<u>Nov 1, 2023</u>

Random Variables $\{Y\}_{i=1}^{n}$ — Realized value of R.V.

Numbers $\{y_i\}_{i=1}^{n}$ (No randomness)

$Z \rightarrow$ to create randomness, generate a new random var.

$$P(Z = y_i) = \frac{1}{n}$$

what is this?

Risk $R(\theta) = E[\ell(Y,\theta)]$ — what is a loss f^n

Empirical Risk $\hat{R}(\theta) = E_n[\ell(Z,\theta)] = \frac{1}{n}\sum_{i=1}^{n} \ell(y_i,\theta)$

$\theta \in \Theta \subset \mathbb{R}^p$ — Capital Theta

$$\theta_0 = \arg\min R(\theta) \qquad \hat{\theta} = \underset{\theta \in \Theta}{\arg\min}\, \hat{R}(\theta)$$

$$\hat{R}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(y_i, \theta)$$

Uncertainty in $\hat{R}(\theta)$: How can we evaluate how far $R(\theta)$ is from $\hat{R}(\theta)$. Let's look at the data collection process $\rightarrow$ reverse engineer the r.v. from which the data was collected. This reverse engineering is imagined. You can't actually do it.

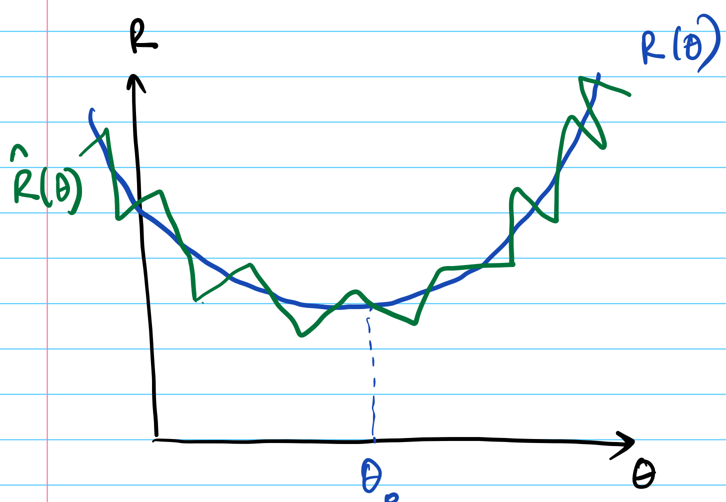
$$\hat{R}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(Y_i,\theta) \longrightarrow \text{R.V.}$$

By law of large number;

$$\hat{R}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(Y_i,\theta) \xrightarrow{P} E[\ell(Y,\theta)] = R(\theta)$$

(*) Theory of empirical process

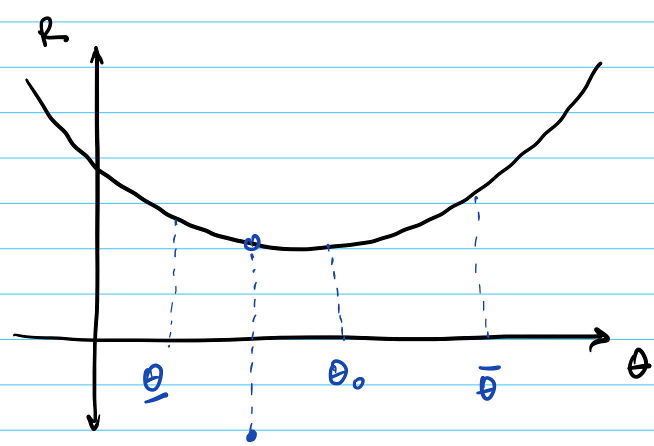
$\theta \to$ single dimensional parameter



Let $\hat{R}^*(\theta)$ be close to $R(\theta)$

$\hat{R}(\theta) = R(\theta) - 1000 \cdot 1\{\theta^*\}$; $\theta^* \sim U[\Theta]$ (spoiled version)

$\underset{\theta \in \Theta}{\text{Argmin}} \ \hat{R}(\theta) = \theta^* \sim U[\Theta]$



$P(\hat{R}(\theta) = R(\theta)) = 1$

Uniform Convergence

Def: $\hat{R}(\cdot)$ converges to $R(\cdot)$ uniformly in probability if

$$X_n = \sup_{\theta \in \Theta} |\hat{R}_n(\theta) - R(\theta)| \xrightarrow{P} 0$$

(How supremum behaves as $n \to \infty$)

for every n.

L_∞ — L infinity norm : $\sup_{x \in R} |f(x)|$

sample size.

(Converge uniformly in probability → Converge in probability in the worst case)

Def: $\hat{R}(\cdot)$ converges to $R(\cdot)$ uniformly a.s: If

$$X_n = \sup_{\theta \in \Theta} |\hat{R}_n(\theta) - R(\theta)| \xrightarrow{a.s} 0$$

(*) We are ultimately interested in proximity of $\hat{\theta}$ to θ_0

What are the conds. when $\hat{\theta} \xrightarrow{P} \theta_0$?

Theorem :— Suppose following conditions are satisfied :—

1. Θ is a compact set.
2. $R(\cdot)$ is continuous
3. $\hat{R}(\cdot)$ is converging to R uniformly in probability
4. (Identification) $R(\cdot)$ attains a unique global min at θ_0.

pop^n risk

Then, $\hat{\theta} \xrightarrow{P} \theta_0$ [converge of sequence of r.v. to θ_0 as $n \to \infty$]

LOSS FUNCTION :—

empirical

$$\ell = \frac{1}{2}(Y_i - \theta)^2$$

Parameter of the dist^n. solve (say your guess of pop^n mean) using theory

$$\Rightarrow \hat{R}(\theta) = \frac{1}{2n}\sum_{i=1}^{n}(Y_i - \theta)^2$$

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}}\ \hat{R}(\theta) = \frac{1}{n}\sum_{i=1}^{n} Y_i ,$$

How close your estimated θ is to actual θ.