Lecture 1

Estimation theory.

1.1 Introduction

Let us consider a set $\mathcal{X}$ (probability space) which is the set of possible values that some random variables (random object) may take. Usually X will be a subset of $\mathbb{R}$, for example $\{0, 1\}$, $[0, 1]$, $[0, \infty)$, $\mathbb{R}$, etc.

I. Parametric Statistics.

We will start by considering a family of distributions on $\mathcal{X}$:

- $\{\mathbb{P}_\theta, \theta \in \Theta\}$, indexed by parameter θ. Here, Θ is a set of possible parameters and probability $\mathbb{P}_\theta$ describes chances of observing values from subset of X, i.e. for $A \subseteq X$, $\mathbb{P}_\theta(A)$ is a probability to observe a value from A.
- Typical ways to describe a distribution:
  - probability density function (p.d.f.),
  - probability function (p.f.),
  - cumulative distribution function (c.d.f.).

For example, if we denote by $N(\alpha, \sigma^2)$ a normal distribution with mean α and variance σ^2, then $\theta = (\alpha, \sigma^2)$ is a parameter for this family and $\Theta = \mathbb{R} \times [0, \infty)$.

Next we will assume that we are given $X = (X_1, \cdots, X_n)$ - independent identically distributed (i.i.d.) random variables on $\mathcal{X}$, drawn according to some distribution $\mathbb{P}_{\theta_0}$ from the above family, for some $\theta_0 \in \Theta$, and suppose that θ_0 is unknown. In this setting we will study the following questions.

1. **Estimation Theory.**

   Based on the observations $X_1, \cdots, X_n$ we would like to estimate unknown parameter θ_0, i.e. find $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$ such that $\hat{\theta}$ approximates θ_0. In this case we also want to understand how well $\hat{\theta}$ approximates θ_0.

2. **Hypothesis Testing.**

   Decide which of the hypotheses about θ_0 are likely or unlikely. Typical hypotheses:

   - $\theta_0 = \theta_1$? for some particular θ_n?
   - $\theta_0 \gtrless \theta_1$
   - $\theta_0 \neq \theta_1$

Example: In a simple yes/no vote (or two candidate vote) our variable (vote) can take two values, i.e. we can take the space $\mathcal{X} = \{0, 1\}$. Then the distribution is described by

$$\mathbb{P}(1) = p, \;\; \mathbb{P}(0) = 1 - p$$

for some parameter $p \in \Theta = [0, 1]$. The true parameter p_0 is unknown. If we conduct a poll by picking n people randomly and if $X_1, \cdots, X_n$ are their votes then:

1.Estimation theory. What is a natural estimate of p_0?

$$\hat{p} = \frac{\#(1's \text{ among } X_1, \cdots, X_n)}{n} \sim p_0$$

How close is $\hat{p}$ to p_0?

2. Hypothesis testing. How likely or unlikely are the following:

- Hypothesis 1: $p_0 > \frac{1}{2}$
- Hypothesis 2: $p_0 < \frac{1}{2}$

II. Non-parametric Statistics

In the second part of the class the questions that we will study will be somewhat different. We will still assume that the observations $X = (X_1, \cdots, X_n)$ have unknown distribution $\mathbb{P}$, but we won't assume that $\mathbb{P}$ comes from a certain parametric family $\{\mathbb{P}_\theta, \theta \in \Theta\}$. Examples of questions that may be asked in this case are the following:

- Does $\mathbb{P}$ come from some parametric family $\{\mathbb{P}_\theta, \theta \in \Theta\}$?
- Is $\mathbb{P} = \mathbb{P}_0$ for some specific $\mathbb{P}_0$?

If we have another sample $X' = (X'_1, \cdots, X'_m)$ then,

- Do X and X' have the same distribution?

If we have paired observations $(X_1, Y_1), \cdots, (X_n, Y_n)$:

- Are X and Y independent of each other?
- Classification/regression problem: predict Y as a function of X; i.e.,

$$Y = f(X) + \text{ small error term }.$$