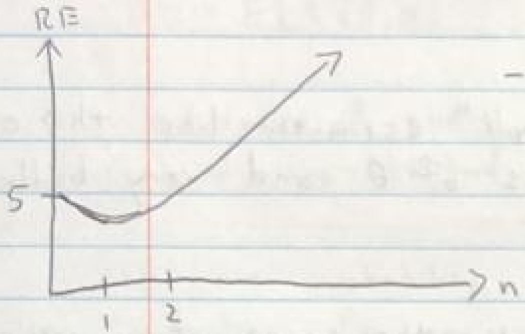
Lecture 09 10/5/20

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 when estimating θ for a given DGP?

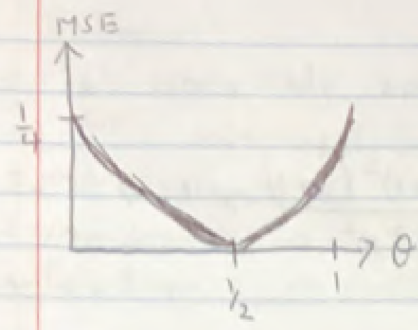
(2) If (1) is true, then for any DGP / θ, is there a procedure for locating that estimator with the best MSE?

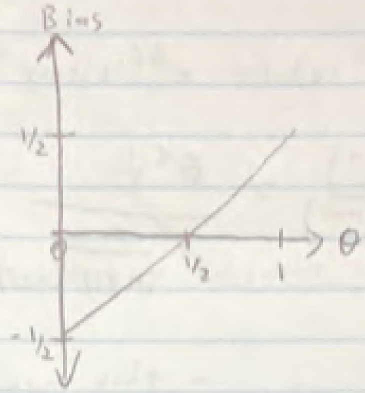
The answer to both is... No! (p 334 C&B) why? Because the class of "all" estimators is too big. For example...

$X_1, \ldots, X_n \overset{iid}{\sim} Bern(\theta)$

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

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





I can always create a "counter example" 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 counter example)...

(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 $\hat{\theta}^*$ such that for all θ and all other unbiased estimators $\hat{\theta}$,

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

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.

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

DGP iid Normal

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

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

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

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

$I(\theta) := E[\ell'(\theta; X)^2]$ - 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, Cauchy - Schwartz Inequality for any two r.v.'s Q and S:

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

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

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

Define the "score function" S as:

$$S := \frac{\partial}{\partial \theta}[\ln(f_X(x_1, \ldots, x_n; \theta))] \quad \text{(def. 1.)}$$

$$\overset{\text{chain rule}}{=} \frac{\frac{\partial}{\partial \theta}[f(x_1, \ldots, x_n; \theta)]}{f(x_1, \ldots, x_n; \theta)} \quad \text{(def 2.)}$$

$$\overset{\text{b/c iid multiplication rule}}{=} \underbrace{\frac{\partial}{\partial \theta}\left[\ln \prod_{i=1}^{n} f(x_i; \theta)\right]}_{\text{(def. 3.)}} \overset{\text{precalc}}{=} \underbrace{\frac{\partial}{\partial \theta}\left[\sum_{i=1}^{n} \ln f(x_i; \theta)\right]}_{\text{(def. 4.)}} \overset{\text{linearity of derivative}}{=} \underbrace{\sum_{i=1}^{n} \frac{\partial}{\partial \theta}[\ln f(x_i; \theta)]}_{\text{(def. 5.)}}$$

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

$$= \underset{\text{(def. 6)}}{\frac{\partial}{\partial\theta}\left[\ell(\theta; x_1, \dots x_n)\right]} = \underset{\text{(def. 7)}}{\ell'(\theta; x_1, \dots, x_n)} \overset{\text{def 5}}{=} \underset{\text{(def 8)}}{\sum_{i=1}^{n} \ell'(\theta; x_i)}$$

Note: S is a r.v., hence all x_i's are also r.v.'s hence capital X.

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

$$E[S] \overset{\text{def 2}}{=} E\left[\frac{\frac{\partial}{\partial\theta}[f(x_1, \dots x_n; \theta)]}{f(x_1, \dots, x_n; \theta)}\right] = \underbrace{\overbrace{\int \cdots \int}^{n}}_{\substack{\text{Support of the} \\ \text{n-dim r.v. } X_1, \dots X_n}} \frac{\frac{\partial}{\partial\theta}[f(x_1, \dots x_n; \theta)]}{\cancel{f(x_1, \dots, x_n; \theta)}} \cancel{f(x_1, \dots x_n; \theta)}\, dx_1, \dots dx_n$$

If you can interchange the derivative with the integral

$$\overset{\Downarrow}{=} \frac{\partial}{\partial\theta}\left[\overbrace{\int_{Supp[X]} \cdots \int f(x_1, \dots, x_n; \theta)\, dx_1 \dots dx_n}^{1}\right] = \frac{\partial}{\partial\theta}[1] = 0 \qquad \text{(Fact 1a)}$$

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

$$E[S] \overset{\text{def 8}}{=} E\left[\sum \ell'(\theta; x_i)\right] \overset{\text{iid}}{=} n E[\ell'(\theta; x_i)] \overset{\text{Fact 1a}}{=} 0 \Rightarrow E[\ell'(\theta; x_i)] = 0$$

$$Var[S] = E[S^2] - \cancel{E[S]}^{2} \quad (=0)$$

$$E[S^2] \overset{\text{def 8}}{=} E\left[\Big(\underbrace{\sum_{i=1}^{n} \ell'(\theta; x_i)}_{}\Big)^2\right] = \qquad (a_1 + a_2 + \dots a_n)^2 = \sum_{i=1}^{n} a_i^2 + \sum_{i \neq j} a_i a_j$$

and linearity of expectation

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

to be continued...