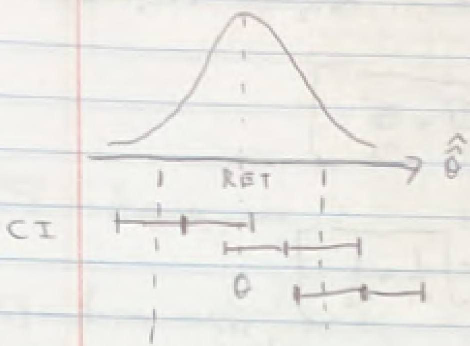
Lecture 13 10/21/20



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

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

p421 C2B: 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 and ask "is the estimate $\hat{\hat{\theta}}$ in agreement?" If no $\Rightarrow$ 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.:

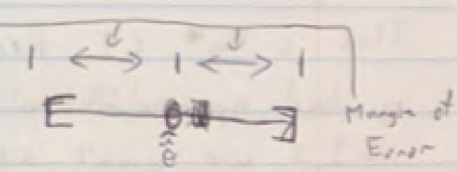
$$CI_{L,\theta,1-\alpha} := [w_L(x_1,\dots,x_n), \infty) \text{ or } CI_{R,\theta,1-\alpha} := (-\infty, w_R(x_1,\dots x_n)]$$

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.

$$CI_{\theta,1-\alpha} = \left[\hat{\hat{\theta}} \pm t_{1-\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right] =$$



$$\hat{\theta}_1 - \hat{\theta}_2 = \bar{X}_1 - \bar{X}_2$$

DGP: $X_{11}, \dots, X_{1n_1} \overset{iid}{\sim} N(\theta_1, \sigma_1^2)$ indep of $X_{21}, \dots, X_{2n_2} \overset{iid}{\sim} N(\theta_2, \sigma_2^2)$

if σ_1^2, σ_2^2 known

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

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

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

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

see lec 6.

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

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

$$\approx \left[(\hat{\theta}_1 - \hat{\theta}_2) \pm t_{1-\frac{\alpha}{2}, df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

see lec 7. for Satterthwaite approximation

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

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

using this fact and following through

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

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

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

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

this is a fail ... I don't know θ!

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

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 modern research.

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

from lec 11, $\dfrac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \xrightarrow{d} N(0,1) \xRightarrow[\text{Slutsky's}]{\text{Thm. 5.5.4.2}} \dfrac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}}$

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

e.g. from medical study, $n_1 = 81$, $\hat{\hat{\theta}}_1 = 0.333$, $n_2 = 79$, $\hat{\hat{\theta}}_2 = 0.152$

$$CI_{\theta_1-\theta_2, 95\%} \approx \left[(0.333 - 0.152) \pm 1.96\sqrt{\frac{0.333(.667)}{81} + \frac{(.152)(.848)}{79}}\right] = [.181 \pm 1.96(0.066)]$$

$$= [0.051, 0.311]$$

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

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

DGP $\overset{iid}{\sim}$ 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!"

$\dfrac{\hat{\theta} - \theta}{\frac{s}{\sqrt{n}}} \xrightarrow{d} N(0,1)$

Prob 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 $\overset{iid}{\sim} f(\theta)$ where $\hat{\theta} = \hat{\theta}^{MLE}$

from lec 11, $\dfrac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} \xrightarrow{d} N(0,1) \Rightarrow \dfrac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\hat{\theta}^{MLE})^{-1}}{n}}} \xrightarrow{d} N(0,1)$

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

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

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

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

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

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

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

(1) Point Estimation

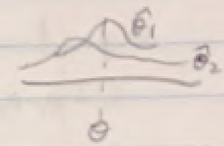
The lower the MSE, the closer $\hat{\hat{\theta}}$ is to θ on average.

(2) Hypothesis testing

Most estimators we discussed with exactly or approximately normal distributed. Thus the retainment region for a 2-sided test looks like:

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

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



$\hat{\theta}_1$ better than $\hat{\theta}_2$ b/c $\hat{\theta}_1$ closer to θ.

(3) Confidence Intervals

For exactly or approximately normally distributed estimators,

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

or $\sqrt{\widehat{MSE}}$

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

$$CI_{\theta,95\%} = [0.49, 0.51] \quad vs. \quad CI_{\theta,95\%}\ [0.4999, 0.5001]$$

Let's picture all three goals: