efficieny" RE as the ratio of Variances:

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

$$= \frac{(n+1)^2(n+2)}{n^2}$$

$$> 1.$$

$\Rightarrow$ MLE is "better" as measured by variance.



This means the higher the Sample size (n) the bigger the MLE's advantage is over the MM estimator.

Maybe we should be comparing the ratio of MSE's? True...

in this case the tiny amount of bias in the MLE (see simulation) won't matter if n is large.

two really important Question:

(1) is there a theorical 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 least MSE?

The answer to both... NO! (P 334 CSB)

Why? because the class of "all" estimators is too big. For example;

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

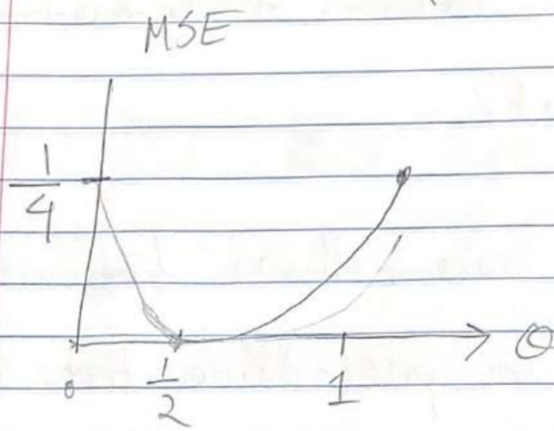
$\hat{\theta}_{BAD} = \frac{1}{2}$

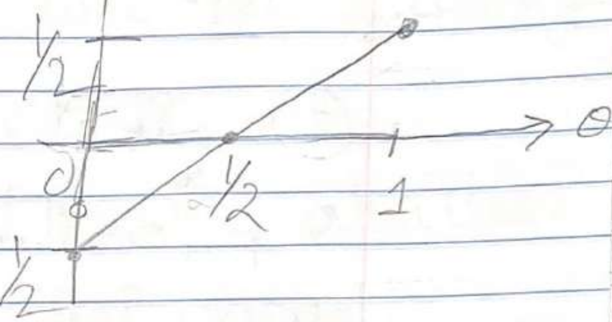
$$MSE[\hat{\theta}_{BAD}]\left(\theta = \frac{1}{2}\right) = E[(\hat{\theta}_{BAD} - \theta)^2]$$

$$= E\left[\left(\frac{1}{2} - \frac{1}{2}\right)^2\right]$$

$$= 0$$

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





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

For "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 given DGP?

② if 1 is true, then for any DGP/ θ, is there a procedure for locating that estimator with the least MSE?

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

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

Rephrase the 2 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 iid normal, $Var[\bar{x}] = \frac{\sigma^2}{n}$.

CRLB. $X_1, \ldots X_n \overset{iid}{\sim} DGP(\theta)$, Continuous...

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

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

$I(\theta)^{-1} \rightarrow$ The numerator is an irreducible core quantity based on the DGP and based on θ.

expectation of the squared log-likelihood.

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

Proof: This pure probability fact is proved in 368,

Cauchy-Schwartz Inequality for any two r.v's Q and S in

$$Cov[Q, S]^2 \leq Var[R] \, Var[S]$$

$$\Rightarrow Var[R] \geq \frac{Cov[R, S]^2}{Var[S]}$$

$$= \frac{(E[QS] - E[Q]E[S])^2}{E[S^2] - E[S]^2}$$

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

th:

Define the "Score function" S as:

$$S := \frac{\partial}{\partial \theta}\left[\ln f(x_1 \ldots x_n; \theta)\right] \quad (\text{def } 1)$$

chainrule $\downarrow$

$$= \frac{\frac{\partial}{\partial \theta}\left[f(x_1, \ldots x_n; \theta)\right]}{f(x_1, \ldots x_n; \theta)} \quad (\text{def } 2)$$

by iid, multiplication rule $\downarrow$

$$= \frac{\partial}{\partial \theta}\left[\ln \prod_{i=1}^{n} f(x_i, \theta)\right] \quad (\text{def } 3)$$

Precalc $\downarrow$

$$= \frac{\partial}{\partial \theta}\left[\sum_{i=1}^{n} \ln f(x_i, \theta)\right] \quad (\text{def } 4)$$

linearity of derivative $\downarrow$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \theta}\left[\ln f(x_i, \theta)\right] \quad (\text{def } 5)$$

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

$$= \frac{\partial}{\partial \theta}\left[\ell(\theta; x_1 \ldots x_n)\right] \ [\text{def } 6]$$

$$= \ell'(\theta, x_1 \ldots x_n) \ [\text{def } 7) = \sum_{i=1}^{n} \ell'(\theta; x_i) \ \text{def } [8]$$

Note: S is a rv, hence all X_i's are also rv's. hence capital letters.

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

$$E[S] \overset{def}{=} E\left[\frac{\frac{\partial}{\partial \theta}[f(x_1, \dots x_n; \theta)]}{f(x_1, \dots x_n, \theta)}\right]$$

$$= \overbrace{\int \cdots \int}^{n}_{\substack{\text{Support of} \\ n \text{ dim RV } X_1, \dots X_n}} \frac{\frac{\partial}{\partial \theta}[f(x_1, \dots x_n; \theta)]}{\cancel{f(x_1, \dots x_n; \theta)}} \cancel{f(x_1, \dots x_n, \theta)} \, dx_1 \dots dx_n$$

if you can interchange the derivative with Integral

$$= \frac{d}{d\theta}\left[\overbrace{\int \cdots \int_{Sup[\vec{x}]} f(x_1, \dots x_n; \theta) \, dx_1 \dots dx_n}^{1}\right]$$

$$= \frac{\partial}{\partial \theta}[1] = 0 \quad \text{(Fact 1a))}$$

$$E[S] \overset{def\ 7}{=} E[\ell'(\theta; x_1 \dots x_n)] = 0$$

$$E[S] \overset{def\ 8}{=} E[\Sigma \, \ell'(\theta; x_i)]$$

(th)

$$\overset{iid}{=} n E[l'(\theta; x_i)] = 0 \quad \leftarrow \text{fact (1a)}$$

$$\Rightarrow E[l'(\theta, x_i)] = 0 \quad (\text{fact } 1b)$$

$$Var[S] = E[S^2] - \cancel{E[S]^2} \rightarrow 0$$

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

$$\overset{def\ 8}{=} E\left[\left(\sum_{i=1}^{n} l'(\theta; x_i)\right)^2\right]$$

linearity of expectation.

$$= \sum_{i=1}^{n} E[l'(\theta, x_i)]^2 + \sum_{i \neq j} E[l'(\theta; x_i) \, l'(\theta; x_j)]$$