The Statistical Set-Up.

Population. e.g., all the people in this class

Sample. e.g., a committee of 4 students

We assume that it's a **REPRESENTATIVE SAMPLE.**

Use the same word for the results of measurement (or polling) from the population w/ an unknown but COMMON dist'n.

Def'n. A random sample of size n from distribution D is a random vector

$$(Y_1, Y_2, \ldots, Y_n)$$

such that:

1. $Y_1, \ldots, Y_n$ are independent,
2. every Y_i has the distribution D.

Example. Consider 10 measurements: $Y_1, \ldots, Y_{10}$
Care was taken so that they are independent.
It's a standard model to assume that

Y_i are normally dist'd

w/ an unknown mean μ.

Scenario #1. We know the std deviation $\sigma = 0.01$

Then, $Y_i \sim N(\mu, \sigma = 0.01)$, $i = 1, \ldots, 10$ $\leftarrow \bar{Y}$? ✓

Scenario #2. We don't know the std deviation σ

Then, $Y_i \sim N(\mu, \sigma)$, $i = 1 \ldots 10$ $\leftarrow \bar{Y}$? ✓

$$\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)$$

Def'n. Any function of the random sample is called a **STATISTIC**.

A **POINT ESTIMATOR** is any function (rule, procedure) of the random sample which include only known constants w/ the purpose of estimating a model parameter.

An interval estimator is a pair of point estimators.

ESTIMATORS MUSTN'T CONTAIN UNKNOWN PARAMETERS IN THEIR EXPRESSION.

e.g.)

- $\dfrac{\#\text{ of people on the "committee" who like ice cream}}{\#\text{ of people on "committee"}}$,

  i.e., a sample proportion

- a sample mean : $\boxed{\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)}$

M378K Introduction to Mathematical Statistics

Problem Set #13

Order Statistics.

Problem 13.1. *An insurance company is handling claims from two categories of drivers: the good drivers and the bad drivers. The waiting time for the first claim from a* **good** *driver is modeled by an exponential random variable T_g with mean 6 (in years). The waiting time for the first claim from a* **bad** *driver is modeled by an exponential random variable T_b with mean 3 (in years). We assume that the random variables T_g and T_b are independent.*

What is the distribution of the waiting time T until the first claim occurs (regardless of the type of driver this claim was filed by)?

→:

$T = \min(T_g, T_b)$

$S_T = [0, \infty)$

$t > 0$: $F_T(t) = \mathbb{P}[T \le t]$

$= \mathbb{P}[\min(T_g, T_b) \le t]$

$= 1 - \mathbb{P}[\min(T_g, T_b) > t]$

$= 1 - \mathbb{P}[T_g > t, T_b > t]$ (independence)

$= 1 - \mathbb{P}[T_g > t] \cdot \mathbb{P}[T_b > t]$

$= 1 - e^{-\frac{t}{\tau_g}} \cdot e^{-\frac{t}{\tau_b}}$

$= 1 - e^{-t\left(\frac{1}{\tau_g} + \frac{1}{\tau_b}\right)}$

$\tau = \dfrac{1}{\frac{1}{\tau_g} + \frac{1}{\tau_b}} = \dfrac{1}{\frac{1}{6} + \frac{1}{3}}$

$\tau = 2$

$T \sim E(\tau = 2)$

□