Lec 10 MATH 341/641

2-populations:

$X_{1,1}, \ldots, X_{1,n_1} \overset{iid}{\sim} \text{Bern}(\theta_1)$

$X_{2,1}, \ldots, X_{2,n_2} \overset{iid}{\sim} \text{Bern}(\theta_2)$

$\hat{\theta}_1 = \bar{X}_1, \quad \hat{\theta}_2 = \bar{X}_2$

$$\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \longrightarrow N(0,1) \text{ by CLT}$$

$\Downarrow$ Inverting the test

$$\hat{CI}_{\theta_1 - \theta_2, 1-\alpha} = \left[\bar{X}_1 - \bar{X}_2 \pm z_{1-\frac{\alpha}{2}} \underbrace{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}}_{\text{unknown! So you cannot compute this CI!}}\right]$$

BTW I can prove that

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\sqrt{\hat{\theta}_0(1-\hat{\theta}_0)}} \text{ has better power than } \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{n_1} + \frac{\bar{X}_2(1-\bar{X}_2)}{n_2}}}$$

but it has a closer size to α.

Why? less variance $\Rightarrow$ $\hat{z}$ closer to $N(0,1)$ than this by [illegible] by CMT

$$\bar{X}_1 \xrightarrow{p} \mu \Rightarrow \bar{X}_1(1-\bar{X}_1) \xrightarrow{p} \theta_1(1-\theta_1)$$

$$\bar{X}_2 \xrightarrow{p} \mu \Rightarrow \bar{X}_2(1-\bar{X}_2) \xrightarrow{p} \theta_2(1-\theta_2)$$

$$\underbrace{\frac{1}{n_1}}_{a}\underbrace{\bar{X}_1(1-\bar{X}_1)}_{C} + \underbrace{\frac{1}{n_2}}_{b}\underbrace{\bar{X}_2(1-\bar{X}_2)}_{D} \longrightarrow \underbrace{\frac{1}{n_1}}_{a}\underbrace{\theta_1(1-\theta_1)}_{c} + \underbrace{\frac{1}{n_2}}_{b}\underbrace{\theta_2(1-\theta_2)}_{d}$$

by Slutsky (B)

$$\Rightarrow \sqrt{\frac{1}{n_1}\bar{X}_1(1-\bar{X}_1) + \frac{1}{n_2}\bar{X}_2(1-\bar{X}_2)} \Rightarrow \sqrt{\frac{1}{n_1}\theta_1(1-\theta_1) + \frac{1}{n_2}\theta_2(1-\theta_2)}$$

by CMT

$$\Rightarrow \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{n_1} + \frac{\bar{X}_2(1-\bar{X}_2)}{n_2}}} = \underbrace{\frac{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{n_1} + \frac{\bar{X}_2(1-\bar{X}_2)}{n_2}}}}_{A} \cdot \underbrace{\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}}}_{B} \xrightarrow{d} N(0,1)$$

$A \xrightarrow{p} 1$ by CMT

$B \xrightarrow{d} N(0,1)$ by CLT

$\Downarrow$ invert test

$$\hat{CI}_{\theta_1-\theta_2, 1-\alpha} = \left[\bar{X}_1 - \bar{X}_2 \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{n_1} + \frac{\bar{X}_2(1-\bar{X}_2)}{n_2}}\right]$$

DGP: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ both unknown! But only θ is inferential target

$$\frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

neither testing $H_0: \theta = \theta_0$ nor building a CI is possible as σ is unknown!

In MATH 340, we showed:

$$\frac{\bar{X} - \theta}{\frac{S}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

which means both are possible!

due to the fact that $S \xrightarrow{p} \sigma$

$$RET = \left[\theta_0 \pm z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$$ doubly approximate!

$$\hat{CI}_{\theta, 1-\alpha} = \left[\bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$$ doubly approximately!

This is actually more powerful than you realize...

DGP: $X_1, \ldots, X_n \overset{iid}{\sim}$ $\theta = E(X) < \infty$, $\sigma^2 < \infty$

$$\frac{\bar{X} - \theta}{\frac{S}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

this is called the "Wald Test" after Abraham Wald. WW2 story...

These RET region and CI's work for any iid DGP with finite variance! Crazy power!

DGP: $X_{1,1}, \dots X_{1,n_1} \overset{iid}{\sim}$ with mean θ_1, variance σ_1^2, both unknown

$X_{2,1}, \dots X_{2,n_2} \overset{iid}{\sim}$ with mean θ_2, variance σ_2^2, both unknown

But only θ_1, θ_2 are the inferential targets

let $\hat{\theta}_1 = \bar{X}_1$, $\hat{\theta}_2 = \bar{X}_2$

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow{d} N(0,1) \quad \text{by CLT}$$

(not the problem in 340)

If $H_0: \theta_1 = \theta_2$ or $\theta_1 \leq \theta_2$ or $\theta_1 \geq \theta_2$ then...

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow{d} N(0,1)$$

but σ_1^2, σ_2^2 unknown!

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \underbrace{\frac{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}}_{A \xrightarrow{p} 1} \underbrace{\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}_{B \xrightarrow{d} N(0,1)} \xrightarrow{d} N(0,1)$$

Also

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow{d} N(0,1)$$

$\Downarrow$ Invert test

$$\hat{CI}_{\theta_1 - \theta_2, 1-\alpha} = \left[\bar{X}_1 - \bar{X}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]$$

Also $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$. Can we use $\frac{\bar{X}-\theta}{\hat{\sigma}/\sqrt{n}} \xrightarrow{d} N(0,1)$? Yes... 4b

Why use S_n^2 over $\hat{\sigma}_n^2$? Because its unbiased! $\hat{\sigma}_n^2 = \frac{1}{n}\sum(X_i-\bar{X})^2$

$S_n^2 = \frac{1}{n-1}\sum(X_i-\bar{X})^2$

$$E[S^2] = \frac{1}{n-1} E\left[\sum(X_i-\bar{X})^2\right]$$

$$= \frac{1}{n-1} E\left[\sum X_i^2 - 2X_i\bar{X} + \bar{X}^2\right]$$

$$= \frac{1}{n-1} E\left[\sum X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right]$$

$$= \frac{1}{n-1} E\left[\sum X_i^2 - n\bar{X}^2\right]$$

$$= \frac{1}{n-1}\left(nE[X_i^2] - nE[\bar{X}^2]\right)$$

$$= \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right)$$

$$= \frac{1}{n-1}\left(n\sigma^2 \cancel{+n\mu^2} - \sigma^2 \cancel{-n\mu^2}\right)$$

$$= \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

Note: $S_n^2 = \underbrace{\frac{n}{n-1}}\hat{\sigma}^2 = \frac{\cancel{n}}{n-1}\frac{1}{\cancel{n}}\sum(X_i-\bar{X})^2$

Bessel's correction

$MSE = \underbrace{Bias^2}_{small} + Var$

then transform a biased estimator to an unbiased estimator. This is an example of a "finite-sample bias correction"

Which is better? $\hat{\sigma}_n^2$ or S_n^2? It has different MSE's for different DGP's! But...

S_n^2 is a generally preferred default estimator. Also: the difference is slight regardless unless n is very small.

let's return to the "core MLE thm":

$$\hat{\theta}^{MLE} \sim N\left(\theta, \sqrt{\frac{(I(\theta))^{-1}}{n}}^{\,2}\right).$$

We will also make use of the fact which we will not prove generally:

$$\hat{\theta}^{MLE} \xrightarrow{p} \theta.$$

Recall Taylor series formula for $h(y)$ "centered at" a:

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

letting $h = \ell'$, the derivative of the log-likelihood (the score function)

$y = \hat{\theta}^{MLE}$,

$a = \theta$, we obtain:

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

Assuming technical conditions on p516 of C&B and large n, we use the first order approx:

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

Recall how to solve for MLE. Find $\ell'(\theta; X_1 \dots X_n) \overset{set}{=} 0$ and solve for θ

$\Rightarrow$ rhs is zero

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

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

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

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

$$\Rightarrow \frac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} = \underbrace{\frac{I(\theta)}{-\frac{1}{n}\, l''(\theta; X_1 \ldots X_n)}}_{A} \cdot \underbrace{\frac{\frac{1}{n}\, l'(\theta; X_1 \ldots X_n)}{\sqrt{\frac{I(\theta)}{n}}}}_{B} \overset{WTS}{\xrightarrow{d}} N(0,1)$$

If $A \xrightarrow{p} 1$, $B \xrightarrow{d} N(0,1) \Rightarrow AB \xrightarrow{d} N(0,1)$

by Slutsky's Thm. And we're done!

First prove $A \xrightarrow{p} 1$

Recall definition of score function

$$l'(\theta; X_1 \ldots X_n) = \sum_{i=1}^{n} l'(\theta; X_i) \Rightarrow l''(\theta; X_1 \ldots X_n) = \sum_{i=1}^{n} l''(\theta; X_i)$$

$$\Rightarrow -\frac{1}{n}\, l''(\theta; X_1 \ldots X_n) = \frac{1}{n} \sum_{i=1}^{n} -l''(\theta; X_i) = \frac{1}{n} \sum Y_i = \bar{Y} \xrightarrow{p} E(Y) = I(\theta)$$

let $Y_i := -l''(\theta; X_i)$, $E(Y_i) = \overset{HW}{\cdots} = I(\theta)$ by WLLN

$$A = \frac{I(\theta)}{L}$$

$L \xrightarrow{P} I(\theta)$

a continuous function in x

let $g(x) = \left(\frac{I(\theta)}{x}\right)$, $A = g(L)$

by CMT $g(L) \xrightarrow{P} g(I(\theta)) \Rightarrow A \xrightarrow{P} \frac{I(\theta)}{I(\theta)} = 1$ ✓

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

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

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

we proved this during the CRLB derivation

$$E[w_i] = \int_{\mathbb{R}} \frac{d}{d\theta}\left[\ln(f(x_i;\theta))\right] f(x_i;\theta)\, dx_i = \int_{\mathbb{R}} \frac{\frac{d}{d\theta}[f(x_i;\theta)]}{\cancel{f(x_i;\theta)}} \cancel{f(x_i;\theta)}\, dx_i = \frac{d}{d\theta} \int_{\mathbb{R}} f(x_i;\theta)\, dx_i = \frac{d}{d\theta}(1) = 0$$

$$Var[w_i] = E\left[l'(\theta;x_i)^2\right] - \cancel{E[l'(\theta;x_i)]^2}^{\,0} = I(\theta) \Rightarrow Var[\bar{w}] = \frac{I(\theta)}{n}$$

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

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