Lecture 12 10/19/20

In Lec. 10 $\frac{\hat{\theta}-\theta}{SE[\hat{\theta}]} \xrightarrow{d} N(0,1) \Rightarrow \frac{\hat{\theta}-\theta}{\widehat{SE}[\hat{\theta}]} \xrightarrow{d} N(0,1)$

We can use this now in our situation:

$$\frac{\hat{\theta}_1-\hat{\theta}_2}{SE[\hat{\theta}_1-\hat{\theta}_2]} \xrightarrow{d} N(0,1) \Rightarrow \frac{\hat{\theta}_1-\hat{\theta}_2}{\widehat{SE}[\hat{\theta}_1-\hat{\theta}_2]} \xrightarrow{d} N(0,1)$$

$$SE[\hat{\theta}_1-\hat{\theta}_2] = \sqrt{\theta_{shared}(1-\theta_{shared})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$\widehat{SE}[\hat{\theta}_1-\hat{\theta}_2] = \sqrt{\hat{\theta}_{shared}(1-\hat{\theta}_{shared})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$ if $\hat{\theta}_{shared}$ is consistent

$$\hat{\theta}_{shared} = \text{avg. of both samples} = \frac{\sum X_{1i} + \sum X_{2i}}{n_1+n_2}$$

$$\Rightarrow \frac{\hat{\theta}_1-\hat{\theta}_2}{\sqrt{\frac{\sum X_{1i}+\sum X_{2i}}{n_1+n_2}\left(1-\frac{\sum X_{1i}+\sum X_{2i}}{n_1+n_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \stackrel{\cdot}{\sim} N(0,1)$$

e.g.: $H_a: \theta_1-\theta_2 \neq 0$, $H_0: \theta_1-\theta_2 = 0$, $\alpha = 5\%$

control $n_1 = 81$, $\sum X_{1i} = 27 \Rightarrow \hat{\theta}_1 = \frac{27}{81} = 0.333$

experiment $n_2 = 79$, $\sum X_{2i} = 12 \Rightarrow \hat{\theta}_2 = \frac{12}{79} = 0.152$

$\Rightarrow \hat{\theta}_{shared} = \frac{27+12}{81+79} = 0.244$

$$(\hat{\theta}_1-\hat{\theta}_2)_{Std.} = \frac{0.333-0.152}{\sqrt{0.244(1-0.244)\left(\frac{1}{81}+\frac{1}{79}\right)}} = 2.66 \notin [-1.96, 1.96]$$

$\Rightarrow$ REJ

Another (obvious) Wald Test: If $X_1, \dots, X_n$ iid DGP with mean θ and variance σ^2 and the estimator $\hat{\theta} = \bar{X}$, then the CLT implies that:

$$\frac{\hat{\theta}-\theta}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$ if σ is known

If σ is unknown ... I can replace σ with any consistent estimator e.g. $S, \hat{\sigma}$ and $\frac{1}{n}\sum(x_i - \theta)^2$

$$\Rightarrow \frac{\hat{\theta} - \theta}{\frac{S}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

Are you allowed to just use the T-test here? Many people just use the T-test here. Technically it's wrong because you need the DGP to be normal iid. But if you use the T-test ... it's "not so bad". I did this on problem 11 of the midterm;

$H_a: \theta > 2, \; n = 30, \; \bar{x} = 2.57, \; s = 1.00$

$$\hat{\theta}_{std} = \frac{2.57 - 2}{\frac{1.00}{\sqrt{30}}} = 3.12 \notin (-\infty, 1.645] \Rightarrow \text{REJ } H_0$$

Another Wald test for two independent samples with unknown variances and you wish to test a difference in means.

$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow{d} N(0,1) \Longrightarrow \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow{d} N(0,1)$$

from last class

If you use the Satterthwaite t-test, it "wouldn't be so bad" because unless your population distributions were so very skewed, it should be fine.

Let's use the asymptotic normality of the MLE thm (last class) to do a Wald test. HW 4 l,m has DGP: $X_1, \ldots, X_n$ iid Gumbel $(\theta, 1)$. The Gumbel is a r.v. model for "extreme events" think maximum rainfall per month.

$$\ell'(\theta; x_1, \dots, x_n) = n - e^{-\theta}\sum e^{-x_i} \overset{set}{=} 0 \Rightarrow \hat{\theta}^{MLE} = \ln\left(\frac{n}{\sum e^{-x_i}}\right)$$

$\Downarrow$

$$\ell'(\theta; x) = 1 - e^{-\theta}e^{-x} \Rightarrow \ell''(\theta; x) = -e^{-\theta}e^{-x} \quad ?$$

$$I(\theta) = E[-\ell''(\theta; x)] = E[e^{\theta}e^{-x}] = e^{\theta}E[e^{-x}] = e^{-2\theta}$$

$$\frac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} = \frac{\hat{\theta}^{MLE} - \theta}{\frac{e^{\theta}}{\sqrt{n}}} = \frac{\ln\left(\frac{n}{\sum e^{-x}}\right) - \theta}{\frac{e^{\theta}}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

$x_1 = 2.15$, $x_2 = 1.91$, $x_3 = 3.66$, $x_4 = 4.85$, $x_5 = 3.03$, $x_6 = 1.03$, $x_7 = 3.58$
$n = 7$

$\hat{\theta}^{MLE} = 2.26$ Test $H_a: \theta > 2$, $\alpha = 5\%$

$$\hat{\theta}^{MLE}_{std.} = \frac{2.26 - 2}{\frac{e^2}{\sqrt{7}}} = \frac{0.26}{2.79} = 0.09 \in (-\infty, 1.645]$$

$\Rightarrow$ RET H_0

There are three goals of statistical inference

(1) Point Estimation

Goal here is to provide a best guess, $\hat{\theta}$ of the value of θ. You don't know if your specific guess is good, is close, is bad, is far... How do we ask the question "is it good / bad"? We imagined $\hat{\theta}$ coming from a distribution $\hat{\theta}$, the "sampling distribution." There are properties about the sampling distribution e.g. some good properties are unbiasedness, consistency, low MSE, low risk (for general loss functions).

(2) Testing

Goal here is to test a theory about a specific θ. We used hypothesis testing. What makes a "good test"? One property is "power." There are other properties we didn't discuss.

(3) Confidence Sets

The goal here is to create a set of values for θ that you're "confident in." The approach we use here is the "confidence interval."

Definition: an "interval estimate" are two statistics: $W_L(x_1,\dots,x_n)$ & $W_u(x_1,\dots,x_n)$ such that $W_L < W_u$ for all data sets. combined in an interval: $[W_L(x_1,\dots,x_n), W_u(x_1,\dots,x_n)]$

e.g. $[1.789, 2.463]$

and of course, the "interval estimator" is:

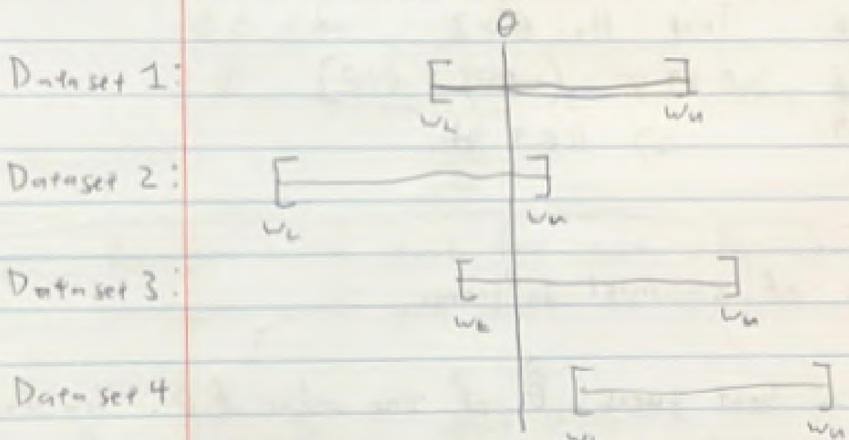
$$[W_L(X_1,\dots X_n), W_u(X_1,\dots X_n)]$$

which is a "random interval".

Definition: An interval estimator has "coverage probability"

$$P(\theta \in [W_L(X_1,\dots X_n), W_u(X_1,\dots,X_n)] \mid \theta).$$

An illustration



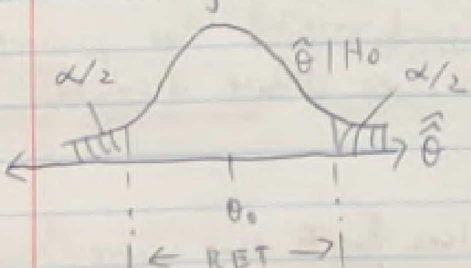
The coverage probability is computed over every dataset. For these four datasets, the coverage probability would be $3/4 = 75\%$

We define the "confidence interval" with coverage probability $1-\alpha$ for parameter θ as this interval estimate and interval estimator (depending on context). Denoted $CI_{\theta, 1-\alpha}$

Given α, how do we find the confidence interval?

Let's begin with the DGP iid normal mean θ, variance σ^2 and variance known and the estimator $= \bar{X}$.

Consider testing: $H_a: \theta \neq \theta_0$ vs. $H_0: \theta = \theta_0$ at size α



$$Z_{1-\frac{\alpha}{2}} := F_Z^{-1}\left(1-\frac{\alpha}{2}\right)$$

$$P(\hat{\theta} \in RET \mid H_0) = 1-\alpha$$

$$\parallel$$

$$P\left(\hat{\theta} \in \left[\theta_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \theta_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] \mid \theta = \theta_0\right)$$

$$= P\left(\hat{\theta} - \theta_0 \in \left[-z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, +z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] \middle| \theta = \theta_0\right)$$

$$= P\left(\theta_0 - \hat{\theta} \in \left[-z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, +z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] \middle| \theta = \theta_0\right)$$

$$= P\left(\theta_0 \in \left[\underbrace{\hat{\theta} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{W_L}, \underbrace{\hat{\theta} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{W_U}\right] \middle| \theta = \theta_0\right)$$

$$= P\left(\theta_0 \in \left[W_L(X_1, \ldots, X_n), W_U(X_1, \ldots, X_n)\right] \middle| \theta = \theta_0\right)$$ since valid $\forall \theta_0 \ldots$

$$\Rightarrow CI_{\theta, 1-\alpha} = \left[\hat{\theta} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \hat{\theta} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

We constructed our first confidence interval by "inverting the test."