Lecture 12

10/19/2020

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

We can use this now in our situation:

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

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

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

$\hat{\theta}_{shared}$ = average over both samples $= \frac{\sum X_{1i}+\sum X_{2i}}{n_1+n_2}$

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

e.g. $H_a: \theta_1-\theta_2 \neq 0$, $H_0: \theta_1-\theta_2=0$, $\alpha=5\%$ // 2-proportion z-test

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

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

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

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

$\Rightarrow$ Reject H_0

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

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

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

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

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 i.i.d. But if you use the T-test... it's "not so bad". I did this on problem 11 of the midterm:

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

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

Another Wald Test for 2 independent samples with unknown variances and you wish to test a different in means.

$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow[\text{from last class}]{d} N(0,1) \Rightarrow \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow{d} N(0,1)$$

If you use the Satterthwaite T-test, it "wouldn't be so bad" because unless your population distr 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: $X_1, \dots, X_n$ i.i.d. Gumbel $(\theta, 1)$. The Gumbel is a R.V. model for "extreme events" think maximum rainfall per month.

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

$$\Downarrow$$

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

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

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

$X_1 = 2.15$, $X_2 = 1.91$, 3.66, 4.85, 3.03, 1.03, 3.58, $n = 7$ //given

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

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

$\Rightarrow$ Retain H_0

There are 3 goals of statistical inference:

1) Point Estimation

Goal here is to provide a best guess, $\hat{\hat{\theta}}$ 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 $\hat{\hat{\theta}}$ coming from a distr $\hat{\theta}$, the "sampling distr". There are properties about the sampling distr 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

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".

Def an "interval estimate" are 2 statistics:

$$W_L(X_1,\dots,X_n) \;\&\; W_u(X_1,\dots,X_n) \text{ s.t. } W_L < W_u \text{ for all data sets}$$

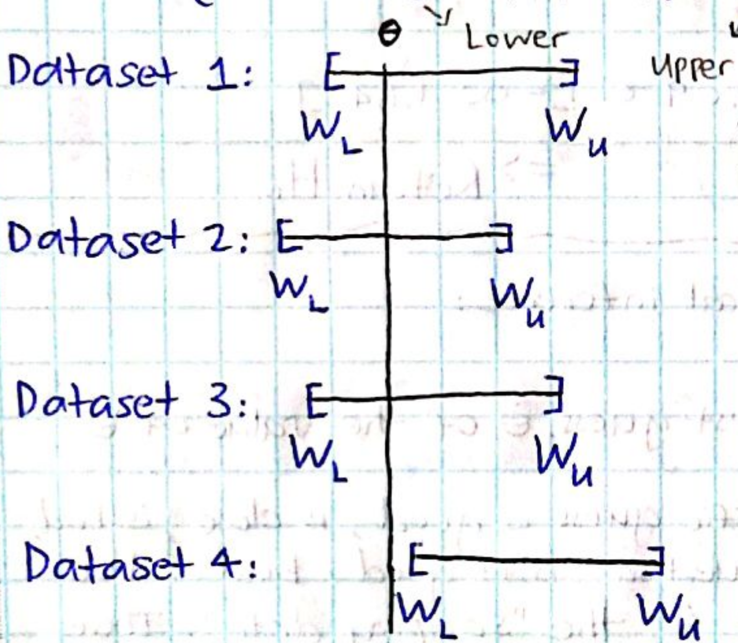
combined in an interval: $[W_L(X_1,\dots,X_n), W_u(X_1,\dots,X_n)]$

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".

Def 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 4 datasets, the coverage probability would be $\frac{3}{4} = 75\%$.

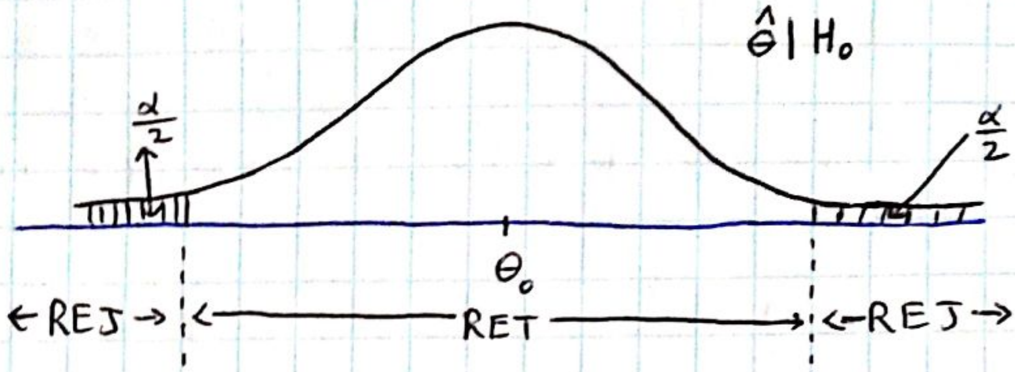
We define the "confidence interval" with coverage probability $1-\alpha$ for parameter θ as this interval estimate and interval estimator (depending on context).

Denoted $CI_{\theta, 1-\alpha}$

Given α, how do we find the CI? — confidence interval abbreviation

Let's begin with the DGP i.i.d. 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



$P(\hat{\theta} \in RET \mid H_0) = 1 - \alpha$ $\quad\quad$ $Z_{1-\frac{\alpha}{2}} := F_Z^{-1}(1-\frac{\alpha}{2})$

$\Downarrow$

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

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

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

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

$= P(\theta_0 \in [W_L(X_1, \dots, X_n), W_u(X_1, \dots, X_n)] \mid \theta = \theta_0)$ since valid $\forall \theta_0 \dots$

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

We constructed our first CI by "inverting the test"

$\downarrow$

confidence interval abbreviation