STAT 5010: Advanced Statistical Inference

Lecturer: Tony Sit
Scribe: XU Weiwei

Lecture #1
07 September 2020

$$\left.\begin{array}{l}\text{Different disciplines:}\\ \text{- Networks, images, recordings, ...}\\ \text{- Economics, Finance, ...}\\ \text{- Biology (Genetic), medical, earth science, ...}\end{array}\right\} \to \text{Data} \xrightarrow{\text{Statistics}} \text{Inference/Conclusions}$$

Question Concerned:

- Modelling How to build a mathematical/statistical model that captures the *uncertainty* in our data?
- Methodology tools that allows us to deduce these statistical conclusions.
- Analysis: Optimal inference

  e.g. Mode: sample mean, sample median.

  * Finite sample optionality
  1. Given observations $(x_1, x_2, ..., x_n)$.
  2. Asymptotic properties $(n \to \infty)$.

1 Decision Theory (Wald, 1939)

Random element X takes values in a sample space χ. X can also be a vector (or matrix).

$$(X_1, X_2, ..., X_n) \overset{i.i.d}{\sim} f \text{ (distribution)}$$

I
- A statistical model is a family of distribution $\mathbb{P}$ indexed by a parameter θ. we denote

$$\mathbb{P} = \{P_\theta : \Theta \in \Omega\},$$

  where θ is the parameter, $\Omega \in \mathbb{R}^k$ is the parameter space and P_θ is a distribution.
- We assume that the data X come from some $P_\theta \in \mathbb{P}$ but the true value of θ is unknown.

Example 1 *Observe a sequence of coin flips $x_1, x_2, ..., x_n \in \{0, 1\}$. The objective is to estimate the probability of heads given the observations. (with 1 denotes a head). One can write*

$$\mathbb{P} = \left\{\textit{Bernoulli}(\theta) : \theta \in [0,1] \triangleq \Omega\right\}$$

$$P_\theta(X_i = 1) = \theta.$$

Estimating Procedure:

$$\left.\begin{array}{r}\text{Estimator} \xrightarrow{\text{Observations}} \text{Estimates} \\ +\text{Testing}\end{array}\right\} \to \text{Inference}$$

II A Decision Procedure

δ (estimator) is a map from χ to the decision space $\mathbb{D}$.

Example: Take $\mathbb{P} = \{\text{Bernoulli}(\theta)\}$ as shown above, we may be interested in estimating θ or testing θ based on:

(a) Estimating θ

The decision space is $\mathbb{D} = [0,1]$ and the decision procedure might be

$$\delta(X) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n.$$

(b) Hypothesis Testing

Accepting/rejecting the hypothesis that $\theta = 0.5$; The corresponding decision space is $\mathbb{D} = \{\text{accept}, \text{reject}\}$, one possible decision procedure is

$$\delta(X) = \text{“Reject if } \bar{X}_n > 0.5'' \text{ and accept otherwise.}$$

(c) A loss function is a mapping $L : \Omega \times \mathbb{D} \to \mathbb{R}^+$, $L(\theta, d)$ represents the penalty for making the decision d when θ is in fact the true parameter for the distribution generating the data.

Example 2 *: [Squared-error Loss] For estimating θ with decision $d \in \mathbb{R} = \mathbb{D}$, a common loss function is the squared-error loss:* $L(\theta, d) = (\theta - d)^2$.

Risk Function:

$$\text{Average loss incurred} \to \; R(\theta, \delta) = E_\theta(L(\theta, \delta(X)))$$

Admissiability:

δ is inadmissible if there exists δ' such that $R(\theta, \delta') \leq R(\theta, \delta)$ for all θ and $R(\theta', \delta') < R(\theta', \delta)$ for some θ'.