Victoria Lombardi

Math 369 — 10/5/20

Lecture 9

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

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

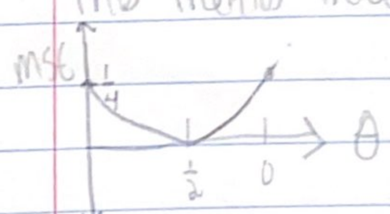
② If (1) is true, then for any DGP/ θ, is there a procedure for locating that estimator with the best MSE?

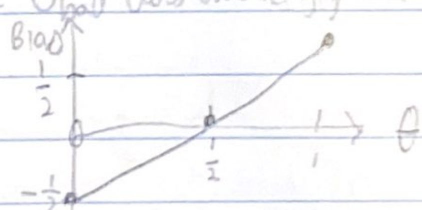
The answer to both is... NO! (p. 334 C+B) Why? Because the class of "all" estimators is too big. For ex...

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

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

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 amazing 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 & closes the loophole of the above counterex):

(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? (2)

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

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...

CRLB: $X_1, \ldots, X_n \overset{iid}{\sim} DGP(\theta)$ continuous... for any unbiased estimator $\hat{\theta}$,

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

the numerator is an illegauble core quantity based on the DGP and based on θ

$\boxed{DGP \text{ iid normal } Var[\bar{X}] = \frac{\sigma^2}{n}}$

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

expectation of the squared log-likelihood

Proof: Cauchy-Schwarz Inequality for any two rv's Q & S is:

$$Cov[Q,S]^2 \leq Var[Q]Var[S] \quad \text{(pure prob fact)}$$

$$\Rightarrow Var[Q] \geq Cov[Q,S]^2 / Var[S]$$

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

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

Define the "score function" S as:

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

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

by iid, mult rule $= \frac{\partial}{\partial\theta}\left[\ln \prod_{i=1}^{n} f(x_i;\theta)\right] \underset{\text{(def 3)}}{=} \frac{\partial}{\partial\theta} \sum_{i=1}^{n} \ln f(x_i;\theta)\Big] \underset{\text{(def 4)}}{=}$ linearity of derivative

pre calc

③

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

$L = f,\ l := \ln(L) = \ln(f)$, def 1

$= \frac{\partial}{\partial\theta}[l(\theta; x_1,\dots,x_n)] \underset{\text{(def 6)}}{=} l'(\theta; x_1,\dots,x_n) \underset{\text{(def 7)}}{=} \sum_{i=1}^{n} l'(\theta; x_i)$ (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!

$E[S] \underset{\text{def 2}}{=} E\left[\frac{\frac{\partial}{\partial\theta}[f(X_1,\dots X_n;\theta)]}{f(X_1,\dots,X_n;\theta)}\right] =$

$= \underbrace{\int\cdots\int}_{n} \frac{\frac{\partial}{\partial\theta}[f(x_1,\dots x_n;\theta)]}{f(x_1,\dots,x_n;\theta)} f(x_1,\dots x_n;\theta)\, dx_1\dots dx_n$

Support of the n-dim rv $X_1,\dots,X_n$

if you can interchange the derivative with the integral

(FACT 1A) $= \frac{\partial}{\partial\theta}\left[\int\cdots\int_{supp(\vec{x})} f(x_1,\dots,x_n;\theta)\,dx_1\dots dx_n\right] = \frac{\partial}{\partial\theta}[1] = 0$

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

$E[S] \underset{\text{def 8}}{=} E[\sum l'(\theta; X_i)] \overset{\text{iid}}{=} n E[l'(\theta; X_i)] \overset{\text{fact 1a}}{=} 0$

$\Rightarrow E[l'(\theta; X_i)] = 0$ (fact 1b)

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

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