Proof next class...

Lecture 11 10/14/2020

We want to prove the *asymptotic normality and asymptotic efficiency of the MLE thm*.
This means we want to show:

$$\frac{\hat{\theta}_{MLE} - \theta}{\sqrt{CRLB}} \xrightarrow{d} N(0,1) \Rightarrow \hat{\theta}^{MLE} \dot{\sim} N\left(\theta, \sqrt{\frac{I(\theta)^{-1}}{n}}^{\,2}\right)$$

$$CRLB := \frac{I(\theta)^{-1}}{n}$$

The asymptotic normality of the MLE is very useful but the asymptotic efficiency is like a huge bonus. The MLE estimates with approximately the theoretically guaranteed minimum variance.

The proof mostly follows from p. 472 of C&B.
Recall the Taylor series formula for $f(y)$ "centered at" a.

$$f(y) = \underbrace{f(a) + (y-a)f'(a)}_{\text{first order approx}} + (y-a)^2 \frac{f''(a)}{2} + \ldots$$

let $f = \ell'$, $y = \hat{\theta}^{MLE}$, $a = \theta$, we obtain:

$$\ell'(\hat{\theta}^{MLE}; X_1, \ldots, X_n) = \ell'(\theta; X_1, \ldots, X_n) + (\hat{\theta}^{MLE} - \theta)\ell''(\theta; x_1, \ldots, x_n) + \frac{(\hat{\theta}^{MLE} - \theta)^2}{2}\ell'''(\ \ ) + \ldots$$

Scanned with CamScanner

If you assume the technical conditions on p. 516 of C&B and a large enough sample size n, then the first order approximation can be employed:

$$\underbrace{\ell'(\hat{\theta}^{MLE}; X_1, \dots, X_n)}_{0} = \ell'(\theta; X_1, \dots, X_n) + (\hat{\theta}^{MLE} - \theta)\ell''(\theta; X_1, \dots, X_n)$$

$$\hat{\theta}^{MLE} := \operatorname{argmax}\{\mathcal{L}(\theta; X_1, \dots, X_n)\} = \operatorname{argmax}\{\ell(\theta; X_1, \dots, X_n)\}$$

$\Rightarrow$ Solve for θ in: $\ell'(\theta; X_1, \dots, X_n) \overset{set}{=} 0$

$$\Rightarrow 0 = \ell'(\theta; X_1, \dots, X_n) + (\hat{\theta}^{MLE} - \theta)\ell''(\theta; X_1, \dots, X_n)$$

$$\Rightarrow \hat{\theta}^{MLE} - \theta = -\frac{\ell'(\theta; X_1, \dots, X_n)}{\ell''(\theta; X_1, \dots, X_n)} \cdot \frac{\frac{1}{n}}{\frac{1}{n}} = \frac{\frac{1}{n}\ell'(\theta; X_1, \dots, X_n)}{-\frac{1}{n}\ell''(\theta; X_1, \dots, X_n)}$$

multiply both sides by $\frac{1}{\sqrt{\frac{I(\theta)^{-1}}{n}}}$

$$\Rightarrow \frac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} = \frac{\frac{1}{n}\ell'(\theta; X_1, \dots, X_n)}{-\frac{1}{n}\ell''(\theta; X_1, \dots, X_n)\sqrt{\frac{I(\theta)^{-1}}{n}}} \cdot \frac{I(\theta)}{I(\theta)}$$

$$= \underbrace{\frac{I(\theta)}{-\frac{1}{n}\ell''(\theta; X_1, \dots, X_n)}}_{\hat{A}} \cdot \underbrace{\frac{\frac{1}{n}\ell'(\theta; X_1, \dots, X_n)}{\sqrt{\frac{I(\theta)^{-1}}{n}}}}_{\hat{B}}$$

If we can prove that $\hat{A} \xrightarrow{p} 1$, $\hat{B} \xrightarrow{d} N(0,1)$, then we're done by Slutsky's thm.

Proof $\hat{A} \xrightarrow{p} 1$

Recall $\ell'(\theta; X_1, \dots, X_n) = \sum_{i=1}^{n} \ell'(\theta; X_i)$ Lec 9, def 7, 8 of score function

$$\Rightarrow \ell''(\theta; X_1, \dots, X_n) = \sum_{i=1}^{n} \ell''(\theta; X_i)$$ MATH 368 Law of Large Numbers

$$-\frac{1}{n}\ell''(\theta; x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{-\ell''(\theta; x_i)} = \frac{1}{n}\sum Y_i = \bar{Y} \xrightarrow{P} E[Y] = I(\theta)$$

Let $Y_i := -\ell''(\theta; X_i)$

$E[Y_i] = E[-\ell''(\theta; X_i)] \overset{HW}{=} \dots = I(\theta)$

By Thm 5.5.4, $\hat{A} \xrightarrow{P} 1$

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

$$\frac{1}{n}\ell'(\theta; X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n} \ell'(\theta; X_i) = \frac{1}{n}\sum W_i = \bar{W}$$

Let $W_i := \ell'(\theta; X_i)$

By the CLT, $\frac{\bar{W} - E[\bar{W}]}{SE[\bar{W}]} \xrightarrow{d} N(0,1)$.

by Fact 1b, lec 9.

$E[\bar{W}] = E[W] = E[\ell'(\theta; X_i)] = 0$

$$SE[\bar{W}] = \sqrt{\frac{Var[W]}{n}} = \sqrt{\frac{I(\theta)}{n}}$$

$$Var[W] = E[W^2] - \cancel{E[W]}^{2} = E[\ell'(\theta; X_i)^2] = I(\theta)$$

$$\hat{B} = \frac{\bar{W}}{\sqrt{\frac{I(\theta)^{-1}}{n}}} = \frac{\bar{W} - E[\bar{W}]}{SE[\bar{W}]} \xrightarrow{d} N(0,1).$$

This concludes the proof of the asymptotic normality and the asymptotic efficiency of the MLE.

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

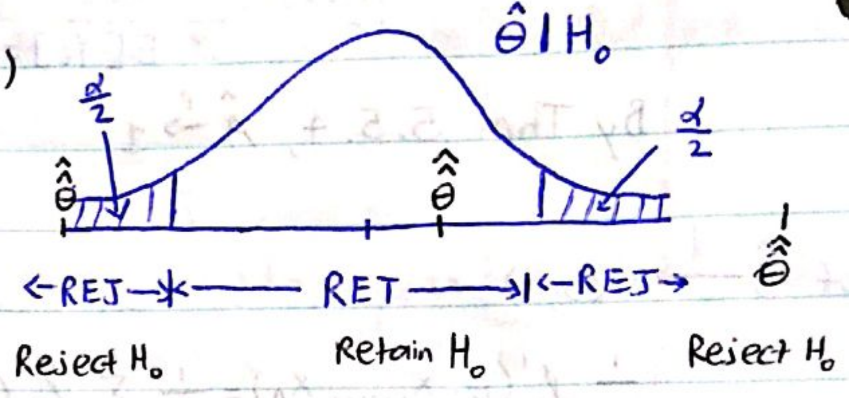
By one more use of Slutsky's thm, the above implies:

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

$$\frac{\hat{\theta}^{MLE} - \theta}{\sqrt{I(\hat{\theta}^{MLE})^{-1}/n}} \xrightarrow{d} N(0,1).$$

Let's use these thms to do "statistical inference", the name of the class. Recall we defined an "asymptotically normal estimator" as one which:

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



Using an asymptotically normal estimator (whether the normality comes from the CLT directly or from the fact that the MLE is asymptotically normal) to create an approximate z-test is called a "Wald Test" (p. 153 AoS). We've seen a Wald test before: the 1-proportion z-test. Let's review that.

Lec 1: $n = 20$, 12 had iphones $\Rightarrow \hat{\theta} = \bar{x}$, $\hat{\hat{\theta}} = 0.6$

Lec 4: $H_a: \theta \neq 0.524$, $H_0: \theta = 0.524$ DGP: $\overset{i.i.d.}{\sim}$ Bern (θ)

Generally $$\frac{\hat{\theta} - \theta}{SE[\theta]} = \frac{\hat{\theta} - \theta}{\sqrt{\theta(1-\theta)/n}} \xrightarrow{d} N(0,1)$$

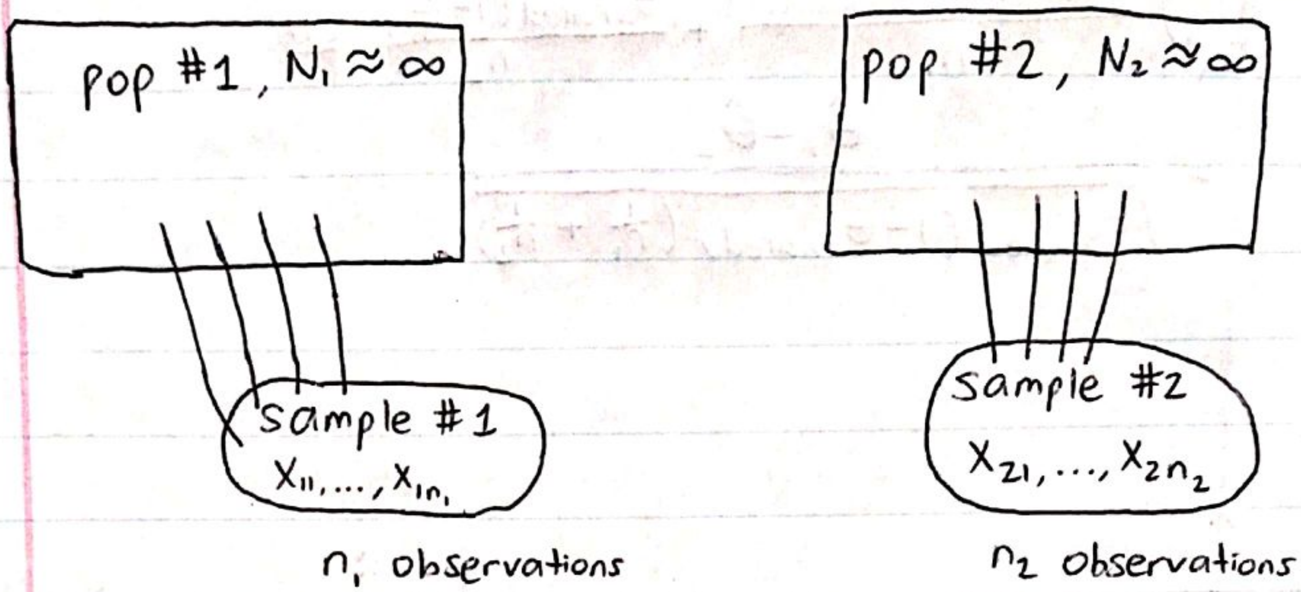
Under H_0, $$\frac{\hat{\theta} - 0.524}{\underbrace{\sqrt{\frac{0.524(1-0.524)}{20}}}_{0.112}} \mathrel{\dot\sim} N(0,1)$$

At $\alpha = 5\%$

$$\hat{\hat{\theta}}_{std} = \frac{\hat{\hat{\theta}} - \theta}{SE[\theta]} = \frac{0.6 - 0.524}{0.112} = 0.678 \in RET = [-1.96, 1.96]$$

$\Rightarrow$ Retain H_0

We never saw a 2-proportion test. We will now derive the approximate 2-proportion z-test as a wald test.



DGP: $X_{11}, \dots, X_{1n_1} \overset{i.i.d.}{\sim} Bern(\theta_1)$ independent of $X_{21}, \dots, X_{2n_2} \overset{i.i.d.}{\sim} Bern(\theta_2)$

$H_a: \theta_1 \neq \theta_2 \iff \theta_1 - \theta_2 \neq 0$, $H_0: \theta_1 = \theta_2 \iff \theta_1 - \theta_2 = 0$.

Now we pick an estimate that can reflect a departure from H_0. Why not $\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2$?

We need another fact from probability theory.

"$X_1, \dots, X_{n_1} \overset{i.i.d.}{\sim}$ with mean μ_1, variance σ_1^2, independent of
$Y_1, \dots, Y_{n_2}$ μ_2, σ_2^2," then...

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow[d]{} N(0,1)$$

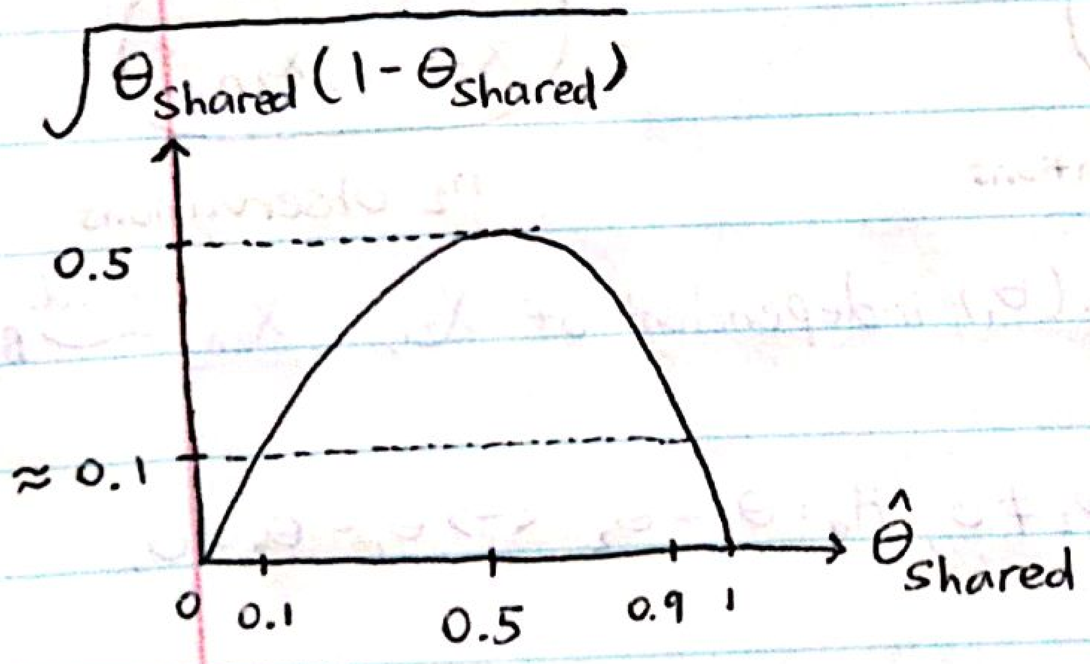
if n_1, n_2 are large

$$\frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \xrightarrow{d} N(0,1)$$

Under $H_0: \theta_1 = \theta_2 = \theta_{shared} \longleftrightarrow \theta_1 - \theta_2 = 0$

$$\Rightarrow \frac{(\hat{\theta}_1 - \hat{\theta}_2) - (0)}{\sqrt{\frac{\theta_{shared}(1-\theta_{shared})}{n_1} + \frac{\theta_{shared}(1-\theta_{shared})}{n_2}}}$$

$$= \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\theta_{shared}(1-\theta_{shared})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$



$$(\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2)_{std} = \frac{0.093}{\sqrt{?\left(\frac{1}{37} + \frac{1}{43}\right)}} = ?$$