# STAT 5430

## Lecture 01, W, Jan 22

- No new homework this week
  (assigned on Monday)
- Office hours to be announced

# STAT 5430: Theory of Probability and Statistics II

STAT 5420

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

  .

STAT 5430

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

- 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

connection between STAT 5420 & STAT 5430

$F(x) = P(X \leq x)$

$$\begin{tikzcd} \boxed{\text{Population with \emph{some} distribution}} \arrow[d, bend right, "\text{STAT 5420 ``probability''}"'] \\ \boxed{\text{Sample } X_1, X_2, \ldots, X_n} \arrow[u, bend right, "\text{STAT 5430}"'] \end{tikzcd}$$

what values can occur & how likely these are to occur

e.g. $P(X=3)$

$P(X_1 + X_2 > \sqrt{5})$

STAT 5430 stat. inference: to use sample to say something about population distribution $F(x)$)

# Introduction to Statistical Inference

Notation & Definitions

*Some Notational Conventions*

- r.v. $\equiv$ random variable (vector)
- $X, Y, W, Z$ $\leftarrow$ denote r.v.'s
- $x, y, w, z$ (lower case) $\leftarrow$ observed values of r.v.'s
- cdf $\equiv$ cumulative distribution function $F(x) = \mathrm{P}(X \leq x)$ $\leftarrow$ gives a probability of an event "$X \leq x$" for each real number $x \in \mathbb{R}$ (can vary $x$ to get different probabilities)
- pdf $\equiv$ probability density function $f(x)$
  continuous r.v.
- pmf $\equiv$ probability mass function $f(x)$
  discrete r.v.
- 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 $f(x)$.

   both equivalent for describing pop. distributions

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

- Above $\theta$ is called the **parameter** and $\Theta$ is the **parameter space**.

*Examples:*

normal

$\mathcal{F} = \{ N(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma > 0 \}$

$\rightarrow$ parametric inference problem

$\theta = (\mu, \sigma) \leftarrow$ parameters

$\Theta = \mathbb{R} \times (0, \infty) \leftarrow$ parameter space

$\mathcal{F} = \{ F(x) :$ $F(x)$ is continuous & symmetric around zero $\}$

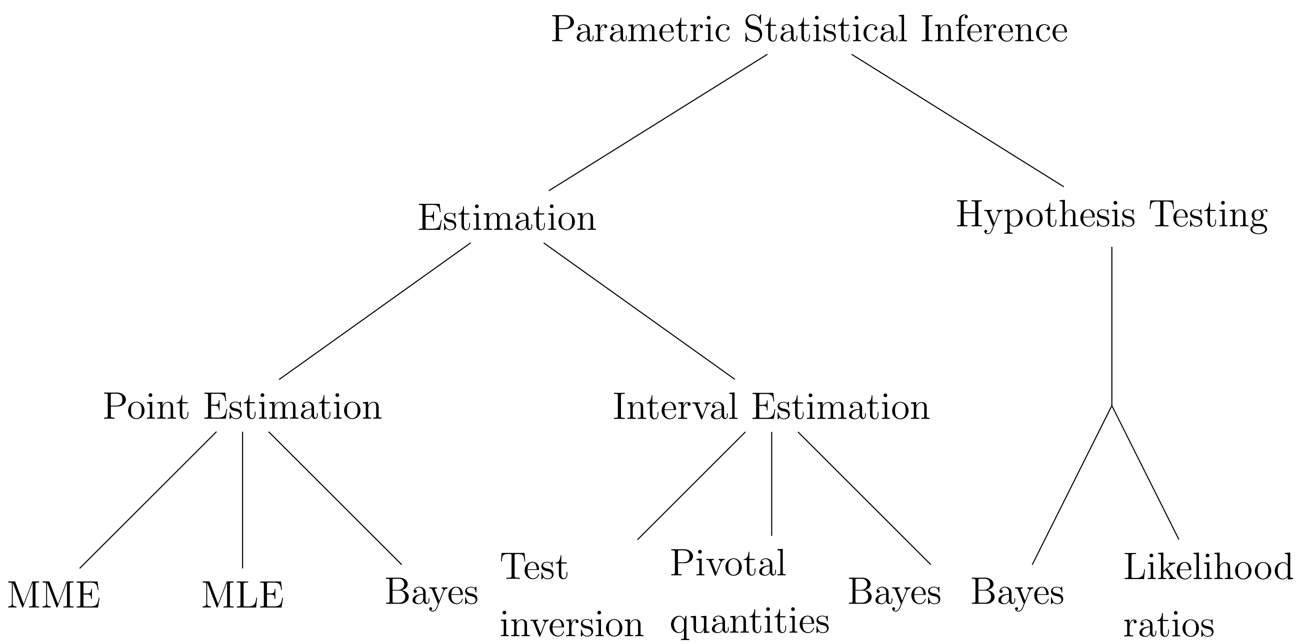
$\rightarrow$ nonparametric inference problem

( $F(x)$ is symmetric around 0 means $F(x) = 1 - F(-x)$ )

# 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?