Lecture 7 9/21/2020

DGP: $X_{11}, \ldots, X_{1n_1} \overset{i.i.d.}{\sim} N(\theta_1, \sigma_1^2)$ independent of

$X_{21}, \ldots, X_{2n_2} \overset{i.i.d.}{\sim} N(\theta_2, \sigma_2^2)$

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 $H_0: \theta_1 - \theta_2 = 0$

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

degree of freedom

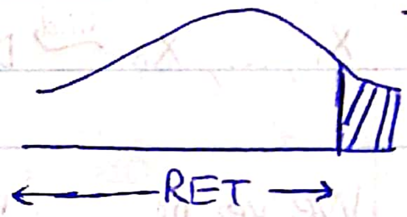
This was pointed by Behrens (1929) and Fisher (1935). Because they discovered this distribution, it's called the Behrens - Fisher distribution (and this is called the Behrens - Fisher problem).

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

In 1946 / 1947 Welch and Satterthwaite found a T approximation which is very good and still used today:

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

Using this T_{df} is known as Welch's T-test or "unequal variance T test".



T_{df} low

(male) $n_1 = 10$, $\bar{X}_1 = 70.5$, $S_1 = 2.07$

(female) $n_2 = 6$, $\bar{X}_2 = 62.3$, $S_2 = 2.25$

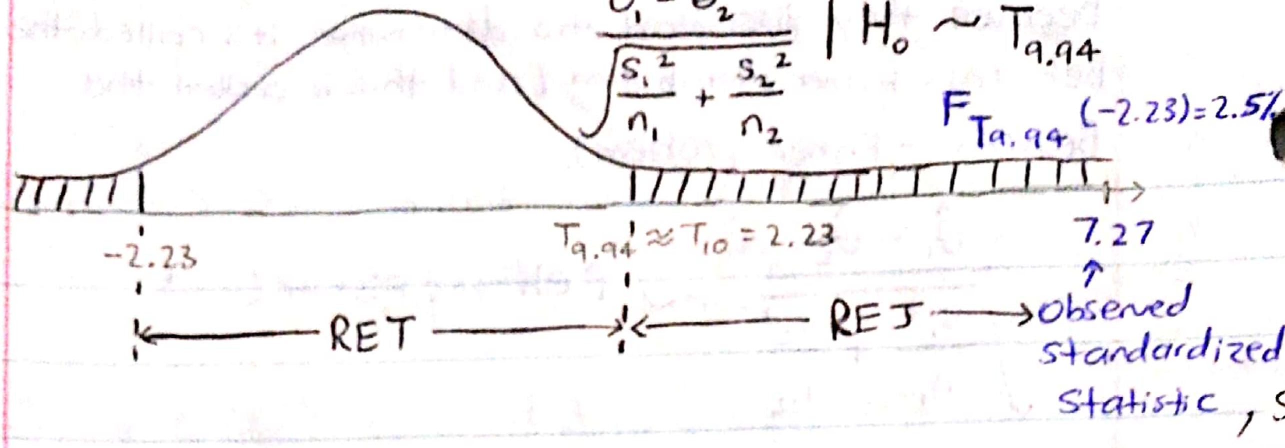
T_{df} high

$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim T_{df} \rightarrow df = \frac{(1.27)^2}{\frac{2.07^4}{10^2(9)} + \frac{2.25^4}{6^2(5)}} = \frac{1.62}{0.163} = 9.94$$

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

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

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



, so we ↓ ⇒ Reject H_0

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{2.07^2}{10} + \frac{2.25^2}{6}}} = \frac{70.5 - 62.3}{\sqrt{1.27}} = \frac{8.2}{1.13} = 7.27$$

Midterm 1 ↑

Midterm 2 ↓

$X_1, \ldots, X_n \overset{iid}{\sim} DGP(\theta_1, \theta_2, \ldots, \theta_K)$ where K is # parameters

We've previously seen estimators $\hat{\theta} = w(X_1, \ldots, X_n)$

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

How did we get this function w? Where did it come from? There are many strategies to create estimators.

We know the DGP and we know which θ's we want to estimate. We now need an algorithm to generate w. The first we'll study is called "Method of Moments" (MM) and it was used by Karl Pearson in the late 1890's.

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),

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

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

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

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

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

$X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\underset{\mu}{\theta_1}, \underset{\sigma^2}{\theta_2})$ We want the MM estimators for both $\theta_1(\mu)$ and $\theta_2(\sigma^2)$ in the i.i.d. normal DGP.

$$\theta_1 = E[X] = \gamma_1(\mu_1, \mu_2) = \mu_1 \Rightarrow \hat{\theta}_1^{MM} = \hat{\mu}_1 = \bar{X}$$

$$Var[X] = \theta_2 = \gamma_2(\mu_1, \mu_2) = \mu_2 - \mu_1^2 \Rightarrow \hat{\theta}_2^{MM} = \hat{\mu}_2 - \hat{\mu}_1$$

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

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

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

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Bin}(\overset{n}{\theta_1}, \overset{p}{\theta_2})$ 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 it's part of the "capture-recapture" problem to estimate population size of wildlife.

Each data point is the result of catching a certain # of fish in a time interval (e.g. 1hr of fishing). Once you catch a fish you re-bait and re-cast. Every time a fish encounters the hook it's a Bern(θ_2) that it bites and you catch it.

θ_2 is the propensity to bite and θ_1 is the # of individual fish-hook encounters in the time period (e.g. 1 hr).

Let's develop MM estimators for both θ_1 and θ_2.

Solve for γ_1 and γ_2

$$E[X] = \mu_1 = \alpha_1(\theta_1, \theta_2) = \theta_1\theta_2 \Rightarrow \theta_1 = \frac{\mu_1}{\theta_2}$$

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

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

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

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

$$\mu_2 = \mu_1 - \mu_1\theta_2 + \mu_1^2 \quad // \text{ solve for } \theta_2$$

$$\mu_2 - \mu_1^2 - \mu_1 = -\mu_1\theta_2$$

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

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

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

$$= \frac{\bar{x}^2}{\bar{x} - \hat{\sigma}^2} \qquad\qquad = \frac{\bar{x} - \hat{\sigma}^2}{\bar{x}}$$

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

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

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

$$\hat{\theta}_1^{MM} = \frac{6.4^2}{6.4 - 10.56} = -9.8, \quad \hat{\theta}_2^{MM} = \frac{6.4 - 10.56}{6.4} = -0.56$$

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...