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

Now we don't assume we know sigsq_1 and sigsq_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 $H_0: \theta_1 - \theta_2 = 0$

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

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).

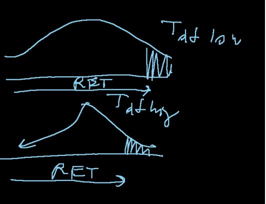
$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim \text{BehrensFisher}(\ldots)$$

They tried to work out its PDF but they couldn't and at some point they gave up and conjectured that it was impossible. In 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".

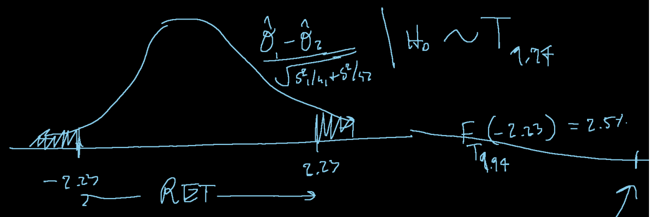


$n_1 = 10, \; \bar{x}_1 = 70.5, \; s_1 = 2.07$ } Male

$n_2 = 6, \; \bar{x}_2 = 62.3, \; s_2 = 2.25$ } Female

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

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



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

Obs. Std. statistic

$\Rightarrow$ Reject!

Midterm I ↑

Midterm II ↓

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

We've previously seen estimators thetahat = w(X_1, ..., X_n) e.g.

$\hat{\theta} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \bar{X})^2, \ldots$

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 theta(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 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: $\hat{\mu}_K := \frac{1}{n} \sum_{i=1}^{n} X_i^K$

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...,

$$\begin{aligned} \mu_1 &= \alpha_1(\theta_1, \ldots, \theta_K), \\ \mu_2 &= \alpha_2(\theta_1, \ldots, \theta_K), \\ &\vdots \\ \mu_K &= \alpha_K(\theta_1, \ldots, \theta_K) \end{aligned} \quad \text{and} \quad \begin{aligned} \theta_1 &= \gamma_1(\mu_1, \ldots, \mu_K) \\ \theta_2 &= \gamma_2(\mu_1, \ldots, \mu_K) \\ &\vdots \\ \theta_K &= \gamma_K(\mu_1, \ldots, \mu_K) \end{aligned}$$

a system of equations

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

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

$X_1, \ldots, X_n \overset{iid}{\sim} N(\overset{=\mu}{\theta_1}, \overset{=\sigma^2}{\theta_2})$

We want the MM estimators for both theta_1 (mean) and theta_2 (variance) in the iid normal DGP

true for all DGP's

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

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

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

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

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

We want to estimate both theta_1 (which is commonly denoted n) and theta_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. 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(theta_2) that it bites and you catch it.

Theta_2 is the propensity to bite and theta_1 is the number of individual fish-hook encounters in the time period (e.g. 1hr).

Let's develop MM estimators for both theta_1 and theta_2.

solve for gamma_1, gamma_2

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

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

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

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

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

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

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

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

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

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

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

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

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...