Lecture - 07.

09/21/2020

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

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

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

Under $H_0: \theta_1 - \theta_2 = 0$.

$$\Rightarrow \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim T_{df} \ \overset{?}{} \quad \text{But now.}$$

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

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

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

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

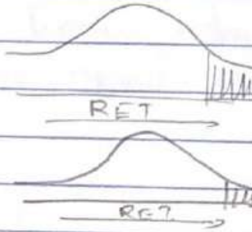
using this T_df is known as Welch's t-test or "unequal variances t test."

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

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

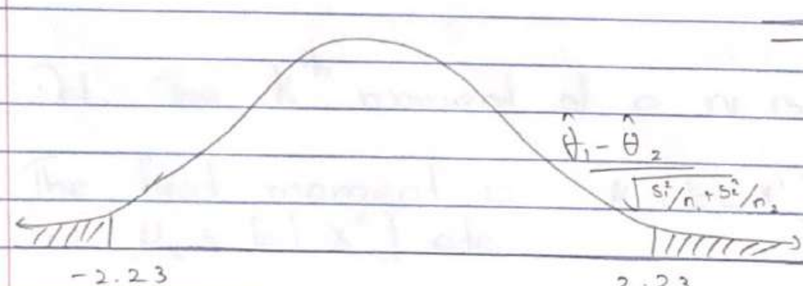
$$\underbrace{\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}}_{} \sim T_{df} \nearrow df = \frac{1.27^2}{\dfrac{2.07^4}{10^2(9)} + \dfrac{2.25}{6^2(5)}} = \frac{1.62}{0.163} = 9.94$$

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



thicker tails T_{df} low

thinner tails T_{df} high



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

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

obs. std statistic

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

$\Longrightarrow$ Reject!

Midterm II

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

We've previously seen estimators. $\hat{\theta} = \omega(X_1, \dots, X_n)$ e.g.

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

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

(We know the DGP and we know which $\theta(s)$ we want to estimate We now need an algorithm to generate ω.)

Def. The k^{th} moment of a rv 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:

$$\hat{\mu}_k \equiv \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

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

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

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

$$\mu_1 = \alpha_1(\theta_1, \ldots, \theta_k), \qquad \theta_1 = \gamma_1(\mu_1, \ldots, \mu_k)$$

$$\mu_2 = \alpha_2(\theta_1, \ldots, \theta_k), \text{ and } \theta_2 = \gamma_2(\mu_1, \ldots, \mu_k)$$

$$\vdots \qquad\qquad \vdots$$

$$\mu_k = \alpha_k(\theta_1, \ldots, \theta_k) \qquad \theta_k = \gamma_k(\mu_1, \ldots, \mu_k)$$

a system of equations.

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

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

$$X_1, \ldots, X_n \overset{iid}{\sim} N(\overset{\mu}{\overset{\parallel}{\theta_1}}, \overset{\sigma^2}{\overset{\parallel}{\theta_2}})$$ We want the MM estimators for both θ_1 (mean) & θ_2 (Variance) in the iid normal DGP.

$$\boxed{\theta_1 = E[X] = \gamma_1(\mu_1, \mu_2) = \mu_1}$$ true for all DGP's

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

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

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

$$= \hat{\sigma}^2.$$

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

$$= \frac{1}{n}\sum x_i^2 - \frac{1}{n} 2\bar{x}(n\bar{x}) + \frac{1}{n} n \bar{x}^2$$

$$= \frac{1}{n}\sum x_i^2 - \bar{x}^2.$$

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

We want to estimate both $\theta_1 (n)$ 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 number of fish in a time interval (e.g. 1 hr of fishing). Once you catch a fish you re-bait and re-cast. Every time a fish encounters the hook it's a $Bern(\theta_2)$ 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. 1hr).

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

Solve for θ_1, θ_2

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

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

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

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

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

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

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

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

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

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

$$\hat{\theta}_1^{MM} = \frac{\bar{X}^2}{\bar{X} - \hat{\sigma}^2} \qquad \hat{\theta}_2^{MM} = \frac{\bar{X} - \hat{\sigma}^2}{\bar{X}}$$

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

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

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

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

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