Lecture 10 10/07/2020

due to $X_1, \dots, X_n \overset{iid}{\sim}$ (fact 1b)

$$\overset{\downarrow}{=} \underbrace{n \, \overbrace{E[\ell'(\theta; x)^2]}^{I(\theta)}}_{I_n(\theta)} + \underbrace{\sum_{i \neq j} E[\cancelto{0}{\ell'(\theta; x_i)}] E[\cancelto{0}{\ell'(\theta; x_j)}]}_{0}$$

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

$\hat{\theta} = w(X_1, \dots, X_n)$

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

assuming we can interchange differentiation and integration ....

$$\overset{\downarrow}{=} \frac{\partial}{\partial\theta}\left[\underbrace{\int \dots \int \hat{\theta} f(x_1, \dots x_n; \theta) dx_1, \dots dx_n}_{E[\hat{\theta}] = \theta}\right] = \frac{\partial}{\partial\theta}[\theta] = 1$$

putting it all together

$$Var[\hat{\theta}] \geq \frac{(\cancelto{1}{E[\hat{\theta} S]} - E[\hat{\theta}] \cancelto{0}{E[S]})^2}{\underbrace{E[S^2]}_{n I(\theta)} - \cancel{E[S]^2}} = \underbrace{\frac{I(\theta)^{-1}}{n}}_{CRLB}$$

This allow you to compute the variance of the best estimator (UMVUE) for most iid DGPs (which means you can then assess if an estimator is a UMVUE). How? You calculate the CRLB and calculate the variance of the estimator. If the two are the same, then it is truely the best. Let's do some examples. First, we need a fact ...

$$I(\theta) := E[\ell'(\theta; X)^2] \underset{\uparrow}{=} \underset{HW}{\ldots} = E[-\ell''(\theta; X)]$$

need to assume differentiation & integration can be interchanged just like in the proof of the CRLB.

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(\theta)$. $\hat{\theta} = \bar{X}$. Is $\hat{\theta}$ the UMVUE?

$$L(\theta; x) = \theta^x (1-\theta)^{1-x}$$

$$\ell(\theta; x) = x \ln(\theta) + (1-x)\ln(1-\theta)$$

$$\ell'(\theta; x) = \frac{x}{\theta} - \frac{1-x}{1-\theta}.$$

$$\ell''(\theta; x) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}.$$

$$-\ell''(\theta; x) = \frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}.$$

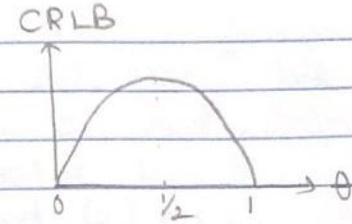
$$I(\theta) = E[-\ell''(\theta; X)]$$

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

$$= \frac{E[X]}{\theta^2} + \frac{1-E[X]}{(1-\theta)^2}.$$

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

$$\Rightarrow I(\theta)^{-1} = \theta(1-\theta)$$



$$\Rightarrow CRLB = \frac{\theta(1-\theta)}{n}$$

$$Var[\hat{\theta}] = Var[\bar{X}] = \frac{Var[X]}{n} = \frac{\theta(1-\theta)}{n}$$

$\hat{\theta}$ is the UMVUE!

$X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \theta_2)$ $\hat{\theta} = \bar{X}$ Is this the UMVUE?

$$L(\theta; x) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2\theta_2}(x-\theta)^2}$$

$$l(\theta; x) = -\tfrac{1}{2}\ln(2\pi\theta_2) - \frac{1}{2\theta_2}(x^2 - 2\theta x + \theta^2)$$

$$= -\tfrac{1}{2}\ln(2\pi\theta_2) - \frac{x^2}{2\theta_2} + \frac{\theta x}{\theta_2} - \frac{\theta^2}{2\theta_2}$$

$$l'(\theta; x) = \frac{x}{\theta_2} - \frac{\theta}{\theta_2}$$

$$l''(\theta; x) = -\frac{1}{\theta_2} \Rightarrow -l''(\theta; x) = 1/\theta_2$$

$$I(\theta) = E[1/\theta_2] = 1/\theta_2$$

$$\Rightarrow I(\theta)^{-1} = \theta_2 \qquad Var[\hat{\theta}] = \frac{Var[X]}{n} = \frac{\theta_2}{n}$$

$$\Rightarrow CRLB = \theta_2/n \qquad \Rightarrow \hat{\theta} \text{ is the UMVUE!}$$

Where did we come from so far? We started with the question "given a DGP, how do we come up with an estimator for θ?" We had two procedures (1) MM and (2) MLE. Then we observed that sometimes they have different performance (in MSE). And we asked "What's the best performance?" Assuming an estimator is unbiased, we proved the best performance is given by the CRLB formula. If an estimator has the CRLB variance, it is the UMVUE (i.e. the very very best).

Let's go back to testing. Let's say you found the MM or the MLE and you want to test H_0. What do you need to do this? You need the "sampling distribution" (the distribution of $\hat{\theta}$) either approximately (for an approximate test) or exactly (for an exact test). We need to derive it.

Definition:-

an estimator $\hat{\theta}$ is "asymptotically normal" if:

$$\hat{\theta}_{std} = \frac{\hat{\theta} - \theta}{SE[\hat{\theta}]} \xrightarrow{d} N(0,1)$$

This means as n gets large the $\hat{\theta}$-standardized distribution looks more & more like the $Z \sim N(0,1)$.

Is this possible to use the above as-is? Hardly ever. Here's why.

* DGP $\overset{iid}{\sim}$ Bern(θ), $\hat{\theta} = \bar{X}$, $SE[\hat{\theta}] = \sqrt{\frac{\theta(1-\theta)}{n}}$

By CLT

$$\Longrightarrow \frac{\hat{\theta} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \xrightarrow{d} N(0,1)$$

What's wrong with the above expression? you donot know θ. In a testing setting, the null hypothesis will assume it. But in general, it is unknown. In general,

$SE[\hat{\theta}](\theta_1, \ldots \theta_K)$ A quantity you need to know is a function of things you can never know.

* DGP $\stackrel{iid}{\sim}$ $N(\theta, \theta_2^2)$, $\hat{\theta} = \bar{X}$, $SE = \frac{\theta_2}{\sqrt{n}}$ ← unknown.

We need an estimate of the standard error without assuming we know the θ's:

$\hat{SE}[\hat{\theta}](\hat{\theta}_1, \ldots \hat{\theta}_K)$ function of estimates which come from data. SE-hat is an estimate of SE.

* DGP $\stackrel{iid}{\sim}$ Bern(θ) $\hat{\theta} = \bar{x}$, $SE[\hat{\theta}] \approx \hat{SE}[\hat{\theta}] = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$

wouldn't it be nice if the following were true - - -

$$\frac{\hat{\theta} - \theta}{\hat{SE}[\hat{\theta}]} \xrightarrow{d} N(0,1)$$

It's true if the estimators employed in $\hat{SE}$ are "consistent".

Definition for this class: an estimator $\hat{\theta}$ is consistent if you can estimate it for any degree of precision you wish given large enough sample size (n).

$$\hat{\theta} \xrightarrow{P} \theta$$

This type of convergence is called "convergence in probability" and it's done at the end of 368. But we're not going to need to know it.

Here are 2 technical theorems.

Thm 5.5.4 p233 C&B. Let $\hat{A}$ be a rv and c is a constant.

if $\hat{A} \xrightarrow{P} c$ then $h(\hat{A}) \xrightarrow{P} h(c)$ for h continuous

$$\Rightarrow \frac{\hat{A}}{c} = h(\hat{A}) \xrightarrow{P} h(c) = c/c = 1$$

$$\Rightarrow \frac{\hat{A}}{c} \xrightarrow{P} 1 \text{ (fact 2)}$$

$$\hat{SE}[\hat{\theta}] = h(\hat{\theta}) \xrightarrow{P} h(\theta) = SE[\hat{\theta}] \Rightarrow \hat{SE}[\hat{\theta}] \xrightarrow{P} SE[\hat{\theta}].$$

if $\hat{\theta} \xrightarrow{P} \theta$.

(fact 1)

Slutsky's Thm (Thm 5.5.17 p 239-240 C&B).
Let $\hat{A}$, $\hat{B}$ be rv's.

If $\hat{A} \xrightarrow{P} c$, $\hat{B} \xrightarrow{d} B \Rightarrow \hat{A}\hat{B} \xrightarrow{d} cB$

$$\underbrace{\hat{A} \xrightarrow{P} 1}\ (\text{by fact } 1,2)$$

$$\frac{\hat{\theta} - \theta}{\hat{SE}[\hat{\theta}]} = \underbrace{\frac{SE[\hat{\theta}]}{\hat{SE}[\hat{\theta}]}}_{\hat{A}} \underbrace{\frac{\hat{\theta} - \theta}{SE[\hat{\theta}]}}_{\hat{B}} \xrightarrow{d} 1\, N(0,1) = N(0,1)$$

Assume $\hat{B} \xrightarrow{d} N(0,1)$

We just proved that if $\hat{\theta}$ is asymptotically normal, then $\hat{\theta}$ standardized with a consistent estimate of its standard error is ALSO asymptotically normal

One of the most fundamental results in this class is the following:
Under some technical conditions,

① $\hat{\theta}^{MM}$, $\hat{\theta}^{MLE}$ are consistent

② $\hat{\theta}^{MM}$, $\hat{\theta}^{MLE}$ are asymptotically normal where

$SE[\hat{\theta}^{MM}]$, $\hat{SE}[\hat{\theta}^{MM}]$ are too difficult for this class but ....

$$SE[\hat{\theta}^{MLE}] = \sqrt{\frac{I(\theta)^{-1}}{n}} \quad \text{i.e. the CRLB!!}$$

$$\hat{SE}[\hat{\theta}^{MLE}] = \sqrt{\frac{I(\hat{\theta}^{MLE})^{-1}}{n}} \; *$$

③ $\hat{\theta}^{MLE}$ is called "asymptotically efficient" bc as n gets large, it provides the SMALLEST possible variance. The MM doesnt.

Proof next class ....