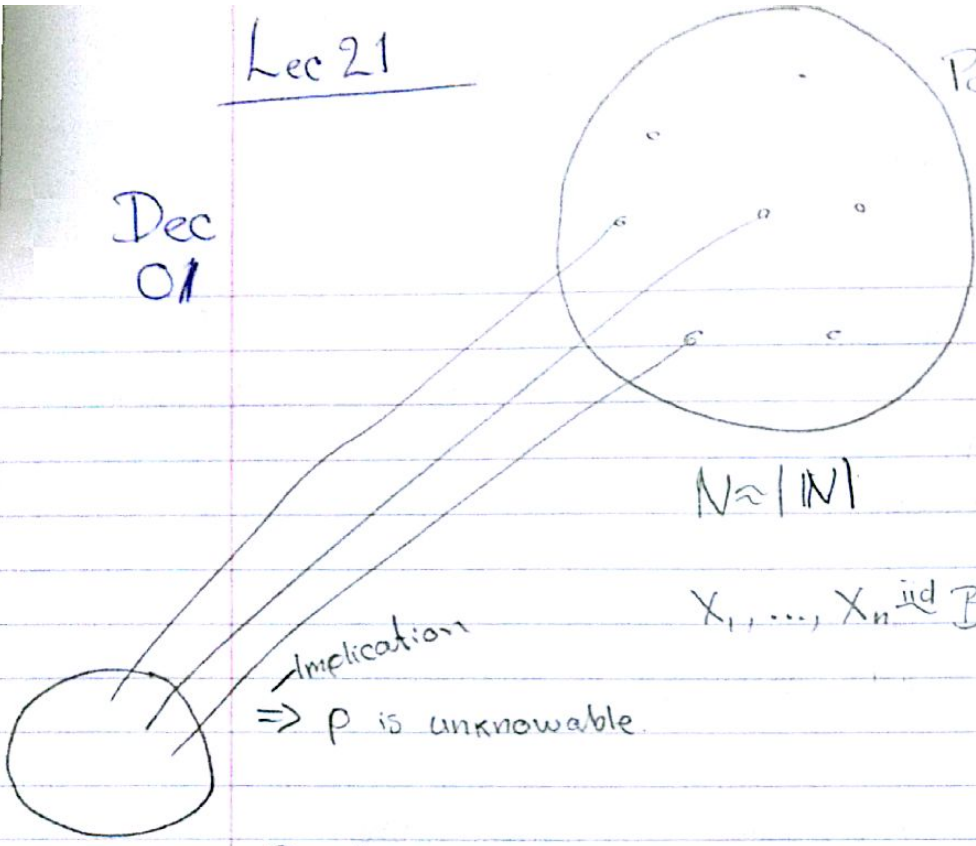
Lec 21

Dec 01



$p := \dfrac{\sum x}{N}$ → Population Proportion of Mushroom Lovers.

$N \approx |N|$

$X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ → Probability Of Being a Mushroom Lover.

Implication

$\Rightarrow$ p is unknowable.

Goal: (1) Estimate p

(2) Test theories about p

(3) Make decisions about p.

Sample size is n. $n << N$, but large enough for CLT to kick in.

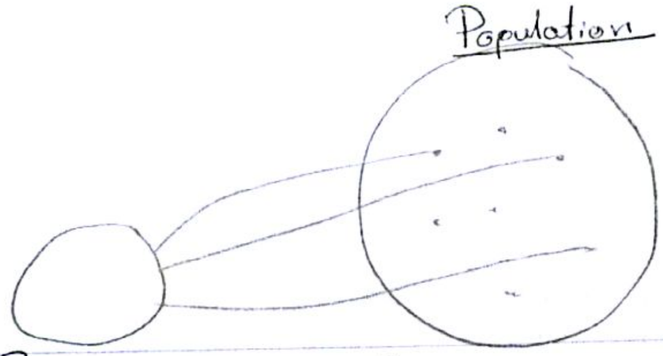
$X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ → Unknowable

Need representation Sample.

Non-Representative Or Biased.

$X_1 \mid C_1 = 1, \dots X_n \mid C_n = 1 \overset{iid}{\sim} \text{Bernoulli}(p_c)$

$\overset{iid}{\sim} \text{Bernoulli}(p_c)$, such that $p_c > p$



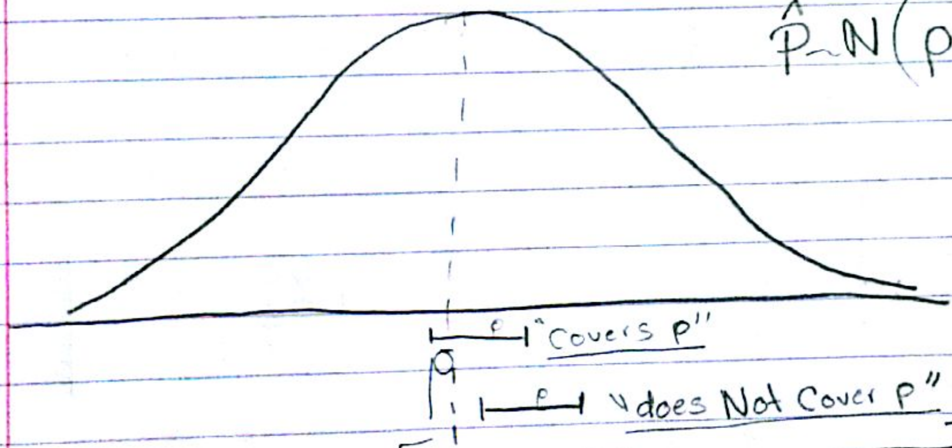
examples

Simple Random Sample gives non-biased estimates.

Point-Estimation

$$\hat{p} = \frac{\sum x_i}{n}$$

Interval Estimate (range of possible values).



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

by Central Limit Theorem

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

Margin Of Error

Estimate

$$[3 \pm 1] = [2, 4]$$

Range Of Error

(Coverage - Confidence) Interval

= CI

$$P(p \in \text{interval}) =$$

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

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

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

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

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

$$= \mathbb{P}(Z \in [-1, 1]) = 0.68 = 68\%$$

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

$Z_{\frac{\alpha}{2}} := z$ such that $F_Z(z) = 1 - \frac{\alpha}{2}$

$\alpha := 5\%$

$\alpha := 2.5\%$

$Z_{5\%} = 2$ → Up to 2

$Z_{\frac{\alpha}{2}} = \pm Z$

$\mathbb{P}(Z \in [$

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

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

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

$$\boxed{\alpha \in (0,1)}$$

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

$$= 2 - \alpha - 1$$

$= 1 - \alpha$ Coverage/confidence → Probability Of Capturing "p" is 70%

Let $\alpha = 0.3 = 30\%$
$1-\alpha = 0.7 = 70\%$

→ 30% do not cover $\hat{p}$
With what % are we comfortable <u>not</u> covering p?

One Sample - One Proportion

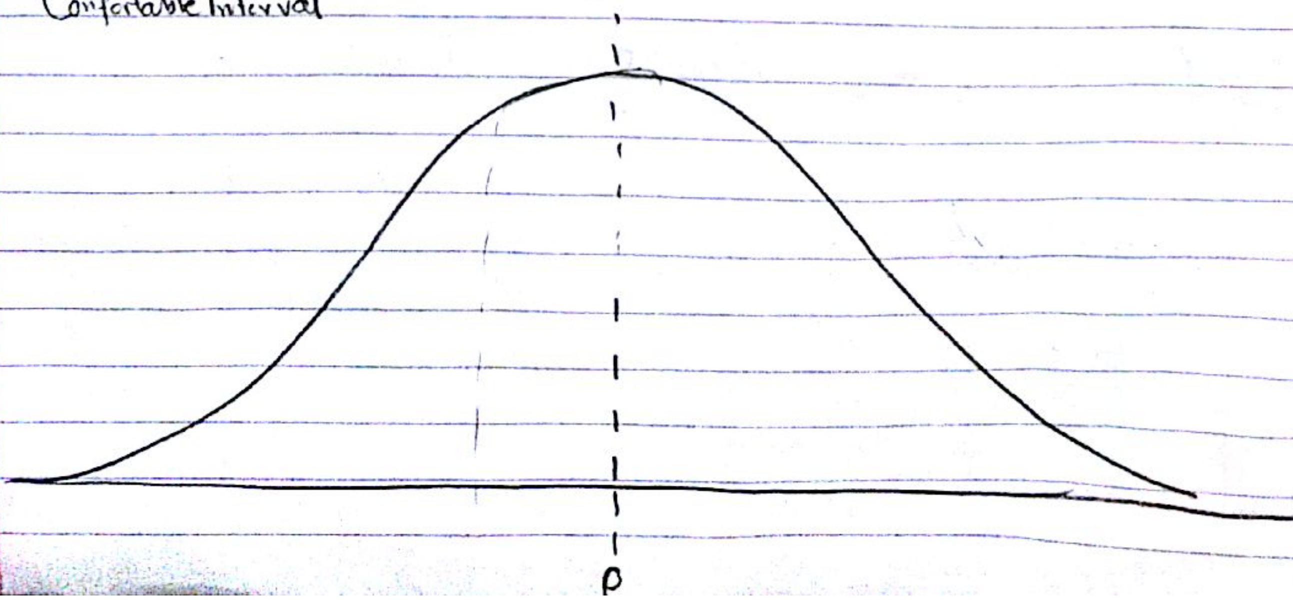
$$\left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right] \approx \left[\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Works well if n large but p is not [illegible] 0 or 1

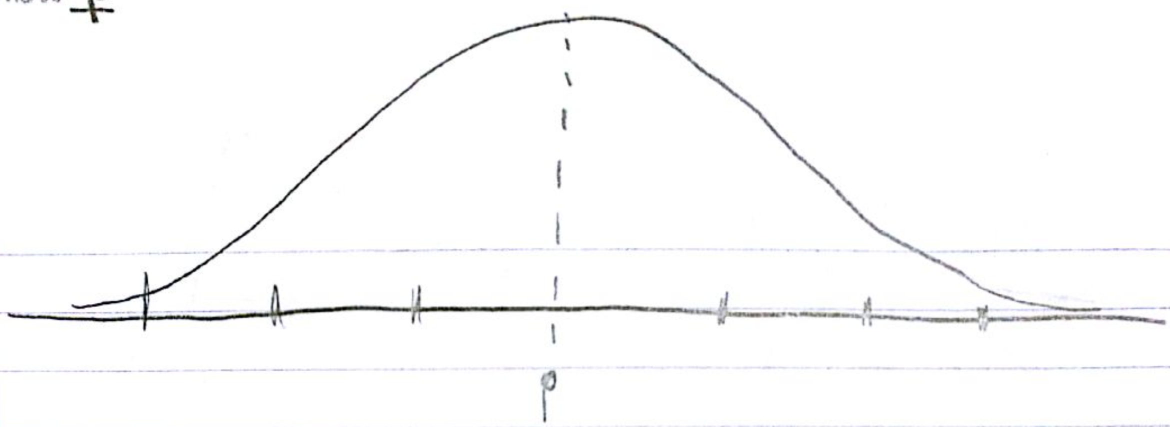
$\downarrow$

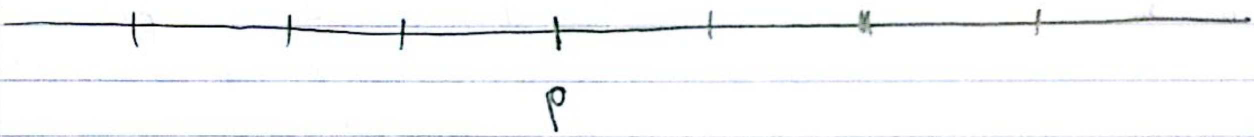
$CI_{p,1-\alpha}$

Comfortable Interval



We don't know p





① - According to Frequentist/Objectivist View of Probability

$\mathbb{P}(p \in CI) = 1-\alpha$ Before Experiment

② If Many Experiments Are Performed:

$$\left.\frac{\sum_{i=1}^{m} \mathbb{1}_{p \in CI}}{m} \approx 1-\alpha \right\} \text{Long Run Frequency}$$

③ After the experiment $\mathbb{P}(p \in CI) \in \{0,1\} \neq 1-\alpha$

Subjectivist/Bayessian Interpretation Under specific prior information/beliefs.

$$\underbrace{\mathbb{P}(p \in CI) = 1-\alpha}_{\text{After the experiment}}$$