FoDA L6

Central Limit Theorem
and
Estimation

Data set : $P = \{p_1, p_2, \dots p_n\}$

observations iid from distribution f

iid: Identically and Independantly Distributed

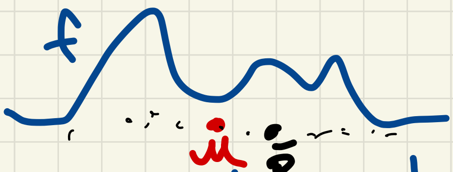
each p_i comes from same distribution

one observation does not influence another

Random Variable X w/ pdf f_X

$X \sim f_X$

drawn from



interested in μ mean $f_X = E[X]$

observations $P = \{P_1, P_2 \dots P_n\}$ from f_X

sample $\bar{P} = \frac{1}{n} \sum_{P_i \in P} P_i$

R.V.

$\bar{P} = \frac{1}{n}\sum_{i=1}^{n} \{P_i\} \leftarrow \{X_i\} \underset{\text{iid}}{\sim} f_X$

realization

Central Limit Theorem

$X_1, X_2, \dots X_n$ drawn iid from f

$\{X\}_i \underset{iid}{\sim} f$

mean $f = \mu$

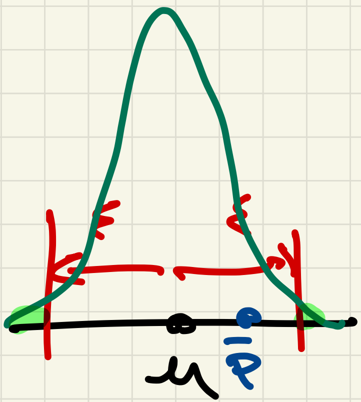
variance $f = \underset{\text{bounded}}{\sigma^2}$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- as $n \to \infty$ $f_{\bar{X}}$ converges to a normal
- $E[\bar{x}] = E[X_i] = \mu$
- $\text{var}[\bar{x}] = \frac{\text{Var}[x_i]}{n} = \frac{\sigma^2}{n}$ $\rightarrow$ $\text{sd-dev}(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

How to measure accuracy?

- ~~$\bar{P}$ is at most ε far from μ.~~
- ~~With high $(1-\delta)$ probability~~



$\cancel{\bar{P} = \mu}$ $\cancel{\Pr[\bar{P} = \mu] = 0}$

Probably Approximately Correct (PAC)

$$\Pr\left[\left|\bar{X} - \underbrace{E[\bar{X}]}_{\mu}\right| \geq \varepsilon\right] \leq \delta$$