Lecture 11:

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

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

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

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

The Proof mostly follows from P472 of C&B. Recall the Taylor Series formula for Y "Centered at" a.

$$f(Y) = f(a) + (Y-a)\, f'(a) + (Y-a)^2 \frac{f''(a)}{2} + \dots$$

Let $f = l'$, $y = \hat{\theta}^{MLE}$, $a = \theta$, we obtain;

$$l'(\hat{\theta}^{MLE}, x_1 \dots x_n) = l'(\theta; x_1 \dots x_n) + (\hat{\theta}^{MLE} - \theta) \cdot$$

$$l''(\theta; x_1 \dots x_n) + \left(\frac{\hat{\theta}^{MLE} - \theta}{2}\right) l'''(\ ) + \dots$$

if you assume the technical conditions on p516 of C&B and a large enough sample size n, then the first order approximation can be employed;

$$l'(\hat{\theta}^{MLE}; x_1 \dots x_n) = l'(\theta; x_1, \dots x_n) + (\hat{\theta}^{MLE} - \theta) \, l''(\theta; x_1 \dots x_n)$$

$$\hat{\theta}^{MLE} = \arg\max \{ L(\theta; x_1, \dots x_n) \}$$

$$= \arg\max \{ l(\theta; (x_1 \dots x_n)) \}$$

$\Rightarrow$ Solve for θ in: $l'(\theta; x_1 \dots x_n) \stackrel{set}{=} 0$

$$\Rightarrow 0 = l'(\theta; x_1 \dots x_n) + (\hat{\theta}^{MLE} - \theta) \, l''(\theta; x_1 \dots x_n)$$

$$\hat{\theta}^{MLE} - \theta = -\frac{l'(\theta; x_1 \dots x_n)}{l''(\theta; x_1 \dots x_n)} \cdot \frac{\frac{1}{n}}{\frac{1}{n}}$$

$$= \frac{\frac{1}{n} l'(\theta; x_1 \dots x_n)}{\frac{1}{n} l''(\theta; x_1 \dots x_n)}$$

multyply both side by $1 / \sqrt{\frac{I(\theta)^{-1}}{n}}$

$$\Rightarrow \frac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} = \frac{\frac{1}{n} l'(\theta; x_1 \dots x_n)}{\frac{1}{n} l''(\theta; x_1 \dots x_n)} \cdot \frac{I(\theta)}{I(\theta)}$$

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

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 $l'(\theta; x_1 \dots x_n) = \sum_{i=1}^{n} l'(\theta, x_i)$ lec 9 def 7,9 of Score func.

$$\Rightarrow l''(\theta; x_1 \dots x_n) = \sum_{i=1}^{n} l''(\theta; x_i)$$

Previous page:

$$-\frac{1}{n} l''(\theta; x_1 \ldots x_n) \overset{\downarrow}{=} -\frac{1}{n} \sum_{i=1}^{n} - l''(\theta; x_i)$$

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

$$= -\frac{1}{n} \sum Y_i$$

Math 368, Law of Large Numbers

$$= \overline{Y} \xrightarrow{P} E[Y] = I(\theta)$$

$$E[Y_i] = E[-l''(\theta; x_i)] = \ldots \overset{H.W}{\ldots} = I(\theta)$$

By theorm 2.5.4, $\hat{A} \xrightarrow{P} 1$.

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

$$\frac{1}{n} l'(\theta; x_1 \ldots x_n) = \frac{1}{n} \sum_{i=1}^{n} l'(\theta; x_i)$$

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

$$= \frac{1}{n} \sum w_i$$

$$= \overline{w}$$

By the CLT; $\dfrac{\overline{W} - E[W]}{SE[\overline{W}]} \xrightarrow{d} N(0,1)$

$$E[\bar{W}] = E[W] + E[\cancel{n}\, l'(\theta; x_i)] \overset{\text{by Fact 1b, 4}}{=} 0$$

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

$$Var[W] = E[W^2] - \cancel{E[W]^2}^{\,0} = E[l'(\theta; X_i)^2] = I(\theta)$$

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

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

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

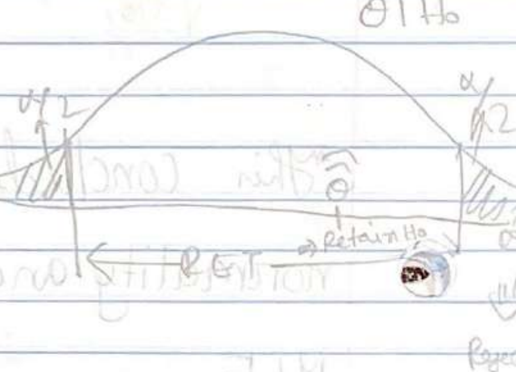
by one more use of Slutsky's thm, the above implies:

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

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

Let's use these theorems to do "Statistically 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" (P153 AoS). We have seen a Wald test before: the 1-proportion Z-test. Let's review that lec1: $n=20$, 12 had iphone $\Rightarrow \hat{\theta}=\bar{x}$, $\hat{\theta}=0.6$

lec4: $H_a: \theta \neq 0.524$ $\quad H_0: \theta = 0.524$.

Generally, DGP: $\overset{iid}{\sim}$ Bern(θ).

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

Under H_0,

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

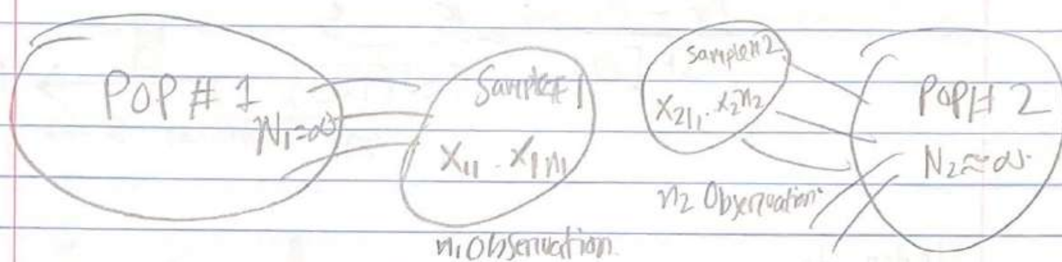
$$\hat{\theta}_{std} = \frac{\hat{\theta}-\theta}{SE[\theta]} = \frac{0.6-0.524}{0.112} \qquad \text{At } \alpha = 5\%$$

$$= 0.678 \in RET = [-1.96, 1.96]$$

$\Rightarrow$ 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: $X_{11}, \dots, X_{1n} \overset{iid}{\sim} Bern(\theta_1)$ independent of $X_{21} \dots X_{2n_2} \overset{iid}{\sim} Bern(\theta_2)$.

$H_a: \theta_1 \neq \theta_2 \iff \theta_1 - \theta_2 \neq 0$

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

Now we pick an estimator that can reflect a departure from H_0

Why not $\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2$

Hwin368

We need another fact from probability theory

$X_1 \ldots X_n$ iid with mean μ_1, variance σ_1^2, independent of

$Y_1 - Y_n$ iid with 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: $\theta_1 = \overset{\theta_{shared}}{\theta_2} \iff \theta_1 - \theta_2 = 0$

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

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