Lecture 21 5/5/15 Math 241

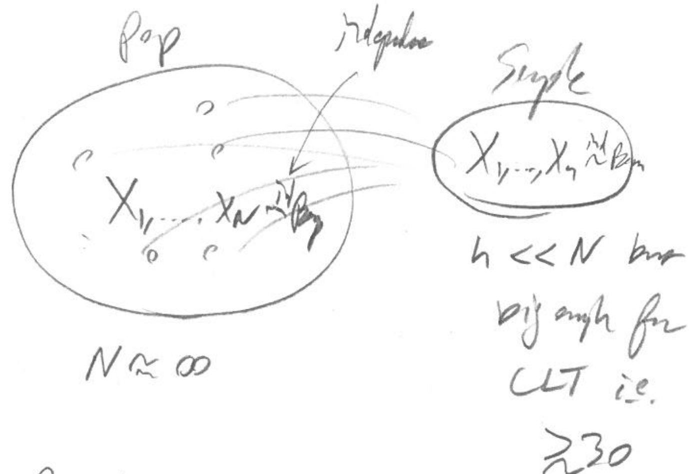
Review CLT

$X_1,\ldots,X_n \overset{iid}{\sim}$ mean μ, SE σ. If n large...

$\Rightarrow \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

$\Rightarrow \bar{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right) \longrightarrow \hat{p} \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$ if $X_1,\ldots,X_n \overset{iid}{\sim} \text{Bern}(p)$

$\Rightarrow T_n \sim N\left(n\mu, (\sqrt{n}\,\sigma)^2\right)$



if sample taken as SRS, everything we're going to do is valid now.

Pop has parameters

$$p := \int\cdots\int_{C_1\cdots C_K} P(Y|C_1,\ldots,C_K)\,P(C_1,\ldots,C_K)\,dC_1\cdots dC_K$$ i.e. the pop prob of "success"

anything you care about

Even when $X_1,\ldots,X_n \overset{iid}{\not\sim} \text{Bern}(p)$, we still get

$\hat{p} \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$ it just goes slower (beyond scope of course)

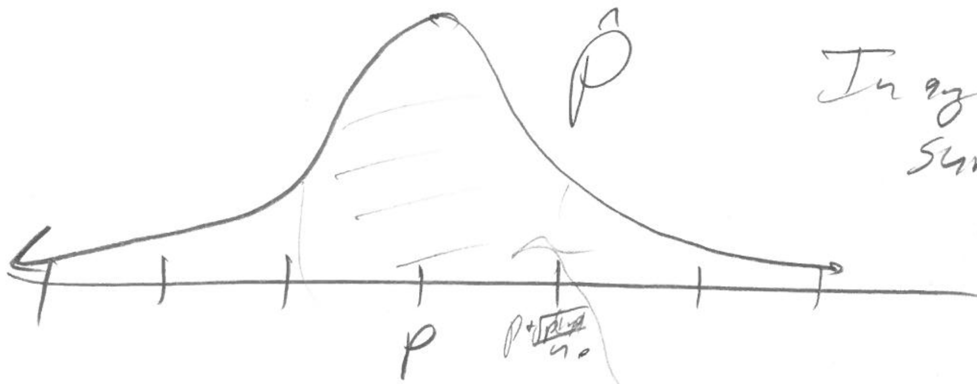
What does this mean?

e.g. let me $\hat{p} = \frac{\#\text{mushroom lovers}}{\#\text{students}} = ?$

Goal: Inference/Estimation: Use $\hat{p}$ to "infer" p, the unknown, inaccessible parameter.

Best guess — just use $\hat{p}$ itself. It's unbiased $E(\hat{p}) = p$ since $E(\bar{X}) = \mu$

But what if we want to give a range of possible values of p?



In any one experiment/poll/survey, n items get tallied to create a $\hat{p}$ which is a draw from $\hat{P}$.

Imagine if I do the following... create an interval:

abbreviation

$$\left[\hat{p} - \sqrt{\frac{p(1-p)}{n}},\ \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right] =: \left[\hat{p} \pm \sqrt{\frac{p(1-p)}{n}}\right]$$

margin of error

$\hat{p}$

How often does this interval capture the true pop prop?

$$P(\text{coverage}) := P\left(p \in \left[\hat{P} - \sqrt{\frac{p(1-p)}{n}},\ \hat{P} - \sqrt{\frac{p(1-p)}{n}}\right]\right) = P\left(\hat{P} - \sqrt{\frac{p(1-p)}{n}} \le p \le \hat{P} + \sqrt{\frac{p(1-p)}{n}}\right)$$

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

$$P\left(p + \sqrt{\frac{p(1-p)}{n}} \ge \hat{P} \ge p - \sqrt{\frac{p(1-p)}{n}}\right) = P\left(\hat{P} \in \left[p - \sqrt{\frac{p(1-p)}{n}},\ p + \sqrt{\frac{p(1-p)}{n}}\right]\right) = P(|Z| \le 1) \approx 0.68$$

two more steps

this covers p as long as $\hat{p}$ is within one SE away.

what about? $\left[\hat{p} \pm 2\sqrt{\frac{p(1-p)}{n}}\right]$? $\Rightarrow P(|Z| \le 2) = 0.95$

$\left[\hat{p} \pm 3\sqrt{\frac{p(1-p)}{n}}\right] \Rightarrow P(|Z| \le 3) = 0.997$

if I want to NOT cover p α of the time...

3

Generally, intervals of the form...

$$\left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \quad \text{where } z_{\frac{\alpha}{2}} := \text{the solution to } 1 - F_Z(z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$$

have coverage $1-\alpha$ so it does not cover α.

e.g. if $\alpha = 5\% \Rightarrow \frac{\alpha}{2} = 2.5\%$

$$\Rightarrow 2.5\% = \int_{z_{\frac{\alpha}{2}}}^{\infty} f_Z(x)\,dx \rightarrow z_{\frac{\alpha}{2}} = 2 \text{ by empirical rule}$$

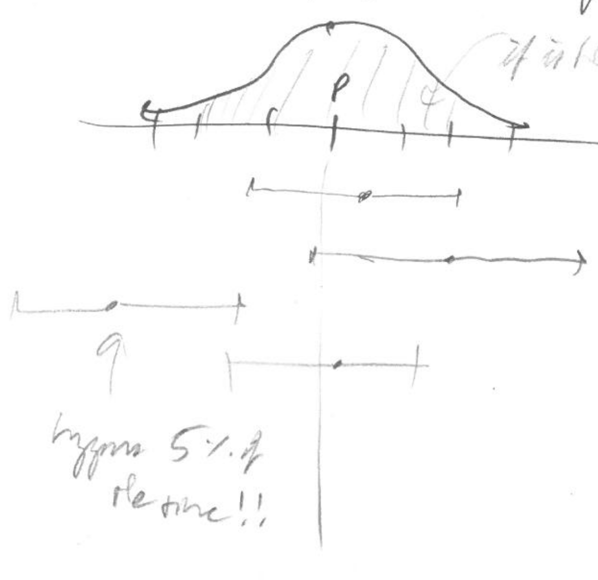
We introduce ~~this is the defn of~~ the "Confidence Interval"

$CI_{p,\,1-\alpha} :=$

what you are inferring

coverage level determined by α the "user-set" P(not covering)

Remember: Best guess: $\hat{p}$ but $\hat{p} \neq p$. Why? $SE(\hat{p}) > 0 \Rightarrow$ r.v. is random!



Let's say $\alpha = 5\%$. Look what happens!

(4)

But there's a problem! $\hat{p} \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$

$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$ we don't know p!! this's what we're trying to find in the first place!!!!

So... Assume $\sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

PhD ~~theses~~ have been written about this approx. ~~It~~ is a pretty good approx. Accept it as fact.

$\Rightarrow CI_{p,1-\alpha} := \left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$ (technically a 2-sided 1-prop CI)

Frequentist / Objectivist Interpretation

either it's inside (1) or outside (0).

① Before you begin experiment, the $P(CI \text{ cont } p) = 1-\alpha$
Think about a coin: before it is flipped $P(H) = \frac{1}{2}$
(determined ~~by "tails"~~ or propensity theory)

② If my samples/models/experiments are ~~taken~~, in the long run, 95% of the CI's will contain the true prop. Just like r.v.s.

Problem: given a CI already computed for a $\hat{p}$ sampled, there is no prob. statement! $P(p \in [0.3, 0.7])$ not a r.v.! It's degenerate (0 or 1)

Big problem with interpretation! Why care?
Imagine a coin already flipped? $P(H)$?

P(H) is 0 or 1 according to the frequentist. There is no physical process to generate a random state in the [illegible].

P(OJ guilty) is 0 or 1. Not random!

But he doesn't think like a frequentist! he thinks subjectively

$P(H) = \frac{1}{2}$, $P(\underset{\text{guilty}}{OJ \text{ ~~[illegible]~~}}) = 0.99$

$\Rightarrow$ a subjective interpretation of CI is $P(\theta \in CI) = 1-\alpha$.
I will let you believe this even though it is not classic.

$\underset{\text{confidence}}{1-\alpha} = \underset{\text{prob}}{1-\alpha}$ or $\underset{\text{confid}}{1-\alpha} \neq \underset{\text{prob}}{1-\alpha}$ (Philosophical distinction)

CI's solve an important prob. Pt. estimates are flawed since you don't know how sure you are. CI's give you a way to quantify uncertainty in your pt. estimate. Uncertainty decreases as function of n. E.g.

$\hat{p} = 0.27$ $n=100$ $CI_{p,95\%} = \ldots$

$\hat{p} = 0.27$ $n=1000$ $CI_{p,95\%} = \ldots$

more complex..