Lec 11 10/14/2020

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

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

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

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 p470 of C&B. Recall the Taylor series formula for $f(y)$ "centered at" a.

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

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

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

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:

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

$$\hat{\theta}^{MLE} := \text{argmax}\left[\mathcal{L}(\theta; X_1, \dots, X_n)\right] = \text{argmax}\left[\ell(\theta; X_1, \dots, X_n)\right]$$

$$\implies \text{Solve for } \theta \text{ in}: \ell'(\theta; X_1, \dots, X_n) \stackrel{set}{=} 0$$

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

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

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

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

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

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 $\ell'(\theta; x_1, \dots, x_n) = \sum_{i=1}^{n} \ell'(\theta; x_i)$ Lec 9, def 7,8 of score function.

$$\Rightarrow \ell''(\theta; X_1, \dots, X_n) = \sum_{i=1}^{n} \ell''(\theta; X_i)$$

$$-1/n\, \ell''(\theta; X_1, \dots, X_n) = 1/n \sum_{i=1}^{n} \underbrace{-\ell''(\theta; X_i)} = 1/n \sum Y_i = \bar{Y} \xrightarrow{P} E[Y] = I(\theta)$$

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, $\dfrac{\bar{w} - E[\bar{w}]}{SE[\bar{w}]} \xrightarrow{d} N(0,1)$

by fact 1b, lec 9.

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

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

$$Var[w] = E[w^2] - \cancelto{0}{E[w]}^2 = E[\ell'(\theta; x_i)^2] = \mathcal{I}(\theta)$$

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

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

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

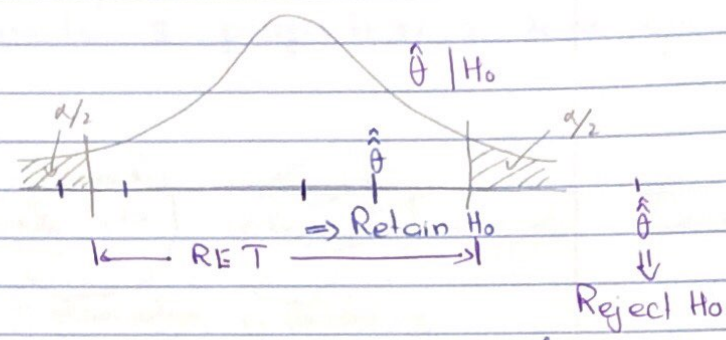
By one more use of Slutsky's thm, the above implies:

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

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

Let's use these theorems 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" (p153 AoS). We've seen a Wald test before: the 1-proportion z test. Let's review that.

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

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

Generally,

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

Under H_0,

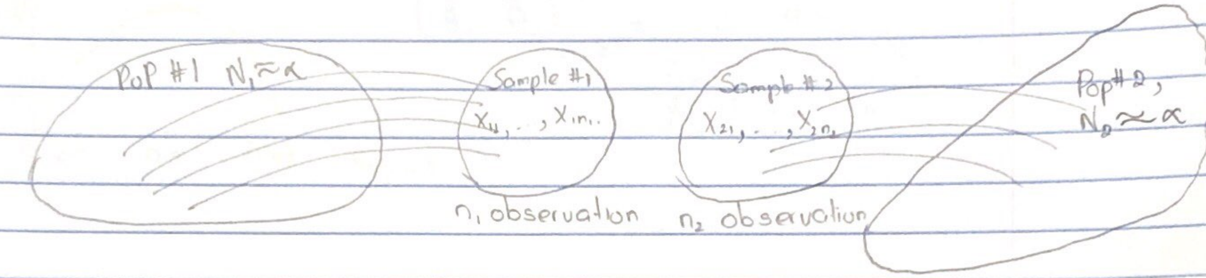
$$\frac{\hat{\theta}-0.524}{\underbrace{\sqrt{\frac{0.524(1-0.524)}{20}}}_{0.112}} \;\dot{\sim}\; N(0,1)$$

at $\alpha = 5\%$,

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

$\Rightarrow$ 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: $X_{11}, \dots, X_{1n_1} \overset{iid}{\sim} Bern(\theta_1)$ 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 estimate that can reflect a departure from H_0.
Why not $\hat{\theta}_1 - \hat{\theta}_2$?

We need another fact from probability theory. HW in 368

$X_1, \dots, X_n \overset{iid}{\sim}$ with mean μ_1, Variance σ_1^2, independent of $Y_1, \dots, Y_{n_2} \overset{iid}{\sim}$ with mean μ_2, variance σ_2^2, then ...

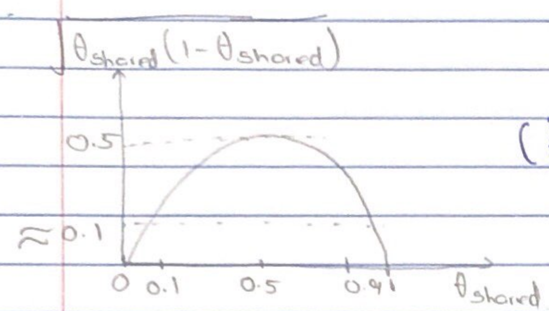
$$\frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow{d} N(0,1)$$

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

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



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