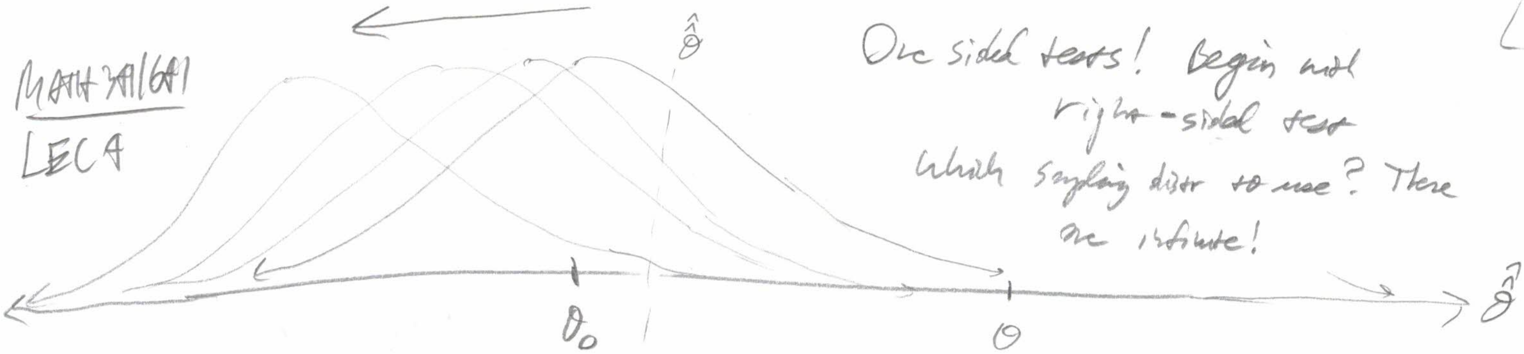
MATH 341/641

LEC 4



One sided tests! Begin with right-sided test

Which sampling distr to use? There are infinite!

$$H_0 : \theta \leq \theta_0 \Rightarrow \Theta_0 = (-\infty, \theta_0]$$

$$H_a : \theta > \theta_0 \Rightarrow \Theta_a = (\theta_0, \infty)$$

Reject H_0 if $p_{val} < \alpha \Rightarrow$ there is significant evidence against H_0.

Depending on the Null Distr used, we can always reject H_0!

There is something wrong somewhere. C&B p397-398.

For a one-sided test, we define p-val a bit more exactly:

$$p_{val} = \max_{\theta \in \Theta_0} \left\{ \underset{\alpha}{\text{argmax}} \left\{ \alpha : \hat{\hat{\theta}} \in RET_{\alpha, \theta_0} \right\} \right\}$$

the retainment region changes with the value of θ_0

We want to make sure to be as conservative as possible

Effectively then...

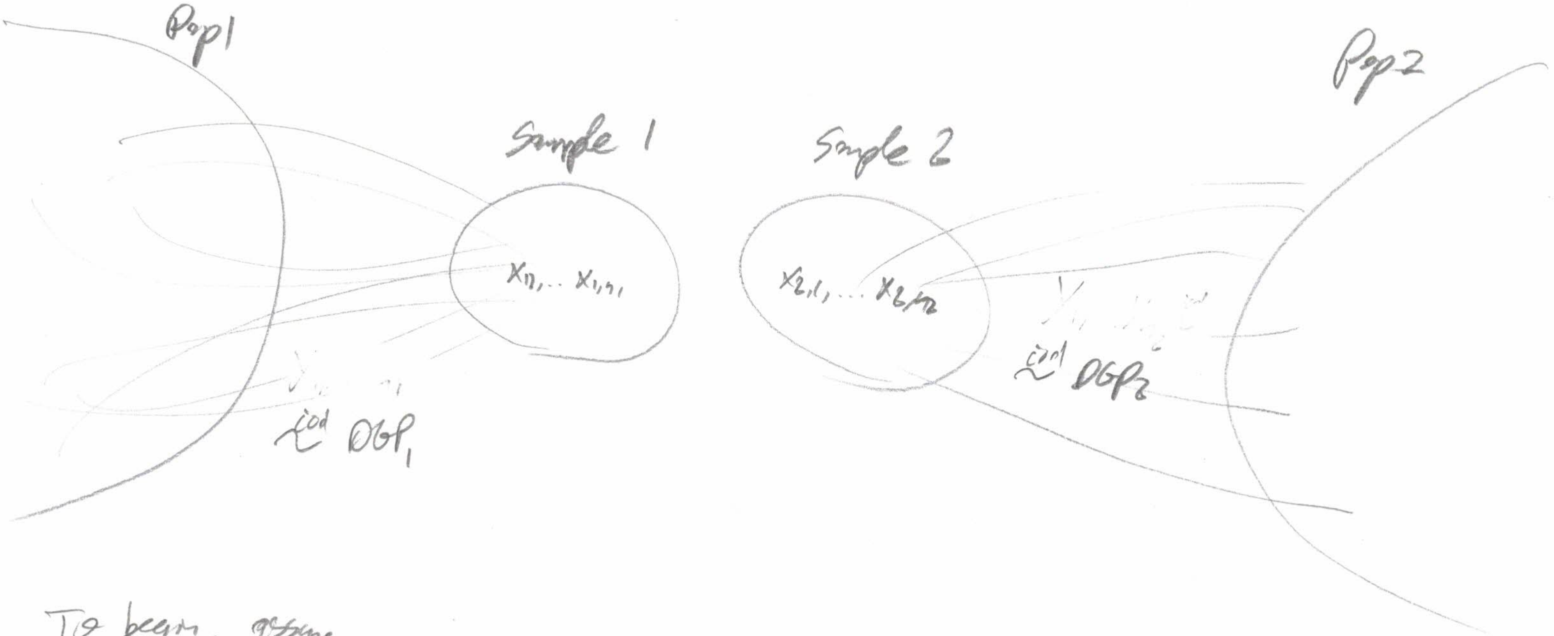
$$H_0 : \theta = \theta_0$$

$$H_a : \theta > \theta_0$$

Back to Joe's question... why Retain H_0 and Accept H_a?

This is how scientific theories work. Every theory is an exact mathematical statement about reality. E.g. $F = G\frac{m_1 m_2}{r^2}$ which has one way to be right and infinite ways to be wrong. We never "accept" our theories (i.e. calling them absolutely true), we only retain them as "the best we got right now". If evidence surfaces that they're wrong, we reject them. This is akin to accepting that it's wrong. E.g. $H_0: \theta = \theta_0$, $H_a: \theta \neq \theta_0$. Accepting H_a is not the same as accepting a new, exact theory.

New setting: two infinitely sized populations (or DGP's)



To begin, assume

$$X_{1,1}, \dots, X_{1,n_1} \overset{iid}{\sim} N(\theta_1, \sigma_1^2) \quad \text{ind. of} \quad X_{2,1}, \dots, X_{2,n_2} \overset{iid}{\sim} N(\theta_2, \sigma_2^2)$$

Just like before, we assume σ_1^2 and σ_2^2 are known.

There are three common tests:

$H_a: \theta_1 \neq \theta_2 \iff H_0: \theta_1 = \theta_2 \quad (H_a: \theta_1 - \theta_2 \neq 0 \iff H_0: \theta_1 - \theta_2 = 0)$ Two-sided, two sample test

$H_a: \theta_1 < \theta_2 \iff H_0: \theta_1 \geq \theta_2 \quad (H_a: \theta_1 - \theta_2 < 0 \iff H_0: \theta_1 - \theta_2 \geq 0)$ Left-sided, two-sample test

$H_a: \theta_1 > \theta_2 \iff H_0: \theta_1 \leq \theta_2 \quad (H_a: \theta_1 - \theta_2 > 0 \iff H_0: \theta_1 - \theta_2 \leq 0)$ Right-sided, ~~two-sample test~~

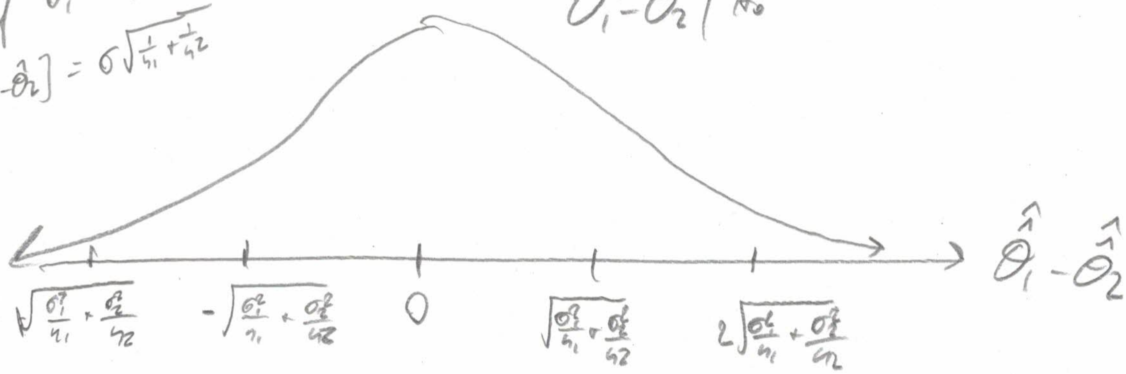
How to run the test? Find a test statistic that captures departure from H_0. $\hat{\theta}_1 - \hat{\theta}_2 = \bar{X}_1 - \bar{X}_2$ or $\frac{\hat{\theta}_1}{\hat{\theta}_2} = \frac{\bar{X}_1}{\bar{X}_2}$. Now we need its sampling distr under the null. For all 1-sided and 2-sided, the sampling distr uses $\theta_1 - \theta_2 = 0$

$$\hat{\theta}_1 - \hat{\theta}_2 \mid H_0 \sim N\left(0, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}^{\,2}\right)$$

amazing: we don't need to know anything about the the values θ_1, θ_2. Very convenient!!

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

$SE[\hat{\theta}_1 - \hat{\theta}_2] = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$



RET$_{5\%}$

Let's test if mean male height $\neq$ mean female height in 300-level STEM classes in QC!

Assume: $\sigma^2_{MALE} = 4''^2$, $\sigma^2_{FEMALE} = 3.5''^2$

$\hat{\theta}_1 = 68.85''$ $n_1 = 13$

$\hat{\theta}_2 = 63.33''$ $n_2 = 6$

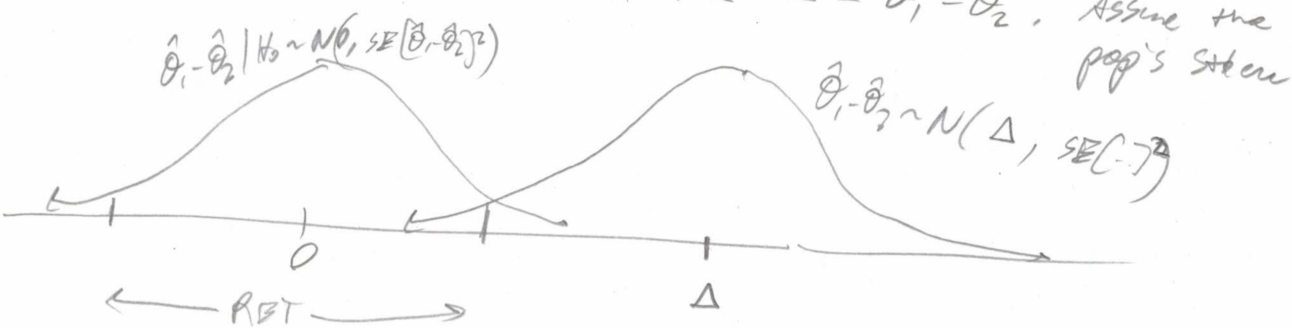
Let pop 1 be male, pop 2 be female

$\Rightarrow H_a: \theta_1 - \theta_2 \neq 0$

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

$SE[\hat{\theta}_1 - \hat{\theta}_2] = \sqrt{\frac{4^2}{13} + \frac{3.5^2}{6}} = 1.81$

How does power work? Assume a true $\Delta = \theta_1 - \theta_2$. Assume the pop's std err



We will return to Hypothesis tests later!

We return now to estimator theory...

Let $X_1, \ldots, X_n \overset{iid}{\sim} DGP(\theta_1, \theta_2, \ldots, \theta_k)$ $k = \#$ of parameters

e.g. $k=2$ if $DGP = N(\underbrace{\theta}_{\text{param 1}}, \underbrace{\sigma^2}_{\text{param 2}})$

estimator for any parameter is a function of the data

$$\hat{\theta}_j = w(X_1, \ldots, X_n)$$

How do we pick the function w?

Recall $X_1, \ldots, X_n \overset{iid}{\sim} Bern(\theta)$

$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ was our default estimator.

We now know, due to LLN $\hat{\theta} \xrightarrow{d} \theta$

So as n becomes large, this $\hat{\theta}$ gets better and better.

Can we extend this idea? ~~Yes~~.

Define $\mu_k := E[X^k]$, the k^{th} moment of r.v. X, $k \in \mathbb{N}$.

Not all moments exist for all r.v.'s. Some r.v.'s don't even have μ_1!

Define the sample moment to be

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \Rightarrow \bar{X} = \hat{\mu}_1$$

If there are k parameters, there exist a system of k equations:

$$\begin{aligned} \mu_1 &= \alpha_1(\theta_1, \dots, \theta_k) \\ \mu_2 &= \alpha_2(\theta_1, \dots, \theta_k) \\ &\vdots \\ \mu_k &= \alpha_k(\theta_1, \dots, \theta_k) \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \theta_1 &= \beta_1(\mu_1, \dots, \mu_k) \\ \theta_2 &= \beta_2(\mu_1, \dots, \mu_k) \\ &\vdots \\ \theta_k &= \beta_k(\mu_1, \dots, \mu_k) \end{aligned}$$

The Method of Moments (MM) estimator is then just finding the β_jth function and estimating its inputs with sample moments:

$$\hat{\theta}_j^{MM} = \beta_j(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_k)$$

Goes back to Karl Pearson in 1890's.

MM usually works but rarely produces "great" estimators.

If $k=1$, $\theta = E[X] = \mu_1 \Rightarrow \theta = \beta(\mu_1) = \mu_1$

$$\Rightarrow \hat{\theta}^{MM} = \hat{\mu}_1 = \bar{X}$$

If $k=2$, $\theta_1 = E[X] = \mu_1$, $\theta_2 = Var[X] = E[X^2] - E[X]^2 = \mu_2 - \mu_1^2$

$$\Rightarrow \hat{\theta}_1 = \hat{\mu}_1 = \bar{X}$$

$$\hat{\theta}_2 = \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$$

Note: $\sum(x_i-\bar{x})^2 = \sum(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$

$\sum x_i = n\bar{x}$

$$\begin{aligned} &= \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum x_i^2 - n\bar{x}^2 \end{aligned}$$

$$\Rightarrow \hat{\sigma}^2 := \frac{1}{n}\sum(x_i-\bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2 = \hat{\mu}_2 - \hat{\mu}_1^2$$

Param Space

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bin}(\overset{n}{\theta_1}, \overset{p}{\theta_2})$ Both unknown!! , $\theta_1 \in \mathbb{N}$, $\theta_2 \in (0,1)$

Does this model correspond to anything remotely real?

Imagine X_i are the # of crimes reported each night in a city district.
Let θ_1 be the total # of crimes each night (unknown),
Let θ_2 be the prob. a crime gets called in.
Both params are interesting to estimate!
Use MM to get estimates:

$$\mu_1 = E[X] = \theta_1 \theta_2 \qquad = \alpha_1(\theta_1, \theta_2)$$

$$\mu_2 = E[X^2] = \text{Var}[X] + E[X]^2 = \theta_1\theta_2(1-\theta_2) + \theta_1^2\theta_2^2 = \alpha_2(\theta_1, \theta_2)$$

$$= \theta_1\theta_2 - \theta_1\theta_2^2 + \theta_1^2\theta_2^2$$

$$\Rightarrow \theta_1 = \frac{\mu_1}{\theta_2}$$

Now we need β_1, β_2 functions!

$$\Rightarrow \mu_2 = \frac{\mu_1}{\cancel{\theta_2}}\cancel{\theta_2} - \frac{\mu_1}{\cancel{\theta_2}}\theta_2^{\cancel{2}} + \frac{\mu_1^2}{\cancel{\theta_2^2}}\cancel{\theta_2^2} = \mu_1 - \mu_1\theta_2 + \mu_1^2$$

$$\Rightarrow \mu_2 - \mu_1 - \mu_1^2 = -\mu_1\theta_2 \Rightarrow \theta_2 = \frac{\mu_1^2 + \mu_1 - \mu_2}{\mu_1} = \frac{\mu_1 - (\mu_2 - \mu_1^2)}{\mu_1} = \beta_2(\mu_1, \mu_2)$$

$$\Rightarrow \mu_1 = \theta_1 \frac{\mu_1 - (\mu_2 - \mu_1^2)}{\mu_1} \Rightarrow \theta_1 = \frac{\mu_1^2}{\mu_1 - (\mu_2 - \mu_1^2)} = \beta_1(\mu_1, \mu_2)$$

$$\Rightarrow \hat{\theta}_1^{mm} = \frac{\hat{\mu}_1 - (\hat{\mu}_2 - \hat{\mu}_1^2)}{\hat{\mu}_1} = \frac{\bar{X} - \hat{\sigma}^2}{\bar{X}}$$

$$\hat{\theta}_2^{mm} = \frac{\hat{\mu}_1^2}{\hat{\mu}_1 - (\hat{\mu}_2 - \hat{\mu}_1^2)} = \frac{\bar{X}}{\bar{X} - \hat{\sigma}^2}$$

Let data be $\vec{x} = \langle 3, 7, 5, 5, 6 \rangle \Rightarrow \bar{x} = 5.2,\ \hat{\sigma}^2 = 2.64$

$\hat{\theta}_1^{mm} = 10.56$, $\hat{\theta}_2^{mm} = 0.49$ makes sense

But what if... $\vec{x} = \langle 3, 7, 5, 11, 6 \rangle \Rightarrow \bar{x} = 6.4,\ \hat{\sigma}^2 = 10.56$

$\Rightarrow \hat{\theta}_1^{mm} = -9.8,\ \hat{\theta}_2^{mm} = -0.65$ Both estimates not in param space!

This is what I mean by "rarely very good"!

$X_1, \dots, X_n \overset{iid}{\sim} U(0, \theta)$. Find $\hat{\theta}^{mm}$

$$\mu_1 = E[X] = \frac{\theta - 0}{2} = \frac{\theta}{2} \Rightarrow \theta = 2\mu_1 \Rightarrow \hat{\theta}^{mm} = 2\hat{\mu} = 2\bar{X}$$

let $x = \langle 1, 2, 3, 10 \rangle \Rightarrow \bar{x} = 4 \Rightarrow \hat{\theta}^{mm} = 2 \cdot 4 = 8$ Stupid! We already observed an $x_i > \hat{\theta}$!

Is there a better way to find estimators??