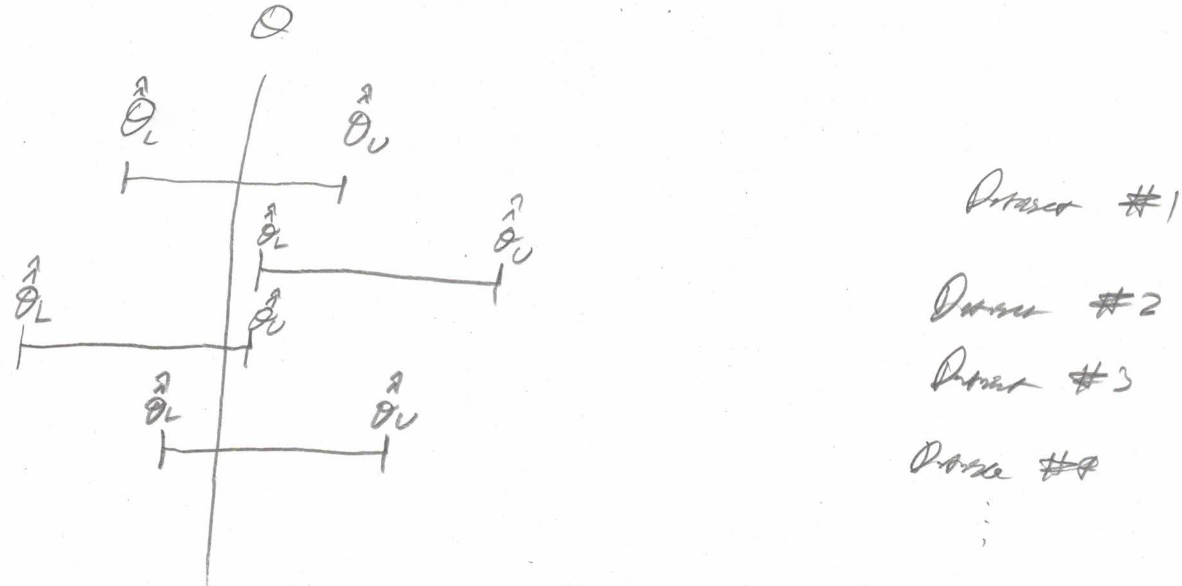
MATH 341/641 Lec 7

Coverage Prob. is best illustrated as follows:



The coverage prob. is computed over every possible dataset. If this were every dataset, cov. prob = 75%.

Def: A two-sided confidence interval estimator with cov. prob. $1-\alpha$ for param θ

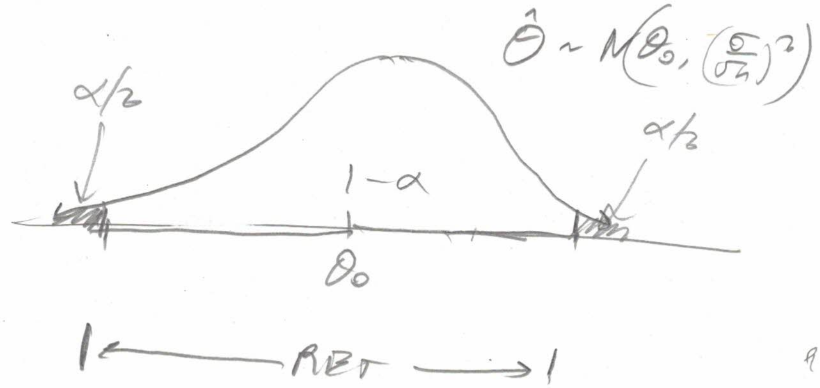
$$\hat{CI}_{\theta,1-\alpha} := \left[\hat{\theta}_{\frac{\alpha}{2}}, \hat{\theta}_{1-\frac{\alpha}{2}}\right]$$

A two-sided confidence interval estimate (just "confidence interval") corresponding to the above confidence interval estimator is

$$\hat{\hat{CI}}_{\theta,1-\alpha} := \left[\hat{\hat{\theta}}_{\frac{\alpha}{2}}, \hat{\hat{\theta}}_{1-\frac{\alpha}{2}}\right]$$

How do we derive the CI? As usual, we need to begin with a DGP, e.g. $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\theta, \sigma^2)$ where σ^2 is known and θ is the inferential target. And $\hat{\theta} = \bar{X}$. Just like in hypothesis testing, you need to pick level of confidence $1-\alpha$ usually 95%, 99%.

Now imagine $\theta = \theta_0$ just like the null sampling distr for $H_0: \theta = \theta_0$



$a \in [3, 7]$
$-a \in [-7, -3]$

$$1-\alpha = P(\hat{\theta} \in RET \mid \theta = \theta_0)$$

$$= P\left(\hat{\theta} \in \left[\theta_0 - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \theta + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \mid \theta = \theta_0\right)$$

where $z_{1-\frac{\alpha}{2}} := \{z : \Phi(z) = 1-\frac{\alpha}{2}\}$

$\Phi^{-1}(1-\frac{\alpha}{2})$

$$= P\left(\hat{\theta} - \theta_0 \in \left[-z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \mid \theta = \theta_0\right)$$

for $\alpha = 5\% \Rightarrow z_{1-\frac{\alpha}{2}} = 2$

$$= P\left(\theta_0 - \hat{\theta} \in \left[-z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \mid \theta = \theta_0\right)$$

$$= P\left(\theta_0 \in \left[\underbrace{\hat{\theta} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}_{\hat{\theta}_L}, \underbrace{\hat{\theta} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}_{\hat{\theta}_U}\right] \mid \theta = \theta_0\right)$$

valid for every $\theta_0 \in \Theta$!

$$\Rightarrow \hat{CI}_{\theta, 1-\alpha} = \left[\hat{\theta} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \hat{\theta} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \Rightarrow \hat{CI}_{\theta, 1-\alpha} = \left[\hat{\theta} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \hat{\theta} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

$$= \left[\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

We constructed an interval estimator with the correct coverage prob. by "inverting the test"!

why did this work?

$$\hat{\theta} \in RET_\alpha \iff \theta_0 \in \hat{CI}_{\theta,1-\alpha}$$

$$RET_\alpha = \left[\theta_0 \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \qquad \hat{CI}_{\theta,1-\alpha} = \left[\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

p421 C&B: Hypothesis tests fix the parameter value and asks "which estimates are consistent?"
Confidence sets fixes the estimate and asks "which parameter values are consistent"?

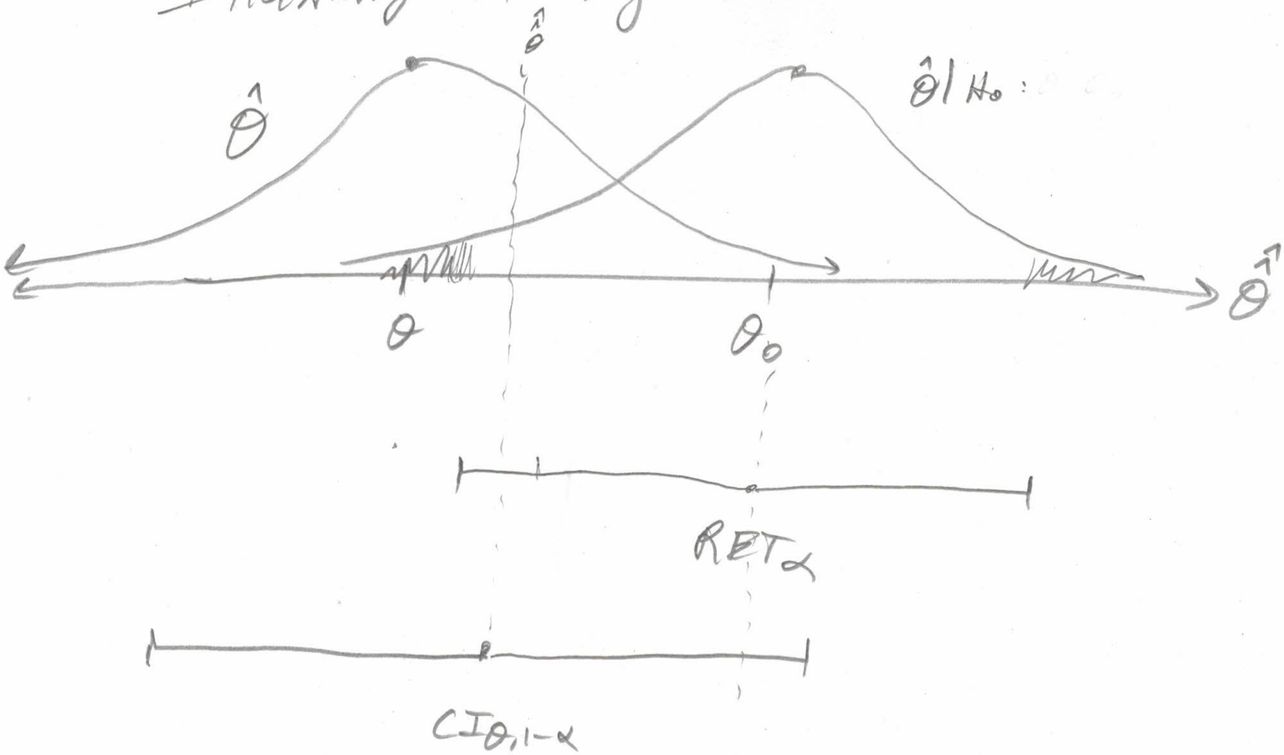
Just like there are 1-sided (left or right) hypothesis tests, there are also 1-sided CI's:

$$CI_{L,\theta,1-\alpha} = [\hat{\theta}_L, \infty) \quad \text{and} \quad CI_{R,\theta,1-\alpha} = (-\infty, \hat{\theta}_R]$$

They're not difficult ... but we will skip them

3 interpretations of CI's:
(1) Before you run the data experiment, the probability of θ being inside the CI is 1 - α
(2) If you run many experiments, the proportion of CI's that contain θ is approximately 1 - α
(3) The CI contains θ with probability zero or one.
All three are unsatisfactory!

Illustrating all 3 goals:

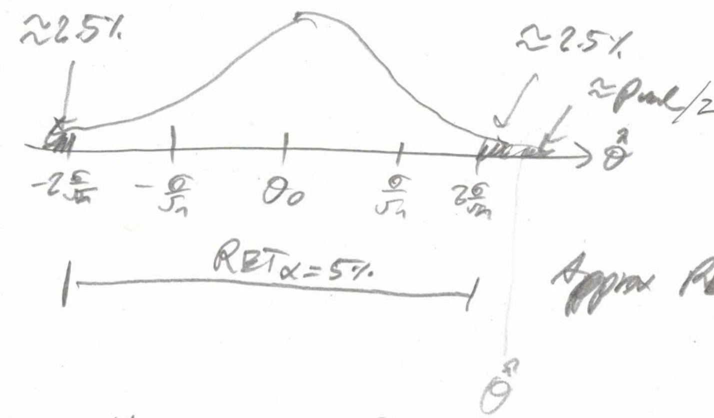


Consider DGP: $X_1, \ldots, X_n \overset{iid}{\sim}$ f or p where mean θ is unknown but variance σ^2 is known

Consider $\underbrace{\hat{\theta}}_{\text{"Asymptotically Normal estimator"}} = \bar{X} \overset{\cdot}{\underset{\text{by CLT}}{\sim}} N\left(\theta, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

E.g. Bern(θ). Under H_0, σ^2 known! But this test is more powerful

If testing $H_a: \theta \neq \theta_0 \Leftrightarrow H_0: \theta = \theta_0 \Rightarrow \hat{\theta} | H_0 \overset{\cdot}{\sim} N\left(\theta_0, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$



Approx RET region

$\Rightarrow$ Approx Test, How approximate? depends on how fast CLT converges for the DGP's PDF/PMF

What about CI?

$$1 - \alpha \approx P(\hat{\theta} \in RET \mid \theta = \theta_0)$$
$$= P\left(\theta_0 \in \left[\hat{\theta} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \hat{\theta} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]\right)$$

$\Rightarrow \hat{CI}_{\theta, 1-\alpha} \approx \left[\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$, an approx CI!

A true CI is impossible since PDF/PMF of DGP is unknown!!

$\Rightarrow$ The more you know about your DGP, the more accurate your hypothesis tests' decisions and p-vals and power will be, and the more accurate the coverage of your CI's. Point estimation will be more accurate too. We will see an example of this later in the semester...

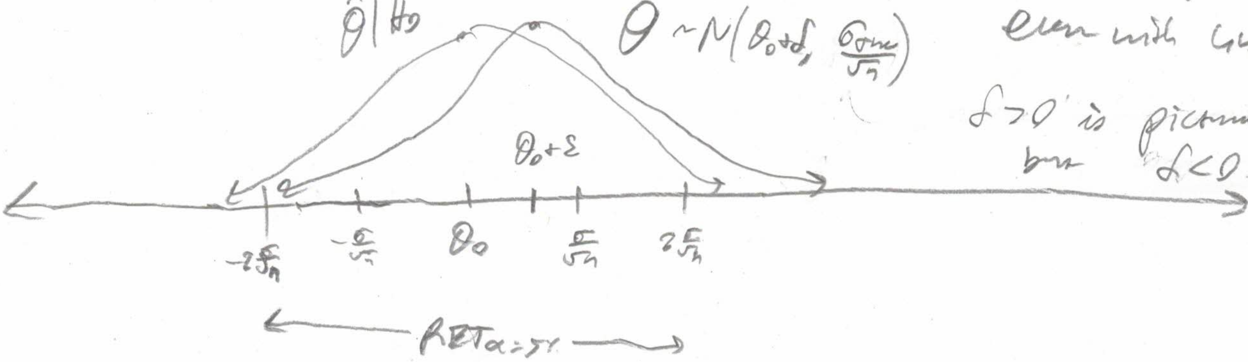
Let's review hypothesis testing and power

Consider $H_a: \theta \neq \theta_0$, $H_0: \theta = \theta_0$

We talked about how a failure to reject $H_0 \not\Leftrightarrow$ accept H_0 since the pt. θ_0 can be slightly perturbed to yield the same decision.

There is another implication of this...

Consider DGP: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ although the following will work for any DGP even with unknown variance



$\delta > 0$ is pictured here but $\delta < 0$ works too!

What if the true $\theta = \theta_0 + \delta$ where $|\delta| > 0$ but $|\delta| \approx 0$

With small $n \Rightarrow$ Only $\approx \frac{\alpha}{2}$ power to detect this difference $\Rightarrow$ near zero chance!

But what if $n \to \infty$?

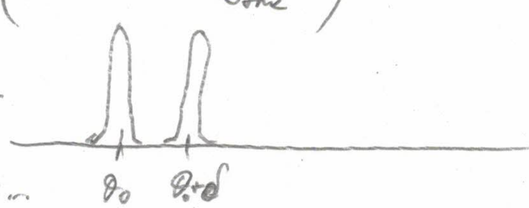
$$\text{Pow} = P\left(\hat{\theta} > \theta_0 + 2\frac{\sigma}{\sqrt{n}}\right) = P\left(\frac{\hat{\theta} - (\theta_0+\delta)}{\frac{\sigma_{true}}{\sqrt{n}}} > \frac{\theta_0 - (\theta_0+\delta) + 2\frac{\sigma}{\sqrt{n}}}{\frac{\sigma_{true}}{\sqrt{n}}}\right)$$

if $\delta > 0$

$$= P\left(Z > \frac{-\delta\sqrt{n} + 2\sigma}{\sigma_{true}}\right) \longrightarrow 1 \quad \boxed{\text{OR}} \quad = P\left(Z < \frac{\delta\sqrt{n} + 2\sigma}{\sigma_{true}}\right) \longrightarrow 1$$

if $\delta < 0$

For n large... this is the picture...



$\Rightarrow$ Since no null point hypothesis in the world is absolutely true, this means all H_0's will be rejected with sufficient sample size i.e. all estimates will become "statistically significant". (This holds true for right sided tests for $\delta > 0$ and left-sided tests for $\delta < 0$.)

Although estimates will be "statistically significant", they may not be "clinically significant" or "practically significant". The degree of significance is determined by the value of δ.

E.g. test if coin is unfair $H_a: \theta \neq 50\% \Leftrightarrow H_0: \theta = 50\%$

Since more metal on one side $\theta = 50\% + \delta$, $\delta = 0.001\%$

With enough flips, H_0 is rejected. Is coin unfair?? No...

Maybe let $\delta = 2\%$?. If $|\hat{\theta} - \theta| > \delta$ and H_0 rejected

$\Rightarrow$ conclude coin is unfair.

"δ" is sometimes called "margin of equivalence".

MID I $\uparrow$

MID II $\downarrow$

Recall the two type of errors for the hypothesis test:

		Decision	
		Retain H_0	Reject H_0
H_0	true	✓	Type I error
	false	Type II error	✓

$P(\text{Type I error}) = \alpha$ assuming continuous exact test

Let's say you were doing m different tests (a "family" of tests)

Lets say you reject H_0 in r of the m tests and thus retain $m-r$ of the tests. But for v of the r tests, you make type I errors. Let R, V represent the rv's that r, v are realized from. Let m_0 be the number of tests where H_0 is true.

If $m=1$ and $m_0=1$, $P(R=1) = P(V=1) = \alpha$ i.e. $P(\text{Type I err})$ which is controlled by you! What if we want control for $m>1$ tests? Can we use α for each test?

If $m = m_0$

$V = R \sim \text{Binom}(m, \alpha)$.

$\Rightarrow$ The chance of making at least one Type I err is

$P(V \geq 1) = 1 - P(V=0) = 1 - (1-\alpha)^m$

for $\alpha = 5\%$, $m = 30$, $P(V \geq 1) = 76\%$ Too High!!

for $\alpha > 0$, $m \to \infty$ $P(V \geq 1) \longrightarrow 1$

Let's talk about "family-wise error control". Let:

$FWER := P(V \geq 1)$ over all m tests

A family is "any logical collection of inferences for which it is meaningful to take into account some combined measure of error" or "a collection which you wish to prevent data dredging" (342 concept) or to "ensure a correct decision on a collection of tests".

If you can prove

$FWER \leq FWER_0 \overset{e.g.}{=} 5\%$ for any $m_0 \leq m$

this is called "strong control of family-wise error rate".

If you can prove

$FWER \leq FWER_0$ for $m_0 = m$

this is called "weak control of family-wise error rate".