Estimation

Using Data to approximate a fixed number

Ex: estimate the expected value (mean) of an unfair coin

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

$$P_X(1) = \alpha$$

$$\mathbb{E}[X] = 1 \cdot \alpha + 0 \cdot (1-\alpha) = \alpha$$

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

Flip the coin n times and take the average

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

$X_i \sim \text{Bernoulli}(\alpha)$ and independent for all $i \in \{1, \ldots, n\}$

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

$$\mathbb{E}[\hat{\alpha}] = \mathbb{E}\left[\frac{1}{n}\cdot(X_1 + X_2 + \cdots + X_n)\right]$$

$$= \frac{1}{n}\mathbb{E}[(X_1 + X_2 + \cdots + X_n)]$$

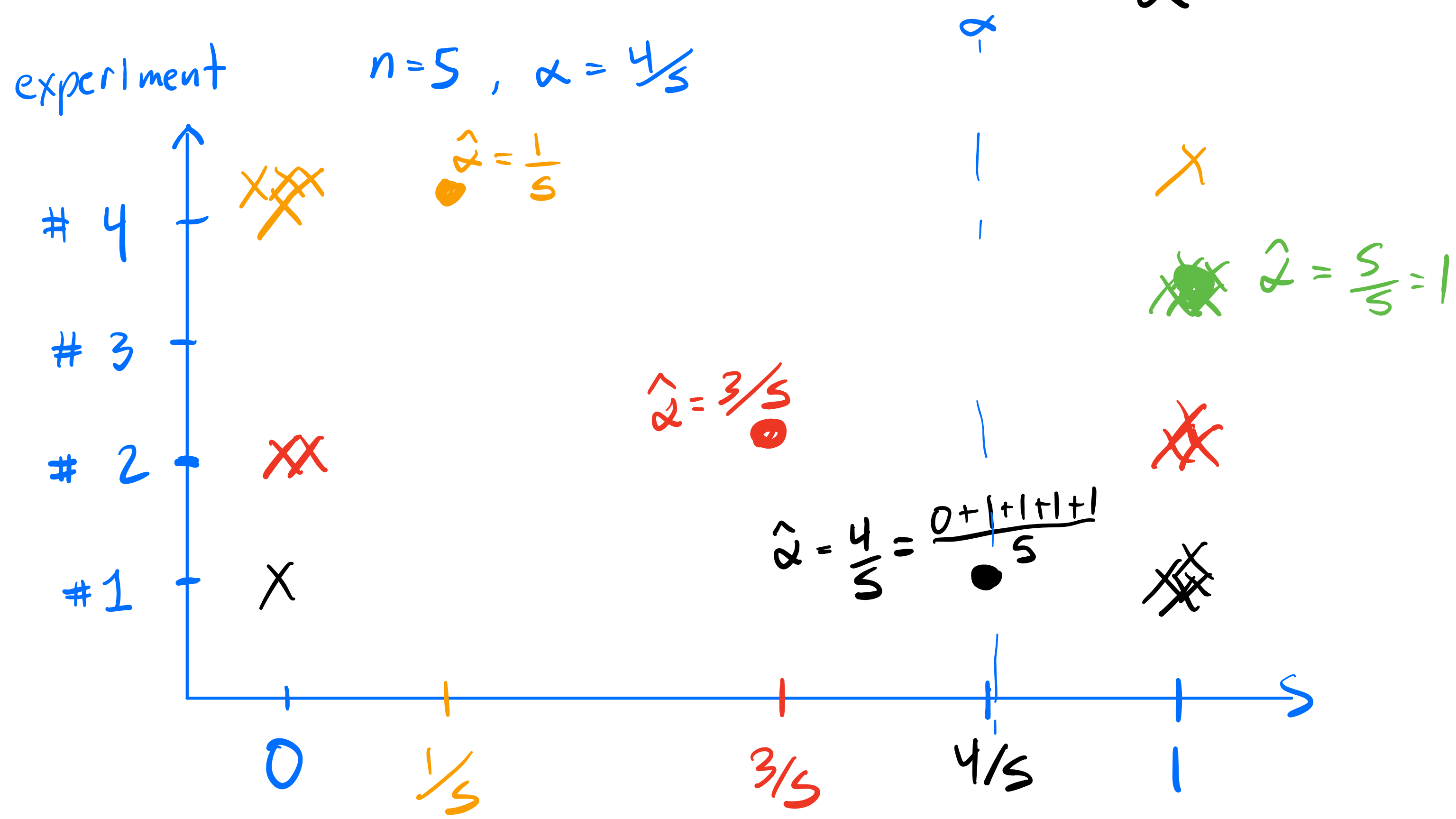
linearity of Expectation $= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i]$

identically distributed $= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_1] = \frac{n}{n}\mathbb{E}[X_1] = \mathbb{E}[X_1]$

$$= \mathbb{E}[X_2] = \cdots = \mathbb{E}[X_n]$$

$$\sum_{i=1}^{n} 2 = n \cdot 2$$

$$= \alpha$$



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

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

independence

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

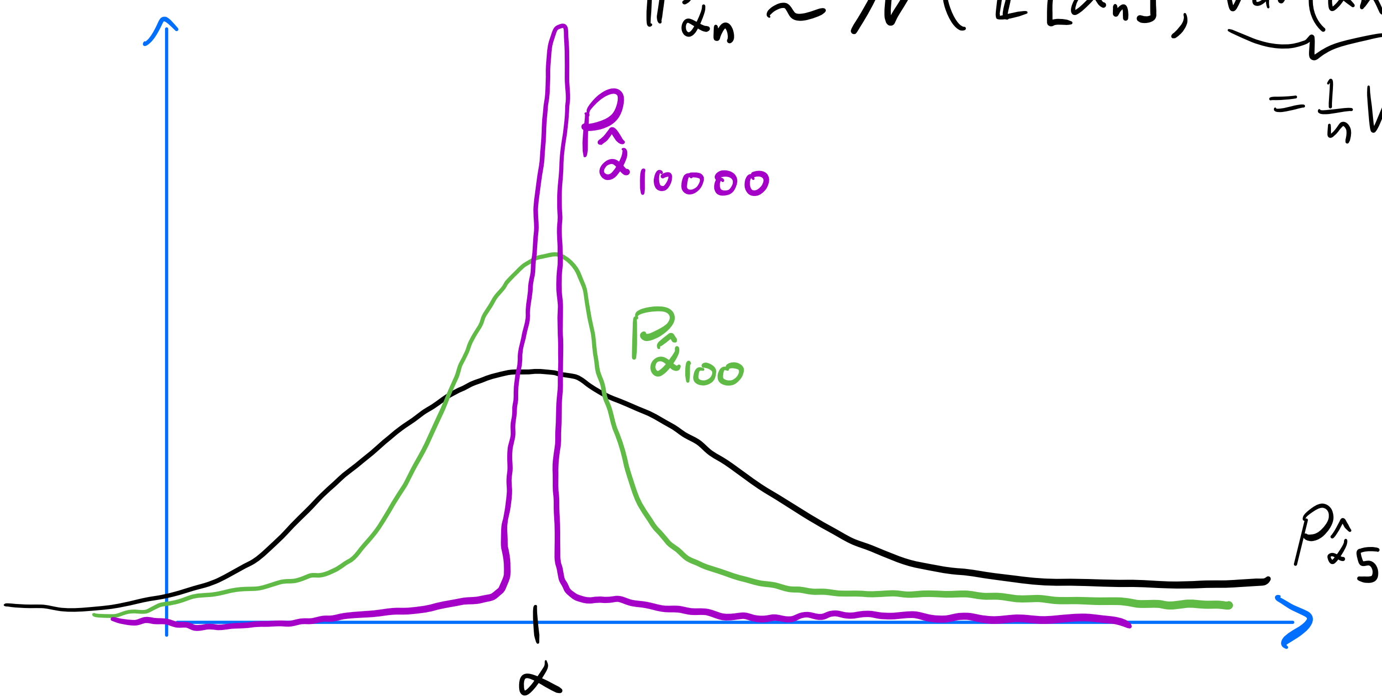
$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_1]$$

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

$$= \frac{1}{n}\underbrace{\mathrm{Var}[X_1]}_{\alpha(1-\alpha)} = \dots = \frac{1}{n}\mathrm{Var}[X_n]$$

Central Limit theorem

$$\mathbb{P}_{\hat{\alpha}_n} \approx \mathcal{N}\Big(\mathbb{E}[\hat{\alpha}_n],\ \underbrace{\mathrm{Var}(\hat{\alpha}_n)}_{=\frac{1}{n}\mathrm{Var}[X_1]}\Big)$$



Estimating the Risk/Expected loss

$\underbrace{(\vec{X}, Y)}_{\text{r.v.}} \sim \mathbb{P}_{\vec{X},Y}$ fix $f: \mathcal{X} \to \mathcal{Y}$

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

$\ell(f(\vec{X}), Y)$ r.v.

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

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

$(\vec{X}_i, Y_i) \sim \mathbb{P}_{\vec{X},Y}$ and independent

$(\underbrace{\ell(f(\vec{X}_1), Y_1)}_{=Z_1}, \dots, \underbrace{\ell(f(\vec{X}_n), Y_n)}_{=Z_n})$

$= (Z_1, \dots, Z_n)$

$\ell(f(\vec{X}_i), Y_i) \sim \mathbb{P}_{\ell(f(\vec{X}), Y)}$ and independent

$\ell(\mathcal{A}(D)(\vec{X}_i), Y_i)$

$$\hat{L}(f) = \frac{\sum_{i=1}^{n} \ell(f(\vec{X}_i), Y_i)}{n}$$

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

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

We want n large

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

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

It will depend on $\mathcal{F}$ and n