## W8: Wednesday

Ch 5.3
~~5.6~~, ~~5.7~~, ~~5.8~~

Last time we saw that

$$CI_{95\%} \text{ for } \mu : \Big[ \underset{\uparrow}{\bar{X}} - \frac{\sigma^2}{\sqrt{n}} \underbrace{z_{0.025}}, \ \bar{X} + \frac{\sigma^2}{\sqrt{n}} \cdot z_{1-0.025}$$

centering

radius: std error of centering estimator

---

<u>EX</u>: CI for binomial p.

$X_1, \ldots, X_n \sim \text{Bin}(n, p)$

HW7, p.5

MLE estimator of $p$ is $\boxed{\hat{p} = \frac{1}{n} \sum_{i=1}^{n} \xi_i = \frac{1}{n^2} \sum_{i=1}^{n} X_i}$

Bernouli random variables.

By CLT we know that

$$\hat{p} \longrightarrow N(\,?,\ ??\,)$$

? $p(1-p)$

$$? = E\hat{p} = E\xi_1 = p.$$

$$\text{var}\,\hat{p} = \frac{1}{n^2} \cdot \sum_{i=1}^{n} \text{var}\,\xi_i = \frac{\text{var}\,\xi_1}{n} = \frac{p(1-p)}{n}.$$

$$\hat{p} \longrightarrow N, \qquad N \sim N\left(p, \frac{p(1-p)}{n}\right)$$

$$\Rightarrow \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \longrightarrow Z, \qquad Z \sim N(0,1).$$

→ Fix level ~~&~~ $\alpha$

then.

$$\mathbb{P}\left( -z_{1-\frac{\alpha}{2}} \le \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \le z_{1-\alpha/2} \right) = 1-\alpha.$$

Isolate $p$ in the midle

Not so easy now as to isolate $p$ we have to solve quadratic inequality:

$$\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \le z_{1-\alpha/2}$$

$$\hat{p}-p \le z_{1-\alpha} \cdot \sqrt{\frac{p(1-p)}{n}}.$$

$$\therefore (\hat{p}-p)^2 \le z_{1-\alpha}^2 \cdot \frac{p(1-p)}{n} = \frac{p-p^2}{n} \cdot z_{1-\alpha}^2$$

$$\vdots$$

quite messy.

Insted it is justifiable to replace $p$ by $\hat{p}$ in the denominator:

$$\mathbb{P}\left( -z_{1-\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \le z_{1-\alpha/2} \right) = 1-\alpha.$$

Because $\hat{p} \to p$ and $\hat{p}(1-\hat{p}) \to p(1-p)$.

these two numbers are close

→ CI ~~$\alpha$~~ $(1-\alpha)\%$ for $p$ is:

$$\left[ \hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \cdot z_{1-\alpha/2},\ \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \cdot z_{1-\alpha/2} \right]$$

Radius of the CI for p for a 95%-interval

is $\overset{2}{\cancel{1.96}} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

→ How do you choose a sample size to get a $(1-\alpha)$ CI to have radius $= \varepsilon$

(to be reasonably sure of knowing p to be within $\varepsilon$ either way).

$$\varepsilon = \underset{\sim}{z_{1-\alpha}} \sqrt{\frac{p(1-p)}{n}}$$

$$\varepsilon^2 \approx 4 \cdot \sqrt{\frac{p(1-p)}{n}} \Rightarrow n = \frac{4 \cdot p(1-p)}{\varepsilon^2}$$

If $\varepsilon = 10\%$ and $p = \frac{1}{2} \Rightarrow n = \frac{1}{\varepsilon^2} = 100$

Note that CI for p gets shorter when p is near 0 or 1. (but then CLT approximation is less reliable).

## Ch. 7.4

$X \sim N(\mu, \sigma^2)$ $\mu$ - unknown
$\sigma$ - unknown

$$\boxed{\frac{\bar{X} - \mu}{S_n / \sqrt{n}} = t_{n-1}}$$

t-distribution with $(n-1)$ df.
- continuous
- symmetric.

$$\mathbb{P}\left(-t_{n-1, \alpha/2} < \frac{\bar{X} - \mu}{S_n / \sqrt{n}} < t_{n-1, \alpha/2}\right) = 1 - \alpha.$$

(in the same way as when $\sigma$ was known)

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$ - unbiased estimator of $\sigma^2$

## Ch. 7.5

$X \sim N(\mu, \sigma^2)$ . $\mu$ - unknown
$\sigma$ - unknown.

$$\boxed{\frac{(n-1) S^2}{\sigma^2} = \chi^2_{n-1}.}$$

$$\Rightarrow \mathbb{P}\left(\chi^2_{\alpha/2, n-1} \leq \frac{(n-1) S^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2}, n-1}\right) = 1 - \alpha.$$

$$CI_{\sigma^2} = \left[\frac{(n-1) S_n^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}, \frac{(n-1) S_n^2}{\chi^2_{\frac{\alpha}{2}, n-1}}\right]$$