Lecture 21 Math 281 12/1/15

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

p ← ~~the~~ proportion of mushroom lovers

~~the prob of someone liking mushrooms~~

pop.

Consider $N \approx |\mathbb{N}|$

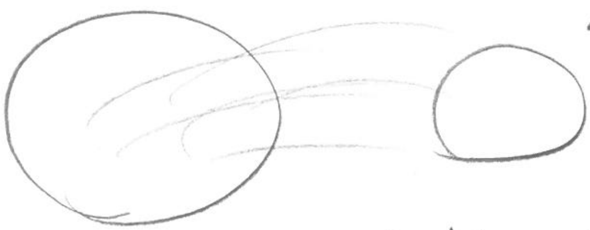
$\Rightarrow$ Knowing p is not possible

Goal:
1. Estimate p
2. Test hypothesis of p
3. Make decisions based on p

$\left.\right\}$ Inference

↖ what we do $\Rightarrow$

$\Rightarrow$



Take sample of size $n << N$ but large enough for the CLT "to kick in"

$\Rightarrow X_1, \ldots, X_n \overset{iid}{\sim} Bern(p)$

Sample must be "representative" of entire population.

What is pop here?

"Simple random sample": best means to select a sample to be representative

All: choice

Most likely...

$$X_1 \mid C_1 = 1, \ldots, X_n \mid C_n = 1 \stackrel{iid}{\sim} \text{Bern}(p_c^{*}) \qquad p_c^{*} > p$$

$\Rightarrow$ Sample biased

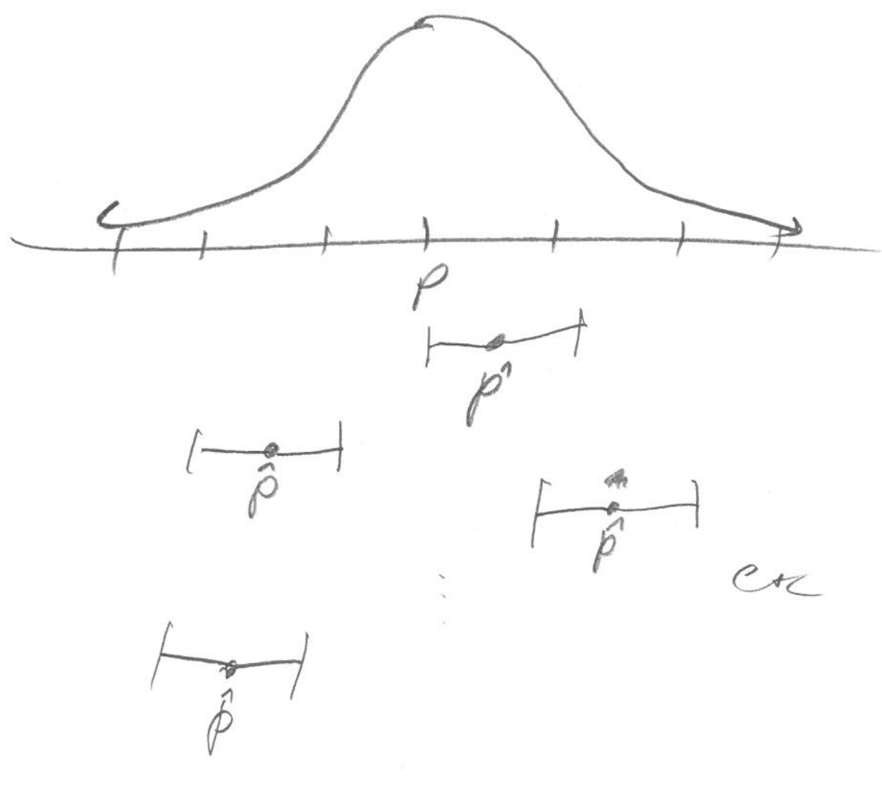
SRS balances all other dependent r.v.'s!

Sample $\Rightarrow$ inference on pop. param. $\quad \hat{p} := \frac{\#1\text{'s}}{n} = \frac{1}{n}\sum x_i$

Point est. $\hat{p}$ realization from $\hat{p} \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$

Interval est. Imagine I did the following

How?
CLT
later
(lecture)



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

$\hat{p}$ est — Margin of error (MOE)

How often does this capture the true pop. p? ("coverage")

the r.v.

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

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

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

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

$$= P\left(1 \ge Z \ge -1\right) = P\left(Z \in (-1,1)\right) = .68$$

$z_{\frac{\alpha}{2}}$ is the z s.t. $F(z) = 1 - \frac{\alpha}{2}$

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

e.g. $\alpha = 5\%$

z s.t. $F(z) = .975$

$\Rightarrow z = 2$

$\alpha = 32\%$

z s.t. $F(z) = .84$

$\Rightarrow z = 1$

etc...

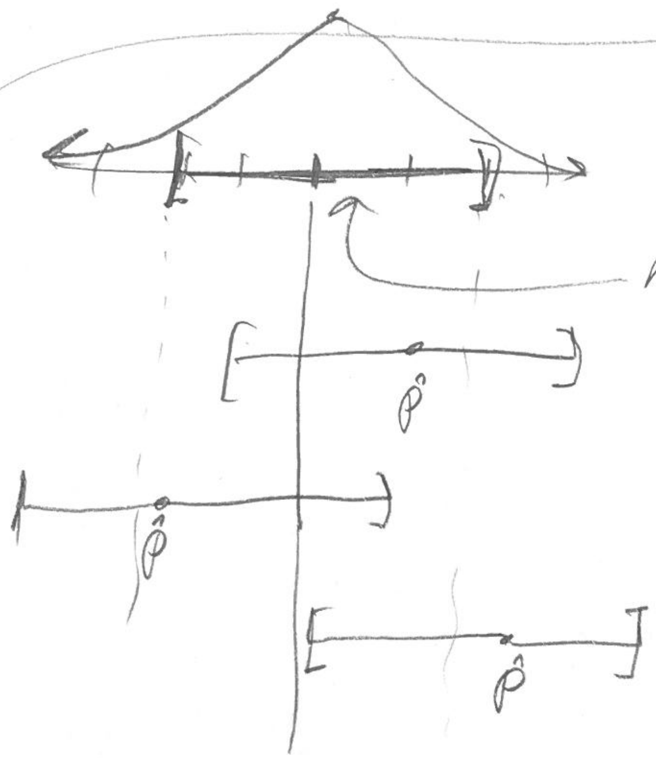
$$P\left(z_{\frac{\alpha}{2}} \ge Z \ge -1\right) = P\left(Z \in \left[-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right]\right)$$

$= 2 - \alpha - 1 = \boxed{1-\alpha}$

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

$\boxed{4}$

Pick $\alpha \Rightarrow \overset{\text{Compute}}{z_{\frac{\alpha}{2}}} \Rightarrow$ Calc $\left(\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right) \Rightarrow P(\text{coverage}) = 1-\alpha$.



m.o.e.

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

$\Rightarrow$ covers

o/w no

How often does it fall outside?

α of the time.

But Big problem: p is unknown...

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

as long as p isn't ≈ 0 or ≈ 1

Very debated!!

$P(\text{coverage}) \approx 1-\alpha$ → "Confidence Interval"

→ param p

→ confidence $1-\alpha$

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

Frequentist / Objective Interpretation

Recall: p is the pop. prop. It is a #, one #!

① Before you take sample OR if you take many samples...

$$\underbrace{P(p \in CI)}_{\text{r.v.}} = P\left(p \in \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]\right) = 1-\alpha$$

$$\frac{\#\ \hat{p} \in CI}{n} \xrightarrow{LLN} p$$

② After you take sample

$$P(p \in CI) = P\left(p \in \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]\right) \in \{0,1\}$$ $\Rightarrow$ no prob. statement can be made so any given CI is useless!

$P(H) = \frac{1}{2}$ before flip $\quad P(H) \in \{0,1\}$ after flip

③ Subjectivist Interpretation / Bayesian interpretation

$\Rightarrow$ prior belief on where p should be $\Rightarrow$ p is a r.v.

$\Rightarrow P(\hat{p} \in CI) = 1-\alpha$ under certain assumptions... Math 390.03!

Objectivist $P(\text{OJ Simpson guilty}) \in \{0,1\}$ past event & no way to measure using propensity or l.r.f.

Subjectivist $P(\text{OJ Simpson guilty}) \in [0,1]$ your subjective belief matters

$1-\alpha$ confidence $\neq$ $1-\alpha$ probability

except if you're a subjectivist with specific prior knowledge

mushroom lovers = 24   n = 47 students

$\hat{p} = 24/47 = 0.511$

$\approx Z_{2.5\%}$

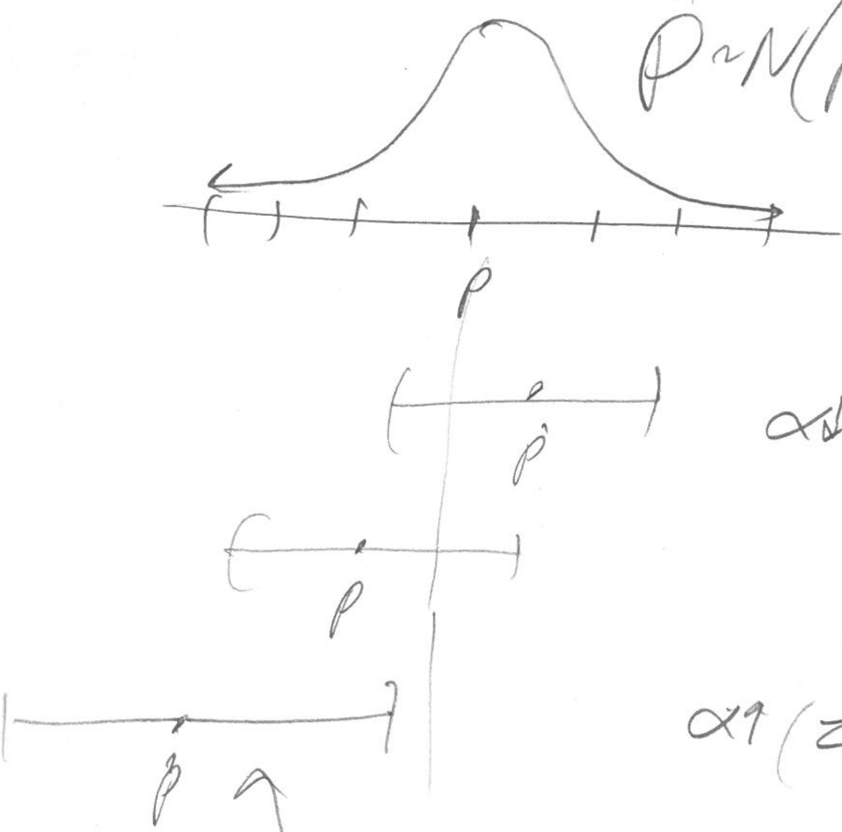
$CI_{p,95\%} = \left[\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = \left[0.511 \pm 2\sqrt{\frac{.511 \cdot .489}{47}}\right]$

$= [0.365, 0.657]$

A 95% CI for the pop. prop. of true mushroom lovers is

Interpret...

$\hat{p} \sim N\left(p, \left(\sqrt{\frac{p(1-p)}{n}}\right)^2\right)$



with bigger n, CI's get tighter

What's another way to get tighter?

$\alpha\downarrow$ ($Z_{\frac{\alpha}{2}}\downarrow$) Cons: lower coverage, less interesting

How to guarantee high coverage?

$\alpha\uparrow$ ($Z_{\frac{\alpha}{2}}\uparrow$) Cons: intervals very big, not useful

don't know if we're covered or not!! That's the reality

New paradigm: testing: is a theory about p true or not?

No longer care about ~~estimating~~ p....

theory: male/females born in same prop

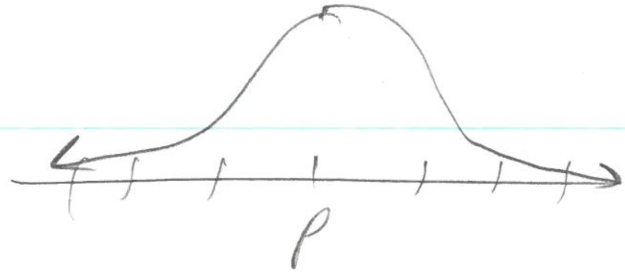
Let $p := P(\text{male}) = 0.5$ ← arbitrary

Now use sample to test theory!

Under theory

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(p = 0.5)$

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



Known distr. if we assume theory is true...