But we can "estimate" or "infer" P from a sample.

P is known as a "parameter"
Best "point estimate" is $p = \hat{p}$

$\bar{X} \approx \mu$ by LLN

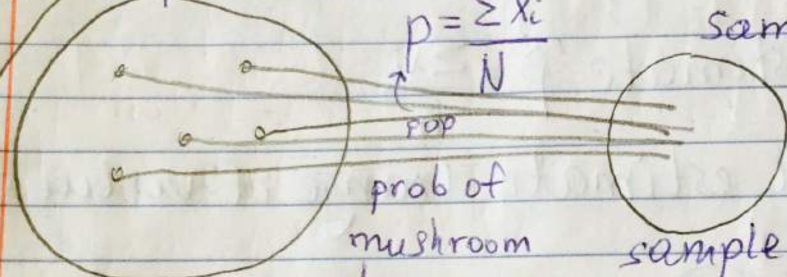
What about an interval estimate?

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

Population

$p = \frac{\sum X_i}{N}$ pop

sample (survey) → simple random (SRS)



prob of mushroom lovers

sample size $n << N$

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

probability (population level)

Knowing p is impossible

Goals

① Estimate p (CI)

② Test Theorem about p (HT's)

③ Make decisions using p

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

• How to take a sample? (SRS)

$X_1 | C_1 = 1, \dots, X_n | C_n = 1 \overset{iid}{\sim} Bern(p_c)$

↑ chinese yes/NO

s.t. $p_c > p$

Biased sample

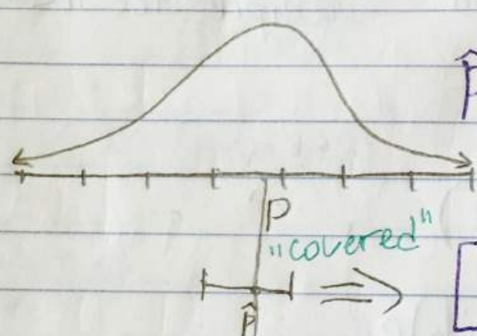
Non-repetative sample

Estimates

① Point estimate $\hat{p} = \frac{\sum x_i}{n}$

② interval estimate (range of values)

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



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

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

coverage or confidence := p(p ∈ interval)

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

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

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

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

$$= p\left(1 \ge \underbrace{\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}}_{\to \text{standard normal}} \ge -1\right)$$

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

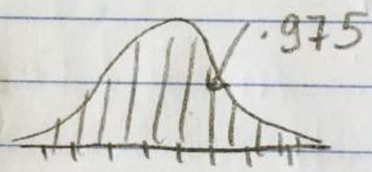
$$= p(Z \in [-1,1]) = .68$$

$Z_{\frac{\alpha}{2}} :=$ the z s.t.

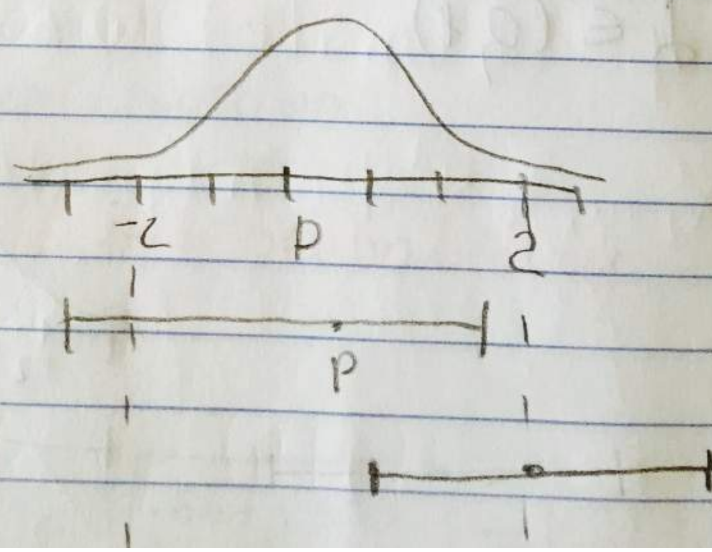
$F_Z(z) = 1 - \frac{\alpha}{2}$

$\alpha = 5\%$

$F_Z(z) = .975$



$Z_{2.5\%} = z$



• from previous problem instead of 1 is Z

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

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

$$= P(Z \geq Z \geq -Z)$$

$$= P\left(Z \in \left(-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}\right)\right) = F\left(Z_{\frac{\alpha}{2}}\right) - F\left(-Z_{\frac{\alpha}{2}}\right)$$

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

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

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

$$= 2 - \alpha - 1$$

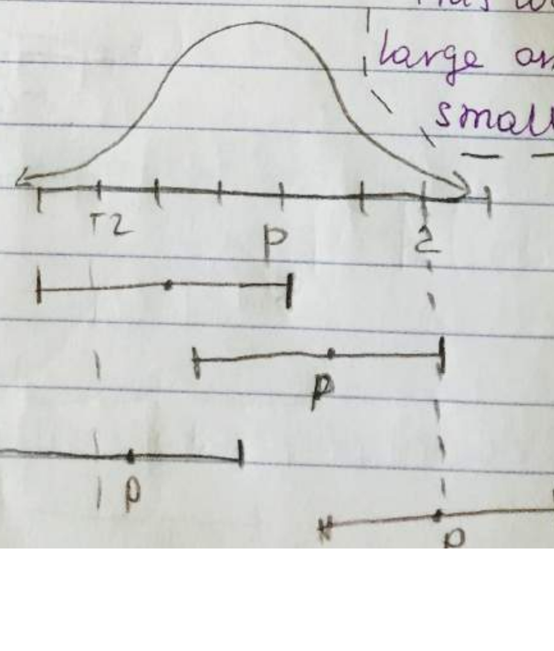
$$= \boxed{1-\alpha}$$

$\alpha \in (0,1)$

point estimate

$$\left[\hat{p} \pm \underbrace{Z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}}_{\text{margin of error}}\right] \approx \left[\hat{p} + \underbrace{Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-}{n}}}_{\text{marg of er}}\right.$$

This works if n
large and p not
small not lar



confidence interval

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

parameter confidence level

Objectivist (Frequetist Interpretation)

① before you make a CI

$P(p \in CI) = 1-\alpha$

② If I repeat the problem many times

$1-\alpha \approx \frac{\sum \mathbb{1}_{p \in CI_i}}{n}$ L.r.6

③ After CI is computed

$P(p \in CI) \in \{0,1\} \neq 1-\alpha$

Subjectivist / Bayesian

Under certain prior beliefs about p

$P(p \in CI) = 1-\alpha$
after construction

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

$\hat{p} = \frac{10}{15} = .66$

$$CI_{p,95\%} = \left[.667 \pm 2\sqrt{\frac{.667 \cdot .333}{15}}\right] = [.424, .910]$$

0.12 $.133$
$.243$

- $\alpha = 5\% \Rightarrow z = 2$
  $\alpha = 0.3\% \Rightarrow z = 3$

$$CI_{p,95\%} = \left[.667 \pm \underbrace{3\sqrt{\frac{.667 \cdot .333}{15}}}_{\substack{.133 \\ .365}}\right] = [.300, 1.00]$$

as goes down
$\alpha \downarrow$ interval is
Useless

- $\hat{p} = \frac{100}{150} = .667$

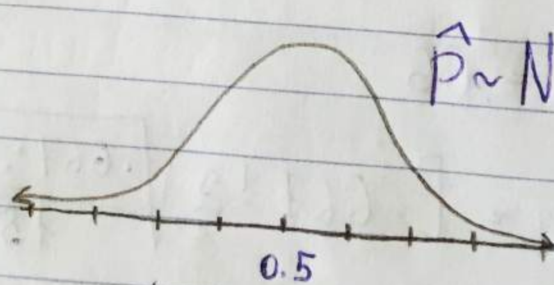
$$CI_{p,95\%} = \left[.667 \pm \underbrace{2\sqrt{\frac{.667 \cdot .333}{150}}}_{\substack{0.38 \\ 0.77}}\right] = [.59, .74]$$

Is my theory about p true?

Theory: male/female births are same proportion

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

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



$$\hat{p} \sim N\left(0.5, \left(\sqrt{\frac{0.25}{n}}\right)^2\right)$$