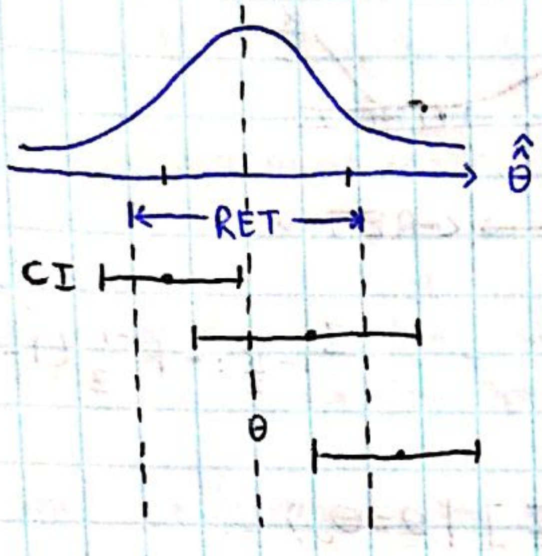
Lecture 13

10/21/2020



θ_0 being in the confidence interval with coverage probability $1-\alpha$, is equivalent to the test at size α retaining.

$$\hat{\hat{\theta}} \in RET_{\theta_0,\alpha}$$
$$\Updownarrow$$
$$\theta_0 \in CI_{\theta,1-\alpha}$$

p. 421 C&B: both hypothesis testing and interval construction look for consonance between the sample statistic ($\hat{\hat{\theta}}$) and the population parameter (θ).

Hypothesis tests fix the value of the parameter θ (H_0) and ask "is the estimate $\hat{\hat{\theta}}$ in agreement?" If no => Reject.

Confidence sets fixes the estimate ($\hat{\hat{\theta}}$) and asks "which values of the parameter (θ) are in agreement?"

We inverted a 2-sided hypothesis test to get a 2-sided CI. You can also have a 1-sided CI e.g.

$$CI_{L,\theta,1-\alpha} := [W_L(X_1,\dots,X_n), \infty) \text{ or } CI_{R,\theta,1-\alpha} := (-\infty, W_u(X_1,\dots,X_n)]$$

but we won't do this in class only for the interest of saving time and moving on to other topics.

Sometimes the sampling distribution was approximate. Inverting that test will yield CI's with approximate coverage i.e. "approximate CI's". Let's build some popular CI's!

DGP: $\overset{iid}{\sim} N(\theta, \sigma^2)$ with σ^2 unknown

$$CI_{\theta,1-\alpha} = [\hat{\hat{\theta}} \pm \underbrace{T_{1-\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}}}_{\text{margin of error}}] =$$

DGP: $X_{11}, \dots, X_{1n_1} \overset{i.i.d.}{\sim} N(\theta_1, \sigma_1^2)$ independent of
$X_{21}, \dots, X_{2n_2} \overset{i.i.d.}{\sim} N(\theta_2, \sigma_2^2)$, $\hat{\theta}_1 - \hat{\theta}_2 = \bar{X}_1 - \bar{X}_2$

if σ_1^2, σ_2^2 known

$$CI_{\theta_1-\theta_2, 1-\alpha} \overset{\downarrow}{=} \left[(\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

if $\sigma_1^2 = \sigma_2^2 = \sigma^2$ known

$$\overset{\downarrow}{=} \left[(\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2) \pm Z_{1-\frac{\alpha}{2}} \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

see lecture 6

if $\sigma_1^2 = \sigma_2^2$ but unknown

$$\overset{\downarrow}{=} \left[(\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2) \pm T_{1-\frac{\alpha}{2}, n_1+n_2-2}\, S_{pooled} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

if $\sigma_1^2 \neq \sigma_2^2$ and unknown

$$\overset{\downarrow}{\approx} \left[(\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2) \pm T_{1-\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right]$$

see lecture 7 for the Satterthwaite approximation

DGP: $\overset{i.i.d.}{\sim} \text{Bern}(\theta)$, $\hat{\theta} = \bar{x}$. via the CLT, $\dfrac{\hat{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \xrightarrow{d} N(0,1)$.

$$\Rightarrow \frac{\hat{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \overset{\cdot}{\sim} N(0,1) \xrightarrow[\text{Slutsky's}]{\text{via Thm 5.5.4 \&}} \frac{\hat{\theta}-\theta}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}} \xrightarrow{d} N(0,1)$$

Using this fact and following through...

$$\Rightarrow P\left(\frac{\hat{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \in \left[-Z_{1-\frac{\alpha}{2}}, +Z_{1-\frac{\alpha}{2}}\right]\right) \approx 1-\alpha$$

$$\Rightarrow P\left(\frac{\theta-\hat{\theta}}{\sqrt{\frac{\theta(1-\theta)}{n}}} \in \left[-Z_{1-\frac{\alpha}{2}}, +Z_{1-\frac{\alpha}{2}}\right]\right) \approx 1-\alpha$$

$$\Rightarrow P\left(\theta \in \left[\hat{\theta} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\theta(1-\theta)}{n}}, \hat{\theta} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\theta(1-\theta)}{n}}\right]\right) \approx 1-\alpha$$

$$\Rightarrow CI_{\theta, 1-\alpha} \approx \left[\hat{\hat{\theta}} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\theta(1-\theta)}{n}}, \hat{\hat{\theta}} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\theta(1-\theta)}{n}}\right]$$

this is a fail... I don't know θ!

Scanned with CamScanner

$$\Rightarrow CI_{\theta, 1-\alpha} \approx \left[\hat{\hat{\theta}} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\hat{\theta}}(1-\hat{\hat{\theta}})}{n}}, \hat{\hat{\theta}} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\hat{\theta}}(1-\hat{\hat{\theta}})}{n}} \right]$$

this is *a* CI for the binomial proportion. It is actually a bad approximation for low n and θ near 0 or 1. There are other CI's we won't study and it is actually an area of the modern research.

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

From Lecture 11,

$$\frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \xrightarrow{d} N(0,1) \overset{\text{Thm 5.5.4 \& Slutsky's}}{\Rightarrow} \frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}} \xrightarrow{d} N(0,1)$$

$$\Rightarrow CI_{\theta_1 - \theta_2, 1-\alpha} \approx \left[(\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\hat{\theta}}_1(1-\hat{\hat{\theta}}_1)}{n_1} + \frac{\hat{\hat{\theta}}_2(1-\hat{\hat{\theta}}_2)}{n_2}} \right]$$

e.g. from the medical study,

$n_1 = 81$, $\hat{\hat{\theta}}_1 = 0.333$, $n_2 = 79$, $\hat{\hat{\theta}}_2 = 0.152$

$$CI_{\theta_1 - \theta_2, 95\%} \approx \left[(0.333 - 0.152) \pm 1.96 \sqrt{\frac{0.333(0.667)}{81} + \frac{(0.152)(0.048)}{79}} \right]$$

$$= [0.181 \pm 1.96(0.066)] = [0.051, 0.311]$$

"You're 95% confident that the true proportion difference is between 5.1% and 31.1%.

By CLT, $\frac{\hat{\theta} - \theta}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1) \Rightarrow \frac{\hat{\theta} - \theta}{S/\sqrt{n}} \xrightarrow{d} N(0,1)$

DGP $\overset{i.i.d.}{\sim}$ Some R.V. with mean θ, variance σ^2 unknown. $\hat{\theta} = \bar{X}$

$CI_{\theta, 1-\alpha} \approx \left[\hat{\hat{\theta}} \pm z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right]$ if you use the T it won't be "so bad"

Problem 11 on Midterm I: $\bar{x} = 2.57$, $s = 1.00$

$$CI_{\theta, 95\%} \approx \left[2.57 \pm (1.96) \frac{1.00}{\sqrt{30}} \right] = [2.212, 2.928]$$

DGP $\xrightarrow{i.i.d.} f(\theta)$ where $\hat{\theta} = \hat{\theta}^{MLE}$

(From Lec 11), $\dfrac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} \xrightarrow{d} N(0,1) \implies \dfrac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\hat{\theta}^{MLE})^{-1}}{n}}} \xrightarrow{d} N(0,1)$

$$\Rightarrow CI_{\theta, 1-\alpha} \approx \left[\hat{\hat{\theta}} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{I(\hat{\theta}^{MLE})^{-1}}{n}}\right]$$

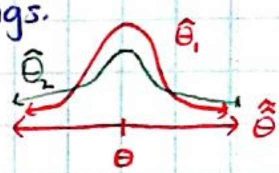
example from last class. DGP: i.i.d. Gumbel$(\theta, 1)$ and the data is $\langle 2.15, 1.91, 3.66, 4.85, 3.03, 1.03, 3.58 \rangle$ and $n=7$. Find a 95% CI for θ:

$\hat{\theta}^{MLE} = \ln\left(\frac{n}{\sum e^{-x_i}}\right)$, $\hat{\hat{\theta}}^{MLE} = 2.26$

$\sqrt{I(\theta)^{-1}} = e^{\theta} \qquad \Rightarrow \sqrt{I(\hat{\theta}^{MLE})^{-1}} = 9.57$

$CI_{\theta, 95\%} \approx \left[2.26 \pm (1.96)\left(\frac{9.57}{\sqrt{7}}\right)\right] = [0.58, 3.93]$

Now that we've been properly introduced to statistical inference (all 3 goals), let's talk about some big picture things.



For an unbiased estimator, MSE (being small) is KING. Why?

1) Point Estimation The lower the MSE, the closer $\hat{\hat{\theta}}$ is to θ on average.

2) Hypothesis Testing Most estimators we discussed with exactly or approximately normally distributed. Thus the RET for a 2-sided test looks like:

$$RET = [\theta_0 \pm z_{1-\frac{\alpha}{2}} \underbrace{\sqrt{MSE}}_{= SE[\hat{\theta}]}]$$

with a smaller MSE $\Rightarrow$ smaller RET $\Rightarrow$ higher power!

3) Confidence Intervals For exactly or approximately normal distributed estimators,

$$CI_{\theta, 1-\alpha} \overset{=}{\underset{\approx}{\text{or}}} \left[\hat{\hat{\theta}} \pm z_{1-\frac{\alpha}{2}} \sqrt{MSE}\right]$$

or $\sqrt{\hat{MSE}}$

A lower MSE means a tighter / smaller CI which means you're more confidence about where θ lies e.g.

$$CI_{\theta, 95\%} = [0.49, 5.1] \text{ vs. } CI_{\theta, 95\%} = [0.4999, 0.5001]$$

Let's picture all 3 goals: