STAT 5430: Theory of Probability and Statistics II

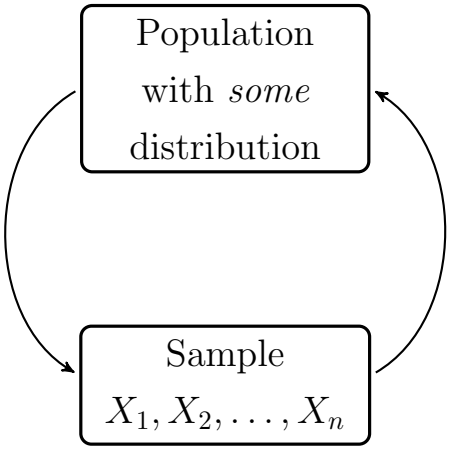
- **Probability** is a branch of mathematics concerned with the study of *random* phenomenon (e.g., experiments, models of populations)

- **Statistical inference** is the science of drawing inferences about populations based on only a part of the population (i.e., a sample)

- Developing statistical inference with probability is the topic of STAT 5430.

  Learning the probability notions useful for statistical inference is topic of STAT 5420.

- Little picture here



Introduction to Statistical Inference

Notation & Definitions

Some Notational Conventions

- r.v. $\equiv$ random variable (vector)
- X, Y, W, Z $\quad \leftarrow$ denote r.v.'s
- x, y, w, z (lower case) $\quad \leftarrow$ observed values of r.v.'s
- cdf $\equiv$ cumulative distribution function $F(x) = \mathrm{P}(X \leq x)$
- pdf $\equiv$ probability density function $f(x)$
- pmf $\equiv$ probability mass function $f(x)$
- iid $\equiv$ independent and identically distributed

Definition: Let $X_1, X_2, \ldots, X_n$ be iid r.v.'s with common cdf $F(x)$ and pdf/pmf $f(x)$. Then, we say,

1. $X_1, \ldots, X_n$ is a **random sample** (r.s.), $F(x)$ is the population cdf and $f(x)$ is the population pmf/pdf;
2. $X_1, \ldots, X_n$ is a r.s. from $F(x)$ or from $f(x)$.

Introduction to Statistical Inference

Problem Statement

- Statistical inference is about *making statements about population distributions based on samples.*

- For a collection $\mathcal{F}$ of cdf's, let $F(x) \in \mathcal{F}$ be the underlying population cdf.

  Given $X_1, \ldots, X_n$, our objective is to draw inferences about $F(x)$.

- *Definition*: If $\mathcal{F} \equiv \{F(x|\theta) : \theta \in \Theta\}$, $\Theta \in \mathbb{R}^k$, $1 \leq k < \infty$, then the inference problem is called **parametric**; otherwise, it is nonparametric.

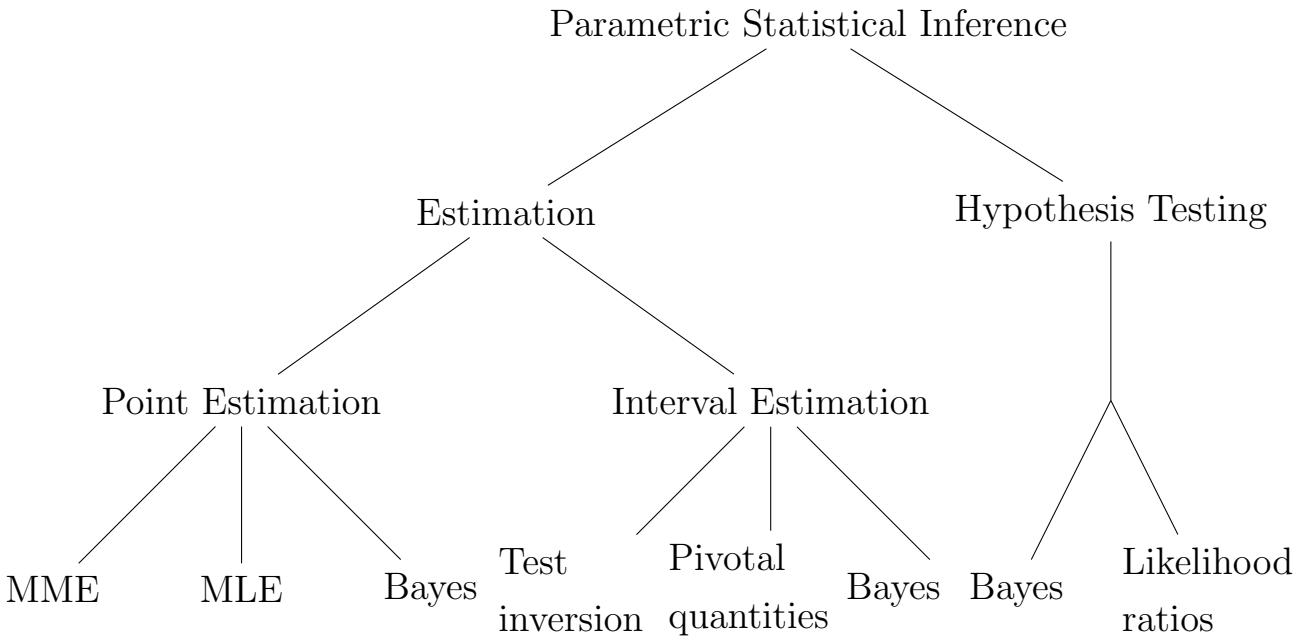
- Above θ is called the **parameter** and Θ is the **parameter space**.

Examples:

Introduction to Statistical Inference

High-level Overview of STAT 5430

- We focus on *parametric* statistical inference and develop the following inference topics:



- We will answer the following types of questions:

  1. What are some strategies for finding estimators or tests?

  2. What are "good" properties of an estimator or a test?

  3. What general statistical principles exist, if any, to guarantee that we can actually find estimators/tests with good properties?