Topic 18: Parameter Estimation Basics

02-680: Essentials of Mathematics and Statistics

November 15, 2024

What is a "Statistic"?

> Definition: Anything that can be computed from the collected data
> (i.e., must be observable).

Statistics deals with data.

Goal: Make inferences based on data

This process can be divided into three overlapping phases

(1) **Collecting data** — collect data in experiments; Preceded by forming hypotheses about phenomena of interest

(2) **Describing data** — describe the results

(3) **Analyzing data** — infer from the results the strength of the evidence with respect to the hypotheses

Generally two types of statistics

> **Point statistic** — a single value computed from data (e.g., $\overline{X_n}$)

> **Interval or range statistics** — an interval $a \leq x \leq b$ computed from the data

Note that a statistic is itself a random variable because a new experiment will produce new data to compute it.

1 Some Basics

Consider a dataset $X_1, X_2, \cdots, X_n$ of i.i.d. (independent and identically distributed) random variables from the same unknown distribution. That is, for each X_is, the underlying distributions have the same μ and σ.

Let $\overline{X_n}$ be the average of the actual value of the observations:

$$\overline{X_n} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Note this is different from the expected value of the underlying distribution μ. Note also that $\overline{X_n}$ is also a random variable in and of itself.

Example.   Consider a dataset $C_1, C_2, \cdots, C_{100}$ of 100 fair coin flips (chosen iid), with the corresponding random variables taking on values of 1 for Heads and 0 for Tails. Assume 48 of the 100 coins are heads, then $\overline{C_{100}} = \frac{48}{100}$.

Law of Large Numbers.   The *law of large numbers* says that as n grows, the probability that $\overline{X_n}$ is close to μ approaches 1.

$$\lim_{n\to\infty} p\left(\overline{X_n} = \mu\right) \mapsto 1.$$

Central Limit Theorem (brief introduction).   While similar to the law of large numbers the **Central Limit Theorem** says that n grows, the distribution of $\overline{X_n}$ converges to the Gaussian (normal) distribution $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

$$\lim_{n\to\infty} \overline{X_n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

These both say something similar and speak to the fact that more data will make the empirical mean will approach the underlying (theoretical) mean. But the Central Limit Theorem is more precise, it allows us to approximate the probability that the sample size is not adequate (gives some confidence on the probability from the LLN). We will see much more about this over the next couple weeks.