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

$$= \int_{Supp[X_1]}\cdots\int_{Supp[X_n]} \hat\theta \frac{\frac{\partial}{\partial\theta}[f(X_1,\ldots,X_n;\theta)]}{\cancel{f(X_1,\ldots,X_n;\theta)}} \cancel{f(X_1,\ldots,X_n;\theta)}\, dx_1\cdots dx_n$$

Assume we can interchange differentiation and integration

$$= \frac{\partial}{\partial\theta}\left[\int_{Supp[X_1]}\cdots\int_{Supp[X_n]} \hat\theta f(X_1,\ldots,X_n;\theta)\,dx_1\ldots dx_n\right] = \frac{\partial}{\partial\theta}\left[E[\hat\theta]\right]$$

$\hat\theta$ unbiased

$$= \frac{\partial}{\partial\theta}[\theta] = 1$$

$$\Rightarrow Var[\hat\theta] \ge \frac{((1)-\theta(0))^2}{(nI(\theta))-0^2} = \frac{I(\theta)^{-1}}{n} \checkmark$$

Let's go prove some of our estimators are UMVUE's.

$$X_1,\ldots,X_n \overset{iid}{\sim} Bern(\theta), \quad \hat\theta = \bar X, \text{ unbiased}, \quad Var[\bar X] = \frac{\theta(1-\theta)}{n}$$

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

Is that variance the CRLB variance?? If so, it's the UMVUE.

We need a more useful expression for I(θ). Before...

$$I(\theta) := E_X[\ell'(\theta;X)] = \overset{(HW)}{\cdots\cdots\cdots} = E_X[-\ell''(\theta;X)]$$

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

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

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

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

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

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

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

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

$$\Rightarrow \bar X \text{ is a UMVUE}$$

$$X_1,\ldots,X_n \overset{iid}{\sim} N(\theta_1,\theta_2), \quad \hat\theta_1 = \hat\theta = \bar X$$ Is it the UMVUE?

$$Var[\bar X] = \frac{\theta_2}{n} \overset{?}{=} \frac{I(\theta_1)^{-1}}{n} \quad \text{YES!}$$

$$\mathcal{L}(\theta_1;x) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2\theta_2}(x-\theta_1)^2}$$

$$\ell(\theta_1;x) = -\frac{1}{2}\ln(2\pi\theta_2) - \frac{1}{2\theta_2}(x^2 - 2x\theta_1 + \theta_1^2)$$

$\partial/\partial\theta_1[\ldots]$

$$= -\frac{1}{2}\ln(2\pi\theta_2) - \frac{x^2}{2\theta_2} + \frac{x\theta_1}{\theta_2} - \frac{\theta_1^2}{2\theta_2}$$

$$\ell'(\theta_1;x) = \frac{x}{\theta_2} - \frac{\theta_1}{\theta_2}, \quad \ell''(\theta_1;x) = -\frac{1}{\theta_2}$$

$$I(\theta_1) = E[-\ell''(\theta;x)] = E_X\left[\frac{1}{\theta_2}\right] = \frac{1}{\theta_2} \Rightarrow I(\theta_1)^{-1} = \theta_2$$

We now return to testing. First recall the CLT:

$X_1,\ldots,X_n \overset{iid}{\sim}$ DGP with mean θ and variance σ², $\hat\theta = \bar X$

$$\overset{CLT}{\Rightarrow} \frac{\hat\theta-\theta}{\frac{\sigma}{\sqrt n}} = \frac{\hat\theta-\theta}{SE[\hat\theta]} \xrightarrow{d} N(0,1)$$

This is the definition of $\hat\theta$ being "asymptotically normal". Thus Xbar is asymptotically normal (always if expectation exists).

If you don't know σ, you need to estimate it. You can use S, the sample std deviation for example:

$$\text{e.g. } \hat{SE}[\hat\theta] = \frac{S}{\sqrt n} \approx \frac{\sigma}{\sqrt n} = SE[\hat\theta]$$

$$S = \sqrt{\frac{1}{n-1}\sum(X_i - \hat\theta_1)^2}$$

Generally

$$SE[\hat\theta](\theta_1,\ldots,\theta_K),$$

$$\hat{SE}[\hat\theta](\hat\theta_1,\ldots,\hat\theta_K)$$

e.g. iid $Bern(\theta)$

$$\hat{SE}[\hat\theta](\hat\theta) = \sqrt{\frac{\hat\theta(1-\hat\theta)}{n}}$$

Is the following true?

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

For iid normal DGP, we know that

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

What kind of estimators for the parameters should we employ inside the estimator for the standard error? Beautiful theory follows if you employ "consistent estimators".

Def: an estimator is consistent if $\hat\theta \xrightarrow{p} \theta$. This means it "converges in probability" to θ. This convergence is not in this class (it's in 368). Convergence in probability means you can estimate θ with an degree of precision you wish given sufficient n. This usually implies asymptotic unbiasedness.

Thm 5.5.4 p222 Casella and Berger. Let $\hat A$ be a rv indexed by n and c ∈ ℝ. If $\hat A \xrightarrow{p} c$ then $h(\hat A) \xrightarrow{p} h(c)$ for any continuous function h.

Prove: $\frac{\hat A}{c} \xrightarrow{p} 1$ if $\hat A \xrightarrow{p} c$ (corr 5.5.4a)

$$h(\hat A) = \frac{\hat A}{c} \xrightarrow{p} h(c) = \frac{c}{c} = 1$$

Thm 5.5.4

$\frac{c}{\hat A} \xrightarrow{p} 1$

$$h(\hat A) = \frac{c}{\hat A} \xrightarrow{p} h(c) = 1$$

corr 5.5.4b

$$\text{If } \hat\theta \xrightarrow{p} \theta \Rightarrow \hat{SE}[\hat\theta] \xrightarrow{p} SE[\hat\theta]$$

assuming the SEhat function is continuous

Thm 5.5.17 (Slutsky's thm) p239-240 Casella & Berger.

Let $\hat A$, $\hat B$ be rv's indexed by n. If $\hat A \xrightarrow{p} c$ and $\hat B \xrightarrow{d} B$ (a distribution),

then $\hat A\hat B \xrightarrow{d} cB$

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

Slutsky's

$$\hat A \xrightarrow{p} 1 \text{ by corr 5.5.4b}$$

$$\hat B \xrightarrow{d} \overset{Z}{\overset{\shortparallel}{N(0,1)}} \text{ by CLT}$$

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

As long as the estimates employed internally in SEhat are consistent.

Here's the monster thm (we will prove parts of it).
Under some technical conditions...

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

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

$SE[\hat\theta^{MM}] =$ very difficult and we won't do it in our class

$\hat{SE}[\hat\theta^{MM}] =$ very difficult and we won't do it in our class

$SE[\hat\theta^{MLE}] = \sqrt{\frac{I(\theta)^{-1}}{n}}$ which means it has the CRLB for variance

$\hat{SE}[\hat\theta^{MLE}] = \sqrt{\frac{I(\hat\theta^{MLE})^{-1}}{n}}$ where $I(\hat\theta^{MLE})$ is an estimator for I(θ)

3. $\hat\theta^{MLE}$ is "asymptically efficient" which means as n gets larger, it provides the smallest possible variance. $\hat\theta^{MM}$ does not.

We will prove
* the asymptotic normality result and
* the asymptotic efficiency result

This means we want to show:

$$\frac{\hat\theta^{MLE}-\theta}{\sqrt{CRLB}} \xrightarrow{d} N(0,1)$$