10/14/20

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

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

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

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 p472 of C&B. Recall the taylor series formula for $f(y)$ "centered at" $a$.

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

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

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

If you assume the technical condition on p516 of C&B and a large enough sample size $n$, then the first order approx can be employed:

$$\underbrace{l'(\hat{\theta}^{MLE}; X_1, \ldots X_n)} = l'(\theta; X_1 \ldots X_n) + (\hat{\theta}^{MLE} - \theta) l''(\theta; X_1 \ldots X_n)$$

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

$\Rightarrow$ Solve for $\theta$ in: $l'(\theta; X_1 \ldots X_n) \overset{set}{=} 0$

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

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

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

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

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

If we can prove $\hat{A} \xrightarrow{P} 1$; $\hat{B} \xrightarrow{d} N(0,1)$ then we're by slutsky's thm

Lec 9, def 7,8 of score function

$\boxed{\text{proof } \hat{A} \xrightarrow{P} 1}$

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

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

Math 368 Law of Large numbers

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

$= I(\theta)$

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

$E[Y_i] = E[-l''(\theta, x_i)] \overset{HW}{=} \dots = I(\theta)$

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

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

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

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

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

by Fact 1b, lec 9

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

$SE[\overline{w}] = \sqrt{\frac{Var[w]}{n}}$

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

$$Var[w] = E[w^2] - \cancel{E[w]}^{\,0\,2} = E\left[\ell'(\theta; X_i)^2\right] = I(\theta)$$

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

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

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

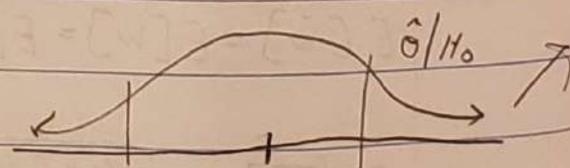
By one more use of slutsky's thm the above implies

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

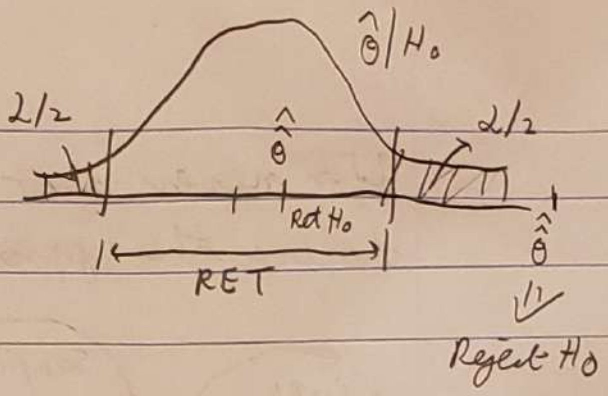
$$\frac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\hat{\theta}^{MLE})^{-1}}{n}}} \xrightarrow{d} N(0,1)$$

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

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



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




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 iphone $\Rightarrow \hat{\theta} = \bar{X}$, $\hat{\hat{\theta}} = 0{,}6$

Lec 4: $H_a: \theta \neq 0{,}524$, $H_0: \theta = 0{,}524$

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

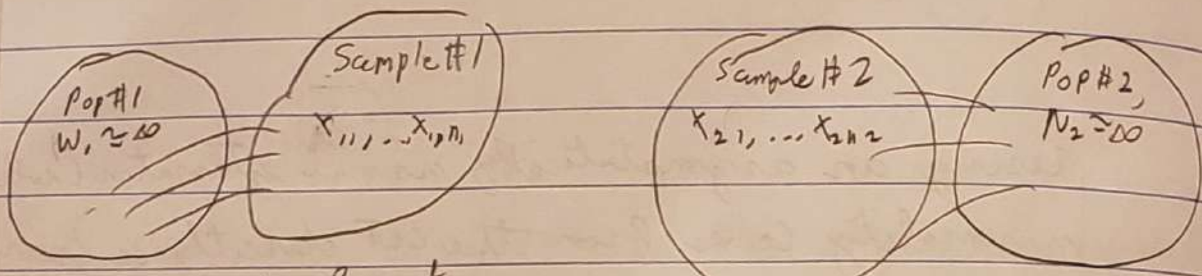
under $H_0$, $$\frac{\hat{\theta} - 0.524}{\underbrace{\sqrt{\frac{0.524(1-0.524)}{20}}}_{0.112}} \dot{\sim} N(0,1)$$

at $\alpha = 5\%$

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

$\Rightarrow$ Retain $H_0$

We never saw a 2-proportion test. We will now derive the approximate 2-proportion Z-test as a Wald test




$n_1$ observation $\qquad$ $n_2$ observation

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

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

now we pick an estimator that can reflect a departure from $H_0$

<u>why not $\hat{\theta}_1 - \hat{\theta}_2$?</u> $\qquad$ HW in 368

We need another fact from probability theory

$X_1, \dots X_n \overset{iid}{\sim}$ with mean $\mu_1$, variance $\sigma_1^2$ independent of $Y_1, \dots Y_n \overset{iid}{\sim}$ with mean $\mu_2$ " " $\sigma_2^2$, 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

mplies

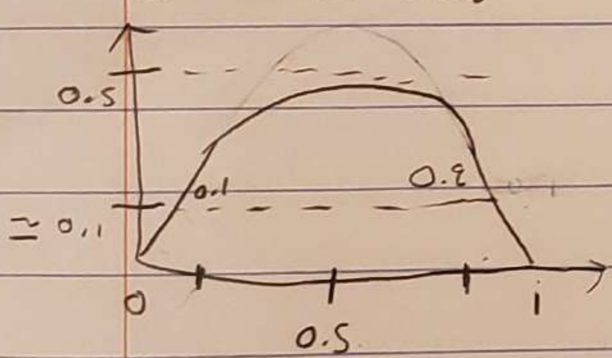
$$\frac{(\hat{\theta}_1-\hat{\theta}_2)-(\theta_1-\theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1}+\frac{\theta_2(1-\theta_2)}{n_2}}} \xrightarrow{d} N(0,1)$$

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

$= \theta_{shared}$

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

$\sqrt{\theta_{shared}(1-\theta_{shared}}$




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

can't get a number

didn't

assume in $\theta$ value