Statistical Inference and Multivariate Analysis (MA324)

Lecture Slides

Lecture 08

Point Estimation

Indian Institute of Technology Guwahati

Jan-May 2023

The logic of Probability and Statistics:

The words probability and statistics are often linked together in books, but they are different.

- Probability: starts with the general case (the population or model) and then predict what would happen in many samples.

$$\text{Genreal} \rightarrow \text{Specific}$$

$$\text{Population} \rightarrow \text{Sample}$$

$$\text{Model} \rightarrow \text{Data}$$

- Statistics: works in opposite direction; Start with one set of data (the sample) and make inference about the overall population or model.

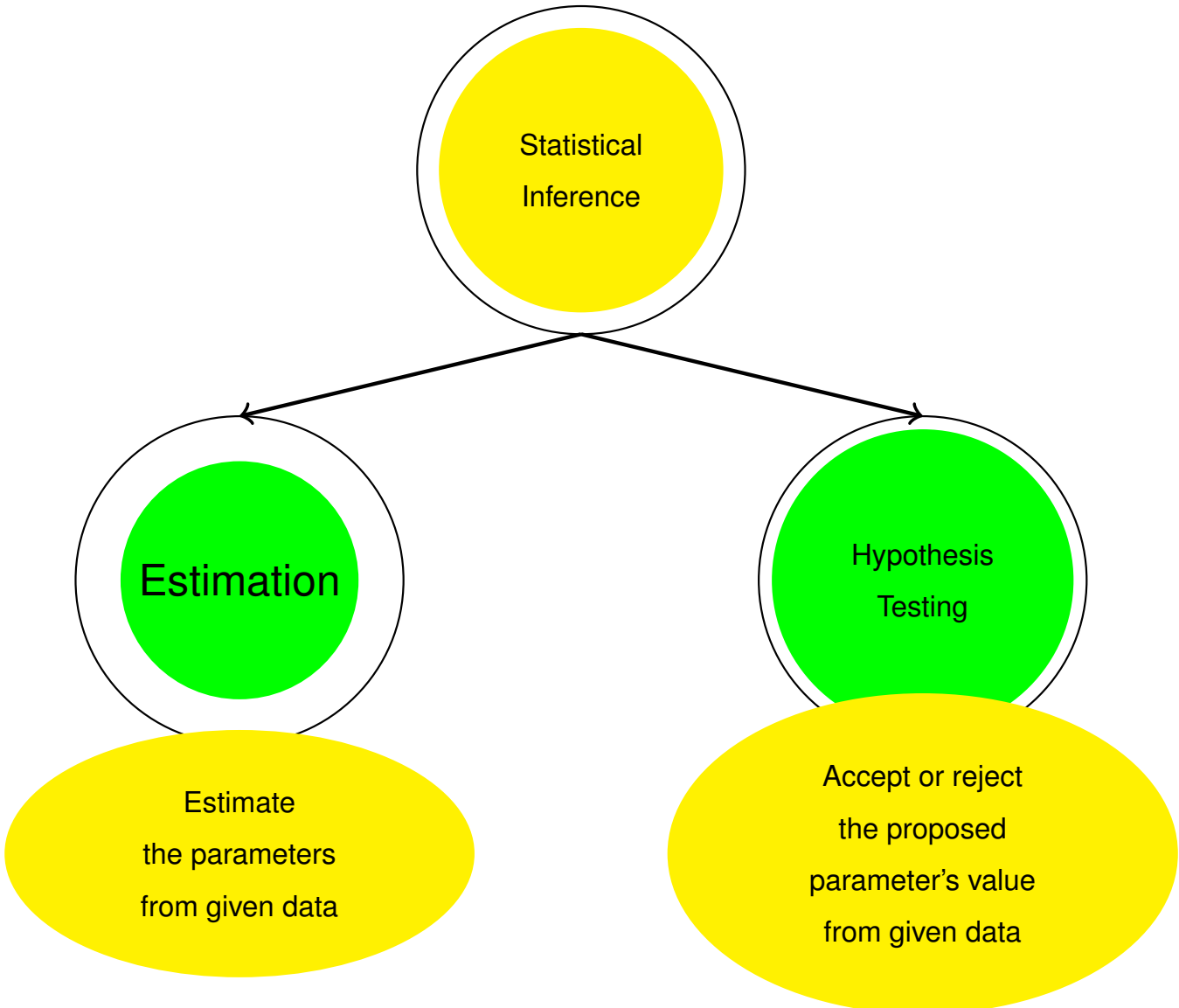
$$\text{Genreal} \leftarrow \text{Specific}$$

$$\text{Population} \leftarrow \text{Sample}$$

$$\text{Model} \leftarrow \text{Data}$$

Ref: Intuitive Biostatistics: Harvey Motulsky

Statistical Inference:



Point and Interval estimation

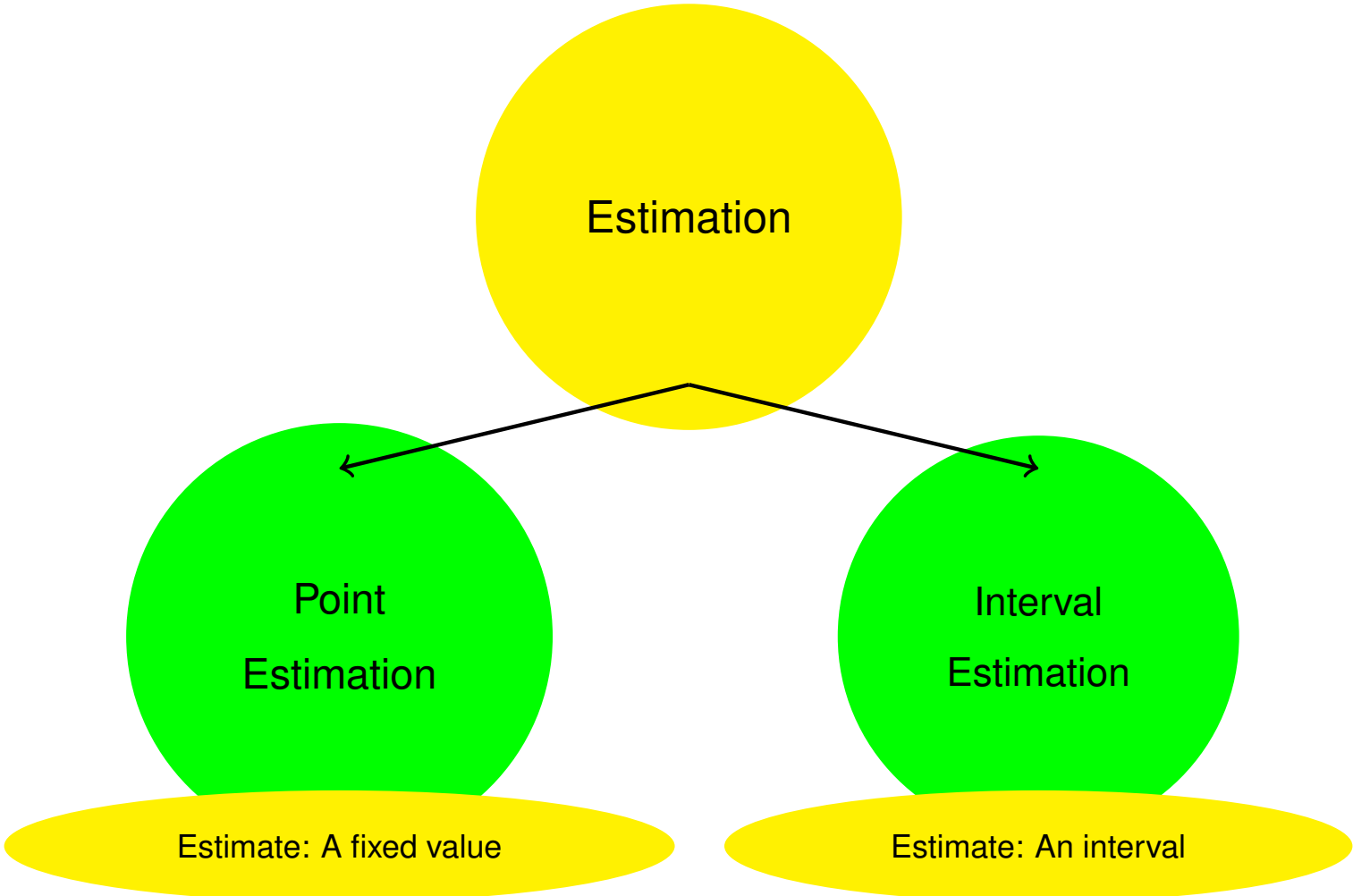
Why Estimation?

Questions:

- What is the **average height** of IITG M&C students?
- Assume that: height **follows a normal** distribution with mean μ and variance σ^2: $N(\mu, \sigma^2)$
- What is the value of μ and σ?
- To know the average height of IITG M&C students, you need to estimate μ from a sample (random or non-random?) obtained from the **source** population (all IITG M&C students).
- The last step is called estimation.
- Intuitively: Estimation is a way of **knowing** the unknown model (that you assumed) **parameters** from a sample.

Why Estimation?

- Want to measure the **amount of time** it takes for a certain cellular protein to degrade.
- How do you construct a probability model that does a good job of describing degradation for this particular protein
- Let X is a time measurement, we are likely to consider probability models that take into account the non-negative real numbers.
- **Gamma** distribution and the **exponential** distribution (which is just a special case of the gamma). Both are actually very common distributions to use to model waiting times (in this case, **waiting time** until the protein degrades).
- Let sample: $X_1, \cdots, X_n \sim$ Exponential(λ): need to **estimate** λ.
- Note that: **assumption** of a right parametric model is the key; if you choose a wrong model, it will give you **wrong information**!


Estimation
Point
Estimation
Interval
Estimation
Estimate: A fixed value
Estimate: An interval

1. X has a CDF F with known functional form except perhaps some parameters. Here our aim is to (educated) guess value of the parameters.

   For example, in some case we may have $X \sim N(\mu, \sigma^2)$, where the functional form of the PDF is known, but the parameters μ and/or σ^2 may be unknown. In this case, we need to find value of the unknown parameters based on a sample.

   This is known as ***parametric inference***. In this course, we will mainly consider parametric inference.

2. X has a CDF F who's functional form is unknown. This is known as ***nonparametric inference***. We will not discuss nonparametric inference in this course.