Inference (learning)

Given a sample $X_1, X_2, \ldots, X_n \sim F$,

how do we infer F?

Def
Statistical model = set of distributions
(or densities, regression fxns, etc)

Def
parametric model = set of distributions
parametrized by a finite # of parameters

Ex.

$$\mathcal{F} = \left\{ f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\pi\sigma} \right) : \mu \in \mathbb{R}, \sigma^2 > 0 \right\}$$

In general, we write

parameter space

$$\mathcal{F} = \{ f_\theta(x) : \theta \in \Theta \}$$

$$\mathbb{P}_\theta[A] := \int_A f_\theta(x)\,dx \qquad \mathbb{E}_\theta[r(X)] = \int r(x) f_\theta(x)\,dx$$

Def
non-parametric model = model that's not parametric

Ex. $X_1, X_2, \ldots, X_n$ correspond to flips of a coin (i.e. Bernoulli RV). The goal is to estimate parameter p = prob. of heads

Ex. $X_1, X_2, \ldots, X_n \sim F$, Goal is to approximate F

Ex. $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ pairs of data

Goal is to approximate $r(x) = \mathbb{E}[Y \mid X = x]$

X_i = "features" (image of pixels)

Y_i = "output" (age)

May assume $r(x)$ has the form of a neural network n_θ

$\left(n_\theta = \sigma(Ax + b), \quad \theta = [A, b] \right)$

Point estimation

point estimate = "best guess" of some quantity of interest

- θ : unknown <u>fixed</u> parameter
- $\hat{\theta}_n$: point estimate of θ from data $X_1, \dots, X_n$
    - this is a <u>random variable</u>

Assumption
Samples $X_1, X_2, \dots, X_n$ drawn iid from some $f_\theta(x)$ where $\theta \in \Theta$

Def
We say $\hat{\theta}_n$ is <u>unbiased</u> if $\mathbb{E}[\hat{\theta}_n] = \theta$

Def
We say $\hat{\theta}_n$ is <u>consistent</u> if $\hat{\theta}_n \xrightarrow{P} \theta$

Ex. $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

- $\mathbb{E}[S_n^2] = \sigma^2 \Rightarrow S_n^2$ is unbiased
- $\mathbb{V}[S_n^2] \to 0 \Rightarrow S_n^2$ is consistent

Def The <u>standard error</u> of a point estimator $\hat{\theta}_n$ for θ is

$$se(\hat{\theta}_n) = \sqrt{\mathbb{V}[\hat{\theta}_n]}.$$

Def The <u>mean squared error</u> of a point estimate $\hat{\theta}_n$ for θ is

$$MSE = \mathbb{E}_\theta\left[(\hat{\theta}_n - \theta)^2\right]$$

$\uparrow$

expectation w.r.t. distribution $f_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$

<u>Ex.</u> $X_1, \ldots, X_n \sim Ber(p) = \begin{cases} 1 & w.p.\ p \\ 0 & o.w. \end{cases}$

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- is $\hat{p}_n$ unbiased? ✓

$$\mathbb{E}[\hat{p}_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n}\sum_i p = p$$

- what is $se(\hat{p}_n)$?

$$\sqrt{\mathbb{V}[\hat{p}_n]} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{V}[X_i]} = \sqrt{\frac{1}{n}(1-p)(p)}$$

Theorem

$$\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] = \underbrace{\mathbb{E}[\hat{\theta}_n - \theta]^2}_{(\text{bias})^2} + \mathbb{E}[(\hat{\theta}_n - \mathbb{E}\hat{\theta}_n)^2]$$

$(\text{bias})^2 + (\text{variance})$

Theorem

$$\text{bias} \to 0, \ se \to 0 \Rightarrow \text{consistent}$$

Def

An estimator is asymptotically normal if

$$\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{d} N(0,1)$$

Ex $\hat{p}_n$ from previous example (CLT)

Confidence Sets

Def. A $1-\alpha$ confidence interval for a parameter θ is an interval $C_n = (a_n, b_n)$, where a_n, b_n are functions of $X_1, \dots, X_n$ st.

$$\mathbb{P}_\theta[\theta \in C_n] \geq 1 - \alpha \quad \forall \theta \in \Theta$$

C_n = random

θ = fixed