(Lec 12)

We want to prove the asymptotic normality & asymptotic efficienly of the MLE thrm^{d}. means we want to show

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

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

The asymptotic normality of the MLE is very usefull but the asymptotic efficiency is like a huge bonus. The MLE estimates approximatly the theoretically guaranteed minimum Varience

The Proof mostly follows from (Pg 472, C & B). Recall taylor series formula for $f(y)$ "centered at" a.

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

Let $f = l'$, $y = \hat{\theta}^{mle}$, $a = \theta$

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

If you assumed the technical conditions on Pg 516 of C & B & a large enough sample size n, then the first order approx can be employed:

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

$$\hat{\theta}^{mle} := \text{argument}\{\mathcal{L}(\theta; x_1 \ldots x_n) = \text{argmax}\{l; x_1 \ldots x_n)$$

$$\Rightarrow \text{Solve for } \theta \text{ in } l'(\theta; x_1 \ldots x_n) \overset{s.t}{=} 0$$

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

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

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

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

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

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

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

Recall $\ell'(\theta; x_1 \dots x_n) = \sum_{i=1}^{n} \ell'(\theta; x_i)$ (lec 9, def 7.8 ("score func"))

Math 368

$\Rightarrow \ell''(\theta; x_1 \dots x_n) = \sum_{i=1}^{n} \ell''(\theta; x_i)$ laws of large numbers

$\rightarrow -\frac{1}{n} \ell''(\theta; x_1 \dots x_n) = \frac{1}{n} \sum_{i=1}^{n} -\ell''(\theta; x_i) = \frac{1}{n} \sum y_i = \bar{Y} \xrightarrow{P} E(y) = I(\theta)$

let $y_i = -\ell''(\theta; x_i)$

$E[y_i] = E[-\ell''(\theta; x_i)] = \dots I(\theta)$

By thrm 5.5.4 $\hat{A} \xrightarrow{?} 1$

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 CLT, $\frac{\bar{W} - E(w)}{SE} \xrightarrow{d} N(0,1)$

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

$SE[\bar{w}] = \sqrt{\frac{Var(w)}{n}} = \sqrt{\frac{I(\theta)}{n}}$, $Var[w] = E[w^2] - \cancel{E[w]}^{0\,2} = E[\ell'(\theta; x_i)^2]$

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

This concludes the Proof of the asymptotic normality & the asymptotic efficency 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

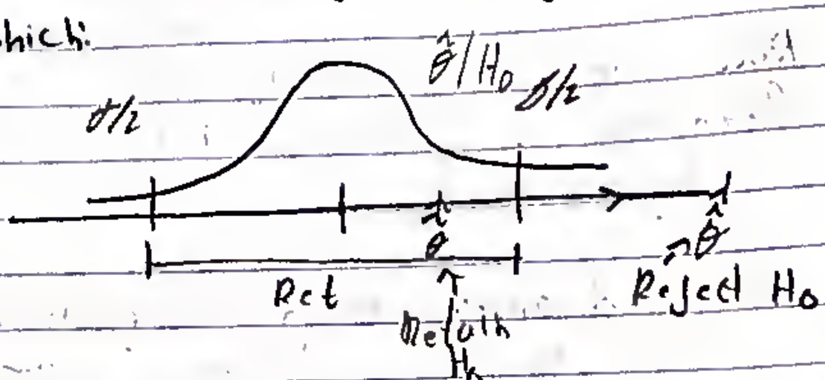
$\frac{\hat{\theta} - \theta}{SE[\theta]} \xrightarrow{d} N(0,1) \Rightarrow \frac{\hat{\theta} - \theta}{SE[\hat{\theta}]} \xrightarrow{d} N(0,1)$

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

Scanned with CamScanner

Lets use these thms to do "statistical infrences", 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 asyptotically normal estimator (whether the normality comes from CLT directly or from the fact that the MLE is asyptotically normal) to create an approx z test is called a "wald Test" (pg 153 AoS). We've seen a wald test before: the 1-Proportion z test

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

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

Generally $\frac{\hat{\theta}-\theta}{SE[\theta]} = \frac{\hat{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \xrightarrow{d} N(0,1)$ $\alpha = 5\%$

under H_0, $\frac{\hat{\theta}-.524}{\sqrt{\frac{.524(1-.524)}{20}}} \approx N(0,1)$ $\hat{\theta}_{std} = \frac{\hat{\theta}-\theta}{SE[\theta]} = \frac{.6-.524}{.112} = .678 \in Ret$

0.112 → Retain H_0 $[-1.96, 1.96]$

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

Pop #1 $N_1 \approx \infty$ → sample 1 $(X_{11} \cdots X_{n1})$ n_1 observed

Pop #2 $N_2 \approx \infty$ ↓ Sample 2 $(X_{21} \cdots X_{n_2})$ n_2 observed

DGP: $X_{11} \cdots X_{1n} \overset{iid}{\sim} Bern(\theta_1)$ independent of $X_{21} \cdots X_{2n_2} \overset{iid}{\sim} Bern(\theta_2)$

We want

nt ue $\rightarrow$ $H_a: \theta_1 \neq \theta_2 \iff \theta_1 - \theta_2 \neq 0$

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

Now we Pick estimates that can reflect a departure from H_0

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

We need another fact from probability theory

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

$Y_1 \dots Y_n$ iid " " μ_2 " " σ_2^2 then

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

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

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

θ_{shared}

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

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

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

$\theta_{shared}(1-\theta_{shared})$: