9/21/20

DGP: $x_1, \ldots, x_n \sim$ iid $N(\theta, \sigma^2)$ indep of
$x_{21}, \ldots x_{1 n_2} \overset{iid}{\sim} N(\theta_2, \sigma_2^2)$

Now we don't assume we know σ_1^2 and σ_2^2
and we use the sample variances to estimate them

$$S_1^2 := \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2, \quad S_2^2 := \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2$$

under $H_0: \theta_1 - \theta_2 = 0$

$$\Rightarrow \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim T_{df_1}? \quad \text{but no} \ldots$$

This was pointed by Behrens (1929) and Fisher (1935).
Because they discovered this distribution, it's called the Behrens-Fisher distribution (and this is callded the Behrens-Fishner problem).

$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \text{Behrensfisher}(\ldots)$$

They tried to work out its PDF but they couldn't and at some point they gave up and conjectured that it was impossible, zen 1966, it was proven that it has a closed form solution, And, it was published in 2018

In 1946/7 Welch and Satterthwaite found a T approximation which is very good and still used today (p 314 (C&B));

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}$$

using this T_{df} is known as welch's t-test or "unequal variances t test".

lower $\Rightarrow$ thicker tails

$n_1 = 10$, $\bar{x}_1 = 70.5$, $S_1 = 2.07$ } male

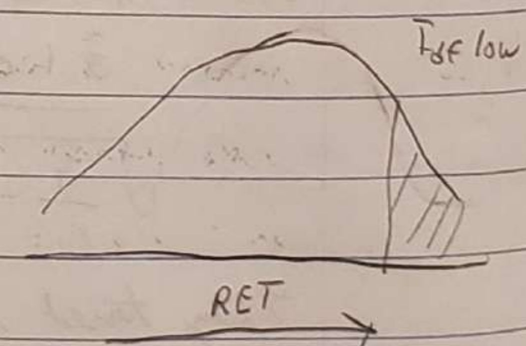
$n_2 = 6$, $\bar{x}_2 = 62.3$, $S_2 = 2.25$ } female

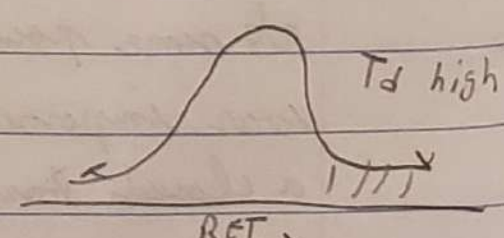
$$d.f = \frac{1.27^2}{\frac{2.07^4}{10^2(9)} + \frac{2.25^4}{6^2(5)}} = \frac{1.62}{0.163} = 9.94$$

$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \dot{\sim} T_{df}$$

$$SE = \sqrt{\frac{2.07^2}{10} + \frac{2.25^2}{6}} = \sqrt{1.27} = 1.13$$

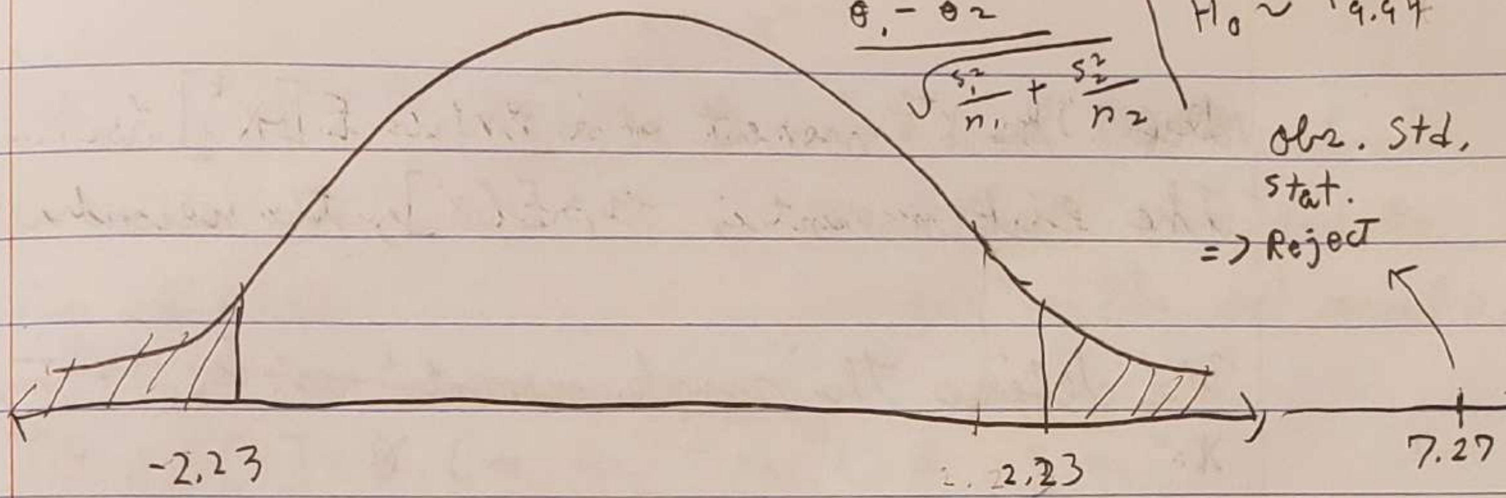
Harder to reject ↓





$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \Big| H_0 \sim T_{9.94}$$

Obs. Std. stat. ⟹ Reject



$F_{T_{9.94}}(-2.23) = 2.5$

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2.07^2}{10} + \frac{2.25^2}{6}}} = \frac{8.2}{1.13} = 7.27$$

Midterm I ↑

Midterm II ↓

$x_1, \ldots, x_n \overset{iid}{\sim} DGD(\theta_1, \theta_2, \ldots, \theta_k)$ K is # parameters

we've previously seen estimators $\hat{\theta} = W(x_1 \ldots x_n)$ eg

$\hat{\theta} = \bar{x}$, $\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ ...

How did we get this function w? Where did it come from? There are many strategies to create estimators. The first we'll study is called "method of moments" (MM and it was used by Karl Pearson in the late 1890's

We know the DGP and we know which θs we want to estimate. We now need an algorithm to generate W.

Def: The k^{th} moment of a r.v. is $E[X^k]$

The first moment is $\mu_1 = E[X^1]$, the second is $\mu_2 := E[X^2]$ etc

We define the "sample moments" as: $\hat{\mu}_k := \frac{1}{n} \sum_{i=1}^{n} X_i^k$

The first sample moment is the "sample average" (Sample mean)

$$\hat{\mu}_1 = \frac{1}{n} \sum X_i = \overline{X}$$

Pearson's idea is to "match moments to parameters" If...

$$\begin{aligned} \mu_1 &= \alpha_1(\theta_1, \dots, \theta_k), & \theta_1 &= \gamma_1(\mu_1, \dots, \mu_k) \\ \mu_2 &= \alpha_2(\theta_1, \dots, \theta_k) \quad \text{and} & \theta_2 &= \gamma_2(\mu_1, \dots, \mu_k) \\ \vdots & & \vdots & \\ \mu_k &= \alpha_k(\theta_1, \dots, \theta_k) & \theta_k &= \gamma_k(\mu_1, \dots, \mu_k) \end{aligned}$$

a system of equation

$$\Rightarrow \hat{\theta}_j^{mm} = \gamma_j(\hat{\mu}_1, \dots \hat{\mu}_k)$$

mm pretty much always gives you an estimator, But it is rarely a "great" estimator and sometimes produces totally wrong answers

$X_1, \dots, X_n \overset{iid}{\sim} N(\underset{=\mu}{\theta_1}, \underset{=\sigma^2}{\theta_2})$ We want the MM estimator for both θ_1 (mean) and θ_2 (varience) in the iid normal DGP

True for all DGP's

$$\boxed{\begin{aligned} \theta_1 &= E[X] = g_1(\mu_1, \mu_2) = \mu_1 \\ Var[X] = \theta_2 &= g_2(\mu_1, \mu_2) = \mu_2 - \mu_1^2 \end{aligned}}$$

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

$$\Rightarrow \hat{\theta}_2^{mm} = \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{1}{n} \sum X_i^2 - \bar{X}^2$$

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

$$= \frac{1}{n} \sum X_i^2 - \bar{X}^2 = \hat{\sigma}^2$$

$X_1, \dots X_n \overset{iid}{\sim} Bin(\underset{=n}{\theta_1}, \underset{=p}{\theta_2})$ both θ_1, θ_2 unknown

We want to estimate both θ_1 (which is commonly denoted n) and θ_2 (which is commonly denoted p). Ecologists love this estimation problem because its part of the "capture-recapture" problem to estimate population size of wildlife

Each data point is the result of catching a certain number of fish in a time interval (eg 1 hr of fishing). Once you catch a fish you re-bait and re-cast. Every time a fish encounters the hook its a $Bern(\theta_2)$ that it bites and you catch it

θ_2 is the propensity to bite and θ_1 is the number of individual fish-hook encounters in the time period (eg 1 hr)

Let's delvop MM estimators for both θ_1 and θ_2

$E[X]$ solve for gamma 1, gamma 2

$\mu_1 = \alpha_1(\theta_1, \theta_2) = \theta_1\theta_2 \Rightarrow \theta_1 = \frac{\mu_1}{\theta_2}$

$\mu_2 = Var[X] + \mu_1^2 = \theta_1\theta_2(1-\theta_2) + \theta_1^2\theta_2^2 = \alpha_2(\theta_1, \theta_2)$

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

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

$= \mu_1 - \mu_1\theta_2 + \mu_1^2$

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

Plug in

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

$$\hat{\theta}_1^{MM} = \frac{\hat{\mu}_1^2}{\hat{\mu}_1 - (\hat{\mu}_2 - \hat{\mu}_1^2)}, \quad \hat{\theta}_2^{mm} = \frac{\hat{\mu}_1 - (\hat{\mu}_2 - \hat{\mu}_1^2)}{\hat{\mu}_1}$$

$$\hat{\theta}_1^{mm} = \frac{\bar{X}^2}{\bar{X} - \hat{\sigma}^2}, \quad \hat{\theta}_2^{mm} \frac{\bar{X} - \hat{\sigma}^2}{\bar{X}}$$

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

$$\hat{\theta}_1^{mm} = \frac{5.2^2}{5.2-2.64} = 10.56,\quad \hat{\theta}_2^{mm} = \frac{5.2-2.64}{5.2} = 0.49$$

different lake

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

$$\hat{\theta}_1^{mm} = \frac{6.4^2}{6.4-10.56} = -9.8,\quad \hat{\theta}_2^{mm} = \frac{6.4-10.56}{6.4} = -0.65$$

Obviously, n can't be negative and p must be a probability so these estimates are nonsensical, MM estimators are sometimes really bad... but they make for a nice place to start