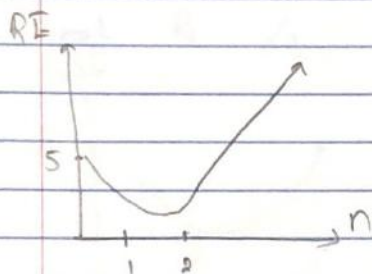


Define "relative efficiency" (RE) as the ratio of variances:

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

\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 ratios of MSE's? True... but in this case the tiny amount of bias in the MLE (see simulation) don't matter if n is large.

Two really important questions:

I Is there a theoretical minimum MSE when estimating θ for a given DGP?

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

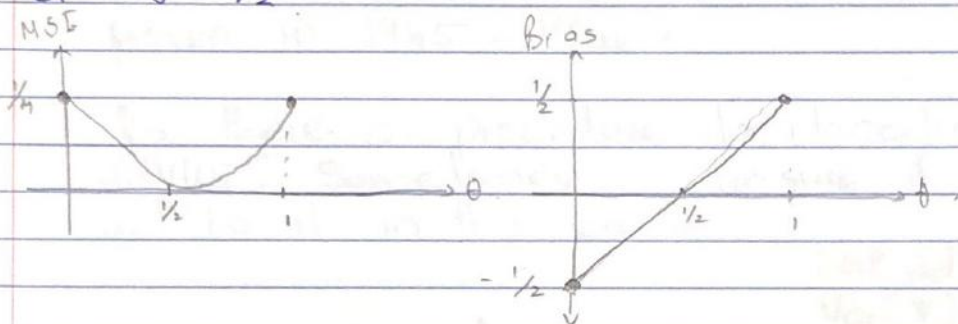
The answer to both is ... NO! (P334 CLB)

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

$$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bern}(\theta)$$

$$\begin{aligned}\hat{\theta}_{\text{BAD}} &= \frac{1}{2} \quad \text{MSE}[\hat{\theta}_{\text{BAD}}](\theta = \frac{1}{2}) = E[(\hat{\theta}_{\text{BAD}} - \theta)^2] \\ &= E[(\frac{1}{2} - \frac{1}{2})^2] = 0.\end{aligned}$$

This means that $\hat{\theta}_{\text{BAD}}$ 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 $\hat{\theta}$,

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

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.

DGP iid normal!

$$\text{Var}[X] = \sigma^2/n$$

CRLB $X_1, \dots, X_n \text{ iid DGP}(\theta)$, continuous
for any unbiased estimator $\hat{\theta}$,

$\text{Var}[\hat{\theta}] \geq \frac{I(\theta)}{n}$ 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
↑
Expectation of the squared log-likelihood. Fisher in 1922

Proof

The pure probability fact is proved is 368, Cauchy-Schwarz inequality for any two rv's Q and S is:

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

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

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

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

Define the "score function" S as:

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

chain rule \searrow

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

by prod. rule.

$$= \frac{\partial}{\partial \theta} \left[\ln \prod_{i=1}^n f(x_i; \theta) \right] \stackrel{\text{Pre-Calcu.}}{=} \frac{\partial}{\partial \theta} \left[\sum_{i=1}^n \ln f(x_i; \theta) \right]$$

linearity of derivative \searrow

$$= \sum_{i=1}^n \frac{\partial}{\partial \theta} [\ln f(x_i; \theta)] \quad (\text{def 3}) \quad (\text{def 4}) \quad (\text{def 5})$$

$$\mathcal{L} = \mathcal{L}, \quad \mathcal{L} := \ln(\mathcal{L}) = \ln(f), \quad \text{def 1}$$

$$= \frac{\partial}{\partial \theta} [\mathcal{L}(\theta; x_1, \dots, x_n)] = \mathcal{L}(\theta; x_1, \dots, x_n) = \sum_{i=1}^n \mathcal{L}'(\theta, x_i)$$

(def 6) (def 7) (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're done!

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

$$= \int \int \frac{\partial/\partial \theta [f(x_1, \dots, x_n; \theta)]}{f(x_1, \dots, x_n; \theta)} \cdot$$

Supp[x] Supp[x]

support of the

n -dim rv x_1, \dots, x_n .

$$[f(x_1, \dots, x_n; \theta) dx_1 \dots dx_n]$$

if you can interchange the derivative
with integral

$$= \frac{\partial}{\partial \theta} \left[\int \int f(x_1, \dots, x_n; \theta) dx_1 \dots dx_n \right]$$

Supp[x]

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

$$E[S] \stackrel{\text{def 7}}{=} E[\mathcal{L}'(\theta; x_1, \dots, x_n)] = 0.$$

$$E[S] \stackrel{\text{def 8}}{=} E\left[\sum \mathcal{L}'(\theta; x_i)\right]$$

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

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

$$\text{Var}[S] = E[S^2] - E[S]^2 \stackrel{\text{del } 8}{=} E\left[\left(\sum_{i=1}^n l'(\theta; x_i)\right)^2\right]$$

and linearity of expectation

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