Estimation

Using data to approximate some fixed number

Ex: estimating the Expected value (mean) of a unfair coin

Suppose we have an unfair coin

$X \in \{0, 1\}$ ← Tails $\qquad X \sim \mathbb{P}_X = \text{Bernoulli}(\alpha)$

$p(1) = \alpha, \quad p(0) = 1 - \alpha$

$$\mathbb{E}[X] = \sum_{x \in \{0,1\}} x\, p(x) = 0(1-\alpha) + 1(\alpha) = \alpha$$

What if we don't know $\mathbb{E}[X]$

how do we estimate it?

Flip the coin n times and use that

$(X_1, X_2, \ldots, X_n) \in \{0,1\}^n$

$X_i \sim \mathbb{P}_X = \text{Bernoulli}(\alpha)$ and independent

for all $i \in \{1, \ldots, n\}$

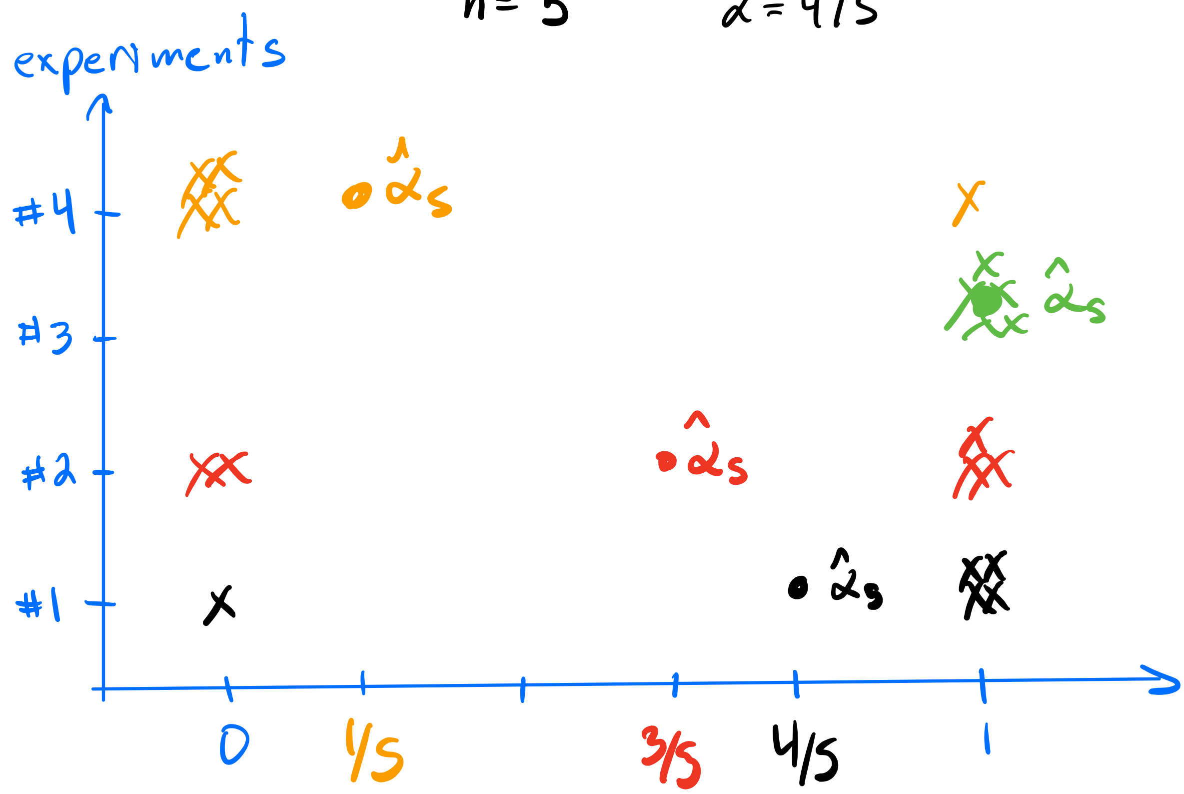
$$\hat{\alpha}_n = \bar{X} = g(X_1, \ldots, X_n) = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\mathbb{E}[\hat{\alpha}_n] = \mathbb{E}\left[\frac{1}{n}(X_1 + \ldots + X_n)\right]$$

$$= \frac{1}{n}\left(\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]\right)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} \mathbb{E}[X_1]\right) = \mathbb{E}[X_1] = \alpha$$

$n = 5 \qquad \alpha = 4/5$



$$\mathrm{Var}[\hat{\alpha}_n] = \mathrm{Var}\left[\tfrac{1}{n}(X_1 + \cdots + X_n)\right]$$

$$= \frac{1}{n^2}\mathrm{Var}[X_1 + \cdots + X_n]$$

independence

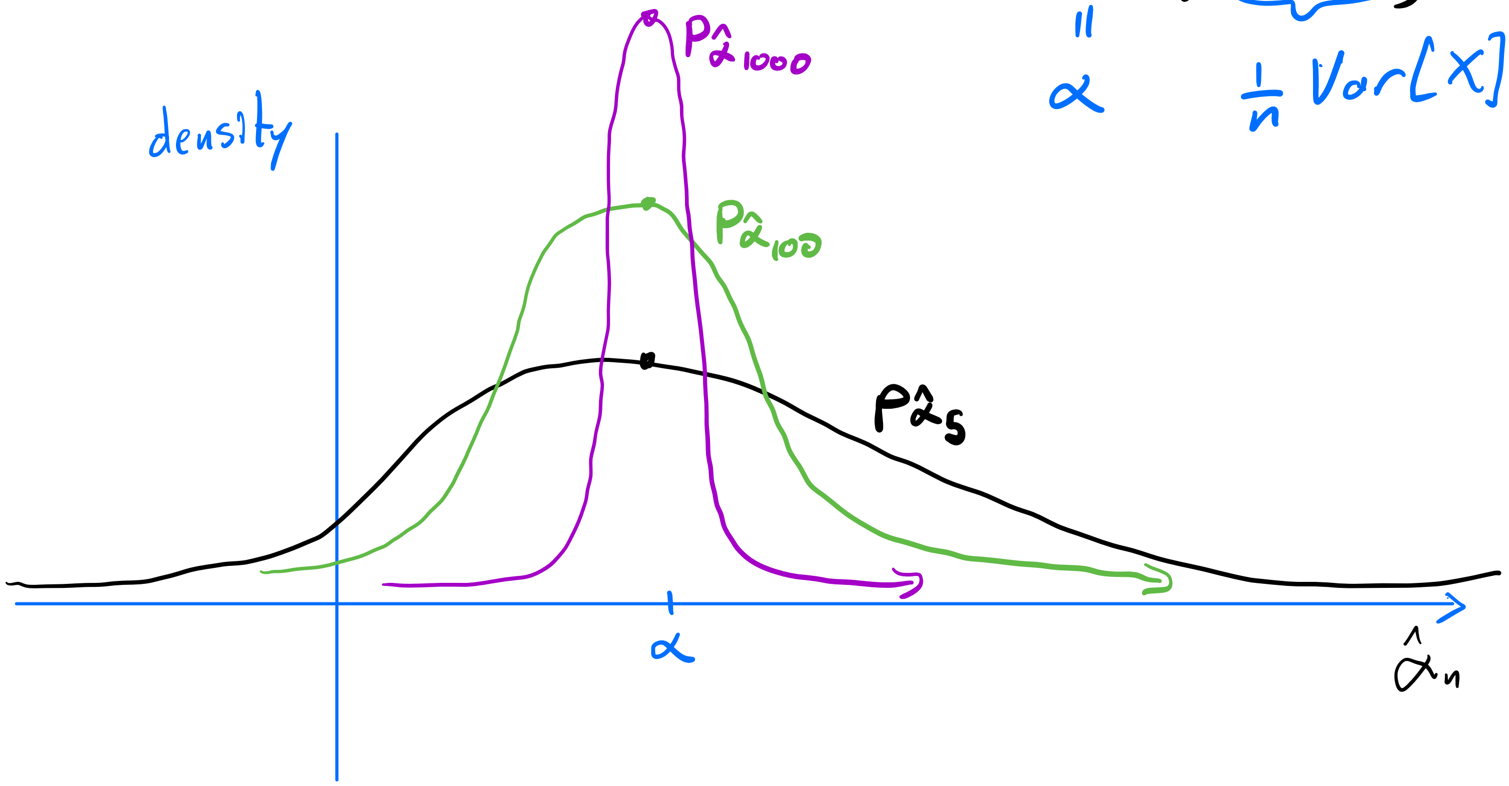
$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i]$$

identically distributed

$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_1] = \frac{1}{n}\mathrm{Var}[X_1]$$

as n increases the variance goes down

$$\mathbb{P}_{\hat{\alpha}_n} \overset{\text{central limit theorem}}{\approx} \mathcal{N}\Big(\underset{\alpha}{\underset{\parallel}{\mathbb{E}[\hat{\alpha}_n]}}, \underbrace{Var[\hat{\alpha}_n]}_{\frac{1}{n} Var[X]}\Big)$$



Estimating the loss function

$(\vec{X}, Y) \sim \mathbb{P}_{\vec{X},Y}$ is a r.v.

$f(\vec{X})$ is a r.v.

$Z = \ell(f(\vec{X}), Y)$ is a r.v.

We don't know $L(f) = \mathbb{E}[\overbrace{\ell(f(\vec{X}), Y)}^{Z}]$

$D = ((\vec{X}_1, Y_1), \ldots, (\vec{X}_n, Y_n))$ where

$(\vec{X}_i, Y_i) \sim \mathbb{P}_{\vec{X},Y}$ and independent for all $i \in \{1, \ldots, n\}$

$$\hat{L}(f) = \frac{\overbrace{\ell(f(\vec{X}_1), Y_1)}^{Z_1} + \cdots + \overbrace{\ell(f(\vec{X}_n), Y_n)}^{Z_n}}{n}$$

$Z_1, \ldots, Z_n$ are also i.i.d

$$\mathbb{E}[\hat{L}(f)] = L(f), \quad Var[\hat{L}(f)] = \frac{1}{n} Var[\ell(f(\vec{X}_1), Y_1)]$$

if n is large enough

$$\hat{L}(f) \approx L(f)$$

ERM: $\mathcal{A}(D) = \hat{f}_D$

$$\hat{L}(\hat{f}_D) \approx L(\hat{f}_D) \ ?$$

Sometimes, it will depend on $\mathcal{F}$

for now we assume $\hat{L}$ is a good estimate