Lec 21 Math 241 12/1/16

Previously

CLT is: $X_1, \dots, X_n \overset{iid}{\sim}$ w/ mean μ, s.e. σ and n large

II $\quad \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{d}{\approx} Z \sim N(0,1)$

III $\quad \bar{X} \overset{d}{\approx} N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

IV $\quad T \overset{d}{\approx} N\left(n\mu, \left(\sqrt{n}\,\sigma\right)^2\right)$

$f(t)$ Lightbulb failure time

$\mu \approx 1000$ hr, $\sigma \approx 500$ hr

buy 50 lightbulbs

prob avg is more than 1300 hr?

$$P(\bar{X} \geq 1300) = P\left(\frac{\bar{X} - 1000}{\frac{500}{\sqrt{50}}} \geq \frac{1300 - 1000}{\frac{500}{\sqrt{50}}}\right) \underset{\text{CLT II}}{\approx} P(Z \geq 4.24) \approx 0$$

standardize $\bar{X}$

Suppose we lose 2% of orders. In 10,000 orders, what's prob more than 3% lose?

$$P(\bar{X} > 3\%) = P\left(\frac{\bar{X} - .02}{.0014} > \frac{.03 - .02}{.0014}\right) \overset{\text{by CLT}}{\approx} P(Z > 7.14) \approx 0$$

$$X_1, \dots, X_{10000} \overset{iid}{\sim} \text{Bern}(0.02) \overset{\text{by CLT}}{\Rightarrow} \bar{X} \approx N\left(\mu, \left(\tfrac{\sigma}{\sqrt{n}}\right)^2\right) = N(.02, .0014^2)$$

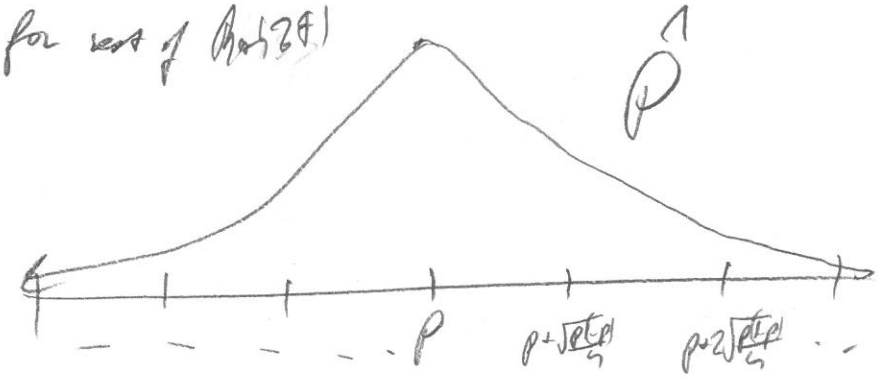
$$\Rightarrow \mu = p = 0.02, \quad \sigma = \sqrt{p(1-p)} = \sqrt{.02 \cdot .98}, \quad \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{.02 \cdot .98}{10000}} = .0014$$

$\bar{X}$ gets a special notation if $X_1, \dots, X_n \overset{iid}{\sim}$ Bern, $\bar{x}$ denotes

$$\underbrace{\hat{P}}_{\text{sample proportion r.v.}} = \bar{X} \qquad \underbrace{\hat{p}}_{\text{sample proportion}} := \bar{x} := \frac{\sum x_i}{n} = \frac{\#1\text{'s}}{n}$$

$$\bar{X} \approx N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

$$\Rightarrow \underbrace{\hat{P} \approx N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)}_{\text{focus for rest of Math 241}} \quad \text{for Bernoullis}$$



Who likes mushrooms? $\hat{p} = \dots$ what is p? INVERSE problem!!

PROB ↑
STAT ↓

$X_1, \ldots, X_N \overset{iid}{\sim} \text{Bern}(p)$

$|N| \approx \infty$

p is unknown as it is

$p := \frac{\sum x_i}{N}$

Goal:

1) Estimate p as a single pt
2) Estimate a range of p's which make sense
3) Test theories about what p is.

How? Take a "finite sample" or "small sample" of size $\underbrace{n << N}_{\text{very less than}}$

Sample

$X_1, \ldots X_n$

Sample must be "representative" which means it preserves iid property. How? Simple random sample. All males? All college students? No... completely at random using random # generator.

How to get ~~best~~ guess of p? $\hat{p} := \frac{\sum x_i}{n} = \frac{\# 1\text{'s}}{n}$ sample proportion.

Where does $\hat{p}$ come from? $\hat{p}$ is a realization from $\hat{P}$

lowercase uppercase

Interval procedure:

What if I take $\left[\hat{p} \pm \sqrt{\frac{p(1-p)}{n}}\right] = \left[\hat{p} - \quad , \hat{p} + \quad \right]$

What is the probability of this many repeats of this interval procedure ~~of~~ covers p?

$$P\left(p \in \left[\hat{P} \pm \sqrt{\frac{p(1-p)}{n}}\right]\right)$$

lots of P's: $P(\,)$, p, $\hat{p}$, $\hat{P}$

$$= P\left(\hat{P} - \sqrt{\frac{p(1-p)}{n}} \le p \le \hat{P} + \sqrt{\frac{p(1-p)}{n}}\right)$$

$$= P\left(-\sqrt{\frac{p(1-p)}{n}} \le p - \hat{P} \le + \sqrt{\frac{p(1-p)}{n}}\right)$$

$$= P\left(-1 \le \frac{p - \hat{P}}{\sqrt{\frac{p(1-p)}{n}}} \le 1\right)$$

$$= P(-1 \le -Z \le 1)$$

$$= P(1 \ge Z \ge -1)$$

$$= P(-1 \le Z \le 1)$$

$$= P(Z \in [-1, 1]) = .68$$

What if I ~~do~~ $\left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$

St $z_{\frac{\alpha}{2}} := F_Z^{-1}\left(1-\frac{\alpha}{2}\right) \iff 1-\frac{\alpha}{2} = \int_{-\infty}^{z_{\frac{\alpha}{2}}} f_Z(x)\,dx$

what does this mean?

$\alpha = 5\% \Rightarrow \frac{\alpha}{2} = 2.5\% \Rightarrow 1-\frac{\alpha}{2} = 97.5\%$

what is $F_Z^{-1}(97.5\%) = 2 = z_{2.5\%}$

$1-\frac{\alpha}{2} = \int_{-\infty}^{-z_{\frac{\alpha}{2}}} f_Z(x)\,dx + \int_{-z_{\frac{\alpha}{2}}}^{z_{\frac{\alpha}{2}}} f_Z(x)\,dx$

by symmetry

$1-\frac{\alpha}{2} = \int_{-z_{\frac{\alpha}{2}}}^{\infty} f_Z(x)\,dx = \int_{-z_{\frac{\alpha}{2}}}^{z_{\frac{\alpha}{2}}} f_Z(x)\,dx + \int$

$\Rightarrow \frac{\alpha}{2} = \int_{z_{\frac{\alpha}{2}}}^{\infty} f_Z(x)\,dx$

via symmetry

$= \int_{-\infty}^{-z_{\frac{\alpha}{2}}} f_Z(x)\,dx$

$$P\left(p \in \left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]\right)$$

$$= P\left(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$$= P\left(\hat{p} \in \left[-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right]\right) = F_Z\left(z_{\frac{\alpha}{2}}\right) - F_Z\left(-z_{\frac{\alpha}{2}}\right)$$

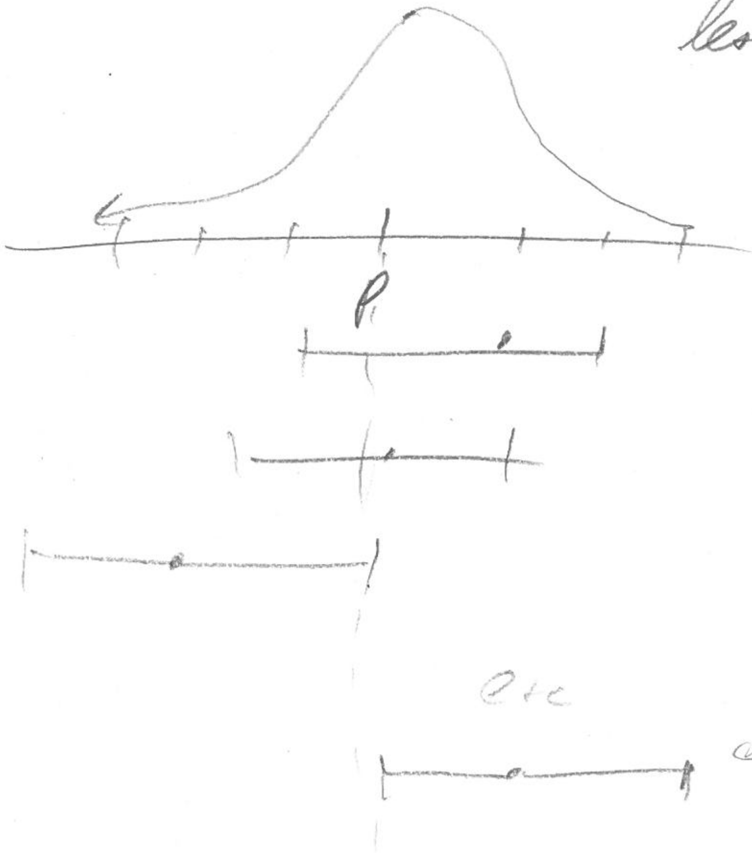
$$= \left(1-\frac{\alpha}{2}\right) - \left(\frac{\alpha}{2}\right)$$

$$= 1-\alpha$$

So the multiple $z_{\frac{\alpha}{2}}$ allows me to pick any prob of the interval.

So $\left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$ is a procedure, when repeated, gives $1-\alpha$ prob that p is contained within it.

let $\alpha = 5\% \Rightarrow z_{\frac{\alpha}{2}} = 2$



5% of the time this happens

Big prob: p unknown! But..

$$\left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right] \approx \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

as long as $p \not\approx 0$ and $p \not\approx 1$

Hotly debated

$$\underbrace{CI_{p,1-\alpha}}_{\text{Confidence Interval}} := \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Said to have $1-\alpha$ coverage of p. What does this mean?