Estimation

Using data to approximate some fixed object

Ex: estimating the mean or variance of a r.v.

Suppose we have an unfair coin

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

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

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

Suppose you don't know $\mathbb{E}[X] = \alpha$

How do we estimate α?

Ans: flip the coin multiple times and take the average

$Z = (X_1, \dots, X_n) \in \{0,1\}^n$ represents n coin flips

$X_i \sim \mathbb{P}_X = \text{Bernoulli}(\alpha)$ and independent for all $i \in \{1, \dots, n\}$

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

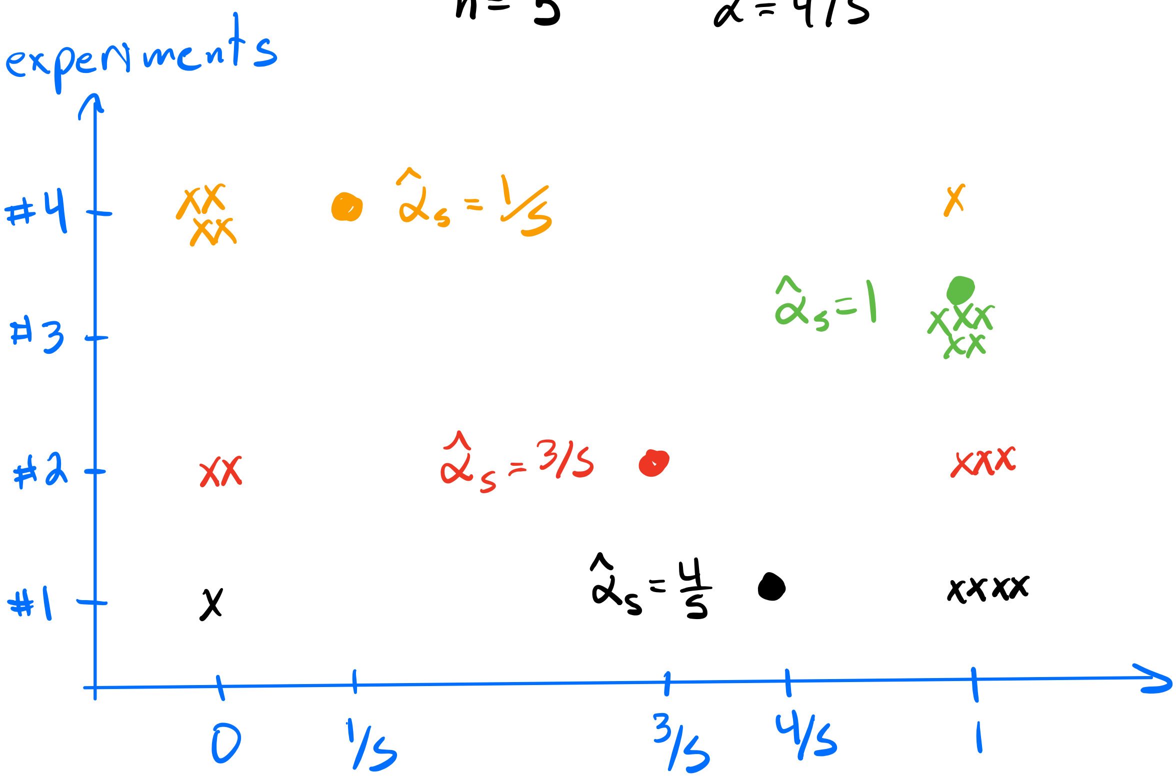
$$\mathbb{E}[\hat{\alpha}_n] = \mathbb{E}\left[\tfrac{1}{n}(X_1 + \dots + X_n)\right] = \frac{1}{n}\left(\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]\right)$$

$$= \frac{1}{n}\left(\mathbb{E}[X] + \dots + \mathbb{E}[X]\right)$$

$$= \frac{1}{n} n \alpha = \alpha$$

$$\mathbb{E}[\hat{\alpha}_n] = \mathbb{E}[\bar{X}] = \mathbb{E}[X] = \alpha$$

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



independence

$$\mathrm{Var}[\hat{\alpha}_n] = \mathrm{Var}\left[\tfrac{1}{n}(X_1 + \dots + X_n)\right] = \frac{1}{n^2}\left(\mathrm{Var}[X_1] + \dots + \mathrm{Var}[X_n]\right)$$

$$\mathrm{std}[Z] = \sqrt{\mathrm{Var}[Z]}$$

$$= \frac{1}{n^2}\left(\mathrm{Var}[X] + \dots + \mathrm{Var}[X]\right)$$

$$\mathrm{Var}\left[\tfrac{1}{n} Z\right] = \frac{1}{n^2}\mathrm{Var}[Z]$$

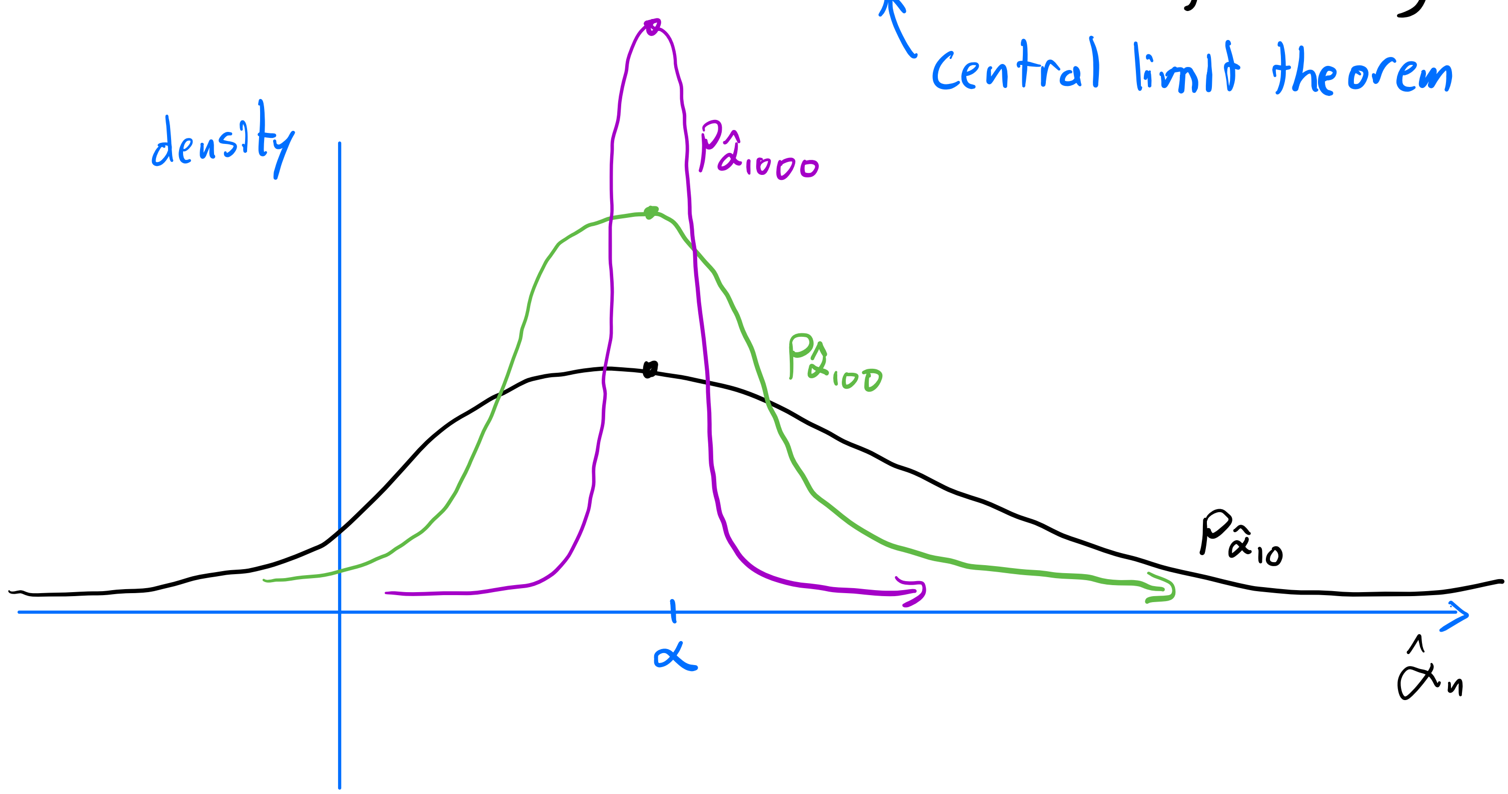
$$= \frac{1}{n^2} n \,\mathrm{Var}[X] = \frac{1}{n}\mathrm{Var}[X]$$

$$\mathrm{Var}[Z] = \mathbb{E}\left[\left(\tfrac{1}{n}Z - \mathbb{E}[Z]\right)^2\right] = \mathbb{E}\left[\tfrac{1}{n^2} Z^2 - \dots + \tfrac{1}{n^2}\mathbb{E}[Z]^2\right]$$

as n increases the variance of $\bar{X}$ decreases

$$\mathbb{P}_{\hat{\alpha}_n} \approx \mathcal{N}(\mathbb{E}[\hat{\alpha}_n], Var[\hat{\alpha}_n])$$

central limit theorem



Estimating the loss function

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

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

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

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

We can estimate $L(f)$ with n feature-label pairs

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

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

$$\hat{L}(f) = \frac{\ell(f(\vec{X}_1), Y_1) + \ldots + \ell(f(\vec{X}_n), Y_n)}{n}$$

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

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

for a predictor f the variance of $\hat{L}(f)$ decreases as n increases

We need an estimate of $L(f)$ for all $f \in \mathcal{F}$

Is it okay to use $\hat{L}(f)$ based on the same D to estimate $L(f)$ for all $f \in \mathcal{F}$?

Ans: It depends... on how large n is

and how large $\mathcal{F}$

But we will discuss

For now we use $\hat{L}(f)$ for all $f \in \mathcal{F}$ and see what happens