Victoria Lombardi

Math 369 10/14/2020

Lecture 11

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

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

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

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

The proof mostly follows from p472 of C&B. Recall the Taylor series formula for f(y) "centered at" a.

$$f(y) = \underbrace{f(a) + (y-a)f'(a)}_{1^{st}\text{ order approx}} + (y-a)^2 \frac{f''(a)}{2} + \ldots$$

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

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

If you assume the technical conditions on p516 and a large enough sample size n, then the first order approximation can be employed:

$$\ell'(\hat{\theta}^{mle}; X_1, \ldots, X_n) = \ell'(\theta; X_1, \ldots, X_n) + (\hat{\theta}^{mle} - \theta)\,\ell''(\theta, X_1, \ldots, X_n)$$

$\hat{\theta}^{mle} := \text{argmax}\{L(\theta; x_1, \ldots, x_n)\} = \text{argmax}\{\ell(\theta; x_1, \ldots, x_n)\}$ ②

$\Rightarrow$ Solve for θ : $\ell'(\theta; x_1, \ldots, x_n) \overset{set}{=} 0$

$\Rightarrow 0 = \ell'(\theta; x_1, \ldots, x_n) + (\hat{\theta}^{mle} - \theta)\,\ell''(\theta; x_1, \ldots, x_n)$

$$\Rightarrow \hat{\theta}^{mle} - \theta = \frac{-\ell'(\theta; x_1, \ldots, x_n)}{\ell''(\theta; x_1, \ldots, x_n)} \cdot \frac{1/n}{1/n} = \frac{\frac{1}{n}\ell(\theta; x_1, \ldots, x_n)}{-\frac{1}{n}\ell''(\theta; x_1, \ldots, x_n)}$$

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

$$= \frac{\hat{\theta}^{mle} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} = \frac{\frac{1}{n}\ell'(\theta; x_1, \ldots, x_n)}{-\frac{1}{n}\ell''(\theta; x_1, \ldots, x_n)\sqrt{\frac{I(\theta)^{-1}}{n}}} \cdot \frac{I(\theta)}{I(\theta)}$$

$$= \underbrace{\frac{I(\theta)}{-\frac{1}{n}\ell''(\theta; x_1, \ldots, x_n)}}_{\hat{A}} \cdot \underbrace{\frac{\frac{1}{n}\ell'(\theta; x_1, \ldots, x_n)}{\sqrt{\frac{I(\theta)}{n}}}}_{\hat{B}}$$

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 $\ell'(\theta; x_1, \ldots, x_n) = \sum_{i=1}^{n} \ell'(\theta; x_i)$ Lec 9, def 7.8 of score function

$\Rightarrow \ell''(\theta; x_1, \ldots, x_n) = \sum_{i=1}^{n} \ell''(\theta; x_i)$

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

let $Y_i := \ell''(\theta; x_i)$

$E[Y_i] = E[\ell''(\theta; x_i)] \overset{HW}{=} \cdots = I(\theta)$

__By Thm 5.5.4 $\hat{A} \xrightarrow{p} 1$__

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

$$\frac{1}{n}l'(\theta; x_1,\dots,x_n) = \frac{1}{n}\sum_{i=1}^{n} l'(\theta; x_i) = \frac{1}{n}\sum w_i = \bar{w}$$

let $w_i := l'(\theta; x_i)$

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

$E[\bar{w}] = E[w] = E[l'(\theta; x_i)] = 0$ by Fact 1b, lec 9

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

$$var[w] = E[w^2] - \overbrace{E[w]}^{0}{}^2 = E[l'(\theta; x_i)^2] = I(\theta)$$

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

This concludes the proof of the asymptotic normality & the asymptotic efficiency of the MLE

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

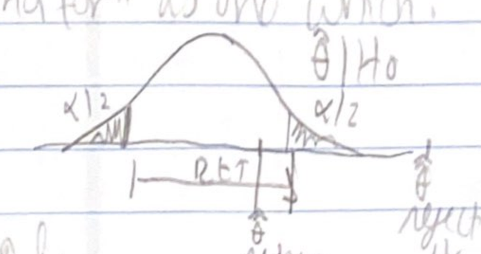
By one more use of Slutsky's thm, the above implies:

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

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

Lets use these theorems to do "statistical inference", the name of the class. Recall we define 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" (p153 AoS). We've seen a Wald test before: the 1 proportion z test. Let's review that. (4)

Lec 1: $n = 20$, 12 had iphones $\Rightarrow \hat{\theta} = \bar{x}$, $\hat{\theta} = .6$

Lec 4: $H_a: \theta \neq .524$, $H_0: \theta = .524$ DGP: $\overset{iid}{\sim} \text{Bern}(\theta)$

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

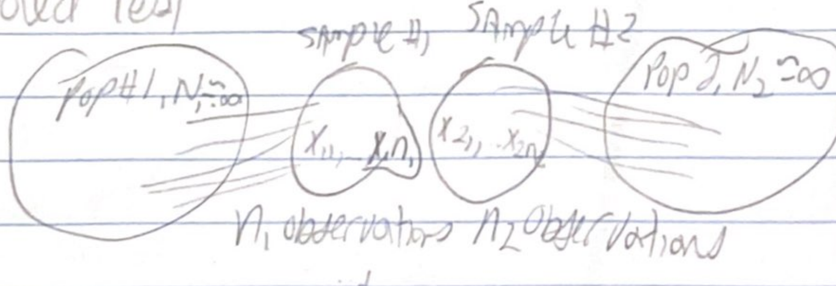
Under H_0, $\dfrac{\hat{\theta} - 0.524}{\underbrace{\sqrt{\dfrac{0.524(1-0.524)}{20}}}_{0.112}} \dot{\sim} N(0,1)$

At $\alpha = 5\%$

$\hat{\hat{\theta}}_{STD} = \dfrac{\hat{\theta} - \theta}{se(\theta)} = \dfrac{0.6 - 0.524}{0.112} = 0.678 \in RET = [-1.96, 1.96]$

$\Rightarrow$ 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: $X_{11}, \ldots, X_{1n_1} \overset{iid}{\sim} \text{Bern}(\theta_1)$ independent of
$X_{21}, \ldots, X_{2n_2} \overset{iid}{\sim} \text{Bern}(\theta_2)$

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

Now we pick an estimator 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. ⑤

$X_1, \dots, X_{n_1} \overset{iid}{\sim}$ w/ mean μ_1, variance σ_1^2; indep of

$Y_1, \dots, Y_{n_2} \overset{iid}{\sim}$ w/ mean μ_2, variance σ_2^2, then

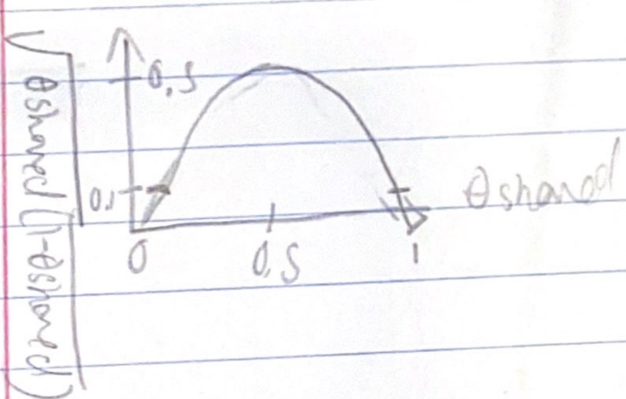
$$\frac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow{d} N(0,1)$$

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 $H_0: \theta_1 = \theta_2 \Leftrightarrow \theta_1 - \theta_2 = 0$

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



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